frontiers | Research Topics

Genetic and proteomic biomarkers in solid tumor detection and treatment

Edited by
Jitian Li, Apeng Chen, Ruowen Zhang, Changwei Li and Jian-Guo Zhou

Published in
Frontiers in Genetics



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-2545-6
DOI 10.3389/978-2-8325-2545-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genetic and proteomic biomarkers in solid tumor detection and treatment

Topic editors

Jitian Li — Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Apeng Chen — Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, China

Ruowen Zhang — Jiahehongsheng (Shenzhen) Health Industry Group, China

Changwei Li — Shanghai Jiao Tong University, China

Jian-Guo Zhou — University of Erlangen Nuremberg, Germany

Citation

Li, J., Chen, A., Zhang, R., Li, C., Zhou, J.-G., eds. (2023). *Genetic and proteomic biomarkers in solid tumor detection and treatment*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2545-6

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Table of contents

10 Editorial: Genetic and proteomic biomarkers in solid tumor detection and treatment
Jitian Li, Apeng Chen, Ruowen Zhang, Changwei Li and Jianguo Zhou

14 Identification and Validation of a Novel Pyroptosis-Related Gene Signature for Prognosis Prediction in Soft Tissue Sarcoma
Lin Qi, Ruiling Xu, Lu Wan, Xiaolei Ren, WenChao Zhang, Keming Zhang, Chao Tu and Zhihong Li

27 A Prognostic Model for Breast Cancer Based on Cancer Incidence-Related DNA Methylation Pattern
Zhenchong Xiong, Lin Yang, Juan Ao, Jiarong Yi, Xiazi Zouxu, Wenjing Zhong, Jikun Feng, Weiling Huang, Xi Wang and Zeyu Shuang

40 SERPINH1 is a Potential Prognostic Biomarker and Correlated With Immune Infiltration: A Pan-Cancer Analysis
Yu Wang, Weigang Gu, Weiwei Wen and Xiaofeng Zhang

55 Hsa-miR-557 Inhibits Osteosarcoma Growth Through Targeting KRAS
Zhi Qiao, Jinfeng Li, Hongwei Kou, Xiangrong Chen, Deming Bao, Guowei Shang, Songfeng Chen, Yanhui Ji, Tian Cheng, Yisheng Wang and Hongjian Liu

65 Global Characterization of Metabolic Genes Regulating Survival and Immune Infiltration in Osteosarcoma
Zhongpei Zhu, Min Zhang, Weidong Wang, Peng Zhang, Yuqiang Wang and Limin Wang

78 Case Report and Literature Review: Diagnosis, Tailored Genetic Counseling and Cancer Prevention for a Locally Advanced dMMR/MSI-H/TMB-H Lung Cancer Patient With Concurrent Lynch Syndrome Mediated by a Rare *PMS2* Splicing Variant (c.1144+1G>A)
Quanli Han, Si Liu, Zhi Cui, Qi Wang, Tonghui Ma, Liwen Jiang, Xiaomo Li and Guanghai Dai

86 The Role of RAB GTPases and Its Potential in Predicting Immunotherapy Response and Prognosis in Colorectal Cancer
Xuefei Jiang, Lanlan Yang, Qianling Gao, Yiting Liu, Xingzhi Feng, Shubiao Ye and Zihuan Yang

101 Hypoxia-Induced miR-378a-3p Inhibits Osteosarcoma Invasion and Epithelial-to-Mesenchymal Transition *via BYSL* Regulation
Junlei Zhang, Haijun Tang, Xiaohong Jiang, Nenggan Huang and Qingjun Wei

112 THBS2 is Closely Related to the Poor Prognosis and Immune Cell Infiltration of Gastric Cancer
Shiyu Zhang, Huiying Yang, Xuelian Xiang, Li Liu, Huali Huang and Guodu Tang

130 Effects of miR-103a-3p Targeted Regulation of TRIM66 Axis on Docetaxel Resistance and Glycolysis in Prostate Cancer Cells
Qiang Yi, Junfeng Wei and Yangzhou Li

142 A Novel Ras--Related Signature Improves Prognostic Capacity in Oesophageal Squamous Cell Carcinoma
Hao-Shuai Yang, Wei Liu, Shao-Yi Zheng, He-Yuan Cai, Hong-He Luo, Yan-Fen Feng and Yi-Yan Lei

152 A Comprehensive Pan-Cancer Analysis for Pituitary Tumor-Transforming Gene 1
Siming Gong, Changwu Wu, Yingjuan Duan, Juyu Tang and Panfeng Wu

165 KDF1 Promoted Proliferation and Metastasis of Epithelial Ovarian Cancer *via* Wnt/Beta-Catenin Pathway: TCGA-Based Data Mining and Experimental Validation
Changyu Zhu, Yilong Liu, Rongsheng Tong and Jianmei Guan

178 SNRPD1/E/F/G Serve as Potential Prognostic Biomarkers in Lung Adenocarcinoma
Gaohua Liu, Fuping Li, Meichun Chen, Yang Luo, Yinhai Dai and Peifeng Hou

192 ASF1B is a Promising Prognostic Biomarker and Correlates With Immunotherapy Efficacy in Hepatocellular Carcinoma
Shirong Zhang, Longwen Xu, Jinteng Feng, Deli Tan, Yue Zhu, Jia Hou, Wenyuan Li, Kejia Lv, Wenjuan Wang, Lili Jiang, Min Jiao and Hui Guo

210 One Ferroptosis-Related Gene-Pair Signature Serves as an Original Prognostic Biomarker in Lung Adenocarcinoma
Lei Li and Buhai Wang

219 Identification of lncRNA/circRNA-miRNA-mRNA ceRNA Network as Biomarkers for Hepatocellular Carcinoma
Shanshan Chen, Yongchao Zhang, Xiaoyan Ding and Wei Li

235 A Novel Signature Integrated of Immunoglobulin, Glycosylation and Anti-Viral Genes to Predict Prognosis for Breast Cancer
Shengshan Xu, Yuchen Liu, Hansu Ma, Shuo Fang, Shoupeng Wei, Xiaoping Li, Zhuming Lu, Youbin Zheng, Tong Liu, Xiaojian Zhu, Dongming Xu and Yihang Pan

250 Identification and Validation of Chromobox Family Members as Potential Prognostic Biomarkers and Therapeutic Targets for Human Esophageal Cancer
Xuefen Fang, Junjun Wang, Jiabing Chen, Mingkai Zhuang, Tingxuan Huang, Zhixin Chen, Yuehong Huang, Biyun Zheng and Xiaozhong Wang

268 Prognostic and Diagnostic Values of Semaphorin 5B and Its Correlation With Tumor-Infiltrating Immune Cells in Kidney Renal Clear-Cell Carcinoma
Junping Ding, Shubin Zhao, Xianhua Chen, Changjun Luo, Jinjian Peng, Jiantan Zhu, Yongqi Shen, Zhou Luo and Jianlin Chen

286 Identification of a Gene Signature to Aid Treatment Decisions by Integrated Analysis of Mutated Genes Between Primary and Metastatic Prostate Cancer
Qinyu Li, Xueyan Xiao, Bingliang Chen, Guoda Song, Kai Zeng and Jianping Miao

299 Construction of a Necroptosis-Related miRNA Signature for Predicting the Prognosis of Patients With Hepatocellular Carcinoma
Tongyu Meng, Qingfeng Wang, Yufeng Yang, Yanling Ren and Yan Shi

310 Evaluating a Panel of Autoantibodies Against Tumor-Associated Antigens in Human Osteosarcoma
Manli Luo, Songmei Wu, Yan Ma, Hong Liang, Yage Luo, Wentao Gu, Lijuan Fan, Yang Hao, Haiting Li and Linbo Xing

318 Molecular Portrait of GISTs Associated With Clinicopathological Features: A Retrospective Study With Molecular Analysis by a Custom 9-Gene Targeted Next-Generation Sequencing Panel
Haoran Qian, Na Yan, Xiaotong Hu, Junchang Jiang, Zhengzheng Cao and Dan Shen

330 Anti-POSTN and Anti-TIMP1 Autoantibodies as Diagnostic Markers in Esophageal Squamous Cell Carcinoma
Weihong Xie, Guiying Sun, Jicun Zhu, Huimin Wang, Zhuo Han and Peng Wang

341 Altered Expressions of NF1 and NF1-Related microRNAs as Biomarkers in the Diagnosis of Undifferentiated Pleomorphic Sarcoma
Peng Zhang, Lingling Huang, Pengwei Ma and Xiaoying Niu

352 Pan-Cancer Analyses Reveal Oncogenic and Immunological Role of PLOD2
Qiqi Xu, Na Kong, Yiguo Zhao, Quan Wu, Xin Wang, Xiaodong Xun and Pengji Gao

366 DEAD-Box Helicase 27 Triggers Epithelial to Mesenchymal Transition by Regulating Alternative Splicing of Lipoma-Preferred Partner in Gastric Cancer Metastasis
Yirong Jin, Suzhen Yang, Xiaoliang Gao, Di Chen, Tingting Luo, Song Su, Yanting Shi, Gang Yang, Lei Dong and Jie Liang

379 Transmembrane Protein 170B is a Prognostic Biomarker and Associated With Immune Infiltrates in Pancreatic Adenocarcinoma
Zilong Zhang, Jin Shang, Zonglin Dai, Yutong Yao, Ying Shi, Deyuan Zhong, Yuxin Liang, Chunyou Lai, Qinyan Yang, Tianhang Feng and Xiaolun Huang

395 Prognostic Value of *SPOCD1* in Esophageal Squamous Cell Carcinoma: A Comprehensive Study Based on Bioinformatics and Validation
Zhizhong Lin, Lin Chen, Tingting Wu, Yiping Zhang, Xinyi Huang, Yuanmei Chen, Junqiang Chen and Yuanji Xu

412 The Landscape of the Tumor-Infiltrating Immune Cell and Prognostic Nomogram in Colorectal Cancer
Jiateng Zhong, Yu Qin, Pei Yu, Weiyue Xia, Baoru Gu, Xinlai Qian, Yuhan Hu, Wei Su and Zheying Zhang

424 Systems-Level Mapping of Cancer Testis Antigen 1b/a to Sarcoma Pathways Identifies Activated Ran Binding-2 E3 SUMO-Protein Ligase and Transducin-Like Enhancer Protein 1
Nikolaos A. Papanikolaou, Prodromos Hytiroglou, Pavlina Pantelidou, Athanasios G. Papavassiliou and Lloyd L. Old

437 Identification of Candidate Therapeutic Genes for More Precise Treatment of Esophageal Squamous Cell Carcinoma and Adenocarcinoma
Aneta Polewko-Klim, Sibo Zhu, Weicheng Wu, Yijing Xie, Ning Cai, Kexun Zhang, Zhen Zhu, Tao Qing, Ziyu Yuan, Kelin Xu, Tiejun Zhang, Ming Lu, Weimin Ye, Xingdong Chen, Chen Suo and Witold R. Rudnicki

448 NAP1L1 Functions as a Novel Prognostic Biomarker Associated With Macrophages and Promotes Tumor Progression by Influencing the Wnt/β-Catenin Pathway in Hepatocellular Carcinoma
Bingbing Shen, Wenjie Zhu, Xinyuan Liu and Jianxin Jiang

464 Biological and Clinicopathological Characteristics of OPN in Cervical Cancers
Shuhang Qin, Li Yi, Yanchun Liang, Yili Chen, Wei Wang, Yuandong Liao, Chunyu Zhang, Hua Huang, Jiaming Huang and Shuzhong Yao

476 *CNTN-1* Upregulation Induced by Low-Dose Cisplatin Promotes Malignant Progression of Lung Adenocarcinoma Cells *via* Activation of Epithelial-Mesenchymal Transition
Ruijie Zhang, Shengjin Li, Jian Lan, Changyi Li, Xianzhi Du, Weijie Dong, Qian Yu and Daoxin Wang

487 Strategies of LncRNA DLX6-AS1 on Study and Therapeutics
Yanyan Zhao and Pei Li

495 Clinical Benefits of Olaparib in Mexican Ovarian Cancer Patients With Founder Mutation *BRCA1*-Del ex9-12
Dolores Gallardo-Rincón, Edgar Montes-Servín, Gabriela Alamilla-García, Elizabeth Montes-Servín, Antonio Bahena-González, Lucely Cetina-Pérez, Flavia Morales Vásquez, Claudia Cano-Blanco, Jaime Coronel-Martínez, Ernesto González-Ibarra, Raquel Espinosa-Romero, Rosa María Alvarez-Gómez, Abraham Pedroza-Torres and Denisse Castro-Eguiluz

505 Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment
Fazhan Li, Huijuan Wen, Ihtisham Bukhari, Bin Liu, Chenxu Guo, FeiFei Ren, Youcai Tang, Yang Mi and Pengyuan Zheng

519 Identification and Validation of Necroptosis-Related LncRNA Signature in Hepatocellular Carcinoma for Prognosis Estimation and Microenvironment Status
Cong Chen, Yumeng Wu, Kang Chen, Zicong Xia, Xiaokan Liu, Chaojie Zhang, Hui Zhao and Aiguo Shen

537 Circular RNAs as Prognostic Biomarkers in Renal Cell Carcinoma: A Systematic Review and Meta-Analysis
Dan Liao, Qiu Lin, Huan Xiao, Fenglian Zhang and Qin Han

545 Systematic Construction and Validation of a Novel Macrophage Differentiation–Associated Prognostic Model for Clear Cell Renal Cell Carcinoma
Chen Liu, Xuhui Zhang, Caoyang Hu, Xuezhi Liang, Xiaoming Cao and Dongwen Wang

558 High Expression of FCRLB Predicts Poor Prognosis in Patients With Colorectal Cancer
Xiaopeng Wang, Ruirong Lin, Yi Zeng, Yi Wang, Shenghong Wei, Zhitao Lin, Shu Chen, Zaisheng Ye and Luchuan Chen

573 The m7G-Related Long Noncoding RNA Signature Predicts Prognosis and Indicates Tumour Immune Infiltration in Colon Cancer
Li Liu, Yukang Wu, Wenzheng Chen, Yebei Li, Jiahe Yu, Guoyang Zhang, Pengcheng Fu, Liu Huang, Jianbo Xiong and Zhigang Jie

589 An Autophagy-Related Gene Signature can Better Predict Prognosis and Resistance in Diffuse Large B-Cell Lymphoma
Xuan Zhou, Ying-Zhi He, Dan Liu, Chao-Ran Lin, Dan Liang, Rui Huang and Liang Wang

605 BTN3A2 Expression Is Connected With Favorable Prognosis and High Infiltrating Immune in Lung Adenocarcinoma
Yuansheng Lin, Hao Zhou and Shengjun Li

616 RPP30 is a novel diagnostic and prognostic biomarker for gastric cancer
Ying Kan, Xia Lu, Lijuan Feng, Xu Yang, Huan Ma, Jianhua Gong and Jigang Yang

626 Common Core Genes Play Vital Roles in Gastric Cancer With Different Stages
Zhiyuan Yu, Chen Liang, Huaiyu Tu, Shuzhong Qiu, Xiaoyu Dong, Yonghui Zhang, Chao Ma and Peiyu Li

639 An inflammation-related signature could predict the prognosis of patients with kidney renal clear cell carcinoma
Qingxin Yu, Facai Zhang, Dechao Feng, Dengxiong Li, Yuhui Xia and Mei-Fu Gan

653 Identification of an autophagy-related 12-lncRNA signature and evaluation of NFYC-AS1 as a pro-cancer factor in lung adenocarcinoma
Fang Tong, Lifa Xu, Sheng Xu and Mingming Zhang

672 Association of *FLG* mutation with tumor mutation load and clinical outcomes in patients with gastric cancer
Fu Yicheng, Liu Xin, Yu Tian and Liu Huilin

690 Integrated Analysis Reveals the Potential Significance of HDAC Family Genes in Lung Adenocarcinoma
Congkuan Song, Weichen Lin, Heng Meng, Ning Li and Qing Geng

704 Characterization of aging cancer-associated fibroblasts draws implications in prognosis and immunotherapy response in low-grade gliomas
Zijian Zhou, Jinhong Wei, Lijun Kuang, Ke Zhang, Yini Liu, Zhongming He, Luo Li and Bin Lu

725 NcRNA-mediated upregulation of CAMK2N1 is associated with poor prognosis and tumor immune infiltration of gastric cancer
Kaipeng Peng, Xiangqing Ren and Qian Ren

741 Exploration of differentially expressed mRNAs and miRNAs for pediatric acute myeloid leukemia
Qing Wang, Chao Yue, Qin Liu and Xuchun Che

750 A cancer stem cell associated gene signature for predicting overall survival of hepatocellular carcinoma
Xin-Yi Liang, Yue Zhang, Ya-Nan He, Xue-Yi Liu, Zhi-Hao Ding, Xiao-Dong Zhang, Ming-You Dong and Run-Lei Du

765 Transcriptome profiles of fatty acid metabolism-related genes and immune infiltrates identify hot tumors for immunotherapy in cutaneous melanoma
Yunxian Dong, Zirui Zhao, Maijimi Simayi, Chufen Chen, Zhongye Xu, Dongming Lv and Bing Tang

781 Comprehensive analysis of liquid-liquid phase separation-related genes in prediction of breast cancer prognosis
Huang Yu-Qing, Li Peng-Ping, Sun Ke, Yin Ke-Xing, Zhang Wei-Jun and Wang Zhen-Yu

796 Delta-catenin attenuates medulloblastoma cell invasion by targeting EMT pathway
Yuanjun Hu, Sihan Zhu, Rizhen Xu, Manxia Wang, Furong Chen, Zeshun Zhang, Binghong Feng, Jian Wang, Zhongping Chen and Jing Wang

808 Light at the end of the tunnel: Clinical features and therapeutic prospects of KRAS mutant subtypes in non-small-cell lung cancer
Liyuan Gao and Weizhang Shen

818 Identification of biomarkers for hepatocellular carcinoma based on single cell sequencing and machine learning algorithms
Weimin Li, Jixing Liu, Wenjuan Zhu, Xiaoxin Jin, Zhi Yang, Wenzhe Gao, Jichun Sun and Hongwei Zhu

828 High expression of B4GALT1 is associated with poor prognosis in acute myeloid leukemia
Zhihong Ren, Xiaoyu Huang, Qing Lv, Yiming Lei, Haiqiang Shi, Fanping Wang and Mingyong Wang

846 N6-methyladenosine related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer
Lili Yu, Lijuan Wang, Jing Sun, Xuan Zhou, Yeting Hu, Lidan Hu, Yazhou He, Chunqing Lin, Jie Chen, Xiaolin Xu, Malcolm G. Dunlop, Evropi Theodoratou, Kefeng Ding and Xue Li

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 11 May 2023
DOI 10.3389/fgene.2023.1200660

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Richard D. Emes,
Nottingham Trent University,
United Kingdom

*CORRESPONDENCE
Jitian Li,
jitianlee@hotmail.com

RECEIVED 05 April 2023
ACCEPTED 17 April 2023
PUBLISHED 11 May 2023

CITATION
Li J, Chen A, Zhang R, Li C and Zhou J (2023), Editorial: Genetic and proteomic biomarkers in solid tumor detection and treatment.
Front. Genet. 14:1200660.
doi: 10.3389/fgene.2023.1200660

COPYRIGHT
© 2023 Li, Chen, Zhang, Li and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Genetic and proteomic biomarkers in solid tumor detection and treatment

Jitian Li[1]*, Apeng Chen[2], Ruowen Zhang[3], Changwei Li[4] and Jianguo Zhou[5]

[1]Laboratory of Molecular Biology, Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), Zhengzhou, China, [2]Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, China, [3]Jiahehongsheng (Shenzhen) Health Industry Group, Shenzhen, China, [4]School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [5]University of Erlangen Nuremberg, Erlangen, Germany

KEYWORDS

solid tumor, genetic biomarker, proteomic biomarker, detection, treatment

Editorial on the Research Topic
Genetic and proteomic biomarkers in solid tumor detection and treatment

Introduction

Cancer is a major public health problem worldwide, which has become one of the leading causes of death, and it remains a great burden for the society (Siegel et al., CA Cancer J Clin, 2023, 73, 17–48). That is to say, it is urgent to search for suitable diagnostic, therapeutic, and prognostic methods such as biomarkers to affect the outcome of cancer patients as early as possible. The earlier the diagnosis and treatment, the better the recovery of the patient. An ideal cancer biomarker has high sensitivity and specificity and can reflect the status of the cancer, based on a molecular or process-based change. Biomarkers can help to identify cancer patients at an earlier stage and guide personalized therapy. In recent years, the advances in the genomics, transcriptomics, proteomics, and metabolomics have significantly improved our understanding of cancer biology. Based on these histological techniques, lots of biomarkers of cancer could be screened out and validated, such as nucleic acids, proteins, sugars, small metabolites, and entire tumor cells found in the body fluid, which could be used for diagnosis, treatment efficacy, prognosis, recurrence, and risk assessment (Wu, Chem Soc Rev, 2015, 44, 2963–2997). The evolutionary transformation in personalized cancer therapy and the prognosis of cancer have greatly improved. Recently, molecular biomarkers have played a unique role in cancer management and treatment.

Through research in the field of solid tumor diagnosis and treatment, this Research Topic covers a wide spectrum of the biomarkers from gene to protein. Research from laboratories, results from clinical trials, or reports of interesting cases are within the scope of this current Research Topic. The subtopics include:

1. Identification, characterization, and validation of genetic and proteomic biomarkers that are associated with solid tumor susceptibility, diagnosis, and therapeutics.
2. Novel techniques for identifying or capturing genetic or proteomic biomarkers in solid tumors.
3. Translational research bridging the gap between our incremental knowledge on the association of solid tumor biomarkers and characteristics and the outcome of cancer patients in clinical practice.
4. Case reports of rare cancer types or interesting cases in solid tumors associated with genetic or proteomic biomarkers in clinical practice.

The Research Topic we organized not only included the techniques to identify novel biomarkers and genetic and proteomic biomarkers that are associated with solid tumor susceptibility, diagnosis, and therapeutics, but it also included clinical translational and interesting case reports. Finally, this topic accepted 63 articles, including two brief research reports, one case report, one mini review, one review, one systemic review, and 57 original research articles covering a wide spectrum of studies.

In this Research Topic, many of the studies focused on the identifying, characterizing, and validating of corresponding genetic or proteomic biomarkers associated with tumor detection and therapy. Zhang et al. drew a conclusion that NF1 and NF1-related microRNAs, including hsa-miR-199a-3p and hsa-miR-34a-5p, may be novel biomarkers in diagnosing undifferentiated pleomorphic sarcoma (UPS). Kan et al. identified the hub gene in the tumorigenesis of gastric cancer (GC), obtaining a biomarker prediction model in which the high expression level of ribonuclease P protein subunit p30 (RPP30) was correlated with cancer progression and poor survival, and thus it was considered to be a novel diagnostic and prognostic biomarker. Through bioinformatic analysis, Chen et al. identified the lncRNA/circRNA-miRNA-mRNA ceRNA network as potential diagnostic and prognostic biomarkers in hepatocellular carcinoma (HCC). Luo et al. established a predictive model including three anti-tumor-associated antigen (ENO1, GAPDH, and TPI1) autoantibodies based on SERPA, which could be used as a promising and powerful tool in detecting osteosarcoma (OS). As for esophageal squamous cell carcinoma (ESCC), based on the Gene Expression Profiling Interactive Analysis (GEPIA) platform, Xie et al. concluded that the expression levels of anti-POSTN and anti-TIMP1 autoantibodies were higher in ESCC patients with areas under the ROC (AUC)s of 0.638 and 0.585, respectively, which could distinguish affected patients and non-affected ones, thus acting as potential biomarkers in diagnosing ESCC. Xu et al. showed that PLOD2 plays a significant role in tumorigenesis and maybe serves as a potential biomarker for diagnosis and prognosis through pan-cancer analyses. Early detection of tumors could improve survival, avoiding their progress and lethality as much as possible (Crosby et al., 2022). Sensitive and specific early detection technologies are essential in this research field.

As for the treatment biomarkers, most of the studies chose bioinformatics to screen the potential biomarkers at the first stage and then validated them with experiments, which is a commonly used and convenient research routine. Based on a comprehensive study of bioinformatics and experimental validation, Lin et al. indicated that SPOCD1 might be an independent prognostic factor for ESCC patients in the occurrence and development of the cancer. Li et al. obtained significantly differentially expressed mutant genes between primary and metastatic prostate cancer from the COSMIC database, which showed potential in aiding the treatment strategy in clinical practice. Through the TCGA database, Zhang et al. focused on THBS2, which is related to the poor prognosis and immune infiltration of GC. Wang et al. analyzed the expression of SERPINH1 in the Cancer Genome Atlas and the Genotype-Tissue Expression dataset, after which it was validated and correlated with immune infiltration as a potential prognostic biomarker in a pan-cancer analysis. Based on the bioinformatics analysis and an *in vitro* and *in vivo* study, Zhu et al. indicated that KDF1 plays an important role in ovarian cancer (OC) progression, which might be a therapeutic target for OC patients. Jiang et al. showed that RAB GTPases have an important function in regulating the cell cycle and immune-related pathways based on bioinformatics and functional science, demonstrating their potential as biomarkers in predicting prognosis and immunotherapy response in colorectal cancer (CRC).

A number of studies identified genetic biomarkers for therapy in various solid tumors, including gene mutations, DNA methylation, lncRNAs, miRNAs, and circRNAs. Gao and Shen discussed a common mutation in the KRAS gene, glycine 12 mutated to cysteine (G12C), which could be used in the treatment of non-small-cell lung cancer (NSCLC). Fu et al. explored the association of the *FLG* mutation with tumor mutation load and clinical outcomes, showing that the *FLG* gene mutation might be a protective factor, thus validating its usage as a novel therapeutic target and biomarker for stomach adenocarcinoma (STAD) treatment. As for copy number variations (CNVs), Li et al. identified and validated potential therapeutic and prognostic targets of GC by analyzing single-cell sequencing data, drawing a conclusion that *CPVL* could be a potential prognostic and therapeutic biomarker in GC. One novel pyroptosis-related gene signature consisting of five key DEGs in predicting the prognosis of soft tissue sarcoma was identified and validated, and Qi et al. suggested that it could be an independent prognostic factor and may be an important research direction in future research. Zhu et al. concluded that the regulation of potential candidate gene PAICS (phosphoribosyl aminoimidazole carboxylase, phosphoribosyl aminoimidazole succinocarboxamide synthetase), the metabolic-related gene, is associated with the development and metastasis of OS. Gong et al. discussed the function of PTTG1 in a pan-cancer analysis, which could act as a potential oncogene and associate with immune infiltration, immune checkpoints, tumor mutational burden, and microsatellite instability, indicating that PTTG1 could be a potential biomarker for both the prognosis and outcomes of tumor treatment as well as a promising target in tumor therapy. Polewko-Klim et al. identified four potential drug-targeted genes (ERBB3, ATP7B, ABCC3, and GALNT14) and five drug-related genes for the more precise treatment of esophageal squamous cell carcinoma and adenocarcinoma. The DNA methylation pattern is also a kind of biomarker. Based on 361 breast cancer (BC) incidence-related DNA methylation patterns, Xiong et al. developed a nomogram to quantify the survival probability of BC. A mini review by Zhao and Li concluded that lncRNA DLX6-AS1 could play crucial regulatory roles in various tumors, contributing to pre-clinical therapeutics.

Tong et al. constructed a lncRNA-based risk signature including 12 lncRNAs with important prognostic values in predicting lung adenocarcinoma (LUAD). Moreover, they found that NFYC-AS1 and BIRC6 may be potential therapeutic targets. Chen et al. discovered that three necroptosis-related high-risk lncRNAs had the prognostic value of HCC, helping to provide treatments for the patients. Liu et al. built a risk model with 21 signature m7G-related lncRNAs and evaluated the prognosis of colon cancer (CC) patients, showing that this panel could be of help to diagnosis and therapy in the future. Through their analysis of the GEO database, Qiao et al. showed that hsa-miR-557, a kind of miRNA, could inhibit OS growth by modulating the expression of KRAS both *in vivo* and *in vitro*, and it can thus be used as a therapeutic target. Furthermore, circRNAs can also be potential prognostic biomarkers of tumors, such as in renal cell carcinoma (RCC), as analyzed by Liao et al. The metabolism-related genetic biomarker also has the potential to be a therapy target. Dong et al. reported that transcriptome profiles of fatty acid metabolism-related genes are effective for distinguishing cutaneous melanoma (CM) into hot–cold tumors, providing valuable therapy strategies for the effective immunotherapy of patients.

In addition to genetic markers, there are also many studies on various proteomic biomarkers. As proteins are the biological workhorse responsible for most cellular processes, searching for protein biomarkers could be more accurate in reflecting the mechanism of cancer cellular status in disease progression (Tan et al., 2012). There is also plenty of research related to kinds of protein therapy biomarker in this topic. It was concluded by Qin et al. that osteopontin (OPN) is a promising biomarker for cervical cancer, as it acts as a potential therapeutic target involved in immunological activities and multiple tumor processes. Zhang et al. discussed the expression level of anti-silencing function 1B (ASF1B), a histone H3-H4 chaperone, and concluded that it is a promising independent prognostic biomarker and that it may serve as a potential immunotherapeutic target in HCC. Ding et al. drew a conclusion that increased expression of semaphorin 5B (SEMA5B) is associated with immune cell infiltration, and it can be served as a novel diagnostic biomarker and prognostic factor for kidney renal clear cell carcinoma (KIRC). The study by Fang et al. posited that CBX1/3/7/8, a family number of chromobox family proteins (CBXs), could serve as a potential therapeutic target and prognostic biomarker for esophageal carcinoma (ESCA). An inflammation-related signature could also be used as a biomarker in predicting the prognosis of patients. Yu et al. constructed an inflammation-related signature (IRS), which might be sensitive to immune drugs and serve as a biomarker to predict survival in KIRC. Sm proteins (SNRPD1/E/F/G) independently predict the prognostic outcome of LUAD, and Liu et al. explored the function of which ones act as treatment targets. Chemotherapeutic agents can elevate not only the therapeutic effects but also the malignancy of cancer cells. Zhang et al. concluded that through upregulating CNTN-1 in lung adenocarcinoma cells, low-dose cisplatin can activate the epithelial–mesenchymal transition (EMT) process and the resulting malignant progression.

There was also included reported biomarkers about prognosis. Shen et al. discussed the function of NAP1L1 that, through influencing the Wnt/β-Catenin pathway in HCC, may act as a novel prognostic biomarker associated with macrophages. The high expression of Fc receptor-like B (FCRLB) was first reported by Wang et al. to have the function of predicting the poor prognosis of CRC, which could be a potential prognostic biomarker. Hu et al. reported that delta-catenin could attenuate medulloblastoma cell invasion by targeting the EMT pathway, which might be a positive prognostic biomarker. Zhang et al. elucidated that transmembrane protein 170B is a prognostic biomarker, associated with the poor prognosis of pancreatic adenocarcinoma (PAAD), and considered it to be a therapeutic target. As ferroptosis was discovered as a type of regulated cell death recently, the ferroptosis-related gene pair index (FRGPI) was concluded to be an independent prognostic biomarker guiding individualized tumor therapy by Li and Wang. Liu et al. constructed and validated a prognostic model about macrophage differentiation-associated genes (MDGs) to predict the outcomes of clear cell renal cell carcinoma (ccRCC). Zhou et al. developed a gene signature to predict prognosis and resistance in diffuse large B-cell lymphoma (DLBCL), which can not only predict guide individualized treatment but also can predict survival and resistance. Song et al. revealed the important function of histone deacetylases (HDACs) family genes in the efficacy of immunotherapy and chemotherapy of LUAD. Liang et al. identified a novel model consisting of three cancer stem cell-related genes that could accurately and independently predict the clinical outcomes of HCC patients. Xu et al. mainly focused on identifying a novel signature that integrated immunoglobulin (IGHA2), a glycosylation-related gene (SLC35A2), and an anti-viral gene (BST2) as an independent prognostic indicator for BC.

Compared with single biomarkers, a combined panel of biomarkers may have a higher clinical value. The panel established by Luo et al. showed that the sensitivity, specificity, and AUC in diagnosing OS were 70.59%, 86.27%, and 0.798, respectively, higher than the single one. Yang et al. built a new prognostic prediction model based on three ras-related genes with the AUC of 0.932, contributing to providing new insight into both diagnosis and treatment in ESCC. Meng et al. constructed a necroptosis-related miRNA signature consisting of five miRNAs (miR-139-5p, hsa-miR-326, miR-10b-5p, miR-500a-3p, and miR-592) as a powerful tool in predicting the prognosis of HCC of AUC>0.7. A model consisting of liquid–liquid phase separation (LLPS)-correlated genes (LCGs) combined with clinical factors was constructed by Huang et al., who showed that it is a good tool in predicting the prognosis of BC with receiver operating characteristic curve (ROC) values in all cohorts (1/3/5-year ROC values were 0.89, 0.79, and 0.75, respectively).

Novel techniques to identify genetic or proteomic biomarkers were also included. High-throughput sequencing and screening based on identification technologies are particularly suitable for cancer research for diagnosis, prognosis, and therapeutics. Qian et al. detected nine gastrointestinal stromal tumor (GIST)-related gene mutations through targeted next-generation sequencing (NGS), which provide a reference for individualized diagnosis and treatment. The study by Papanikolaou et al. demonstrated that mapping unknown genes to functional pathways by network reconstruction could be used as a powerful tool to identify candidate oncoproteins as biomarkers. Zhong et al. used the CIBERSORT method to establish a prognostic model, revealing that infiltrating immune cells in CRC may be an important determinant of prognosis and immunotherapy. Through single-cell sequencing and machine learning algorithms, Li et al. identified 21 potential essential

biomarkers for HCC cells, providing diagnostic and therapeutic value for HCC pathogenesis. Luo et al. used the serological proteome analysis (SERPA) approach to select candidate anti-TAA autoantibodies as biomarkers for OS.

However, a tumor-associated biomarker cannot be applied in a clinical setting until it is confirmed accurate, reproducible, and reliable and has clinical utility (Hayes, 2015). Gallardo-Rincón et al. explored the survival rate of Mexican OC patients with the founder mutation (*BRCA1*-Del ex9-12) after treatment with olaparib. They showed that treatment increased progression-free survival (PFS). A biomarker needs to undergo a lengthy experimental validation process from research to clinical application to accomplish transformation. Only by achieving clinical application can a biomarker truly realize its value.

A case report was contained in this Research Topic. Han et al. presented the diagnosis, tailored genetic counseling, and cancer prevention of a locally advanced lung cancer patient with dMMR/MSI-H/TMB-H tumor and *PMS2*-LS, who received four cycles of nivolumab plus chemotherapy, achieving a disease-free survival of 16 months. Further efforts are required to investigate the efficacy of targeted therapy in rare tumor patients. In a future topic, we hope more case reports of rare cancer types or some other interesting cases in solid tumors associated with biomarkers in clinical practice will be reported for researchers.

The sources of cancer biomarkers are extensive, including blood, urine, feces, cerebrospinal fluid, ascites, pleural fluid, saliva, sweat, skin tissue, oral mucosa, biopsy and puncture tissue, and intraoperative specimens, among others. In this topic, blood is the most common source. Meanwhile, precision medicine acting on a specific target, whether for diagnosis or therapy, highlights the clinical application of biomarkers. However, the occurrence and progression of cancer is complex, which requires more effective and multiple kinds of biomarkers to gain an in-depth understanding of it.

In summary, we are proud of this Research Topic at *Frontiers in Genetics*, which covers many aspects of the biomarkers in solid tumors, especially relating to genetic and proteomic biomarkers, from discovery to clinical application. We hope this Research Topic can provide useful and helpful advice for researchers studying the detection and treatment of solid tumors. Moreover, we thank all authors and reviewers contributed to this Research Topic.

Author contributions

JL organized the Research Topic as guest editor. JL and AC drafted and finalized the editorial. RZ, CL, and JZ revised and suggested intellectual modifications. All authors contributed to the article and approved the submitted version.

Funding

This editorial was supported by the Innovation Fund of National Clinical Research Center for Orthopedics, Sports Medicine, and Rehabilitation (Grant No. 2021-NCRC-CXJJ-PY-13), the Young Elite Scientists Sponsorship Program by CAST (Grant No. 2021-QNRC2-A06), and the project of health guiding plan in Luoyang (Grant No. 2101043A).

Acknowledgments

The guest editors are grateful to all the authors contributing to this special issue papers collection and thank all the reviewers who helped improve the manuscripts.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., et al. (2022). Early detection of cancer. *Science* 375, eaay9040. doi:10.1126/science.aay9040

Hayes, D. F. (2015). Biomarker validation and testing. *Mol. Oncol.* 9, 960–966. doi:10.1016/j.molonc.2014.10.004

Siegel, R. L., Miller, K. D., Wagle, N. S., and Jemal, A. (2023). Cancer statistics, 2023. *CA Cancer J. Clin.* 73, 17–48. doi:10.3322/caac.21763

Tan, H. T., Lee, Y. H., and Chung, M. C. (2012). Cancer proteomics. *Mass Spectrom. Rev.* 31, 583–605. doi:10.1002/mas.20356

Wu, L., and Qu, X. (2015). Cancer biomarker detection: Recent achievements and challenges. *Chem. Soc. Rev.* 44, 2963–2997. doi:10.1039/c4cs00370e

frontiers in Genetics

ORIGINAL RESEARCH
published: 01 December 2021
doi: 10.3389/fgene.2021.773373

Identification and Validation of a Novel Pyroptosis-Related Gene Signature for Prognosis Prediction in Soft Tissue Sarcoma

Lin Qi[1,2], *Ruiling Xu*[1,2], *Lu Wan*[1,2], *Xiaolei Ren*[1,2], *WenChao Zhang*[1,2], *Keming Zhang*[3,4], *Chao Tu*[1,2] *and Zhihong Li*[1,2]*

[1]Department of Orthopaedics, The Second Xiangya Hospital, Central South University, Changsha, China, [2]Hunan Key Laboratory of Tumor Models and Individualized Medicine, The Second Xiangya Hospital, Changsha, China, [3]Department of Dermatology and Venereology, Changzheng Hospital, Second Military Medical University, Shanghai, China, [4]Shanghai Key Laboratory of Molecular Medical Mycology, Shanghai Institute of Medical Mycology, Changzheng Hospital, Second Military Medical University, Shanghai, China

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

Reviewed by:
Xiaoxiao Sun,
University of Arizona, United States
Haoyun Lei,
Carnegie Mellon University, United States

***Correspondence:**
Zhihong Li
lizhihong@csu.edu.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 09 September 2021
Accepted: 05 November 2021
Published: 01 December 2021

Citation:
Qi L, Xu R, Wan L, Ren X, Zhang W, Zhang K, Tu C and Li Z (2021) Identification and Validation of a Novel Pyroptosis-Related Gene Signature for Prognosis Prediction in Soft Tissue Sarcoma. Front. Genet. 12:773373. doi: 10.3389/fgene.2021.773373

Soft tissue sarcoma (STS) represents an uncommon and heterogenous group of malignancies, and poses substantial therapeutic challenges. Pyroptosis has been demonstrated to be related with tumor progression and prognosis. Nevertheless, no studies exist that delineated the role of pyroptosis-related genes (PRGs) in STS. In the present study, we comprehensively and systematically analyzed the gene expression profiles of PRGs in STS. The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases were utilized to identify differentially expressed PRGs. In total, 34 PRGs were aberrantly expressed between STS and normal tissues. Several PRGs were validated with RT-qPCR. Consensus clustering analysis based on PRGs was conducted to divide STS patients into two clusters, and significant survival difference was observed between two distinct clusters ($p = 0.019$). Differentially expressed genes (DEGs) were identified between pyroptosis-related clusters. Based on the least absolute shrinkage and selection operator (LASSO) COX regression analysis, the pyroptosis-related gene signature with five key DEGs was constructed. The high pyroptosis-related risk score group of TCGA cohort was characterized by poorer prognosis ($p < 0.001$), with immune infiltration and function significantly decreased. For external validation, STS patients from Gene Expression Omnibus (GEO) were grouped according to the same cut-off point. The survival difference between two risk groups of GEO cohort was also significant ($p < 0.001$). With the combination of clinical characteristics, pyroptosis-related risk score was identified to serve as an independent prognostic factor for STS patients. In conclusion, this study provided a comprehensive overview of PRGs in STS and the potential role in prognosis, which could be an important direction for future studies.

Keywords: soft tissue sarcoma, pyroptosis, gene signature, prognosis, immune

INTRODUCTION

Soft tissue sarcomas (STSs) comprise a rare group of heterogenous tumor cells, which account for only 1% of all adult malignancies (Gamboa et al., 2020). It was estimated that there were 13,460 cases with STS in the United States in 2021 (Siegel et al., 2021). STS originated from mesenchymal tissues with more than 100 different subtypes according to the histology and genetic alterations, which displayed various clinical behaviors (Vilanova, 2017). Although STS could arise in any body site, it exhibited a predilection to occur in the extremity and intra-abdominal region (Brennan et al., 2014). Surgical procedures are the cornerstones of STS treatment (Crago and Brennan, 2015). The past few decades also have witnessed the evolution of therapeutic strategies for STS, with the collaboration of multidisciplinary team (MDT) including radiologists, pathologists, oncologists and surgical specialists (Gamboa et al., 2020). However, for elderly patients with STS, 5-years relative survival was below 50% (Hoven-Gondrie et al., 2016). Additionally, STS was also characterized by the susceptibility to distant metastasis and recurrence. Nearly half of patients with localized STS developed distant metastasis, especially to the lung and leading to poor prognosis (Navarria et al., 2015; Gamboa et al., 2020). Therefore, novel therapeutic targets and reliable prognostic model need to be identified for effective and personalized treatment.

Pyroptosis, recognized as caspase 1-dependent programmed cell death (PCD), features the prompt perforation of plasma membrane along with releasing intracellular properties of pro-inflammatory function (Bergsbaken et al., 2009; Shi et al., 2017). Pyroptosis is usually triggered by the activation of pattern recognition receptors (PRRs) and then activated caspase one upon inflammasomes (Xue et al., 2019). It was reported that pyroptosis-associated PRRs consist of intracellular nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs), Toll-like receptors (TLRs) and absent in melanoma 2 (AIM2)-like receptors (ALRs) (Lamkanfi and Dixit, 2014). In recent years, gasdermin D (GSDMD) was reported as the executioner of pyroptosis, as it released N-terminal fragment (GSDMD-cNT) to induce cell swelling after caspase cleavage (Shi et al., 2015; Ding et al., 2016). Likely, other gasdermin family genes including GSDMA, GSDMB, GSDMC and GSDME also take part in the process of pyroptosis (Ding et al., 2016). In the pathophysiologic process of diseases, pyroptosis is competitively regulated between the host and pathogen, and the outcomes determine the fate of the host in turn (Bergsbaken et al., 2009).

The tight relationship between pyroptosis and cancers has been reported in recent years, while controversy still exists regarding the dual role of pyroptosis (Xia et al., 2019). Various signaling pathways activated by pyroptosis may promote tumorigenesis and chemoresistance (Thi and Hong, 2017; Zhou and Fang, 2019). Nevertheless, pyroptosis may also exert tumor-suppressive effect by inhibiting tumor growth and angiogenesis (Nagarajan et al., 2019). In hepatocellular carcinoma (HCC), significant downregulation of NLR family pyrin domain containing 3 (NLRP3) was observed, which was inversely correlated with clinical stage (Wei et al., 2014). Highly expressed GSDMB was associated with poor prognosis and high metastatic potential in breast cancer (Hergueta-Redondo et al., 2014). Furthermore, there has been no relevant study focusing on the role of pyroptosis in STS. With genomic and clinical information integrated, the current study aims to systematically analyze pyroptosis-related genes (PRGs) in STS, develop and validate pyroptosis-related risk score and establish the novel prognostic model for STS.

MATERIALS AND METHODS

Data Collection and Sources

The UCSC Xena browser (https://xenabrowser.net/datapages/) was used to download gene expression profiles of The Cancer Genome Atlas (TCGA)—sarcoma (SARC) cohort and normal tissues in the Genotype-Tissue Expression (GTEx) dataset (Goldman et al., 2020). FPKM values of RNA sequencing (RNA-Seq) data from TCGA and GTEx were normalized through $\log_2$(FPKM+1) transformation. RNA-Seq data from two database were then processed and unified following sufficiently rigorous procedures, including the uniform realignment, the quantification of gene expression and the correction of batch effect (Wang et al., 2018). Clinical information of TCGA-SARC cohort has been made available for download at cBioPortal (https://www.cbioportal.org/) (Cerami et al., 2012). Within TCGA-SARC cohort, a total of 259 patients with STS were screened, composed of 104 patients having leiomyosarcoma (LMS), 59 patients having dedifferentiated liposarcoma (DDLPS), 49 patients having undifferentiated pleomorphic sarcoma (UPS), 25 patients having myxofibrosarcoma (MFS) and 22 patients having other STS. In GTEx, gene expression profiles of 911 normal human adipose and muscle were integrated with that of TCGA-cohort owing to the lack of normal tissues in TCGA. Additionally, the RNA-Seq profiles with clinical characteristics of GSE30929 were available in the GEO data repository for further validation.

Identification of Differentially Expressed genes Between Pyroptosis-Related Clusters

In total, 37 PRGs were identified based on previous studies (Kang et al., 2014; Zhu et al., 2017; Feng et al., 2018; Tsuchiya et al., 2019; Xue et al., 2019; Zhang et al., 2020; Zheng and Kanneganti, 2020), and were listed in **Supplementary Table S1**. Differential expression of PRGs between SARC and corresponding normal tissue was conducted, utilizing the "limma" R package (Version 3.48.3). In clusters based on consensus clustering analysis of PRGs, DEGs between cluster one and cluster two were identified if $|\log_2(\text{fold change})| > 2$ in expression value and false discovery rate (FDR) < 0.05. We established correlation network of PRGs using "corr" R package (Version 0.4.3) with the correlation coefficient set of 0.4. Somatic mutation of PRGs were visualized utilizing "Maftools" R package (Version 2.8.0). Circos plot was present to demonstrate the location of PRGs by "Circos" R package (Version 1.2.1).

Protein–Protein Interaction Network for PRGs

The PPI network of PRGs was constructed with the minimum required interaction score set of 0.9 to ensure high confidence, utilizing STRING database (https://string-db.org/) (Szklarczyk et al., 2021). Moreover, PPI network of PRGs was further analyzed using Cytoscape software (version 3.8.2). Hub genes were then screened based on MCODE, with the indicators set as degree cutoff = 2, node score cutoff = 0.2, K-core = 2 and maximum depth = 100.

Pyroptosis-Based Consensus Clustering Analysis

The "ConsensusClusterPlus" R package (Version 1.56.0) was introduced to conduct consensus clustering analysis, so as to identify pyroptosis-related subtypes of STS. Because k-means clustering analysis was stochastic, repetitions was set to 1,000 to ensure stable clustering (Wilkerson and Hayes, 2010). Differences in survival among clusters were visualized with Kaplan-Meier (KM) plots based on the R packages of "survival" (Version 3.2–11) with "survminer" (Version 0.4.9).

Gene Set Enrichment Analysis

DEGs (| $\log_2$ (fold change) | > 2 in expression value and FDR <0.05) between cluster one and cluster two based on pyroptosis-related consensus clustering were collected. With the "clusterProfiler" R package (Version 4.0.4) introduced, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed.

Development of Pyroptosis-Based Prognostic Model

The prognostic significance of DEGs within different clusters was evaluated by utilizing univariate COX regression analysis. In order to narrow down gene selection, DEGs with significant impact on survival ($p < 0.01$) were subsequently incorporated into the least absolute shrinkage and selection operator (LASSO) Cox regression analysis based on the "glmnet" R package (Version 4.1–2). The risk scores were further established according to the formula: $\sum_{i}^{n} Xi*Yi$ (*Xi*: coefficients of the gene *i*, *Yi*: expression values of the gene *i*).

The risk scores were subsequently discretized to divide TCGA-SARC cohort into high- and low-risk groups. Survival between two risk groups was analyzed with KM plot in TCGA-SARC cohort. For further external validation of risk scores, gene expression profiles of GSE30929 were entered into the formula, and we sorted these patients into high- and low-risk groups according to the same cut-off point. The "timeROC" R package (Version 0.4) was introduced for evaluating the predictive accuracy. Furthermore, the risk scores integrated with clinical characteristics including age, sex, race, histology, tumor site, tumor multifocality, surgical margin, tumor depth and radiotherapy were analyzed using multivariate COX regression analysis. The nomogram was performed to illustrate the prognostic model, which was further evaluated by the calibration curve.

TABLE 1 | Sequences of the primers used in RT-qPCR.

Gene	Sequence of primer
CASP3	F: CATGGAAGCGAATCAATGGACT
	R: CTGTACCAGACCGAGATGTCA
DHX9	F: GCAGCAGAGTGTAACATCGTAG
	R: ACTCAAATCGAACGCTGTAGC
IL1B	F: ATGATGGCTTATTACAGTGGCAA
	R: GTCGGAGATTCGTAGCTGGA
GADPH	F: CAGGAGGCATTGCTGATGAT
	R: GAAGGCTGGGGCTCATTT

UPS, undifferentiated pleomorphic sarcoma.

Single Sample Gene Set Enrichment Analysis and Immune Infiltration Analysis

The 16 immune cells infiltration and 13 related functions were quantified through ssGSEA in different pyroptosis-related risk groups of TCGA-SARC cohort and GSE30929, by utilizing the R package of "GSVA" (Version 1.40.1) and "GSEABase" (Version 1.54.0). The correlation between expression of DEGs in the gene signature and immune infiltrates was analyzed by the Tumor Immune Estimation Resource 2.0 database (TIMER2.0) (Li et al., 2020).

Cell Lines and Cell Culture

The human synovial sarcoma cell line (SW-982) was purchased from the American Type Culture Collection (ATCC). The human skin fibroblast cell line (HSF) with related media were purchased from Fenghui Biotechnology Co., Ltd (Hunan, China). The primary human synovial sarcoma cells (hSS-005R) were also established for the validation of PRGs. The human synovial sarcoma cells SW-982 and hSS-005R were cultured in Dulbecco's modified Eagle medium (DMEM) (Gibco, United States) supplemented with 10% fetal bovine serum (FBS) (Gibco, United States) and 1% penicillin-streptomycin (NCM Biotech, China). Cells were cultured at 37°C with 5% CO_2 in a humidified incubator (Thermo Fisher Scientific, United States).

Real-Time Quantitative PCR

Total cellular RNA was extracted using the RNA Express Total RNA Kit (M050, NCM Biotech, China). For cDNA synthesis, reverse transcription was conducted with the RevertAid First Strand cDNA Synthesis kit (K1622, Thermo Fisher Scientific, United States). Subsequently, RT-qPCR was performed on the StepOne Plus (Applied Biosystems, United States) by utilizing SYBR Green qPCR Master Mix (2×) (Bimake, United States). The primers used for the RT-qPCR were listed in **Table 1**.

Statistical Analysis

Statistical Analysis were conducted by utilizing R (Version 4.1.0). Differential gene expression between two groups was identified using Wilcoxon rank sum test, with *p* value calculated for each gene. Spearman's correlation test and matrix were conducted to compare gene expression with each other. Survival differences were compared utilizing log-rank test with KM curve. The $\chi 2$ test

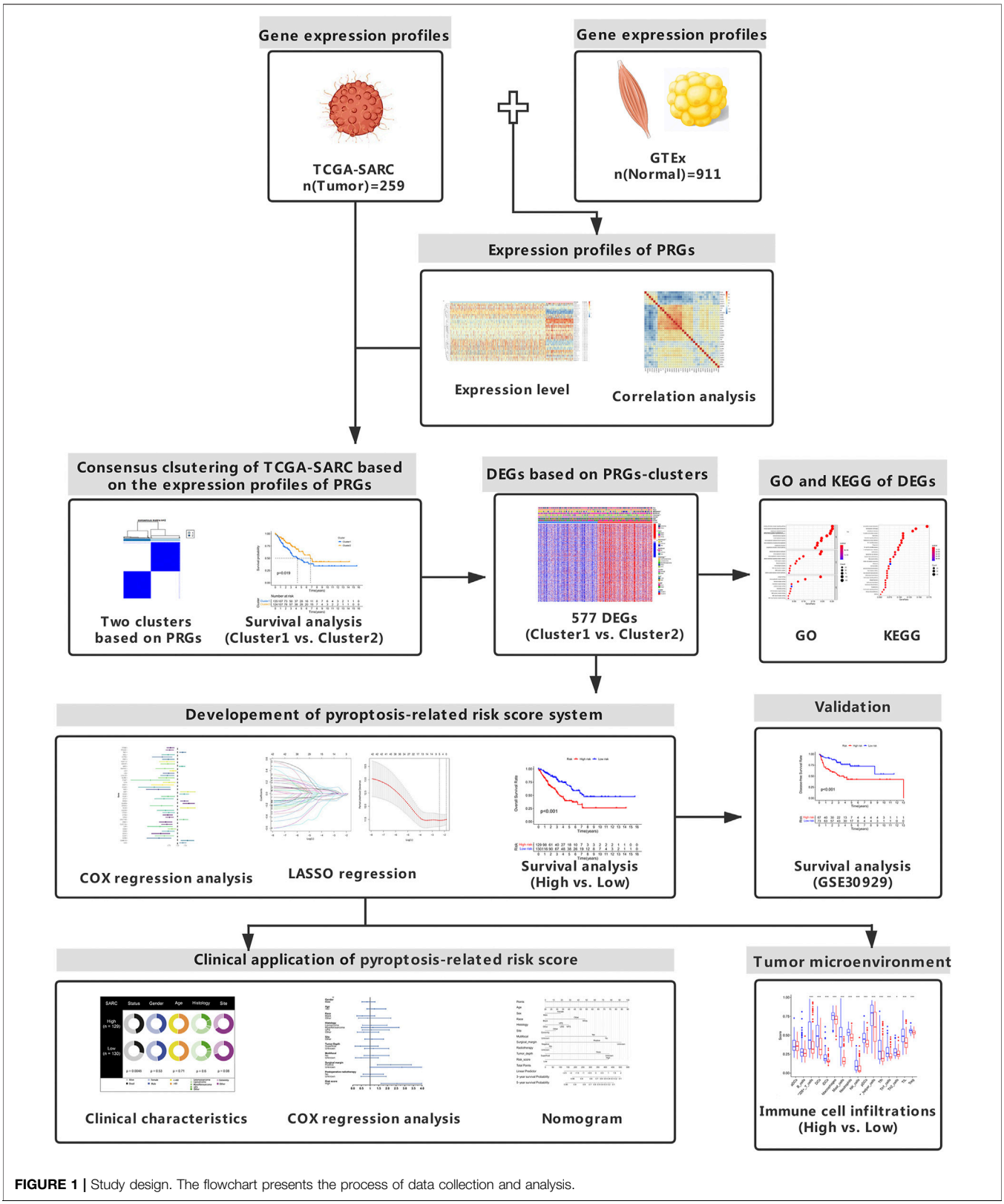


FIGURE 1 | Study design. The flowchart presents the process of data collection and analysis.

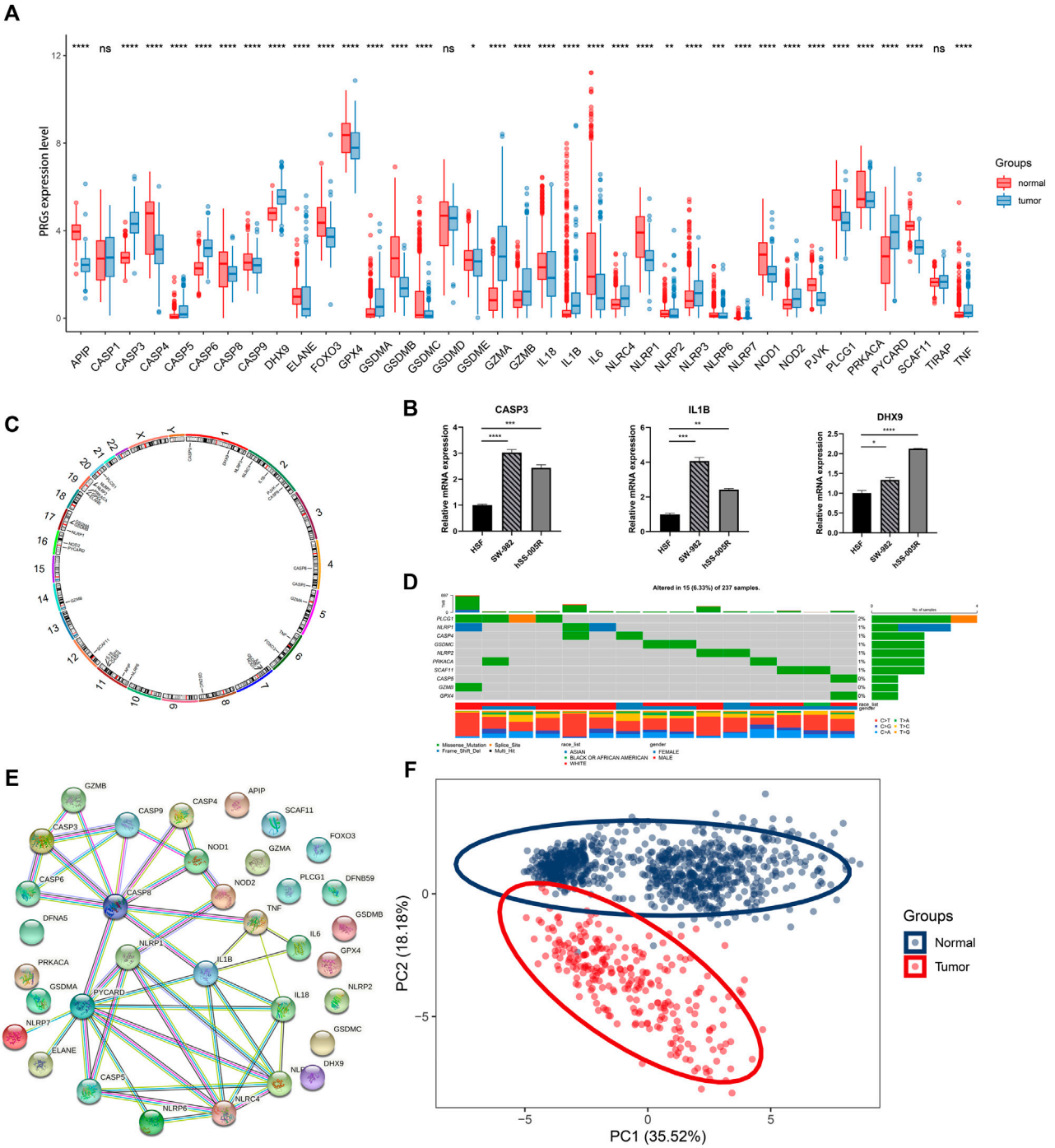


FIGURE 2 | Expression and association of PRGs. **(A)** The expression of PRGs between STS (blue) and normal tissues (red). Box plot represents the median (center horizontal line), upper and lower quartiles (top and bottom horizontal line). **(B)** Validation of mRNA expression of key PRGs in cell lines. **(C)** The location of PRGs on chromosomes. **(D)** Schematic overview of mutation frequency and type in PRGs. **(E)** PPI network constructed by PRGs-encoded proteins (interaction score: 0.9). **(F)** Principal component analysis (PCA) for discriminating STS and normal tissues based on PRGs. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, ns: nonsignificant.

or Fisher's exact test was introduced to evaluate clinical characteristics between high- and low-risk groups. COX regression analysis was conducted to identify prognostic factors, and hazard ratio (HR) with 95% confidence interval (CI) were also computed. Statistical difference of $p < 0.05$ was defined significant.

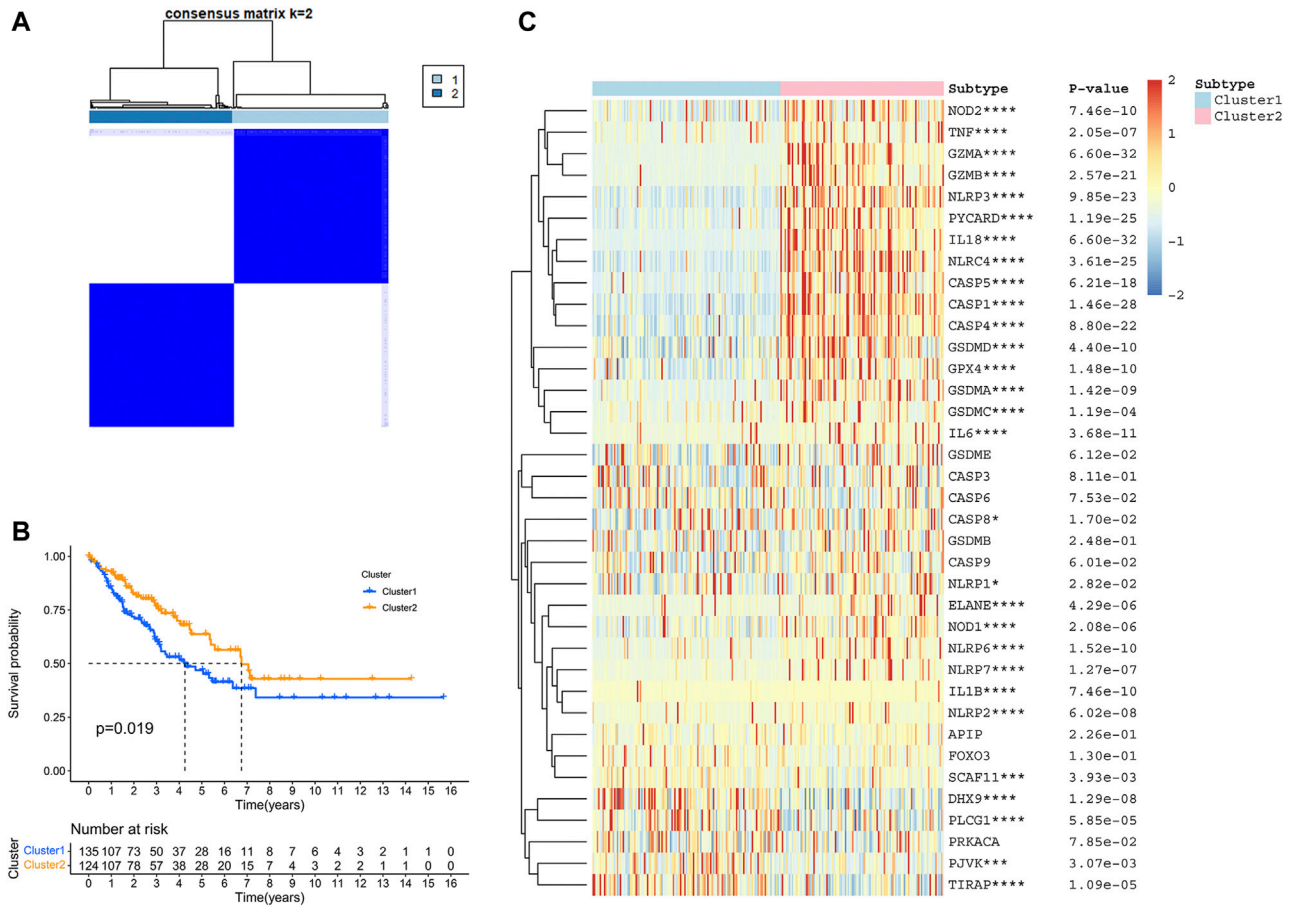


FIGURE 3 | Identification of TCGA-SARC cluster based on PRGs. **(A)** TCGA-SARC cohort was divided into two distinct clusters using pyroptosis-based consensus clustering analysis (k = 2, repetition = 1,000). **(B)** Overall survival (OS) curve comparing survival of patients in cluster 1 (blue) and cluster 2 (orange). **(C)** Heatmap of PRGs between two clusters (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

RESULTS

Identification of PRGs Between STS and Normal Tissues

The study design was illustrated in **Figure 1**. Totally 37 PRGs were included to detect differential expression between STS and normal tissues from TCGA-SARC and GTEx dataset (**Supplementary Table S1**). We found that 34 PRGs were differentially expressed (p < 0.05), among which 14 genes (CASP3, CASP5, CASP6, DHX9, GSDMA, GZMA, GZMB, IL1B, NLRC4, NLRP3, NLRP7, NOD2, PYCARD, TNF) were upregulated and 20 genes (APIP, CASP4, CASP8, CASP9, ELANE, FOXO3, GPX4, GSDMB, GSDMC, GSDME, IL18, IL6, NLRP1, NLRP2, NLRP6, NOD1, PJVK, PLCG1, PRKACA, SCAF11) were downregulated in STS group (**Figure 2A**, **Supplementary Figure S1A**). To validate PRGs in related cell lines, we performed RT-qPCR analysis (**Figure 2B**). The expression levels of several key PRGs including CASP3, IL1B and DHX9 were significantly higher in the human synovial sarcoma cells SW-982 and hSS-005R, compared with those in the human skin fibroblast cell line (HSF). The correlation network of 34 PRGs was present in **Supplementary Figure S1B** and **Supplementary Figure S1C**. The chromosome location of 34 PRGs was illustrated in **Figure 2C**. We then analyzed copy number variations (CNVs) of PRGs (**Figure 2D**). It could be found that 15 of 237 (6.33%) SARC samples displayed pyroptosis-related mutations. And the majority of mutations were missense mutations. The PPI network was subsequently established and seven hub genes (NLRP1, IL18, NLRC4, NLRP3, PYCARD, CASP5, IL1B) were identified (**Figure 2E**). Furthermore, these 34 PRGs were efficient to discriminate STS and normal tissues on the expression level (**Figure 2F**).

Identification of TCGA-SARC Cluster Based on PRGs

To elucidate different STS subtypes and corresponding clinical characteristics and prognosis, the TCGA-SARC cohort was clustered into two distinct clusters based on PRGs through consensus clustering analysis (**Figure 3A**, **Supplementary Figure S2A–F**). Within pyroptosis-related cluster 1, There were 135 STS patients and 124 STS patients were in pyroptosis-related cluster 2. Remarkably, overall survival (OS) curves of these two clusters indicated significantly great survivorship difference (p =

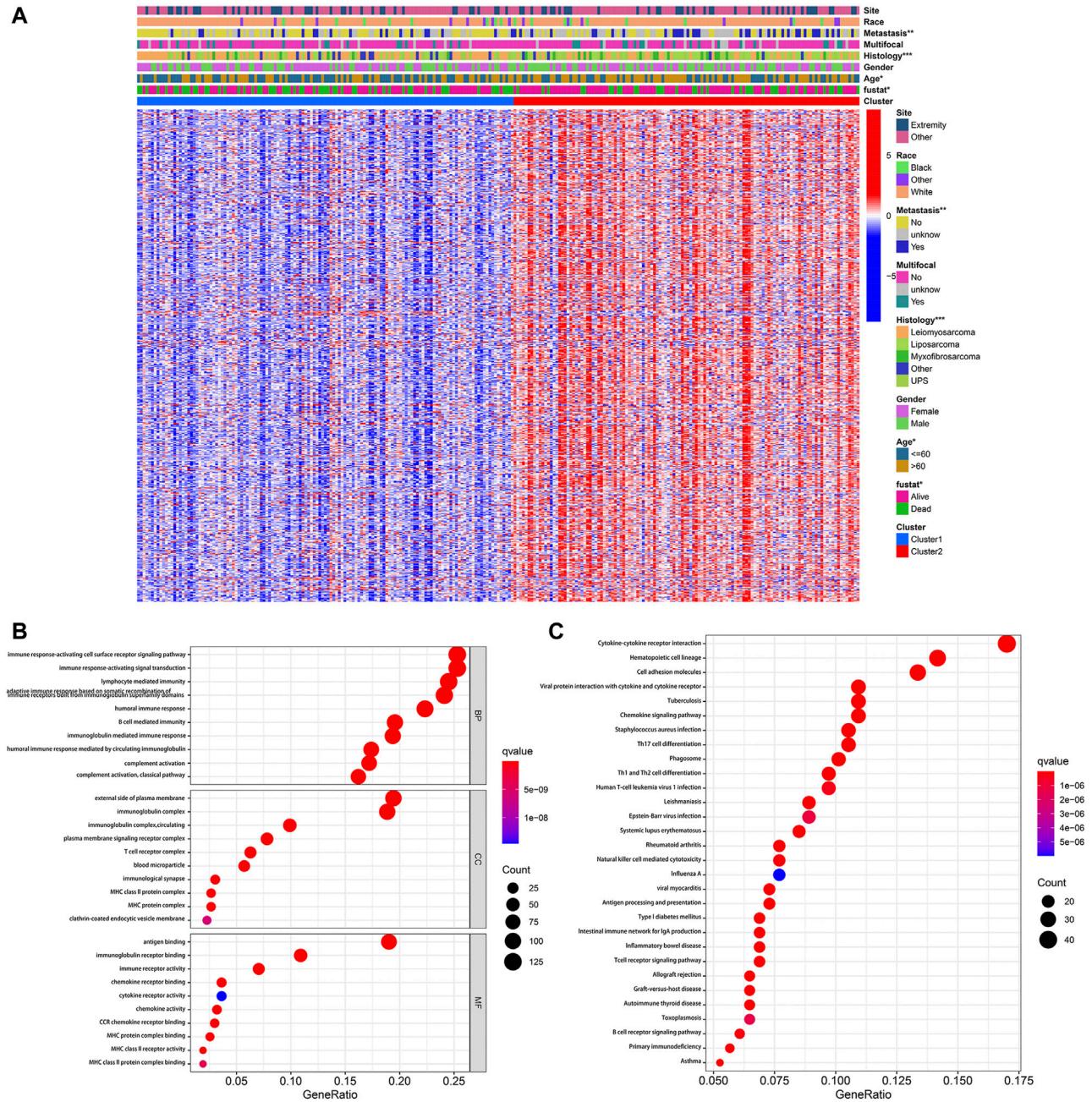


FIGURE 4 | Identification of DEGs between pyroptosis-related clusters. **(A)** Heatmap depicting DEGs of pyroptosis-related clusters and corresponding clinical characteristics. **(B)** GO enrichment analysis including biological process (BP), cellular component (CC), and molecular function (MF). **(C)** KEGG enrichment analysis indicating related genes and pathways. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

0.019, **Figure 3B**). In **Figure 3C**, the gene expressing level of PRGs in two distinct clusters were displayed.

Profiling DEGs Between Pyroptosis-Related Clusters

According to the stringent selecting criterion of $|\log_2$ (fold change)$| > 2$ in expression value and FDR <0.05, a total of 577 DEGs were identified between pyroptosis-related cluster one and cluster 2 (**Figure 4A**). There were also significant differences in clinicopathological characteristics including age, histology, metastatic status and survival between two distinct clusters ($p < 0.05$).

Subsequently, these 577 DEGs were subjected to the GO and KEGG enrichment analysis, in order to reveal biological processes and mechanisms of pyroptosis-related clusters. GO

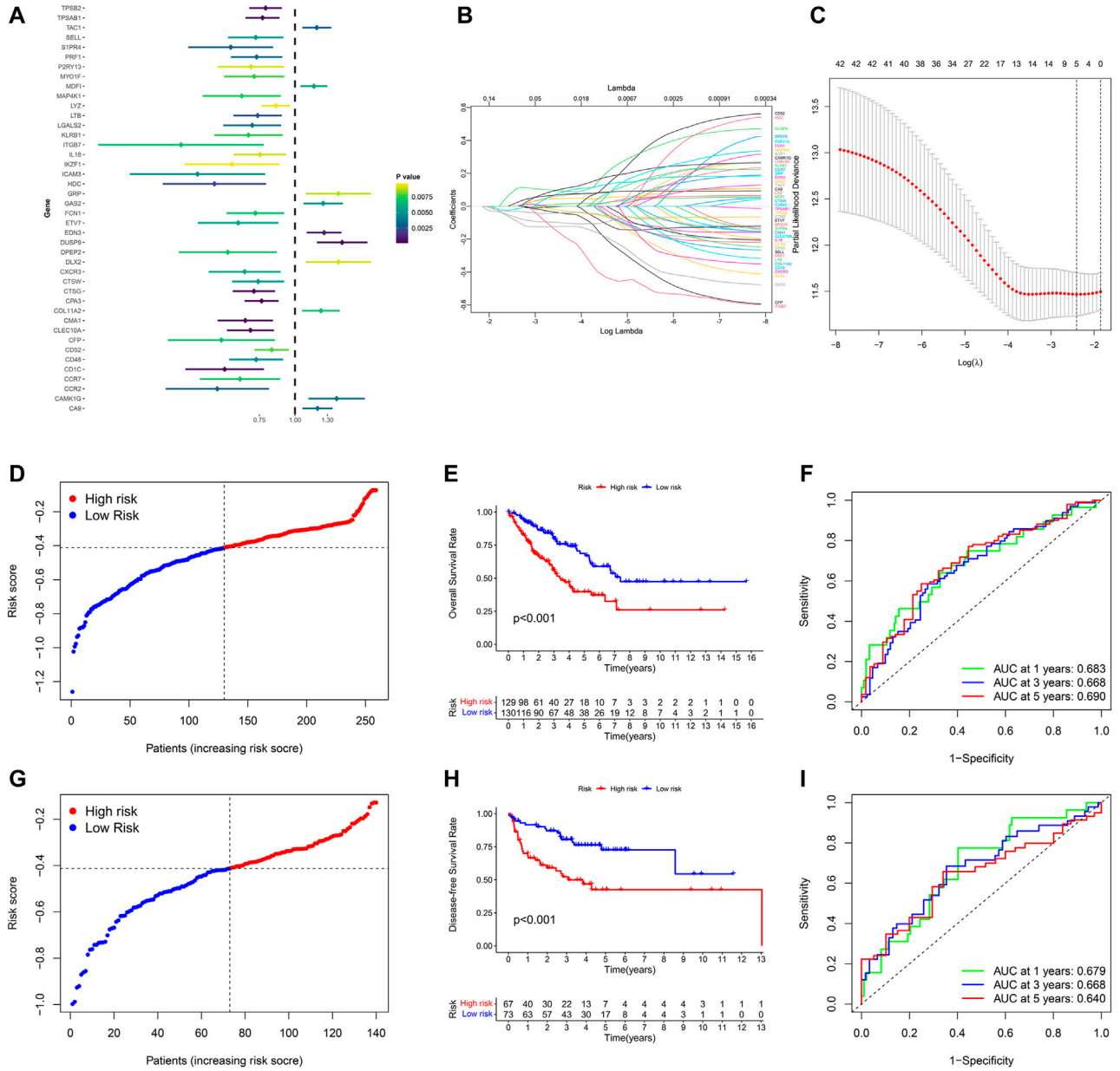


FIGURE 5 | Development and validation of the pyroptosis-related gene signature in STS. **(A)** Univariate COX regression analysis of DEGs between pyroptosis-related cluster one and cluster 2, and all 42 genes with $p < 0.001$. **(B)** LASSO regression analysis of 42 DEGs. **(C)** Cross validation method to select optimal genes. **(D)** Distribution of TCGA-SARC cohort based on the risk score. **(E)** OS curve of TCGA-SARC cohort in low-risk and high-risk groups. **(F)** Time-dependent ROC to evaluate the prognostic performance of the risk score in TCGA-cohort. **G** Distribution of GSE30929 cohort based on the risk score. **(H)** Disease-free survival (DFS) curve of GSE30929 cohort in low-risk and high-risk groups. **(I)** Time-dependent ROC to evaluate the prognostic performance of risk score in GSE30929 cohort.

enrichment analysis indicated that DEGs were predominantly enriched in immune response-activating cell surface receptor signaling pathway, immune response-activating signal transduction, external side of plasma membrane and antigen binding (**Figure 4B**). Moreover, these DEGs were also significantly enriched in cytokine-cytokine receptor interaction, hematopoietic cell lineage and cell adhesion molecules (**Figure 4C**).

Development and Validation of Pyroptosis-Related Gene Signature in STS

The prognostic significance of 577 DEGs between cluster one and cluster two were analyzed by utilizing univariate COX regression analysis. Accordingly, 42 genes were preserved based on the strict criteria ($p < 0.01$) and processed for subsequent analysis (**Figure 5A**). The LASSO COX regression analysis was then performed, and five key genes

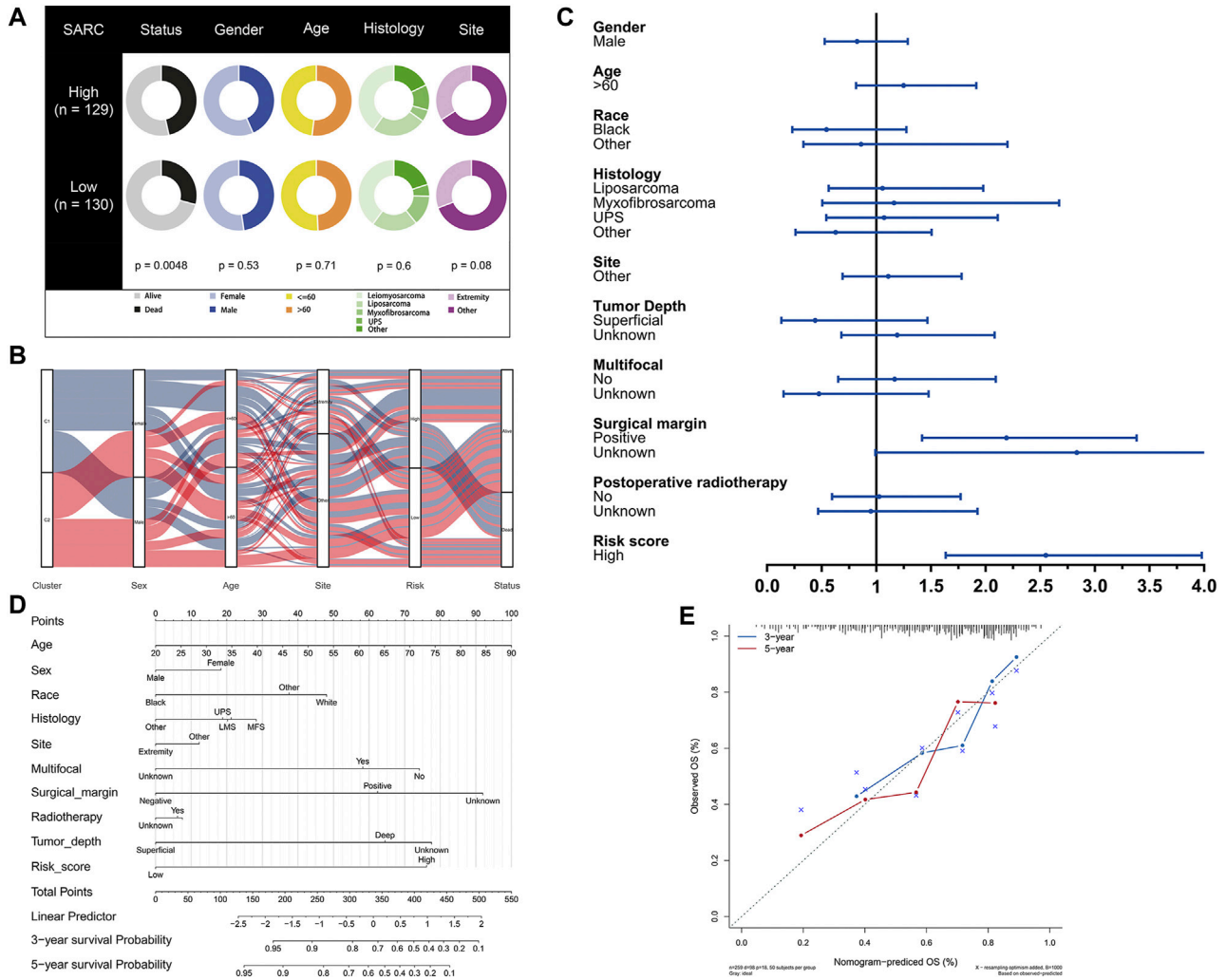


FIGURE 6 | Development of pyroptosis-based prognostic model. **(A)** Clinical characteristic between low-risk and high-risk groups. **(B)** Alluvial diagram illustrating the relationship of pyroptosis-based cluster distribution, clinical characteristics, different risk groups and survival outcomes. **(C)** Multivariate COX regression analysis of clinical characteristics and pyroptosis-based risk score. **(D)** Nomogram predicting 3-years and 5-years survival rate of STS patients. **(E)** Calibration curve for predicting OS rate of STS patients.

were eventually identified with the pyroptosis-related gene signature constructed (**Figures 5B,C**). The risk score = (-0.05167*CTSG exp.) + (0.06184*DUSP9 exp.) + (-0.02483*CLEC10A exp.) + (-0.02979*CPA3 exp.) + (-0.06014*CD1C exp.). Based on the median of the risk scores in TCGA-SARC cohort, 259 STS patients were divided into the low-risk group (n = 130) and the high-risk group (n = 129) (**Figure 5D**, **Supplementary Figure S3A**). Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) indicated that low-risk and high-risk group could be clearly distinguished (**Supplementary Figure S3C**, **Supplementary Figure S3E**). The KM plot of OS rate demonstrated significant difference between two risk scores groups of TCGA-SARC cohort ($p < 0.001$, **Figure 5E**). Time-dependent receiver operating characteristics (ROC) curves were introduced for assessing model performance, and the area under curve (AUC) of 1-year, 3-year and 5-year OS rate were 0.683, 0.668 and 0.690, accordingly (**Figure 5F**).

For further validating the accuracy of the pyroptosis-related gene signature, corresponding data of GSE30929 was retrieved and the risk score of was calculated respectively. According to the same cut-off point of TCGA-SARC cohort, GSE30929 cohort was divided into different risk groups (**Figure 5G**, **Supplementary Figure S3B**). PCA and t-SNE also illustrated optimal degree of discrimination between high- and low-risk groups of GSE30929 (**Supplementary Figure S3D**, **Supplementary Figure S3F**). Remarkably, the disease-free survival (DFS) of two distinct risk groups demonstrated significant discrepancy ($p < 0.001$, **Figure 5H**). AUC were 0.679 for 1-year, 0.668 for 3-year and 0.640 for 5-year (**Figure 5I**).

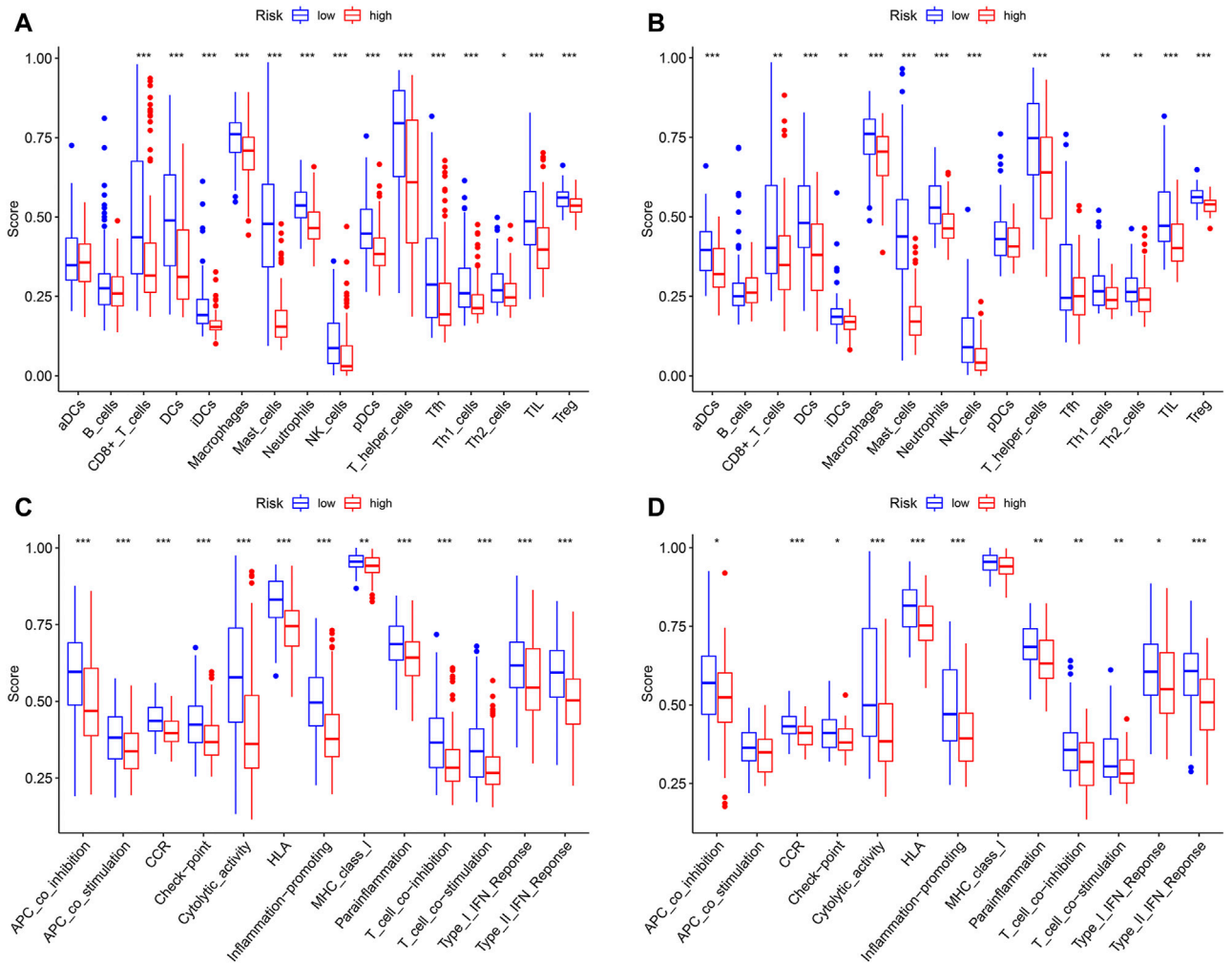


FIGURE 7 | Analysis of immune status based on pyroptosis-related risk score. **(A, C)** Comparisons of immune cells and immune functions between different risk groups in TCGA-SARC cohort. **(B, D)** Comparisons of immune cells and immune functions between different risk groups in GSE30929 cohort (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Development of Pyroptosis-Based Prognostic Model

The potential clinical utility of pyroptosis-based risk score was further investigated. Clinical characteristics of gender, age, tumor histology and tumor site between high- and low-risk groups were visualized in **Figure 6A**. Alluvial diagram also illustrated the relationship of pyroptosis-based cluster distribution, clinical characteristics, different risk groups and survival outcomes (**Figure 6B**). Furthermore, multivariate COX regression analysis integrated with clinical characteristics and pyroptosis-based risk score were performed to establish the prognostic model (**Figure 6C**, **Supplementary Table S2**). Based on the established prognostic model, a novel nomogram was subsequently constructed for predicting the survival probability of STS patients (**Figure 6D**). The 3-year and 5-year OS rate have proven to be relatively well predicted by the calibration curve of the nomogram (**Figure 6D**).

Analysis of Immune Status Based on Pyroptosis-Related Risk Score

In order to compare the immune activity, the ssGSEA was applied to analyze the immune infiltration and functions of high- and low-risk groups. The ssGSEA score calculated could indicate the infiltration degrees of immune cells and pathways within TCGA-SARC cohort and GSE30929 cohort. In TCGA-SARC cohort, the infiltration degrees of $CD8^+$ T cell, dendritic cell (DC), immature DC (iDC), macrophage, mast cell, neutrophil, natural killer (NK) cell, plasmacytoid DC (pDC), T helper (Th) cell, T follicular helper (Tfh) cell, Th1 cell, Th2 cell, tumour-infiltrating lymphocyte (TIL) and regulatory T cell (Treg) were significantly lower in the high pyroptosis-related risk group ($p < 0.05$, **Figure 7A**). All 13 related immune functions were also significantly decreased in the high pyroptosis-related risk group ($p < 0.05$, **Figure 7C**). GSE30929 dataset was subsequently included to analyze the immune activity. The results were similar

with the majority of immune infiltration and function significantly decreased in the high-risk group (**Figure 7B**, **Figure 7D**). The correlation of key DEGs in the gene signature and immune infiltration was also illustrated (**Supplementary Figure S4A–E**).

DISCUSSION

For a long period of time, apoptosis has been traditionally considered as the predominant mode that regulating cell death (Lowe and Lin, 2000; Reed, 2000). With further studies on novel forms of cell death, pyroptosis has aroused increasing attention due to its morphological and mechanistical distinction from others (Bergsbaken et al., 2009). Besides, pyroptosis was reported to chemically mediate multiple processes of malignancy progression (Wang et al., 2019; Ruan et al., 2020). However, no study has yet investigated the role of pyroptosis in STS. In this study, we comprehensively analyzed gene expressing profiles of PRGs in STS. Moreover, pyroptosis-related risk scoring system was established and validated.

The gene expression of 34 in 37 predefined PRGs was found to be significantly different between STS and normal tissues in the current study. Similarly, PRGs included in this study were relatively consistent with those in studies focusing on pyroptosis in other tumor types (Lin et al., 2021; Ye et al., 2021). Due to constraints of the TCGA data, normal tissue samples were extremely limited in TCGA-SARC cohort. Thus, GTEx including expression profiles of normal human tissues was the optimal data resource (Carithers et al., 2015). The gene expression data of TCGA and GTEx were merged through sufficiently rigorous procedures (Wang et al., 2018), and this novel method has been confirmed by several studies concerning TCGA-SARC cohort (Hu et al., 2020; Huang et al., 2021). Besides, PRGs identified in this study showed good ability to distinguish STS from normal samples. Based on the differentially expressed PRGs sets, tumor-infiltrated and normal tissues could be clearly distinguished, which was strongly suggestive of the role in tumor diagnosis. Consensus clustering analysis was a proven method to demonstrate distinct subtypes and survival patterns of malignant tumors (Brannon et al., 2010; Wang et al., 2014). The clustering of subtypes of TCGA-SACR cohort provided an opportunity to identify biological differences of STS based on PRGs. STS patients in pyroptosis-related cluster two had substantially better prognosis, with most PRGs significantly upregulated within this cluster. Remarkably, the previous study has demonstrated that the ATP releasing by dying tumor cells would act on $P2X_7$ purinergic receptors and subsequently trigger NLRP3-CASP1 complex (inflammasome) to mediate the innate and adaptive immune responses against dying tumor cells (Ghiringhelli et al., 2009). In keeping with previous findings, the current study also identified significant upregulation of NLRP3 and CASP1 in the pyroptosis-related cluster 2, which was probably related with underlying mechanisms of different prognosis in STS clusters.

The DEGs analysis between two pyroptosis-related clusters was further conducted to establish the risk score system, which was also a proven method of identifying different risk tumor patterns (Bai et al., 2021). Gene enrichment analysis revealed the significant enrichment in several immune-related pathways, and these findings were also coincident with the role of pyroptosis in mediating immune system (Hachim et al., 2020). Importantly, pyroptosis-related signature was established based on five key DEGs by LASSO COX regression analysis. Besides, solely relying on a single gene for diagnosis and prognosis prediction was inaccurate (Ju et al., 2021). The utility of pyroptosis-related risk score was confirmed by the significant survival difference in TCGA-SARC cohort. As relevant STS dataset in Gene Expression Omnibus (GEO) was extremely limited, there was no OS status recorded in GSE30929, and only DFS status were available. However, to our surprise, this pyroptosis-related risk score was also significantly efficient to predict the DFS of GSE30929 cohort for external validation. There were studies demonstrating that DFS could be considered as the acceptable surrogate of OS in a variety of tumors (Fajkovic et al., 2013; Oba et al., 2013), which hinted potential relationship between OS and DTS in STS.

After adjusting for clinical characteristics, pyroptosis-related risk score turned out to be the independent prognostic factor for overall survival of TCGA-SARC cohort. Besides, the nomogram integrated with pyroptosis-related risk score and clinical indicators has been developed for the clinical application. The score of each indicator could be added to estimate the OS rate of patients with STS. In the high pyroptosis-related risk group, the immune infiltration degrees were significantly lower, also indicating abnormal immune functions (Pages et al., 2010). Dual-specificity phosphatase 9 (DUSP9), one gene of the pyroptosis-related signature, is a dual-specificity phosphatase inhibiting mitogen-activated protein kinases (MAPKs) with preference for ERK (Caunt and Keyse, 2013). Multiple studies have revealed the relationship between DUSP9 and different types of tumors (Lu et al., 2018; Chen et al., 2021). In the current study, the coefficient of DUSP9 was positive in the formula of the pyroptosis-related risk score, which contributed the most to the increasing of the risk score and may suggest several directions towards relevant fields.

To our best knowledge, this is the first study identifying pyroptosis-related gene signature in STS, which is of great significance in diagnosis and survival prediction. However, the sample size of STS was limited due to disease characteristics, which was one of the deficiencies of this study. K-means clustering performed with $k = 2$ was also based on the limited sample size. Besides, most clinical data were collected retrospectively, and several important clinical characteristics including tumor grade, tumor stage, and surgy information were not available, which leading to some inevitable bias. Therefore, findings in this study should be viewed as the resource for future studies.

In conclusion, this study comprehensively and systematically analyzed the gene expression profiles of PRGs in STS. A total of 34 differentially expressed PRGs were identified, which were efficient to discriminate STS and normal tissues. Distinct pyroptosis-related clusters were divided with corresponding DEGs analyzed. Furthermore, pyroptosis-related risk scoring system with five key DEGs was established and served as an independent prognostic factor for STS patients. There was a significant difference in the levels of immune infiltration between low and high pyroptosis-related risk groups.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA-SARC and GSE30929.

AUTHOR CONTRIBUTIONS

ZL and CT designed the study and made final approval of the version. LQ performed study concept and design and wrote the manuscript. RX, LW, XR and WZ helped with data analysis. KZ interpretated results and helped to write the manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (NSFC; No. 81902745, No.82172500, No.82103228), Hunan Provincial Research and Development Program in Key Areas (2020DK 2003), and China Postdoctoral Science Foundation (No. 2021M693557).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.773373/full#supplementary-material

Supplementary Figure S1 | Expression and association of PRGs. A The heatmap of PRGs between STS and normal tissues (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). B and C The correlation network of PRGs.

Supplementary Figure S2 | Consensus clustering of TCGA-SARC cohort based on PRGs. A-D Consensus clustering based on PRGs (k = 3-6). E Consensus cumulative distribution function (CDF) Plot based on PRGs. F Delta area plot of consensus clustering based on PRGs.

Supplementary Figure S3 | Identification of low-risk and high-risk group based on risk score. A The survival status of TCGA-SARC cohort based on the risk score (low-risk group: left-hand side of vertical dotted line, high-risk group: right-hand side of vertical dotted line). B The survival status of GSE30929 cohort based on the risk score. C PCA for TCGA-SARC cohort based on the risk score. D PCA for GSE30929 cohort based on the risk score. E t-SNE for TCGA-SARC cohort based on the risk score. F t-SNE for GSE30929 cohort based on the risk score.

Supplementary Figure S4 | The correlation of prognostic DEGs of the gene signature and immune infiltration. The correlation between the abundance of immune cells and the expression of A CTSG, B DUSP9, C CLEC10A, D CPA3, ECD1C in TCGA-SARC cohort.

REFERENCES

Bai, D., Feng, H., Yang, J., Yin, A., Lin, X., Qian, A., et al. (2021). Genomic Analysis Uncovers Prognostic and Immunogenic Characteristics of Ferroptosis for clear Cell Renal Cell Carcinoma. *Mol. Ther. - Nucleic Acids* 25, 186–197. doi:10.1016/j.omtn.2021.05.009

Bergsbaken, T., Fink, S. L., and Cookson, B. T. (2009). Pyroptosis: Host Cell Death and Inflammation. *Nat. Rev. Microbiol.* 7, 99–109. doi:10.1038/nrmicro2070

Brannon, A. R., Reddy, A., Seiler, M., Arreola, A., Moore, D. T., Pruthi, R. S., et al. (2010). Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. *Genes & Cancer* 1, 152–163. doi:10.1177/1947601909359929

Brennan, M. F., Antonescu, C. R., Moraco, N., and Singer, S. (2014). Lessons Learned from the Study of 10,000 Patients with Soft Tissue Sarcoma. *Ann. Surg.* 260, 416–422. doi:10.1097/sla.0000000000000869

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., et al. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking* 13, 311–319. doi:10.1089/bio.2015.0032

Caunt, C. J., and Keyse, S. M. (2013). Dual-specificity MAP Kinase Phosphatases (MKPs). *FEBS J.* 280, 489–504. doi:10.1111/j.1742-4658.2012.08716.x

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.cd-12-0095

Chen, K., Gorgen, A., Ding, A., Du, L., Jiang, K., Ding, Y., et al. (2021). Dual-Specificity Phosphatase 9 Regulates Cellular Proliferation and Predicts Recurrence after Surgery in Hepatocellular Carcinoma. *Hepatol. Commun.* 5, 1310–1328. doi:10.1002/hep4.1701

Crago, A. M., and Brennan, M. F. (2015). Principles in Management of Soft Tissue Sarcoma. *Adv. Surg.* 49, 107–122. doi:10.1016/j.yasu.2015.04.002

Ding, J., Wang, K., Liu, W., She, Y., Sun, Q., Shi, J., et al. (2016). Pore-forming Activity and Structural Autoinhibition of the Gasdermin Family. *Nature* 535, 111–116. doi:10.1038/nature18590

Fajkovic, H., Cha, E. K., Xylinas, E., Rink, M., Pycha, A., Seitz, C., et al. (2013). Disease-free Survival as a Surrogate for Overall Survival in Upper Tract Urothelial Carcinoma. *World J. Urol.* 31, 5–11. doi:10.1007/s00345-012-0939-5

Feng, S., Fox, D., and Man, S. M. (2018). Mechanisms of Gasdermin Family Members in Inflammasome Signaling and Cell Death. *J. Mol. Biol.* 430, 3068–3080. doi:10.1016/j.jmb.2018.07.002

Gamboa, A. C., Gronchi, A., and Cardona, K. (2020). Soft-tissue Sarcoma in Adults: An Update on the Current State of Histiotype-specific Management in an Era of Personalized Medicine. *CA A. Cancer J. Clin.* 70, 200–229. doi:10.3322/caac.21605

Ghiringhelli, F., Apetoh, L., Tesniere, A., Aymeric, L., Ma, Y., Ortiz, C., et al. (2009). Activation of the NLRP3 Inflammasome in Dendritic Cells Induces IL-1β-dependent Adaptive Immunity against Tumors. *Nat. Med.* 15, 1170–1178. doi:10.1038/nm.2028

Goldman, M. J., Craft, B., Hastie, M., Repečka, K., Mcdade, F., Kamath, A., et al. (2020). Visualizing and Interpreting Cancer Genomics Data via the Xena Platform. *Nat. Biotechnol.* 38, 675–678. doi:10.1038/s41587-020-0546-8

Hachim, M. Y., Khalil, B. A., Elemam, N. M., and Maghazachi, A. A. (2020). Pyroptosis: The Missing Puzzle Among Innate and Adaptive Immunity Crosstalk. *J. Leukoc. Biol.* 108, 323–338. doi:10.1002/jlb.3mir0120-625r

Hergueta-Redondo, M., Sarrió, D., Molina-Crespo, Á., Megias, D., Mota, A., Rojo-Sebastian, A., et al. (2014). Gasdermin-B Promotes Invasion and Metastasis in Breast Cancer Cells. *PLoS One* 9, e90099. doi:10.1371/journal.pone.0090099

Hoven-Gondrie, M. L., Bastiaannet, E., Ho, V. K. Y., Van Leeuwen, B. L., Liefers, G.-J., Hoekstra, H. J., et al. (2016). Worse Survival in Elderly Patients with Extremity Soft-Tissue Sarcoma. *Ann. Surg. Oncol.* 23, 2577–2585. doi:10.1245/s10434-016-5158-7

Hu, C., Chen, B., Huang, Z., Liu, C., Ye, L., Wang, C., et al. (2020). Comprehensive Profiling of Immune-Related Genes in Soft Tissue Sarcoma Patients. *J. Transl Med.* 18, 337. doi:10.1186/s12967-020-02512-8

Huang, W., Duan, Y., Yang, X., Shang, C., Chen, X., Zhang, H., et al. (2021). Identification of Novel Prognostic Risk Signatures of Soft Tissue Sarcoma Based on Ferroptosis-Related Genes. *Front. Oncol.* 11, 629868. doi:10.3389/fonc.2021.629868

Ju, A., Tang, J., Chen, S., Fu, Y., and Luo, Y. (2021). Pyroptosis-Related Gene Signatures Can Robustly Diagnose Skin Cutaneous Melanoma and Predict the Prognosis. *Front. Oncol.* 11, 709077. doi:10.3389/fonc.2021.709077

Kang, W., Hong, S. H., Lee, H. M., Kim, N. Y., Lim, Y. C., Le, L. T. M., et al. (2014). Structural and Biochemical Basis for the Inhibition of Cell Death by APIP, a Methionine Salvage Enzyme. *Proc. Natl. Acad. Sci. USA* 111, E54–E61. doi:10.1073/pnas.1308768111

Lamkanfi, M., and Dixit, V. M. (2014). Mechanisms and Functions of Inflammasomes. *Cell* 157, 1013–1022. doi:10.1016/j.cell.2014.04.007

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48, W509–w514. doi:10.1093/nar/gkaa407

Lin, W., Chen, Y., Wu, B., Chen, Y., and Li, Z. (2021). Identification of the Pyroptosis related Prognostic Gene Signature and the Associated Regulation axis in Lung Adenocarcinoma. *Cell Death Discov.* 7, 161. doi:10.1038/s41420-021-00557-2

Lowe, S. W., and Lin, A. W. (2000). Apoptosis in Cancer. *Carcinogenesis* 21, 485–495. doi:10.1093/carcin/21.3.485

Lu, H., Tran, L., Park, Y., Chen, I., Lan, J., Xie, Y., et al. (2018). Reciprocal Regulation of DUSP9 and DUSP16 Expression by HIF1 Controls ERK and P38 MAP Kinase Activity and Mediates Chemotherapy-Induced Breast Cancer Stem Cell Enrichment. *Cancer Res.* 78, 4191–4202. doi:10.1158/0008-5472.can-18-0270

Nagarajan, K., Soundarapandian, K., Thorne, R. F., Li, D., and Li, D. (2019). Activation of Pyroptotic Cell Death Pathways in Cancer: An Alternative Therapeutic Approach. *Translational Oncol.* 12, 925–931. doi:10.1016/j.tranon.2019.04.010

Navarria, P., Ascolese, A. M., Cozzi, L., Tomatis, S., D'Agostino, G. R., De Rose, F., et al. (2015). Stereotactic Body Radiation Therapy for Lung Metastases from Soft Tissue Sarcoma. *Eur. J. Cancer* 51, 668–674. doi:10.1016/j.ejca.2015.01.061

Oba, K., Paoletti, X., Alberts, S., Bang, Y.-J., Benedetti, J., Bleiberg, H., et al. (2013). Disease-free Survival as a Surrogate for Overall Survival in Adjuvant Trials of Gastric Cancer: a Meta-Analysis. *J. Natl. Cancer Inst.* 105, 1600–1607. doi:10.1093/jnci/djt270

Pagès, F., Galon, J., Dieu-Nosjean, M.-C., Tartour, E., Sautès-Fridman, C., and Fridman, W.-H. (2010). Immune Infiltration in Human Tumors: a Prognostic Factor that Should Not Be Ignored. *Oncogene* 29, 1093–1102. doi:10.1038/onc.2009.416

Reed, J. C. (2000). Mechanisms of Apoptosis. *Am. J. Pathol.* 157, 1415–1430. doi:10.1016/s0002-9440(10)64779-7

Ruan, J., Wang, S., and Wang, J. (2020). Mechanism and Regulation of Pyroptosis-Mediated in Cancer Cell Death. *Chem. Biol. Interact* 323, 109052. doi:10.1016/j.cbi.2020.109052

Shi, J., Gao, W., and Shao, F. (2017). Pyroptosis: Gasdermin-Mediated Programmed Necrotic Cell Death. *Trends Biochem. Sci.* 42, 245–254. doi:10.1016/j.tibs.2016.10.004

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by Inflammatory Caspases Determines Pyroptotic Cell Death. *Nature* 526, 660–665. doi:10.1038/nature15514

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49, D605–d612. doi:10.1093/nar/gkaa1074

Thi, H. T. H., and Hong, S. (2017). Inflammasome as a Therapeutic Target for Cancer Prevention and Treatment. *J. Cancer Prev.* 22, 62–73. doi:10.15430/jcp.2017.22.2.62

Tsuchiya, K., Nakajima, S., Hosojima, S., Thi Nguyen, D., Hattori, T., Manh Le, T., et al. (2019). Caspase-1 Initiates Apoptosis in the Absence of Gasdermin D. *Nat. Commun.* 10, 2091. doi:10.1038/s41467-019-09753-2

Vilanova, J. C. (2017). "WHO Classification of Soft Tissue Tumors," in *Imaging of Soft Tissue Tumors* (Springer), 187–196. doi:10.1007/978-3-319-46679-8_11

Wang, C., Machiraju, R., and Huang, K. (2014). Breast Cancer Patient Stratification Using a Molecular Regularized Consensus Clustering Method. *Methods* 67, 304–312. doi:10.1016/j.ymeth.2014.03.005

Wang, Q., Armenia, J., Zhang, C., Penson, A. V., Reznik, E., Zhang, L., et al. (2018). Unifying Cancer and normal RNA Sequencing Data from Different Sources. *Sci. Data* 5, 180061. doi:10.1038/sdata.2018.61

Wang, Y.-Y., Liu, X.-L., and Zhao, R. (2019). Induction of Pyroptosis and its Implications in Cancer Management. *Front. Oncol.* 9, 971. doi:10.3389/fonc.2019.00971

Wei, Q., Mu, K., Li, T., Zhang, Y., Yang, Z., Jia, X., et al. (2014). Deregulation of the NLRP3 Inflammasome in Hepatic Parenchymal Cells during Liver Cancer Progression. *Lab. Invest.* 94, 52–62. doi:10.1038/labinvest.2013.126

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Xia, X., Wang, X., Cheng, Z., Qin, W., Lei, L., Jiang, J., et al. (2019). The Role of Pyroptosis in Cancer: Pro-cancer or Pro-"host"? *Cell Death Dis* 10, 650. doi:10.1038/s41419-019-1883-8

Xue, Y., Enosi Tuipulotu, D., Tan, W. H., Kay, C., and Man, S. M. (2019). Emerging Activators and Regulators of Inflammasomes and Pyroptosis. *Trends Immunol.* 40, 1035–1052. doi:10.1016/j.it.2019.09.005

Ye, Y., Dai, Q., and Qi, H. (2021). A Novel Defined Pyroptosis-Related Gene Signature for Predicting the Prognosis of Ovarian Cancer. *Cel Death Discov.* 7, 71. doi:10.1038/s41420-021-00451-x

Zhang, Z., Zhang, Y., Xia, S., Kong, Q., Li, S., Liu, X., et al. (2020). Gasdermin E Suppresses Tumour Growth by Activating Anti-tumour Immunity. *Nature* 579, 415–420. doi:10.1038/s41586-020-2071-9

Zheng, M., and Kanneganti, T. D. (2020). The Regulation of the ZBP1-NLRP3 Inflammasome and its Implications in Pyroptosis, Apoptosis, and Necroptosis (PANoptosis). *Immunol. Rev.* 297, 26–38. doi:10.1111/imr.12909

Zhou, C.-B., and Fang, J.-Y. (2019). The Role of Pyroptosis in Gastrointestinal Cancer and Immune Responses to Intestinal Microbial Infection. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1872, 1–10. doi:10.1016/j.bbcan.2019.05.001

Zhu, S., Ding, S., Wang, P., Wei, Z., Pan, W., Palm, N. W., et al. (2017). Nlrp9b Inflammasome Restricts Rotavirus Infection in Intestinal Epithelial Cells. *Nature* 546, 667–670. doi:10.1038/nature22967

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Qi, Xu, Wan, Ren, Zhang, Zhang, Tu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fgene.2021.814480

A Prognostic Model for Breast Cancer Based on Cancer Incidence-Related DNA Methylation Pattern

Zhenchong Xiong[1†], Lin Yang[2†], Juan Ao[3†], Jiarong Yi[1], Xiazi Zouxu[1], Wenjing Zhong[1], Jikun Feng[1], Weiling Huang[1], Xi Wang[1] and Zeyu Shuang[1]**

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Department of Breast Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China, [2]Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, China, [3]Department of Neurology, Guangzhou First People's Hospital, Guangzhou, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Tianzhu Lu,
Jiangxi Cancer Hospital, China
San-Gang Wu,
First Affiliated Hospital of Xiamen University, China
Si-Qi Qiu,
Shantou Central Hospital, China

***Correspondence:**
Xi Wang
wangxi@sysucc.org.cn
Zeyu Shuang
shuangzy@sysucc.org.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 13 November 2021
Accepted: 13 December 2021
Published: 03 January 2022

Citation:
Xiong Z, Yang L, Ao J, Yi J, Zouxu X, Zhong W, Feng J, Huang W, Wang X and Shuang Z (2022) A Prognostic Model for Breast Cancer Based on Cancer Incidence-Related DNA Methylation Pattern. Front. Genet. 12:814480. doi: 10.3389/fgene.2021.814480

Breast cancer (BC) is the most diagnosed cancer and the leading cause of cancer-related deaths in women. The purpose of this study was to develop a prognostic model based on BC-related DNA methylation pattern. A total of 361 BC incidence-related probes (BCIPs) were differentially methylated in blood samples from women at high risk of BC and BC tissues. Twenty-nine of the 361 BCIPs that significantly correlated with BC outcomes were selected to establish the BCIP score. BCIP scores based on BC-related DNA methylation pattern were developed to evaluate the mortality risk of BC. The correlation between overall survival and BCIP scores was assessed using Kaplan–Meier, univariate, and multivariate analyses. In BC, the BCIP score was significantly correlated with malignant BC characteristics and poor outcomes. Furthermore, we assessed the BCIP score-related gene expression profile and observed that genes with expressions associated with the BCIP score were involved in the process of cancer immunity according to GO and KEGG analyses. Using the ESTIMATE and CIBERSORT algorithms, we discovered that BCIP scores were negatively correlated with both T cell infiltration and immune checkpoint inhibitor response markers in BC tissues. Finally, a nomogram comprising the BCIP score and BC prognostic factors was used to establish a prognostic model for patients with BC, while C-index and calibration curves were used to evaluate the effectiveness of the nomogram. A nomogram comprising the BCIP score, tumor size, lymph node status, and molecular subtype was developed to quantify the survival probability of patients with BC. Collectively, our study developed the BCIP score, which correlated with poor outcomes in BC, to portray the variation in DNA methylation pattern related to BC incidence.

Keywords: breast cancer, prognosis, DNA methylation, immune, nomogram

Abbreviations: BC, breast cancer; BCIPs, breast cancer incidence-related probes; c-index, the concordance index; CAFs, cancer-associated fibroblasts; Chr, Chromosome; DMP, differentially methylated probe; DNMTs, DNA methyltransferases; knn, k-nearest neighbors; ER, estrogen receptor; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; HRs, hazard ratios; ICI, immune checkpoint inhibitor; IFNG, IFN-γ signature; IFNG.GS, IFNG hallmark gene set; IFN.ISGs, IFN stimulated genes resistance signature; KEGG, Kyoto Encyclopedia of Genes and Genomes; MDSCs, myeloid-derived suppressor cells; PR, progesterone receptor; TNBC, triple-negative breast cancer; SNPs, single nucleotide polymorphisms; TCGA, The Cancer Genome Atlas.

INTRODUCTION

Breast cancer (BC), the most diagnosed cancer globally, is the leading cause of cancer-related deaths in women (Siegel et al., 2020). As the incidence of BC has continuously increased over recent years, BC has become a major public health problem, and one in eight women would be affected by BC by the age of 85 years in high-income countries (Britt et al., 2020). With a deep understanding of cancer biology, BC is subdivided into four molecular subtypes (luminal A, luminal B, HER2, and TNBC) based on the expression of the estrogen receptor (ER), progesterone receptor (PR), HER2 receptor, and Ki-67 (Abreu et al., 2020). Although BC therapy guided by molecular subtyping greatly reduces BC mortality, cancer recurrence inevitably occurs (Alwan and Tawfeeq, 2019). Investigations of prognostic methods for predicting outcomes in BC may provide clues for improving cancer treatment.

DNA methylation, an epigenetic modification, plays an important role in cancer development (Jones et al., 2016). Although previous studies have shown that BC is a genetic disease, genome-wide variations in DNA methylation have been observed in cancer cells (Flavahan et al., 2017; Widschwendter et al., 2018). DNA methylation induces chromatin structure changes and inhibits gene expression mediated by DNA methyltransferases. Aberrations in DNA methylation, which are affected by environmental, lifestyle-related, and heritable factors, may induce cancer development by silencing tumor suppressors and/or re-activating oncogenes (Baylin and Jones, 2011; Koo et al., 2015; Muvarak et al., 2016). In BC, hypomethylation of stemness- and proliferation-associated genes in circulating tumor cells promotes stemness and metastasis (Gkountela et al., 2019). Thienpont et al. indicated that tumor hypoxia induces hypermethylation and promotes BC progression by inactivating TET enzymes (Thienpont et al., 2016). Thus, tracking the changes in DNA methylation pattern associated with BC is helpful for developing prognostic methods for BC.

The purpose of our study was to discover DNA methylation pattern and develop a prognostic model based on BC-related DNA methylation pattern. We identified 361 breast cancer incidence-related probes (BCIPs) that were differentially methylated in blood samples from women at a high risk of BC and BC tissues. Twenty-nine of the 361 BCIPs that were significantly correlated with BC outcomes were included to establish the BCIP score. In BC, the BCIP scores were significantly correlated with malignant BC characteristics and poor outcomes. Furthermore, we assessed the BCIP score-related gene expression profile and observed that the BCIP score-related gene profile participated in the process of cancer immunity. BCIP scores were negatively correlated with immune cell infiltration and the immune checkpoint inhibitor (ICI) response in BC tissues. Finally, a nomogram comprising the BCIP score, tumor size, lymph node status, and molecular subtype was developed to quantify the survival probability of patients with BC.

MATERIALS AND METHODS

Data Collection and Processing

For the GSE51057, GSE72308, and TCGA-BRCA DNA methylation datasets, genome-wide methylation data were profiled using Illumina Infinium HumanMethylation450 BeadChips Assay (Illumina 450 K platform). For the GSE57285 DNA methylation dataset, genome-wide methylation data were profiled using Illumina Infinium HumanMethylation27 BeadChips Assay (Illumina 27 K platform). The DNA methylation level of each probe was calculated using β values ranging from 0 (no DNA methylation) to 1 (complete DNA methylation). Probes containing missing values in over half of the samples in each dataset were removed, while missing values of the remaining probes were imputed with the k-nearest neighbors imputation method. Probes located on the sex chromosome and probes containing known single-nucleotide polymorphisms were removed (Price et al., 2013). Finally, 23,614 probes were selected for further investigation. The above process was performed using the R package Champ (Tian et al., 2017).

For gene expression data, mRNA expression data were obtained from the Cancer Genome Atlas (TCGA) database. Background correction and normalization of mRNA expression data were performed using the R package limma (Ritchie et al., 2015). Expression data for protein-encoding genes were included in further analysis.

Calculation of the BCIP Score

GSE51057 (including 177 blood samples from normal women and 146 blood samples from women diagnosed with BC after sample donation) and GSE57285 (including 49 blood samples from normal women and 35 blood samples from women diagnosed with BC after sample donation) were selected to identify differentially methylated probes (DMPs) that correlated with a high risk of BC. In addition, 76 cases with matched tumor and tumor-adjacent breast tissues from TCGA database were enrolled to identify DMPs in BC tissues. CpG probes that were commonly demethylated in blood samples from women with high risk of BC and BC tissues were defined as BCIPs.

DMPs were identified using the R package limma. Differential hyper/hypo-methylation probe was defined according to logFc value. Hyper-methylation probe are defined as logFc >0, p value <0.05 (blood/cancer sample of BC patients VS blood/cancer sample of non-BC patients). Hypo-methylation probe are defined as logFc <0, p value <0.05 (blood/cancer sample of BC patients VS blood/cancer sample of non-BC patients). The distribution of BCIPs on chromosomes, CpG islands, and TSS regions was investigated using the R package Champ. The hazard ratios (HRs) of BCIPs with respect to OS were evaluated using TCGA-BRCA data.

Univariate Cox regression was used to calculate the HR of each BCIP, and BCIPs significantly correlated cancer survival in BC were included to develop the BCIP score model. The BCIP score model was assessed as follows: BCIP score = [(transformed HR^1 *β value of $BCIP^1$) + (transformed HR^2 *β value of $BCIP^2$) + ····· (transformed HR^n *β value of $BCIP^n$)]/[abs (transformed HR^1) +

TABLE 1 | Characteristics of patients in the training and validation cohort.

Variate	Training cohort (n = 587) Num (%)	Validation cohort (n = 231) Num (%)
Age		
0–39	37 (6.3)	28 (12.1)
≥40	550 (93.7)	203 (87.8)
Tumor size		
≤2 cm	154 (26.2)	117 (50.6)
>2 cm	433 (73.8)	114 (49.4)
Lymph node status		
Non-metastasis	261 (44.5)	130 (56.3)
Metastasis	326 (55.5)	101 (43.7)
Subtype (IHC)		
Luminal A/B	504 (85.9)	111 (48.1)
HER2	16 (2.7)	55 (23.8)
TNBC	67 (11.4)	65 (28.1)
BCIP score		
Low	425 (72.4)	146 (63.2)
High	162 (27.6)	85 (36.8)
Survival status		
Alive	536 (91.3)	189 (81.8)
Dead	51 (8.7)	42 (18.2)

abs (transformed HR^2) + ····· (transformed HR^n)]. The cutoff value of BCIP score was 0.2, identified using x-tile (https://medicine.yale.edu/lab/rimm/research/software/). Samples with a BCIP score of <0.2 were assigned to the low BCIP group, while those with a BCIP score of ≥0.2 were assigned to the high BCIP group.

Functional and Clinical Characteristics Analysis of the BCIP Score in BC

BCIP scores of TCGA-BRCA tissues (76 cases with matched tumor and tumor-adjacent breast tissues and 699 cases with unmatched tumor tissues) were calculated using DNA methylation data. The correlation between the BCIP score and BC-related characteristics (tumor size, oncogene copy number variation, and oncogene expression) was assessed using linear regression and Spearman's correlation coefficient. The correlation between gene mRNA expression and the BCIP score was analyzed using Spearman's correlation. The gene expression profile associated with BCIP score was identified based on Spearman's coefficient (cutoff value: 0.1), and functional study of the related gene expression was performed using the GO and KEGG databases. The above procedure was performed using R software.

Correlation Analysis Between the BCIP Score and Immune Microenvironment in BC

The degree of infiltration of immune cells and stromal cells in each sample was assessed using the ESTIMATE algorithm (Yoshihara et al., 2013). The proportion of 22 immune cells in each tissue was evaluated using the CIBERSORT algorithm (http://cibersort.stanford.edu/) (Gentles et al., 2015). Correlations between the BCIP score and ESTIMATE and CIBERSORT scores were assessed using linear regression and Spearman's correlation coefficient.

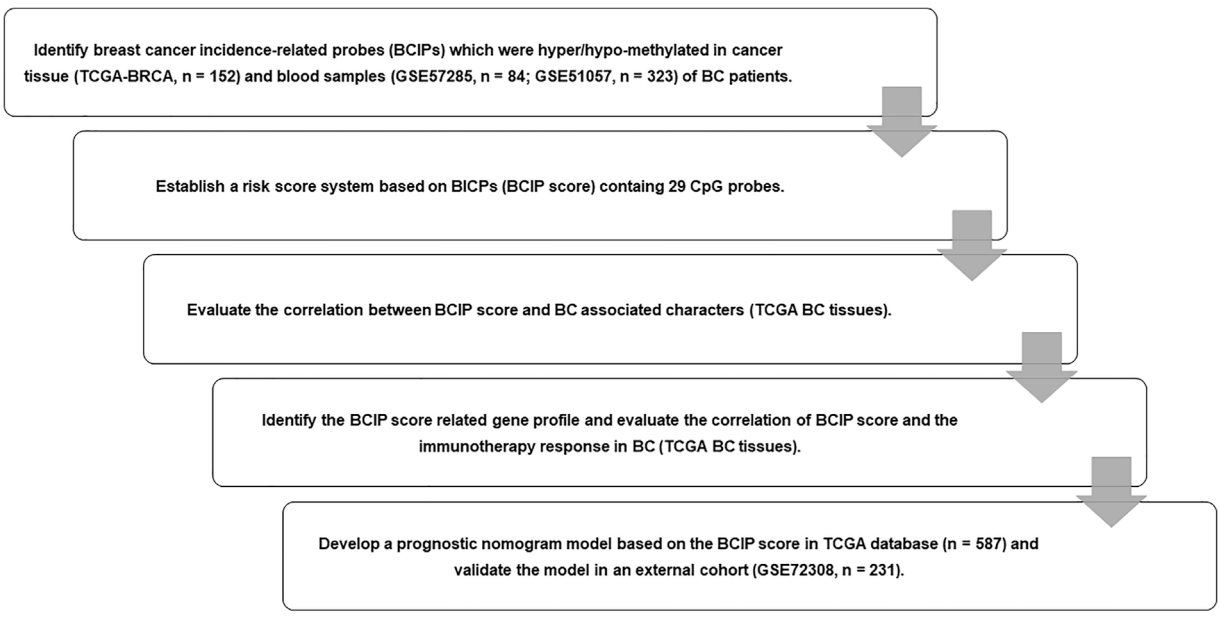


FIGURE 1 | Flow chart of study design. We identified BCIPs that were commonly hypermethylated or hypomethylated in blood samples and cancer tissues in patients with BC; 29 BCIPs that significantly correlated with BC patient survival were selected to develop the BCIP score. The correlation between the BCIP score and clinical characteristics of BC was assessed. We then evaluated the BCIP score-related gene profile and the relationship between the BCIP score and tumor immune response in BC tissues. Furthermore, we assessed the prognostic effect of the BCIP score and developed a prognostic prediction model based on the BCIP score using TCGA database. The efficacy of the prognostic prediction model was validated using the GEO cohort.

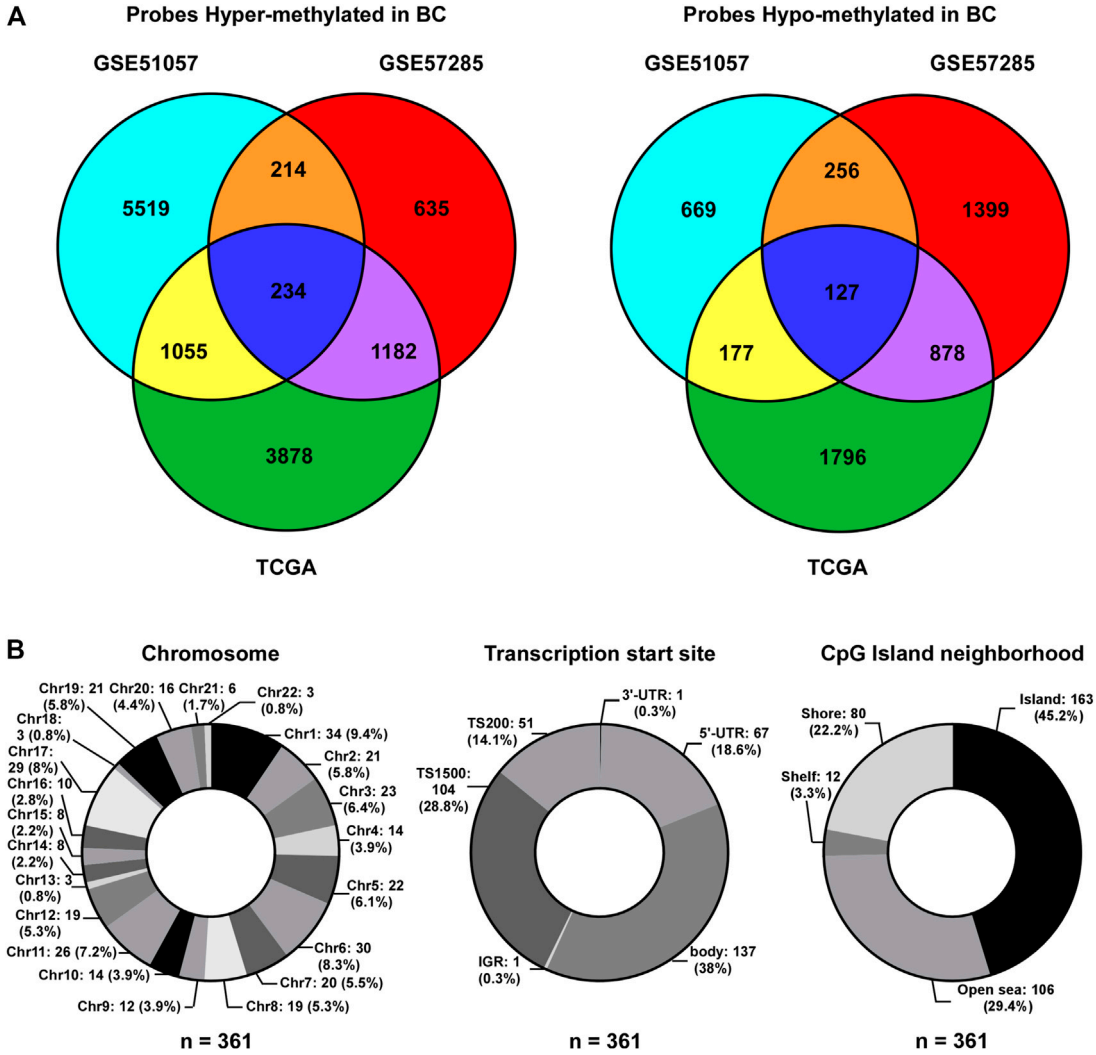


FIGURE 2 | Identification of BCIPs. **(A)** Venn diagram of probes commonly hypermethylated (left panel) or hypomethylated (right panel) in blood samples and cancer tissues from patients with BC. For the GSE51057 and GSE57285 datasets: cases diagnosed with BC vs. normal cases; for TCGA: BC tissues vs. tumor-adjacent tissues. **(B)** Distribution of BCIPs referring to **(Left)** chromosome, **(Middle)** transcription Start Sites, CpG island neighborhood are listed as number (proportion).

Correlation Analysis Between the BCIP Score and Cancer Immunotherapy Response

Four biomarkers were used to assess the response to immunotherapy—CD274, CD8, IFN-γ signature (IFNG) (Ayers et al., 2017), and IFNG hallmark gene set (IFNG.GS) (Benci et al., 2019). Three biomarkers were used to assess resistance to immunotherapy—IFN-stimulated gene resistance signature (ISG.RS) (Benci et al., 2019), myeloid-derived suppressor cells (MDSCs), and cancer-associated fibroblasts (CAFs) (Joyce and Fearon, 2015). IFNG was calculated by averaging six genes (IFNG, STAT1, IDO1, CXCL9, CXCL10, and HLA-DRA). IFNG. GS was calculated as previous reported (Benci et al., 2019). CD274, CD8, ISG. RS, MDSCs, and CAFs were assessed using a web application (http://tide.dfci.harvard.edu).

Establishment and Validation of the Nomogram

A total of 587 BC cases with DNA methylation data, clinical characteristics, and complete follow-up in TCGA database were enrolled as the training cohort, while 231 BC cases with data of DNA methylation, clinical characteristics, and complete follow-up in the GSE72308 dataset were selected as the external validation cohort. The clinical and pathological characteristics of patients with BC are listed in **Table 1**. Tumor size was defined as ≤2 cm or >2 cm. Lymph node status was defined as non-metastasis or lymph node metastasis. The molecular subtypes were defined based on the IHC assessment of ER, PR, and HER2 as follows: luminal A/B (ER+/PR+/HER2−, ER−/PR+/HER2−, ER+/PR−/HER2−, ER+/PR+/HER2+, ER-/PR+/HER2+, ER+/PR-/HER2+),

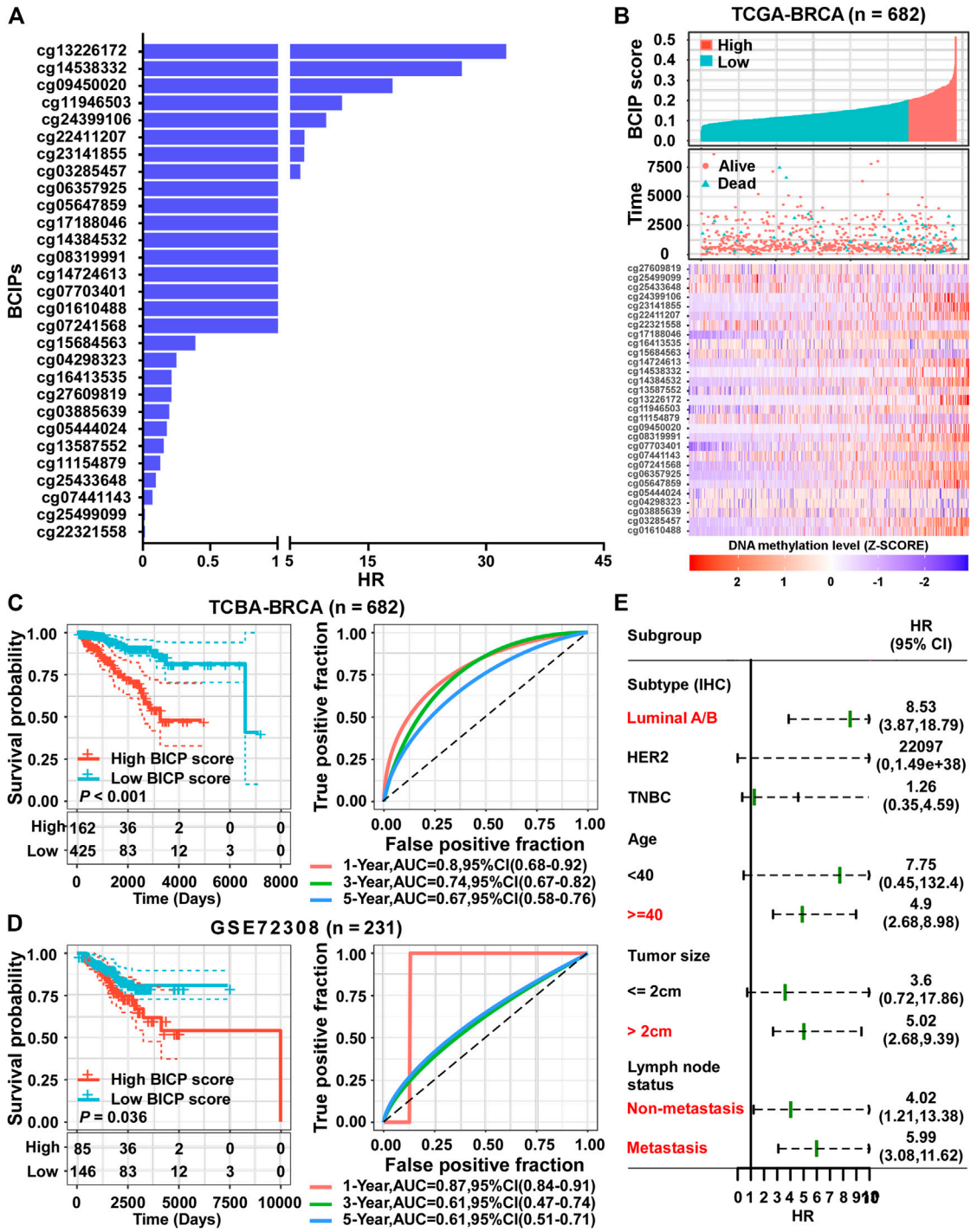


FIGURE 3 | Establishment of a prognostic risk score based on BCIPs for BC. **(A)** HRs of BCIPs were calculated using univariate analysis in TCGA-BRCA cohort (*n* = 682). **(B)** Distribution of BCIP scores (upper panel) in the low-score and high-score subgroups, the death incidence of patients (middle panel), and heatmap of the 29 BCIPs methylation profiles (lower panel) in TCGA cohort. The cutoff value of the BCIP score was identified using x-tile and determined to be 0.2. The DNA methylation levels of BCIPs were normalized using z-score. **(C)** (left panel) Survival analysis of BCIP scores and (right panel) survival prediction ROC curve of the BCIP score in TCGA cohort. **(D)** (left panel) Survival analysis of the BCIP score and (right panel) survival prediction ROC curve of the BCIP score in the GSE72308 cohort. **(E)** Forest plot depicting HRs of BCIP scores in subgroups of TCGA cohort. In subgroups labeled in red, BCIP scores were significantly correlated with overall survival of BC patients. For **(C–D)** (left panel), *p*-values were determined using log-rank test.

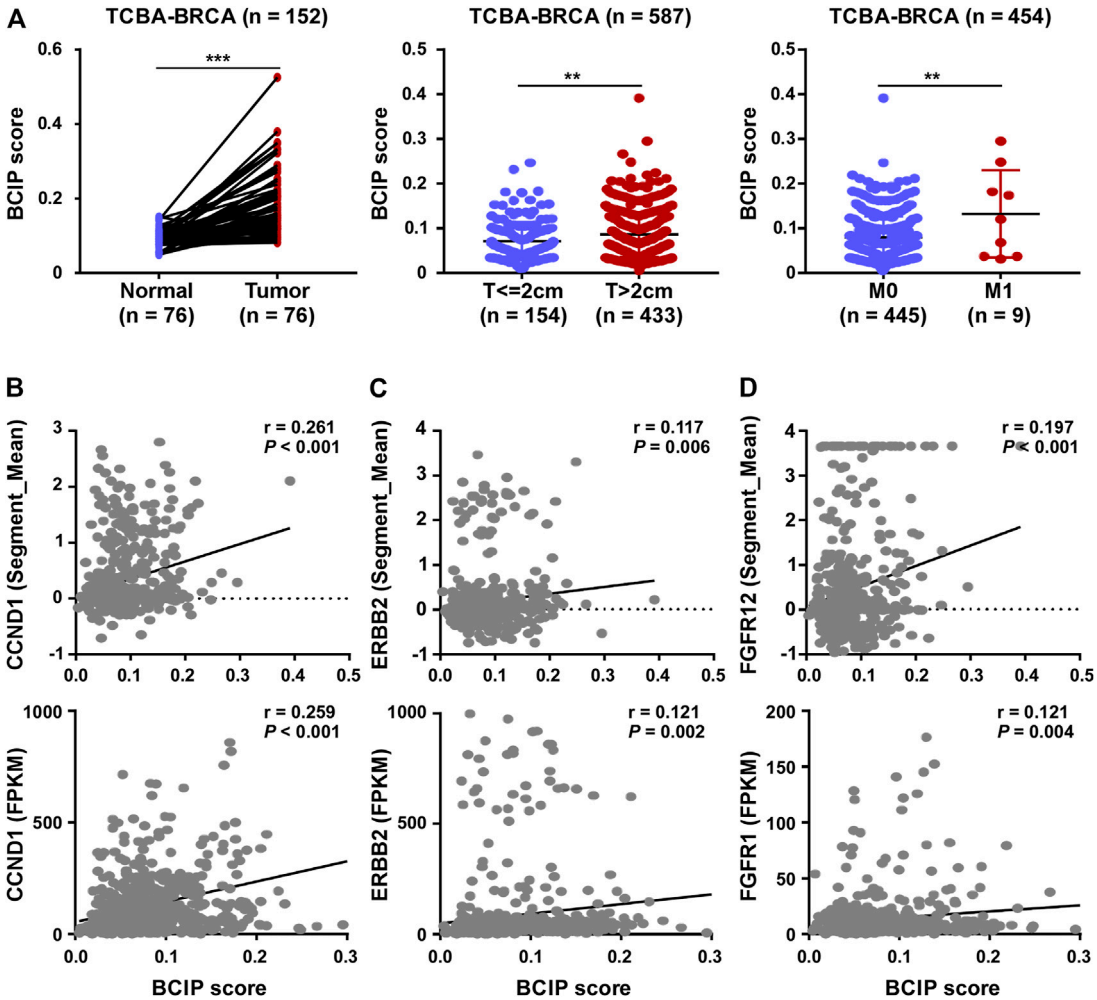


FIGURE 4 | Correlation of BCIP score with BC related characteristics. **(A)** Analysis of BCIP score differences between (left panel) tumor tissues and matched tumor-adjacent breast tissues; (middle panel) cases with tumor size ≤2 cm and cases with >2 cm (right panel) cases with local regional disease (M0) and cases with *de novo* metastasis disease (M1) in TCGA-BRCA cohort. **(B)** Correlations between the BCIP score and (upper panel) copy number and (lower panel) mRNA expression of CCND1. **(C)** Correlations between the BCIP score and (upper panel) copy number and (lower panel) mRNA expression of ERBB2. **(D)** Correlations between the BCIP score and (upper panel) copy number and (lower panel) mRNA expression of FGFR1. For A (left panel), *p*-values were determined by paired *t*-test; for A (middle and right panel), *p*-values were determined by *t*-test; for **(B–D)**, *p*-values were determined by r and Spearman's correlation coefficient.

HER2 (ER−/PR−/HER2+), and TNBC (ER−/PR−/HER2−). Age was defined as 0–39 and ≥40 years.

BC features, which were significantly correlated with BC survival in the multivariate Cox regression, were selected to establish the nomogram model. The patients' survival probability was assumed by summing the scores of the variates, with a higher score corresponding to a higher mortality risk. The efficiency of the model was evaluated with regard to discrimination and calibration. The concordance index (c-index) was used to quantify discrimination ranging from 0 to 1 (<0.5, absolute discordance; 0.5, equal concordance to chance; and 1, best concordance). The calibration curve was used to compare the predicted survival probability with the observed survival probability at 3, 5, and 10 years in the training and validation cohorts.

RESULTS

Identification of BCIPs

To identify CpG probes associated with BC incidence, two GEO datasets (GSE51057 and GSE57285), including 226 blood samples from healthy women (177 cases in GSE51057 and 49 cases in GSE57285) and 181 blood samples from women diagnosed with BC after sample donation (146 cases in GSE51057 and 35 cases in GSE57285; **Figure 1**), were used. The 448 and 383 CpG probes were commonly hypermethylated or hypomethylated in samples from women with BC compared to those in samples from healthy women (**Figure 2A**). Further, we included 76 paired tumor-adjacent breast tissues and tumor tissues from TCGA-BRCA database and identified 9,327 differentially methylated probes (tumor vs. tumor-adjacent;

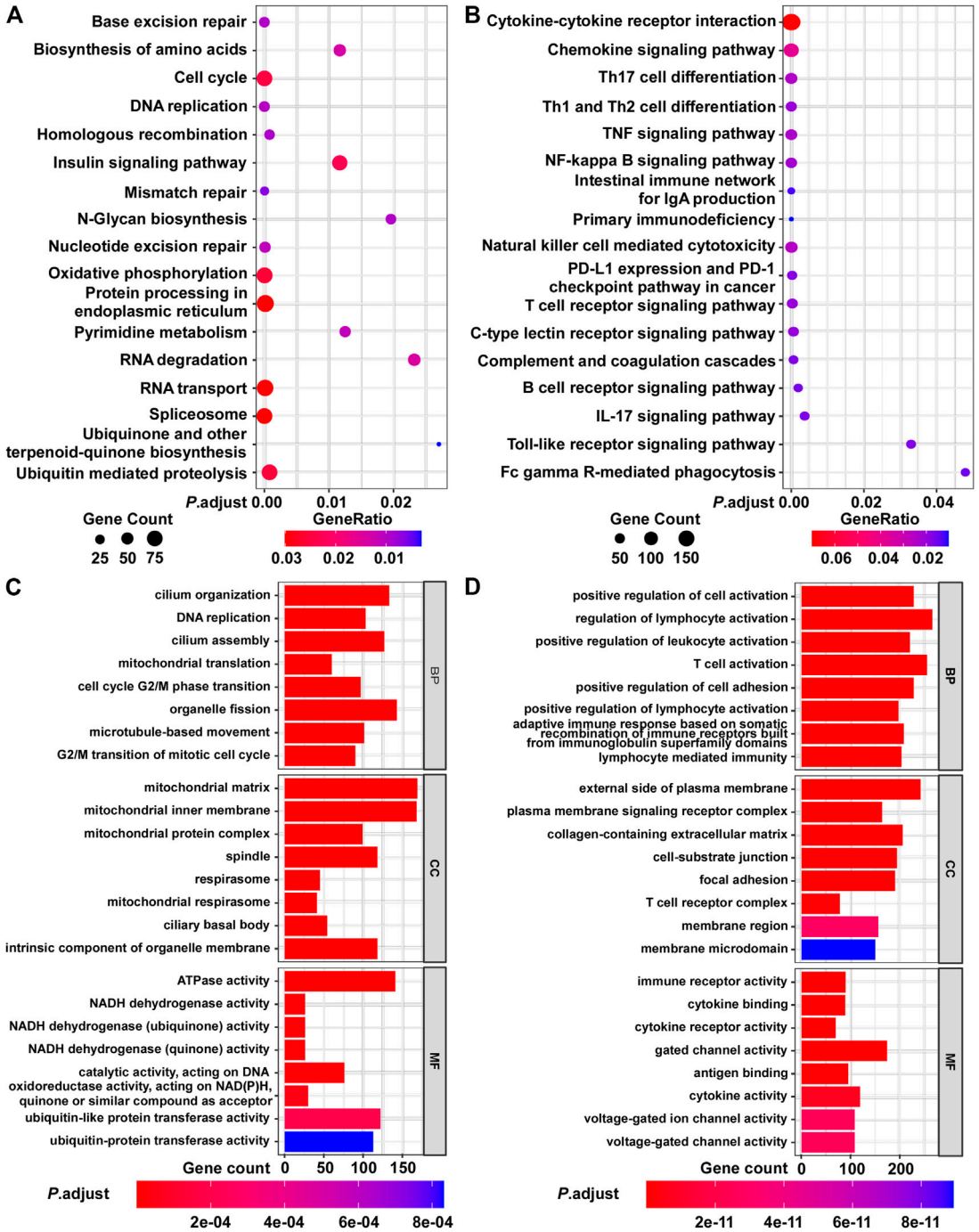


FIGURE 5 | Functional analysis of BCIP score-related gene profile in BC. **(A,B)** Analysis of BCIP score-related gene enrichment in the KEGG pathway. **(A)** Analysis of genes whose mRNA expressions were positively correlated with BCIP score; **(B)** analysis of genes whose mRNA expressions were negatively correlated with BCIP score. Gene ratio was defined as the number of genes enriched to target pathway/number of BCIP score-related gene included in the KEGG dataset. **(C,D)** GO function analysis of BCIP score-related gene. **(C)** Analysis of genes whose mRNA expressions were positively correlated with the BCIP score. **(D)** Analysis of genes whose mRNA expressions were negatively correlated with the BCIP score.

hypermethylated: 6,349; hypomethylated: 2,978). By integrating the results from the GEO and TCGA cohorts, 234 hypermethylated and 127 hypomethylated CpG probes were identified as BCIPs in blood samples from women with high risk of BC and BC tissues (**Figure 2A**). Next, we assessed the distribution of BCIPs based on chromosomes, transcription start sites, and CpG islands. Among the 22 pairs, chromosomes (Chr) 1, 6, 11, and 17 were the most

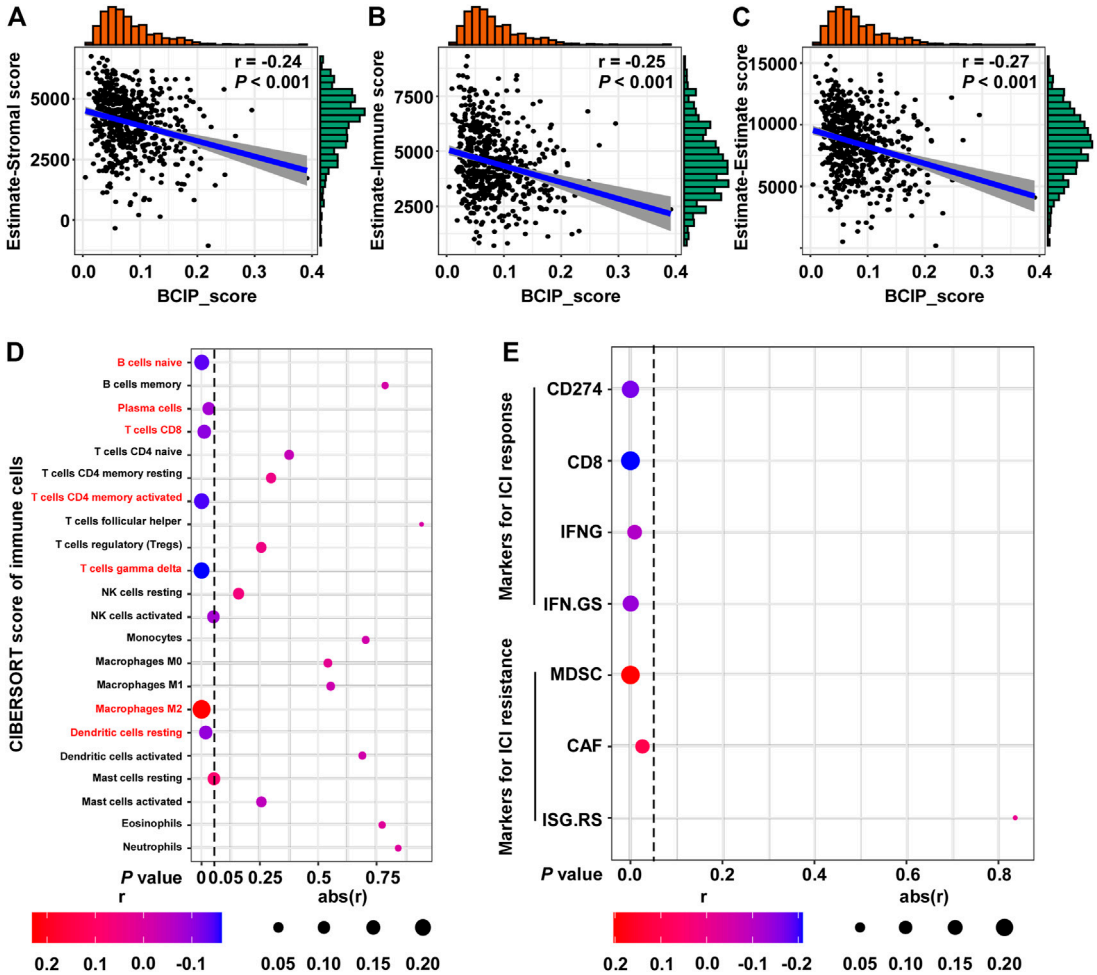


FIGURE 6 | Correlations between the BCIP score and immune microenvironment and ICI response in BC. **(A)** Correlation between the BCIP score and the level of stromal cells (estimate-Stromal score) in BC tissues (TCGA-BRCA, *n* = 587). Numerical distribution of BCIP scores and estimate-Stromal scores are shown on the above the *x*-axis and on the right of the *y*-axis, respectively. **(B)** Correlation between the BCIP score and the level of stromal cells (estimate-Immune score) in BC tissues (TCGA-BRCA, *n* = 587). Numerical distribution of BCIP scores and estimate-Immune scores is shown on the above the *x*-axis and on the right of the *y*-axis, respectively. **(C)** Correlation between the BCIP score and the level of stromal cells (estimate-Estimate score) in BC tissues (TCGA-BRCA, *n* = 587). Numerical distribution of BCIP scores and estimate-Estimate scores are shown on the above the *x*-axis and on the right of the *y*-axis, respectively. **(D)** Correlation between the BCIP score and the 22 type of immune cell components is shown by dot plot. Cell contents correlated with the BCIP score are labeled in red. **(E)** Correlation between DM-BMI and markers for ICI response/resistance is shown by dot plot. r, Spearman's correlation coefficient.

common regions for BCIP distribution; 45.2% were distributed in CpG islands and 42.9% were distributed in the promoter regions (TS200 and TS1500; **Figure 2B**).

Establishment of a Prognostic Risk Score Based on BCIPs for BC

Using a univariate Cox regression model, we evaluated the association between methylation levels of BCIPs and overall survival in TCGA-BRCA cohort. DNA methylation levels of 29/361 BCIPs significantly correlated with OS in BC, and these probes were selected to establish the BCIP score model (**Figure 3A** and **Supplemental Table S1**). Using X-Tile analysis, the cutoff value for the BCIP score was set at 0.2; patients with a BCIP score ≤0.2 were assigned to the low score group, while patients with a BCIP score of >0.2 were assigned to the high score group (**Figure 3B**). In TCGA cohort, patients with low BCIP scores had better survival rates than those in the high score group (**Figure 3C**). The ROC curve showed that the BCIP score exhibited a high predictive efficacy for OS in BC (**Figure 3C**). Likewise, a high BCIP score predicted a high mortality risk for BC in the GEO cohort (**Figure 3D**). Further, subgroup analysis indicated that the BCIP score was a negative prognostic factor in subgroups of BC with luminal A/B subtype, old age (≥40 years), larger tumor size, and lymph node metastasis status (**Figure 3E** and **Supplemental Table S2**). These data show that the BCIP score is an efficient prognostic model for BC.

TABLE 2 | Univariate and multivariate analysis for patients with BC in TCGA cohort.

Variate	Univariate			Multivariate		
	HR	95% CI	*p* Value	HR	95% CI	*p* Value
Age			0.339			
Age	1					
0–39	1.799	0.54–5.989				
≥40			0.04			0.045
Tumor size	1			1		
≤2 cm	2.467	1.044–5.83		2.41	1.021–5.688	
>2 cm			<0.001			<0.001
Lymph node status	1			1		
Non-metastasis	4.18	2.087–8.371		3.959	1.997–7.848	
Subtype (IHC)						<0.001
Luminal A/B	1		<0.001	1		
HER2	3.39	0.789–14.561		3.445	0.801–14.816	
TNBC	6.034	3.016–12.073		5.561	2.833–10.918	
BCIP score			<0.001			<0.001
Low	1			1		
High	6.618	3.628–12.073		6.748	3.701–12.304	

Functional and Clinical Characteristic Analysis of the BCIP Score in BC

Compared to tumor-adjacent breast tissues, BC tissues exhibited a significantly higher BCIP score (**Figure 4A**). In BC, patients with a larger tumor size and *de novo* metastatic disease had a higher BCIP score (**Figure 4A**). By assessing the gene copy number in BC tissues, the BCIP score was associated with an increased copy number of several oncogenes (including *CCND1*, *ERBB2*, and *FGFR1*; **Figures 4B–D**). Consistently, the BCIP score was positively correlated with the mRNA levels of *CCND1*, *ERBB2*, and *FGFR1* (**Figures 4B–D**).

Next, we evaluated the biological significance of the BCIP score in BC. By assessing the correlation between the BCIP score and gene expression, we identified the BCIP score-related gene expression profile. The results of correlation analysis are shown in **Supplementary Table S3**. Genes with expression correlated with the BCIP score were included in GSEA analysis using the KEGG and GO databases. KEGG and GO analyses showed that genes with expression that positively correlated with the BCIP score were significantly enriched in pathways involving cell cycle regulation, DNA replication, DNA repair (base excision repair, mismatch repair, nucleotide excision repair, and homologous recombination), and energy metabolism (**Figures 5A,C**). Interestingly, both KEGG and GO analyses showed that genes with expressions that negatively correlated with the BCIP score were significantly involved in cancer-immunity-related pathways (including antigen binding, T cell differentiation and activation, the PD-1 checkpoint pathway, NK cell-mediated cytotoxicity, and immune receptor activity) (**Figures 5B,D**).

Correlations Between the BCIP Score and Immune Microenvironment and ICI Response in BC

Further, we used the ESTIMATE algorithm to evaluate the correlation between the BCIP score and degree of immune cell infiltration in BC tissues. The BCIP score was significantly correlated with decreased levels of immune and stromal cell content, indicating that the BCIP score was negatively correlated with immune cell infiltration in BC tissues (**Figures 6A–C**). The CIBERSORT algorithm was used to evaluate the association between the BCIP score and 22 immune cell contents in BC tissues (TCGA-BRCA cohort). The BCIP score was negatively correlated with the level of several immune cells with antitumor activity (including plasma, $CD8^+$ T, $CD4^+$ T, and gamma delta T cells; **Figure 6D**). In addition, the level of resting dendritic cells, which play a critical role in antigen phagocytosis and processing, decreased in BC tissues with increased BCIP scores (**Figure 6D**). In contrast, the level of M2 macrophages, which promote tumor progression, was positively correlated with the BCIP score (**Figure 6D**). These results indicate that the BCIP score was correlates with poor antitumor immunity. As a previous study showed that ICIs (Shah et al., 2012) significantly improved cancer survival through T cell immunity in BC, we further investigated whether the BCIP score correlated with the ICI response in BC. Four markers for ICI sensitivity and three markers of ICI resistance were selected to evaluate the ICI response in BC tissues. The BCIP score was negatively correlated with ICI-sensitive markers (CD274, CD8, IFNG, and IFN. GS; **Figure 6E**). In contrast, two of the three ICI resistance markers (MDSC and CAF) were positively correlated with the BCIP score in BC (**Figure 6E**). Collectively, the BCIP score was a negative marker of immune cell infiltration and ICI response in BC.

Establishment of a BCIP Score-Based Nomogram Model to Predict Overall Survival in BC

Univariate and multivariate Cox regression analyses showed that tumor size, molecular subtype, lymph node status**,** and the BCIP score were significant prognostic factors for BC (**Table 2**). After including the above variables, we developed a comprehensive prognostic nomogram based on TCGA-BRCA cohort

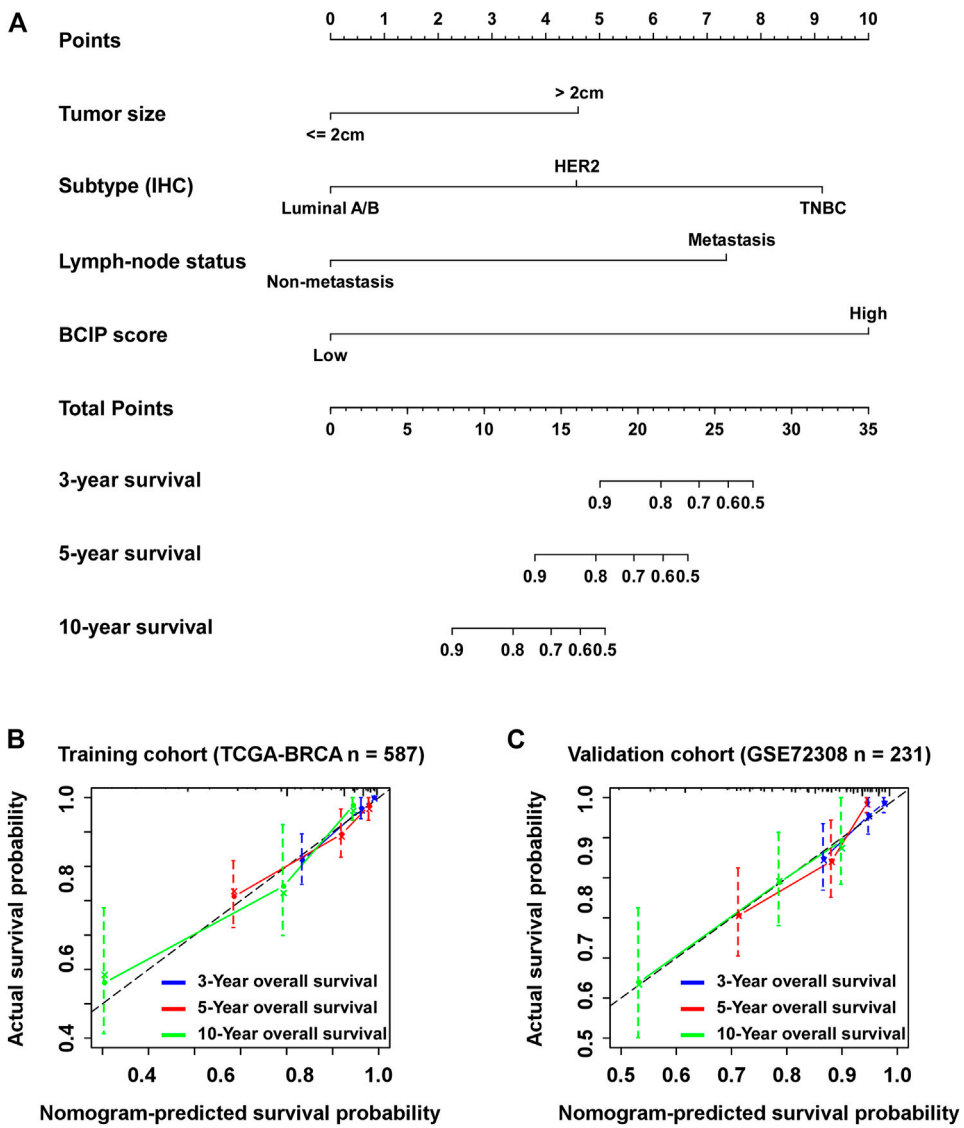


FIGURE 7 | Establishment of the BCIP score-based nomogram model to predict overall survival in BC. **(A)** Prognostic nomogram for patients with BC with factors including tumor size, molecular subtype, lymph node status, and the BCIP score. Points are defined based on the prognostic contribution each factor. Points summing the contribution of tumor size, molecular subtype, lymph node status, and the BCIP score are translated to the survival probability at 3, 5, and 10 years **(B,C)** Calibration plots for predicting patient overall survival at 3, 5, and 10 years in **(B)** TCGA and **(C)** the GSE72308 cohorts. Probability of survival based on the nomogram is listed on the *x*-axis, while the actual probability of survival is listed on the *y*-axis.

(**Figure 7A**). Factors correlated with high mortality risk (larger tumor size, TNBC subtype, metastatic lymph node status, and high BCIP score) were scored higher than those correlated with low mortality risk (smaller tumor size, luminal A/B subtype, non-lymph node metastasis, and low BCIP score). The c-index of this model was 0.831 (95% CI: 0.774–0.888) in TCGA-BRCA cohort. Furthermore, data from the GSE72308 dataset were selected for external validation of the nomogram. The c-index of the model was 0.734 (95% CI: 0.665–0.803) in the external validation cohort. The calibration curve was constructed to evaluate the accuracy of model prediction and indicated that the BCIP score-based nomogram exhibited good consistency in the prediction of 3-, 5-, and 10-year survival probabilities in both TCGA-BRCA and the GSE72308 cohorts (**Figures 7B,C**).

DISCUSSION

For years, prognostic prediction in patients with BC has been mostly based on the pathological features of the tumor (including tumor size, lymph node status, distant metastatic status, and molecular subtype) (Harbeck and Gnant, 2017). BC therapy guided by these prognostic indicators has significantly improved cancer survival and avoids the therapeutic side effects caused by overtreatment in BC (DeSantis et al.,

20192019). With advances in BC treatment, more precise prognostic methods are required (Denkert et al., 2017). Several prognostic biomarkers, including gene sequencing, gene copy number, and circulating tumor cells, have been used in clinical practice for BC (Lord and Ashworth, 2016; Tsai et al., 2018; Radovich et al., 2020). For instance, the 21-gene assay (Oncotype Dx) has been used to identify patients with low recurrence risk who could be exempted from chemotherapy in T1b/c and T2, HR+, HER2-, and lymph node-negative BC (Giuliano et al., 2017). Detection of *BRCA1/2* mutation status is helpful in identifying the potential benefits of PARP inhibitors (Tarsounas and Sung, 2020). Recently, Fackler et al. identified hypermethylation signatures that were correlated with cancer recurrence in TNBC (Fackler et al., 2020). In this study, we identified DNA methylation signatures related to BC incidence and developed a nomogram based on DNA methylation to calculate the 3-, 5-, and 10-year survival probabilities for BC.

An increasing amount of data has shown that aberrations in DNA methylation correlated with breast malignancy development is a prerequisite for the transformation of normal cells into BC cells, and these changes in DNA methylation accumulate in malignant cells, inducing an enhanced ability for proliferation and self-renewal (Zhuang et al., 2012; Wienken et al., 2016). As DNA methylation is regulated by environmental factors, genetic predisposition, and individual lifestyle, variation in DNA methylation pattern can be a reflection of the individual response to exposure to BC risk (Widschwendter et al., 2018). Identification of cancer incidence-related DNA methylation pattern is of great value for prognostic prediction in BC. Hence, we developed the BCIP score based on BC-related DNA methylation pattern and observed that the BCIP score was significantly correlated with poor outcomes in patients with BC.

In BC tissues, the BCIP score not only acted as a prognostic biomarker but also significantly correlated with aggressive BC features (such as larger tumor size, distant metastatic disease, and oncogene amplification). In addition to being a reflection of biological changes, changes in DNA methylation play a critical role in tumor progression through the regulation of gene expression. By assessing the BCIP score-related gene expression profile, we discovered that genes with expressions that were negatively correlated with the BCIP score were also significantly involved in the cancer immunity-related pathway. As a high BCIP score correlated with increased mortality, an aberration of cancer immunity might account for the poor outcome of patients with a high BCIP score.

In our study, we observed that BCIP scores were negatively correlated with the degree of T cell infiltration in BC tissues, indicating that the T-cell-mediated immune response (cellular immunity) was aberrantly inactivated. Cellular immunity is the main type of tumor immune response, which is mediated by the direct killing effect of T cells and release of cytokines by T cells (Joyce and Fearon, 2015). When activated by tumor antigens, T cells become immunoreactive and acquire the ability to recognize and kill tumor cells. However, T cell activation is often blocked by immune checkpoint molecules (including PD-1/PDL1, CTLA-4, and TIM-3) in individuals with cancer (Joyce and Fearon, 2015; Voorwerk et al., 2019). In BC, ICIs targeting PD1/PDL1 have been proven to improve survival in some patients, while most patients exhibit a poor response to ICIs (Voorwerk et al., 2019). To improve the efficacy of ICIs, effective biomarkers are required to identify potential beneficiaries. As our data showed that the BCIP score was negatively correlated with ICI response markers, the BCIP score is a potential biomarker to predict the sensitivity of BC to ICIs.

It is well known that DNA methylation correlates with suppressed gene expression, indicating that DNA methylation alterations can be exploited in cancer diagnosis. Compared with other genetic approaches (such as mutational analysis), DNA methylation-based approaches present advantages with regard to their clinical application (Heyn and Esteller, 2012). For instance, DNA methylation detection mostly focuses on a specific promoter region containing CpG islands, while mutational studies can cope with large regions since point mutations are located throughout the length of the gene (Heyn and Esteller, 2012). Moreover, alterations in DNA methylation were detected in a higher proportion of tumor tissue than genetic alterations, leading to higher sensitivity in prognostic analysis (Esteller et al., 2001). In our study, we developed a BCIP scoring system based on BC-related DNA methylation variation and observed that the BCIP score was positively correlated with mortality in patients with BC. Further, we adjusted for the confounding effects of tumor size, lymph node status, and molecular subtypes using Cox regression and found that BCIP was an independent prognostic factor for BC. By integrating the BCIP score, tumor size, lymph node status, and molecular subtypes, we established a nomogram model that could accurately quantify the survival probability of patients with BC. Based on our nomogram, patients with BC and a low mortality risk could be identified and might be exempted from aggressive and excessive medical treatment. However, patients with a high mortality risk should undergo more intensive surveillance for cancer recurrence.

Collectively, our study identified the variation in DNA methylation pattern related to BC incidence and developed a BCIP score model depicting BC-related DNA methylation variation. In BC, the BCIP score was significantly correlated with malignant BC characteristics and poor outcomes. Furthermore, we observed that the BCIP score was negatively correlated with immune cell infiltration and ICI response markers in BC tissues. Finally, a nomogram comprising the BCIP score, tumor size, lymph node status, and molecular subtype was developed to quantify the survival probability of patients with BC.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

ZS and XW designed the overall project; ZX, LY, and JA wrote the manuscript; JF, WZ, and WH revised and polished the manuscript. ZX, ZS, JY, and XZ performed the statistical analysis of the data; All the authors have read and approved the final manuscript.

FUNDING

This work was supported by project grants from the National Natural Science Foundation of China (81802663 and 82002557).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.814480/full#supplementary-material

REFERENCES

Abreu, M., Cabezas-Sainz, P., Pereira-Veiga, T., Falo, C., Abalo, A., Morilla, I., et al. (2020). Looking for a Better Characterization of Triple-Negative Breast Cancer by Means of Circulating Tumor Cells. *Jcm* 9 (2), 353. doi:10.3390/jcm9020353

Alwan, N. A. S., and Tawfeeq, F. N. (2019). Comparison of Clinico-Pathological Presentations of Triple-Negative versus Triple-Positive and HER2 Iraqi Breast Cancer Patients. *Open Access Maced J. Med. Sci.* 7 (21), 3534–3539. doi:10.3889/oamjms.2019.808

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ-related mRNA Profile Predicts Clinical Response to PD-1 Blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/JCI91190

Baylin, S. B., and Jones, P. A. (2011). A Decade of Exploring the Cancer Epigenome - Biological and Translational Implications. *Nat. Rev. Cancer* 11 (10), 726–734. doi:10.1038/nrc3130

Benci, J. L., Johnson, L. R., Choa, R., Xu, Y., Qiu, J., Zhou, Z., et al. (2019). Opposing Functions of Interferon Coordinate Adaptive and Innate Immune Responses to Cancer Immune Checkpoint Blockade. *Cell* 178 (4), 933–948. doi:10.1016/j.cell.2019.07.019

Britt, K. L., Cuzick, J., and Phillips, K.-A. (2020). Key Steps for Effective Breast Cancer Prevention. *Nat. Rev. Cancer* 20 (8), 417–436. doi:10.1038/s41568-020-0266-x

Denkert, C., Liedtke, C., Tutt, A., and von Minckwitz, G. (2017). Molecular Alterations in Triple-Negative Breast Cancer-The Road to New Treatment Strategies. *The Lancet* 389 (10087), 2430–2442. doi:10.1016/s0140-6736(16)32454-0

DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., et al. (20192019). Breast Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (6), 438–451. doi:10.3322/caac.21583

Esteller, M., Corn, P. G., Baylin, S. B., and Herman, J. G. (2001). A Gene Hypermethylation Profile of Human Cancer. *Cancer Res.* 61 (8), 3225–3229. doi:10.1016/S0165-4608(00)00403-9

Fackler, M. J., Cho, S., Cope, L., Gabrielson, E., Visvanathan, K., Wilsbach, K., et al. (2020). DNA Methylation Markers Predict Recurrence-free Interval in Triple-Negative Breast Cancer. *NPJ Breast Cancer* 6, 3. doi:10.1038/s41523-020-0145-3

Flavahan, W. A., Gaskell, E., and Bernstein, B. E. (2017). Epigenetic Plasticity and the Hallmarks of Cancer. *Science* 357 (6348). doi:10.1126/science.aal2380

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The Prognostic Landscape of Genes and Infiltrating Immune Cells across Human Cancers. *Nat. Med.* 21 (8), 938–945. doi:10.1038/nm.3909

Giuliano, A. E., Connolly, J. L., Edge, S. B., Mittendorf, E. A., Rugo, H. S., Solin, L. J., et al. (2017). Breast Cancer-Major Changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual. *CA: A Cancer J. Clinicians* 67 (4), 290–303. doi:10.3322/caac.21393

Gkountela, S., Castro-Giner, F., Szczerba, B. M., Vetter, M., Landin, J., Scherrer, R., et al. (2019). Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. *Cell* 176 (1-2), 98–112. doi:10.1016/j.cell.2018.11.046

Harbeck, N., and Gnant, M. (2017). Breast Cancer. *The Lancet* 389 (10074), 1134–1150. doi:10.1016/S0140-6736(16)31891-8

Heyn, H., and Esteller, M. (2012). DNA Methylation Profiling in the Clinic: Applications and Challenges. *Nat. Rev. Genet.* 13 (10), 679–692. doi:10.1038/nrg3270

Jones, P. A., Issa, J.-P. J., and Baylin, S. (2016). Targeting the Cancer Epigenome for Therapy. *Nat. Rev. Genet.* 17 (10), 630–641. doi:10.1038/nrg.2016.93

Joyce, J. A., and Fearon, D. T. (2015). T Cell Exclusion, Immune Privilege, and the Tumor Microenvironment. *Science* 348 (6230), 74–80. doi:10.1126/science.aaa6204

Koo, G.-B., Morgan, M. J., Lee, D.-G., Kim, W.-J., Yoon, J.-H., Koo, J. S., et al. (2015). Methylation-dependent Loss of RIP3 Expression in Cancer Represses Programmed Necrosis in Response to Chemotherapeutics. *Cell Res* 25 (6), 707–725. doi:10.1038/cr.2015.56

Lord, C. J., and Ashworth, A. (2016). BRCAness Revisited. *Nat. Rev. Cancer* 16 (2), 110–120. doi:10.1038/nrc.2015.21

Muvarak, N. E., Chowdhury, K., Xia, L., Robert, C., Choi, E. Y., Cai, Y., et al. (2016). Enhancing the Cytotoxic Effects of PARP Inhibitors with DNA Demethylating Agents - A Potential Therapy for Cancer. *Cancer Cell* 30 (4), 637–650. doi:10.1016/j.ccell.2016.09.002

Price, E. M., Cotton, A. M., Lam, L. L., Farré, P., Emberly, E., Brown, C. J., et al. (2013). Additional Annotation Enhances Potential for Biologically-Relevant Analysis of the Illumina Infinium HumanMethylation450 BeadChip Array. *Epigenetics & Chromatin* 6 (1), 4. doi:10.1186/1756-8935-6-4

Radovich, M., Jiang, G., Hancock, B. A., Chitambar, C., Nanda, R., Falkson, C., et al. (2020). Association of Circulating Tumor DNA and Circulating Tumor Cells after Neoadjuvant Chemotherapy with Disease Recurrence in Patients with Triple-Negative Breast Cancer. *JAMA Oncol.* 6, 1410. doi:10.1001/jamaoncol.2020.2295

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Shah, S. P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., et al. (2012). The Clonal and Mutational Evolution Spectrum of Primary Triple-Negative Breast Cancers. *Nature* 486 (7403), 395–399. doi:10.1038/nature10933

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Tarsounas, M., and Sung, P. (2020). The Antitumorigenic Roles of BRCA1-BARD1 in DNA Repair and Replication. *Nat. Rev. Mol. Cel Biol* 21 (5), 284–299. doi:10.1038/s41580-020-0218-z

Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., et al. (2016). Tumour Hypoxia Causes DNA Hypermethylation by Reducing TET Activity. *Nature* 537 (7618), 63–68. doi:10.1038/nature19081

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., et al. (2017). ChAMP: Updated Methylation Analysis Pipeline for Illumina BeadChips. *Bioinformatics* 33 (24), 3982–3984. doi:10.1093/bioinformatics/btx513

Tsai, M., Lo, S., Audeh, W., Qamar, R., Budway, R., Levine, E., et al. (2018). Association of 70-Gene Signature Assay Findings with Physicians' Treatment Guidance for Patients with Early Breast Cancer Classified as Intermediate Risk by the 21-Gene Assay. *JAMA Oncol.* 4 (1), e173470. doi:10.1001/jamaoncol.2017.3470

Voorwerk, L., Slagter, M., Horlings, H. M., Sikorska, K., van de Vijver, K. K., de Maaker, M., et al. (2019). Immune Induction Strategies in Metastatic Triple-Negative Breast Cancer to Enhance the Sensitivity to PD-1 Blockade: the TONIC Trial. *Nat. Med.* 25 (6), 920–928. doi:10.1038/s41591-019-0432-4

Widschwendter, M., Jones, A., Jones, A., Evans, I., Reisel, D., Dillner, J., et al. (2018). Epigenome-based Cancer Risk Prediction: Rationale, Opportunities and Challenges. *Nat. Rev. Clin. Oncol.* 15 (5), 292–309. doi:10.1038/nrclinonc.2018.30

Wienken, M., Dickmanns, A., Nemajerova, A., Kramer, D., Najafova, Z., Weiss, M., et al. (2016). MDM2 Associates with Polycomb Repressor Complex 2 and

Enhances Stemness-Promoting Chromatin Modifications Independent of P53. *Mol. Cel* 61 (1), 68–83. doi:10.1016/j.molcel.2015.12.008

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhuang, J., Jones, A., Lee, S.-H., Ng, E., Fiegl, H., Zikan, M., et al. (2012). The Dynamics and Prognostic Potential of DNA Methylation Changes at Stem Cell Gene Loci in Women's Cancer. *Plos Genet.* 8 (2), e1002517. doi:10.1371/journal.pgen.1002517

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xiong, Yang, Ao, Yi, Zouxu, Zhong, Feng, Huang, Wang and Shuang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fgene.2021.756094

SERPINH1 is a Potential Prognostic Biomarker and Correlated With Immune Infiltration: A Pan-Cancer Analysis

Yu Wang[1,2,3†], *Weigang Gu*[1,2,3†], *Weiwei Wen*[4] *and Xiaofeng Zhang*[1,2,3*]

[1]Department of Gastroenterology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2]Hangzhou Institute of Digestive Diseases, Hangzhou, China, [3]Key Laboratory of Integrated Traditional Chinese and Western Medicine for Biliary and Pancreatic Diseases of Zhejiang Province, Hangzhou, China, [4]Department of Dermatology, Third People's Hospital of Hangzhou, Hangzhou, China

Background: Serpin peptidase inhibitor clade H, member 1 (SERPINH1) is a gene encoding a member of the serpin superfamily of serine proteinase inhibitors. The upregulated of SERPINH1 was associated with poor prognosis in breast cancer, stomach adenocarcinoma, and esophageal carcinoma. However, the role of SERPINH1 in pan-cancer is largely unexplored.

Methods: SERPINH1 expression and the correlation with prognosis in human pan-cancer were analyzed by the Cancer Genome Atlas and the Genotype-Tissue Expression dataset. Pearson correlation analysis was applied to evaluate the role of SERPINH1 expression in tumor mutation burden (TMB), microsatellite instability (MSI), mismatch repair (MMR), DNA methyltransferase, and common immunoregulators. Spearman's correlation test was used to analysis SERPINH1 expression in tumor immune infiltration and infiltrating immune cells via the Tumor Immune Evaluation Resource database. Furtherly, immunohistochemistry staining of SERPINH1 was acquired from the Human Protein Atlas database for validation.

Results: SERPINH1 was abnormally expressed in fourteen cancers. The high expression of SERPINH1 significantly reduced the overall survival (OS), disease-specific survival, and progression free interval in eleven cancers. Moreover, SERPINH1 expression was correlated with MMR, MSI, TMB, and DNA methylation in multiple types of cancer. Also, SERPINH1 expression showed strong association with immunoregulators and immune checkpoint markers in testicular germ cell tumors, brain lower grade glioma (LGG), pheochromocytoma and paraganglioma. In addition, SERPINH1 expression was related to immune cell infiltration in multiple cancers, particularly in breast invasive carcinoma, LGG, and liver hepatocellular carcinoma. The result of immunohistochemistry verification shown that SERPINH1 staining was higher in tumor samples than in normal tissue in colon adenocarcinoma, head and neck squamous cell carcinoma, kidney renal papillary cell carcinoma and cervical squamous cell carcinoma, which was consistent with the result of OS.

OPEN ACCESS

Edited by:
Changwei Li,
Shanghai Jiao Tong University, China

Reviewed by:
Emil Bulatov,
Kazan Federal University, Russia
Daiwei Wan,
The First Affiliated Hospital of Soochow University, China

***Correspondence:**
Xiaofeng Zhang
zxf837@tom.com

†*These authors share first authorship*

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *16 September 2021*
Accepted: *06 December 2021*
Published: *04 January 2022*

Citation:
Wang Y, Gu W, Wen W and Zhang X (2022) SERPINH1 is a Potential Prognostic Biomarker and Correlated With Immune Infiltration: A Pan-Cancer Analysis. Front. Genet. 12:756094. doi: 10.3389/fgene.2021.756094

Conclusion: Overall, these results indicate that SERPINH1 may serve as an important prognostic biomarker and correlate with tumor immunity in human pan-cancer.

Keywords: SERPINH1, prognosis, immune infiltrate, pan-cancer, biomarker

INTRODUCTION

Cancer is a major public health problem and the second leading cause of mortality around the world (Beheshtirouy et al., 2021). In the treatment of cancer, immune response, protein dysfunction, specific gene mutations, disorders of intracellular signal transduction pathways had been targeted in the past 2 decades (Yeger et al., 2021). Among them, tumor immunotherapy, as a new type of treatment, has played an anti-tumor effect through immune regulation, and shown significant clinical effects (Dashti et al., 2016; Shamseddine et al., 2021). In addition, accumulating evidence indicates that the tumor growth, development and patient prognostic outcomes are related to immune infiltration, and the research on tumor immune mechanism is particularly important (Kamensek et al., 2021).

Serpin peptidase inhibitor, clade H, member 1 (SERPINH1) encodes a member of the serpin superfamily of serine proteinase inhibitors. SERPINH1 is localized to the endoplasmic reticulum and plays a role in collagen biosynthesis as a collagen-specific molecular chaperone. Abnormal expression of this gene may be a marker for cancer, and the nucleotide polymorphisms in this gene may be associated with preterm birth caused by preterm premature rupture of membranes.

The role of SERPINH1 in cancer is largely unexplored, and its functions appear to be inconsistent. For example, SERPINH1 has been reported to be upregulated, and associated with poor prognosis in breast cancer, stomach adenocarcinoma (STAD), and esophageal carcinoma (ESCA) (Lee et al., 2016; Strack et al., 2020; and Yang et al., 2020). Suppression of SERPINH1 expression with short interfering RNAs has been shown significantly inhibit cell proliferation, migration, and invasion in cervical squamous cell carcinoma (Yamamoto et al., 2013). In contrast, in human neuroblastoma, the aberrant methylation of the promoter CpG island makes the expression level of SERPINH1 relatively low (Yang et al., 2004). Therefore, the molecular mechanism through which SERPINH1 exerts its function and its value for prognostic diagnosis need further study.

In the present study, we systematically analyzed 33 types of cancer to investigate the correlation between SERPINH1 expression and the patients' prognosis. Moreover, this study totally explored the correlation of SERPINH1 with immunoregulators, immune checkpoints, and tumor-infiltrating immune cells in 33 tumor microenviroments. Taken together, our results suggest that SERPINH1 affects the prognosis of patients and is associated with immune infiltration in many cancers, especially in brain lower grade glioma (LGG).

METERIALS AND METHODS

Data Source

Transcriptome data and clinical information for patients of 33 types of cancers were downloaded from the TCGA data portal (http://cancergenome.nih.gov/), and the Genotype-Tissue Expression (GTEx) portal. 33 cancer types including ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma.

Analysis of SERPINH1 Expression in Human Pan-Cancer

The RNA sequencing data in TPM (transcripts per million reads) format were downloaded from TCGA and GTEx portal and processed by the Toil (Vivian et al., 2017). SERPINH1 expression data were normalized by log2 conversion. The Wilcoxon rank sum test was for comparing two independent samples, and Wilcoxon signed rank test was for paired sample.

Analysis of Survival and Prognosis

The pan-cancer samples were divided into SERPINH1 high and SERPINH1 low expression groups based on the minimum *p*-value approach (Kouyama et al., 2019). The Kaplan–Meier survival curves were utilized to exhibit the correlation of SERPINH1 expression with the prognosis of patients' overall survival (OS), disease-specifific survival (DSS), progression-free interval (PFI).

Analysis of Biological Function and Pathway

Interaction network analysis was obtained by employing STRING v11.5 database (http: //string-db.org/), keeping default

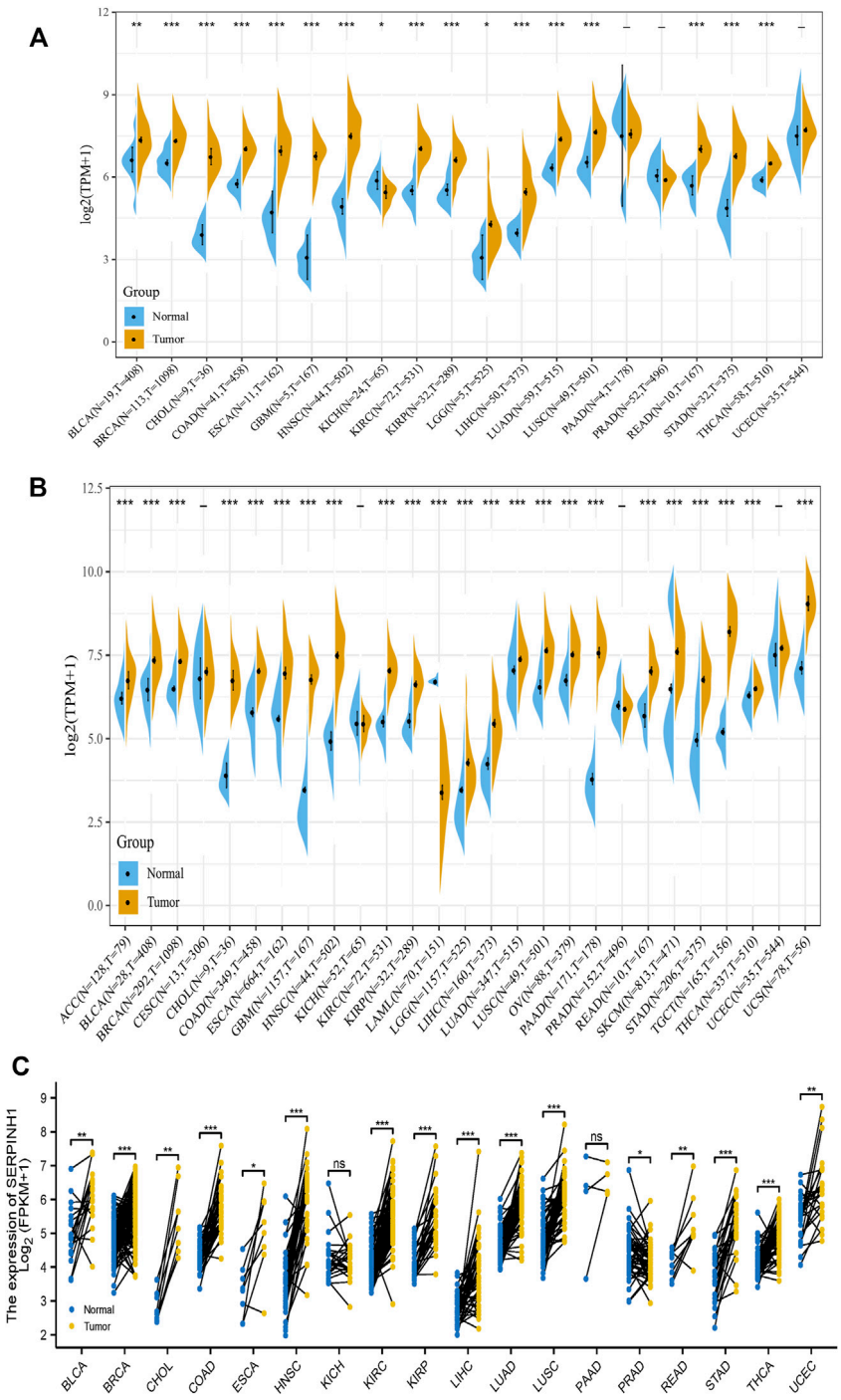


FIGURE 1 | SERPINH1 is abnormally expressed in human pan-cancer. **(A)** Differential expression of SERPINH1 in cancers and normal tissues from the TCGA dataset. **(B)** Data from TCGA and GTEx dataset showed differential expression of SERPINH1 in mutiple cancers. **(C)** Differential expression of SERPINH1 in cancers and normal tissues from TCGA dataset (paired samples, n = 11,093). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

parameters. The topological properties of the PPI network included average shortest path length, betweenness centrality, closeness centrality, degree, eccentricity, neighborhood connectivity, radiality, stress, and topological coefficient. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed to identify significant pathways via the "cluster Profiler" package in R (Yu et al., 2012).

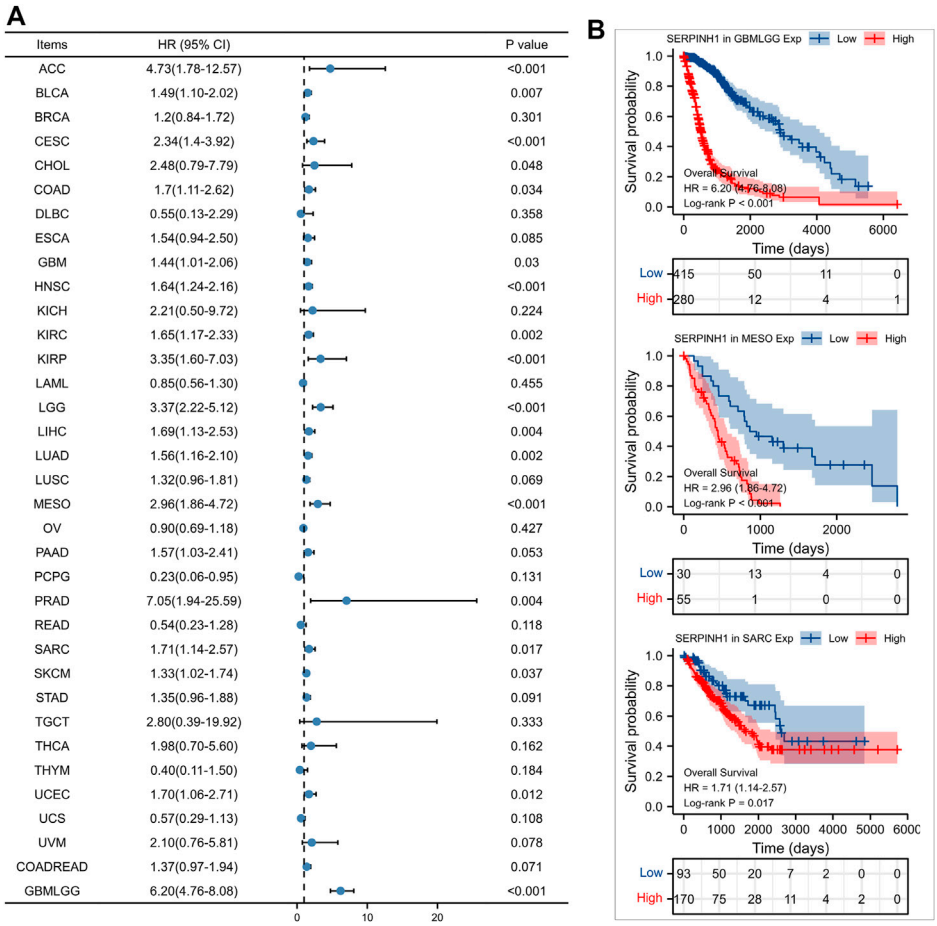


FIGURE 2 | Relationship of SERPINH1 expression level with patients' OS. **(A)** Forset plot of hazard ratio of SERPINH1 in human pan-cancer. **(B)** Kaplan-Meier OS curves for patients stratified by different expression levels of SERPINH1 in three cancer types.

Analysis of SERPINH1 Genetic Alteration and Correlations With TMB and MSI

We applied the cBioPortal website (https://www.cbioportal.org/) to query the genetic variation characteristics of SERPINH1 (Cerami et al., 2012; Gao et al., 2013). The correlations of SERPINH1 expression with TMB and MSI in human pan-cancer were based on the TCGA pan-cancer atlas, and analyzed by spearman's rank correlation method.

Analysis of MMR Gene Mutation and DNA Methylation

Mismatch repair (MMR) is a post-replicative repair mechanism that can repairs errors in DNA replication and attenuate chromosomal rearrangements, and is closely related to tumorigenesis (Iyer et al., 2015; Baretti and Le, 2018). In this study, the expression levels of MMR genes (MLH1, MSH2, MSH6, EPCAM, and PMS2) in human pan-cancer were obtained from the TCGA database. The correlation between MMR genes and SERPINH1 expression level were analyzed by pearson's correlation method. Another epigenetic factor affecting gene expression is DNA methylation. We also evaluated the correlation of four methyltransferases (DNMT1, DNMT2, DNMT3A, and DNMT3B) with SERPINH1 expression by the pearson's correlation method.

Analysis the Correlation Between SERPINH1 and Immunoregulators

The relationship between SERPINH1 and immunomodulators (such as immunostimulators, immunoinhibitors, MHC molecules, TILs, receptors, and chemokines) across multiple types of human cancer were evaluated through the TISIDB portal (http://cis.hku.hk/TISIDB/). The correlation analysis was performed to explore the relationship between SERPINH1 and 47 common immune checkpoint genes. To explore the prognostic value of SERPIHN1 on immunotherapy, we analyzed the relationship between SERPINH1 and immunotherapy targets (PD-1, PD-L1, PD-L2, and CTLA4). We utilized the Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) database

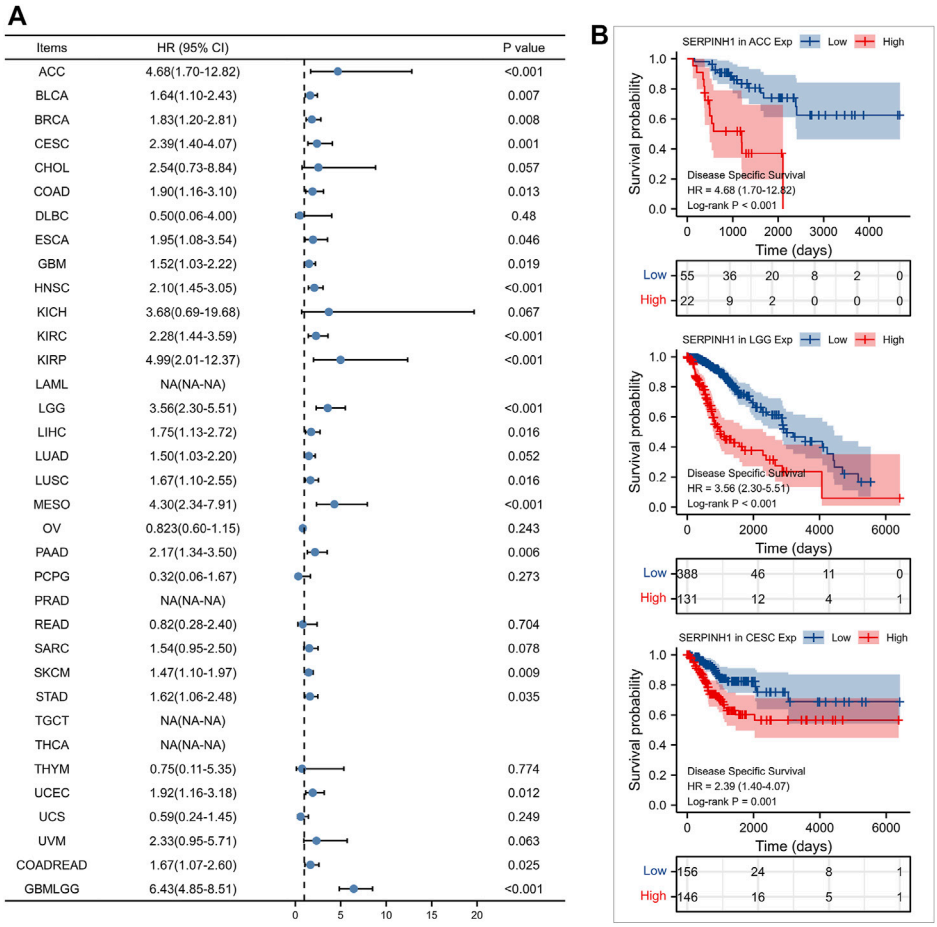


FIGURE 3 | Relationship of SERPINH1 expression level with patients' DSS. **(A)** Forset plot of hazard ratio of SERPINH1 in human pan-cancer. **(B)** Kaplan-Meier DSS curves for patients stratified by different expression levels of SERPINH1 in three cancer types.

to explore potential associations between SERPINH1 and the infiltrating immune cells across human cancers.

Analysis of Immune Infiltration

The TISIDB web portal (http://cis.hku.hk/TISIDB/index.php) was applied to explore the association between SERPINH1 and multiple immune regulatory factors across multiple cancer types (Ru et al., 2019). The TIMER database (https://cistrome.shinyapps.io/timer) was used to analyze the correlation between SERPINH1 expression and immune cell infiltration level across human cancers (Li et al., 2017). Spearman correlation analysis was applied to evaluate the correlation between SERPINH1 expression and the scores of multiple immune cells, including B cells, $CD4^+$ T cells, $CD8^+$ T cells, macrophages, neutrophils, and myeloid denedritic cells.

Statistical Analysis

All statistical analysis and visualization were conducted by R (version 3.6.3). Kaplan-Meier analysis was carried out using the survival package, correlation heatmap was visualized by ggplot2 package, radar chart was illustrated by ggradar and ggplot2 package. Tumor purity is a major confounding factor in this analysis, since most immune cell types are negatively correlated with tumor purity. Therefore, we use the partial spearman's correlation to perform purity adjustment association analysis, and explore the relationship between infiltrates estimation value and SERPINH1 expression. $p < 0.05$ was considered statistically significant.

RESULTS

SERPINH1 Is Abnormally Expressed in Human Pan-Cancer

In this study, we aimed to analysis the expression of SERPINH1 across various cancer types. Based on the TCGA database, we found that SERPINH1 expression is relatively higher in BRCA, CHOL, COAD, ESCA, GBM, HNSC, KIRP, LIHC, LUAD, LUSC, READ, STAD, and THCA (all $p < 0.001$), BLCA ($p < 0.01$), and LGG (all $p < 0.05$) than those in nomal tissues (**Figure 1A**). Taking into account the lack of normal samples in the TCGA database for some cancer types, we integrated the GTEx database for further analysis and the results showed that the expression level of SERPINH1 in the tumor tissues of ACC, BLCA, BRCA, CHOL, COAD, DLBC, ESCA, GBM, HNSC, KIRC, KIRP, LAML,

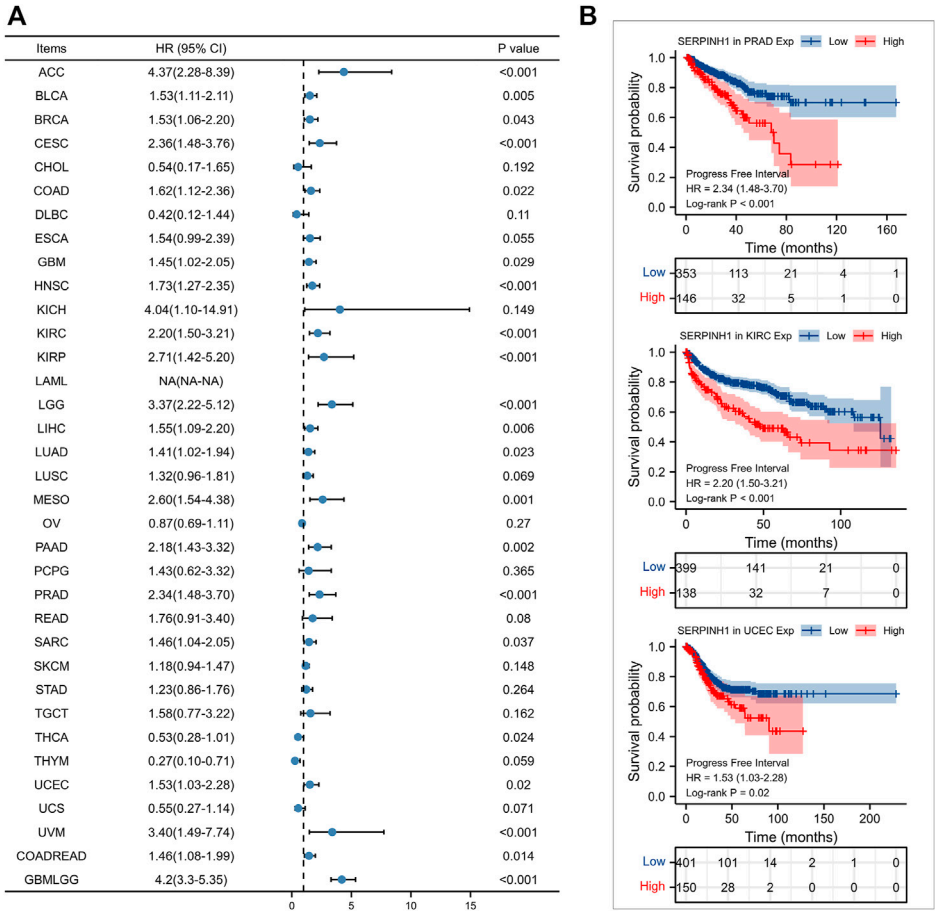


FIGURE 4 | Relationship of SERPINH1 expression level with patients' PFI. **(A)** Forset plot of hazard ratio of SERPINH1 in human pan-cancer. **(B)** Kaplan-Meier PFI curves for patients stratified by different expression levels of SERPINH1 in three cancer types.

LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD, READ, SKCM, STAD, TGCT, THYM, and UCS (all $p < 0.001$); UCEC ($p < 0.01$); and SARC, THCA (all $p < 0.05$) are much higher than the corresponding control tissues (**Figure 1B**). In paired samples, the expression of SERPINH1 in tumor tissues of BRCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, STAD, and THCA (all $p < 0.001$); BLCA, CHOL, and KICH (all $p < 0.01$); and ESCA, READ, and UCEC (all $p < 0.05$) are significantly higher than the corresponding control tissues (**Figure 1C**). The above results showed that SERPINH1 expressed abnormally across multiple cancers. Interestingly, we noticed that in PRAD, compared with control tissues, the expression of SERPINH1 has no significant difference in independent samples, but there is a significant difference in paired samples ($p < 0.05$).

Prognostic Potential of SERPINH1 in Human Pan-Cancer

Furtherly, we evaluated the relationship between the expression of SERPINH1 and the prognosis of pan-cancer patients. According to the Log-rank analysis, we found that in 33 cancers, the high expression of SERPINH1 significantly reduces the OS in 18 types of cancers including ACC, BLCA, CESC, CHOL, COAD, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, MESO, PRAD, SARC, SKCM, UCEC, and GBMLGG (**Figures 2A,B**). Since there may be non-tumor death factors during follow-up, we further explored the relationship between SERPINH1 expression and patients' DSS. Results indicated that high expression of SERPINH1 significantly impacted DSS in 20 types of cancers including ACC, BLCA, BRCA, CESC, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUSC, MESO, PAAD, SKCM, STAD, UCEC, COADREAD, and GBMLGG (**Figures 3A,B**). In addition, we also explored the correlation of SERPINH1 with patients' PFI, which indicated that high expression of SERPINH1 unfavorably impacted PFI in 21 types of cancers including ACC, BLCA, BRCA, CESC, COAD, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, MESO, PAAD, PRAD, SARC, THCA, UCEC, UVM, COADREAD, and GBMLGG (**Figures 4A,B**).

PPI Network Construction and Functional Enrichment Analysis of SERPINH1

The PPI network of SERPINH1 was shown in **Figure 5A**. The results of GO enrichment analysis showed that SERPINH1

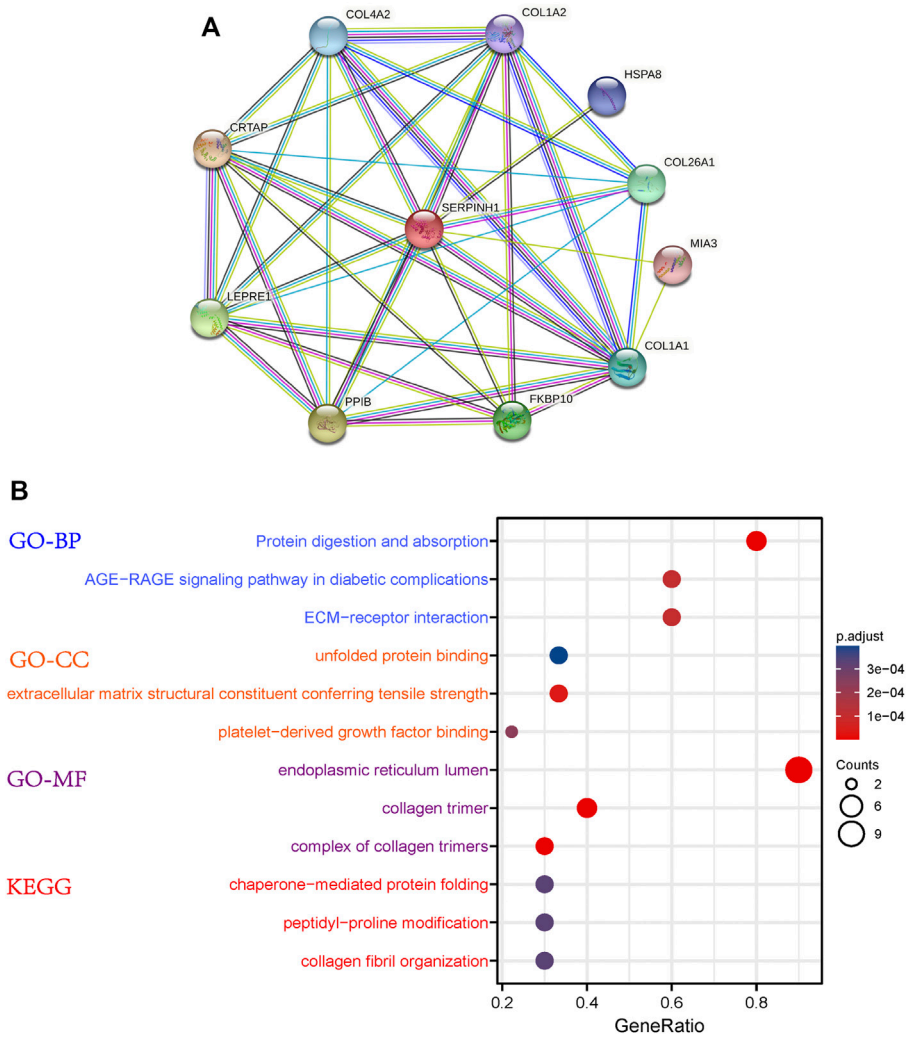


FIGURE 5 | Protein-protein interaction (PPI) network construction and enrichment analysis. **(A)** PPI for SERPINH1. **(B)** GO and KEGG enrichment analysis for SERPINH1.

significantly focused on protein digestion and absorption, AGE-RAGE signanling pathway in diabetic complications, ECM-receptor interaction (biological process, BP); unfolded protein binding, extracellular matrix structural constituent conferring tensile strength (cell components, CC); and endoplasmic reticulum lumen, collagen trimer (molecular function, MF); According to the *p*-value, the top three items from the three categories were selected to plot a bubble (**Figure 5B**). KEGG enrichment analysis (**Figure 6B**) indicated that SERPINH1 was significantly enriched with chaperone-mediated protein folding and peptidyl-proline modification *etc*.

Correlation Between SERPINH1 Expression and TMB or MSI in Human Pan-Cancer

The genetic alteration status of SERPINH1 in different cancer samples were observed via the TCGA cohorts. As shown in **Figure 7A**, the highest alteration frequency of SERPINH1 is "amplification" type, accounting for 7% of 13 patients with undifferentiated STAD. The "mutation" was the only type in 48 patients with mature B-cell neoplasms, which show an alteration frequency of 4.17%. 63 seminoma cases with genetic alteration (1.59% frequency) had copy number deletion of SERPINH1.

Tumor mutation burden (TMB), which counts the number of somatic mutations per megabase (mut/Mb), is an emerging potential biomarker for immunotherapy (Aggarwal et al., 2020). Microsatellite instability (MSI) is closely association with the development of most cancers (Matsuno et al., 2019). Here, we evaluated the correlation of SERPINH1 expression with TMB and MSI. Results showed that SERPINH1 expression was positively related to TMB in BRCA, KIRC, LAML, LGG, LUAD, and THYM; while negatively associated with TMB in ESCA, HNSC, LIHC, and SKCM (**Figure 7B**). In terms of MSI, the SERPINH1 expression was positively related to MSI in SARC, and TGCT (**Figure 7C**).

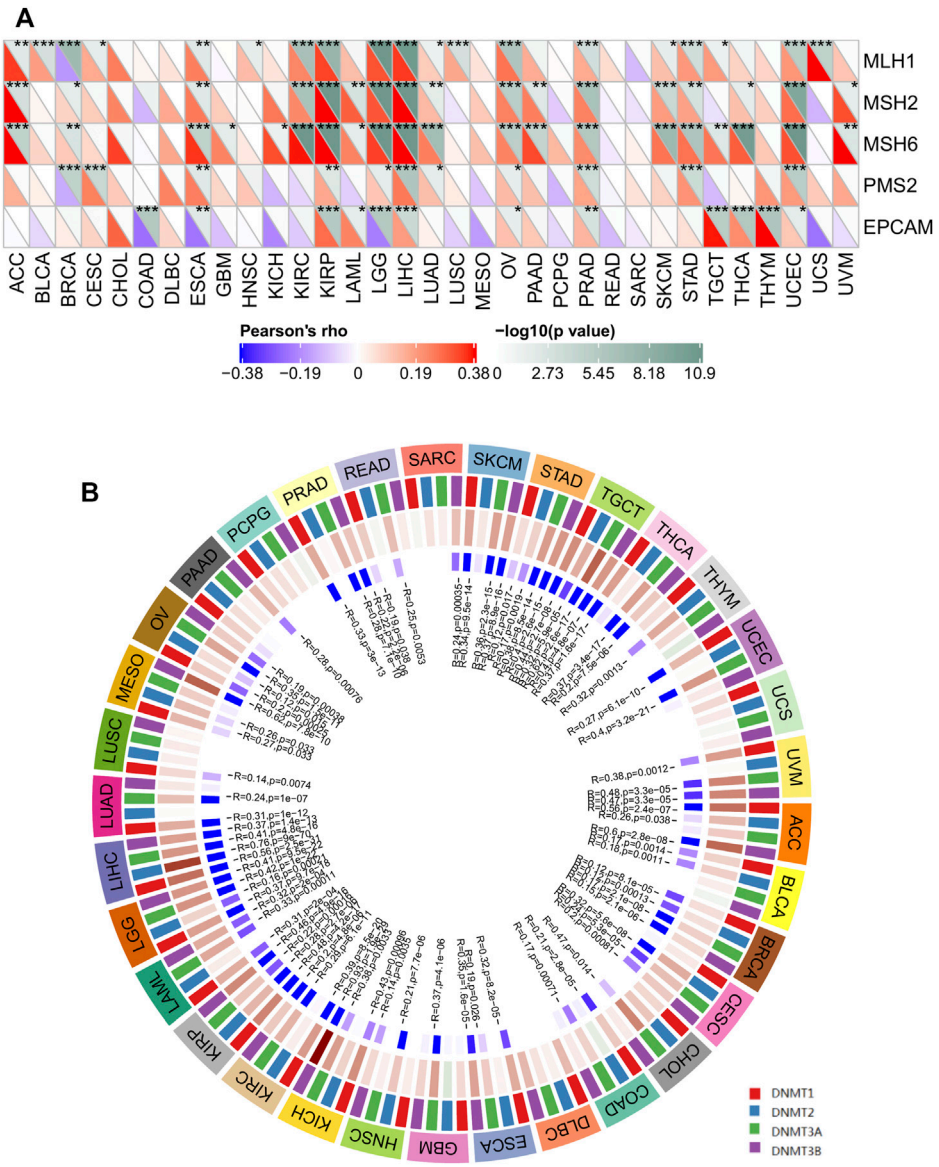


FIGURE 6 | The correlations of SERPINH1 expression with five MMR genes and four DNA methyltransferases in human pan-cancer. **(A)** The correlation of SERPINH1 expression with expression levels of five MMR genes across cancers. **(B)** The correlation of SERPINH1 expression with four DNA methyltransferases across cancers. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

SERPINH1 Is Associated With MMR Gene and DNA Methylation Across Cancers

Correct replication of the genome is a prerequisite for the maintenance of genome stability (Barra and Fachinetti, 2018). MMR serves as an important factor in maintaining genome stability in response to spontaneous DNA damage. MMR-deficient cells usually show microsatellite instability but display low levels of genome instabilities (Jia et al., 2017). To determine the role of SERPINH1 in tumor progression, we evaluated the association of the SERPINH1 expression with five MMR genes mutation levels. As results shown in **Figure 6A**, SERPINH1 expression was highly related to five MMR genes in seven cancers, including ESCA, KIRP, LGG, LIHC, OV, PRAD, and UCEC.

Global DNA hypomethylation and focal DNA hypermethylation are associated with tumorigenesis. Methylation of DNA is catalyzed by the DNA methyltransferase (DNMT) family of enzymes, with DNMT3A and DNMT3B catalyzing *de novo* DNA methylation and DNMT1 mediating both *de novo* and maintenance methylation of DNA (Suppiah et al., 2019). We therefore sought to determine whether SERPINH1 influences methyltransferases -mediated DNA repair, then we evaluated the relationships between SERPINH1 and four DNMTs. As shown in **Figure 6B**, SERPINH1 expression was

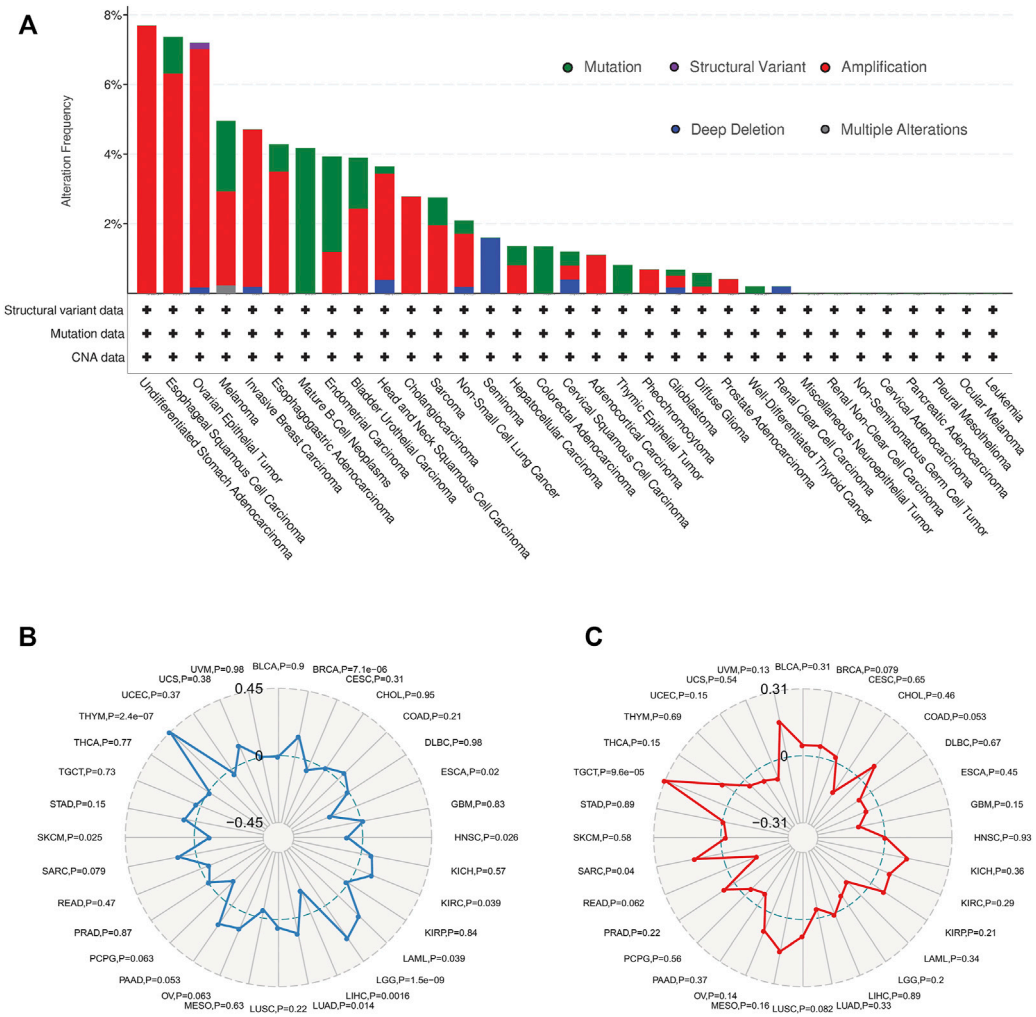


FIGURE 7 | Overview of SERPINH1 mutation information and correlations with TMB and MSI in human pan-cancer. **(A)** SERPINH1 mutations in 30 cancer types. **(B)** The correlations of SERPINH1 expression with TMB in human pan-cancer. **(C)** The correlations of SERPINH1 expression with MSI in human pan-cancer.

highly associated with these four DNMTs in multiple cancers, specifically in ACC, BLCA, BRCA, CESC, COAD, ESCA, GBM, HNSC, KICH, KIRP, KIRC, LAML, LGG, LIHC, MESO, OV, PRAD, SKCM, STAD, TGCT, THCA, and UVM. Results indicate that SERPINH1 may regulate the tumor progress by MMR-mediated DNA repair and DNA methylation.

Relationship Between SERPINH1 Expression and Immunoregulators in Human Pan-Cancer

The relationship between SERPINH1 and immunomodulators were shown in **Figures 8A–F**). Take COAD for example, we founded that SERPINH1 was significant positive correlation with several immunomodulators, including immunostimulator CD 276 and CD70 expressions (spearman correlation of 0.554 and 0.448 respectively), immunoinhibitor TGFB1 expression (spearman correlation of 0.553), and CCL21 expression (spearman correlation of 0.38) (**Figures 8G–I**). Based on the results, we can infer that the immunomodulators CD 276, TGFB1, CD70, and CCL21 might be regulated by SERPINH1 in COAD. Thus it can be seen that SERPINH1 interacts with immune regulation and may become a potential biomarker which has an important impact on the development of cancers and prognosis of patients.

Furtherly, we performed a correlation analysis to explore the relationship between SERPINH1 and 47 common immune checkpoint genes. Results showed that SERPINH1 expression correlated with 38 immune checkpoint genes in TGCT, and 35 immune checkpoint genes in LGG (**Figure 9A**). The relationship between SERPINH1 and immunotherapy targets (PD-1, PD-L1, PD-L2, and CTLA4) was shown in **Figure 9B**. Results indicated that SERPINH1 was positively correlated with immunotherapy targets in BLCA, BRCA, COAD, LGG, PAAD, PCPG, PRAD,

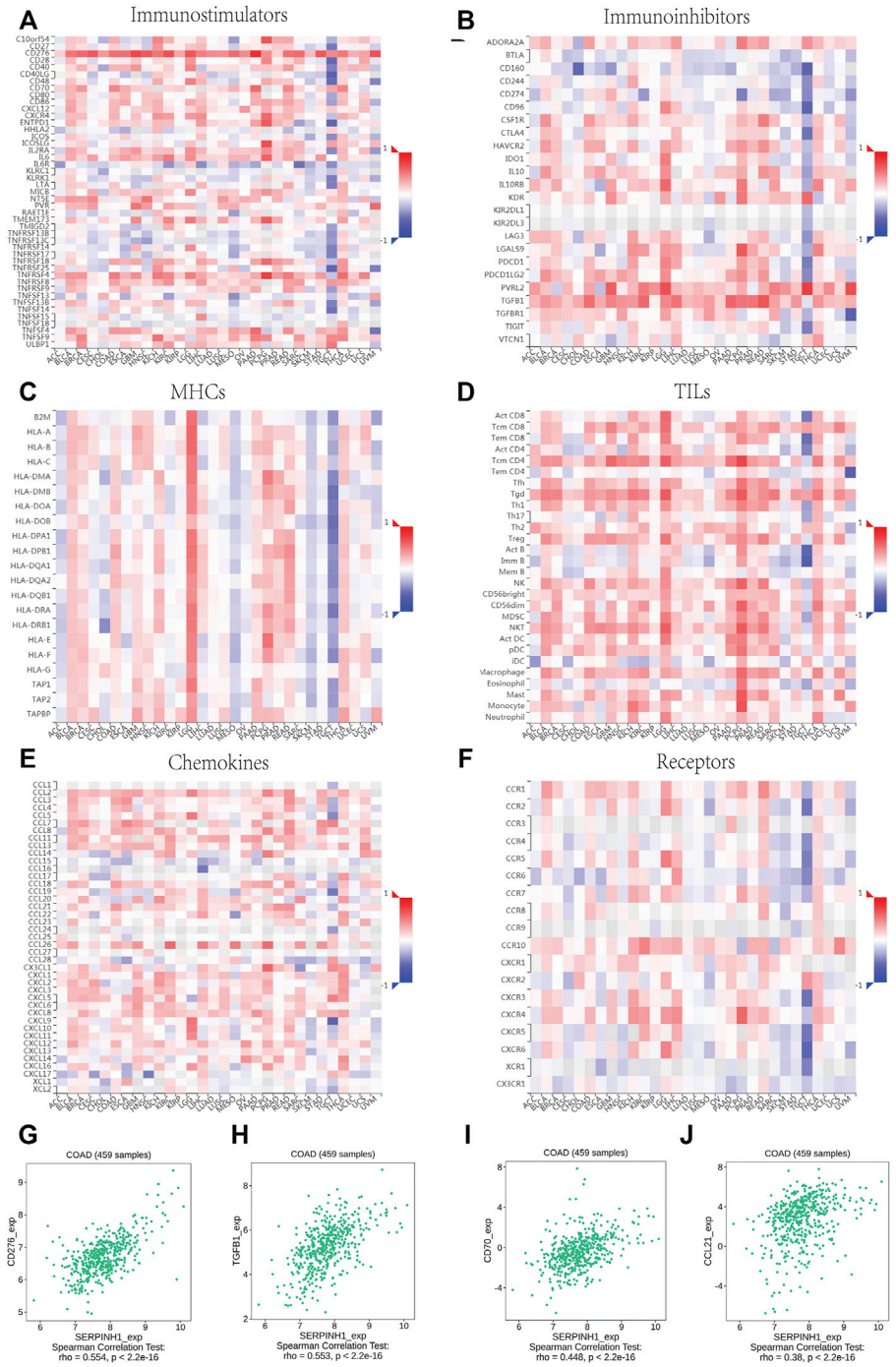


FIGURE 8 | Relationship of SERPINH1 expression and immunoregulators. **(A–F)**: Spearman correlations between immunostimulators, immunoinhibitors, MHC molecules, lymphocyte (TILs), receptors, and chemokines respectively across multiple types of human cancers. **(G–J)**: The correlation between the expression of SERPINH1 and the expression of CD276, TGFB1, CD70, and CCL21, respectively, in COAD.

READ, COADREAD, and LUADLUSC; and negatively correlated with immunotherapy targets in SKCM, TGCT, and THYM.

The tumor microenvironment consists of tumor cells, infiltrating immune cells, and stromal cells (Zhang et al., 2020). We utilized the TIMER database to explore potential associations between SERPINH1 and the infiltrating immune cells across human cancers (**Figure 10A**). This study focuses on showing the correlation between SERPINH1 expression and six immune infiltrating cells, including B cell, $CD4^+$ T cell, $CD8^+$

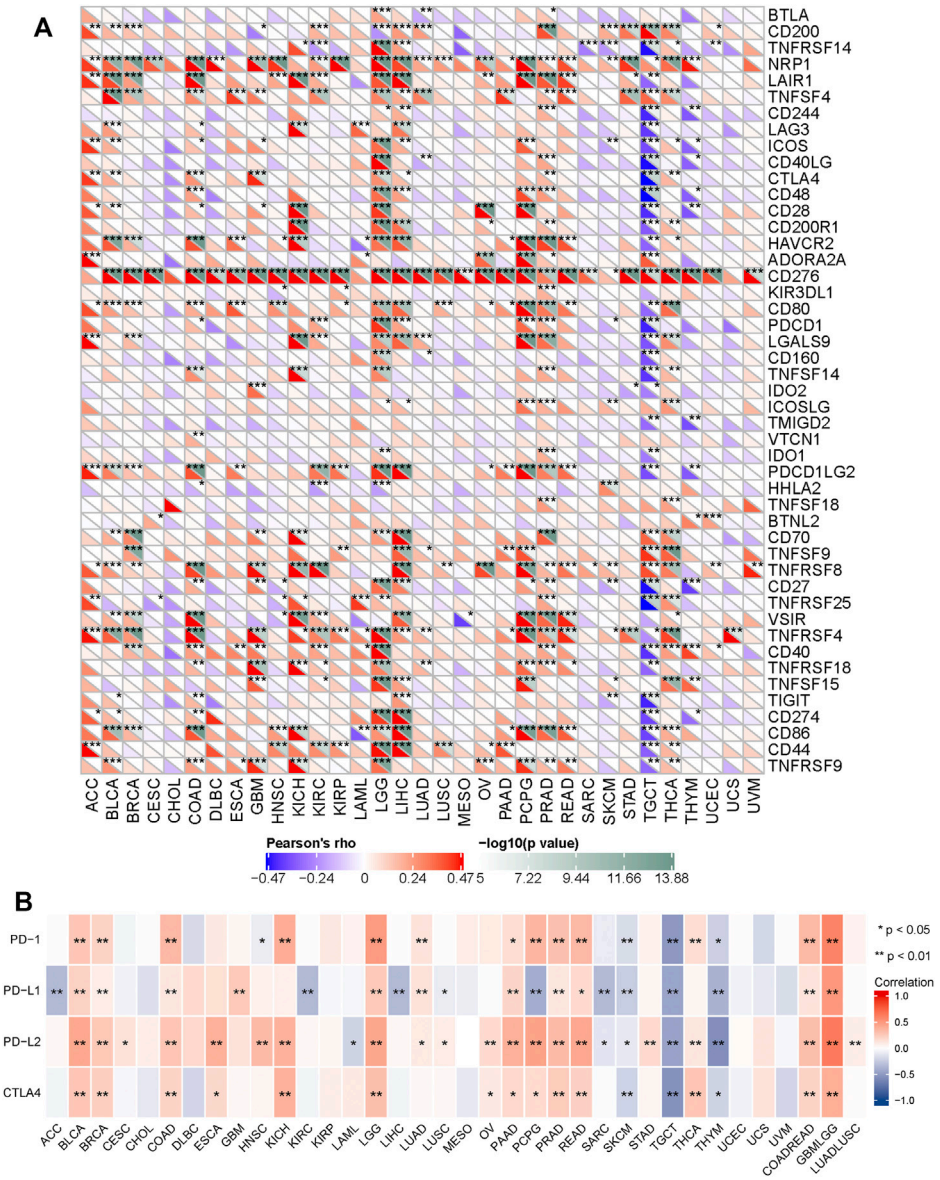


FIGURE 9 | Correlation of SERPINH1 with common immunue checkpoints and immunotherapy targets. **(A)** SERPINH1 is correlated with 47 common immunue checkpoints across human cancers. **(B)** SERPINH1 is correlated with immunotherapy targets across human cancers. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

T cell, dendritic cell macrophage cell and neutrophil cell in BRCA, LGG, and LIHC (**Figure 10B**). From the above results, we can infer that SERPINH1 acts significantly in tumor immunity.

Validation for SERPINH1 by Immunohistochemistry

The immunohistochemistry data of SERPINH1 was acquired from the Human Protein Atlas database (http://www.proteinatlas.org) for validation. As shown in **Figure 11**, SERPINH1 staining was higher in tumor samples than in normal tissue, which was consistent with the result of survival analysis, indicating that high expression of SERPINH1 is a risk factor in COAD, HNSC, KIRP, and CESC.

DISCUSSION

Serpins (serine-protease inhibitors) are a superfamily of proteins that share a conserved tertiary structure (Choi et al., 2019). Among 16 phylogenetic clades, SERPINH1 localized to the endoplasmic reticulum and plays a role in collagen biosynthesis as a collagen-specific molecular chaperone. Alternatively spliced transcript variants have been observed for

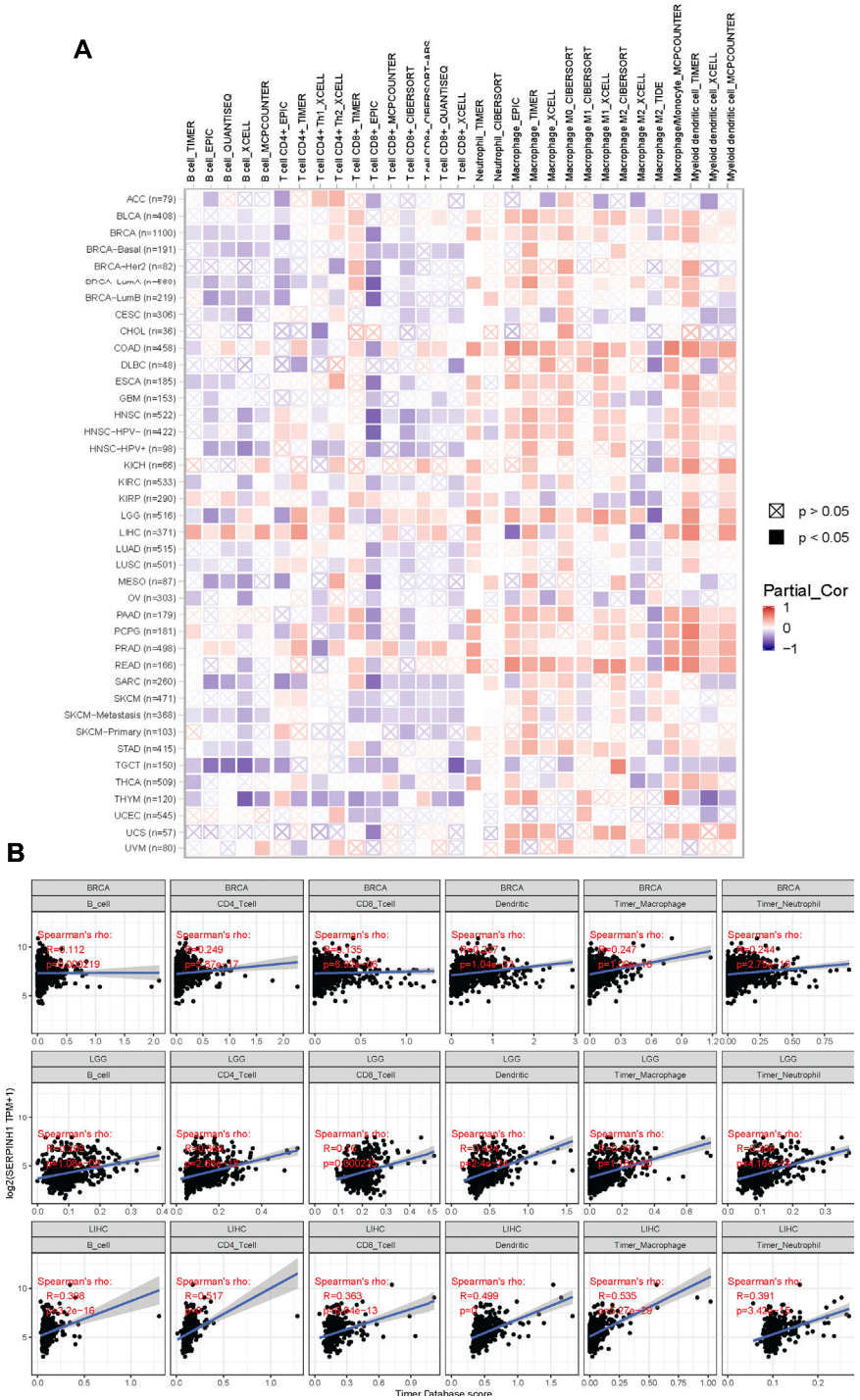


FIGURE 10 | Correlation of SERPINH1 with immune infiltration level in human pan-cancer. **(A)** SERPINH1 expression is correlated with infiltrating immune cells across human cancers. **(B)** SERPINH1 expression is positively correlated with immune infiltration in BRCA, LGG, and LIHC.

this gene, and a pseudogene of this gene is located on the short arm of chromosome 9.

In this study, the exploratory analysis of SERPINH1 in pan-cancers was carried out in the TCGA database, and the results showed that in both independent samples and paired samples, SERPINH1 were highly expressed in a variety of cancers (including BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, READ, STAD, and UCEC) compared with normal tissues. We further analyzed whether the high expression of SERPINH1 was related to the poor

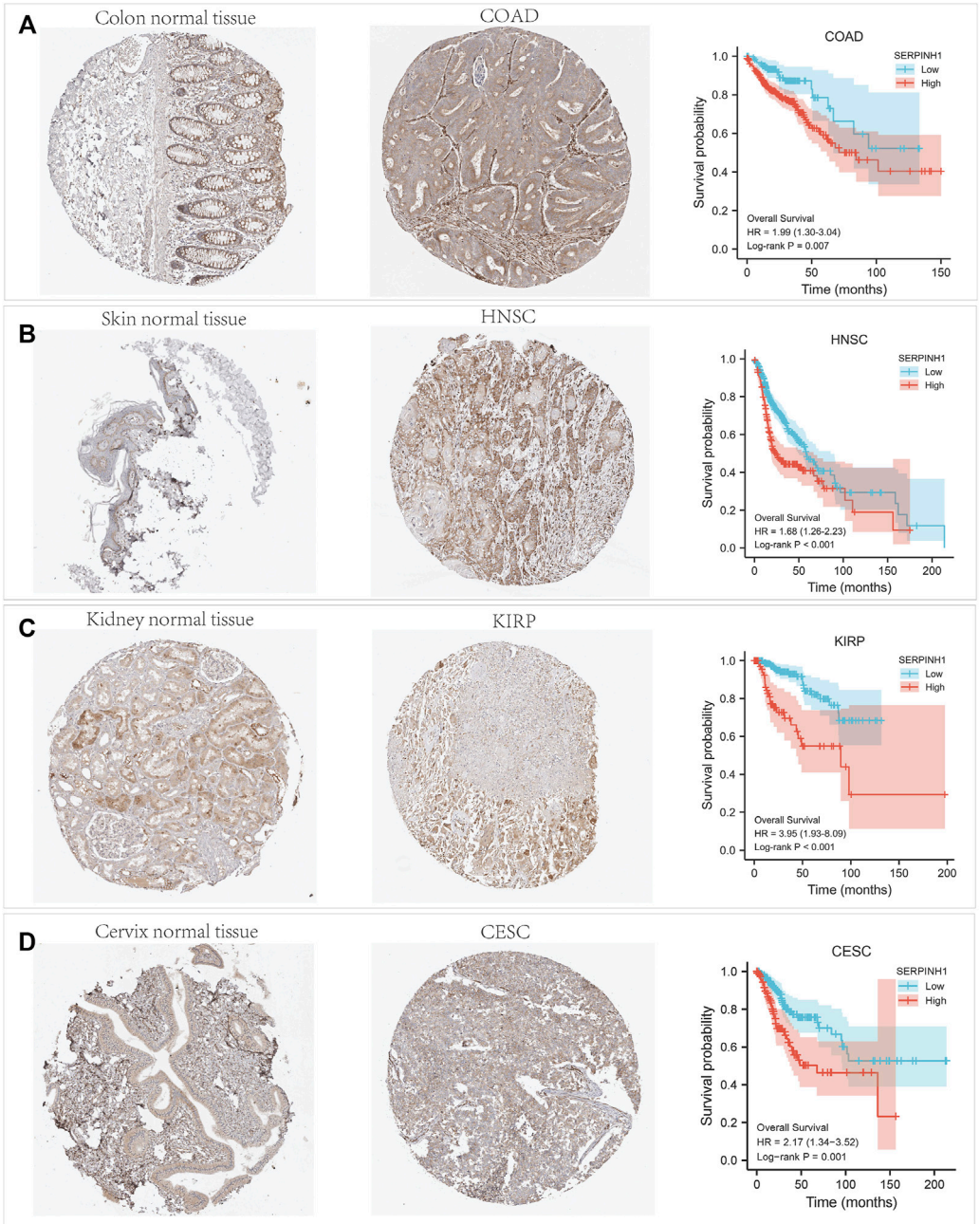


FIGURE 11 | Immunohistochemistry (IHC) staining of SERPINH1 based on the Human Protein Atlas. **(A)** IHC staining of SERPINH1 in colon normal tissue and COAD, Kaplan-Meier OS curves for patients stratified by different expression levels of SERPINH1 in COAD. **(B)** IHC staining of SERPINH1 in skin tumor tissue and HNSC, Kaplan-Meier OS curves for patients stratified by different expression levels of SERPINH1 in HNSC. **(C)** IHC staining of SERPINH1 in kidney normal tissue and KIRP, Kaplan-Meier OS curves for patients stratified by different expression levels of SERPINH1 in KIRP. **(D)** IHC staining of SERPINH1 in cervix tumor tissue and CESC, Kaplan-Meier OS curves for patients stratified by different expression levels of SERPINH1 in CESC.

prognosis of pan-cancer, and the results showed that high expression of SERPINH1 significantly reduced the OS, DSS, and PFI in 11 types of cancers including ACC, BLCA, CESC, COAD, GBM, HNSC, KIRC, KIRP, LGG, LIHC, and MESO.

TMB is an emerging biomarker in cancer characterized by MSI. TMB has been described as a powerful predictor of tumor behavior and response to immunotherapy (Chan et al., 2019; Chen et al., 2019). In this study, we observed the genetic alteration status of SERPINH1 in different cancers and evaluated the correlation of SERPINH1 expression with TMB and MSI. Results showed that SERPINH1 was positively related to TMB in BRCA, KIRC, LAML, LGG, LUAD, and THYM; while

negatively associated with TMB in ESCA, HNSC, LIHC, and SKCM. In terms of MSI, the SERPINH1 expression was positively related to MSI in SARC and TGCT.

SERPINH1 encodes HSP47, a chaperone located in the endoplasmic reticulum that appears to preferentially recognize and help maintain the folded state of the type I procollagen trimer (Macdonald and Bächinger, 2001). Two osteogenesis imperfecta mutations have been reported in SERPINH1, one in dachshunds (Drögemüller et al., 2009) and one in human moderately severe osteogenesis imperfecta (Duran et al., 2015 and Christiansen et al., 2010). In both instances, the homozygosity of the missense mutation corresponding main functional domain of SERPINH1 was determined, and the serine-type endopeptidase inhibitor domain recognizing Xaa–Arg–Gly-containing procollagen sequences were confirmed (Ono et al., 2012). However, the mutations and mutation sites of SERPINH1 in cancer are rarely reported. In this study, we investigated the SERPINH1 mutation information and evaluated the association of the SERPINH1 with five MMR and four DNMTs genes. Results showed that the highest alteration frequency of SERPINH1 was 7.69% in STAD, and 7.37% in ESCA. In addition, SERPINH1 was highly related to five MMR and four DNMTs genes in 7 and 22 cancers, respectively, which indicated that the mutant SERPINH1 may play a key role in the occurrence and prognosis of patients with related cancers.

Tumor immune microenvironment is known to play a pivotal role in the occurrence and development of cancer (Yu et al., 2019). Considering that the up-regulation of SERPINH1 expression level was related to shorter OS, DSS and PFI, we speculated that SERPINH1 may involved in tumor immune response. Here, we used the TIMER database to verify the hypothesis. The results showed that SERPINH1 was related to the immunostimulators, immunoinhibitors, MHC molecules, TILs, receptors, and chemokines in multiple types of human cancers. We further found the co-expression of SERPINH1 with immune checkpoint markers across cancers, specifically in TGCT (38/47), LGG (35/47), and PRAD (32/47). Tumor-infiltrating immune cells are considered to be a marker of host antitumor immune response and prognostic feature (Noman et al., 2020; and; Väyrynen et al., 2020). $CD8^+$ T-cells and NK cells play a predominant role in antitumoral immune response via immune checkpoints (Titov et al., 2021). In our present study, we found that SERPINH1 was significantly correlated with six immune infiltrating cells including B cell, $CD4^+$ T cell, $CD8^+$ T cell, dendritic cell macrophage cell and neutrophil cell in multiple cancers, particularly in BRCA, LGG, and LIHC. All this results infer that SERPINH1 may recruit and regulate infiltrating immune cells to inhibit or promote the progression of cancers, which strongly suggest that SERPINH1 serves as a key factor in cancer immunity.

In summary, this study revealed that SERPINH1 was highly expressed and related to poor prognosis in a variety of cancers, especially in BLCA, COAD, HNSC, KIRC, KIRP, and LIHC. Furthermore, SERPINH1 expression was found to be associated with MMR, MSI, TMB, and DNA methylation in multiple cancers. In addition, SERPINH1 was correlated with multiple immunoregulators and immune infiltration level in a variety of cancers, particularly in BRCA, LGG, and LIHC. To conclude, SERPINH1 may play a important role as a prognostic biomarker for human pan-cancer and the results of this study may provide a novel and effective immunological antitumor strategy for tumor immunity research. Since this research was based on data analysis, further experimental verification needs to be carried out.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

AUTHOR CONTRIBUTIONS

XZ designed the study. YW and WG analyzed the data and wrote the original manuscript. WW contributed to the manuscript revision. All the authors have read and approved the final manuscript.

FUNDING

This study was supported by grants from the Zhejiang Provincial National Science Fundation (LY21H270013 and LGF21H310004), China Postdoctoral Science Foundation (2020M671818), Zhejiang Medical and Health Science and Technology Plan (WKJ-ZJ-2136 and 2019RC068), and Hangzhou Medical and Health Science and Technology Plan (2016ZD01, OO20190610, and A20200174).

REFERENCES

Aggarwal, C., Thompson, J. C., Chien, A. L., Quinn, K. J., Hwang, W.-T., Black, T. A., et al. (2020). Baseline Plasma Tumor Mutation Burden Predicts Response to Pembrolizumab-Based Therapy in Patients with Metastatic Non-small Cell Lung Cancer. *Clin. Cancer Res.* 26 (10), 2354–2361. doi:10.1158/1078-0432.CCR-19-3663

Baretti, M., and Le, D. T. (2018). DNA Mismatch Repair in Cancer. *Pharmacol. Ther.* 189, 45–62. doi:10.1016/j.pharmthera.2018.04.004

Barra, V., and Fachinetti, D. (2018). The Dark Side of Centromeres: Types, Causes and Consequences of Structural Abnormalities Implicating Centromeric DNA. *Nat. Commun.* 9 (1), 4340. doi:10.1038/s41467-018-06545-y

Beheshtirouy, S., Mirzaei, F., Eyvazi, S., and Tarhriz, V. (2021). Recent Advances in Therapeutic Peptides for Breast Cancer Treatment. *Cpps* 22 (1), 74–88. doi:10.2174/1389203721999201117123616

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Chen, Y., Liu, Q., Chen, Z., Wang, Y., Yang, W., Hu, Y., et al. (2019). PD-L1 Expression and Tumor Mutational burden Status for Prediction of Response to

Chemotherapy and Targeted Therapy in Non-small Cell Lung Cancer. *J. Exp. Clin. Cancer Res.* 38 (1), 193. doi:10.1186/s13046-019-1192-1

Choi, Y. J., Kim, S., Choi, Y., Nielsen, T. B., Yan, J., Lu, A., et al. (2019). SERPINB1-mediated Checkpoint of Inflammatory Caspase Activation. *Nat. Immunol.* 20 (3), 276–287. doi:10.1038/s41590-018-0303-z

Christiansen, H. E., Schwarze, U., Pyott, S. M., AlSwaid, A., Al Balwi, M., Alrasheed, S., et al. (2010). Homozygosity for a Missense Mutation in SERPINH1, Which Encodes the Collagen Chaperone Protein HSP47, Results in Severe Recessive Osteogenesis Imperfecta. *Am. J. Hum. Genet.* 86 (3), 389–398. doi:10.1016/j.ajhg.2010.01.034

Dashti, A., Ebrahimi, M., Hadjati, J., Memarnejadian, A., and Moazzeni, S. M. (2016). Dendritic Cell Based Immunotherapy Using Tumor Stem Cells Mediates Potent Antitumor Immune Responses. *Cancer Lett.* 374 (1), 175–185. doi:10.1016/j.canlet.2016.01.021

Drögemüller, C., Becker, D., Brunner, A., Haase, B., Kircher, P., Seeliger, F., et al. (2009). A Missense Mutation in the SERPINH1 Gene in Dachshunds with Osteogenesis Imperfecta. *Plos Genet.* 5 (7), e1000579. doi:10.1371/journal.pgen.1000579

Duran, I., Nevarez, L., Sarukhanov, A., Wu, S., Lee, K., Krejci, P., et al. (2015). HSP47 and FKBP65 Cooperate in the Synthesis of Type I Procollagen. *Hum. Mol. Genet.* 24 (7), 1918–1928. doi:10.1093/hmg/ddu608

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6 (269), pl1. doi:10.1126/scisignal.2004088

Iyer, R. R., Pluciennik, A., Napierala, M., and Wells, R. D. (2015). DNA Triplet Repeat Expansion and Mismatch Repair. *Annu. Rev. Biochem.* 84, 199–226. doi:10.1146/annurev-biochem-060614-034010

Jia, P., Chastain, M., Zou, Y., Her, C., and Chai, W. (2017). Human MLH1 Suppresses the Insertion of Telomeric Sequences at Intra-chromosomal Sites in Telomerase-Expressing Cells. *Nucleic Acids Res.* 45 (3), 1219–1232. doi:10.1093/nar/gkw1170

Kamensek, U., Ursic, K., Markelc, B., Cemazar, M., Setrajcic Dragos, V., and Sersa, G. (2021). Mutational burden, MHC-I Expression and Immune Infiltration as Limiting Factors for *In Situ* Vaccination by TNFα and IL-12 Gene Electrotransfer. *Bioelectrochemistry* 140, 107831. doi:10.1016/j.bioelechem.2021.107831

Kouyama, Y., Masuda, T., Fujii, A., Ogawa, Y., Sato, K., Tobo, T., et al. (2019). Oncogenic Splicing Abnormalities Induced by DEAD -Box Helicase 56 Amplification in Colorectal Cancer. *Cancer Sci.* 110 (10), 3132–3144. doi:10.1111/cas.14163

Lee, H. W., Kwon, J., Kang, M. C., Noh, M.-K., Koh, J. S., Kim, J. H., et al. (2016). Overexpression of HSP47 in Esophageal Squamous Cell Carcinoma: Clinical Implications and Functional Analysis. *Dis. Esophagus.* 29 (7), 848–855. doi:10.1111/dote.12359

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: a Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Macdonald, J. R., and Bächinger, H. P. (2001). HSP47 Binds Cooperatively to Triple Helical Type I Collagen but Has Little Effect on the thermal Stability or Rate of Refolding. *J. Biol. Chem.* 276 (27), 25399–25403. doi:10.1074/jbc.M102471200

Matsuno, Y., Atsumi, Y., Shimizu, A., Katayama, K., Fujimori, H., Hyodo, M., et al. (2019). Replication Stress Triggers Microsatellite Destabilization and Hypermutation Leading to Clonal Expansion *In Vitro*. *Nat. Commun.* 10 (1), 3925. doi:10.1038/s41467-019-11760-2

Noman, M. Z., Parpal, S., Van Moer, K., Xiao, M., Yu, Y., Arakelian, T., et al. (2020). Inhibition of Vps34 Reprograms Cold into Hot Inflamed Tumors and Improves Anti-PD-1/pd-L1 Immunotherapy. *Sci. Adv.* 6 (18), eaax7881. doi:10.1126/sciadv.aax7881

Ono, T., Miyazaki, T., Ishida, Y., Uehata, M., and Nagata, K. (2012). Direct *In Vitro* and *In Vivo* Evidence for Interaction between Hsp47 Protein and Collagen Triple helix. *J. Biol. Chem.* 287 (9), 6810–6818. doi:10.1074/jbc.M111.280248

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). TISIDB: an Integrated Repository portal for Tumor-Immune System Interactions. *Bioinformatics* 35 (20), 4200–4202. doi:10.1093/bioinformatics/btz210

Shamseddine, A. A., Burman, B., Lee, N. Y., Zamarin, D., and Riaz, N. (2021). Tumor Immunity and Immunotherapy for HPV-Related Cancers. *Cancer Discov.* 11 (8), 1896. doi:10.1158/2159-8290.CD-20-1760

Strack, E., Rolfe, P. A., Fink, A. F., Bankov, K., Schmid, T., Solbach, C., et al. (2020). Identification of Tumor-associated Macrophage Subsets that Are Associated with Breast Cancer Prognosis. *Clin. Translational Med.* 10 (8), e239. doi:10.1002/ctm2.239

Suppiah, S., Nassiri, F., Bi, W. L., Dunn, I. F., Hanemann, C. O., Horbinski, C. M., et al. (2019). Molecular and Translational Advances in Meningiomas. *Neuro-oncology.* 21 (Suppl. 1), i4–i17. doi:10.1093/neuonc/noy178

Titov, A., Zmievskaya, E., Ganeeva, I., Valiullina, A., Petukhov, A., Rakhmatullina, A., et al. (2021). Adoptive Immunotherapy beyond CAR T-Cells. *Cancers* 13 (4), 743, 2021 . undefined. doi:10.3390/cancers13040743

Väyrynen, J. P., Lau, M. C., Haruki, K., Väyrynen, S. A., Dias Costa, A., Borowsky, J., et al. (2020). Prognostic Significance of Immune Cell Populations Identified by Machine Learning in Colorectal Cancer Using Routine Hematoxylin and Eosin-Stained Sections. *Clin. Cancer Res.* 26 (16), 4326–4338. doi:10.1158/1078-0432.CCR-20-0071

Vivian, J., Rao, A. A., Nothaft, F. A., Ketchum, C., Armstrong, J., Novak, A., et al. (2017). Toil Enables Reproducible, Open Source, Big Biomedical Data Analyses. *Nat. Biotechnol.* 35 (4), 314–316. doi:10.1038/nbt.3772

Yamamoto, N., Kinoshita, T., Nohata, N., Yoshino, H., Itesako, T., Fujimura, L., et al. (2013). Tumor-suppressive microRNA-29a Inhibits Cancer Cell Migration and Invasion via Targeting HSP47 in Cervical Squamous Cell Carcinoma. *Int. J. Oncol.* 43, 1855–1863. doi:10.3892/ijo.2013.2145

Yang, B., Zhang, M., and Luo, T. (2020). Identification of Potential Core Genes Associated with the Progression of Stomach Adenocarcinoma Using Bioinformatic Analysis. *Front. Genet.* 11, 517362. doi:10.3389/fgene.2020.517362

Yang, Q., Liu, S., Tian, Y., Hasan, C., Kersey, D., Salwen, H. R., et al. (2004). Methylation-Associated Silencing of theHeat Shock Protein 47Gene in Human Neuroblastoma. *Cancer Res.* 64, 4531–4538. doi:10.1158/0008-5472.CAN-04-0956

Yeger, H., Mokhtari, R. B., and Zhao, O. (2021). Perspective on Dietary Isothiocyanates in the Prevention, Development and Treatment of Cancer. *Jcmt* 2020, 26. doi:10.20517/2394-4722.2020.61

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, H., Chen, Z., Ballman, K. V., Watson, M. A., Govindan, R., Lanc, I., et al. (2019). Correlation of PD-L1 Expression with Tumor Mutation Burden and Gene Signatures for Prognosis in Early-Stage Squamous Cell Lung Carcinoma. *J. Thorac. Oncol.* 14 (1), 25–36. doi:10.1016/j.jtho.2018.09.006

Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. *Mol. Ther. - Nucleic Acids* 22, 298–309. doi:10.1016/j.omtn.2020.08.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Gu, Wen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fgene.2021.789823

Hsa-miR-557 Inhibits Osteosarcoma Growth Through Targeting KRAS

*Zhi Qiao, Jinfeng Li, Hongwei Kou, Xiangrong Chen, Deming Bao, Guowei Shang, Songfeng Chen, Yanhui Ji, Tian Cheng, Yisheng Wang and Hongjian Liu**

Department of Orthopedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Objective: Osteosarcoma is the most common malignancy in the skeletal system; studies showed an important role of miRNAs in tumorigenesis, indicating miRNAs as possible therapeutic molecules. This study found abnormal hsa-miR-557 expression levels in osteosarcoma and tried to explore the potential function and the mechanism.

Methods: Differential expression genes of osteosarcoma were analyzed using GSE28423 from the GEO database. Survival analysis of miRNAs was conducted with data obtained from the TARGET-OS database. STRING and miRDIP were used to predict target genes of hsa-miR-557; KRAS was then verified using dual-luciferase reporter assay. Expression of genes was detected by qPCR, and levels of proteins were detected by Western blot. The proliferation ability of cells was detected by CCK-8 and cell cycle analysis. Tumor formation assay in nude mice was used to detect the influence of osteosarcoma by hsa-miR-557 *in vivo*.

Results: Analysis from the GEO and TARGET databases found 12 miRNAs that are significantly related to the osteosarcoma prognosis, 7 downregulated (hsa-miR-140-3p, hsa-miR-564, hsa-miR-765, hsa-miR-1224-5p, hsa-miR-95, hsa-miR-940, and hsa-miR-557) and 5 upregulated (hsa-miR-362-3p, hsa-miR-149, hsa-miR-96, hsa-miR-744, and hsa-miR-769-5p). CCK-8 analysis and cell cycle analysis found that hsa-miR-557 could significantly inhibit the proliferation of osteosarcoma cells. The tumor formation assay in nude mice showed that tumor sizes and weights were inhibited by hsa-miR-557 transfection. Further studies also proved that hsa-miR-557 could target the 3′UTR of KRAS and modulate phosphorylation of downstream proteins.

Conclusion: This study showed that hsa-miR-557 could inhibit osteosarcoma growth both *in vivo* and *in vitro*, by modulating KRAS expression.

Keywords: hsa-miR-557, miRNA, osteosarcoma, KRAS, survival rate, animal model

OPEN ACCESS

Edited by:
Changwei Li,
Shanghai Jiao Tong University, China

Reviewed by:
Dongmei Tong,
Hudson Institute of Medical Research, Australia
Antonio Giordano,
Temple University, United States

****Correspondence:***
Hongjian Liu
fccliuhj@zzu.edu.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *05 October 2021*
Accepted: *15 December 2021*
Published: *11 January 2022*

Citation:
Qiao Z, Li J, Kou H, Chen X, Bao D, Shang G, Chen S, Ji Y, Cheng T, Wang Y and Liu H (2022) Hsa-miR-557 Inhibits Osteosarcoma Growth Through Targeting KRAS. Front. Genet. 12:789823. doi: 10.3389/fgene.2021.789823

INTRODUCTION

Osteosarcoma is one of the most common malignancies in the skeletal system, with an incidence of 2–3 million/year (Ritter and Bielack, 2010). Currently, around 70% of patients can be cured by surgical resection combined with chemotherapy. However, recurrent and metastatic osteosarcomas have a survival rate of only around 20%, which has remained unchanged for 30 years (Kansara et al., 2014). Thus, new therapies are needed for these osteosarcomas.

MicroRNAs (miRNAs) may be possible therapeutic molecules. Multiple studies showed the abnormal expression of miRNAs in malignancies, which indicated the role of miRNAs in tumorigenesis (Wang et al., 2019). Several *in vivo* studies showed the potential treatment effect of miRNAs in tumors (Wu et al., 2018). In osteosarcoma, molecules like miR-429 and miR-143-3p were found to be potential diagnostic and prognostic markers (Yang et al., 2020). Also, several molecules like miR-233-3p (Ji et al., 2018) and miR-1225-5p (Zhang et al., 2020) were found to be tumor suppressors of osteosarcoma. These results indicated potential diagnostic and treatment roles of miRNAs in osteosarcoma.

The Kirsten rat sarcoma virus (KRAS) gene belongs to the rat sarcoma virus (RAS) gene family, a group of genes that contribute to tumorigenesis. Publications have proved the association of KRAS with osteosarcoma, showing that KRAS could promote proliferation of osteosarcoma cells (Zhang et al., 2018a; Chen et al., 2019). KRAS was also proved to be targets of miRNAs. As previously reported, several miRNAs (miR-548-3p (Chen et al., 2019), miR-422a (Zhang et al., 2018a), and miR-217 (Zhang et al., 2015)) were shown to inhibit osteosarcoma by targeting KRAS. Thus, KRAS could be an important miRNA target that modulates the growth of osteosarcoma.

In this study, we found abnormal expression levels of hsa-miR-557 in osteosarcoma and tried to explore its potential function and mechanism.

MATERIAL AND METHODS

Bioinformatic Analysis

The Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) from the National Center for Biotechnology Information (NCBI) was searched, and data from the GSE28425 (a SuperSeries) were downloaded for miRNA expression analysis. GSE28425 includes GSE28423 (a non-coding RNA profile) and GSE28424 (an expression profile) of 19 osteosarcoma cell lines and 4 normal bones. MiRNA expression analysis was calculated using the R limma package (version 3.42.2). MiRNA expression data and clinical information of 89 osteosarcoma patients were downloaded from tumor alterations relevant for genomics-driven therapy (TARGET (https://xenabrowser.net/datapages/). Survival package (version 3.2–7) and survminer package (version 0.4.8) from R (version 3.5) were used for survival analysis. Psych package (version 2.0.9) was used for gene expression analysis, and the microRNA Data Integration Portal (miRDIP, http://ophid.utoronto.ca/mirDIP/) was used for target gene prediction. Protein–protein interaction (PPI) networks were constructed using STRING (https://string-db.org/).

Cell Culture

One normal osteoblast cell line (HFOB1.19, ATCC, United States) and four osteosarcoma cell lines (U20S (ATCC, United States), MNNG/HOS (ATCC, United States), SAOS-2 (ATCC, United States), and MG63 (ATCC, United States)) were used. Cells were cultured in the complete culture medium (CCM: DMEM (Procell, Wuhan, China) +10% FBS (WISENT, Nanjing, China) +1% P/S (NCM Biotech, Suzhou, China)). Cells were cultured at 37°C with 5% CO_2.

qPCR Analysis

Cellular samples were put into 1 ml Trizol for 5 min and transferred into 1.5 ml Eppendorf for further use. Tissue samples were grinded in liquid nitrogen, treated with 1 ml Trizol, and then transferred into 1.5 ml Eppendorf. Total RNA was extracted with chloroform and cold isopropanol and washed with 75% ethanol 3 times. A reverse transcription kit (Genecopoeia, Guangzhou, China) was used to synthesize cDNAs according to the manufacturer's instructions. BeyoFast™ SYBR Green qPCR Mix (Bio-Rad, Shanghai, China) was used for quantification of cDNAs, with the qTower 3.2G (Jena, Germany) qPCR instrument. The relative expression levels were analyzed using the $2^{-\Delta\Delta Ct}$ method. Triple wells were tested for each sample. Primers are shown in **Table 1**.

Western Blot

Cell samples were lysed with 2×SDS lysis buffer (Beyotime Biotechnology, Shanghai, China). Tissues were grinded in liquid nitrogen and treated with RIPA (Beyotime Biotechnology, Shanghai, China). Quantification of proteins was performed using BCA (bicinchoninic acid assay). Protein samples were isolated in SDS-PAGE gel by electrophoresis and transferred to a PVDF membrane. GAPDH mouse (proteintech, Wuhan, China), *p*-ERK1/2 (Santa Cruz, United States), ERK1/2 (proteintech, Wuhan, China), *p*-JNK (proteintech, Wuhan, China), JNK (proteintech, Wuhan, China), KRAS (proteintech, Wuhan, China), anti-mouse IgG (H + L) (CST, United States), and anti-rabbit IgG (H + L) (CST, United States) were used to incubate the PVDF membrane according to the detecting proteins. Results were read *via* chemiluminescence (Tanon 5200 multi, Shanghai, China) using an ECL kit (Beyotime Biotechnology, Shanghai, China).

CCK-8 Assay

The cell counting kit-8 (CCK-8) assay was used to assess cell viability. Manufactured cells were cultured in a 96-well plate at a density of 5×10^3 cells per well, with 5% CO_2 at 37°C. After 0, 24, 48, and 72 h, 10 μL CCK-8 (Sigma, United States) was mixed and cultured for 2h at 37°C. The optical density (OD) values at 450 nm were recorded for analysis. Three repetitions were measured for each sample.

Cell Cycle Analysis

Manufactured cells were suspended with trypsin (NCM Biotech, Suzhou, China), washed twice with cold PBS, and fixed with 75% ethanol at 4°C for 24 h. After fixation, cells were washed twice with PBS and incubated in RNase A (Beyotime Biotechnology, Shanghai, China) and propidium iodide (PI, Beyotime Biotechnology, Shanghai, China) for 1h. The percentage of cells in different cell cycles was tested using a flow cytometer (ACEA NovoCyte D2040R, United States). The proliferation index (PI) was used for analysis. PI = (S + G2)/(G1+S + G2). Three repetitions were measured for each group.

TABLE 1 | Primers used in qPCR.

GAPDH forward primer	ACAGCCTCAAGATCATCAGC
GAPDH reverse primer	GGTCATGAGTCCTTCCACGAT
hsa-miR-557 RT primer	GTCGTATCCAGTGCGTGTCGT GGAGTCGGCAATTGCACTGGA TACGACAGACAA
hsa-miR-557 forward primer	GTTTGCACGGGTGGGCC
hsa-miR-557 reverse primer	CAGTGCGTGTCGTGGA
KRAS forward primer	GGACTGGGGAGGGCTTTCT
KRAS reverse primer	GCCTGTTTTGTGTCTACTGTTCT

TABLE 2 | Sequences used in plasmid construction.

Hsa-miR-557	GUUUGCACGGGUGGGCCUUGUCU
Hsa-miR-557 inhibitor	AGACAAGGCCCACCCGUGCAAAC
NC miRNA	UCACAACCUCCUAGAAAGAGUAGA
NC miRNA inhibitor	UCUACUCUUUCUAGGAGGUUGUGA
Wt-KRAS forward primer	TCGAGAGGCTCTATTTAACTGAGTCAC ACTGCATAGGAATTTAGGTACCGC
Wt-KRAS reverse primer	GGCCGCGGTACCTAAATTCCTATGCAG TGTGACTCAGTTAAATAGAGCCTC
Mut-KRAS forward primer	TCGAGAGGCTCTATTTAACTGATAGCA GTTGCATAGGAATTTAGGTACCGC
Mut-KRAS reverse primer	GGCCGCGGTACCTAAATTCCTATGCAA CTGCTATCAGTTAAATAGAGCCTC

Plasmid Construction

The PLVX-IRES-ZsGreen1 (ORL-BIO, Hunan, China) plasmid was used as a gene vector, and the psiCHECK-2 (ORL-BIO, Hunan, China) plasmid was used for dual-luciferase reporter assay. The coding sequence (CDS) of KRAS was downloaded from the NCBI, and primers were designed using CE DesignV (version 1.04). cDNAs were replicated by PCR and collected using an Agarose Gel Extraction kit (Omega, Georgia, United States). The QuickCut™ XhoI and the QuickCut™ EcoR I were used to cut plasmids. Plasmids and cDNAs were connected in Seamless Master Mix (Beyotime Biotechnology, Shanghai, China). Gene vectors were transformed into the competent DH5α *Escherichia coli*, cultured in the amp culture medium. Positive bacteria were selected. Plasmids were extracted and sent for sequencing. Sequences used are seen in **Table 2**.

Lentivirus Packaging

For lentivirus packaging, 293 T cells were used. Cells were cultured until 70% coverage. The culture medium was renewed with serum-free DMEM 2 h before transfection. Cells were cocultured with pHelper1, pHelper2, the constructed plasmid (1:1.5:2), and the lipofectamine 2000 for transfection. The cells were washed with PBS after 8 h, the culture medium was renewed, and cells were cultured for 48 h. A fluorescence microscope was used to observe the transfection efficiency.

Tumor Formation in Nude Mice

The animal experiment was proven by the research and clinical trial ethics committee of Zhengzhou University (No.2021-KY-0035–001). Twelve nude mice were separated into two groups, the experimental group and the control group. MG63 cells infected with lentivirus encoding has-miR-557 or empty vectors were harvested and resuspended in the complete culture medium. Cellular density was adjusted to 2×10^7/ml with the complete culture medium and matrigel (3:1). Each mouse was subcutaneously injected with 2×10^6 cells in 0.1 ml. Mice were raised at 24°C, humidity 30–40%, and 12 h light per day, with enough food and water supply. Tumors were sized from the eighth day (when visible) and sized every 5 days afterward. The mice were terminated on the 28th day, and tumors were collected for further analysis.

Dual-Luciferase Reporter Assay

The manufactured psiCHECK-2 plasmid was transfected into 293T cell line (ATCC, United States) by Lipo 2000 (Thermo Fisher Scientific, United States), together with the hsa-miR-557 or hsa-miR-557 inhibitor or the negative control (NC), accordingly. Dual-luciferase reporter assays (Beyotime Biotechnology, Shanghai, China) were performed according to the manufacturer's instruction. Relative light unit (RLU) of firefly and Rinilla luciferase were recorded.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism 8 (GraphPad Software, La Jolla, CA) and R (version 3.5). Graphs were created using GraphPad Prism 8 (GraphPad Software, La Jolla, CA). The Student's *t*-test and Mann–Whitney test were used in comparisons between two groups, ANOVA was used for comparisons between three or more groups. Pearson's correlation was used for correlation analysis. $p < 0.05$ was considered as statistically significant.

RESULTS

Hsa-miR-557 Positively Correlate With the Survival Rate of Osteosarcoma Patients

To investigate the differently expressed miRNAs in osteosarcoma, we searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) from the NCBI and found the GSE28425 (Namlos et al., 2012) dataset. Within this dataset, 4 normal human bone tissues and 19 osteosarcoma cell lines (143B, HAL, HOS, IOR-MOS, IOR-OS10, IOR-OS14, IOS-OS15, IOS-OS18, IOS-OS9, KPD, MG-63, MHM, MNHG-HOS, OHS, SaOS-2, U-2OS, and ZK-58) were analyzed. 153 miRNAs were found to express differently, including 74 upregulated miRNAs (log_2foldchange >2, adjusted p. value <0.05) and 79 downregulated miRNAs (log_2foldchange < -2, adjusted *p*. value <0.05) (**Figures 1A,B**).

Data of 89 osteosarcoma patients' samples from the TARGET database were also analyzed, including 58,389 genes. 314 miRNAs were found that may influence the survival rate of OS patients (hazard ratio >1 or <1, *p*. value <0.05). Taking the intersection with differently expressed miRNAs from the GSE28425 dataset, 12 miRNAs were found that correlate with poor prognosis: 7 miRNAs were downregulated in the GSE28425 dataset and lowly expressed in the TARGET database (hsa-miR-140-3p, hsa-miR-564, hsa-miR-765, hsa-miR-1224-5p, hsa-miR-95, hsa-miR-940,

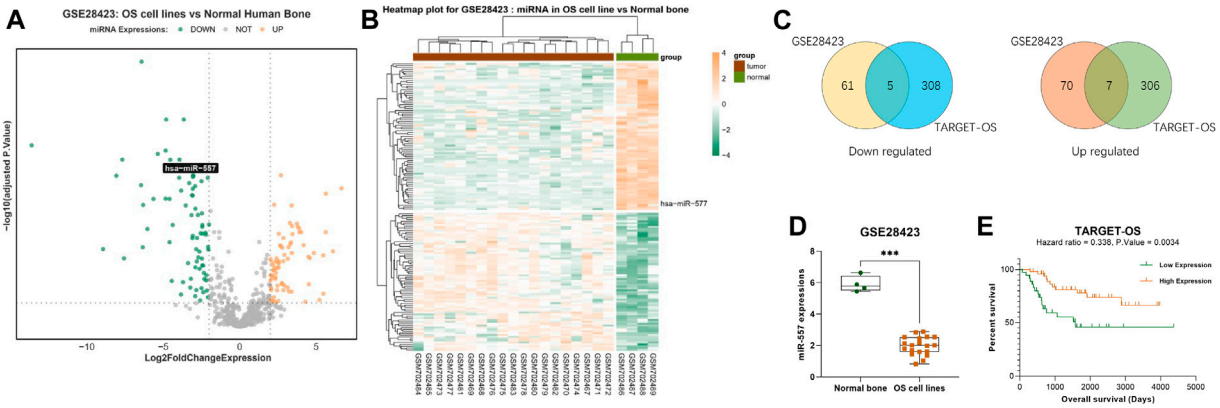


FIGURE 1 | Hsa-miR-557 positively correlate with the survival rate of osteosarcoma patients: **(A,B)** GSE28425 dataset from the GEO showed 74 upregulated miRNAs (log_2foldchange >2, adjusted *p*. value <0.05) and 79 downregulated miRNAs (log_2foldchange < -2, adjusted *p*. value <0.05). **(C)** Combining data from GSE28425 with 89 OS patient samples from TARGET, 7 miRNAs were downregulated in the GSE28425 dataset and lowly expressed in the TARGET database (hsa-miR-140-3p, hsa-miR-564, hsa-miR-765, hsa-miR-1224-5p, hsa-miR-95, hsa-miR-940, and hsa-miR-557), and 5 miRNAs (hsa-miR-362-3p, hsa-miR-149, hsa-miR-96, hsa-miR-744, and hsa-miR-769-5p) were upregulated in the GSE28425 dataset and highly expressed in the TARGET database. **(D)** Significant low levels of hsa-miR-557 were found in osteosarcoma cell lines from GSE28425. **(E)** Data from TARGET-OS showed that hsa-miR-557 positively correlate with the survival rate of osteosarcoma patients (hazard ratio <1, *p*. value < 0.01).

and hsa-miR-557), and 5 miRNAs (hsa-miR-362-3p, hsa-miR-149, hsa-miR-96, hsa-miR-744, and hsa-miR-769-5p) were upregulated in the GSE28425 dataset and highly expressed in the TARGET database (**Figure 1C**). Among these genes, hsa-miR-557 was found to be significantly downregulated (log2foldchange = −3.958, p. value = 1.083e-10, adjusted *p*. value = 4.808e-09) (**Figure 1D**); at the same time, higher expression of hsa-miR-557 positively correlates with a better survival rate of osteosarcoma patients (hazard ratio <1, *p*. value < 0.01) (**Figure 1E**).

Hsa-miR-557 Inhibits Proliferation of Osteosarcoma Cells *In Vitro*

Expression levels of hsa-miR-557 in different osteosarcoma cell lines (MNNG/HOS, SAOS-2, U20S, and MG63) and normal osteoblasts (hFOB1.19) were analyzed. Results showed that MNNG/HOS cells have the highest level of hsa-miR-557 expression, while MG63 cells have the lowest level of hsa-miR-557 expression (**Figure 2A**). Thus, MNNG/HOS and MG63 were selected to analyze the function of hsa-miR-557 on osteosarcoma *in vitro*. MNNG/HOS cells were transfected with the hsa-miR-557 inhibitor to specifically knock down hsa-miR-557, and MG63 cells were transfected with hsa-miR-557 to upregulate has-miR-557 expression. Results of qPCR showed that hsa-miR-557 levels were increased in MG63 cells and decreased in MNNG/HOS cells (MG63: $p < 0.0001$, q = 111.6; MNNG/HOS: $p = 0.02$, q = 5.458) (**Figure 2B**). To test the effect of hsa-miR-557 on cellular proliferation, CCK-8 assay and cell cycle analysis were performed. CCK-8 assay showed that the proliferation of MG63 cells was inhibited by hsa-miR-557 compared with the NC (48 h: $p = 0.0003$, t = 5.205; 72 h: $p < 0.0001$, q = 9.928) (**Figure 2C**), and the activity of MNNG/HOS cells was promoted by the hsa-miR-557 inhibitor, compared with the NC (48 h: $p = 0.0003$, t = 5.305; 72 h: $p < 0.0001$, t = 7.603) (**Figure 2D**). Cell cycle analysis also showed that comparing with the NC, inhibition of hsa-miR-557 promoted MNNG/HOS proliferation ($p = 0.0001$, t = 14.62), while hsa-miR-557 led to an inhibition of MG63 proliferation ($p = 0.0008$, t = 8.999) (**Figures 2E,F**).

Hsa-miR-557 Inhibit Growth of Osteosarcoma *In Vivo*

To test if hsa-miR-557 could inhibit the growth of osteosarcoma *in vivo*, nude mice were used to establish a xenograft mouse model. Results showed that hsa-miR-557 could inhibit the growth of osteosarcoma, as the tumor size and tumor weight were significantly reduced when has-miR-557 was upregulated (**Figures 3A–C**). Expression levels of hsa-miR-557 in tumor tissue samples were tested by qPCR. The tumor tissue transfected with hsa-miR-557 showed significantly high levels of hsa-miR-557 expression compared with the NC group ($p < 0.0001$, t = 7.083) (**Figure 3D**).

Hsa-miR-557 Targets KRAS and Negatively Regulates KRAS Expression

MiRNA was considered to play a critical role in modulating gene expressions through binding to the 3′UTR of its target gene, thus leading to a degradation of the target gene. To disclose the potential mechanism of hsa-miR-557 in osteosarcoma progression, the GSE28425 dataset was used to analyze genes related to has-miR-557. 4,353 negatively expressed genes were found. Furthermore, 2,100 potential targets of has-miR-557 were obtained from the miRDIP (**Figure 4A**). Genes were then subjected to performing PPI networks using STRING. Results showed that the Ras protein signaling transduction pathway was

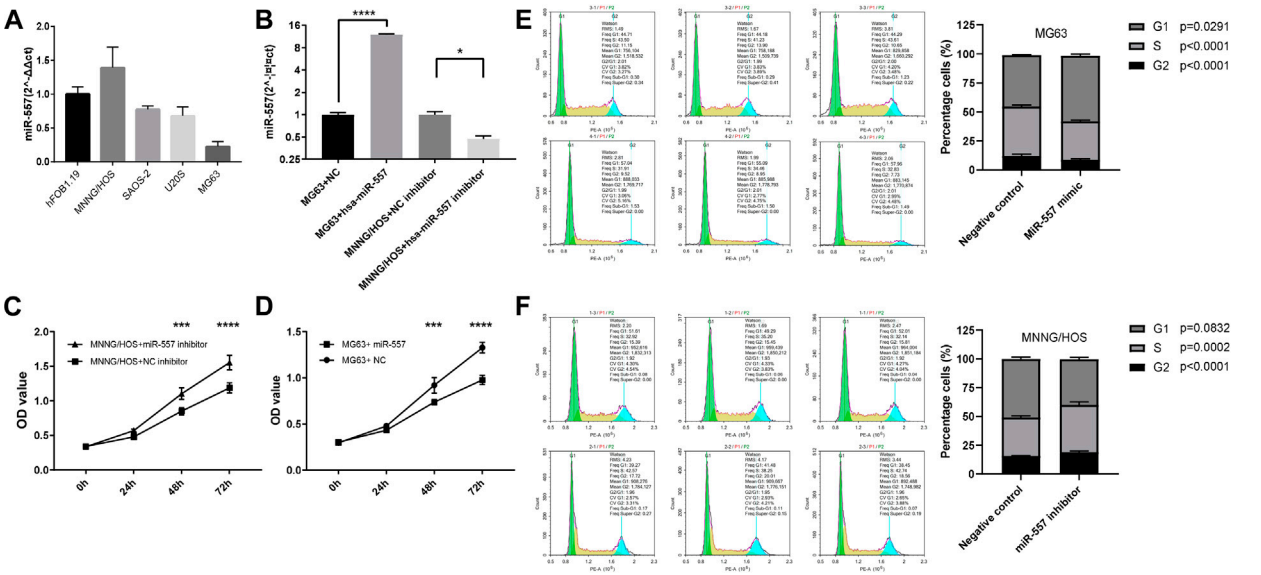


FIGURE 2 | Hsa-miR-557 inhibits the proliferation of osteosarcoma cells *in vitro*. **(A)** Hsa-miR-557 expression in osteosarcoma cell lines (MNNG/HOS, SAOS-2, U20S, and MG63) and normal osteoblasts (hFOB1.19). **(B)** Successfully elevated hsa-miR-557 expression levels by hsa-miR-557 transfection in MG63 and inhibited hsa-miR-557 expression levels by hsa-miR-557 inhibitor transfection in MNNG/HOS. **(C,D)** CCK-8 assay showed that Hsa-miR-557 inhibited MG63 cell proliferation, and hsa-miR-557 inhibitor could promote MNNG/HOS cell proliferation. **(E,F)** Cell cycle analysis showed that comparing with the NC, hsa-miR-557 inhibited the proliferation of MG63 cells ($p = 0.0008$, t = 8.999), and inhibition of hsa-miR-557 promoted MNNG/HOS proliferation ($p = 0.0001$, t = 14.62) (*$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$).

involved in hsa-miR-557 modulation (**Figure 4B**), and KRAS was found to be a target gene of hsa-miR-557, while its expression was found to correlate significantly with hsa-miR-557 levels (**Figure 4C**).

To further verify the potential mechanism of hsa-miR-557 in modulating osteosarcoma cell proliferation, dual-luciferase reporter assays were performed. Gene sequence of KRAS was downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/gene), Inctar (http://www.cuilab.cn/lnctar) was used to predict the potential target site of hsa-miR-557 seed sequence (**Figure 4D**). Wild-type or mutant KRAS 3′UTR (KRAS-3′UTR-wt and KRAS-3′UTR-mut) was integrated into the psiCHECK-2 plasma. Hsa-miR-557, hsa-miR-557 inhibitor, NC miRNA, and NC miRNA inhibitor were constructed and transfected into the 293T cell line together with the KRAS-3′UTR-wt or the KRAS-3′UTR-mut, respectively. Sequences are seen in **Table 2**. Results showed that compared with KRAS-3′UTR-wt + NC miRNA, levels of F/R (Firefly luciferase/Renilla luciferase) in the KRAS-3′UTR-wt + hsa-miR-557 group were significantly lower ($p = 0.0031$, t = 6.366), while the level of F/R in the KRAS-3′UTR-wt + hsa-miR-557 inhibitor group was significantly higher ($p = 0.0006$, t = 9.820) (**Figure 4E**). These results indicated that hsa-miR-557 could target the 3′UTR of KRAS, inhibiting the expression of KRAS.

To verify the correlation of hsa-miR-557 with KRAS, levels of KRAS in osteosarcoma and osteoblast cell lines were tested. KRAS expression levels were higher in osteosarcoma cell lines than the osteoblast cell line hFOB1.19, tested by qPCR ($p < 0.0001$) and Western blot ($p < 0.0001$) (**Figures 4F,G**). Then, MNNG/HOS cells transfected with the hsa-miR-557 inhibitor and NC miRNA inhibitor, as well as MG63 cells transfected with hsa-miR-557 and NC miRNA, were constructed. Results showed that the overexpression of hsa-miR-557 reduced both the protein and mRNA levels of KRAS, while the inhibition of hsa-miR-557 resulted in an upregulation of KRAS levels (**Figures 4H,I**). Similar results were also found *in vivo* (**Figure 4J,K**).

Hsa-miR-557 Inhibit Growth of Osteosarcoma Through Targeting KRAS and the Downstream Molecules

To certify that hsa-miR-557 could inhibit the proliferation of osteosarcoma, the psiCHECK-2 plasma integrated with KRAS 3′UTR (KRAS vector), the psiCHECK-2 plasma without KRAS 3′UTR (empty vector) were constructed. Hsa-miR-557 and NC miRNA were constructed according to **Table 2**. CCK-8 and cell cycle analysis were performed. Results showed that compared with the NC miRNA + empty vector group, KRAS could promote the proliferation of osteosarcoma cells (NC miRNA + KRAS vector, $p = 0.0021$, t = 7.114) and hsa-miR-557 could inhibit the proliferation of osteosarcoma cells (hsa-miR-557 + empty vector, $p = 0.0007$, t = 9.457), while the transfection of the KRAS vector could alleviate the inhibition effect of hsa-miR-557 (hsa-miR-557 + KRAS vector vs hsa-miR-557 + empty vector, $p = 0.0038$, t = 6.039) (**Figures 5A,B**). Results of CCK-8 also showed that compared with the NC miRNA + empty vector group, hsa-miR-557 could inhibit the proliferation of osteosarcoma cells at 48 h ($p = 0.0003$, q = 6.522) and 72 h ($p < 0.0001$, q = 13.84), KRAS could promote the proliferation of osteosarcoma cells (48h: $p < 0.0001$, q = 8.080; 72 h: $p < 0.0001$, q = 12.380), while the

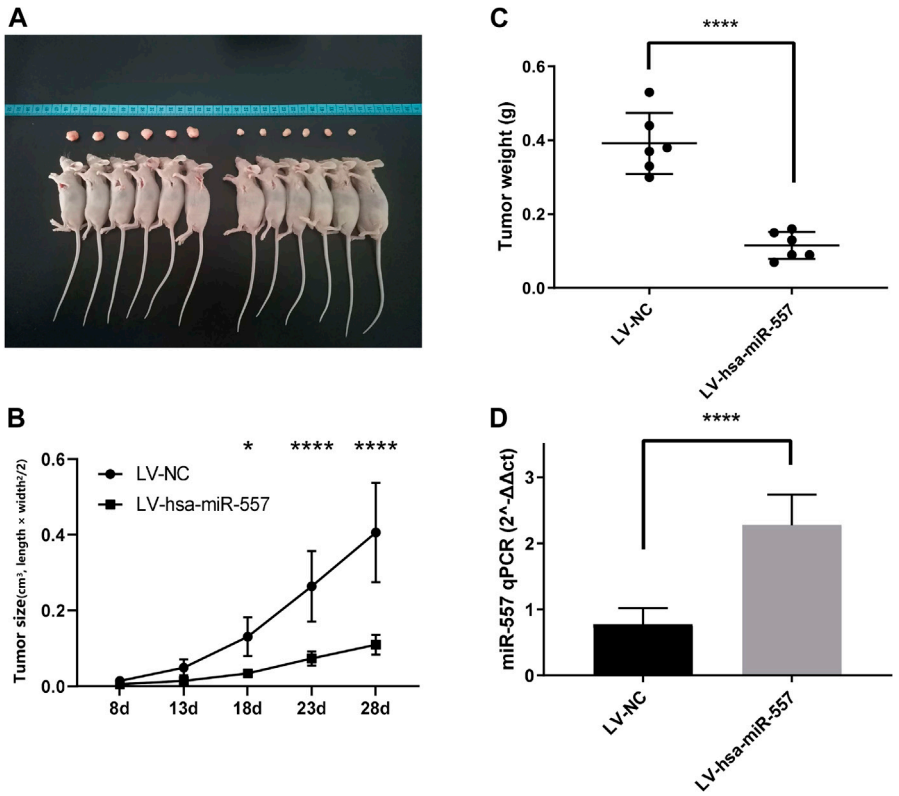


FIGURE 3 | Hsa-miR-557 inhibited the growth of osteosarcoma *in vivo*. MG63 cells transfected with hsa-miR-557 or the negative control (NC) by lentivirus (LV) were injected into BALB/c nude mice subcutaneously. **(A)** Tumor sizes from the hsa-miR-557 group and the negative control group after 28 days. **(B)** Tumor volume was measured every 5 days from the 8th day. **(C)** Significant lighter tumor weight was tested in the hsa-miR-557 group. **(D)** Higher levels of hsa-miR-557 expression in the hsa-miR-557 overexpression group were verified (*$p < 0.05$, ****$p < 0.0001$).

transfection of hsa-miR-557 could inhibit the effect of KRAS (NC miRNA + KRAS vector vs hsa-miR-557 + KRAS vector, 48 h: $p < 0.0001$, q = 7.815; 72 h: $p < 0.0001$, q = 14.72) (**Figure 5C**). These results proved the inhibition role of hsa-miR-557 in osteosarcoma proliferation through KRAS.

In the test of downstream proteins, four groups were built: ①MG63 + NC miRNA + empty vector, ②MG63 + hsa-miR-557 + empty vector, ③MG63 + NC miRNA + KRAS vector, and ④MG63+ hsa-miR-557 + KRAS vector. Downstream proteins (*p*-ERK1/2, ERK, *p*-JNK, JNK) were then tested by Western blot. Results showed that higher levels of *p*-ERK1/2 and *p*-JNK were found in the MG63 + NC miRNA + KRAS group, while lower levels of *p*-ERK1/2 and *p*-JNK were seen in the MG63+hsa-miR-557 + empty vector group (**Figure 5D**). Inhibition of *p*-ERK1/2 and *p*-JNK could be alleviated by KRAS. Results indicated that KRAS could promote the phosphorylation of ERK1/2 and JNK, while hsa-miR-557 could inhibit this process.

DISCUSSION

Hsa-miR-557 was recently noticed by researchers, as it has abnormal expression levels in several malignancies (Katayama et al., 2012; Jiang et al., 2014). We also noticed the abnormality of hsa-miR-557 expression in osteosarcoma and tried to find its function and the mechanism in modulating osteosarcoma. In this study, we found the following results: 1. hsa-miR-557 positively correlates with the survival rate of osteosarcoma patients. 2. hsa-miR-557 inhibits the proliferation of osteosarcoma cells *in vitro*. 3. hsa-miR-557 inhibits the growth of osteosarcoma *in vivo*. 4. hsa-miR-557 inhibits the growth of osteosarcoma through KRAS and its downstream molecules.

Several miRNAs were reported to have potential in osteosarcoma treatment. MiR-328-3p was reported to increase the radiosensitivity of osteosarcoma cells, inhibit the proliferation, and promote apoptosis through H2AX (Yang et al., 2018). Dai et al. reported that miR-513a-5p could enhance the radiosensitivity of OS, by negatively regulating APE1 (Dai et al., 2018). Mir-92a could inhibit the growth and the migration of osteosarcoma cells through targeting Notch1 (Liu et al., 2018). Other publications showed miR-29, miR-590, and miR-140-5p also inhibit proliferation and promote apoptosis of osteosarcoma cells (Wei et al., 2016; Xu et al., 2018; Long and Lin, 2019). On the other hand, several miRNAs were overexpressed in osteosarcoma, assisting in tumorigenesis. Zhang et al. found that miR-19a-3p could inhibit PTEN, thus promoting the proliferation of osteosarcoma cells (Zhang et al.,

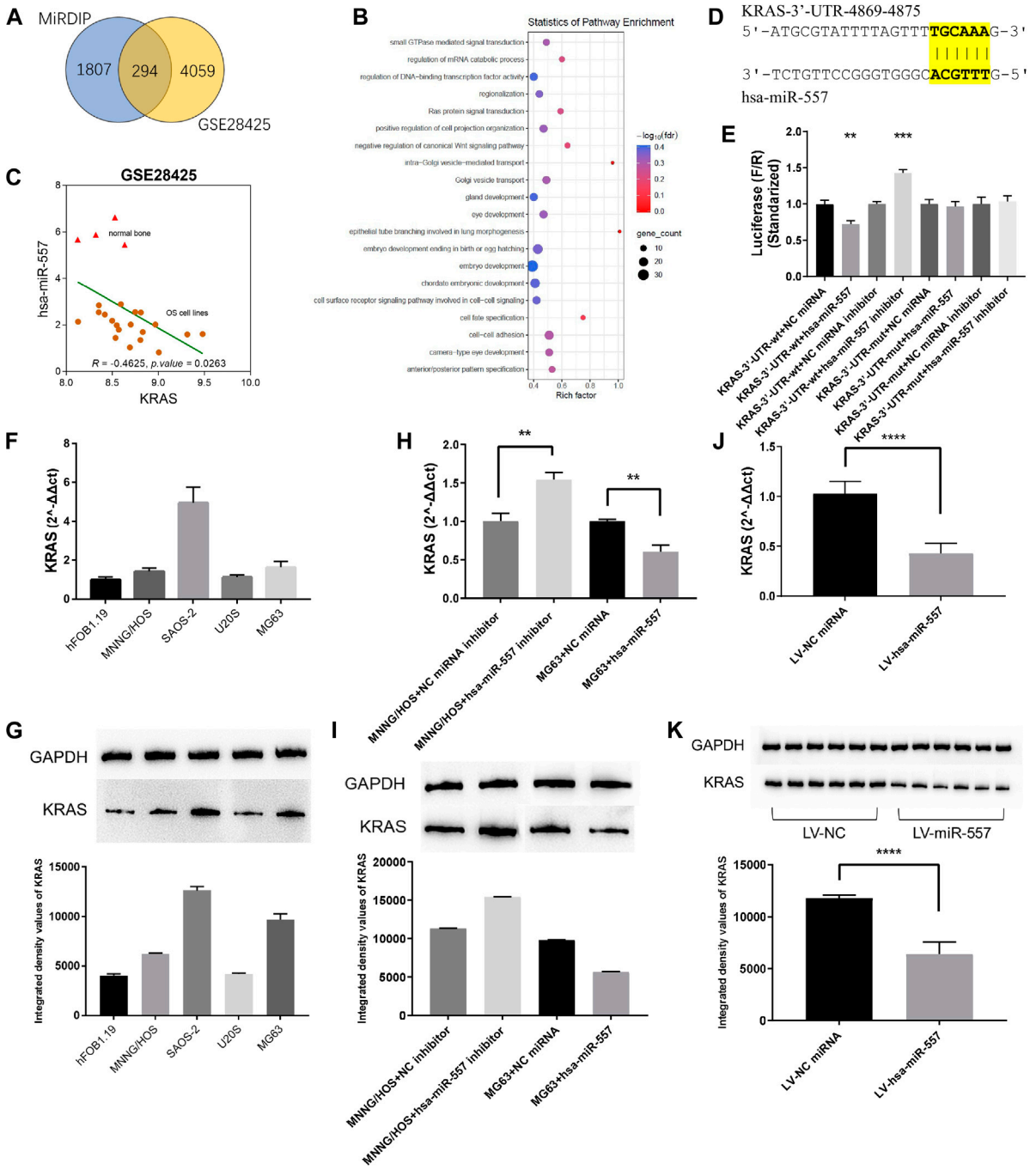


FIGURE 4 | Hsa-miR-557 inhibited the growth of osteosarcoma through KRAS. **(A)** Combing the GSE28425 dataset with the miRDIP database, 294 hsa-miR-557 target genes were found to be significantly downregulated. **(B)** Pathway enrichment using STRING showed that the RAS signaling pathway was significantly involved. **(C)** Data from GSE28425 showed a significant negative correlation between KRAS and hsa-miR-557. **(D)** Simulation of hsa-miR-557 target sequence of KRAS 3′UTR. **(E)** Dual-luciferase reporter assay showed that lower levels of F/R were detected in the KRAS-3′U TR-wt + hsa-miR-557 group, and higher levels of F/R were detected in the KRAS-3′UTR-wt + hsa-miR-557 inhibitor group. **(F,G)** Expression level of KRAS in osteosarcoma cell lines (MNNG/HOS, SAOS-2, U20S, and MG63) and normal osteoblasts (hFOB1.19). **(H,I)** Levels of KRAS were inhibited by hsa-miR-557 in MG63 cells and promoted by hsa-miR-557 inhibitor in MNNG/HOS cells. **(J,K)** In the nude mice model, lower levels of KRAS expression were tested in the hsa-miR-557 vector group (**$p < 0.01$, ****$p < 0.0001$).

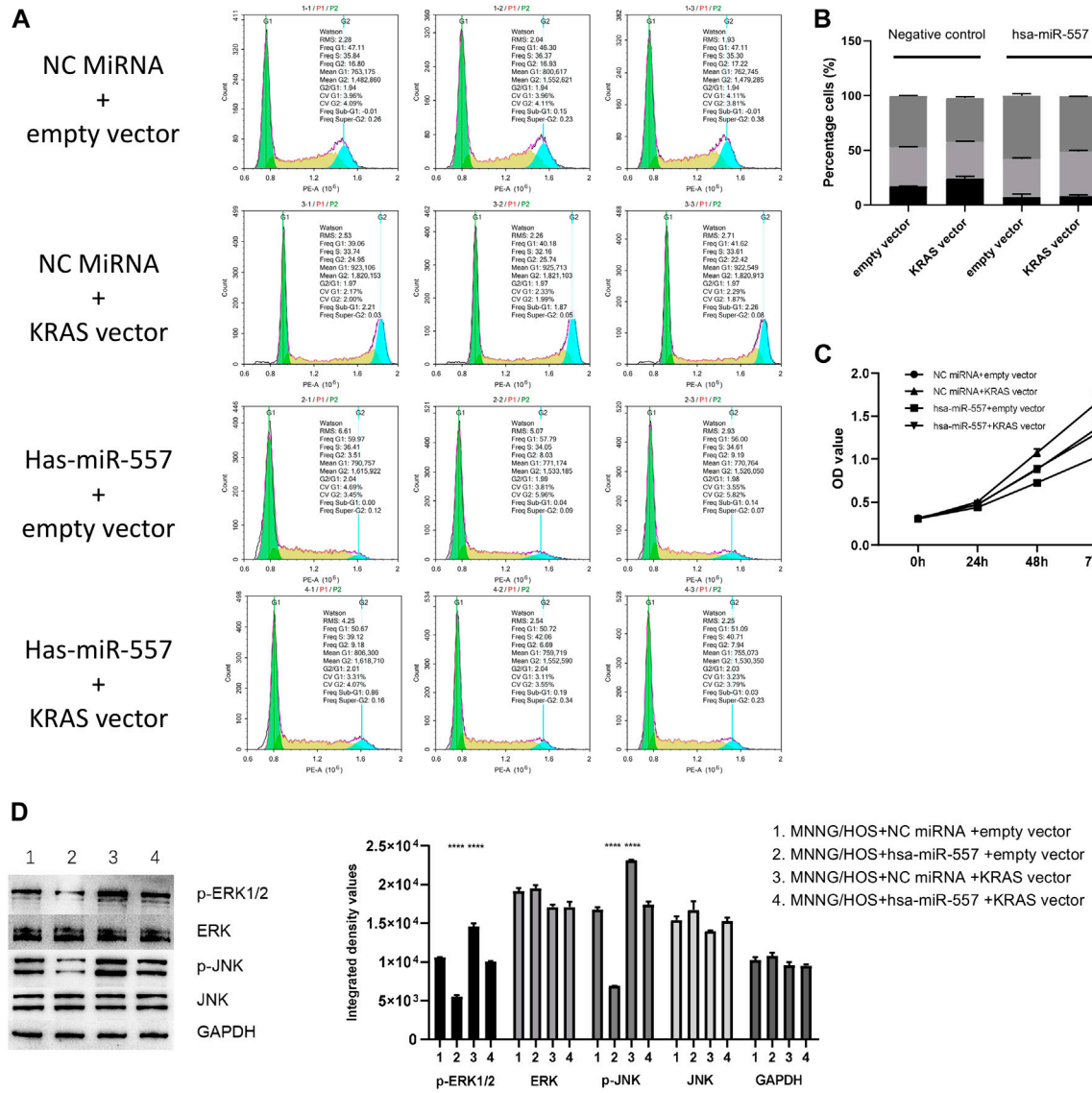


FIGURE 5 | Hsa-miR-557 inhibits the growth of osteosarcoma through targeting KRAS and its downstream molecules. **(A,B)** Cell cycle analysis showed that hsa-miR-557 could inhibit the proliferation of osteosarcoma cells through KRAS (NC miRNA + empty vector vs NC miRNA + KRAS vector: $p = 0.0021$, t = 7.114, NC miRNA + empty vector vs hsa-miR-557 + empty vector: $p = 0.0007$, t = 9.457, hsa-miR-557 + KRAS vector vs hsa-miR-557 + empty vector: $p = 0.0038$, t = 6.039). **(C)** CCK-8 showed that KRAS could alleviate the inhibition effect of hsa-miR-557 on osteosarcoma (NC miRNA + empty vector vs hsa-miR-557 + empty vector: p (48 h) = 0.0003, q (48 h) = 6.522, p (72 h) < 0.0001, q (72 h) = 13.84); NC miRNA + empty vector vs NC miRNA + KRAS: p (48 h) < 0.0001, q (48 h) = 8.080, p (72 h) < 0.0001, q (72 h) = 12.380; NC miRNA + KRAS vector vs hsa-miR-557 + KRAS vector, p (48 h) < 0.0001, q (48 h) = 7.815, p (72 h) < 0.0001, q (72 h) = 14.72). **(D)** Phosphorylation of ERK1/2 and JNK was promoted by KRAS with NC miRNA and inhibited by hsa-miR-557 mimic (****p < 0.0001).

2019). Mir-21 was found to be overexpressed in osteosarcoma patients, and miR-21 knockout could reduce the invasion of the MG63 osteosarcoma cell line (Zhang et al., 2019).

Hsa-miR-557 was previously reported to be related to tumorigenesis and the survival rate. Abnormal hsa-miR-557 expression was found in patients with hepatocellular carcinoma (Katayama et al., 2012), gastric cancer (Jiang et al., 2014), pancreatic ductal adenocarcinoma (Zhang et al., 2018b), and lung cancers (Qiu et al., 2017). In addition, hsa-miR-557 was reported to be a tumor suppressor. Qiu et al. reported that miR-557 could inhibit lung cancer by negatively regulating lymphocyte enhancement factor 1. Razaviyan et al. found that mir-557 could inhibit S6K1 in breast cancer (Razaviyan et al., 2018). MiR-557 was also found to suppress pancreatic cancer cells through miR-557/SLC7A11/PI3K/AKT (Zhang et al., 2021). In this study, we found that hsa-miR-557 was positively correlated with the survival rate of osteosarcoma patients. Through *in vivo* and *in vitro* study, hsa-miR-557 was proved to be a tumor suppressor of osteosarcoma.

In a recent study, Wang et al. noticed the suppressive effect of hsa-miR-557 on osteosarcoma (Wang et al., 2021). Their study found that hsa-miR-557 could inhibit the malignant behavior of osteosarcoma by reducing HOXB9, which can deactivate the epithelial-mesenchymal transition (EMT) process. In our study, KRAS was found to be a target of hsa-miR-557, inhibiting the proliferation of osteosarcoma cells. As known, each miRNA could have multiple targets, which modulate cellular behaviors in the same or different directions. Combining these two studies, more evidence was offered, proving that hsa-miR-557 plays an inhibiting role in osteosarcoma. Therefore, hsa-miR-557 could be a promising treatment molecule in osteosarcoma.

MiRNAs inhibit translation or degrade mRNA molecules through binding to the complementary mRNA sequence (Rupaimoole and Slack, 2017). Several tumor-related genes were reported previously as targets of has-miR-557. Yang et al. found EGFR to be one of the has-miR-557 targets in inhibiting the proliferation and invasion of pancreatic cancer (Yang et al., 2019). Qiu et al. found that miR-557 could suppress lung cancer by regulating LEF1 (Qiu et al., 2017). In this study, KRAS was proved to be one of the hsa-miR-557 target molecules. KRAS is a famous mutated oncogene seen in a variety of tumors, among them is osteosarcoma. The activation of KRAS could switch on downstream pathways causing cell growth, differentiation, and influence survival (Uprety and Adjei, 2020). At the same time, the disturbance of KRAS suppressed tumor growth (Zhang et al., 2018a; Chen et al., 2019). ERK and JNK are important RAS downstream molecules, controlling cell growth and survival. Phosphorylation levels of ERK and JNK are associated with poor prognosis of osteosarcoma patients (Li et al., 2016; Czarnecka et al., 2020). The inhibition of ERK and JNK phosphorylation by hsa-miR-557 indicates its potential in treating osteosarcoma. Additionally, ERK and JNK are downstream molecules of KRAS in tumorigenesis (Cellurale et al., 2011; Samatar and Poulikakos, 2014). Modulation of KRAS downstream molecules (ERK and JNK) by hsa-miR-557 reconfirmed the relationship between them.

In conclusion, the present study demonstrated that hsa-miR-557 could inhibit the proliferation of osteosarcoma cells through modulating KRAS expression.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

ETHICS STATEMENT

The animal study was reviewed and approved by the ethics committee of the First Affiliated Hospital of Zhengzhou University.

AUTHOR CONTRIBUTIONS

ZQ performed experiments and drafted the manuscript. JL, HK, XC, DB, GS, SC, YJ, and TC performed experiments and revised the manuscript. YW and HL designed the study and revised the manuscript.

FUNDING

This project is supported by the Medical science and Technology project of Henan Province (LHGJ20190169).

ACKNOWLEDGMENTS

We would like to thank "Translational Medical Center at the First Affiliated Hospital of Zhengzhou University" for their access of analytic instruments.

REFERENCES

Cellurale, C., Sabio, G., Kennedy, N. J., Das, M., Barlow, M., Sandy, P., et al. (2011). Requirement of C-Jun NH2-Terminal Kinase for Ras-Initiated Tumor Formation. *Mol. Cel Biol* 31, 1565–1576. doi:10.1128/mcb.01122-10

Chen, J., Yan, C., Yu, H., Zhen, S., and Yuan, Q. (2019). miR-548d-3p Inhibits Osteosarcoma by Downregulating KRAS. *Aging* 11, 5058–5069. doi:10.18632/aging.102097

Czarnecka, A. M., Synoradzki, K., Firlej, W., Bartnik, E., Sobczuk, P., Fiedorowicz, M., et al. (2020). Molecular Biology of Osteosarcoma. *Cancers (Basel)* 12. doi:10.3390/cancers12082130

Dai, N., Qing, Y., Cun, Y., Zhong, Z., Li, C., Zhang, S., et al. (2018). miR-513a-5p Regulates Radiosensitivity of Osteosarcoma by Targeting Human Apurinic/apyrimidinic Endonuclease. *Oncotarget* 9, 25414–25426. doi:10.18632/oncotarget.11003

Ji, Q., Xu, X., Song, Q., Xu, Y., Tai, Y., Goodman, S. B., et al. (2018). miR-223-3p Inhibits Human Osteosarcoma Metastasis and Progression by Directly Targeting CDH6. *Mol. Ther.* 26, 1299–1312. doi:10.1016/j.ymthe.2018.03.009

Jiang, H. B., Yang, T. J., Lu, P., and Ma, Y. J. (2014). Gene Expression Profiling of Gastric Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 18, 2109–2115.

Kansara, M., Teng, M. W., Smyth, M. J., and Thomas, D. M. (2014). Translational Biology of Osteosarcoma. *Nat. Rev. Cancer* 14, 722–735. doi:10.1038/nrc3838

Katayama, Y., Maeda, M., Miyaguchi, K., Nemoto, S., Yasen, M., Tanaka, S., et al. (2012). Identification of Pathogenesis-Related microRNAs in Hepatocellular Carcinoma by Expression Profiling. *Oncol. Lett.* 4, 817–823. doi:10.3892/ol.2012.810

Li, Y.-S., Deng, Z.-H., Zeng, C., and Lei, G.-H. (2016). JNK Pathway in Osteosarcoma: Pathogenesis and Therapeutics. *J. Receptors Signal Transduction* 36, 465–470. doi:10.3109/10799893.2015.1122045

Liu, Q., Song, Y., Duan, X., Chang, Y., and Guo, J. (2018). MiR-92a Inhibits the Progress of Osteosarcoma Cells and Increases the Cisplatin Sensitivity by Targeting Notch1. *Biomed. Res. Int.* 2018, 9870693. doi:10.1155/2018/9870693

Long, X., and Lin, X. J. (2019). P65-mediated miR-590 Inhibition Modulates the Chemoresistance of Osteosarcoma to Doxorubicin through Targeting Wild-type P53-induced Phosphatase 1. *J. Cel Biochem* 120, 5652–5665. doi:10.1002/jcb.27849

Namlos, H. M., Meza-Zepeda, L. A., Baroy, T., Ostensen, I. H., Kresse, S. H., Kuijjer, M. L., et al. (2012). Modulation of the Osteosarcoma Expression Phenotype by microRNAs. *PLoS One* 7, e48086.

Qiu, J., Hao, Y., Huang, S., Ma, Y., Li, X., Li, D., et al. (2017). MiR-557 Works as a Tumor Suppressor in Human Lung Cancers by Negatively Regulating LEF1 Expression. *Tumour Biol.* 39, 1010428317709467. doi:10.1177/1010428317709467

Razaviyan, J., Hadavi, R., Tavakoli, R., Kamani, F., Paknejad, M., and Mohammadi-Yeganeh, S. (2018). Expression of miRNAs Targeting mTOR and S6K1 Genes of mTOR Signaling Pathway Including miR-96, miR-557, and miR-3182 in Triple-Negative Breast Cancer. *Appl. Biochem. Biotechnol.* 186, 1074–1089. doi:10.1007/s12010-018-2773-8

Ritter, J., and Bielack, S. S. (2010). Osteosarcoma. *Ann. Oncol.* 21 (Suppl. 7), vii320–vii325. doi:10.1093/annonc/mdq276

Rupaimoole, R., and Slack, F. J. (2017). MicroRNA Therapeutics: towards a new era for the Management of Cancer and Other Diseases. *Nat. Rev. Drug Discov.* 16, 203–222. doi:10.1038/nrd.2016.246

Samatar, A. A., and Poulikakos, P. I. (2014). Targeting RAS-ERK Signalling in Cancer: Promises and Challenges. *Nat. Rev. Drug Discov.* 13, 928–942. doi:10.1038/nrd4281

Uprety, D., and Adjei, A. A. (2020). KRAS: From Undruggable to a Druggable Cancer Target. *Cancer Treat. Rev.* 89, 102070. doi:10.1016/j.ctrv.2020.102070

Wang, J., Liu, S., Shi, J., Li, J., Wang, S., Liu, H., et al. (2019). The Role of miRNA in the Diagnosis, Prognosis, and Treatment of Osteosarcoma. *Cancer Biother. Radiopharm.* 34, 605–613. doi:10.1089/cbr.2019.2939

Wang, Y., Hao, W., and Wang, H. (2021). miR-557 Suppressed the Malignant Behaviours of Osteosarcoma Cells by Reducing HOXB9 and Deactivating the EMT Process. *Artif. Cell Nanomedicine, Biotechnol.* 49, 230–239. doi:10.1080/21691401.2021.1890100

Wei, R., Cao, G., Deng, Z., Su, J., and Cai, L. (2016). miR-140-5p Attenuates Chemotherapeutic Drug-Induced Cell Death by Regulating Autophagy through Inositol 1,4,5-trisphosphate Kinase 2 (IP3k2) in Human Osteosarcoma Cells. *Biosci. Rep.* 36, e00392. doi:10.1042/BSR20160238

Wu, M., Wang, G., Tian, W., Deng, Y., and Xu, Y. (2018). MiRNA-based Therapeutics for Lung Cancer. *Cpd* 23, 5989–5996. doi:10.2174/1381612823666170714151715

Xu, W., Li, Z., Zhu, X., Xu, R., and Xu, Y. (2018). miR-29 Family Inhibits Resistance to Methotrexate and Promotes Cell Apoptosis by Targeting COL3A1 and MCL1 in Osteosarcoma. *Med. Sci. Monit.* 24, 8812–8821. doi:10.12659/msm.911972

Yang, L., Li, H., and Huang, A. (2020). MiR-429 and MiR-143-3p Function as Diagnostic and Prognostic Markers for Osteosarcoma. *Clin. Lab.* 66. doi:10.7754/Clin.Lab.2020.191237

Yang, Y., Sun, K. K., Shen, X. J., Wu, X. Y., and Li, D. C. (2019). miR-557 Inhibits the Proliferation and Invasion of Pancreatic Cancer Cells by Targeting EGFR. *Int. J. Clin. Exp. Pathol.* 12, 1333–1341.

Yang, Z., Wa, Q. D., Lu, C., Pan, W., Lu, Z. M., and Ao, J. (2018). miR-328-3p Enhances the Radiosensitivity of Osteosarcoma and Regulates Apoptosis and Cell Viability via H2AX. *Oncol. Rep.* 39, 545–553. doi:10.3892/or.2017.6112

Zhang, B., Liu, Y., and Zhang, J. (2019). Silencing of miR-19a-3p Enhances Osteosarcoma Cells Chemosensitivity by Elevating the Expression of Tumor Suppressor PTEN. *Oncol. Lett.* 17, 414–421. doi:10.3892/ol.2018.9592

Zhang, H., He, Q. Y., Wang, G. C., Tong, D. K., Wang, R. K., Ding, W. B., et al. (2018). miR-422a Inhibits Osteosarcoma Proliferation by Targeting BCL2L2 and KRAS. *Biosci. Rep.* 38. 38BSR20170339. doi:10.1042/BSR20170339

Zhang, T., Li, M., Lu, H., and Peng, T. (2021). Up-Regulation of circEIF6 Contributes to Pancreatic Cancer Development through Targeting miR-557/SLC7A11/PI3K/AKT Signaling. *Cmar* 13, 247–258. doi:10.2147/cmar.s280307

Zhang, W., Wei, L., Sheng, W., Kang, B., Wang, D., and Zeng, H. (2020). miR-1225-5p Functions as a Tumor Suppressor in Osteosarcoma by Targeting Sox9. *DNA Cel Biol.* 39, 78–91. doi:10.1089/dna.2019.5105

Zhang, X., Guo, Q., Chen, J., and Chen, Z. (2015). Quercetin Enhances Cisplatin Sensitivity of Human Osteosarcoma Cells by Modulating microRNA-217-KRAS Axis. *Mol. Cell* 38, 638–642. doi:10.14348/molcells.2015.0037

Zhang, Y., Zhang, R., Ding, X., and Ai, K. (2018). EFNB2 Acts as the Target of miR-557 to Facilitate Cell Proliferation, Migration and Invasion in Pancreatic Ductal Adenocarcinoma by Bioinformatics Analysis and Verification. *Am. J. Transl Res.* 10, 3514–3528.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Qiao, Li, Kou, Chen, Bao, Shang, Chen, Ji, Cheng, Wang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fgene.2021.814843

Global Characterization of Metabolic Genes Regulating Survival and Immune Infiltration in Osteosarcoma

*Zhongpei Zhu[1†], Min Zhang[1†], Weidong Wang[1], Peng Zhang[2], Yuqiang Wang[1] and Limin Wang[1]**

[1]*Department of Orthopedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China,* [2]*Department of Orthopedics, Tumor Hospital of Henan Province, Zhengzhou, China*

OPEN ACCESS

Edited by:
Apeng Chen, Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Muhammad Riaz Khan, Université de Sherbrooke, Canada
Furhan Iqbal, Bahauddin Zakariya University, Pakistan

***Correspondence:**
Limin Wang
gu2keo@126.com

[†]*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *14 November 2021*
Accepted: *02 December 2021*
Published: *13 January 2022*

Citation:
Zhu Z, Zhang M, Wang W, Zhang P, Wang Y and Wang L (2022) Global Characterization of Metabolic Genes Regulating Survival and Immune Infiltration in Osteosarcoma. Front. Genet. 12:814843. doi: 10.3389/fgene.2021.814843

Background: The alterations in metabolic profile of tumors have been identified as one of the prognostic hallmarks of cancers, including osteosarcoma. These alterations are majorly controlled by groups of metabolically active genes. However, the regulation of metabolic gene signatures in tumor microenvironment of osteosarcoma has not been well explained.

Objectives: Thus, we investigated the sets of previously published metabolic genes in osteosarcoma patients and normal samples.

Methods: We applied computational techniques to identify metabolic genes involved in the immune function of tumor microenvironment (TME) and survival and prognosis of the osteosarcoma patients. Potential candidate gene *PAICS* (*phosphoribosyl aminoimidazole carboxylase, phosphoribosyl aminoimidazole succino carboxamide synthetase*) was chosen for further studies in osteosarcoma cell lines for its role in cell proliferation, migration and apoptosis.

Results: Our analyses identified a list of metabolic genes differentially expressed in osteosarcoma tissues. Next, we scrutinized the list of genes correlated with survival and immune cells, followed by clustering osteosarcoma patients into three categories: C1, C2, and C3. These analyses led us to choose *PAICS* as potential candidate gene as its expression showed association with poor survival and negative correlation with the immune cells. Furthermore, we established that loss of *PAICS* induced apoptosis and inhibited proliferation, migration, and wound healing in HOS and MG-63 cell lines. Finally, the results were supported by constructing and validating a prediction model for prognosis of the osteosarcoma patients.

Conclusion: Here, we conclude that metabolic genes specifically *PAICS* play an integral role in the immune cell infiltration in osteosarcoma TME, as well as cancer development and metastasis.

Keywords: osteosarcoma, immune infiltration, metabolic genes, differentially expressed genes, PAICS

INTRODUCTION

Osteosarcoma is a rare type of cancer that affects the individuals of all ages including children (Yang et al., 2021). However, because of the advancements in chemotherapy, precise radiotherapy, and immunotherapy, the overall 5-year overall survival (OS) rate of patients has substantially increased to 70% (Li et al., 2021). It is noteworthy that only 15% to 20% of the patients were usually diagnosed with metastases, and the OS of these patients was reported to be extremely poor (Durnali et al., 2013; Zhang et al., 2019). Recently, immunotherapy, such as adoptive cellular therapy, vaccination, and checkpoint inhibitors, has been considered as the effective therapies for osteosarcoma (Lettieri et al., 2016; Heymann et al., 2019). Therefore, it is worth studying the underlying molecular mechanisms of occurrence and development of osteosarcoma with a deep focus on the identification of novel diagnostic, therapeutic, and prognostic markers (Yiqi et al., 2020).

In recent years, much attention has been paid to the contents of the tumor microenvironment (TME) and their roles in cancer development. It has been found that TME plays a critical role in the cancer development and recurrence (Ren et al., 2021). In short, the TME contains a variety of cells, including tumor cells, fibroblasts, endothelial cells, immune cells, various signal molecules, and extracellular matrix (Cortini et al., 2016; Chen et al., 2020; Luo et al., 2020). The increasing evidence shows that infiltrating immune cells such as T cells, B cells, macrophages, dendritic cells, monocytes, neutrophils, and mast cells may be involved in cancer development and progression (Du et al., 2021; Gerard et al., 2021; Kemmerer et al., 2021). In TME, tumor cells can invade surrounding tissues or metastasize through lymphatic vessels, and the infiltrated cells can stimulate host immune response for releasing cytokines, cytokine receptors, and other factors, which directly or indirectly promote or inhibit tumor cell proliferation (Federico et al., 2021; Zhang et al., 2021). The osteosarcoma TME is now considered as an essential element of tumor growth and dissemination (Corre et al., 2020). A latest study has shown the association of the tumor immune cell infiltration with clinical outcomes of osteosarcoma patients (Chen et al., 2020). Extensive studies on the TME have shown that infiltrating immune cells play a vital role in tumor growth, recurrence, metastasis, and response to the immunotherapy (Ma et al., 2020; Zhang et al., 2020). However, the detailed profile of immune cells infiltrating in osteosarcoma TME has not been elucidated yet (Zhang et al., 2020).

Previous studies have primarily focused on the one or two kinds of immune cells or key genes, which could bias the osteosarcoma microenvironment exploration. The identification of multiple genes from tumor-infiltrating immune cell profiles can help to construct a gene signature with better and more accurate prognostic potential. So far, no study has explained the role of metabolic genes in immune cell infiltration osteosarcoma TME and cancer development. Here, for the first time, we collected raw RNA-seq data from two different studies (Haider et al., 2016; Peng et al., 2018) and identified the top 10 differentially expressed metabolic genes in osteosarcoma tumors. Furthermore, we performed an *in silico* analysis to find their roles in osteosarcoma tumor progression. We also studied the role of these genes in immune cell infiltration in osteosarcoma TME, and a prediction model was also developed to predict the OS and prognosis of the osteosarcoma patients.

MATERIALS AND METHODS

Clinical Samples and Data Acquisition

The RNA sequencing data of osteosarcoma samples and normal tissues were downloaded from two previously published studies (Haider et al., 2016; Peng et al., 2018). Complete clinical information (age, sex, primary tumor site, metastatic state at diagnosis, survival time, and survival state) were also included in our study. We included only genes expressing more than 50% of the samples and had expression level >1. The detailed workflow of the study is shown in **Supplementary Figure S1**.

DEGs and Enrichment Analysis

The DEGs were identified by the difference of logFC in the expression between tumor and normal samples, genes having logFC >1 were considered for analysis. All genes were analyzed using R package limma. The differences in expression of the genes were presented as heat map. The selected genes were analyzed for the molecular, biological, and cellular functions by Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment programs.

Construction and Validation of Prediction Model

The univariate and multivariate Cox HR regression analyses were performed to estimate the risk score and clinical features of the osteosarcoma patients. The risk score and clinical factors were used to identify the independent risk factors associated with the OS. The area under the receiver operating characteristic (ROC) curve (AUC) was calculated by using survival ROC R package to validate the performance of the prediction model. The 95% confidence intervals for AUCs were obtained using the nonparametric bootstrap via survAccuracy Measures package. For an internal validation, the calibration curve was plotted to assess the performance of the prediction model for prognosis and survival of the osteosarcoma patients.

Analyses of Immune Cell Infiltration in TME

The R package xCELL was used to explore the immune cell infiltration in osteosarcoma TME. Furthermore, we used ESTIMATE to analyze the purity of the osteosarcoma tumors. Based on the metabolic transcriptional profiles of the patients, the stromal and immune scores of the osteosarcoma tissues were estimated. The association among the infiltrating immune cells was also studied, and $p < 0.05$ was set as cutoff for analysis.

Survival Analysis

The DEGs in osteosarcoma patients were studied for OS analysis. The univariate and multivariate Cox regression

analyses were performed using R package. The Kaplan–Meier survival curves were plotted, and ggplot2 was used to visualize the data. The variables associated with the survival were screened out using LASSO (Ichikawa and Brenner, 1981); variables having a regression coefficient more than zero were selected for analyses.

Cell Culture

The human renal cancer cell lines HOS and MG-63 were purchased from the American Type Culture Collection (Manassas, VA, USA) and cultured in modified eagle medium (SH30244; Hyclone, Logan, UT, USA) supplemented with 10% fetal bovine serum (FBS; 10,099,141; Invitrogen, Carlsbad, CA, USA) at 37°C and 5% CO_2 in a humidified incubator.

Real-Time Quantitative Polymerase Chain Reaction

The osteosarcoma cells were first lysed with TRIzol (9,109; TaKaRa, Tokyo, Japan), then whole RNA was extracted using standard chloroform/isopropanol method. The RNA samples were quantified by using NanoDrop, and cDNA was synthesized by using reverse transcriptase enzyme (205,311; Qiagen, Duesseldorf, Germany). Relative expression of genes was studied by performing real-time polymerase chain reaction (PCR) using Quantinova Syber green PCR kit (208,054; Qiagen). GAPDH was used as internal control. Primer sequences are listed in the **Supplementary Table S1**.

Cell Function Assays

PAICS (*phosphoribosyl aminoimidazole carboxylase, phosphoribosyl aminoimidazole succino carboxamide synthetase*) was knocked down using siRNAs in osteosarcoma cells HOS and MG-63. The PAICS siRNAs and negative control were synthesized by GenePharma Corporation (Shanghai, China). The sequences are provided in **Supplementary Table S1**. To assess the cells' viability, after 24 h of transfection, 1,000 cells were seeded into each well of the 96-well plates. Cell Counting Kit-8 (CK04; Dojindo Laboratories, Kumamoto, Japan) was used to examine the proliferation capacity of the cells at 0, 24, 48, and 72 h after seeding cells. Absorbance was measured at 450 nm. Standard Transwell columns were used to study the effect of *PAICS* knockdown on cell migration; an equal number of cells (NC and si-PAICS) were seeded into in upper chamber of the Transwell column containing FBS less media; the lower chamber of the column was supplied with media supplemented with 10% FBS. The migration of cells from the upper column to the lower column was observed after 24 h by simply blocking with 4% paraformaldehyde (PFA) and staining with 0.5% crystal violet, visualized, and photographed by inverted microscope fitted with Nikon camera. For apoptosis, 72 h after transfecting cells with siRNA, cells were treated with an apoptosis detection kit (KGA1013; KeyGEN BioTECH, Nanjing, China) and analyzed by flow cytometry (FACScan; BD Biosciences). The wound healing assay was performed by creating a monolayer at 90% confluence of cells after transfecting with siRNA in six-well plates. The difference in the wound recovery was photographed and measured for 24 h. The colony formation assay was performed by seeding cells transfected with siRNA in six-well plates (1,000 cells/well) and cultured for 10 days. Then, cells were fixed with 4% PFA and stained with 0.1% crystal violet; the colonies were photographed and counted for determining differences between NC and knockdown cells.

Statistical Analysis

The statistical analyses were performed by using SPSS 25.0 (IBM, Inc., Chicago, IL, USA), GraphPad Prism 7 (GraphPad Software, San Diego, CA, USA), and R packages (version 3.6.4). The LASSO regression analyses were performed by using glmnet R package. The comparison of the immune cell infiltration between tumor and normal tissues was analyzed using Wilcoxon test. However, the Cox proportional hazards model by Survival R package and log-rank test were used to calculate the p value for survival analyses. $p < 0.05$ was considered as significant, and all experiments were performed in triplicate.

RESULTS

Identification of Metabolic Genes With OS

To identify the metabolic genes correlated with OS in osteosarcoma patients, we first downloaded the raw RNA-seq data of osteosarcoma patients mentioned in two different studies (Haider et al., 2016; Peng et al., 2018). We intersected only metabolic genes with expression greater than 1 and have 50% coverage across cancer tissues. The overall flowchart of the study is shown in **Supplementary Figure S1**. In total, we identified 876 metabolic genes. Next, we classified the top 10 metabolic genes associated with survival (**Figure 1A**). Among the top 10 genes, 3 metabolic genes (*ALDH1A1*, *HNMT*, and *NUDT7*) were negatively correlated, and 7 (*GL O 1*, *SLC19A1*, *SQLE*, *PAICS*, *PPAT*, *FASN*, *AK2*) were positively correlated with the OS of the patients (**Figure 1A**). Afterward, we performed GO (**Figure 1B**) and KEGG (**Figure 1C**) enrichment analyses for all survival-related metabolic genes. We also constructed a protein–protein interaction (PPI) network for these metabolic genes associated with survival (**Supplementary Figure S2A**). Lastly, we isolated two subgroups across the top 10 hub survival-related metabolic genes that were interconnected with each other (**Supplementary Figure S2B**). Both GO and KEGG enrichment revealed nucleoside biosynthetic process and purine metabolism as top enrichment terms, respectively.

Identification of Metabolic Genes Correlated With $CD8^+$ T Cells

We analyzed all metabolic genes to find correlation with immune cells, especially activated $CD8^+$ T cells, which identified 206 differential metabolic genes correlated to immune cells. However, we found *PLAG2D*, *PIK3R5*, and *INPP5D* as the top three positively associated metabolic genes, with $R > 0.7$ (**Figure 2A**). On the other hand the *CTPS2*, *CACNB3*, and *ADCY6* were the top three negatively correlated metabolic

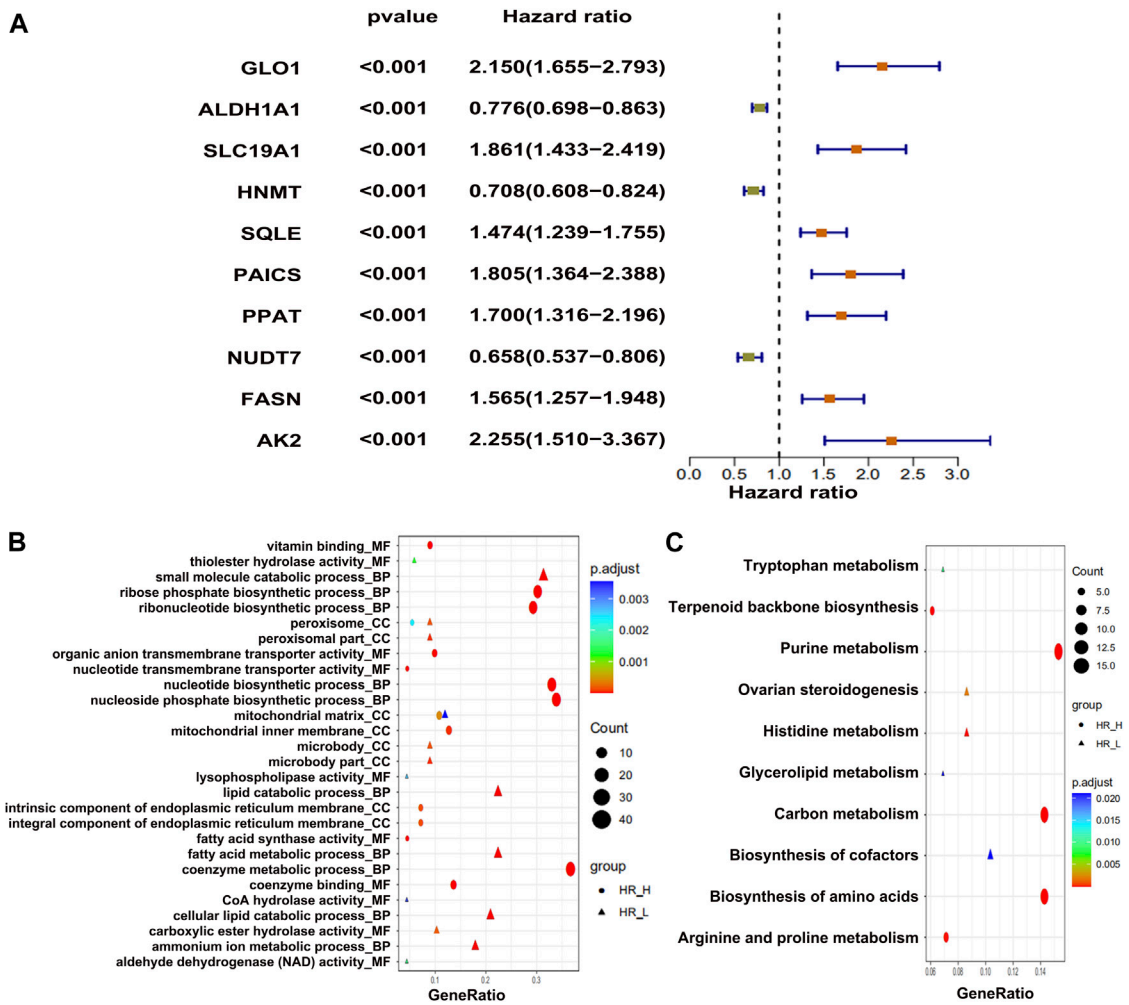


FIGURE 1 | The differentially expressed metabolic genes in osteosarcoma tumors. **(A)** Top 10 metabolic genes, showing positive and negative association with survival. **(B)** GO enrichment and **(C)** KEGG enrichment analysis of top metabolic genes.

genes with $R < -0.50$ (**Figure 2A**). GO enrichment analysis found that small_molecule_catabolic _process is one of the top BP terms associated with positively correlated metabolic genes (**Figure 2B**). On the contrary, the metabolic process of coenzymes was found to be enriched with negatively correlated metabolic genes (**Figure 2B**). KEGG pathway enrichment identified purine metabolism as the top pathway for negatively correlated genes (**Figure 2C**). Alternatively, arachidonic acid metabolism was positively correlated (**Figure 2C**). Next, we constructed PPI network for immune cell–related metabolic genes (**Supplementary Figure S3A**) and identified the top 10 hub immune-related genes (**Supplementary Figure S3B**). In conclusion, a total of 206 metabolic genes showed correlation with $CD8^+$ T immune cells in osteosarcoma patients.

Clustering Osteosarcoma Patients

Of a total 876 metabolic genes associated with OS and 206 associated with immune cells, we intersected common genes (**Figure 3A**). These common genes were then subjected to cluster analysis, and tumors were divided into three clusters: C1, C2, and C3 (**Figure 3B**). Subsequently, the heat map shows the expression of immune-survival related genes (**Figure 3C**). We observed differential expression of immune-survival related genes in three clusters (**Figure 3C**). The high variance between the three clusters was confirmed by principal component analysis (PCA) (**Figure 3D**). Furthermore, clusters C1 and C3 were closely grouped and compared with C2 in the PCA map of osteosarcoma. Lastly, in cluster grouping, we performed survival analyses and found that among the three clusters, C1 has the worst OS (**Figure 3E**).

Immune Status of the Clusters

To further identify the potential prognostic features for osteosarcoma, the stromal immune scores of metabolic gene signatures in clusters were determined (**Figure 4A**). The immune score in C1 was significantly low, whereas C3 had the

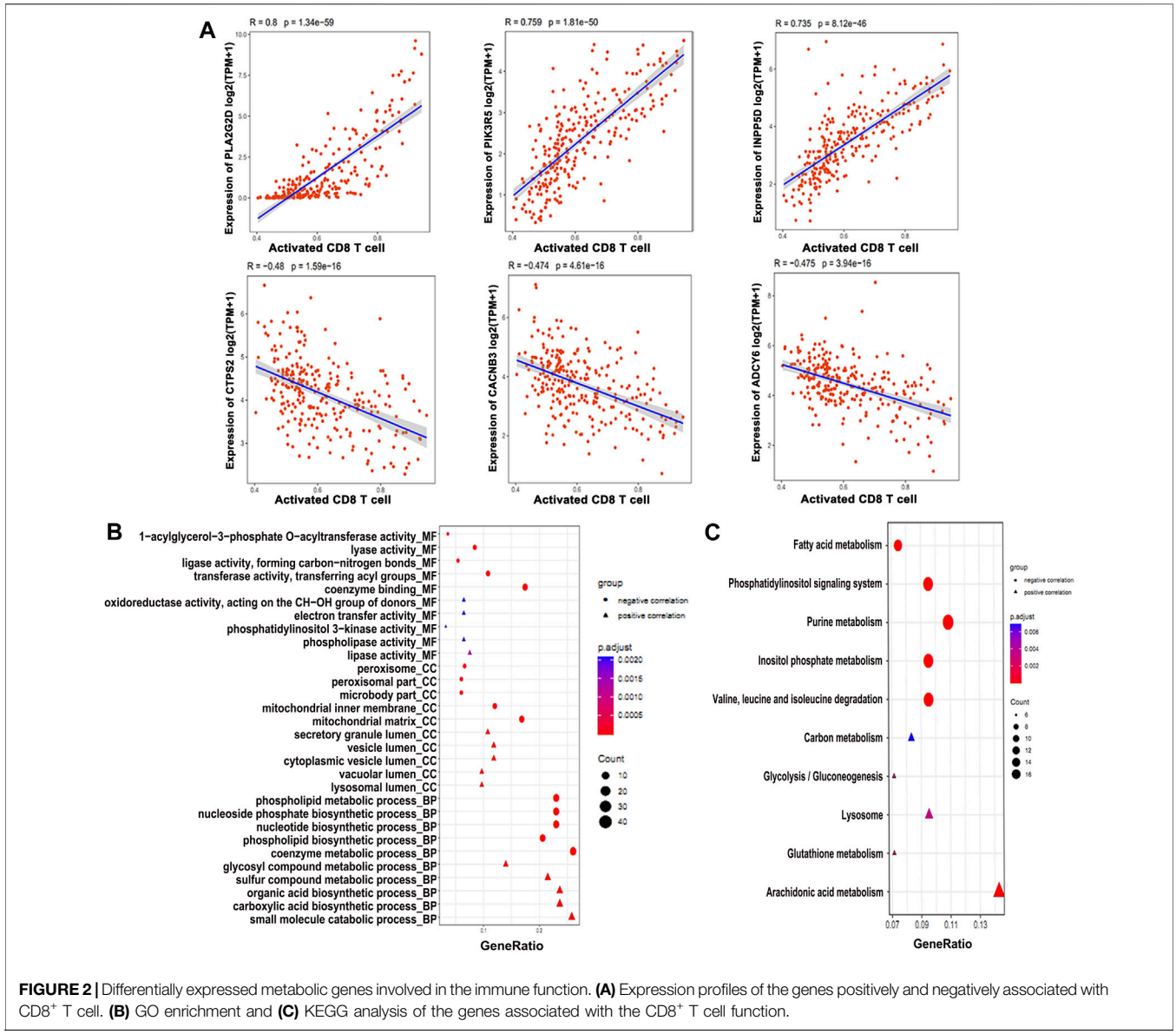


FIGURE 2 | Differentially expressed metabolic genes involved in the immune function. **(A)** Expression profiles of the genes positively and negatively associated with $CD8^+$ T cell. **(B)** GO enrichment and **(C)** KEGG analysis of the genes associated with the $CD8^+$ T cell function.

highest immune score (**Figure 4A**). Next, we determined the level of immune infiltration and types of the immune cells in all osteosarcoma clusters using ssGSEA and MCP-counter, respectively. Both methods are considered as reliable computational methods for estimating the infiltration of immune cells in TME (Barbie et al., 2009; Becht et al., 2016). The data obtained from ssGSEA and MCP-counter analysis were plotted in the form of a hierarchical heat map (**Figures 4B,C**). As expected, C3 showed the highest level of immune infiltration as compared to C1 and C2 clusters. Immune infiltration provided a clear picture of differential patterns of the cells among three clusters. Furthermore, we analyzed differential patterns of the genes in all osteosarcoma clusters and determined highly expressed differential genes in each cluster. The close observation of each cluster revealed that only C3 was enriched with the high expression of genes associated with immune cells (**Figure 4D**). Moreover, we illustrated an interacting map of highly expressed genes in three clusters via hub gene diagram (**Figure 4E**). Finally, we performed GO and KEGG pathway enrichment of highly expressed genes in all clusters (**Figures 4F,G**).

Phosphoribosyl Aminoimidazole Succino Carboxamide Synthetase Contributes to Poor Survival in Osteosarcoma Patients

Among the top 8 metabolic genes, *PAICS* attracted our attention as it is known to be significantly involved in the *de novo* purine biosynthesis and tumorigenesis (Huang et al., 2020). The expression analysis of *PAICS* in several cancers

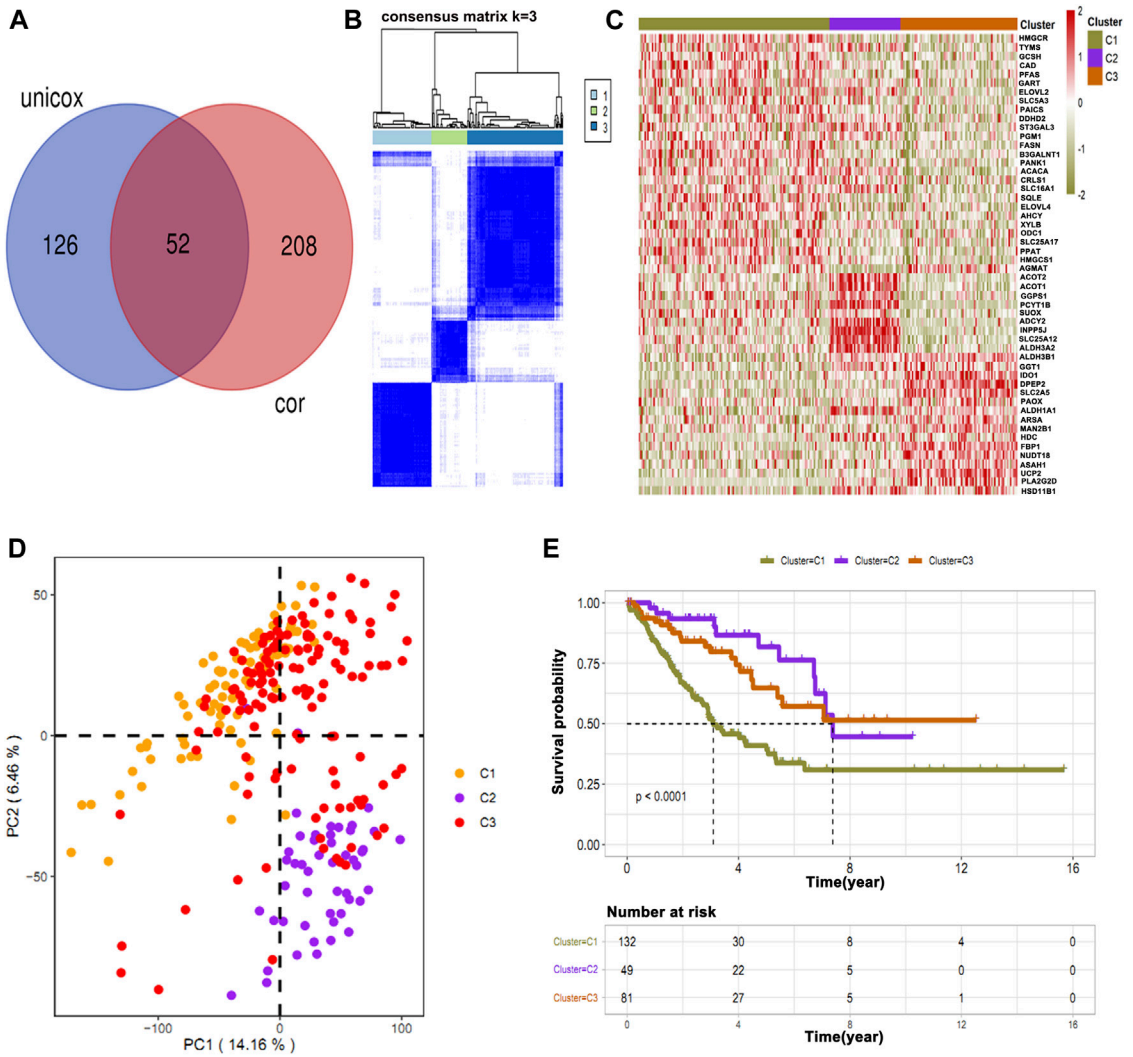


FIGURE 3 | Intersection of common immune-related genes and clustering of osteosarcoma patients. **(A)** Unicox analysis showing 52 common immune-related genes. **(B)** Patients were clustered based on common genes. **(C)** Heat map showing the expression of immune-survival related genes in different clusters. **(D)** PCA showing high variance in C2. **(E)** Survival analysis in three clusters showing C1 has a poor survival rate as compared to other clusters.

(The Cancer Genome Atlas [TCGA]) showed high expression of *PAICS* in most of the cancers including osteosarcoma (SARC) (**Supplementary Figure S4**). Survival analysis showed that patients with high expression of *PAICS* had poor survival rate, and *vice versa* (**Figures 5A,B**). Next, we performed GO and KEGG enrichment analyses of the genes having positive and negative correlations with *PAICS*. GO enrichment revealed positively associated genes linked with catalytic activity, ribosome, mitochondria function, and RNA catabolism processes (**Figure 5C**). Alternatively, negatively correlated genes of *PAICS* were enriched with immune-related functions, immune-related cellular components, and immune-related biological processes (**Figure 5C**). Moreover, KEGG pathway enrichment supplemented the GO enrichment analyses results (**Figure 5D**). Furthermore, we performed correlation analysis of *PAICS* expression with immune cells and found that activated B cells, activated $CD8^+$ T cells, and activated dendritic cells were negatively correlated with *PAICS* (**Figures 5E–H**). These findings explain that patients with high expression of *PAICS* have poor survival, which negatively correlated with the immune cells.

Loss of Phosphoribosyl Aminoimidazole Succino Carboxamide Synthetase Inhibits Osteosarcoma Cell Proliferation and Migration and Induces Apoptosis

After establishing the fact that *PAICS* is associated with poor survival, we next determined the effect of the loss of *PIACS* on osteosarcoma cell lines HOS and MG-63. We first verified the knockdown of *PAIC* using different siRNAs in HOS and MG-

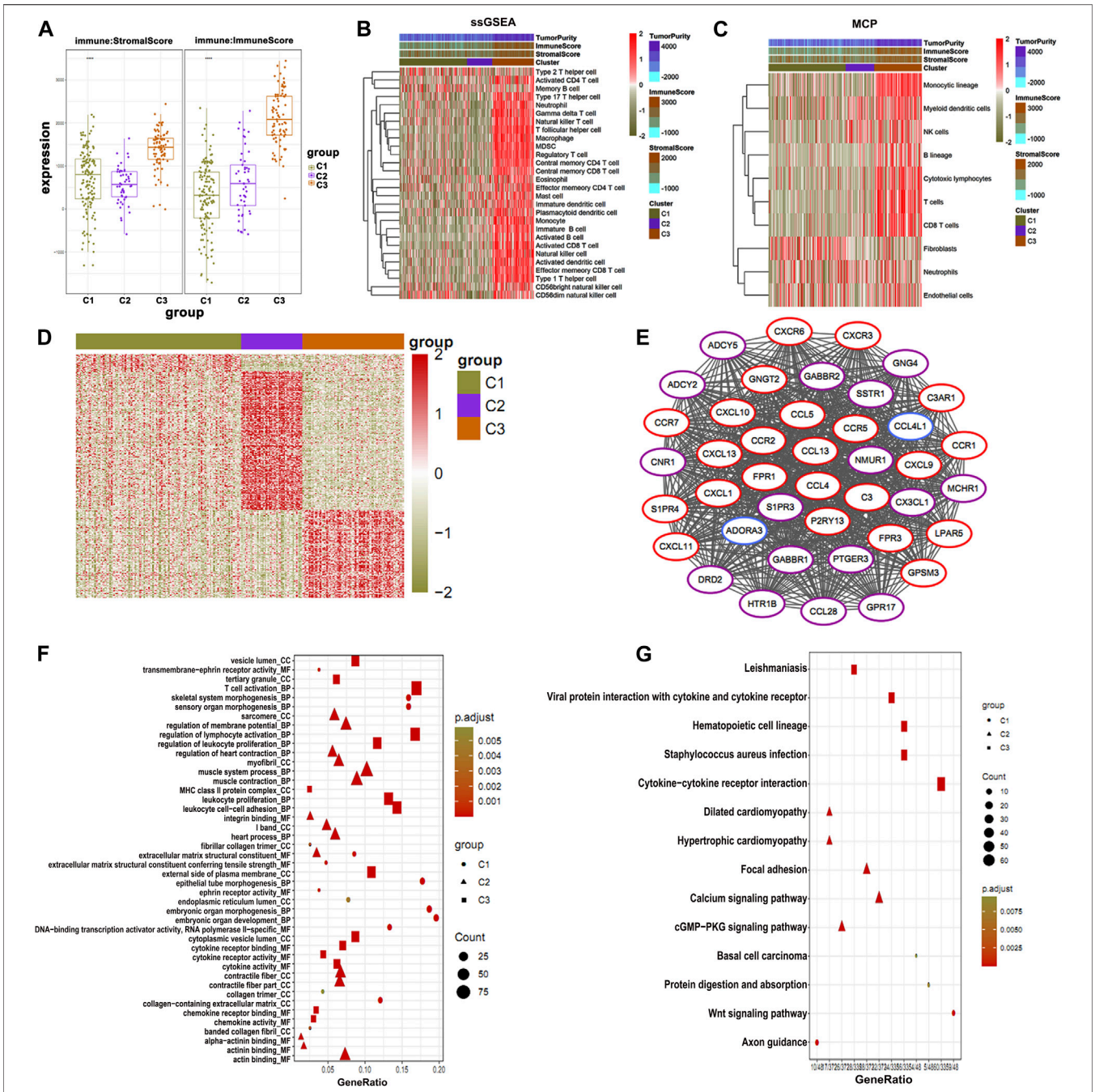


FIGURE 4 | Immune status of osteosarcoma clusters. **(A)** C1 shows the lowest immune and stromal score than the clusters. **(B)** The ssGSEA and **(C)** MCP-counter analysis plotted in the form of a hierarchical heat map (C3) show the highest immune score. **(D)** Highly expressed genes in three clusters. **(E)** PPI of highly expressed genes showing association with immune function. **(F)** GO enrichment and **(G)** KEGG enrichment analysis of the genes associated with immune function.

63 cell lines. We found that si-PAICS-1 and si-PAICS-2 can successfully deplete the expression of *PAICS* to 60% and 70%, respectively (**Figure 6A**). Next, we performed CCK8 assay to determine the effect of *PAICS*'s silencing on cell proliferation. It was observed that depletion of *PAICS* had significantly reduced cell viability and proliferation in both cell lines HOS and MG-63 (**Figures 6B,C**). This prompted us to investigate the cell death upon knockdown of *PAICS*. Then, we subjected both cell lines HOS and MG-63 to flow cytometry analyses using annexin V/propidium iodide (PI) apoptosis staining kit. The data generated by flow cytofluorometry analyses showed a high population of apoptotic cells in si-*PAICS* HOS and MG-63 cells (**Figure 6D**). Then, we quantified these apoptotic cells' population (6%–8%) and plotted

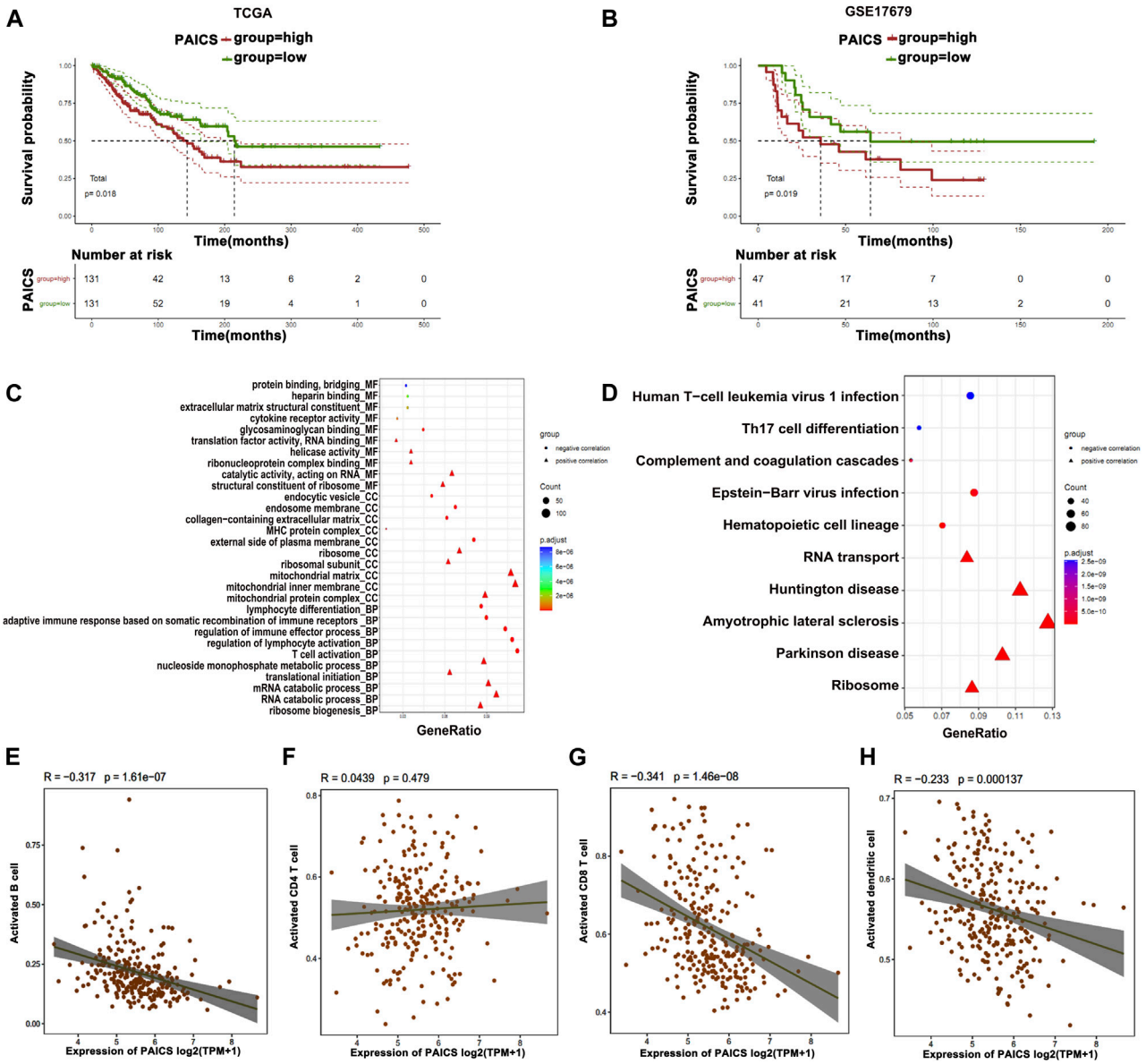


FIGURE 5 | Prognostic features of *PAICS*. **(A, B)** TCGA and GSE17679 databases show overexpression of *PAICS* associated with poor survival. **(C, D)** GO and KEGG enrichment analysis of *PAICS* gene showing its role in immune cells' function. Dot plots showing association of *PAICS* with **(E)** activated B cells, **(F)** activated $CD4^+$ T cells, **(G)** activated $CD8^+$ T cells, and **(H)** activated dendritic cells.

them in a bar chart (**Figure 6E**). Furthermore, cell migration was also reduced in the knockdown cells, which was verified by Transwell cell migration assay (**Figures 6F,G**). Afterward, we performed a wound healing assay upon depleting *PAICS* in both cell lines HOS and MG-63 (**Figures 6H,I**). We found the loss of wound repair ability of cells by 20%–30% in si-*PAICS* cells (**Figure 6J**). Finally, we determined the ability of cells to form colonies upon knockdown of *PIACS* (**Figures 6K,L**), which discovered that *PAICS* loss in HOS and MG-63 cell lines also inhibited the capacity to form colonies (**Figure 6M**). All the above results conclude that *PAICS* promotes tumor growth by inhibiting apoptosis and inducing migration of the cells.

Construction and Validation of Prediction Model

Based on the differential expression of the metabolic genes in clusters and their relationship to immune cells, we constructed and validated the prediction model by utilized metabolic genes. For this purpose, we devised two groups of low and highly expressed metabolic genes in osteosarcoma

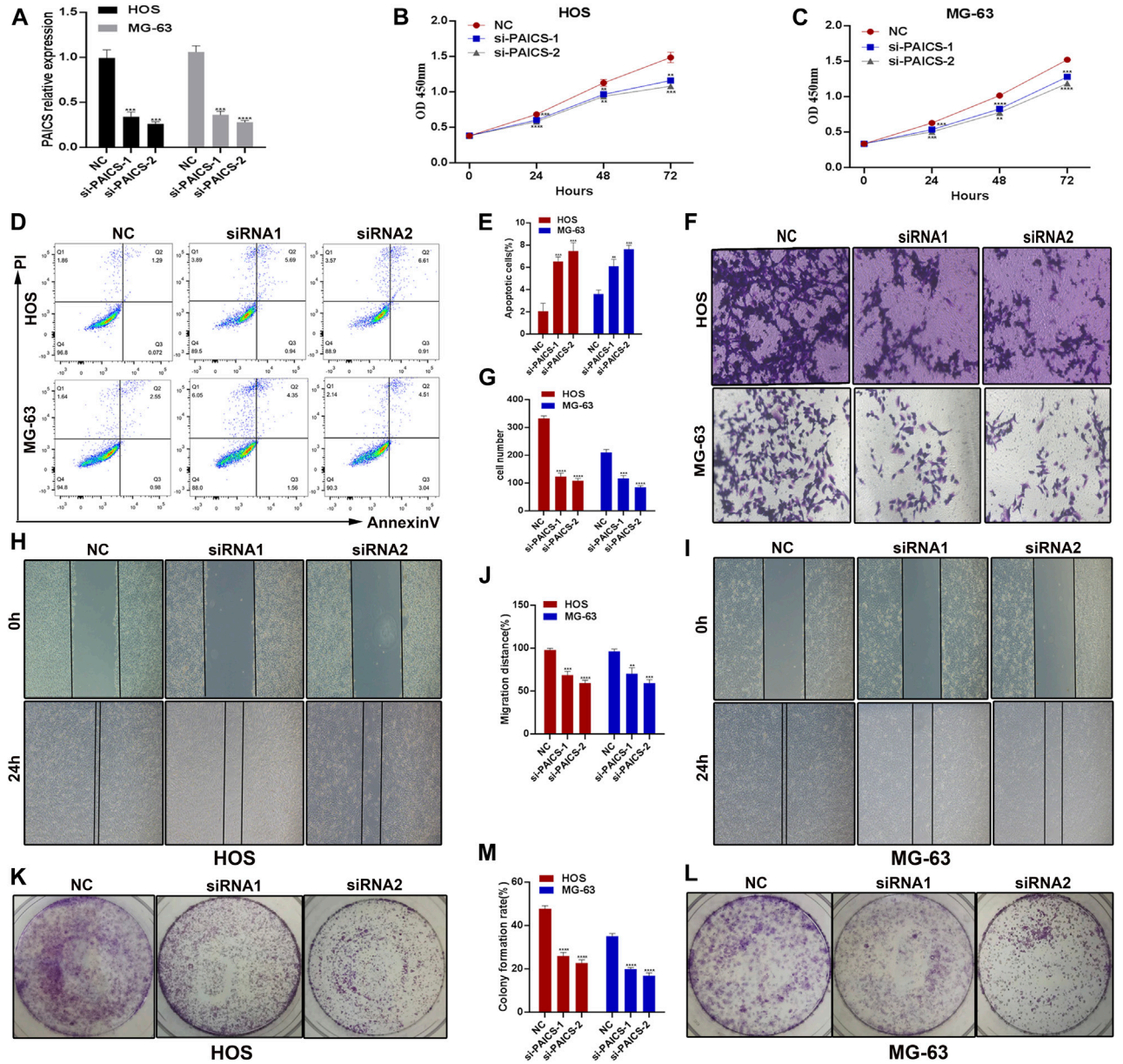


FIGURE 6 | *PAICS* act as an oncogene. *PAICS* was silenced using siRNAs in two osteosarcoma cell lines HOS and MG-63 cells. **(A)** Knockdown efficiency of two siRNAs used to knockdown *PAICS*. **(B, C)** CCK8 assay showing reduced cell proliferation in knockdown cells. **(D, E)** Flow cytometry showing more apoptotic cells in KD cells, **(F, G)** Transwell assay showing reduced migration after siRNA-mediated silencing of *PAICS* in osteosarcoma cells. **(H–J)** Wound healing and **(K–M)** colony formation capability of the osteosarcoma cells were also reduced.

patients in the training, testing, and validation cohorts. We found that the high expression group showed poor OS than the low expression group (**Figures 7A–C**). To further validate our prediction model, we applied ROC curve analysis on training, testing, and validation cohorts, which revealed AUC at 1 year was 0.86 in the validation cohort, which showed strong performance (**Figures 7D–F**). Furthermore, partial likelihood deviance was observed in our prognostic model by Lasso Cox regression analysis (**Supplementary Figure S5A**). We found that eight hub genes could be potential prognostic factors (**Supplementary Figure S5B**). These findings may further contribute to the prognostic improvement of osteosarcoma patients.

DISCUSSION

Recently, the advancements and applications in bioinformatics tools made it easy to mine the potential gene signatures associated with the OS and prognosis of

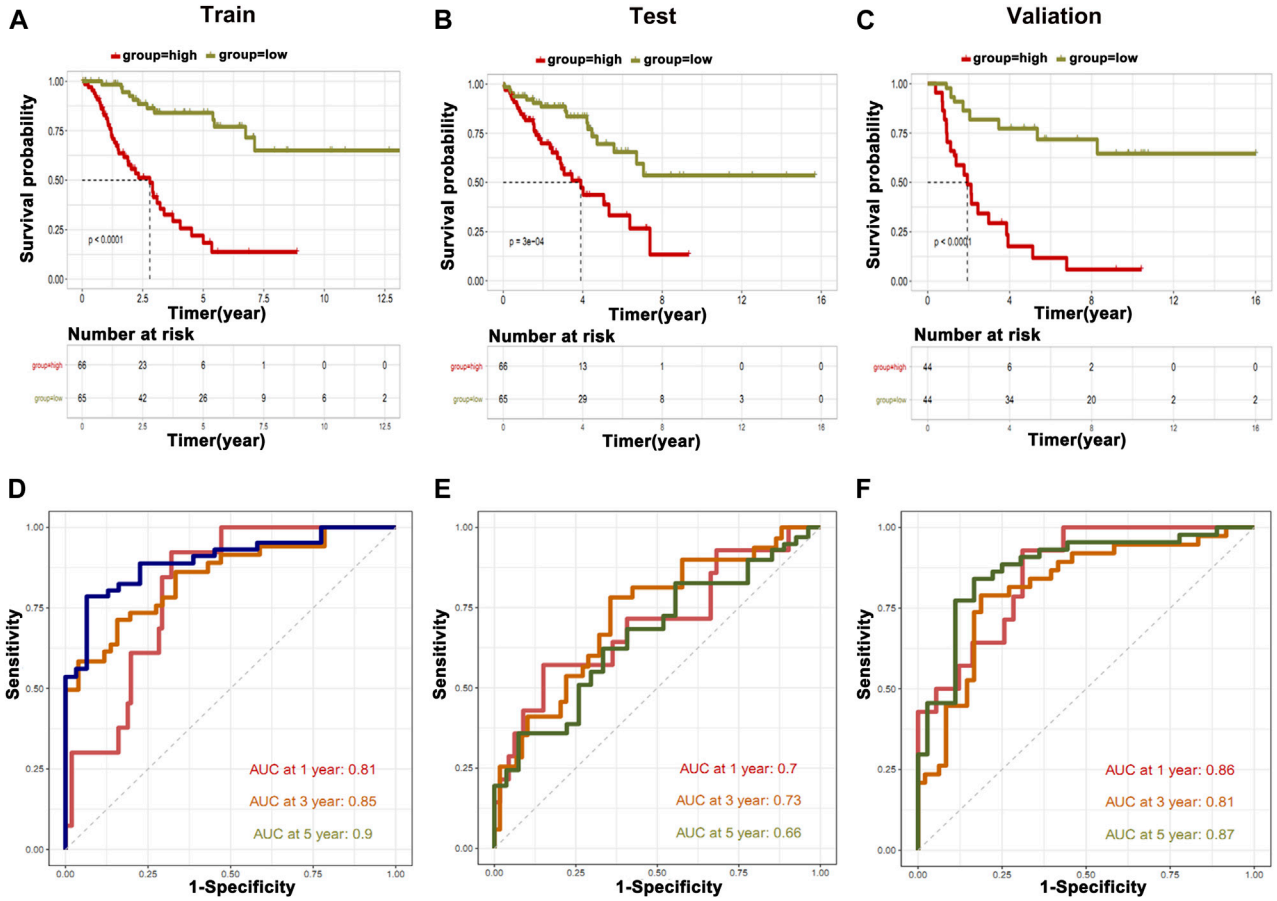


FIGURE 7 | Construction of prediction model. The information of metabolic genes and their association with immune cells were used to construct and validate the prediction model. **(A)** Training cohort, **(B)** testing cohort, and **(C)** validation cohort. ROC curve analysis was applied on **(D)** training cohort, **(E)** testing cohort, and **(F)** validation cohort.

the cancers including osteosarcoma from publically available databases such as TCGA, TARGET, and GEO (Cao et al., 2020; Chen et al., 2020; Song et al., 2020; Wen et al., 2020; Xiao et al., 2020; Yang et al., 2021). Various signature genes have been found associated with the OS of osteosarcoma patients and tumor development. A recent study (Fan et al., 2021) identified the nine novel signature genes potentially predicting the OS and prognosis of osteosarcoma. Another study identified the three gene signature for successful prognostic prediction of osteosarcoma patients through GSEA (Yang et al., 2021). However, using a similar pattern, we identified novel potential metabolic gene signatures through comprehensive RNA-seq analyses in osteosarcoma tissues. In total, 876 genes related to metabolic function, OS, and prognosis of the osteosarcoma patients were identified. However, among them, the top 10 metabolic genes related to OS were prioritized of osteosarcoma patients. Of these top 10 genes, 3 metabolic genes (*ALDH1A1*, *HNMT*, and *NUDT7*), and 7 (*GL O 1*, *SLC19A1*, *SQLE*, *PAICS*, *PPAT*, *FASN*, *AK2*) were negatively and positively correlated with survival, respectively. We independently used these 10 gene signatures for training and validation cohort to construct Kaplan–Meier estimator, Cox proportional hazards model, and ROC curve and found significant association with OS and prognosis of the osteosarcoma.

It is a well-known fact that immune cell infiltration such as $CD4^+$ naive T cells, $CD4^+$ memory T cells, $CD8^+$ T cells, $CD8^+$ Tcm, B cells, and memory B cells in the TME are the key regulators of the tumor development and OS of the GC patients (Ren et al., 2021). Mast cells have also shown a clear association with the signature genes and prognosis of the cancer patients (Fan et al., 2021). Mast cells can expand the tumors by inducing angiogenesis and tissue remodeling, specifically changing the composition of the extracellular matrix, and also by promoting inflammatory pathways (Maciel et al., 2015). Mast cells also enhance the proliferation and migration of the dendritic cells, tumor-associated macrophages, and lymphocytes (Inagaki et al., 2016) and tissue homeostasis of TME, which facilitate the growth and progression of the tumors (Oldford and Marshall, 2015). It has also been confirmed that Mast cells may trigger some mechanisms that can affect the homeostasis of the osteosarcoma

overall, affecting the occurrence and development of osteosarcoma (Campillo-Navarro et al., 2014; Maciel et al., 2015; Inagaki et al., 2016). In the current study, we analyzed the correlation between the infiltration of the immune cells, specifically activated $CD8^+$ T cells and 876 differentially expressed metabolic genes, and found 206 differentially expressed metabolic genes were correlated to immune cells; specifically, the top three genes *PLAG2D*, *PIK3R5*, and *INPP5D* were positively associated, and *CTPS2*, *CACNB3*, and *ADCY6* were the top three negatively correlated metabolic genes with immune cells. Based on the immune scores, patients were classified into three clusters, C1, C2, and C3; patients in cluster 1 showed the lowest immune score as compared to other clusters; however, cluster 3 was found to have the highest immune score. In the next step, we determined the immune infiltration of the osteosarcoma clusters using the ssGSEA package and obtained 28 immune cells in osteosarcoma samples; as expected, cluster 3 was detected with a higher number of immune cells, such as monocytic lineage, T cells and $CD8^+$ T. Previously, it has been suggested that most tumors at an advanced stage may have a higher frequency of mutations in genes related to the tumor immunity as compared to the early-stage tumors, which can activate more T cells and produce a stronger immune response (Ren et al., 2021). Usually, TME's stromal cells express a large number of surface and secretory molecules, which directly inhibit $CD4^+$ and $CD8^+$ T cells, and activate immunosuppressed myeloid cells (Salmon et al., 2012; Turley et al., 2015). $CD4^+$ T cells are the T helper cells, but also assist many other types of cells and act as a catalyst, increasing immune protection through many different pathways (Jaigirdar and Macleod, 2015).

Furthermore, we analyzed the expression of genes in all clusters showing that cluster 3 specifically has a high expression of the genes that are found downregulated in other clusters; thus, these genes were associated with immune function. Among all genes in three clusters, *PAICS* was commonly expressed in all clusters; therefore, we chose it for further validations. *PAICS* is an essential enzyme that has a significant role in *de novo* purine biosynthesis and associated with the formation of various tumors (Huang et al., 2020). *PAICS* has been localized in lung cancer tissues and shows high expression in tumor tissues as compared to normal tissues. High expression of *PAICS* has also been associated with the poor prognosis of lung cancer and gastric cancer patients (Goswami et al., 2015; Huang et al., 2020). Mechanistically, *PAICS* regulates pyruvate kinase activity, cell proliferation, and invasion (Goswami et al., 2015). Furthermore, its high expression has also been associated with the aggressiveness of the prostate cancer; therefore, *PAICS* has been considered essential for the proliferation and invasion of prostate cancer cells (Chakravarthi et al., 2018). Silencing its expression in breast cancer cells significantly reduced cell viability and proliferation (Meng et al., 2018). Increased expression of *PAICS* is involved in the proliferation, migration, invasion, and growth of colorectal cancer (CRC) cells, while depleting *PAICS* in the mice reduced tumor growth and metastasis to the liver, lung, and bones (Agarwal et al., 2020).

We performed survival analysis for the patients with low and high expression of *PAICS*. Interestingly, we observed that patients with high expression of *PAICS* had a poor survival, and *vice versa*. Furthermore, the expression of *PAICS* was negatively correlated with the immune cells, specifically with activated B cells, activated $CD8^+$ T cells, and activated dendritic cells, thus proving that patients with high expression of *PAICS* have poor survival. *PAICS* was highly expressed in 70% of the CRC tissues and associated with poor 5-year survival of the patients regardless of the pathological stages, patients' race, gender, and age (Agarwal et al., 2020). Based on current results, knockdown of *PAICS* in osteosarcoma cells (HOS and MG-63) significantly reduced the cell proliferation, migration, and wound healing ability and induced apoptosis, which revealed the oncogenic role of *PAICS* in the osteosarcoma. Our current research has some limitations; for example, the gene signature still needs to be verified for its clinical use. Accordingly, further studies are required to uncover the relationships between the gene signature and osteosarcoma progression.

CONCLUSION

In summary, the genetic or physiological alterations affect the regulation of metabolic gene signatures in TME, leading to cancer development; however, it has not fully been explained in osteosarcoma. We observed the significant difference of metabolic genes' expression between osteosarcoma and normal samples. Our analyses identified the correlation of a list of metabolic genes with OS and immune cell infiltration in osteosarcoma patients and TME. Through these analyses, we identified *PAICS* as a potential candidate gene for further analyses as it showed association with poor survival and immune cells. Furthermore, we established that loss of *PAICS* induces apoptosis and inhibits the proliferation and migration of HOS and MG-63 cell lines. Thus, we conclude that *PAICS* acts as an oncogene in osteosarcoma and could be used as a potential diagnostic and prognostic marker.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZZ and LW designed this study, ZZ and MZ analyzed data and wrote this paper. ZZ, MZ and WW conducted the experiments. PZ and YW downloaded and collected the data. All authors have read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.814843/full#supplementary-material

REFERENCES

Agarwal, S., Chakravarthi, B. V. S. K., Behring, M., Kim, H.-G., Chandrashekar, D. S., Gupta, N., et al. (2020). PAICS, a Purine Nucleotide Metabolic Enzyme, Is Involved in Tumor Growth and the Metastasis of Colorectal Cancer. *Cancers* 12, 772. doi:10.3390/cancers12040772

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462, 108–112. doi:10.1038/nature08460

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Campillo-Navarro, M., Chavez-Blanco, A., Wong-Baeza, I., Serafín-Lopez, J., Flores-Mejia, R., Estrada-Parra, S., et al. (2014). Mast Cells in Lung Homeostasis: Beyond Type I Hypersensitivity. *Crmr* 10, 115–123. doi:10.2174/1573398X10666141024220151

Cao, M., Zhang, J., Xu, H., Lin, Z., Chang, H., Wang, Y., et al. (2020). Identification and Development of a Novel 4-Gene Immune-Related Signature to Predict Osteosarcoma Prognosis. *Front. Mol. Biosci.* 7, 608368. doi:10.3389/fmolb.2020.608368

Chakravarthi, B. V. S. K., Goswami, M. T., Pathi, S. S., Dodson, M., Chandrashekar, D. S., Agarwal, S., et al. (2018). Expression and Role of PAICS, a De Novo Purine Biosynthetic Gene in Prostate Cancer. *Prostate* 78, 693–694. doi:10.1002/pros.23533

Chen, Y., Zhao, B., and Wang, X. (2020). Tumor Infiltrating Immune Cells (TIICs) as a Biomarker for Prognosis Benefits in Patients with Osteosarcoma. *BMC Cancer* 20, 1022. doi:10.1186/s12885-020-07536-3

Corre, I., Verrecchia, F., Crenn, V., Redini, F., and Trichet, V. (2020). The Osteosarcoma Microenvironment: A Complex but Targetable Ecosystem. *Cells* 9, 976. doi:10.3390/cells9040976

Cortini, M., Massa, A., Avnet, S., Bonuccelli, G., and Baldini, N. (2016). Tumor-Activated Mesenchymal Stromal Cells Promote Osteosarcoma Stemness and Migratory Potential via IL-6 Secretion. *PLoS One* 11, e0166500. doi:10.1371/journal.pone.0166500

Du, M., Liang, Y., Liu, Z., Li, X., Liang, M., Zhou, B., et al. (2021). Identification of Key Genes Related to CD8+ T-Cell Infiltration as Prognostic Biomarkers for Lung Adenocarcinoma. *Front. Oncol.* 11, 693353. doi:10.3389/fonc.2021.693353

Durnali, A., Alkis, N., Cangur, S., Yukruk, F. A., Inal, A., Tokluoglu, S., et al. (2013). Prognostic Factors for Teenage and Adult Patients with High-Grade Osteosarcoma: an Analysis of 240 Patients. *Med. Oncol.* 30, 624. doi:10.1007/s12032-013-0624-6

Fan, L., Ru, J., Liu, T., and Ma, C. (2021). Identification of a Novel Prognostic Gene Signature from the Immune Cell Infiltration Landscape of Osteosarcoma. *Front. Cel Dev. Biol.* 9, 718624. doi:10.3389/fcell.2021.718624

Federico, L., McGrail, D. J., Bentebibel, S.-E., Haymaker, C., Ravelli, A., Forget, M.-A., et al. (2021). Distinct Tumor-Infiltrating Lymphocyte Landscapes Are Associated with Clinical Outcomes in Localized Non-small-cell Lung Cancer. *Ann. Oncol.* S0923-7534, 04519. doi:10.1016/j.annonc.2021.09.021

Gerard, C. L., Delyon, J., Wicky, A., Homicsko, K., Cuendet, M. A., and Michielin, O. (2021). Turning Tumors from Cold to Inflamed to Improve Immunotherapy Response. *Cancer Treat. Rev.* 101, 102227. doi:10.1016/j.ctrv.2021.102227

Goswami, M. T., Chen, G., Chakravarthi, B. V. S. K., Pathi, S. S., Anand, S. K., Carskadon, S. L., et al. (2015). Role and Regulation of Coordinately Expressed De Novo Purine Biosynthetic Enzymes PPAT and PAICS in Lung Cancer. *Oncotarget* 6, 23445–23461. doi:10.18632/oncotarget.4352

Haider, S., McIntyre, A., van Stiphout, R. G. P. M., Winchester, L. M., Wigfield, S., Harris, A. L., et al. (2016). Genomic Alterations Underlie a Pan-Cancer Metabolic Shift Associated with Tumour Hypoxia. *Genome Biol.* 17, 140. doi:10.1186/s13059-016-0999-8

Heymann, M.-F., Lézot, F., and Heymann, D. (2019). The Contribution of Immune Infiltrates and the Local Microenvironment in the Pathogenesis of Osteosarcoma. *Cell Immunol.* 343, 103711. doi:10.1016/j.cellimm.2017.10.011

Huang, N., Xu, C., Deng, L., Li, X., Bian, Z., Zhang, Y., et al. (2020). PAICS Contributes to Gastric Carcinogenesis and Participates in DNA Damage Response by Interacting with Histone Deacetylase 1/2. *Cell Death Dis* 11, 507. doi:10.1038/s41419-020-2708-5

Ichikawa, I., and Brenner, B. M. (1981). The Effector Limb of the Tubuloglomerular Feedback Mechanism. *Trans. Assoc. Am. Physicians* 94, 73–79.

Inagaki, Y., Hookway, E., Williams, K. A., Hassan, A. B., Oppermann, U., Tanaka, Y., et al. (2016). Dendritic and Mast Cell Involvement in the Inflammatory Response to Primary Malignant Bone Tumours. *Clin. Sarcoma Res.* 6, 13. doi:10.1186/s13569-016-0053-3

Jaigirdar, S. A., and MacLeod, M. K. L. (2015). Development and Function of Protective and Pathologic Memory CD4 T Cells. *Front. Immunol.* 6, 456. doi:10.3389/fimmu.2015.00456

Kemmerer, C. L., Schittenhelm, J., Dubois, E., Neumann, L., Häsler, L. M., Lambert, M., et al. (2021). Cerebrospinal Fluid Cytokine Levels Are Associated with Macrophage Infiltration into Tumor Tissues of Glioma Patients. *BMC Cancer* 21, 1108. doi:10.1186/s12885-021-08825-1

Lettieri, C. K., Appel, N., Labban, N., Lussier, D. M., Blattman, J. N., and Hingorani, P. (2016). Progress and Opportunities for Immune Therapeutics in Osteosarcoma. *Immunotherapy* 8, 1233–1244. doi:10.2217/imt-2016-0048

Li, L. Q., Zhang, L. H., Yuan, Y. B., Lu, X. C., Zhang, Y., Liu, Y. K., et al. (2021). Signature Based on Metabolic-related Gene Pairs Can Predict Overall Survival of Osteosarcoma Patients. *Cancer Med.* 10, 4493–4509. doi:10.1002/cam4.3984

Luo, Z.-W., Liu, P.-P., Wang, Z.-X., Chen, C.-Y., and Xie, H. (2020). Macrophages in Osteosarcoma Immune Microenvironment: Implications for Immunotherapy. *Front. Oncol.* 10, 586580. doi:10.3389/fonc.2020.586580

Ma, C., Luo, H., Cao, J., Zheng, X., Zhang, J., Zhang, Y., et al. (2020). Identification of a Novel Tumor Microenvironment-Associated Eight-Gene Signature for Prognosis Prediction in Lung Adenocarcinoma. *Front. Mol. Biosci.* 7, 571641. doi:10.3389/fmolb.2020.571641

Maciel, T. T., Moura, I. C., and Hermine, O. (2015). The Role of Mast Cells in Cancers. *F1000prime Rep.* 7, 09. doi:10.12703/P7-09

Meng, M., Chen, Y., Jia, J., Li, L., and Yang, S. (2018). Knockdown of PAICS Inhibits Malignant Proliferation of Human Breast Cancer Cell Lines. *Biol. Res.* 51, 24. doi:10.1186/s40659-018-0172-9

Oldford, S. A., and Marshall, J. S. (2015). Mast Cells as Targets for Immunotherapy of Solid Tumors. *Mol. Immunol.* 63, 113–124. doi:10.1016/j.molimm.2014.02.020

Peng, X., Chen, Z., Farshidfar, F., Xu, X., Lorenzi, P. L., Wang, Y., et al. (2018). Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. *Cell Rep* 23, 255. doi:10.1016/j.celrep.2018.03.077

Ren, F., Zhao, Q., Zhao, M., Zhu, S., Liu, B., Bukhari, I., et al. (2021). Immune Infiltration Profiling in Gastric Cancer and Their Clinical Implications. *Cancer Sci.* 112, 3569–3584. doi:10.1111/cas.15057

Salmon, H., Franciszkiewicz, K., Damotte, D., Dieu-Nosjean, M.-C., Validire, P., Trautmann, A., et al. (2012). Matrix Architecture Defines the Preferential Localization and Migration of T Cells into the Stroma of Human Lung Tumors. *J. Clin. Invest.* 122, 899–910. doi:10.1172/JCI45817

Song, Y.-j., Xu, Y., Zhu, X., Fu, J., Deng, C., Chen, H., et al. (2020). Immune Landscape of the Tumor Microenvironment Identifies Prognostic Gene Signature CD4/CD68/CSF1R in Osteosarcoma. *Front. Oncol.* 10, 1198. doi:10.3389/fonc.2020.01198

Turley, S. J., Cremasco, V., and Astarita, J. L. (2015). Immunological Hallmarks of Stromal Cells in the Tumour Microenvironment. *Nat. Rev. Immunol.* 15, 669–682. doi:10.1038/nri3902

Wen, C., Wang, H., Wang, H., Mo, H., Zhong, W., Tang, J., et al. (2020). A Three-Gene Signature Based on Tumour Microenvironment Predicts Overall Survival of Osteosarcoma in Adolescents and Young Adults. *Aging* 13, 619–645. doi:10.18632/aging.202170

Xiao, B., Liu, L., Li, A., Xiang, C., Wang, P., Li, H., et al. (2020). Identification and Verification of Immune-Related Gene Prognostic Signature Based on ssGSEA for Osteosarcoma. *Front. Oncol.* 10, 607622. doi:10.3389/fonc.2020.607622

Yang, M., Ma, X., Wang, Z., Zhang, T., Hua, Y., and Cai, Z. (2021). Identification of a Novel Glycolysis-Related Gene Signature for Predicting the Prognosis of Osteosarcoma Patients. *Aging* 13, 12896–12918. doi:10.18632/aging.202958

Yiqi, Z., Ziyun, L., Qin, F., Xingli, W., and Liyu, Y. (2020). Identification of 9-Gene Epithelial-Mesenchymal Transition Related Signature of Osteosarcoma by Integrating Multi Cohorts. *Technol. Cancer Res. Treat.* 19, 153303382098076. doi:10.1177/1533033820980769

Zhang, C., Guo, L., Su, Z., Luo, N., Tan, Y., Xu, P., et al. (2021). Tumor Immune Microenvironment Landscape in Glioma Identifies a Prognostic and Immunotherapeutic Signature. *Front. Cel Dev. Biol.* 9, 717601. doi:10.3389/fcell.2021.717601

Zhang, H., Guo, L., Zhang, Z., Sun, Y., Kang, H., Song, C., et al. (2019). Co-Expression Network Analysis Identified Gene Signatures in Osteosarcoma as a Predictive Tool for Lung Metastasis and Survival. *J. Cancer* 10, 3706–3716. doi:10.7150/jca.32092

Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. *Mol. Ther. - Nucleic Acids* 22, 298–309. doi:10.1016/j.omtn.2020.08.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu, Zhang, Wang, Zhang, Wang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

CASE REPORT
published: 18 January 2022
doi: 10.3389/fgene.2021.799807

Case Report and Literature Review: Diagnosis, Tailored Genetic Counseling and Cancer Prevention for a Locally Advanced dMMR/MSI-H/TMB-H Lung Cancer Patient With Concurrent Lynch Syndrome Mediated by a Rare *PMS2* Splicing Variant (c.1144+1G>A)

Quanli Han[1†], Si Liu[2†], Zhi Cui[1], Qi Wang[1], Tonghui Ma[2], Liwen Jiang[2], Xiaomo Li[2] and Guanghai Dai[1]**

[1]Department of Medical Oncology, the First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, Beijing, China, [2]Genetron Health (Beijing) Technology, Co. Ltd., Beijing, China

Lynch syndrome (LS) is a cancer-predisposing genetic disease mediated by pathogenic mutations in DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Accumulating evidence demonstrates that there is significant biological heterogeneity across MMR genes. Compared to *MLH1* and *MSH2*, *PMS2* variant carriers have a much lower risk for LS-related cancers. Tumors in *MLH1* and *MSH2* variant carriers often display MMR deficiency (dMMR) and/or high microsatellite instability (MSI-H), two predictive biomarkers for immunotherapy efficacy. However, tumors in *PMS2* variant carriers are largely microsatellite stable (MSS) instead of MSI. Therefore, the optimal management of cancer patients with LS requires the integration of disease stage, MMR gene penetrance, dMMR/MSI status, and tumor mutational burden (TMB). In this work, we presented a locally advanced lung cancer patient with dMMR/MSI-H/TMB-H tumor and selective loss of PMS2 by immunohistochemistry. Germline testing revealed a rare *PMS2* splicing variant (c.1144+1G>A) in the proband and his healthy daughter. The diagnosis of LS was made based on genetic analysis of this variant and literature review. Given the incomplete penetrance of *PMS2*, the proband and the carrier received tailored genetic counseling. To reduce cancer risk, the proband received four cycles of nivolumab plus chemotherapy and achieved a disease-free survival of sixteen months.

Keywords: Lynch syndrome, lung cancer, PMS2, splicing variant, incomplete penetrance

OPEN ACCESS

Edited by:
Ruowen Zhang, Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Luke Hesson, Garvan Institute of Medical Research, Australia
Inge Bernstein, Aalborg University Hospital, Denmark

****Correspondence:***
Guanghai Dai
daigh301@vip.sina.com
Xiaomo Li
xiaomo.li@genetronhealth.com

†These authors have contributed equally to this work and share first authorship.

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *22 October 2021*
Accepted: *20 December 2021*
Published: *18 January 2022*

Citation:
Han Q, Liu S, Cui Z, Wang Q, Ma T, Jiang L, Li X and Dai G (2022) Case Report and Literature Review: Diagnosis, Tailored Genetic Counseling and Cancer Prevention for a Locally Advanced dMMR/MSI-H/TMB-H Lung Cancer Patient With Concurrent Lynch Syndrome Mediated by a Rare PMS2 *Splicing Variant (c.1144+1G>A). Front. Genet. 12:799807. doi: 10.3389/fgene.2021.799807*

INTRODUCTION

In the past decade, cancer immunotherapy has shifted the landscape of cancer treatment (Tang et al., 2018). Predictive biomarkers such as PD-L1 expression level have greatly facilitated the selection of patients for immunotherapy in some cancer types (Doroshow et al., 2021). In 2017, FDA approved PD-1 antibody pembrolizumab to treat patients with unresectable or metastatic mismatch repair deficiency (dMMR) and/or microsatellite instability-high (MSI-H) solid tumors, making it the first tissue/site-agnostic predictive biomarker (Marcus et al., 2019).

MMR deficiency is mediated by somatic or germline mutations in MMR genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) and rarely, *EPCAM* (Le et al., 2015). Pathogenic germline variants in MMR genes cause Lynch Syndrome (LS), a genetic disease predisposing patients to multiple types of cancers (Lynch et al., 2015). Therefore, the diagnosis of LS is essential for the treatment and cancer-risk reduction for LS patients and family members harboring pathogenic variants (Yurgelun and Hampel, 2018).

The optimal testing and treatment of LS patients require a good knowledge of the pan-cancer prevalence and heterogeneity of pathogenic MMR gene variants. To explore the prevalence of LS across solid tumors according to MSI status, researchers at the Memorial Sloan Kettering Cancer Center (MSKCC) screened 15,045 patients of more than 50 cancer types (Latham et al., 2019). Among LS patients with MSI tumors, half had tumors other than colorectal and endometrial cancer, including gastric, pancreas, small bowel, and germ cell tumors. However, none of the 1,952 lung cancer patients had LS. While most tumors from *MLH1* and *MSH2* carriers are MSI, more than two-thirds of tumors from *PMS2* carriers are microsatellite stable (MSS) (Latham et al., 2019).

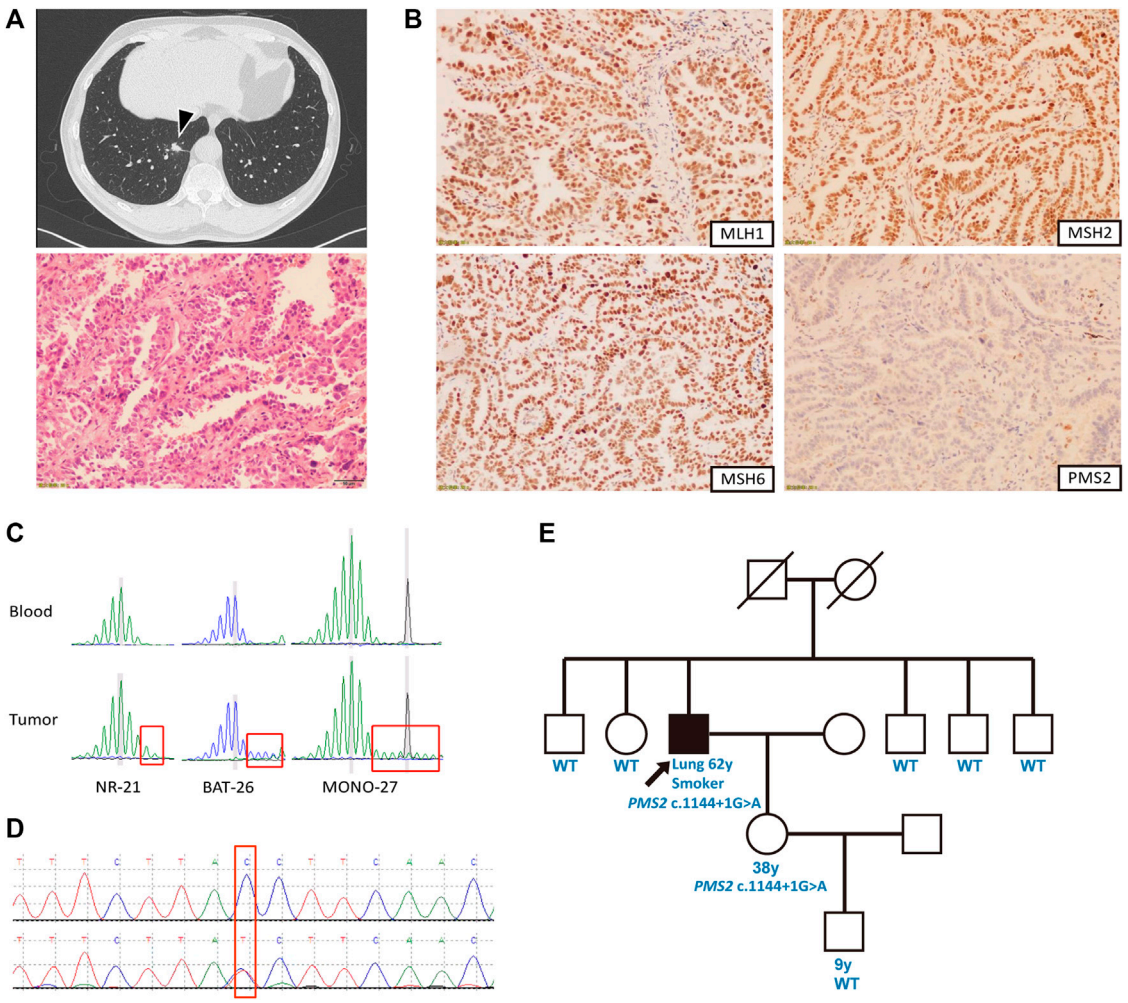


FIGURE 1 | Case summary. **(A)** Clinical findings of the patient. Above: Chest computed tomography revealing a mass in the right lower lobe (arrow); below: Hematoxylin and eosin-stained tumor tissue sections presenting adenocarcinoma (magnification, ×400). **(B)** Immunohistochemical staining showing absence of PMS2 as well as presence of MLH1, MSH2 and MSH6 in the tumor cell nuclei. **(C)** Capillary electrophoresis results showed loss of stability of three microsatellite biomarkers in tumor compared to the blood control, indicating high microsatellite instability. **(D)** Sequence chromatogram of the patient's daughter containing the same *PMS2* c.1144+1G>A mutation. **(E)** Pedigree of the patient's family. The proband is indicated with an arrow and black denotes the cancer-affected individual.

TABLE 1 | Somatic and germline testing results

Mutation type	Gene	Nucleotide change	Amino acid change	Mutation effect	VAF (%)
Germline	*PMS2*	c.1144+1G>A		Splicing	
Somatic	*KRAS*	c.38G>A	p.Gly13Asp	Nonsynonymous	20.9
	ARID1A	c.4555dup C	p.Gln1519ProfsTer13	Frameshift insertion	22.7
	TP53	c.799C>T	p.Arg267Trp	Nonsynonymous	21.1
	JAK3	c.2126G>T	p.Trp709Leu	Nonsynonymous	18.6
	TLR4	c.1325T>C	p.Val442Ala	Nonsynonymous	20.9
	ATM	c.8546G>T	p.Arg2849Leu	Nonsynonymous	20.9
	ATM	c.8549T>C	p.Leu2850Ser	Nonsynonymous	15.5
	EPHA5	c.2250G>A	p.Met750Ile	Nonsynonymous	19.2
	TP63	c.490G>T	p.Ala164Ser	Nonsynonymous	20.9
	CRBN	c.1020A>T	p.Leu340Phe	Nonsynonymous	23.2
	TUBG1	c.476A>C	p.Asp159Ala	Nonsynonymous	18.4
	DAPK1	c.754A>G	p.Ile252Val	Nonsynonymous	14
	NOTCH4	c.5186C>T	p.Ala1729Val	Nonsynonymous	19.2
	TET2	c.1842dup	p.Leu615AlafsTer23	Frameshift insertion	6.8
	ANKRD11	c.1041T>G	p.Tyr347Ter	Stopgain	5.7
MSI-H					
TMB high (13.62 mutations/MB)					

VAF: variant allele frequency; MSI: microsatellite instability; TMB: tumor mutational burden. Variants with VAF > 0.05 were shown.

Recent studies including the Prospective Lynch Syndrome Database (PLSD) and the International Mismatch Repair Consortium revealed that MMR gene variants had distinct penetrance (Dominguez-Valentin et al., 2020; Dominguez-Valentin et al., 2021; International Mismatch Repair Consortium, 2021). Pathogenic *MLH1* and *MSH2* variants caused high penetrance in a broad spectrum of LS cancers, while pathogenic *PMS2* variants were associated with low penetrance in few LS-related cancers. Therefore, the cancer screening, chemoprevention, and risk-reduction surgery protocol for carriers of *MLH1* and *MSH2* variants should be tailored to carriers of *PMS2* variants (Balmaña et al., 2013; Stoffel et al., 2015; Gupta et al., 2019; Seppälä et al., 2021a). Here, we presented the diagnosis, tailored genetic counseling, and cancer prevention of a locally advanced lung cancer patient with dMMR/MSI-H/TMB-H tumor and *PMS2*-LS.

CASE REPORT AND GENETIC ANALYSIS

Our proband was a 62-year-old Chinese man with 40-pack-year smoking history, presented with a consolidation shadow in the lower lobe of the right lung discovered by chest computed tomography (CT) (**Figure 1A**). The CT scan also detected local occlusion of internal small bronchus, subpleural nodular ground-glass opacity on the posterior segment of the right superior lobe, as well as a nodular contour in the right adrenal gland. He underwent thoracoscopic right lower lobectomy and mediastinal lymphadenectomy. Postoperative pathology showed a stage IIIA lung adenocarcinoma (T1aN2M0) with regional lymph node metastases, which had a size of 0.9 × 0.6 × 0.6 cm. Genomic profiling with a multi-gene next-generation sequencing (NGS) panel (Onco Panscan™, Genetron Health) showed a *KRAS* G13D mutation, a *TP53* R267W mutation, and a high tumor mutational burden (TMB-H, 13.62 mutations/Mb) (**Table 1**). PD-L1 immunochemistry (IHC), MMR IHC, and microsatellite instability (MSI) testing revealed that his tumor was PD-L1-positive (TPS 8%), MSI-H, and dMMR (**Figures 1B,C**).

TABLE 2 | Prediction scores of reference sequence and c.1144+1G>A splicing variant in *PMS2* gene.

Tools	Exon intron junction sequence	
	Reference	c.1144+1G > A
	AAGgtaaga	AAGataaga
ASSP	10.959	-
MaxEntScan	10.57	2.39
NetGene2	0.99	-

Based on the dMMR/MSI-H/TMB-H phenotype of the patient, we suspected Lynch syndrome (LS) even though his personal and family history did not fulfill the revised Amsterdam criteria or Bethesda guideline (Cohen et al., 2019). Germline testing revealed a c.1144+1G>A mutation located at the splice donor site of intron 10 of the *PMS2* gene. This variant was not included in the InSiGHT database (http://www.insight-group.org/) (Thompson et al., 2014). In 2019, the two interpretations of this variant in the ClinVar database did not follow the 2015 ACMG–AMP guidelines (Richards et al., 2015) or the refined 2017 version (Sherloc) (Nykamp et al., 2017). To accurately determine its pathogenicity, we conducted a comprehensive genetic analysis. *In silico* analysis with three splice prediction programs including Alternative Splice Site Predictor (ASSP) (Wang and Marín, 2006), MaxEntScan (Yeo and Burge, 2004), and NetGene2 (Hebsgaard et al., 1996) suggested that the *PMS2* c.1144+1G>A variant will generate aberrant splicing transcripts (**Table 2**). Data in the gnomAD database revealed that its population allele frequency is 2/282496 without homozygotes, which falls into the pathogenic range defined by the Sherloc

guideline (Nykamp et al., 2017). The proband had an oncogenic *KRAS* G13D somatic mutation and his 38-years-old daughter with the same *PMS2* variant had no cancer yet (**Figure 1D**). Moreover, Sanger sequencing results showed that this variant was segregated with his daughter but none of his five siblings (**Figure 1E**). We conducted an extensive literature review and found two probands harboring the same *PMS2* variant, a breast cancer patient from a cohort of 480 patients during a genetic screen of germline hereditary breast cancer susceptibility genes (Wang et al., 2018) and a colorectal cancer patient in a cohort of 6,503 patients from the NHLBI Exome Sequencing Project (ESP) (Amendola et al., 2015). According to the ACMG/Sherloc guidelines, we classified the *PMS2* c.1144+1G>A variant as class 4 (likely pathogenic). Interestingly, in a family with germline *PMS2* c.1144+2T>A variant, which disrupts the same splice donor site as c.1144+1G>A, the cosegregation between this variant and LS-related cancers was seen (Hendriks et al., 2006). These reports and our findings indicated that disruption of this specific *PMS2* splicing site can be pathogenic/likely pathogenic and result in LS.

Although the ESMO (2013), ASCO (2015), NCCN (2017) LS management guidelines gave the same recommendations for different MMR genes (Balmaña et al., 2013; Stoffel et al., 2015; Gupta et al., 2017), accumulating evidence demonstrated that the penetrance of *PMS2* was much lower than other MMR genes (Møller et al., 2017; Ten Broeke et al., 2018a; Ten Broeke et al., 2018b). Therefore, we provided *PMS2*-specific genetic counseling to the proband and the carrier. 5-yearly colonoscopic surveillance was recommended to the patient and the carrier. For the female carrier, endometrial/ovarian cancer screen was recommended after she reaches 50 years old. Furthermore, given the dMMR/MSI-H/TMB-H phenotype of the tumor and the recurrence risk of locally advanced NSCLC with concurrent *KRAS* and *TP53* mutations (Yagishita et al., 2015; Hames et al., 2016; Skoulidis and Heymach, 2019), the patient was then treated with cisplatin (75 mg/m^2), pemetrexed (500 mg/m2), and nivolumab (200 mg) every 3 weeks for a total of 4 cycles after surgery for risk reduction. Sixteen months after the discontinuation of treatment, he developed progressive disease and was enrolled in a study with another investigational agent.

DISCUSSION

After the approval of pembrolizumab for patients with advanced dMMR/MSI-H solid tumor in 2017, the testing of dMMR/MSI-H has become a common diagnosis approach for cancer patients (Evans et al., 2021). If dMMR/MSI testing results are positive, germline testing will be recommended for the diagnosis of Lynch syndrome (LS), which is essential for the optimal care for cancer patients and their family members at risk (Yurgelun and Hampel, 2018). Currently, the standard of dMMR and MSI testing is the immunochemistry (IHC) assay of four MMR proteins and PCR-based assays of five microsatellite loci, respectively (André et al., 2020). One interesting observation in MMR IHC testing is that patients with pathogenic *MSH2* (*path_MSH2*) and *MLH1* (*path_MLH1*) mutations display simultaneous loss-of-expression for MSH2/MSH6 and MLH1/PMS2, respectively (Luchini et al., 2019). In contrast, patients with pathogenic *MSH6* (*path_MSH6*) and *PMS2* (*path_PMS2*) mutations retain expression of MSH2 and MLH1, respectively. This is because MSH6 forms a heterodimer complex with MSH2 and PMS2 forms a heterodimer complex with MLH1, which are required to maintain the protein stability of MSH6 and PMS2 but not MSH2 and MLH1, respectively. Therefore, the phenotypes of *path_MSH2* and *path_MLH1* single mutants mimic the phenotypes of *path_MSH2*/*path_MSH6* and *path_MLH1*/*path_PMS2* double mutants, respectively. According to these results, the cancer risk of *path_MSH2* and *path_MLH1* carriers should be significantly higher than *path_MSH6* and *path_PMS2* carriers.

This prediction was supported by multiple studies including the Prospective Lynch Syndrome Database (PLSD), an international, multicenter prospective observational study involving 6,350 *path_MMR* variants carriers and 1,808 observed cancers (Dominguez-Valentin et al., 2020). *Path_MSH2* and *path_MLH1* variants were associated with high penetrance dominant syndrome in colorectal, endometrial, and ovarian cancers while *path_MSH6* variants were associated with high risk in endometrial cancer but modestly increased risk for colorectal cancer. In contrast, the risk of *path_PMS2* variants for these three cancers was not increased before 50 years of age and only nonsignificantly increased after that.

The PLSD study series is a game changer for current cancer surveillance and risk-reduction practice for LS patients (Seppälä et al., 2021a; Dominguez-Valentin et al., 2021). According to the 2019 NCCN guideline for LS management, both the proband and carriers of *path_MMR* variants should take colonoscopy every 1–2 years for colorectal cancer surveillance and aspirin for risk reduction (Gupta et al., 2019). Additionally, the female carriers may consider an endometrial biopsy screen every 1–2 years for endometrial cancer surveillance and hysterectomy for risk-reduction. Given the incomplete penetrance of *PMS2* (Møller et al., 2017; Ten Broeke et al., 2018a; Ten Broeke et al., 2018b), we provided *PMS2*-specific genetic counseling to the proband and carrier, which was different from the general recommendation of the 2013 ESMO/2015 ASCO LS managenment guidelines (Balmaña et al., 2013; Stoffel et al., 2015). For instance, we did not recommend chemoprevention with aspirin or colonoscopy every one or two years for the proband and carrier. Our practice was largely consistent with the 2021 NCCN LS guideline which provided gene-specific cancer surveillance and risk reduction recommendations (NCCN Guideline, 2021). For instance, chemoprevention with 600 mg/daily aspirin for 2 years is recommended for all *path_MMR* carriers except for *path_PMS2* carriers. For *path_PMS2* and *path_MSH6* carriers, the colonoscopy screen age has been changed to 30–35 from 20–25 years old. Ovarian cancer screen or risk-reduction surgery are not recommended for *path_PMS2* carriers, as they do not have increased risk. In addition, the risk of endometrial cancer for *path_PMS2* carriers is only moderately increased compared to *path_MLH1, path_MSH2, and path_MSH6* carriers. Similarly, the 2021 LS guideline developed by the European Hereditary Tumour Group (EHTG) and the European Society of

Coloproctology (ESCP) also revised the colorectal cancer surveillance and risk-reduction procedures for *path_MMR* carriers (Seppälä et al., 2021b). The colonoscopy screen interval time for *path_PMS2* and other *path_MMR* carriers are 5 years and 2–3 years, respectively. Moreover, the colonoscopy screen starting ages for *path_PMS2* and *path_MSH6* carriers are 35, but 25 for *path_MLH1* and *path_MSH2* carriers. Additionally, extended surgery is only recommended for *path_MLH1* and *path_MSH2* carriers but not for *path_PMS2* and *path_MSH6* carriers at the first diagnosis of colorectal cancer.

In addition to cancer surveillance and prevention, dMMR/MSI is also an important biomarker for cancer immunotherapy (Luchini et al., 2019). Due to the functional heterogeneity of MMR gene variants, sometimes we can see discordance between *path_MMR* variants and MSI status. This is well illustrated in the MSKCC pan-cancer MSI study of 15,045 patients including 103 LS cases (Latham et al., 2019). The microsatellite stable (MSS) cases in *path_PMS2*, *path_MSH6*, *path_MSH2*, and *path_MLH1* carriers were 68.2% (15/22), 53.8% (14/26), 13.9% (5/36), and 16.7% (3/18), respectively. This result indicates that the diagnosis of LS alone does not justify the treatment decision of immunotherapy, especially for *path_PMS2* and *path_MSH6* carriers.

Besides the discordance between *path_MMR* variants and MSI status, the discordance between dMMR and TMB is another issue oncologists should consider during the treatment decision-making process for LS patients. Recently, Bielska et al. reported that low TMB level in LS patients with dMMR tumors was a mechanism of immunotherapy resistance (Bielska et al., 2021). Three LS patients developed two primary dMMR tumors, one TMB-H and the other TMB-L. While the dMMR tumors with high TMB responded to immunotherapy, those with low TMB did not.

Next, we briefly discuss the limitations of TMB-H as the pan-cancer biomarkers for immunotherapy. In 2020, FDA approved PD-1 antibody pembrolizumab for the treatment of TMB-H (TMB >10 mutations/megabase) solid tumors. This approval was based on the results of the KEYNOTE-158 trial (Marcus et al., 2021). Pembrolizumab achieved an overall response rate (ORR) of 29% in TMB-H patients (13%, n = 102). However, this approval should not be applied to colorectal cancer (CRC) as subgroup analysis of 137 advanced CRC patients treated with immunotherapy showed that there is no survival benefit in TMB-H patients after the removal of patients with dMMR or *POLD*/*POLE1* mutations (Rousseau et al., 2021). Further pan-cancer analysis of 1,661 cancer patients treated with immunotherapy revealed that TMB-H was associated with improved overall survival in a limited subgroup of pMMR cancers including NSCLC (Rousseau et al., 2021). There are three important lessons that we can learn from these studies: first, TMB is not a pan-cancer immunotherapy biomarker; second, the combination of dMMR/MSI-H and TMB-H can provide better immunotherapy efficacy prediction than either alone; third, dMMR tumors with low TMB may not respond to immunotherapy.

LS is very rare in primary lung cancer. For instance, the MSKCC pan-cancer study did not find LS in 1,952 lung cancer patients including 94 MSI cases (Latham et al., 2019). We reviewed the literature and identified a few primary lung cancer cases associated with LS driven by *path_MSH2* (n = 4), *path_MLH1* (n = 1), *path_MSH6* (n = 1), and *path_PMS2* (n = 1) variants (**Table 3**). The *path_PMS2* lung cancer case was a 74-year old female non-smoker with MSS, pMMR, and TMB-L tumor (Sun et al., 2019). Without actionable mutations, she was treated with gefitinib for two months and then switched to platinum-based chemotherapy. The *path_MSH6* lung cancer case was a 76-year old male smoker with PD-L1-positive, pMMR, and TMB-L tumor (Long et al., 2021). Without actionable mutations, pembrolizumab was administered with SD at 4 cycles and PD at 8 cycles, likely due to an acquired *STK11* mutation. The *path_MLH1* lung cancer case was a 36-year old male non-smoker with dMMR/MSI-H tumor (Masuzawa et al., 2020). He received nivolumab as the fifth-line therapy for 15 cycles with a partial response lasting more than 20 months. Among the four *path_MSH2* lung cancer cases (Canney et al., 2009; Nolan et al., 2009; Kawashima et al., 2019; Sun et al., 2019), only one patient had dMMR/MSI tumor (Kawashima et al., 2019). He received nivolumab as the third-line therapy with a partial response lasting more than 10 months (Kawashima et al., 2019). Our patient was positive for three immune biomarkers (dMMR/MSI-H/TMB-H). These results suggested that he could benefit from immunotherapy.

Despite complete surgical resection, stage III NSCLC patients have high rates of relapse (Evison, 2020). Given the success of PD(L)-1 checkpoint inhibitors in metastatic NSCLC treatment, multiple trials are testing their efficacy in earlier stages of disease. Recently, results of the phase 3 IMpower010 trial showed a disease-free survival benefit with atezolizumab versus best supportive care after adjuvant chemotherapy in patients with PD-L1-positive resected stage II–IIIA NSCLC (Felip et al., 2021). This led to the FDA approval of atezolizumab as the adjuvant therapy for this patient population. Similarly, in the phase 2 KEYNOTE-799 trial, pembrolizumab plus concurrent chemoradiation therapy demonstrated objective response rates of 71% in locally advanced, stage III NSCLC (Jabbour et al., 2021). Because our case was a locally advanced dMMR/MSI-H tumor, we were also interested in the efficacy of adjuvant immunotherapy in this setting. For dMMR/MSI-H solid tumors, results of the phase 3 KEYNOTE-177 trial established that first-line pembrolizumab therapy resulted in significantly longer PFS than chemotherapy for dMMR/MSI-H metastatic CRC (André et al., 2020). Currently, two ongoing randomized phase 3 trials are testing the efficacy of adjuvant immunochemotherapy in patients with resected dMMR/MSI stage III CRC. The ATOMIC trial (NCT02912559) and the POLEM trial (NCT03827044) are evaluating the combination of chemotherapy with atezolizumab or avelumab, respectively (Sinicrope et al., 2019; Lau et al., 2020). Results of these two trials will provide proof-of-concept for the use of immune checkpoint inhibitors in the curative setting of dMMR/MSI-H stage III CRC which could be extended to other LS-associated solid tumors.

TABLE 3 | Clinical information and genetic testing results for seven lung cancer patients associated with LS.

Author, year	Histology Type	Age onset, y/ Gender	Personal history (years)	Family history (years)	Germline mutation	TMB (muts/ Mb)	MSI status	MMR expression	Genes with somatic mutations	Immuno-therapy
S. Sun, et al., 2019	ADC	74/F	No	No	*PMS2* c.943C>T p.R315*	1	MSS	Intact	*PTCH1*	No
Y. Long, et al., 2021	SCC	76/M	No	Brother, gastric cancer; sister, colon cancer	*MSH6* c.2552_2553dupGC p.K852Afs*17	8	N/A	Intact	*ALK, LRP1B, TP53, DNMT3A, HRAS, DDR2, NTM, TCF7L2, POLE*	Yes
K.Masuzawa, et al., 2020	ADC	36/M	No	Colorectal cancer: sister (32 years), father (46 years, 54 years), paternal uncle (40 years), paternal grandmother (41 years)	*MLH1* c.2180_2181del p.H727Hfs*5	N/A	MSI-H	MLH1- MSH6-	N/A	Yes
A. Canney, et al., 2009	Two primary ADCs	59/M	Colorectal cancer (54 years, 58 years)	Father, leukemia died 70 years; brother, prostate cancer (62 years)	*MSH2* c.2109del p.I704Lfs*6	N/A	N/A	1st: MSH2- MSH6-; 2nd: intact	N/A	No
L. Nolan, et al., 2009	ADC	64/M	Colorectal cancer (33 years, 53 years); bladder cancer (51 years)	No	*MSH2* c.342+3C>T	N/A	MSI-H	MSH2-	N/A	No
Y. Kawashima, et al., 2019	ADC	68/M	Colorectal cancer (43 years, 63 years, 64 years); SCC (63 years); prostate cancer (67 years)	Mother, uterine cancer; brother, stomach cancer at a young age	*MSH2* rearrangement	N/A	MSI-H	MSH2- MSH6-	N/A	Yes
S. Sun, et al., 2019	ADC	62/F	No	Mother, colon cancer	*MSH2* c.340delG p.E114Rfs*60	1	N/A	Intact	*MAP2K2*, *GNAS*	No

M, male; F, female; ADC, adenocarcinoma; SCC, squamous-cell carcinoma; TMB, tumor mutational burden; MSI, microsatellite instability; MSS, microsatellite stable; N/A, not available.

In summary, we encountered a locally advanced lung cancer patient with untargetable driver mutations, dMMR/MSI-H/TMB-H tumor, and a germline *PMS2* splicing variant which led to the diagnosis of LS. We provided *PMS2*-specific genetic counseling to the proband and the carrier in his family. The proband received 4 cycles of nivolumab plus chemotherapy for cancer-risk reduction, which led to a disease-free survival time of 16 months. Further efforts are required to investigate the efficacy of adjuvant immunotherapy in LS patients with locally advanced dMMR/MSI-H/TMB-H tumors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Concept and design: GD. Acquisition, analysis, and interpretation of data: QH, SL, XL. Drafting of the manuscript: QH, SL, XL. Critical revision of the manuscript for important intellectual content: TM, XL, GD. Technical and material support: SL, ZC, QW, LJ. Study supervision: GD. All the authors gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

The authors would like to thank Yi Cai for his valuable and constructive suggestions during the preparation of this manuscript. The authors are grateful to the patient and family members for their kind cooperation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.799807/full#supplementary-material

REFERENCES

Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., et al. (2015). Actionable Exomic Incidental Findings in 6503 Participants: Challenges of Variant Classification. *Genome Res.* 25, 305–315. doi:10.1101/gr.183483.114

André, T., Shiu, K.-K., Kim, T. W., Jensen, B. V., Jensen, L. H., Punt, C., et al. (2020). Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N. Engl. J. Med.* 383, 2207–2218. doi:10.1056/NEJMoa2017699

Balmaña, J., Balaguer, F., Cervantes, A., and Arnold, D. (2013). Familial Risk-Colorectal Cancer: ESMO Clinical Practice Guidelines. *Ann. Oncol.* 24 (Suppl. 6), vi73–vi80. doi:10.1093/annonc/mdt209

Bielska, A. A., Chatila, W. K., Walch, H., Schultz, N., Stadler, Z. K., Shia, J., et al. (2021). Tumor Mutational Burden and Mismatch Repair Deficiency Discordance as a Mechanism of Immunotherapy Resistance. *J. Natl. Compr. Cancer Netw.* 19, 130–133. doi:10.6004/jnccn.2020.7680

Canney, A., Sheahan, K., Keegan, D., Tolan, M., Hyland, J., and Green, A. (2009). Synchronous Lung Tumours in a Patient with Metachronous Colorectal Carcinoma and a Germline MSH2 Mutation. *J. Clin. Pathol.* 62, 471–473. doi:10.1136/jcp.2008.063008

Cohen, S. A., Pritchard, C. C., and Jarvik, G. P. (2019). Lynch Syndrome: From Screening to Diagnosis to Treatment in the Era of Modern Molecular Oncology. *Annu. Rev. Genom. Hum. Genet.* 20, 293–307. doi:10.1146/annurev-genom-083118-015406

Dominguez-Valentin, M., Sampson, J. R., Seppälä, T. T., Ten Broeke, S. W., Plazzer, J. P., Nakken, S., et al. (2020). Cancer Risks by Gene, Age, and Gender in 6350 Carriers of Pathogenic Mismatch Repair Variants: Findings from the Prospective Lynch Syndrome Database. *Genet. Med.* 22, 15–25. doi:10.1038/s41436-019-0596-9

Dominguez-Valentin, M., Crosbie, E. J., Engel, C., Aretz, S., Macrae, F., Winship, I., et al. (2021). Risk-Reducing Hysterectomy and Bilateral Salpingo-Oophorectomy in Female Heterozygotes of Pathogenic Mismatch Repair Variants: A Prospective Lynch Syndrome Database Report. *Genet. Med.* 23, 705–712. doi:10.1038/s41436-020-01029-1

Doroshow, D. B., Bhalla, S., Beasley, M. B., Sholl, L. M., Kerr, K. M., Gnjatic, S., et al. (2021). PD-L1 as a Biomarker of Response to Immune-Checkpoint Inhibitors. *Nat. Rev. Clin. Oncol.* 18, 345–362. doi:10.1038/s41571-021-00473-5

Evans, D. G., Lalloo, F., Ryan, N. A., Bowers, N., Green, K., Woodward, E. R., et al. (2021). Advances in Genetic Technologies Result in Improved Diagnosis of Mismatch Repair Deficiency in Colorectal and Endometrial Cancers. *J. Med. Genet.* doi:10.1136/jmedgenet-2020-107542

Evison, M. (2020). The Current Treatment Landscape in the UK for Stage III NSCLC. *Br. J. Cancer* 123, 3–9. doi:10.1038/s41416-020-01069-z

Felip, E., Altorki, N., Zhou, C., Csőszi, T., Vynnychenko, I., Goloborodko, O., et al. (2021). Adjuvant Atezolizumab after Adjuvant Chemotherapy in Resected Stage IB-IIIA Non-small-cell Lung Cancer (IMpower010): A Randomised, Multicentre, Open-Label, Phase 3 Trial. *The Lancet* 398, 1344–1357. doi:10.1016/s0140-6736(21)02098-5

Gupta, S., Provenzale, D., Llor, X., Halverson, A. L., Grady, W., Chung, D. C., et al. (2019). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 2.2019. *J. Natl. Compr. Cancer Netw.* 17, 1032–1041. doi:10.6004/jnccn.2019.0044

Gupta, S., Provenzale, D., Regenbogen, S. E., Hampel, H., Slavin, T. P., Hall, M. J., et al. (2017). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. *J. Natl. Compr. Canc Netw.* 15, 1465–1475. doi:10.6004/jnccn.2017.0176

Hames, M. L., Chen, H., Iams, W., Aston, J., Lovly, C. M., and Horn, L. (2016). Correlation between KRAS Mutation Status and Response to Chemotherapy in Patients with Advanced Non-small Cell Lung Cancer☆. *Lung Cancer* 92, 29–34. doi:10.1016/j.lungcan.2015.11.004

Hebsgaard, S., Korning, P., Tolstrup, N., Engelbrecht, J., Rouzé, P., and Brunak, S. (1996). Splice Site Prediction in *Arabidopsis Thaliana* Pre-mRNA by Combining Local and Global Sequence Information. *Nucleic Acids Res.* 24, 3439–3452. doi:10.1093/nar/24.17.3439

Hendriks, Y. M. C., Jagmohan–Changur, S., van der Klift, H. M., Morreau, H., van Puijenbroek, M., Tops, C., et al. (2006). Heterozygous Mutations in PMS2 Cause Hereditary Nonpolyposis Colorectal Carcinoma (Lynch Syndrome). *Gastroenterology* 130, 312–322. doi:10.1053/j.gastro.2005.10.052

International Mismatch Repair Consortium (2021). Variation in the Risk of Colorectal Cancer in Families with Lynch Syndrome: A Retrospective Cohort Study. *Lancet Oncol.* 22, 1014–1022. doi:10.1016/s1470-2045(21)00189-3

Jabbour, S. K., Lee, K. H., Frost, N., Breder, V., Kowalski, D. M., Pollock, T., et al. (2021). Pembrolizumab Plus Concurrent Chemoradiation Therapy in Patients

with Unresectable, Locally Advanced, Stage III Non-Small Cell Lung Cancer. *JAMA Oncol.* 7, 1351–1359. doi:10.1001/jamaoncol.2021.2301

Kawashima, Y., Nishikawa, S., Ninomiya, H., Yoshida, R., Takano, N., Oguri, T., et al. (2019). Lung Adenocarcinoma with Lynch Syndrome and the Response to Nivolumab. *Intern. Med.* 58, 1479–1484. doi:10.2169/internalmedicine.1673-18

Latham, A., Srinivasan, P., Kemel, Y., Shia, J., Bandlamudi, C., Mandelker, D., et al. (2019). Microsatellite Instability Is Associated with the Presence of Lynch Syndrome Pan-Cancer. *J. Clin. Oncol.* 37, 286–295. doi:10.1200/jco.18.00283

Lau, D., Kalaitzaki, E., Church, D. N., Pandha, H., Tomlinson, I., Annels, N., et al. (2020). Rationale and Design of the POLEM Trial: Avelumab Plus Fluoropyrimidine-Based Chemotherapy as Adjuvant Treatment for Stage III Mismatch Repair Deficient or POLE Exonuclease Domain Mutant colon Cancer: a Phase III Randomised Study. *ESMO open* 5, e000638. doi:10.1136/esmoopen-2019-000638

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.* 372, 2509–2520. doi:10.1056/NEJMoa1500596

Long, Y., Tang, Y., Cai, C., Yu, M., Zhang, M., Chen, R., et al. (2021). The Influence of STK11 Mutation on Acquired Resistance to Immunotherapy in Advanced Non-small Cell Lung Cancer with Lynch Syndrome: a Case Report and Literature Review. *Ann. Palliat. Med.* 10, 7088–7094. doi:10.21037/apm-20-1639

Luchini, C., Bibeau, F., Ligtenberg, M. J. L., Singh, N., Nottegar, A., Bosse, T., et al. (2019). ESMO Recommendations on Microsatellite Instability Testing for Immunotherapy in Cancer, and its Relationship with PD-1/pd-L1 Expression and Tumour Mutational burden: a Systematic Review-Based Approach. *Ann. Oncol.* 30, 1232–1243. doi:10.1093/annonc/mdz116

Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., and Hitchins, M. P. (2015). Milestones of Lynch Syndrome: 1895-2015. *Nat. Rev. Cancer* 15, 181–194. doi:10.1038/nrc3878

Marcus, L., Fashoyin-Aje, L. A., Donoghue, M., Yuan, M., Rodriguez, L., Gallagher, P. S., et al. (2021). FDA Approval Summary: Pembrolizumab for the Treatment of Tumor Mutational Burden-High Solid Tumors. *Clin. Cancer Res.* 27, 4685–4689. doi:10.1158/1078-0432.Ccr-21-0327

Marcus, L., Lemery, S. J., Keegan, P., and Pazdur, R. (2019). FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. *Clin. Cancer Res.* 25, 3753–3758. doi:10.1158/1078-0432.Ccr-18-4070

Masuzawa, K., Asakura, T., Ikemura, S., Yasuda, H., Kawada, I., Takaoka, S., et al. (2020). Long-Lasting Response to Nivolumab for a Patient with Lynch Syndrome-Associated Lung Adenocarcinoma. *JCO Precision Oncol.* 4, 74–78. doi:10.1200/po.19.00156

Møller, P., Seppälä, T., Bernstein, I., Holinski-Feder, E., Sala, P., Evans, D. G., et al. (2017). Cancer Incidence and Survival in Lynch Syndrome Patients Receiving Colonoscopic and Gynaecological Surveillance: First Report from the Prospective Lynch Syndrome Database. *Gut* 66, 464–472. doi:10.1136/gutjnl-2015-309675

Nolan, L., Eccles, D., Cross, E., Crawford, G., Beck, N., Bateman, A., et al. (2009). First Case Report of Muir-Torre Syndrome Associated with Non-Small Cell Lung Cancer. *Fam. Cancer* 8, 359–362. doi:10.1007/s10689-009-9247-7

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y.-Y., et al. (2017). Sherloc: A Comprehensive Refinement of the ACMG-AMP Variant Classification Criteria. *Genet. Med.* 19, 1105–1117. doi:10.1038/gim.2017.37

National Comprehensive Cancer Network (2021). *NCCN Clinical Practice Guidelines in Oncology: Genetic/Familinal High-Risk Assessment: Colorectal.* Available at https://www.nccn.org/professionals/physician_gls/pdf/genetics_colon.pdf [Accessed on: May 11, 2021].

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Rousseau, B., Foote, M. B., Maron, S. B., Diplas, B. H., Lu, S., Argilés, G., et al. (2021). The Spectrum of Benefit from Checkpoint Blockade in Hypermutated Tumors. *N. Engl. J. Med.* 384, 1168–1170. doi:10.1056/NEJMc2031965

Seppälä, T. T., Dominguez-Valentin, M., Sampson, J. R., and Møller, P. (2021). Prospective Observational Data Informs Understanding and Future Management of Lynch Syndrome: Insights from the Prospective Lynch Syndrome Database (PLSD). *Fam. Cancer* 20, 35–39. doi:10.1007/s10689-020-00193-2

Seppälä, T. T., Latchford, A., Negoi, I., Sampaio Soares, A., Jimenez-Rodriguez, R., Sánchez-Guillén, L., et al. (2021). European Guidelines from the EHTG and ESCP for Lynch Syndrome: An Updated Third Edition of the Mallorca Guidelines Based on Gene and Gender. *Br. J. Surg.* 108, 484–498. doi:10.1002/bjs.11902

Sinicrope, F. A., Ou, F.-S., Zemla, T., Nixon, A. B., Mody, K., Levasseur, A., et al. (2019). Randomized Trial of Standard Chemotherapy Alone or Combined with Atezolizumab as Adjuvant Therapy for Patients with Stage III colon Cancer and Deficient Mismatch Repair (ATOMIC, Alliance A021502). *J. Clin. Oncol.* 37, e15169. doi:10.1200/JCO.2019.37.15_suppl.e15169

Skoulidis, F., and Heymach, J. V. (2019). Co-Occurring Genomic Alterations in Non-Small-Cell Lung Cancer Biology and Therapy. *Nat. Rev. Cancer* 19, 495–509. doi:10.1038/s41568-019-0179-8

Stoffel, E. M., Mangu, P. B., Gruber, S. B., Hamilton, S. R., Kalady, M. F., Lau, M. W. Y., et al. (2015). Hereditary Colorectal Cancer Syndromes: American Society of Clinical Oncology Clinical Practice Guideline Endorsement of the Familial Risk-Colorectal Cancer: European Society for Medical Oncology Clinical Practice Guidelines. *J. Clin. Oncol.* 33, 209–217. doi:10.1200/jco.2014.58.1322

Sun, S., Liu, Y., Eisfeld, A.-K., Zhen, F., Jin, S., Gao, W., et al. (2019). Identification of Germline Mismatch Repair Gene Mutations in Lung Cancer Patients with Paired Tumor-Normal Next Generation Sequencing: A Retrospective Study. *Front. Oncol.* 9, 550. doi:10.3389/fonc.2019.00550

Tang, J., Shalabi, A., and Hubbard-Lucey, V. M. (2018). Comprehensive Analysis of the Clinical Immuno-Oncology Landscape. *Ann. Oncol.* 29, 84–91. doi:10.1093/annonc/mdx755

Ten Broeke, S. W., van Bavel, T. C., Jansen, A. M. L., Gómez-García, E., Hes, F. J., van Hest, L. P., et al. (2018). Molecular Background of Colorectal Tumors from Patients with Lynch Syndrome Associated with Germline Variants in PMS2. *Gastroenterology* 155, 844–851. doi:10.1053/j.gastro.2018.05.020

Ten Broeke, S. W., van der Klift, H. M., Tops, C. M. J., Aretz, S., Bernstein, I., Buchanan, D. D., et al. (2018). Cancer Risks for PMS2-Associated Lynch Syndrome. *J. Clin. Oncol.* 36, 2961–2968. doi:10.1200/jco.2018.78.4777

Thompson, B. A., Spurdle, A. B., Spurdle, A. B., Plazzer, J.-P., Greenblatt, M. S., Akagi, K., et al. (2014). Application of a 5-Tiered Scheme for Standardized Classification of 2,360 Unique Mismatch Repair Gene Variants in the InSiGHT Locus-specific Database. *Nat. Genet.* 46, 107–115. doi:10.1038/ng.2854

Wang, M., and Marín, A. (2006). Characterization and Prediction of Alternative Splice Sites. *Gene* 366, 219–227. doi:10.1016/j.gene.2005.07.015

Wang, Y. A., Jian, J.-W., Hung, C.-F., Peng, H.-P., Yang, C.-F., Cheng, H.-C. S., et al. (2018). Germline Breast Cancer Susceptibility Gene Mutations and Breast Cancer Outcomes. *BMC cancer* 18, 315. doi:10.1186/s12885-018-4229-5

Yagishita, S., Horinouchi, H., Sunami, K. S., Kanda, S., Fujiwara, Y., Nokihara, H., et al. (2015). Impact of KRAS Mutation on Response and Outcome of Patients with Stage III Non-squamous Non-Small Cell Lung Cancer. *Cancer Sci.* 106, 1402–1407. doi:10.1111/cas.12740

Yeo, G., and Burge, C. B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.* 11, 377–394. doi:10.1089/1066527041410418

Yurgelun, M. B., and Hampel, H. (2018). Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention. *Am. Soc. Clin. Oncol. Educ. Book* 38, 101–109. doi:10.1200/edbk_208341

Conflict of Interest: SL, XL, TM and LJ are employees of Genetron Health (Beijing) Technology, Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Han, Liu, Cui, Wang, Ma, Jiang, Li and Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fgene.2022.828373

The Role of RAB GTPases and Its Potential in Predicting Immunotherapy Response and Prognosis in Colorectal Cancer

*Xuefei Jiang, Lanlan Yang, Qianling Gao, Yiting Liu, Xingzhi Feng, Shubiao Ye and Zihuan Yang**

Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Guangdong Institute of Gastroenterology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Background: Colorectal cancer (CRC) is the third most common cancer worldwide, in which aberrant activation of the RAS signaling pathway appears frequently. RAB proteins (RABs) are the largest Ras small GTPases superfamily that regulates intracellular membrane trafficking pathways. The dysregulation of RABs have been found in various diseases including cancers. Compared with other members of Ras families, the roles of RABs in colorectal cancer are less well understood.

Methods: We analyzed the differential expression and clinicopathological association of RABs in CRC using RNA sequencing and genotyping datasets from TCGA samples. Moreover, the biological function of RAB17 and RAB34 were investigated in CRC cell lines and patient samples.

Results: Of the 62 RABs we analyzed in CRC, seven (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25) were significantly upregulated, while six (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34) were significantly downregulated in tumor tissues as compared to normal. We found that the upregulated-RABs, which were highly expressed in metabolic activated CRC subtype (CMS3), are associated with cell cycle related pathways enrichment and positively correlated with the mismatch repair (MMR) genes in CRC, implying their role in regulating cell metabolism and tumor growth. While, high expression of the downregulated-RABs were significantly associated with poor prognostic CRC mesenchymal subtypes (CMS4), immune checkpoint genes, and tumor infiltrating immune cells, indicating their role in predicting prognosis and immunotherapy efficacy. Interestingly, though RAB34 mRNA is downregulated in CRC, its high expression is significantly associated with poor prognosis. *In vitro* experiments showed that RAB17

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Xiaofan Ding,
The Chinese University of Hong Kong, Hong Kong SAR, China
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

***Correspondence:**
Zihuan Yang
yzhuan@mail.sysu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *03 December 2021*
Accepted: *10 January 2022*
Published: *28 January 2022*

Citation:
Jiang X, Yang L, Gao Q, Liu Y, Feng X, Ye S and Yang Z (2022) The Role of RAB GTPases and Its Potential in Predicting Immunotherapy Response and Prognosis in Colorectal Cancer. Front. Genet. 13:828373. doi: 10.3389/fgene.2022.828373

Abbreviations: AP, autophagosome; C, cilium; COAD, colon cancer; CRC, colorectal cancer; CMS, consensus molecular subtypes; DCs, dendritic cells; DFS, disease-free survival; dMMR, mismatch repair deficient; E, endosome; EE, early endosome; ER, endocytic recycling; EGFR, epidermal growth factor receptor; G, golgi; GO, gene ontology; GTPases, small guanosine triphosphatases; KEGG, Kyoto encyclopedia of genes and genomes; LE, late endosome; LY, lysosome; MS, melanosome; MMPs, matrix metalloproteinases; MMR, mismatch repair; MSI-H, microsatellite instability-high; OS, overall survival; P, peroxisome; PM, plasma membrane; PS, phagosome; PPI, protein interaction network; RABs, RAB proteins; READ, rectal cancer; RE, recycling endosome; TE, tubular endosome. TGN, *trans*-golgi network; TCGA, The cancer genome atlas; TMB, tumor mutation burden.

overexpression can promote cell proliferation via cell cycle regulation. While, RAB34 overexpression can promote cell migration and invasion and is associated with PD-L1/PD-L2 expression increase in CRC cells.

Conclusions: Our study showed that RABs may play important roles in regulating cell cycle and immune-related pathways, therefore might be potential biomarkers in predicting prognosis and immunotherapy response in CRC.

Keywords: RABs, RAB17, RAB34, colorectal cancer, immune, cell cycle

INTRODUCTION

Colorectal cancer (CRC) is the third most prevalent cancer worldwide (Siegel et al., 2020; Siegel et al., 2021). More than 30% of CRC are driven by mutations of the RAS family of genes (Crockett and Nagtegaal, 2019; Du F. et al., 2020; Harada and Morlote, 2020). Oncogenic RAS mutations result in abnormal activation of the EGFR signaling pathway, hence facilitate tumor growth, progression and immune evasion (Sanchez-Vega et al., 2018; Kerk et al., 2021). Targeting Ras-mutant CRC remains a grand challenge to clinical treatment. It is important to better explore the function and mechanism of the RAS family proteins, complexes of RAS proteins with effector and regulators, pathways that are enriched in cancer cells.

The Ras-associated binding proteins (RABs) comprise the largest family of Ras small guanosine triphosphatases (GTPases) that cycle between an inactive guanosine diphosphate (GDP)-bound state and an active guanosine triphosphate (GTP)-bound state. RABs are the identified to be involved in the regulation of vesicular trafficking, including vesicle formation, transportation, membrane docking, and fusion (Pfeffer, 2017; Guerra and Bucci, 2019). In particular, RABs prescribe the directionality of membrane-bound cargo traffic to ensure that the cargo is delivered to the correct destination (Zhen and Stenmark, 2015). Presently, more than 60 RABs have been identified in humans (Anand et al., 2020). Deregulation of RABs has been reported in several diseases, including various cancers (Sun et al., 2018; Ganga et al., 2021). They function either as an oncogene or tumor suppressors depending on cancer types. Nevertheless, the roles of RABs in colorectal cancer are less-well understood.

In this study, we screened multiple genomic datasets for RABs that were differentially expressed between normal and tumor tissues, and systematically evaluated the role of RABs in CRC, including their association with CRC molecular subtypes, immune response, and prognosis. Moreover, we assessed the functional mechanisms of RAB17 and RAB34 in CRC cells.

MATERIALS AND METHODS

Patients and Sample Collection

This study was approved by the ethics committee of the Sixth Affiliated Hospital of Sun Yat-sen University. All patients provided full consent for the study. A total of 4 cases pairs of tumor and adjacent normal tissues (5 cm away from the tumor border) from CRC patients analyzed in this study were provided from the Tissue Bank, Sixth Affiliated Hospital of Sun Yat-sen University, China.

Plasmid Construction and Transfection

The pcDNA 3.1 (+) plasmid was purchased from Invitrogen. The overexpression plasmids were carried out by inserting the RAB17 and RAB34 sequence into the pcDNA3.1+ plasmid at multiple cloning sites with HindIII restriction enzymes (New England Biolabs, CA, United States) and In-Fusion HD Cloning Kit (Clontech, CA, United States). The overexpression plasmids were transfected into CRC cells with Lipofectamine 3000 and p3000 (Invitrogen, MA, United States) according to the manufacturer's protocol.

Cell Culture

CRC cell lines were purchased from the American Type Culture Collection (ATCC, VA, United States). HCT116 and SW480 cells were maintained in RPMI-1640 (GIBCO, NY, United States), supplemented with 10% (v/v) fetal bovine serum (FBS, GIBCO, NY, United States). Cells were allowed to grow in a humidified incubator with 5% CO_2 at 37°C.

Western Blot Assays

Protein was extracted by T-PER tissue protein extraction reagent (Thermo, Rockford, IL) with protease inhibitor cocktail set III and phosphatase inhibitor cocktail set II (Millipore, Germany) according to the manufacturer's protocol. Primary antibodies included GAPDH as loading controls (Proteintech, #60004-1-Ig), RAB17 (Proteintech, #17501-1-AP), RAB34 (SANTA, #SC376898), HA-tag (CST, #3724), CDK2 (CST, #2546), CyclinB1 (Proteintech, 55004-1-AP), PD-L1 (CST, #13684), PD-L2 (Abcam, #ab187662).

Cell Proliferation, Migration and Invasion Assay

Cell proliferation was constructed with the Incucyte ZOOM device (ESSEN bioscience, United States). 5000 cells/well were seeded into a 96-well plate and automatically monitored and recorded every 2 h by the Incucyte device.

As for the wound healing assay, 1×10^5 cells were seeded into a 12-well plate, and wounds were made by a scratcher (ibidi, WI, United States). The size of the wound was automatically captured every 2 h by the Incucyte ZOOM device and measured by ImageJ software. Each assay was

repeated three times. Transwell assay was constructed with cell culture plates of 24 well 8.0 μm pore size (Falcon, CO, United States) with Matrigel (BD Biosciences, NJ, United States). $1x10^5$ cells were seeded into the upper chamber in 0.1 ml of serum-free RPMI-1640, and 0.5 ml of RPMI-1640 with 20% FBS was placed in the lower chamber as a chemoattractant. Invasion cells on the other side of the membrane were fixed and stained with crystal violet for 5 min after 48 h culture. Each assay was repeated three times.

Cell Cycle Assays

Cell cycle assay was constructed with Cell Cycle Staining Solution (Multi Sciences, China) according to the manufacturer's protocol. The flow cytometry results were measured by the Beckman device (BD Biosciences, NJ, United States) and analyzed with Flow jo 10.0 software. Each assay was repeated three times.

Immunohistochemistry Assay

Each sample was deparaffinized for antigen retrieval using sodium citrate (pH 6.0) for 10 min and subsequent incubation with the respective primary antibody: RAB17 (Proteintech, #17501-1-AP), RAB34 (Affinity, #AF9174). Antibodies were used at 1:100 dilution. The reaction was developed using hematoxylin for counterstaining for 2 min. In all cases, sections from normal colonic mucosa distant from the tumor site were used as negative controls.

Colony Formation Assays

For colony formation assay, cells were seeded in a six-well plate (500 cells per well) and allowed for growth for 2 weeks. The culture medium was refreshed every 5 days. At the end of the experiment, colonies were stained by crystal violet.

Structure Similarity Analysis

Structure similarity of RABs were analyzed through sequence alignment. The sequence of each RABs were obtained from UniProt (https://www.uniprot.org/) and aligned by Clustal Omega online tool (Madeira et al., 2019). The phylogenetic relationships are obtained by the neighbor-joining program in the Jalview package. Finally, sequence conservation, consensus residues, and phylogenetic tree were visualized by Jalview (Waterhouse et al., 2009).

Bioinformatic Analysis

GEPIA (http://gepia.cancer-pku.cn/index.html) was used to analyze RABs expression in normal and tumor tissues, as well as the correlation between RABs expression and patient prognosis. Genetic alterations of RABs were analyzed by the c-Bio Cancer Genomics Portal (https://www.cbioportal.org/) website, and the tab OncoPrint shows an overview of genetic changes for each sample in RABs. GeneMANIA (http://www.genemania.org) was used to predict the related genes of RABs. Metascape (https://metascape.org) was used to enrich the functional pathways of RABs. TIMER (https://cistrome.shinyapps.io/timer/) was used to evaluate the level of tumor-infiltrating immune cells (TIIC). The correlation between CXCL11 and TILs was measured by Spearman's test. Selected RABs were input via the "Immune module" and "Exploration module." Home for Researchers (https://www.home-for-researchers.com) was used to evaluate the correlation between RABs expression and tumor mutation burden (TMB). Gene Set Enrichment Analysis (GSEA) analysis was conducted with GSEA preranked tool (http://software.broadinstitute.org/GSEA/msigdb/annotate.jsp).

Statistical Analysis

Statistical analysis was constructed using the R 3.6.3 version and GraphPad Prism 8.0.1. Statistical significance was analyzed by a two-tailed Student's *t*-test. Spearman's correlation was performed to analyze correlations. Kaplan-Meier Survival analysis was performed with a log-rank test. Univariate and multivariate survival analyses were performed using the Cox regression analyses model. Statistical significance was defined by a two-tailed $p < .05$.

RESULTS

Gene Expression, Genetic Alteration, and Sequence Alignment of RABs in Colorectal Cancer

We first analyzed the mRNA expression profiles of 62 RABs in CRC using The Cancer Genome Atlas (TCGA) database and Genotype-Tissue Expression GTEx database through the GEPIA website portal. 13 RABs showed significantly different expressions in tumor tissues as compared to normal tissues in colon cancer (COAD) and rectal cancer (READ). Among them, seven were significantly upregulated, including RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25 (**Figure 1A**; **Supplementary Figure 1A**), while the other six, RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34, were significantly downregulated (**Figure 1B**; **Supplementary Figure 1B**). The correlation of the RABs with each other was analyzed via the R ggplot 2 (**Supplementary Figure 1C**).

Next, we examined the genetic alterations of RABs in 2544 CRC patients using the c-BioPortal tool. RAB20 possesses the highest alteration frequency (1.3%) which was mainly due to amplification alteration in 34 cases (**Supplementary Figure 2B**). Overall, the mutation and deep deletion alteration occupied a higher proportion in downregulated-RABs, which may contribute to their downregulation at transcriptional level (**Figure 1C**).

Sequence alignment was then performed to evaluate the structural similarity of RABs. The RABs exhibit moderate sequence identity (~30–40%) and possess some highly conserved motif (**Figure 1D**; **Supplementary Figure 2A**). The downregulated RABs and the upregulated RABs formed relative distinct groups in

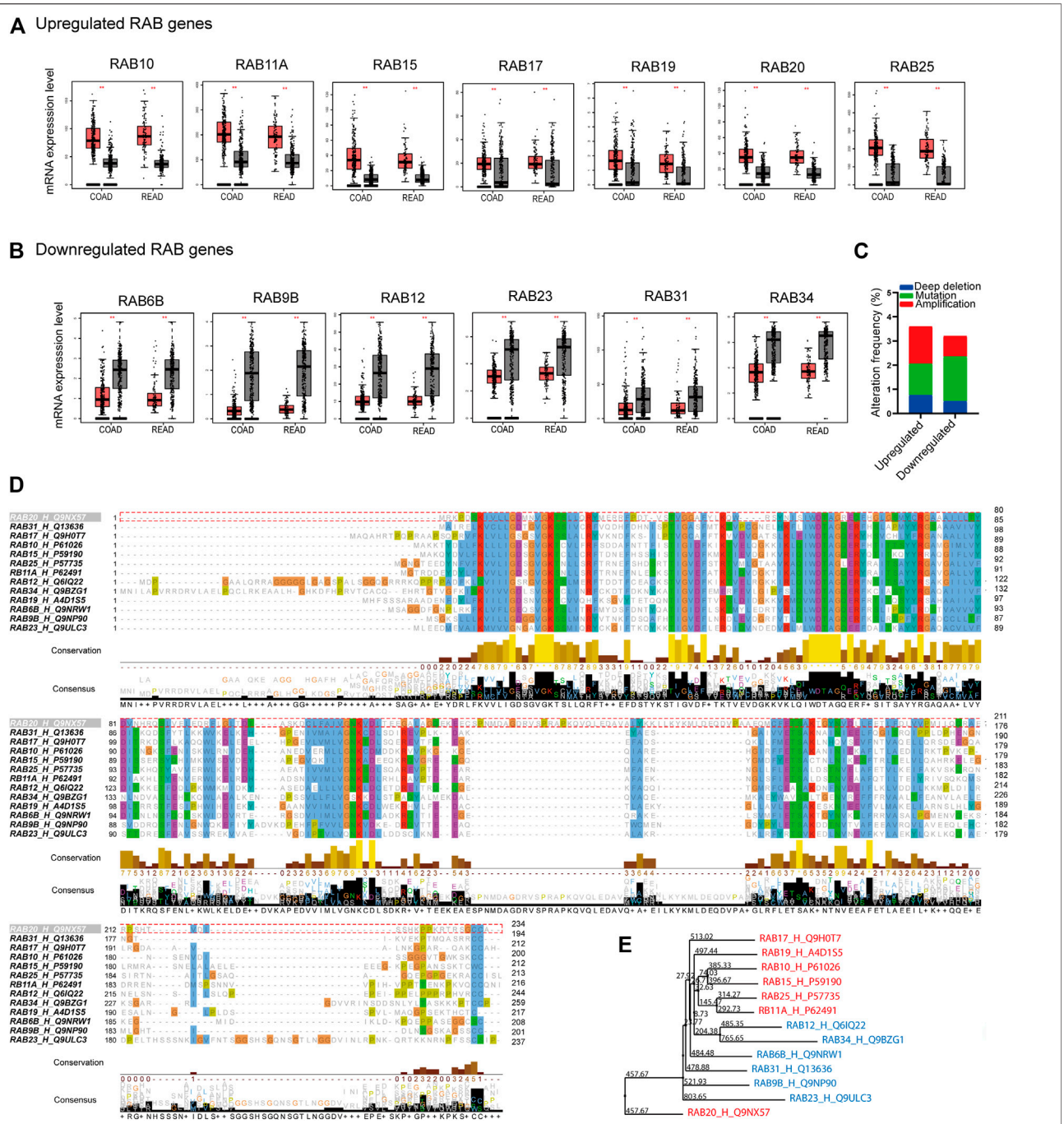


FIGURE 1 | Differentially expressed RABs, genetic alteration, and sequence alignment of the RABs in colorectal cancer. **(A)** 7 RABs (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25) were significantly upregulated in tumor tissues. **(B)** 6 RABs (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34) were significantly downregulated in tumor tissues. Red and grey bar represents tumor and normal tissues, respectively. Tumor tissues (COAD: $n = 275$, READ: $n = 92$) and normal tissues (COAD: $n = 349$, READ: $n = 318$). **(C)** Genetic alterations frequency of each RABs in 2544 CRC patients. **(D)** Sequence alignment of the RABs. **(E)** Phylogenetic tree of the maximum likelihood analysis. $^{**}p < .01$.

phylogenetic analysis (**Figure 1E**). The phylogenetic distance between RAB20 and other RABs is the largest, which may be due to the highest genetic alteration frequency found in RAB20 (**Supplementary Figure 2B**). Previous studies have shown that the up-regulated RABs are usually located in endosome system (E), whereas, the down-regulated RAB genes are usually located in Golgi (G) (Langemeyer et al., 2018; Homma et al., 2021) (**Supplementary Figure 2C**). The above results indicated that though the RABs have a high degree of structural

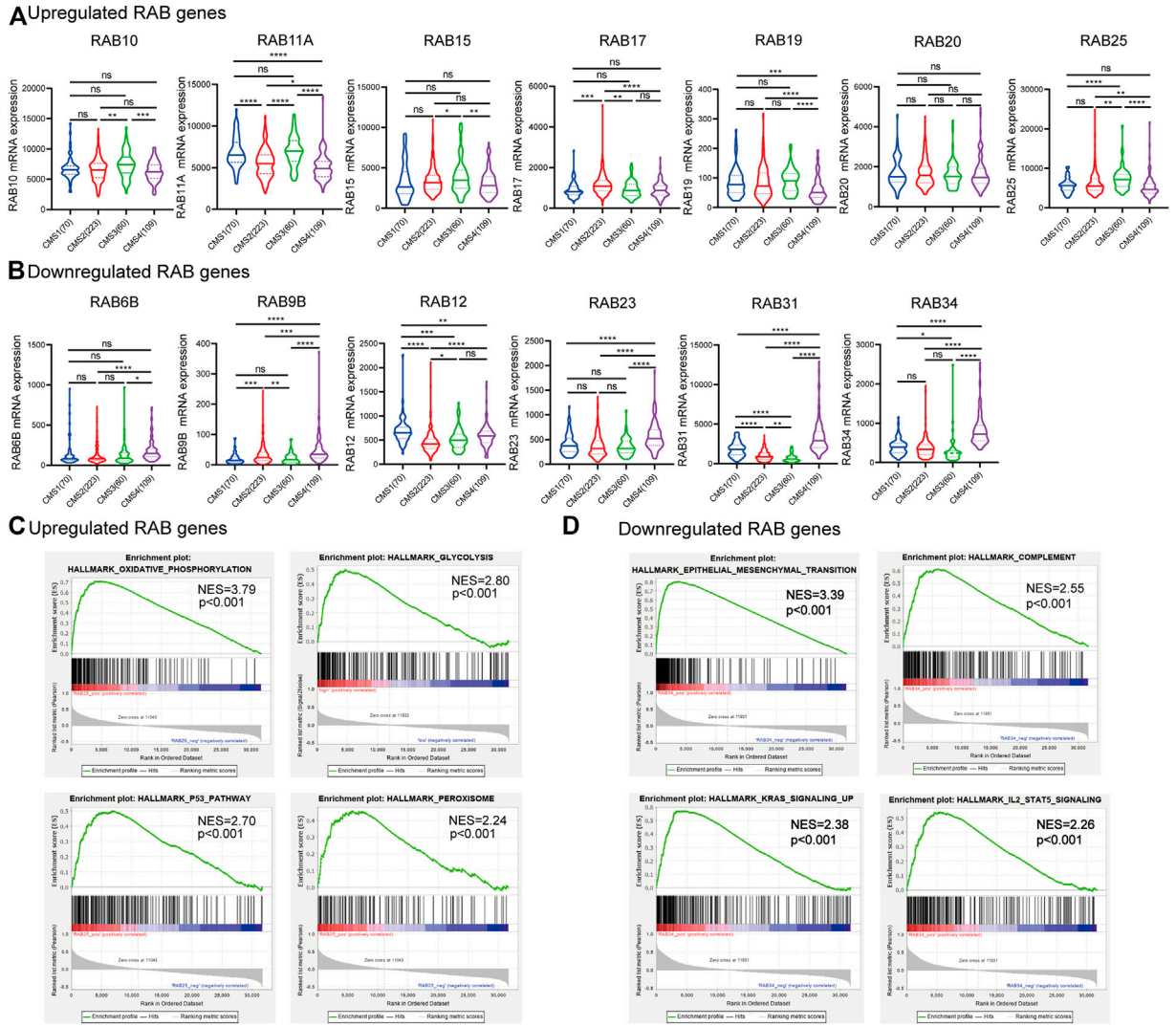


FIGURE 2 | Correlation between RABs expression and CMSs and GESA enrichment analysis in CRC patients. **(A)** 7 RABs (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25) were significantly correlated with CMS3. **(B)** 6 RABs (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34) were significantly correlated with CMS4. Statistical significance was defined by a two-tailed *t*-test. **(C–D)** Gene set enrichment analysis of RABs. **p* < .05, ***p* < .01, ****p* < .001, *****p* < .0001.

homology, the functions may vary depending on their subcellular localization.

Correlation Between RABs Expression and Molecular Signatures, and GESA Enrichment Analysis

Transcriptional profiling has identified four consensus molecular subtypes (CMSs) of CRC possessing distinct molecular signatures and prognostic profiles (Guinney et al., 2015). CMS1 is characterized with microsatellite unstable and strong immune activation; CMS2 is characterized with WNT and MYC signaling activation; CMS3 shows evident metabolic dysregulation; and CMS4 is associated with TGFβ activation, stromal invasion, and angiogenesis. Therefore, we investigated the correlations between RABs expression and CMSs. We found that a high level of the upregulated-RABs was significantly associated with the metabolic CMS3 subtype (**Figure 2A**), implying their involvement in the epithelial and metabolic dysregulation. Interestingly, the downregulated-RABs were significantly highly expressed in mesenchymal CMS4 tumors (**Figure 2B**), which are usually diagnosed at more advanced stages and more aggressive and metastatic than other CMS subtypes. The results indicated that though downregulated-RABs are low expressed in tumor tissues compared to normal tissues, their upregulation might result in a more aggressive cancer progression.

Consistently, GESA analysis showed that the upregulated-RABs mainly were enriched with metabolic-related pathways, while, the downregulated-RABs were enriched with EMT and immune-related pathways (**Figures 2C,D**).

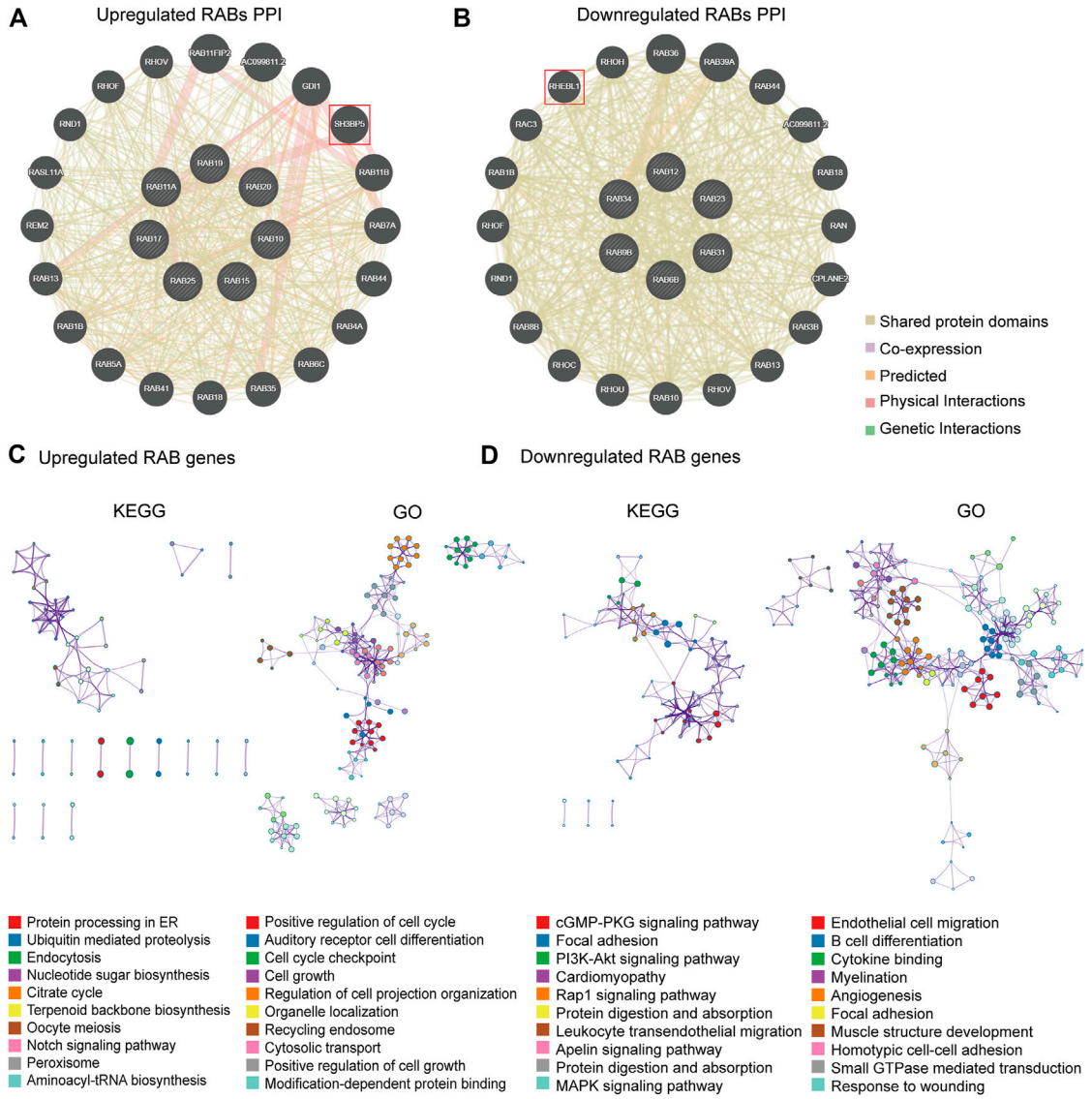


FIGURE 3 | PPI, KEGG, and GO enrichment analysis for RABs. **(A–B)** The protein interaction network of RABs related genes in CRC. **(C–D)** Top 10 enrichment terms in KEGG pathways and GO in RABs (*p*-value of each pathway were shown in the **Supplementary Tables S1–4**).

To further figure out the critical intracellular signaling pathways related to RABs in CRC, we then performed Protein interaction network (PPI), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The PPI results showed that the SH3BP5 and RHEBL1 were most closely associated with upregulated-RABs and downregulated-RABs, respectively (**Figures 3A,B**). SH3BP5 sustained the JNK activity and played a critical role in modulating cell death and cell cycle (Win et al., 2018; Lucero et al., 2019). RHEBL1 activated NFkB-mediated gene transcription and play an important role in the immune process (Yuan et al., 2005). The top 100 genes that were most positively relevant in the RABs were used for KEGG and GO enrichment analyses. Consistent with the PPI analysis results, cell cycle-related pathways were significantly enriched by upregulated-RABs (**Figure 3C**), whereas, immune-related pathways were enriched by downregulated-RABs (**Figure 3D**).

Correlation Between RABs Expression and Immunotherapy Biomarkers

Immunotherapy has shown impressive results in patients with mismatch repair deficient (dMMR) or microsatellite instability-high (MSI-H) CRC (Supek and Lehner, 2015). Tumors that have a dMMR–MSI-H signature are usually associated with a high tumor mutation burden (TMB) and immune cell infiltration (Mandal et al., 2019). Therefore, we firstly investigated the correlation between the expression of RABs and four MMR genes (MLH1, MSH2, MSH6, and PMS2) by the R ggradar

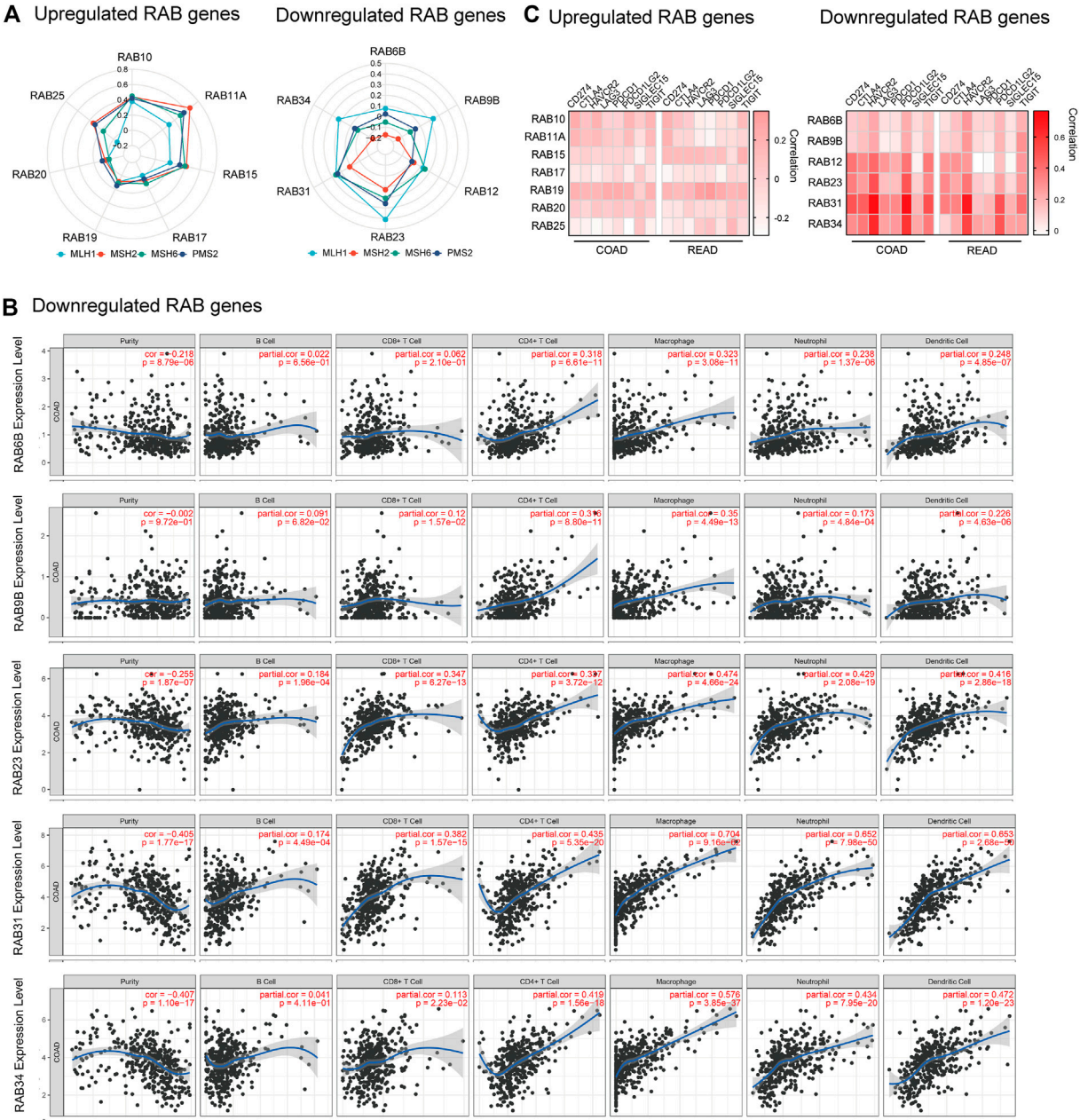


FIGURE 4 | Correlation between RABs expression and immunotherapy biomarkers in CRC patients. **(A)** Correlation between the expression of RABs and MMR genes (*p*-value of each pathway were shown in Supplementary Tables S5–6). **(B)** Correlation between RABs and immune cells infiltration levels in COAD. **(C)** Correlation between the expression of RABs and immune checkpoint genes. Spearman's correlation coefficients were shown above the bar graphs.

and ggplot2. The results showed that the level of upregulated-RABs was positively correlated with the MMR genes in CRC. However, this correlation is not obvious in the downregulated-RABs (**Figure 4A**). Interestingly, we found that only the expression of downregulated-RABs was positively correlated with infiltrating levels of $CD8^+$ T cells, $CD4^+$ T cells, macrophage, neutrophils, and dendritic cells in CRC (**Figure 4B**; **Supplementary Figure S3–4**). Moreover, the overall correlation between RABs expression and TMB was not significant (**Supplementary Figure S5**). MMR is an intracellular process contributing to the fidelity of DNA synthesis and replication and is required for cell cycle regulation (Li et al., 2016). The positive correlation between MMR protein expression and upregulated-RABs may explain the above results that the cell cycle-related pathways were enriched in upregulated-RABs.

To further elucidate the possible role of RABs in regulating immunotherapy response, we then evaluated the correlation

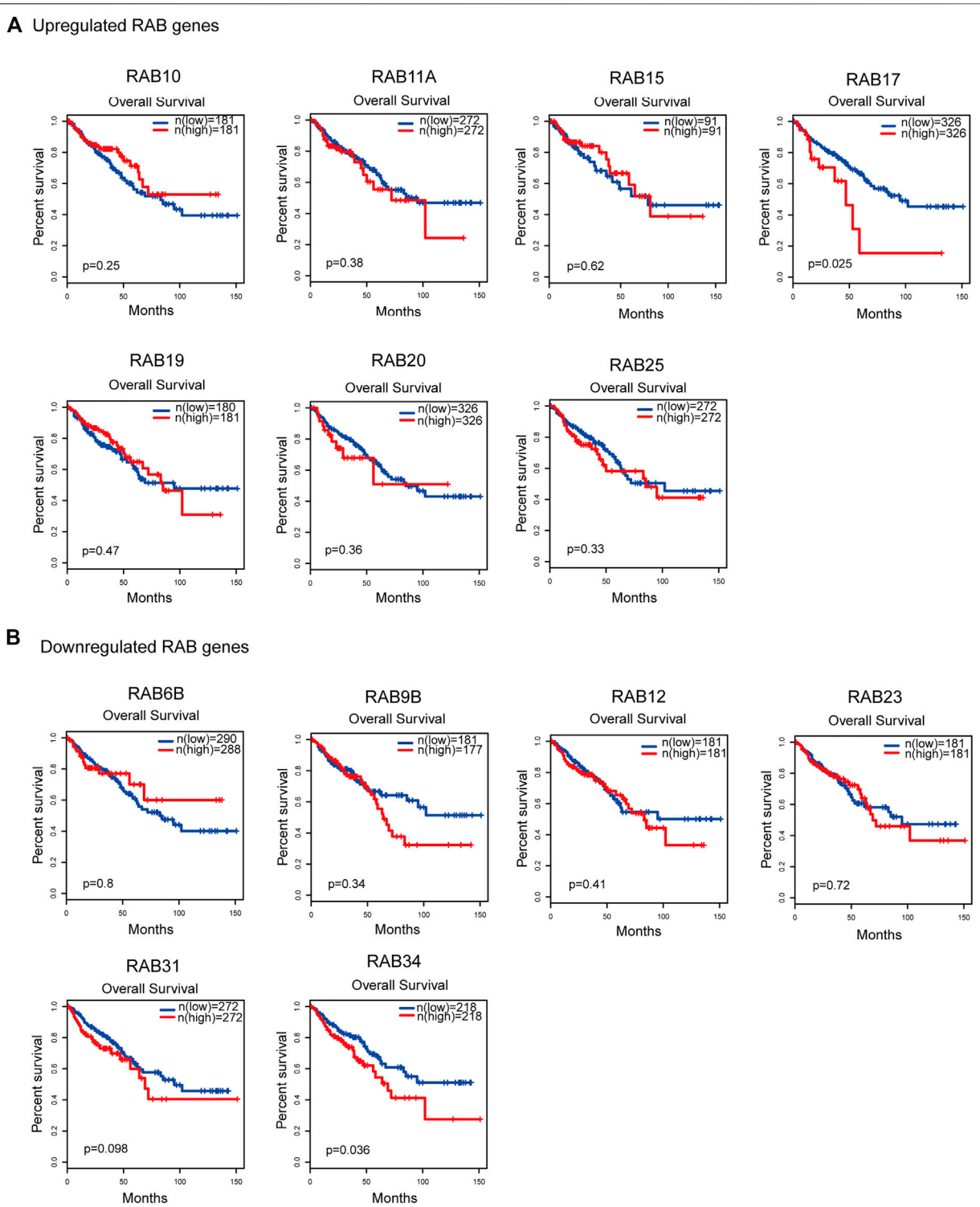


FIGURE 5 | Kaplan–Meier plots for the overall survival of CRC patients stratified by the mRNA expression level of RABs. **(A)** The overall survival curves of the upregulated-RABs (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25). **(B)** The overall survival curves of the downregulated-RABs (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34).

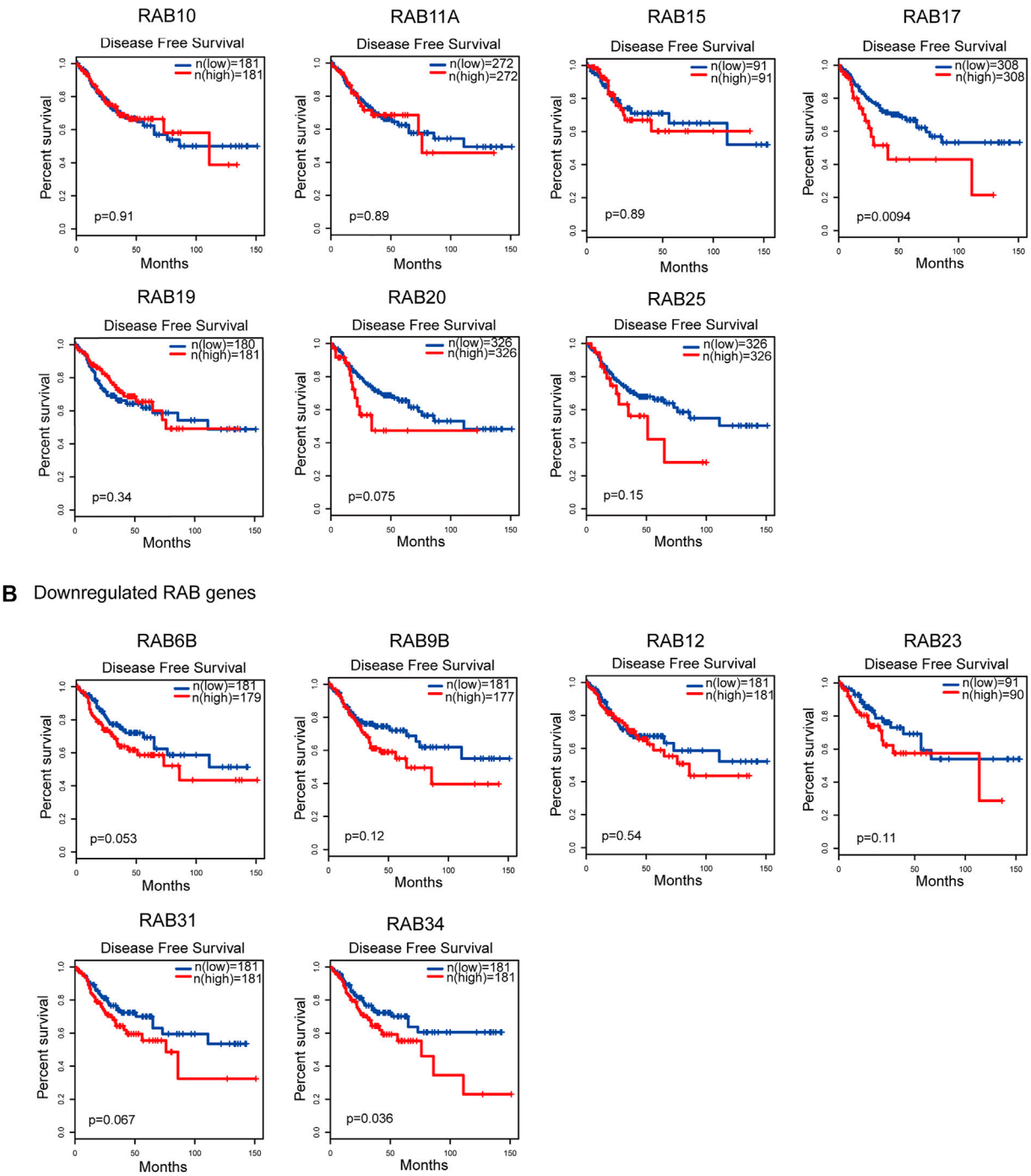


FIGURE 6 | Kaplan–Meier plots for the disease-free survival of CRC patients stratified by the mRNA expression level of RABs. **(A)** The disease-free survival curves of the upregulated-RABs (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25). **(B)** The disease-free survival curves of the downregulated-RABs (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34).

TABLE 1 | Multivariate Cox regression analysis of RAB17 mRNA expression in CRC patients from TCGA.

Variable	OS		DFS	
	HR (95% CI)	*p* value	HR (95% CI)	*p* value
RAB17 expression (high vs. low)	1.792 (1.062–2.816)	0.028*	1.699 (1.016–2.840)	0.043*
Median age (<=68 years vs . >68 years)	0.426 (0.255–0.711)	0.001**		
pN stage (N2 vs. N0+N1)	1.866 (1.086–3.205)	0.024*		
pM stage (M1 vs. M0)	4.109 (2.389–7.066)	0.001***	3.185 (1.724–5.885)	0.001***
Clinical stage (I+II vs. III+IV)				

*p < .05, **p < .01, ***p < .001.

TABLE 2 | Multivariate Cox regression analysis of RAB34 mRNA expression in CRC patients from TCGA.

Variable	OS	*p* value	DFS	*p* value
	HR (95% CI)		HR (95% CI)	
RAB34 expression (high vs. low)			1.944 (1.170–3.231)	0.010*
Median age (<=68 years vs . >68 years)	0.468 (0.282–0.777)	0.003**		
pT stage (T3–T4 vs. T1–T2)				
pN stage (N2 vs. N0+N1)				
pM stage (M1 vs. M0)	3.458 (1.869–6.398)	0.001***	3.419 (1.923–6.078)	0.001***
Clinical stage (I+II vs. III+IV)	0.487 (0.264–0.900)	0.022*		

*p < .05, **p < .01, ***p < .001.

between RABs expression and immune checkpoint genes. We found that the expression of downregulated-RABs was more significantly positive correlated with immune checkpoint genes the R ggplot2 (**Figure 4C**), which is in line with the above analysis that immune-related pathways were enriched by downregulated-RABs.

Correlation Between RABs Expression and Clinical Outcomes

To evaluate the prognostic value of differentially expressed RABs in CRC, we first analyzed the correlations between different RABs and clinical outcomes using the datasets from the GEPIA database. Except for RAB17 and RAB34, the expression of all the other RABs was not significantly associated with overall survival (OS) and disease-free survival (DFS). Kaplan–Meier survival curve showed that high mRNA levels of RAB17 and RAB34 were significantly associated with poor OS and DFS (**Figures 5**, **6**). Thus, we further assessed the correlation between RAB17 and RAB34 mRNA expression and the clinicopathological features of TCGA CRC patients (n = 454). RAB17 was significantly associated with pN status, pM status, and TNM stage (**Supplementary Table S7**). RAB34 was significantly associated with pN status (Supplementary Table 8). Univariate and multivariate Cox regression analyses demonstrated that RAB17 was an independent prognostic factor for OS and DFS (**Table 1**; **Supplementary Table S9**), and RAB34 was an independent predictor for DFS (**Table 2**; Supplementary Table 10).

RAB17 Promotes Cell Proliferation via Cell Cycle Regulation in CRC Cells

We then investigated the role of RAB17 *in vitro*. Immunohistochemical staining showed that RAB17 was highly expressed in tumors (**Figure 7A**; **Supplementary Figure S6A**). Overexpression of RAB17 in HCT116 and SW480 cells significantly promoted cell growth (**Figure 7B**). Similarly, colony formation assay showed that overexpression of RAB17 in HCT116 and SW480 cells increased the colony number compared with the corresponding controls (**Figure 7C**). Our previous analysis showed that upregulated-RABs were highly correlated with the cell cycle, therefore, we investigated the involvement of RAB17 in regulating cell cycle progression using flow cytometry. The results showed that overexpression of RAB17 resulted in an increased percentage of cells in the G2M phase and a decreased percentage of cells in the S phase in HCT116 and SW480

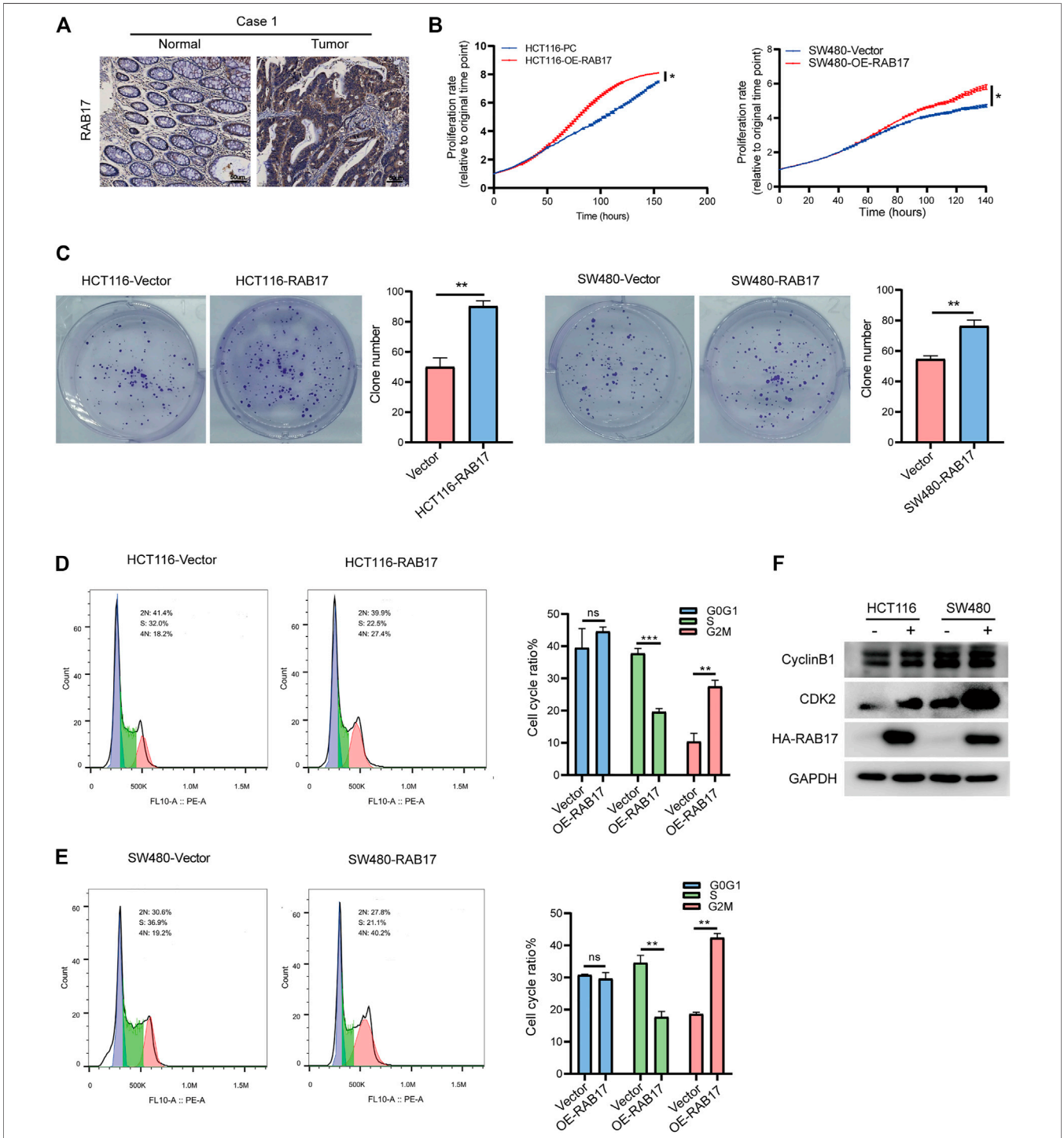


FIGURE 7 | Overexpression of RAB17 promotes cell proliferation of CRC. **(A)** Immunohistochemical staining of patients with colorectal cancer. **(B)** Incucyte assay revealed cell proliferation in HCT116 cells and SW480 cells. **(C)** Overexpression of RAB17 analyzed the colony-forming ability of HCT116 cells and SW480 cells. **(D–E)** Overexpression of RAB17 analyzed cell population both in HCT116 cells **(D)** and SW480 cells **(E)** by flow cytometer. **(F)** Overexpression of RAB17 analyzed the protein levels of CDK2 and cyclin B1 in the HCT16 cells and SW480 cells. *$p < .05$, **$p < .01$, ***$p < .001$.

cells (**Figures 7D,E**). Overexpression of RAB17 significantly increased the protein levels of CDK2 and cyclinB1, which are the cell cycle checkpoint genes regulating S to G2M transition, in the HCT16 cells and SW480 cells (**Figure 7F**). These results demonstrated that RAB17 could promote CRC cells proliferation via regulating cell cycle progression.

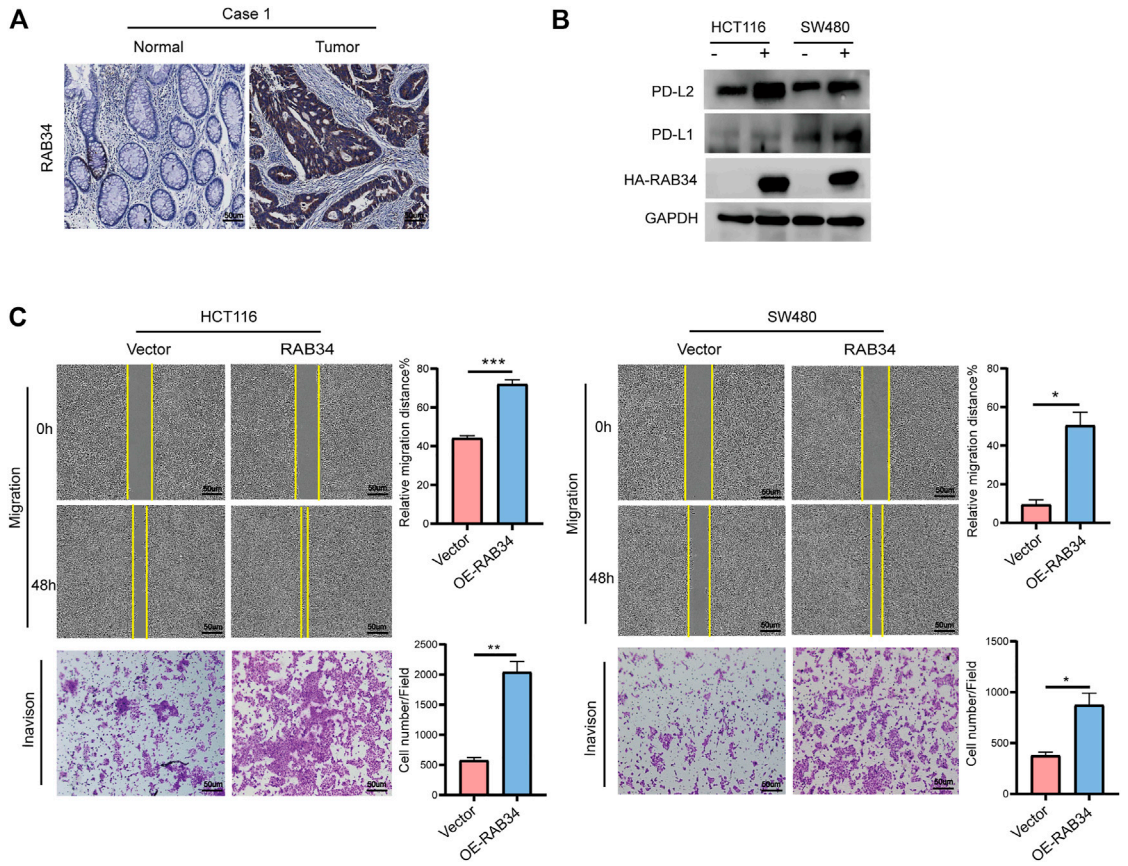


FIGURE 8 | Overexpression of RAB34 increased PD-L1 and PD-L2 expression of CRC. **(A)** Immunohistochemical staining of patients with colorectal cancer. **(B)** Overexpression of RAB34 analyzed the protein levels of PDL-1 and PD-L2 in the HCT16 cells and SW480 cells. **(C)** Overexpression of RAB34 analyzed the migration and invasion in the HCT16 cells and SW480 cells. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

RAB34 Overexpression Promotes Cell Migration and Invasion

Our immunohistochemistry analysis showed strong staining signals of RAB34 in CRC tumor tissues but low in normal tissues (**Figure 8A**; **Supplementary Figure S6B**). Given that RAB34 was significantly upregulated in CMS4 tumors, which showed mesenchymal activation and high invasion ability, we assessed if RAB34 overexpression in CRC cells can promote cell migration and invasion. The wound-healing and transwell assay results showed that overexpression of RAB34 remarkably promotes cell migration and invasion in HCT116 and SW480 cells (**Figure 8C**). As shown in **Figure 4C**, the expression of RAB34 was positively correlated with immune checkpoint genes and RAB34, we chose PD-L1 and PD-L2 for further verification by western blot in CRC cells. The results showed that RAB34 overexpression can significantly increase the protein levels of PD-L1 and PD-L2 in CRC cells (**Figure 8B**). These results combined with the survival analysis indicated that RAB34 is an unfavorable prognostic factor, more importantly, targeting RAB34 might provide a novel therapeutic strategy to improve the responses to anti-PD-1 therapy in CRC.

DISCUSSION

More than 1.2 million patients are diagnosed with colorectal cancer every year and about 600,000 die of the disease (Brenner et al., 2014). Approximately 35–45% of CRC patients are RAS mutant, which is associated with therapy resistance (Wong et al., 2020). It is important to better explore the function of the RAS family proteins and the related pathways driven by activated RAS, therefore, to find new drugable targets for clinical treatment of CRC.

Rab proteins, which belongs to the RAS family, are vital components of the membrane trafficking system that controls secretion, transport, recycling, and degradation of many tumor-associated proteins, such as beta-integrins, epidermal growth factor receptor (EGFR), and matrix metalloproteinases (MMPs) (Dozynkiewicz et al., 2012; Anand et al., 2020). The role of RABs in the shedding of extracellular vesicles has also been reported (Peinado et al., 2012; Recchi and Seabra, 2012). Aberrant expression of RABs have been found in various cancers (Wang et al., 2020), but the roles of RABs in colorectal cancer are less well understood.

In this study, we evaluate the role of RABs in CRC using public datasets and verify their functions *in vitro*. We found that though

the RABs exhibit a high degree of sequence similarity, but the function are vary may be due to their different subcellular location. The RABs play critical roles in regulating cell cycle progression, immune cell infiltration, and might be predictive markers of immunotherapy and patient survival in CRC.

According to the expression level of RABs in tumors compared with that in normal tissues, we divided them into the upregulated and downregulated groups. We found that the upregulated-RABs identified in CRC (RAB10, RAB11A, RAB15, RAB17, RAB19, RAB20, and RAB25) were significantly associated with the metabolic CMS3 subtype and cell cycle regulation. The CMS3 tumors often display remarkable metabolic deregulation with higher KRAS mutations (68%) (Lal et al., 2018; Smeby et al., 2018), which is consistent with the GSEA enrichment analysis. Metabolic plasticity is critical for DNA damage repair and cell cycle regulation, which subsequently affect cancer cell survival, growth, and proliferation (Alcalá et al., 2020; Kumarasamy et al., 2021). Previous studies have shown that RAB10, RAB11A, RAB17, and RAB25 could promote cell proliferation in different cancers (Li et al., 2015; Du J. et al., 2020; Zhang et al., 2020). Elevated expression of Rab25 was correlated with poor prognosis and aggressiveness of renal, lung, breast, ovarian, and other cancers. However, the tumor suppressor function of Rab25 was reported in several cancers, such as colorectal cancer, indicating the tumor type-specific function of Rab25 (Hong et al., 2018; Jeong et al., 2018; Cho and Lee, 2019; Temel et al., 2020). Our results reveal that RAB17-overexpressed CRC cells exhibited an increase in the percentage of cells in the G2/M phase. Furthermore, RAB17 overexpression induced an increase of cyclin B and CDK2 protein levels, which are important checkpoint genes in controlling the S to G2/M transition. More importantly, we found that RAB17 is an independent marker significantly associated with OS and DFS in CRC. It is believed that cytotoxic cancer chemotherapy drugs usually kill dividing cells that proliferate fast (He et al., 2019). Collectively, these observations suggested that the upregulated RABs, especially RAB17, could regulate cell cycle progression and might be useful prognostic markers to be used to stratify patients into a group that would benefit from chemotherapy treatment.

As to the downregulated-RABs (RAB6B, RAB9B, RAB12, RAB23, RAB31, and RAB34), we found their expression levels were significantly correlated with immune cells infiltration. Consistently, the GESA analysis showed that the downregulated-RABs are enriched with immune-related pathways including IL2-STAT5 signaling and complement related pathways. Studies have shown that RAB23 and RAB31 participate in the autophagy process, which play a significant role in immunity (Zheng et al., 2017). RAB34 is associated with lysosomal distribution, hence affecting antigen presentation by dendritic cells and $CD8^+$ T activation (Alloatti et al., 2015). Moreover, the expression levels of immune checkpoint genes are key factors affecting immunotherapy response. Our analysis revealed that the level of downregulated RABs are positively correlated with immune checkpoint genes expression. Furthermore, *in vitro* verification showed that RAB34-overexpression induced a significant upregulation of PD-L1 and PD-L2, the two ligands of PD-1. Therefore, we speculated that the downregulated-RABs are not only be involved in regulating the transportation and membrane translocation of the immune checkpoint genes but may also modulate the tumor microenvironment by recruiting immune cells to the tumor site. The level of down-regulated RABs may help to predict the immunotherapy response.

Among the 13 differentially expressed RABs, RAB17 and RAB34 were independent markers significantly associated with poor OS and DFS. Interestingly, though the mRNA level of RAB34 was downregulated in tumor tissues, a high level of RAB34 was significantly associated with poor OS and DFS. This could have different explanations. First, we found the downregulated RABs are highly expressed in CMS4 CRC, which are usually diagnosed at the late clinical stage and have the worst survival, suggesting that their upregulation is associated with a more aggressive phenotype. Second, there may be unknown post-transcription modification affecting the expression and functions of these RABs. Supporting this notion, though RAB3A mRNA could be detected, a lack of RAB3A immunostaining in pancreatic ductal adenocarcinoma (PDAC) tissues were reported (Dragomir and Calin, 2018). It is also likely that other interacting proteins of these RABs may also be aberrantly expressed in CRC and have an impact on their functions. Specific experiments need to be carried out in further research.

In summary, our study provides initial data elucidating the possible role of RABs in CRC. RABs might be important markers in predicting immunotherapy response and patient survival. Further studies using clinical samples and different models should be considered to explore their utility and underlying mechanisms as therapeutic targets and clinical biomarkers.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) of The Sixth Affiliated Hospital of Sun Yat-sen University. The patients provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XJ conceived this study. XJ and ZY designed the study. LY and QG acquired and analyzed the experimental data. YL and XF contributed analysis tools. XJ and ZY wrote the article. SY and ZY were of immense help in the preparation of the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (No. 81702330 to ZHY), Guangdong Natural Science Foundation (No. 2021A1515011400 to ZHY).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.828373/full#supplementary-material

REFERENCES

Alcalá, S., Sancho, P., Martinelli, P., Navarro, D., Pedrero, C., Martín-Hijano, L., et al. (2020). ISG15 and ISGylation Is Required for Pancreatic Cancer Stem Cell Mitophagy and Metabolic Plasticity. *Nat. Commun.* 11, 2682. doi:10.1038/s41467-020-16395-2

Alloatti, A., Kotsias, F., Pauwels, A.-M., Carpier, J.-M., Jouve, M., Timmerman, E., et al. (2015). Toll-Like Receptor 4 Engagement on Dendritic Cells Restrains Phago-Lysosome Fusion and Promotes Cross-Presentation of Antigens. *Immunity.* 43, 1087–1100. doi:10.1016/j.immuni.2015.11.006

Anand, S., Khan, M. A., Khushman, M., Dasgupta, S., Singh, S., and Singh, A. P. (2020). Comprehensive Analysis of Expression, Clinicopathological Association and Potential Prognostic Significance of RABs in Pancreatic Cancer. *Int. J. Mol. Sci.* 21, 5580. doi:10.3390/ijms21155580

Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal Cancer. *The Lancet.* 383, 1490–1502. doi:10.1016/s0140-6736(13)61649-9

Cho, K. H., and Lee, H. Y. (2019). Rab25 and RCP in Cancer Progression. *Arch. Pharm. Res.* 42, 101–112. doi:10.1007/s12272-019-01129-w

Crockett, S. D., and Nagtegaal, I. D. (2019). Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. *Gastroenterology.* 157, 949–e4. doi:10.1053/j.gastro.2019.06.041

Dozynkiewicz, M. A., Jamieson, N. B., Macpherson, I., Grindlay, J., van den Berghe, P. V. E., von Thun, A., et al. (2012). Rab25 and CLIC3 Collaborate to Promote Integrin Recycling from Late Endosomes/lysosomes and Drive Cancer Progression. *Developmental cell.* 22, 131–145. doi:10.1016/j.devcel.2011.11.008

Dragomir, M., and Calin, G. A. (2018). Circular RNAs in Cancer - Lessons Learned From microRNAs. *Front. Oncol.* 8, 179. doi:10.3389/fonc.2018.00179

Du, F., Cao, T., Xie, H., Li, T., Sun, L., Liu, H., et al. (2020). KRAS Mutation-Responsive miR-139-5p Inhibits Colorectal Cancer Progression and Is Repressed by Wnt Signaling. *Theranostics.* 10, 7335–7350. doi:10.7150/thno.45971

Du, J., Fu, L., Hao, J., Lin, X., and Dong, Q. (2020). Rab11a Is Overexpressed in Gastric Cancer and Regulates FAK/AKT Signaling. *J. Oncol.* 2020, 3494396. doi:10.1155/2020/3494396

Ganga, A. K., Kennedy, M. C., Oguchi, M. E., Gray, S., Oliver, K. E., Knight, T. A., et al. (2021). Rab34 GTPase Mediates Ciliary Membrane Formation in the Intracellular Ciliogenesis Pathway. *Curr. Biol.* 31, 2895–e7. doi:10.1016/j.cub.2021.04.075

Guerra, F., and Bucci, C. (2019). Role of the RAB7 Protein in Tumor Progression and Cisplatin Chemoresistance. *Cancers (Basel).* 11, 1096. doi:10.3390/cancers11081096

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The Consensus Molecular Subtypes of Colorectal Cancer. *Nat. Med.* 21, 1350–1356. doi:10.1038/nm.3967

Harada, S., and Morlote, D. (2020). Molecular Pathology of Colorectal Cancer. *Adv. Anat. Pathol.* 27, 20–26. doi:10.1097/pap.0000000000000247

He, Y., Gao, M., Tang, H., Cao, Y., Liu, S., and Tao, Y. (2019). Metabolic Intermediates in Tumorigenesis and Progression. *Int. J. Biol. Sci.* 15, 1187–1199. doi:10.7150/ijbs.33496

Homma, Y., Hiragi, S., and Fukuda, M. (2021). Rab Family of Small GTPases: an Updated View on Their Regulation and Functions. *Febs J.* 288, 36–55. doi:10.1111/febs.15453

Hong, K. S., Jeon, E.-Y., Chung, S. S., Kim, K. H., and Lee, R.-A. (2018). Epidermal Growth Factor-Mediated Rab25 Pathway Regulates Integrin β1 Trafficking in colon Cancer. *Cancer Cell Int.* 18, 32. doi:10.1186/s12935-018-0526-y

Jeong, B. Y., Cho, K. H., Jeong, K. J., Park, Y.-Y., Kim, J. M., Rha, S. Y., et al. (2018). Rab25 Augments Cancer Cell Invasiveness Through a β1 Integrin/EGFR/VEGF-A/Snail Signaling Axis and Expression of Fascin. *Exp. Mol. Med.* 50, e435. doi:10.1038/emm.2017.248

Kerk, S. A., Papagiannakopoulos, T., Shah, Y. M., and Lyssiotis, C. A. (2021). Metabolic Networks in Mutant KRAS-Driven Tumours: Tissue Specificities and the Microenvironment. *Nat. Rev. Cancer.* 21, 510–525. doi:10.1038/s41568-021-00375-9

Kumarasamy, V., Vail, P., Nambiar, R., Witkiewicz, A. K., and Knudsen, E. S. (2021). Functional Determinants of Cell Cycle Plasticity and Sensitivity to CDK4/6 Inhibition. *Cancer Res.* 81, 1347–1360. doi:10.1158/0008-5472.can-20-2275

Lal, N., White, B. S., Goussous, G., Pickles, O., Mason, M. J., Beggs, A. D., et al. (2018). KRAS Mutation and Consensus Molecular Subtypes 2 and 3 Are Independently Associated with Reduced Immune Infiltration and Reactivity in Colorectal Cancer. *Clin. Cancer Res.* 24, 224–233. doi:10.1158/1078-0432.ccr-17-1090

Langemeyer, L., Fröhlich, F., and Ungermann, C. (2018). Rab GTPase Function in Endosome and Lysosome Biogenesis. *Trends Cell Biology.* 28, 957–970. doi:10.1016/j.tcb.2018.06.007

Li, Y., Jia, Q., Zhang, Q., and Wan, Y. (2015). Rab25 Upregulation Correlates with the Proliferation, Migration, and Invasion of Renal Cell Carcinoma. *Biochem. biophysical Res. Commun.* 458, 745–750. doi:10.1016/j.bbrc.2015.01.144

Li, Z., Pearlman, A. H., and Hsieh, P. (2016). DNA Mismatch Repair and the DNA Damage Response. *DNA Repair (Amst).* 38, 94–101. doi:10.1016/j.dnarep.2015.11.019

Lucero, M., Suarez, A. E., and Chambers, J. W. (2019). Phosphoregulation on Mitochondria: Integration of Cell and Organelle Responses. *CNS Neurosci. Ther.* 25, 837–858. doi:10.1111/cns.13141

Madeira, F., Park, Y. m., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi:10.1093/nar/gkz268

Mandal, R., Samstein, R. M., Lee, K.-W., Havel, J. J., Wang, H., Krishna, C., et al. (2019). Genetic Diversity of Tumors with Mismatch Repair Deficiency Influences Anti-PD-1 Immunotherapy Response. *Science.* 364, 485–491. doi:10.1126/science.aau0447

Peinado, H., Alečković, M., Lavotshkin, S., Matei, I., Costa-Silva, B., Moreno-Bueno, G., et al. (2012). Melanoma Exosomes Educate Bone Marrow Progenitor Cells Toward a Pro-Metastatic Phenotype through MET. *Nat. Med.* 18, 883–891. doi:10.1038/nm.2753

Pfeffer, S. R. (2017). Rab GTPases: Master Regulators that Establish the Secretory and Endocytic Pathways. *MBoC.* 28, 712–715. doi:10.1091/mbc.e16-10-0737

Recchi, C., and Seabra, M. C. (2012). Novel Functions for Rab GTPases in Multiple Aspects of Tumour Progression. *Biochem. Soc. Trans.* 40, 1398–1403. doi:10.1042/bst20120199

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., and La, K. C., (2018). Oncogenic Signaling Pathways in the Cancer Genome Atlas. *Cell* 173 (2), 321–337.e10. doi:10.1016/j.cell.2018.03.035

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA Cancer J. Clin.*, 71 (1), 7–33. doi:10.3322/caac.21654

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70, 145–164. doi:10.3322/caac.21601

Smeby, J., Sveen, A., Merok, M. A., Danielsen, S. A., Eilertsen, I. A., Guren, M. G., et al. (2018). CMS-dependent Prognostic Impact of KRAS and BRAFV600E Mutations in Primary Colorectal Cancer. *Ann. Oncol.* 29, 1227–1234. doi:10.1093/annonc/mdy085

Sun, L., Xu, X., Chen, Y., Zhou, Y., Tan, R., Qiu, H., et al. (2018). Rab34 Regulates Adhesion, Migration, and Invasion of Breast Cancer Cells. *Oncogene.* 37, 3698–3714. doi:10.1038/s41388-018-0202-7

Supek, F., and Lehner, B. (2015). Differential DNA Mismatch Repair Underlies Mutation Rate Variation across the Human Genome. *Nature.* 521, 81–84. doi:10.1038/nature14173

Temel, S. G., Giray, A., Karakas, B., Gul, O., Kozanoglu, I., Celik, H., et al. (2020). RAB25 Confers Resistance to Chemotherapy by Altering Mitochondrial Apoptosis Signaling in Ovarian Cancer Cells. *Apoptosis.* 25, 799–816. doi:10.1007/s10495-020-01635-z

Wang, M., Wang, W., Ding, J., Wang, J., and Zhang, J. (2020). Downregulation of Rab17 Promotes Cell Proliferation and Invasion in Non-Small Cell Lung Cancer through STAT3/HIF-1α/VEGF Signaling. *Thorac. Cancer.* 11, 379–388. doi:10.1111/1759-7714.13278

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2--a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics.* 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Win, S., Than, T. A., and Kaplowitz, N. (2018). The Regulation of JNK Signaling Pathways in Cell Death through the Interplay with Mitochondrial SAB and Upstream Post-Translational Effects. *Int. J. Mol. Sci.* 19, 3657. doi:10.3390/ijms19113657

Wong, C. C., Xu, J., Bian, X., Wu, J.-L., Kang, W., Qian, Y., et al. (2020). Colorectal Cancer Cells with Mutant KRAS, SLC25A22-Mediated Glutaminolysis Reduces DNA Demethylation to Increase WNT Signaling, Stemness, and Drug Resistance. *Gastroenterology.* 159, 2163. doi:10.1053/j.gastro.2020.08.016

Yuan, J., Shan, Y., Chen, X., Tang, W., Luo, K., Ni, J., et al. (2005). Identification and Characterization of RHEBL1, a Novel Member of Ras Family, Which Activates Transcriptional Activities of NF-Kappa B. *Mol. Biol. Rep.* 32, 205–214. doi:10.1007/s11033-005-0984-x

Zhang, X., Wang, S., Lin, G., and Wang, D. (2020). Down-Regulation of Circ-PTN Suppresses Cell Proliferation, Invasion and Glycolysis in Glioma by Regulating miR-432-5p/RAB10 axis. *Neurosci. Lett.* 735, 135153. doi:10.1016/j.neulet.2020.135153

Zhen, Y., and Stenmark, H. (2015). Cellular Functions of Rab GTPases at a Glance. *J. Cell Sci.* 128, 3171–3176. doi:10.1242/jcs.166074

Zheng, L. Q., Chi, S. M., and Li, C. X. (2017). Rab23's Genetic Structure, Function and Related Diseases: a Review. *Biosci. Rep.* 37, BSR20160410. doi:10.1042/BSR20160410

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Jiang, Yang, Gao, Liu, Feng, Ye and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fgene.2021.804952

Hypoxia-Induced miR-378a-3p Inhibits Osteosarcoma Invasion and Epithelial-to-Mesenchymal Transition *via BYSL* Regulation

Junlei Zhang[1], *Haijun Tang*[2], *Xiaohong Jiang*[2], *Nenggan Huang*[1] *and Qingjun Wei*[1]*

[1]Department of Orthopedics, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, [2]Department of Orthopedics, Affiliated Minzu Hospital of Guangxi Medical University, Nanning, China

The bystin-like (*BYSL*) gene is expressed in a wide range of eukaryotes and is closely associated with tumor progression. However, its function and mechanism in osteosarcoma remain unclear. Herein, the protein expression and clinical role of *BYSL* in human osteosarcoma tissues were assessed. High expression of *BYSL* was positively related to the metastasis status and poor patient prognosis. Mechanistically, upregulation of *BYSL* enhanced Nrf2 expression under hypoxia in osteosarcoma cells. MicroRNAs are important epigenetic regulators of osteosarcoma development. Noteworthy, bioinformatics analysis, dual-luciferase reporter and rescue assays showed that miR-378a-3p inhibited *BYSL* expression by binding to its 3′-untranslated region. Analysis of miR-378a-3p function under hypoxia and normoxia showed that its upregulation suppressed osteosarcoma cells invasion and inhibited epithelial-to-mesenchymal transition by suppressing *BYSL*. Collectively, the results show that the miR-378a-3p/BYSL may associate with metastasis risk in osteosarcoma.

Keywords: *BYSL*, osteosarcoma, hypoxia, Nrf2, MiR-378a-3p, epithelial-to-mesenchymal transition

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Javier Gaytan, IMSS, Mexico
Liang-Ting Lin, Hong Kong Polytechnic University, Hong Kong SAR, China

***Correspondence:**
Qingjun Wei
weiqingjungxnn@163.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *29 October 2021*
Accepted: *27 December 2021*
Published: *28 January 2022*

Citation:
Zhang J, Tang H, Jiang X, Huang N and Wei Q (2022) Hypoxia-Induced miR-378a-3p Inhibits Osteosarcoma Invasion and Epithelial-to-Mesenchymal Transition via BYSL Regulation. Front. Genet. 12:804952. doi: 10.3389/fgene.2021.804952

INTRODUCTION

Osteosarcoma is a common malignant bone tumor, with highly invasive and systemic metastasis, that usually occurs in children and young adults (Raymond and Jaffe, 2009; Sevelda et al., 2015). Owing to new adjuvant chemotherapy and improved surgical treatment, the overall survival rate of five-years for osteosarcoma patients has improved up to 70%, while the survival rate of patients with metastasis is still only 10–30% (Ando et al., 2013; Harrison et al., 2018). Several therapeutic approaches have been developed for osteosarcoma in recent years (Li et al., 2021), but their therapeutic effects remain unsatisfactory. In-depth investigations on the molecular mechanisms underlying osteosarcoma metastasis are meaningful for the research of novel treatment approaches. Epithelial-to-mesenchymal transition (EMT) promotes the transformation of early tumors into aggressive malignant tumors, especially during tumor metastasis and invasion, which is an important event in the malignant transformation of cancer cells. EMT is characterized by a complex dynamic change, with a concomitant decline of epithelial cell markers (including β-catenin and E-cadherin) and increased mesenchymal markers (including vimentin and N-cadherin) (Arias, 2001), which has been shown to contribute to cancer metastasis and invasion in osteosarcoma (Buddingh et al., 2011; Wang et al., 2017). Therefore, inhibition of EMT is a suitable therapeutic strategy for preventing osteosarcoma metastasis.

Low oxygen tension state in tissues, known as hypoxia, has become an important factor in tumor pathophysiology. It has been reported that hypoxia triggers EMT in several types of cancers, including oral, nasopharyngeal, and gastric carcinoma (Joseph et al., 2018; Xing et al., 2021). A hypoxic environment is also intimately linked to the invasion and EMT of osteosarcoma (Shi et al., 2020). Bystin-like (*BYSL*) is a protein containing 306 amino acids encoded by the BYSL gene located on chromosome 6p21.1. It is an essential protein for embryo survival, has been identified as a sensitive biomarker for reactive astrocytes induced by ischemia/reperfusion (Aoki et al., 2006; Olczak et al., 2018). It was reported that BYSL plays a role in the biogenesis of the 40S ribosome and cell proliferation by analyzing the expression of BYSL in mouse embryos (Adachi et al., 2007). Generally, BYSL was considered to be involved in cell adhesion and growth, especially in metamorphosis (Aoki et al., 2006). Previous observations demonstrated that *BYSL* is crucial for hepatocellular carcinoma development (Wang et al., 2009). Furthermore, *BYSL* enhances glioblastoma cell migration, invasion, and EMT by controlling the GSK-3β/β-catenin pathway (Sha et al., 2020). However, the clinical role and molecular mechanism of *BYSL* in osteosarcoma metastasis remain unclear.

MicroRNAs are non-coding RNAs (17–23 nucleotides long) that, in combination with the 3′-untranslated region (3′-UTR) of the target mRNA, take part in various biological and pathological processes. MicroRNAs induce translational inhibition and mRNA degradation (Bartel, 2004), playing a critical role in cancer biology due to their involvement in the processes of proliferation, metastasis, and apoptosis (Liu et al., 2020). Therefore, miRNA profiling can be used as a biomarker for cancer prognosis and diagnosis (Lee and Dutta, 2009). Several studies have documented a strong link between miR-378a-3p function and cancer pathogenesis (Pan et al., 2019; Zheng et al., 2020). For example, overexpression of miR-378a-3p was reported to increase ovarian cancer cell development (Chanjiao et al., 2021). Moreover, breast cancer patients with low expression of miR-378a-3p in cancer tissues who were under tamoxifen treatment were found to have a weak prognosis (Ikeda et al., 2015). However, the essential significance of miR-378a-3p in osteosarcoma progression has not yet been uncovered. In the present study, the roles of *BYSL* and miR-378a-3p in osteosarcoma cell lines were examined, and their functional relationship was assessed.

MATERIALS AND METHODS

Bioinformatics Analysis

The Gene Expression Omnibus (GEO) dataset GSE126209, comprising 12 osteosarcoma tissues and 11 paratumor tissues, was used to examine the expression of *BYSL* in osteosarcoma. We then performed bioinformatics analyses using The Cancer Genome Atlas (TCGA) sarcoma dataset to evaluate the prognostic value of *BYSL*.

Patients and Specimens

Tissue specimens from 51 patients with conventional osteosarcoma between March 2011 and December 2019, were retrospectively analyzed after formalin fixation and paraffin embedding. Clinical data were collected from the medical records of each patient. Follow-up procedures were conducted for all patients at least once every 2 years, which included plain film, computed tomography, and magnetic resonance imaging. Further, paraffin sections were collected for this study. Based on the pathological diagnostic criteria, the pathological diagnoses were made by two pathologists. All patients with osteosarcoma agreed to participate in the study and provided informed consent. This experimental research was approved by the Institutional Review Committee of the First Affiliated Hospital of Guangxi Medical University (Nanning, China) (approval number: 2021 KY-E-041).

Immunohistochemical Assay

Tissue samples were dewaxed by dipping in dewaxing solution and 95% ethanol, and then heated in a microwave oven to achieve antigen recovery. Following blocking with 5% goat serum, an anti-BYSL polyclonal antibody (1:300, Novus Biologicals, Littleton, CO, United States) was used for immunohistochemical analysis. Tissue specimens were incubated with BYSL antibody at 4°C for 8 h and a secondary antibody for 60 min at room temperature. Immunoreactions were visualized using diaminobenzidine for 5 min. At least 100 tumor cells were detected, using light microscopy at a magnification of 200× and 400×, in five tissue regions where the anti-BYSL antibody showed the strongest immune response. According to the level of *BYSL* expression, patients were divided into two groups—those with high expression and those with low expression. *BYSL* positivity was assessed independently by two pathologists. Immunohistochemistry results were evaluated using a scoring system as previously described (Gorlick et al., 2014). The final score, which was the product of *BYSL* positivity rate and staining intensity, was classified as low *BYSL* (0–4 points) or high *BYSL* expression (>4 points).

Cell Culture and Transfection

Human osteosarcoma cell lines including MG63 and Saos-2 cells, obtained from the National Collection of Authenticated Cultures (Shanghai, China), were routinely cultured in Dulbecco's Modified Eagle medium with 10% fetal bovine serum (FBS; Gibco, Waltham, MA, United States) plus 1% penicillin-streptomycin and McCoy's 5A medium (Gibco) with 15% FBS and 1% penicillin-streptomycin, respectively. For hypoxic culture, the cells were exposed to 1% oxygen tension (1% O_2) in a hypoxia incubator chamber. Then, miR-378a-3p-mimic, miR-378a-3p-inhibitor, *BYSL* overexpression vector (oe-BYSL), siRNA against *BYSL* (si-*BYSL*), siRNA against *Nrf2* (si-*Nrf2*), and empty vector or relevant negative control (GenePharma, Shanghai, China) were transfected into the MG63 and Saos-2 cells. Transfection experiments in our study were performed using Effectene Transfection Reagent (Qiagen, Hilden, Germany).

Luciferase Reporter Assay

The target microRNA of *BYSL* was confirmed by luciferase reporter assay, wherein the wild-type (WT, 5′-caucUGUGG

CUCCCAGUCCAGg-3′) or mutant (MUT, 5′-caucUGUGG CUCCCAGAATTGg-3′) 3′-UTR of BYSL was integrated into the pmirGLO vector (Promega, Madison, WI, United States). Cells that were cotransfected with WT or MUT *BYSL* and miR-378a-3p mimics were collected after 2 days to measure luciferase activity via a dual-luciferase reporter assay system (Promega).

Cell Viability Assay

To assess cell viability of the experimental and control groups, we use a Cell Counting Kit-8 (CCK-8/WST-8) assay (Solarbio, Beijing, China) to detect the optical density value. Briefly, the seeding of MG63 cells was conducted in a 96-well plate and incubated for 24, 48, and 72 h. Subsequently, CCK-8 reagent was put, and the 20-min incubation of reaction mixture was performed at room temperature. The relative viability of the cells was measured at 450 nm.

Flow Cytometry Analysis

The analysis of cell apoptosis was performed *via* flow cytometry using an apoptosis detection kit (Multi Sciences, Hangzhou, China) on the basis of the manufacturer's specification. Cells in early and late periods of apoptosis were screened and evaluated using a Beckman Coulter CytoFLEX Flow Cytometer (Beckman Coulter, Brea, CA, United States), and FlowJo v10 software (Becton, Dickinson and Company, Franklin Lakes, NJ, United States) was adopted to analyze outcome information.

RNA Analysis

Total RNA was extracted from cultured cells using Rneasy Mini kits (Qiagen) based on the manufacturer's protocol. Next, cDNA was synthesized using the PrimeScript RT Master Mix cDNA synthesis system (Takara, Kusatsu, Japan), and SYBR Green (Bio-Rad, Hercules, CA, United States) was used for quantification. Expression of the gene of interest was normalized to *U6* or *GAPDH* levels. **Supplementary Table S1** shows the primers applied for quantitative reverse transcription polymerase chain reaction (qRT-PCR).

Western Blot

Proteins were extracted from MG63 or Saos-2 cells using lysis buffer (Solarbio), and the process of transferring and incubating were performed on the basis of the manufacturer's protocols. Protein concentration was quantified using a BCA kit (Takara Bio, Inc.). The proteins were separated via SDS-PAGE on 8% gel and then electroblotted onto a PVDF membrane. Following blocking with 5% blocking reagent at room temperature for 1 h, the membranes were incubated with the primary antibodies in 5% BSA overnight at 4°C. Subsequently, the membranes were incubated with secondary antibodies at room temperature for 1 h. Proteins were visualized using ECL reagent. The primary antibodies use were the following: anti-BYSL (1:1,000, NBP1-89501; Novus Biologicals), anti-Bcl-2 (1:1,000, ab196495; Abcam, Cambridge, United Kingdom), anti-Histone H3 (1:1,000, ab18521; Abcam), anti-Bax (1:2000, ab32503; Abcam), anti-E-cadherin (1:5,000, 20874-1-AP; Proteintech, Rosemont, IL, United States), anti-N-cadherin (1:1,000, ab245117; Abcam), anti-Nrf2 (1:1,000, ab137550; Abcam), anti-β-actin (1:1,000, AC026, Abclonal), and ani-vimentin (1:2,000, 10366-1-AP; Proteintech).

Cell Motility Analysis

To test cell migration ability, we performed wound healing assays to evaluate the migration area of cells. Briefly, after the cell density reached 100%, the cells were scratched with a fine tip and then washed with a serum-free medium to remove the unattached cells. The wound healing rate was measured by photography using an inverted microscope. For cell invasion ability, we performed transwell invasion assays to calculate invasive cells rate using transwell chambers with Matrigel (Thermo Fisher Scientific, Waltham, MA, United States).

Osteosarcoma Subcutaneous Tumor Model in Nude Mice

MG63 cells were transfected with lentivirus stably overexpressing miR-378a-3p (miR-378a-3p-OE), and a stable cell line was established using puromycin selection. Ten male BALB/c nude mice aged 4–6 weeks were housed in the Animal Experimental Center of Guangxi Medical University. The Animal Care and Ethics Committee of the Guangxi Medical University (Number: 202108002) approved all protocols. To generate an osteosarcoma tumor *via* miR-378a-3p overexpression, MG63 cells were injected subcutaneously into the right armpit region of the mice. The animals were euthanized 25 days after cell implantation. Tumors were excised, weighed, and photographed. Tumor size in each mouse was evaluated applying the formula: V (mm^3) = 1/6 π × length (mm) × width^2 (mm^2).

Statistical Analysis

Data were analyzed using SPSS 23.0 software (IBM Corp., Armonk, NY, United States) or R software (Version 3.6.3). Univariate and multivariate analyses using the Cox regression model were performed to identify independent risk factors for osteosarcoma that influenced the total survival rate. Two or more groups were compared using Student's t-test or one-way analysis of variance followed by Tukey's post-hoc test, respectively. Differences were regarded as significant at $p < 0.05$.

RESULTS

BYSL Expression is Associated With Survival Using Bioinformatics Analysis

In this study, we used the GSE126209 dataset from the GEO database to evaluate *BYSL* level using the *limma* package in R. Osteosarcoma tissues were found to have higher *BYSL* expression than adjacent paratumor tissues (**Figure 1A**). Next, the prognostic value of *BYSL* in TCGA database was explored. Since there was no clinical data on osteosarcoma in TCGA, the sarcoma dataset was

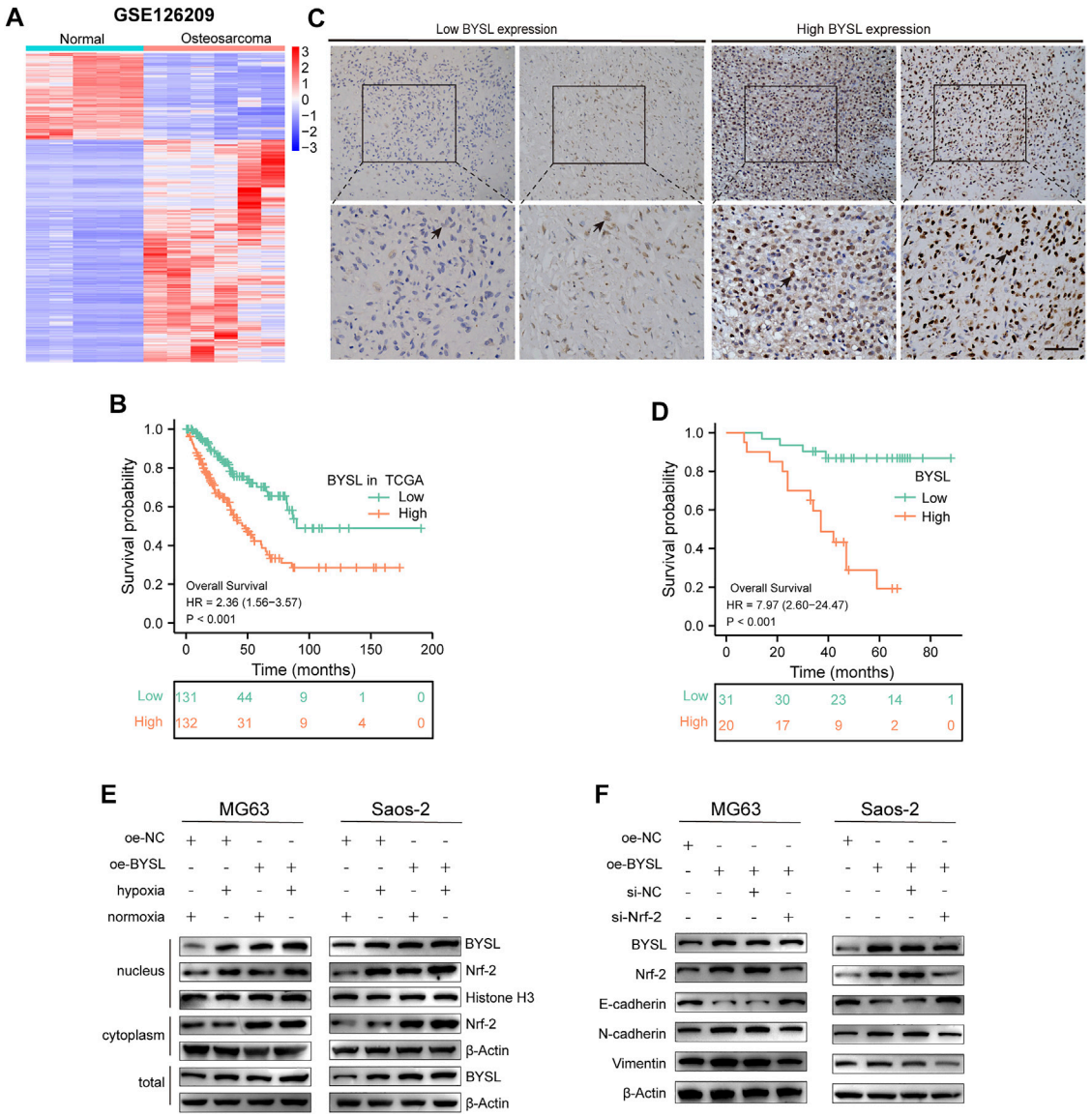


FIGURE 1 | *BYSL* is related to the poor prognosis of patients with osteosarcoma. **(A)** Cluster heatmap of differentially expressed genes in GSE126209. **(B)** Overall survival was compared between osteosarcoma patients with high and low *BYSL* expression in TCGA database. **(C)** Immunocytochemical staining of *BYSL* in osteosarcoma tissues ($n = 51$, scale bar = 50 μm). **(D)** Overall survival was compared between osteosarcoma patients with high and low *BYSL* expression in an in-house cohort. **(E)** MG63 and Saos-2 cells were transfected with the control plasmid (oe-NC) or BYSL overexpression plasmid (oe-BYSL), and then cultured under hypoxic or normoxic conditions. After nuclear and cytosolic separation, protein levels of Nrf2, BYSL, Histone H3, and β-Actin were measured by western blot. **(F)** MG63 and Saos-2 cells were transfected with control plasmid (oe-NC), BYSL overexpression plasmid (oe-BYSL), si-control (si-NC), or si-Nrf2, and then cultured under hypoxic condtions. The protein levels of Nrf2, E-cadherin, N-cadheirn, Vimentin, and β-Actin were measured by western blot.

selected. Kaplan-Meier analysis showed a more adverse prognosis in patients with higher *BYSL* expression (**Figure 1B**).

High *BYSL* Expression is Correlated With Poor Prognosis of Patients With Osteosarcoma

To further identify the role of *BYSL* in osteosarcoma, we first examined its expression in tissue samples of 51 patients with osteosarcoma through histopathological analysis (**Figure 1C**). In 51 cases of osteosarcoma tissues, BYSL was highly expressed in 20 cases and lowly expressed in 31 cases (**Table 1**). Next, the association between clinicopathological characteristics of the osteosarcoma patients and tissue *BYSL* expression was analyzed. Overall, high *BSYL* expression was strongly correlated with tumor-node-metastasis (TNM) stage, relapse, and metastasis (**Table 1**). In addition, Kaplan-Meier analysis showed a longer survival period in patients with low *BYSL* expression than in those with high *BYSL* expression ($p < 0.01$; **Figure 1D**). Moreover, multivariate Cox regression analysis indicated that *BYSL* and metastasis are

TABLE 1 | Associations between clinicopathological characteristics and BYSL expression in patients with osteosarcoma.

Variable	Number of patients	BYSL expression		*p* Value
		High	Low	
Age				
>16 years	24 (47.1%)	11 (55.0%)	13 (41.9%)	0.532
≤16 years	27 (52.9%)	9 (45.0%)	18 (58.1%)	—
Gender				
Female	21 (41.2%)	5 (25.0%)	16 (51.6%)	0.111
Male	30 (58.8%)	15 (75.0%)	15 (48.4%)	—
Relapse				
Yes	15 (29.4%)	11 (55.0%)	4 (12.9%)	**0.004**
No	36 (70.6%)	9 (45.0%)	27 (87.1%)	—
Metastasis				
Yes	21 (41.2%)	13 (65.0%)	8 (25.8%)	**0.013**
No	30 (58.8%)	7 (35.0%)	23 (74.2%)	—
TNM stage				
I	29 (56.9%)	7 (35.0%)	22 (71.0%)	**0.025**
II/III	22 (43.1%)	13 (65.0%)	9 (29.0%)	—
Location				
Femur/Tibia	42 (82.4%)	15 (75.0%)	27 (87.1%)	0.465
Else	9 (17.6%)	5 (25.0%)	4 (12.9%)	—
Tumor size				
≤6 cm	24 (47.1%)	7 (35.0%)	17 (54.8%)	0.272
>6 cm	27 (52.9%)	13 (65.0%)	14 (45.2%)	—

Note: Bold values indicate p < 0.05.

independent risk factors for the total survival rate of patients with osteosarcoma (**Table 2**). Taken together, these data suggest that *BYSL* has considerable clinical significance in the prognosis and metastasis of patients with osteosarcoma.

BYSL Overexpression Enhances Nrf2 Signaling

Next, western blot was conducted to investigate the pathway of how BYSL modulates EMT in osteosarcoma cells. Hypoxia activated the Nrf2 signaling, which has been shown to play an important role in regulation of EMT in cancer cells. Therefore, we investigated whether BYSL expression affects Nrf2 signaling. The results showed that hypoxia enhanced Nrf2 localization within the nucleus, but did not dramatically affect Nrf2 expression in the cytoplasm, indicating that hypoxia activated Nrf2 signaling. It was also found that hypoxia increased the nuclear and total expression of BYSL in osteosarcoma cells. Furthermore, *BYSL* overexpression promoted Nrf2 expression in both nuclear and extranuclear under hypoxic and normoxic conditions (**Figure 1E**; **Supplementary Figure S1**). As shown in **Figure 1F**, BYSL overexpression decreased epithelial marker (E-cadherin) and increased mesenchymal marker (N-cadherin and Vimentin) of osteosarcoma cells in hypoxia. However, *Nrf2* knockdown by siRNA reduced the effect of *BYSL* on EMT markers (**Figure 1F**; **Supplementary Figure S2**). The downregulation of Nrf2 did not affect the expression of BYSL, indicating BYSL was possibly upstream of Nrf2. Together, these results indicate that *BYSL* regulates the EMT of osteosarcoma cells via Nrf2 signaling under hypoxic conditions.

miR-378a-3p Directly Targets BYSL

Next, we used miRTarBase and ENCORI databases to examine the potential *BYSL*-targeting microRNAs and predict the probable functional binding site. Among the candidate microRNAs, we were particularly interested in miR-378a-3p owing to its potential

TABLE 2 | Univariable and multivariable Cox regression analysis of clinical characteristics and BYSL in osteosarcoma.

Variable	Univariate analysis			Multivariate analysis		
	HR	95%CI	*p* value	HR	95%CI	*p* value
BYSL	7.97	2.6–24.47	**0.00029**	6.25	1.17–33.36	**0.032**
High	—	—	—	—	—	—
Low	—	—	—	—	—	—
Age	1.81	0.7–4.69	0.2197	—	—	—
>16 years	—	—	—	—	—	—
≤16 years	—	—	—	—	—	—
Gender	0.75	0.3–1.91	0.54794	—	—	—
Female	—	—	—	—	—	—
Male	—	—	—	—	—	—
Relapse	6.02	2.25–16.13	**0.00036**	2.17	0.48–9.93	0.316
Yes	—	—	—	—	—	—
No	—	—	—	—	—	—
Metastasis	17.08	3.9–74.76	**0.00016**	30.29	2.39–383.43	**0.008**
Yes	—	—	—	—	—	—
No	—	—	—	—	—	—
TNM stage	0.11	0.03–0.39	**0.00056**	3.49	0.41–29.43	0.25
I	—	—	—	—	—	—
II/III	—	—	—	—	—	—
Location	0.5	0.18–1.43	0.19659	—	—	—
Femur/Tibia	—	—	—	—	—	—
Else	—	—	—	—	—	—
Tumor size	0.37	0.13–1.05	0.06074	—	—	—
>6 cm	—	—	—	—	—	—
≤6 cm	—	—	—	—	—	—

Note: Bold values indicate p < 0.05.

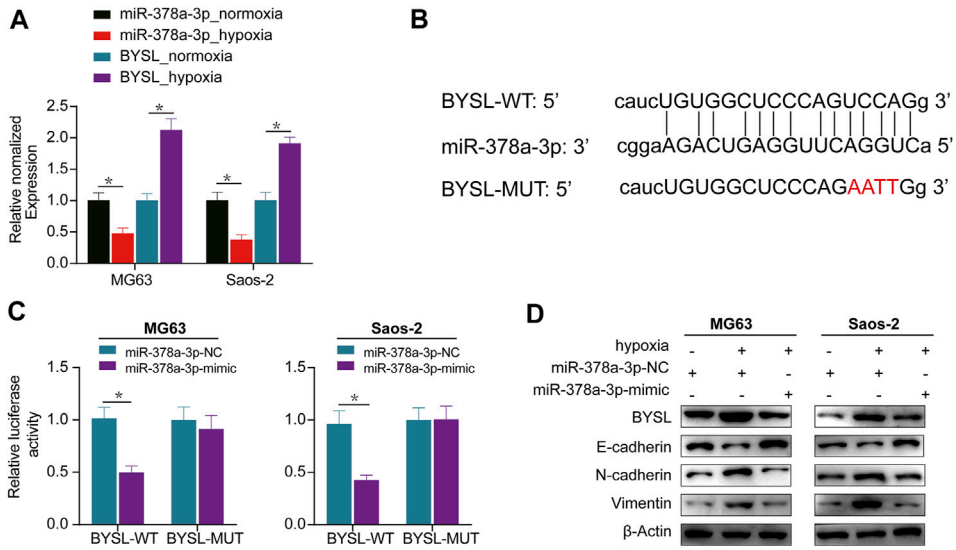


FIGURE 2 | *BYSL* is a direct target of miR-378a-3p. **(A)** MG63 and Saos-2 cells were cultured under hypoxic or normoxic conditions. The RNA levels of miR-378a-3p and BYSL were measured by RT-qPCR. **(B)** The 3′-untranslated region (UTR) of *BYSL* harbor potential miR-378a-3p binding sites. **(C)** The luciferase activity displayed by the luciferase reporter constructs which contained wild-type (WT) or mutant (MUT) 3′-UTR of *BYSL* were co-transfected with miR-378a-3p mimic into MG63 and Saos-2 cells. **(D)** MG63 and Saos-2 cells were transfected with control-mimic (miR-378a-3p-NC) or miR-378a-3p-mimic, and then cultured under hypoxic or normoxic conditions. The protein levels of BYSL, E-cadherin, N-cadherin, and Vimentin were measured by western blot. The data are presented as the mean ± SD. *$p < 0.05$.

tumor suppressing effect in cancer development. Furthermore, qRT-PCR data showed that hypoxia increased *BYSL* expression and reduced that of miR-378a-3p, indicating *BYSL* correlated inversely with miR-378a-3p (**Figure 2A**). Based on these results, we tested whether *BYSL* could be a direct target of miR-378a-3p using a luciferase reporter assay. Cells were cotransfected with luciferase reporters containing *BYSL*-UTR-WT or *BYSL*-UTR-MUT along with miR-378a-3p-mimic. Transfection of miR-378a-3p-mimic markedly inhibited the luciferase reporter activity of WT but not of MUT (**Figures 2B,C**). Taken together, these results indicate the direct binding of miR-378a-3p to the 3′-UTR of *BYSL*. Next, we compared expression of BYSL and EMT markers by western blot under normoxia and hypoxia. Reduced expression of E-cadherin, and increased expressions of BYSL, vimentin, and N-cadherin were observed under hypoxia. Interestingly, miR-378a-3p overexpression rescued the effects of hypoxia on osteosarcoma cells (**Figure 2D**, **Supplementary Figure S3**), indicating that miR-378a-3p decreased hypoxia-induced BYSL expression to inhibit EMT.

miR-378a-3p Knockdown Promotes EMT and Invasion by Elevating *BYSL* Expression Under Normoxia

In osteosarcoma cells, the expression of miR-378a-3p was downregulated after exposure to hypoxia compared to normoxia. Thus, we examined the role of miR-378a-3p in EMT using knockdown or overexpression assays under normoxia or hypoxia, respectively.

qRT-PCR was performed to evaluate the transfection efficiency of miR-378a-3p, confirming that miR-378a-3p expression was significantly decreased in the knockdown group (**Figure 3A**). Under normoxic conditions, western blot analysis showed that miR-378a-3p knockdown resulted in reduced Bax and increased Bcl-2 expression, whereas *BYSL* silencing reversed this effect, suggesting that miR-378a-3p knockdown suppressed apoptosis (**Figures 3B,C**; **Supplementary Figure S4**). This outcome was further confirmed by flow cytometry (**Figure 3D**). Furthermore, miR-378a-3p knockdown promoted N-cadherin and vimentin expression and declined E-cadherin levels, whereas *BYSL* silencing reversed these effects (**Figures 3B,C**; **Supplementary Figure S4**). It was also observed that transfection of miR-378a-3p inhibitor increased BYSL and Nrf2 protein levels, whereas knockdown of BYSL attenuated the effect of miR-378a-3p on Nrf2 expression. We then performed matrigel invasion and wound healing assays to investigate whether miR-378a-3p could regulate the invasive and migratory abilities of osteosarcoma cells. Under normoxic conditions, miR-378a-3p knockdown greatly promoted the invasive (**Figure 3E**) and migratory potential (**Figure 3F**) of human osteosarcoma cells, and knockdown of *BYSL* rescued the miR-378a-3p-induced effects on osteosarcoma cells.

miR-378a-3p Overexpression Inhibits EMT and Invasion by Suppressing *BYSL* Expression Under Hypoxia

qRT-PCR was used to confirm that miR-378a-3p was significantly overexpressed in the overexpression group (**Figure 4A**). Western blot (**Figures 4B,C**; **Supplementary Figure S5**) and apoptosis analyses (**Figure 4D**) demonstrated that miR-378a-3p overexpression increased apoptosis in osteosarcoma cells under hypoxic conditions, which was

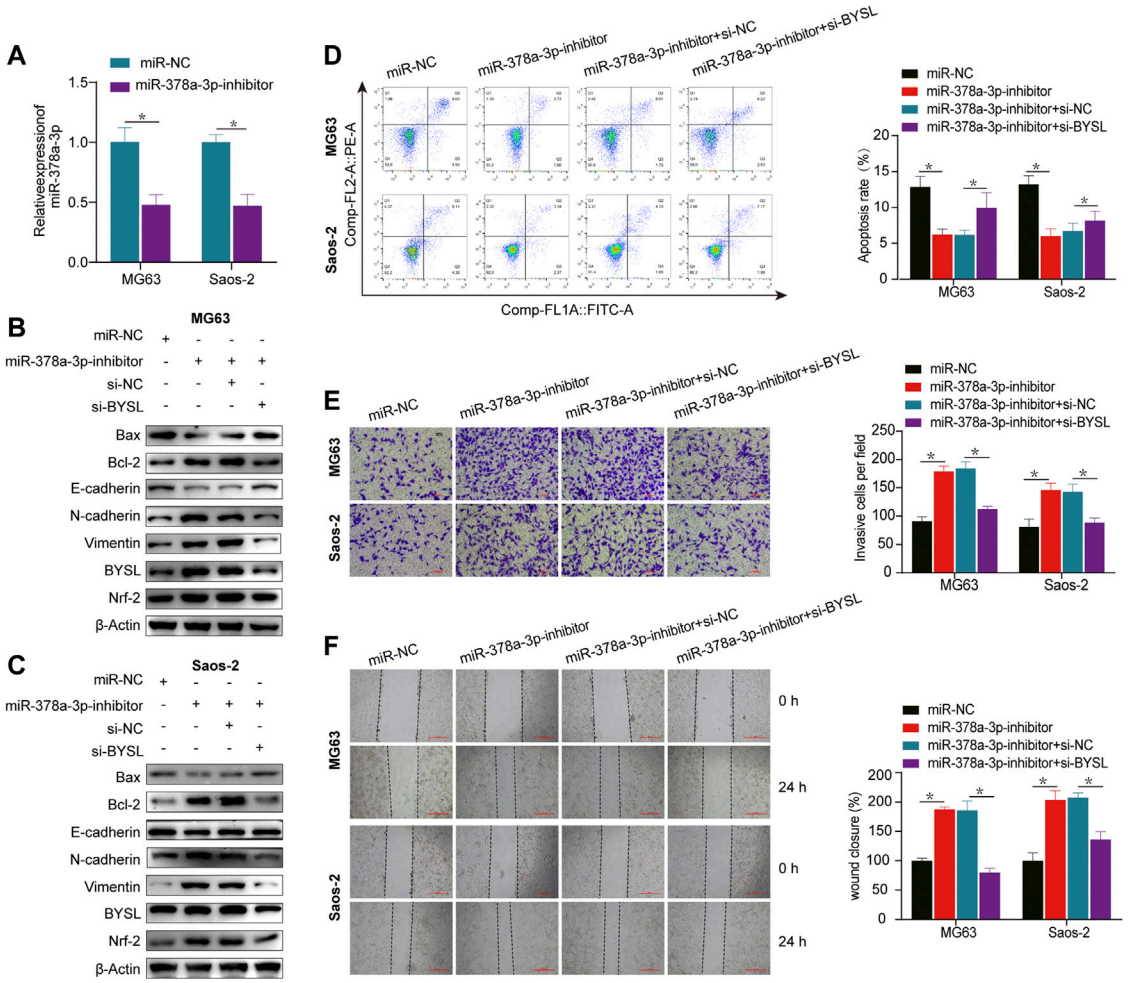


FIGURE 3 | *BYSL* knockdown partially abolishes miR-378a-3p-mediated osteosarcoma cell epithelial-to-mesenchymal transition (EMT), invasion, migration, and apoptosis under normoxia. **(A)** MG63 and Saos-2 cells were transfected with control-inhibitor (miR-NC) or miR-378a-3p-inhibitor, and then cultured under normoxic conditions. The RNA level of miR-378a-3p was measured by RT-qPCR. **(B,C)** MG63 and Saos-2 cells were transfected with control-inhibitor (miR-NC), miR-378a-3p-inhibitor, si-control (si-NC), or si-BYSL, and then cultured under normoxic conditions. The protein levels of Bax, Bcl-2, E-cadherin, N-cadherin, vimentin, BYSL and Nrf2 were measured by western blot. **(D)** MG63 and Saos-2 cells were transfected with control-inhibitor (miR-NC), miR-378a-3p-inhibitor, si-control (si-NC), or si-BYSL, and then cultured under normoxic conditions. Cell apoptosis was measured by flow cytometry. **(E)** MG63 and Saos-2 cells were transfected with control-inhibitor (miR-NC), miR-378a-3p-inhibitor, si-control (si-NC), or si-BYSL, and then cultured under normoxic conditions. Cell invasion was measured by matrigel invasion assay. Scale bar = 100 μm. **(F)** MG63 and Saos-2 cells were transfected with control-inhibitor (miR-NC), miR-378a-3p-inhibitor, si-control (si-NC), or si-BYSL, and then cultured under normoxic conditions. Cell migration was measured by scratch wound healing assay. Scale bar = 500 μm. The data are presented as the mean ± SD. *$p < 0.05$.

inhibited by *BYSL* overexpression. Moreover, the protein levels of N-cadherin and vimentin were decreased, whereas E-cadherin was increased in miR-378a-3p-overexpressing cells. This phenotype can be reversed by *BYSL* overexpression (**Figures 4B,C**; **Supplementary Figure S5**). Osteosarcoma cells transfected with miR-378a-3p mimic had decreased BYSL and Nrf2 expression. However, overexpression of BYSL lead to an increase in Nrf2 expression, which was consistent with the above results (**Figure 1E**). The results of transwell assays (**Figure 4E**) and wound healing (**Figure 4F**) further demonstrated that miR-378a-3p overexpression inhibited osteosarcoma cell migration and invasion in hypoxia. However, upregulation of BYSL markedly attenuated the effect of miR-378a-3p on migration and invasion. In summary, these results indicate that miR-378a-3p regulates hypoxia-induced EMT and invasion by modulating *BYSL* expression.

miR-378a-3p Overexpression Inhibits Osteosarcoma Cell Proliferation *In Vivo* and *In Vitro*

Subsequently, the *in vivo* experiments were conducted to further verify the function of miR-378a-3p in osteosarcoma. CCK-8 cell proliferation assay showed that in osteosarcoma cells, miR-378a-3p overexpression potently decreased CCK-8 optical density, indicating that miR-378a-3p suppressed osteosarcoma cell proliferation (**Figure 5A**). To verify the specificity of miR-378a-3p function in osteosarcoma tumors, an ectopic (subcutaneous) osteosarcoma tumor xenograft model was constructed

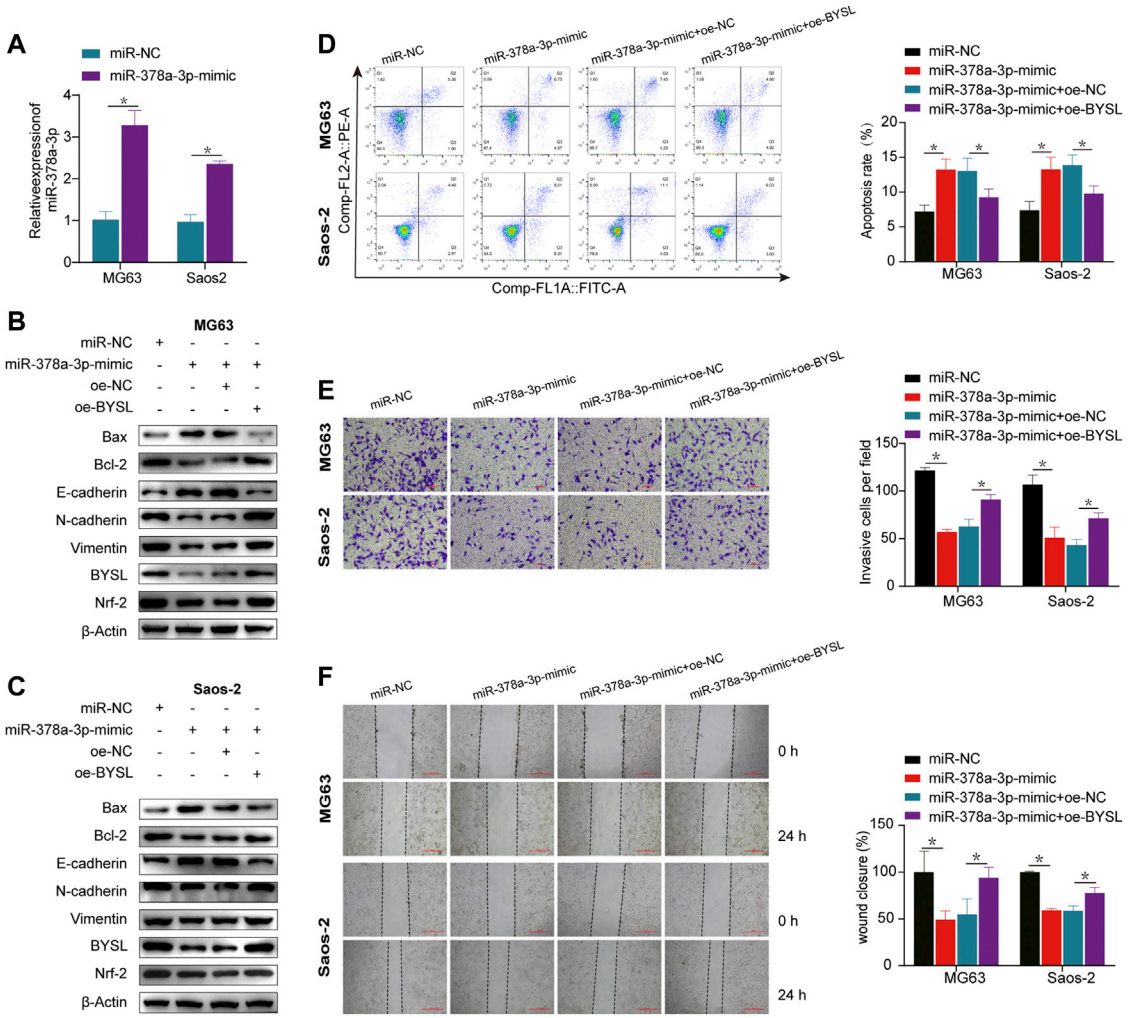


FIGURE 4 | *BYSL* overexpression rescues the effect of miR-378a-3p overexpression on osteosarcoma cells under hypoxia. **(A)** MG63 and Saos-2 cells were transfected with control-mimic (miR-NC) or miR-378a-3p-mimic, and then cultured under hypoxic conditions. The RNA level of miR-378a-3p was measured by RT-qPCR. **(B,C)** MG63 and Saos-2 cells were transfected with control-mimic (miR-NC), miR-378a-3p-mimic, control plasmid (oe-NC), or BYSL overexpression plasmid (oe-BYSL), and then cultured under hypoxic conditions. The protein levels of Bax, Bcl-2, E-cadherin, N-cadherin, vimentin, BYSL and Nrf2 were measured by western blot. **(D)** MG63 and Saos-2 cells were transfected with control-mimic (miR-NC), miR-378a-3p-mimic, control plasmid (oe-NC), or BYSL overexpression plasmid (oe-BYSL), and then cultured under hypoxic conditions. Cell apoptosis was measured by flow cytometry. **(E)** MG63 and Saos-2 cells were transfected with control-mimic (miR-NC), miR-378a-3p-mimic, control plasmid (oe-NC), or BYSL overexpression plasmid (oe-BYSL), and then cultured under hypoxic conditions. Cell invasion was measured by matrigel invasion assay. Scale bar = 100 μm. **(F)** MG63 and Saos-2 cells were transfected with control-mimic (miR-NC), miR-378a-3p-mimic, control plasmid (oe-NC), or BYSL overexpression plasmid (oe-BYSL), and then cultured under hypoxic conditions. Cell migration was measured by scratch wound healing assay. Scale bar = 500 μm. The data are presented as the mean ± SD. *$p < 0.05$.

(**Figure 5B**). Tumor size (**Figure 5C**) and weight (**Figure 5D**) showed that compared with the control group, osteosarcoma tumor growth was decreased in mice with tumors harboring stable miR-378a-3p overexpression. These data indicate that miR-378a-3p inhibits osteosarcoma tumor growth *in vivo*.

DISCUSSION

BYSL is an evolutionarily conserved gene that is expressed in a wide range of eukaryotes (Pack et al., 1998). It was initially identified as a cytoplasmic protein expressed in human trophoblastic embryonal carcinoma cells, forming a complex with theophanic and tastin (Suzuki et al., 1998; Fukuda and Nozawa, 1999). However, the function of *BYSL* in osteosarcoma remains poorly understood. Studies have shown that *BYSL* contributes to tumor cell growth and survival by forming a complex with mTORC2 in gliomas (Gao et al., 2021). Furthermore, high expression of *BYSL* in glioblastoma has been shown to strongly correlate with markers of mesenchymal glioblastoma (Sha et al., 2020). Herein, bioinformatics analysis of TCGA data showed that *BYSL* overexpression is associated with an unsatisfactory prognosis. The present study provides evidence that high *BYSL* expression is

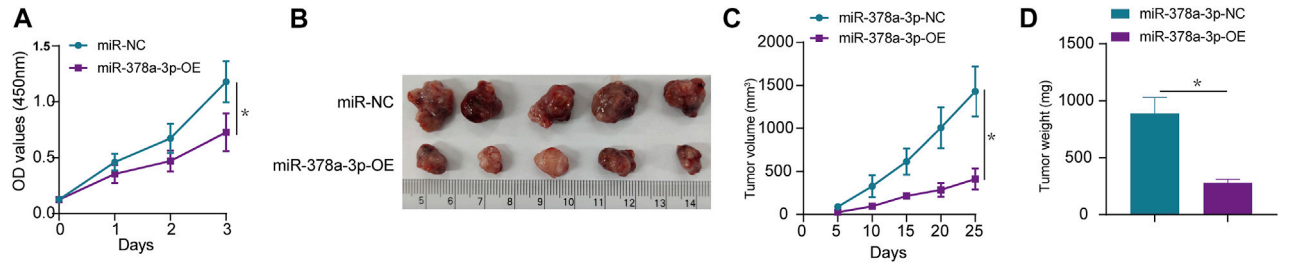


FIGURE 5 | miR-378a-3p inhibits MG63 cell proliferation *in vivo* and *in vitro*. **(A)** MG63 cells were transfected with control-mimic (miR-NC), miR-378a-3p-mimic (miR-378a-3p-OE), and then cultured under hypoxic conditions. Cell proliferation assay using CCK8 assay. **(B–D)** Tumor weight and volume were examined after mice were sacrificed, representative tumors from corresponding treatments groups. (n = 5). The data are presented as the mean ± SD. *p < 0.05.

related to metastasis, TNM stage, and relapse. Moreover, survival analysis showed that patients with high *BYSL* expression have significantly worse overall survival than those with low expression.

EMT, a key developmental process, is frequently activated during embryonic morphogenesis and metastasis. In the present study, *BYSL* overexpression promoted EMT in MG63 and Saos-2 cells, indicating that *BYSL* may exert a critical effect on the invasive and aggressive behavior of osteosarcoma. Many evolutionarily conserved genes seem to correlate with diverse cellular functions during embryonic development and tumorigenesis. For example, an evolutionarily conserved metabolic gene, *LOX*, is involved in gliomagenesis (Chi et al., 2019). Fascin is an evolutionarily highly conserved protein that is involved in cell migration and tumor metastasis (Mattila and Lappalainen, 2008; Sun et al., 2011). Additionally, p53 family members, including p53, p63, and p73, are evolutionarily well conserved, and their main function is to maintain the genomic integrity of germ cells (Hu et al., 2011). A previous study also revealed that *BYSL* is essential for blastocyst formation. Mechanistically, blocking *BYSL* function causes defects in 40S ribosomal subunit biogenesis (Adachi et al., 2007). In the tumor microenvironment, hypoxia is an important hallmark that regulates angiogenesis and tumor growth (Hanahan and Weinberg, 2011). It can drive the EMT process through activation of HIF-1α signaling (Muz et al., 2015). Nrf2 is involved in the stabilization of the hybrid epithelial/mesenchymal phenotype in cancer cells (Bocci et al., 2019). Under hypoxia, HIF-1α signaling increases the expression of Nrf2, which interacts with TRX1, leading to enhanced HIF-1α expression (Toth and Warfel, 2017). Hypoxia activates Nrf2 signaling, which exerts a significant effect on the regulation of EMT in cancer cells (Bocci et al., 2019). Therefore, we investigated whether *BYSL* expression could affect Nrf2 signaling in MG63 and Saos-2 cells. Generally, stabilized Nrf2 migrates into the nucleus in a heterodimer with a small MAF (sMAF) transcription factor (Itoh et al., 1997), promoting cytoprotective genes transcription (Rushmore and Pickett, 1990). In the present study, we observed hypoxia stimulates Nrf2 nuclear translocation. Meanwhile, overexpression of BYSL increased Nrf2 expression and promoted EMT process under hypoxic conditions. Our rescue experiments revealed that knockdown of Nrf2 could reverse the effect of BYSL on EMT. It was thereby inferred, BYSL may modulate EMT by regulating Nrf2 signaling. Together, these findings suggest that *BYSL* is a pro-oncogenic protein and, thus, *BYSL* inhibition may represent a promising molecular therapeutic strategy for treating osteosarcoma in humans.

Several microRNAs have been shown to modulate various biological processes in osteosarcoma cells (Luo et al., 2020). For example, low expression of miR-1225-5p is correlated with poor prognosis in patients with osteosarcoma, and its overexpression inhibits osteosarcoma cell invasion and metastasis by targeting *Sox9* (Zhang et al., 2020). *In vivo* and *in vitro* studies have shown that miR-223-3p upregulation reduces osteosarcoma cell invasion, migration, growth, and proliferation by reducing *CDH6* expression (Ji et al., 2018). Liu et al. also reported that overexpression of miR-95-3p suppresses osteosarcoma cell growth by targeting *HDGF* (Liu et al., 2019). Moreover, miR-378a-3p inhibits the proliferation and migration of glioblastoma (Guo et al., 2019). In esophageal squamous cell carcinoma (ESCC), miR-378a-3p functions as a tumor suppressor to inhibit the migration, proliferation, and invasion of ESCC cells (Ding et al., 2018). However, whether miR-378a-3p participates in osteosarcoma progression remains unclear. In the present study, miR-378a-3p was identified as being downregulated in osteosarcoma cells under hypoxic conditions. Interestingly, miR-378a-3p has been linked to the modulation of ischemia/reperfusion kidney injury (Ding et al., 2020), suggesting that is implicated in hypoxia. Previous reports have shown that hypoxia induces EMT in various tumors (Lu et al., 2020; Xing et al., 2021). Similarly, we found that hypoxia contributed to EMT via miR378a-3p downregulation. Luciferase reporter assays and rescue experiments demonstrated that miR-378a-3p inhibits hypoxia-induced EMT, invasion, and migration of osteosarcoma cells by targeting *BYSL*. Moreover, results of *in vivo* experiments further demonstrated that miR-378a-3p overexpression inhibits osteosarcoma cell proliferation. Therefore, miR-378a-3p maybe associated with osteosarcoma metastasis. Futher clinical studies are required to understand the detailed functions of miR-378a-3p in osteosarcoma (**Supplementary Figure S6**).

In summary, the present study demonstrates that *BYSL* is an independent factor for evaluating the prognosis of patients with osteosarcoma. Additionally, miR-378a-3p can inhibit EMT and invasion by directly targeting *BYSL*. The miR-378a-3p/BYSL axis

may play a role in osteosarcoma, and the cilinical significance of miR-378a-3p in osteosarcoma patients should be explored in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Medical Ethics of First Affiliated Hospital of Guangxi Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. The animal study was reviewed and approved by The Animal Care and Welfare Committee of Guangxi Medical University.

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the research and writing manusript. QW designed the study. JZ and HT performed the experiments. JZ, XJ, and NH collected and analyzed the data. JZ, HT, XJ, and NH drafted and critically revised the manuscript. QW critically revised the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This study was supported by the Guangxi Key R&D Program (Grant Number: GuikeAB17292073).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.804952/full#supplementary-material

Supplementary Figure S1 | Total BYSL, nuclear BYSL, Nrf2, and cytoplasmic Nrf2 in MG63 and Saos-2 were examined using western blot. *$p < 0.05$.

Supplementary Figure S2 | BYSL, Nrf2, E-cadherin, N-cadherin, and vimentin in MG63 and Saos-2 under hypoxia were examined using western blot. *$p < 0.05$.

Supplementary Figure S3 | BYSL, E-cadherin, N-cadherin, and vimentin in MG63 and Saos-2 were examined using western blot. *$p < 0.05$.

Supplementary Figure S4 | Bax, Bcl2, E-cadherin, N-cadherin, and vimentin, BYSL, and Nrf2 were examined using western blot under normoxia. *$p < 0.05$.

Supplementary Figure S5 | Bax, Bcl2, E-cadherin, N-cadherin, and vimentin, BYSL, and Nrf2 were examined using western blot under hypoxia. *$p < 0.05$.

Supplementary Figure S6 | Overall survival was compared between osteosarcoma patients with high and low *miR-378a-3p* expression in TCGA database.

REFERENCES

Adachi, K., Soeta-Saneyoshi, C., Sagara, H., and Iwakura, Y. (2007). Crucial Role of Bysl in Mammalian Preimplantation Development as an Integral Factor for 40S Ribosome Biogenesis. *Mol. Cel. Biol.* 27 (6), 2202–2214. doi:10.1128/mcb.01908-06

Ando, K., Heymann, M.-F., Stresing, V., Mori, K., Rédini, F., and Heymann, D. (2013). Current Therapeutic Strategies and Novel Approaches in Osteosarcoma. *Cancers* 5 (2), 591–616. doi:10.3390/cancers5020591

Aoki, R., Suzuki, N., Paria, B. C., Sugihara, K., Akama, T. O., Raab, G., et al. (2006). TheByslgene Product, Bystin, Is Essential for Survival of Mouse Embryos. *FEBS Lett.* 580 (26), 6062–6068. doi:10.1016/j.febslet.2006.09.072

Arias, A. M. (2001). Epithelial Mesenchymal Interactions in Cancer and Development. *Cell* 105 (4), 425–431. doi:10.1016/s0092-8674(01)00365-8

Bartel, D. P. (2004). MicroRNAs. *Cell* 116 (2), 281–297. doi:10.1016/s0092-8674(04)00045-5

Bocci, F., Tripathi, S. C., Vilchez Mercedes, S. A., George, J. T., Casabar, J. P., Wong, P. K., et al. (2019). NRF2 Activates a Partial Epithelial-Mesenchymal Transition and Is Maximally Present in a Hybrid Epithelial/mesenchymal Phenotype. *Integr. Biol. (Camb)* 11 (6), 251–263. doi:10.1093/intbio/zyz021

Buddingh, E. P., Kuijjer, M. L., Duim, R. A. J., Bürger, H., Agelopoulos, K., Myklebost, O., et al. (2011). Tumor-infiltrating Macrophages Are Associated with Metastasis Suppression in High-Grade Osteosarcoma: a Rationale for Treatment with Macrophage Activating Agents. *Clin. Cancer Res.* 17 (8), 2110–2119. doi:10.1158/1078-0432.ccr-10-2047

Chanjiao, Y., Chunyan, C., Xiaoxin, Q., and Youjian, H. (2021). MicroRNA-378a-3p Contributes to Ovarian Cancer Progression through Downregulating PDIA4. *Immun. Inflamm. Dis.* 9 (1), 108–119. doi:10.1002/iid3.350

Chi, K.-C., Tsai, W.-C., Wu, C.-L., Lin, T.-Y., and Hueng, D.-Y. (2019). An Adult Drosophila Glioma Model for Studying Pathometabolic Pathways of Gliomagenesis. *Mol. Neurobiol.* 56 (6), 4589–4599. doi:10.1007/s12035-018-1392-2

Ding, C., Ding, X., Zheng, J., Wang, B., Li, Y., Xiang, H., et al. (2020). miR-182-5p and miR-378a-3p Regulate Ferroptosis in I/R-induced Renal Injury. *Cell Death Dis.* 11 (10), 929. doi:10.1038/s41419-020-03135-z

Ding, N., Sun, X., Wang, T., Huang, L., Wen, J., and Zhou, Y. (2018). miR-378a-3p Exerts Tumor Suppressive Function on the Tumorigenesis of Esophageal Squamous Cell Carcinoma by Targeting Rab10. *Int. J. Mol. Med.* 42 (1), 381–391. doi:10.3892/ijmm.2018.3639

Fukuda, M., and Nozawa, S. (1999). Trophinin, Tastin, and Bystin: a Complex Mediating Unique Attachment between Trophoblastic and Endometrial Epithelial Cells at Their Respective Apical Cell Membranes. *Semin. Reprod. Med.* 17 (3), 229–234. doi:10.1055/s-2007-1016230

Gao, S., Sha, Z., Zhou, J., Wu, Y., Song, Y., Li, C., et al. (2021). BYSL Contributes to Tumor Growth by Cooperating with the mTORC2 Complex in Gliomas. *Cancer Biol. Med.* 18 (1), 88–104. doi:10.20892/j.issn.2095-3941.2020.0096

Gorlick, S., Barkauskas, D. A., Krailo, M., Piperdi, S., Sowers, R., Gill, J., et al. (2014). HER-2 Expression Is Not Prognostic in Osteosarcoma; a Children's Oncology Group Prospective Biology Study. *Pediatr. Blood Cancer* 61 (9), 1558–1564. doi:10.1002/pbc.25074

Guo, X. B., Zhang, X. C., Chen, P., Ma, L. M., and Shen, Z. Q. (2019). miR-378a-3p Inhibits Cellular Proliferation and Migration in Glioblastoma Multiforme by Targeting Tetraspanin 17. *Oncol. Rep.* 42 (5), 1957–1971. doi:10.3892/or.2019.7283

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

Harrison, D. J., Geller, D. S., Gill, J. D., Lewis, V. O., and Gorlick, R. (2018). Current and Future Therapeutic Approaches for Osteosarcoma. *Expert Rev. Anticancer Ther.* 18 (1), 39–50. doi:10.1080/14737140.2018.1413939

Hu, W., Zheng, T., and Wang, J. (2011). Regulation of Fertility by the P53 Family Members. *Genes & Cancer* 2 (4), 420–430. doi:10.1177/1947601911408892

Ikeda, K., Horie-Inoue, K., Ueno, T., Suzuki, T., Sato, W., Shigekawa, T., et al. (2015). miR-378a-3p Modulates Tamoxifen Sensitivity in Breast Cancer MCF-7 Cells through Targeting GOLT1A. *Sci. Rep.* 5, 13170. doi:10.1038/srep13170

Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., et al. (1997). An Nrf2/small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes through Antioxidant Response Elements. *Biochem. Biophysical Res. Commun.* 236 (2), 313–322. doi:10.1006/bbrc.1997.6943

Ji, Q., Xu, X., Song, Q., Xu, Y., Tai, Y., Goodman, S. B., et al. (2018). miR-223-3p Inhibits Human Osteosarcoma Metastasis and Progression by Directly Targeting CDH6. *Mol. Ther.* 26 (5), 1299–1312. doi:10.1016/j.ymthe.2018.03.009

Joseph, J. P., Harishankar, M. K., Pillai, A. A., and Devi, A. (2018). Hypoxia Induced EMT: A Review on the Mechanism of Tumor Progression and Metastasis in OSCC. *Oral Oncol.* 80, 23–32. doi:10.1016/j.oraloncology.2018.03.004

Lee, Y. S., and Dutta, A. (2009). MicroRNAs in Cancer. *Annu. Rev. Pathol. Mech. Dis.* 4, 199–227. doi:10.1146/annurev.pathol.4.110807.092222

Li, J., Qin, B., Huang, M., Ma, Y., Li, D., Li, W., et al. (2021). Tumor-Associated Antigens (TAAs) for the Serological Diagnosis of Osteosarcoma. *Front. Immunol.* 12, 665106. doi:10.3389/fimmu.2021.665106

Liu, M., Si, Q., Ouyang, S., Zhou, Z., Wang, M., Zhao, C., et al. (2020). Serum MiR-4687-3p Has Potential for Diagnosis and Carcinogenesis in Non-small Cell Lung Cancer. *Front. Genet.* 11, 597508. doi:10.3389/fgene.2020.597508

Liu, X., Ma, W., Ma, J., Xiao, L., and Hao, D. (2019). Upregulation of miR-95-3p I-nhibits G-rowth of O-steosarcoma by T-argeting HDGF. *Pathol. - Res. Pract.* 215 (8), 152492. doi:10.1016/j.prp.2019.152492

Lu, Y., Liu, Y., Oeck, S., Zhang, G. J., Schramm, A., and Glazer, P. M. (2020). Hypoxia Induces Resistance to EGFR Inhibitors in Lung Cancer Cells via Upregulation of FGFR1 and the MAPK Pathway. *Cancer Res.* 80 (21), 4655–4667. doi:10.1158/0008-5472.can-20-1192

Luo, H., Wang, P., Ye, H., Shi, J., Dai, L., Wang, X., et al. (2020). Serum-Derived microRNAs as Prognostic Biomarkers in Osteosarcoma: A Meta-Analysis. *Front. Genet.* 11, 789. doi:10.3389/fgene.2020.00789

Mattila, P. K., and Lappalainen, P. (2008). Filopodia: Molecular Architecture and Cellular Functions. *Nat. Rev. Mol. Cel. Biol.* 9 (6), 446–454. doi:10.1038/nrm2406

Muz, B., de la Puente, P., Azab, F., and Azab, A. K. (2015). The Role of Hypoxia in Cancer Progression, Angiogenesis, Metastasis, and Resistance to Therapy. *Hp* 3, 83–92. doi:10.2147/hp.s93413

Olczak, M., Chutorański, D., Kwiatkowska, M., Samojłowicz, D., Tarka, S., and Wierzba-Bobrowicz, T. (2018). Bystin (BYSL) as a Possible Marker of Severe Hypoxic-Ischemic Changes in Neuropathological Examination of Forensic Cases. *Forensic Sci. Med. Pathol.* 14 (1), 26–30. doi:10.1007/s12024-017-9942-x

Pack, S. D., Pak, E., Tanigami, A., Ledbetter, D. H., and Fukuda, M. N. (1998). Assignment1 of the Bystin Gene BYSL to Human Chromosome Band 6p21.1 by *In Situ* Hybridization. *Cytogenet. Genome Res.* 83 (1-2), 76–77. doi:10.1159/000015131

Pan, X., Zhao, L., Quan, J., Liu, K., Lai, Y., Li, Z., et al. (2019). MiR-378a-5p Acts as a Tumor Suppressor in Renal Cell Carcinoma and Is Associated with the Good Prognosis of Patients. *Am. J. Transl. Res.* 11 (4), 2207–2218.

Raymond, A. K., and Jaffe, N. (2009). Osteosarcoma Multidisciplinary Approach to the Management from the Pathologist's Perspective. *Cancer Treat. Res.* 152, 63–84. doi:10.1007/978-1-4419-0284-9_4

Rushmore, T. H., and Pickett, C. B. (1990). Transcriptional Regulation of the Rat Glutathione S-Transferase Ya Subunit Gene. Characterization of a Xenobiotic-Responsive Element Controlling Inducible Expression by Phenolic Antioxidants. *J. Biol. Chem.* 265 (24), 14648–14653. doi:10.1016/s0021-9258(18)77351-1

Sevelda, F., Mayr, L., Kubista, B., Lötsch, D., van Schoonhoven, S., Windhager, R., et al. (2015). EGFR Is Not a Major Driver for Osteosarcoma Cell Growth *In Vitro* but Contributes to Starvation and Chemotherapy Resistance. *J. Exp. Clin. Cancer Res.* 34, 134. doi:10.1186/s13046-015-0251-5

Sha, Z., Zhou, J., Wu, Y., Zhang, T., Li, C., Meng, Q., et al. (2020). BYSL Promotes Glioblastoma Cell Migration, Invasion, and Mesenchymal Transition through the GSK-3β/β-Catenin Signaling Pathway. *Front. Oncol.* 10, 565225. doi:10.3389/fonc.2020.565225

Shi, C., Huang, C.-M., Wang, B., Sun, T.-F., Zhu, A.-X., and Zhu, Y.-C. (2020). Pseudogene MSTO2P Enhances Hypoxia-Induced Osteosarcoma Malignancy by Upregulating PD-L1. *Biochem. Biophysical Res. Commun.* 530 (4), 673–679. doi:10.1016/j.bbrc.2020.07.113

Sun, J., He, H., Xiong, Y., Lu, S., Shen, J., Cheng, A., et al. (2011). Fascin Protein Is Critical for Transforming Growth Factor β Protein-Induced Invasion and Filopodia Formation in Spindle-Shaped Tumor Cells. *J. Biol. Chem.* 286 (45), 38865–38875. doi:10.1074/jbc.M111.270413

Suzuki, N., Zara, J., Sato, T., Ong, E., Bakhiet, N., Oshima, R. G., et al. (1998). A Cytoplasmic Protein, Bystin, Interacts with Trophinin, Tastin, and Cytokeratin and May Be Involved in Trophinin-Mediated Cell Adhesion between Trophoblast and Endometrial Epithelial Cells. *Proc. Natl. Acad. Sci.* 95 (9), 5027–5032. doi:10.1073/pnas.95.9.5027

Toth, R., and Warfel, N. (2017). Strange Bedfellows: Nuclear Factor, Erythroid 2-Like 2 (Nrf2) and Hypoxia-Inducible Factor 1 (HIF-1) in Tumor Hypoxia. *Antioxidants* 6 (2), 27. doi:10.3390/antiox6020027

Wang, H., Xiao, W., Zhou, Q., Chen, Y., Yang, S., Sheng, J., et al. (2009). Bystin-like Protein Is Upregulated in Hepatocellular Carcinoma and Required for Nucleologenesis in Cancer Cell Proliferation. *Cell Res.* 19 (10), 1150–1164. doi:10.1038/cr.2009.99

Wang, S., Zhang, D., Han, S., Gao, P., Liu, C., Li, J., et al. (2017). Fibulin-3 Promotes Osteosarcoma Invasion and Metastasis by Inducing Epithelial to Mesenchymal Transition and Activating the Wnt/β-Catenin Signaling Pathway. *Sci. Rep.* 7 (1), 6215. doi:10.1038/s41598-017-06353-2

Xing, S., Tian, Z., Zheng, W., Yang, W., Du, N., Gu, Y., et al. (2021). Hypoxia Downregulated miR-4521 Suppresses Gastric Carcinoma Progression through Regulation of IGF2 and FOXM1. *Mol. Cancer* 20 (1), 9. doi:10.1186/s12943-020-01295-2

Zhang, W., Wei, L., Sheng, W., Kang, B., Wang, D., and Zeng, H. (2020). miR-1225-5p Functions as a Tumor Suppressor in Osteosarcoma by Targeting Sox9. *DNA Cel Biol.* 39 (1), 78–91. doi:10.1089/dna.2019.5105

Zheng, S., Li, M., Miao, K., and Xu, H. (2020). lncRNA GAS5-promoted Apoptosis in Triple-negative Breast Cancer by Targeting miR-378a-5p/SUFU Signaling. *J. Cel. Biochem.* 121 (3), 2225–2235. doi:10.1002/jcb.29445

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Tang, Jiang, Huang and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fgene.2022.803460

THBS2 is Closely Related to the Poor Prognosis and Immune Cell Infiltration of Gastric Cancer

Shiyu Zhang[1], *Huiying Yang*[1], *Xuelian Xiang*[1], *Li Liu*[1], *Huali Huang*[2] *and Guodu Tang*[1]*

[1]Department of Gastroenterology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, [2]Department of Gastroenterology, The Fifth Affiliated Hospital of Guangxi Medical University, Nanning, China

Background: The potential functions of Thrombospondin 2 (THBS2) in the progression and immune infiltration of gastric cancer (GC) remain unclear. The purpose of this study was to clarify the role of THBS2 in GC prognosis and the relationship between THBS2 and GC immune cell infiltration.

Material and Methods: The differential expression levels of THBS2 in the GC and cancer-adjacent tissues were identified using the TCGA databases and verified using real-time polymerase chain reaction (PCR), immunohistochemical staining and two datasets from Gene Expression Omnibus (GEO). THBS2 related differential expressed genes (DEGs) were identified and used for further functional enrichment analysis and Gene Set Enrichment Analysis (GSEA). Furthermore, a THBS2-related immune infiltration analysis was also performed. Kaplan-Meier and Cox regression analyses were utilized to illustrate the effects of THBS2 on the prognosis and clinical variables of GC. Finally, a nomogram was constructed to predict the survival probability of patients with GC.

Results: The THBS2 expression in GC was significantly higher than that in cancer-adjacent tissues ($p < 0.001$), which was verified using real-time PCR, immunohistochemical staining and datasets from GEO. The 599 identified DEGs were primarily enriched in pathways related to tumorigenesis and tumor progression, including the focal adhesion pathway, signaling by vascular endothelial growth factor, and Wnt signaling. THBS2 expression was positively correlated with the enrichment of the macrophages ($r = 0.590$, $p < 0.001$), which was also confirmed by immunohistochemistry; however, negatively correlated with the enrichment of Th17 cells ($r = 0.260$, $p < 0.001$). The high expression of THBS2 was significantly correlated with the pathological grade ($p < 0.01$), histological grade ($p < 0.05$), histological type ($p < 0.05$), T stage ($p < 0.001$), and poor overall survival (OS) ($P = 0.003$) of GC. The constructed nomogram can well predict the 1-, 3-, and 5-years OS probability of patients with GC (C-index [95% confidence interval] = 0.725 [0.701–0.750]).

Conclusion: THBS2 is closely related to the poor prognosis and immune infiltration of gastric cancer.

Keywords: gastric cancer, biomarker, immune infiltration, prognostic index, bioinformatics

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
Zunyi Medical University, China

Reviewed by:
Jozsef Dudas,
Innsbruck Medical University, Austria
Sara Pagotto,
University of Studies G. d'Annunzio Chieti and Pescara, Italy

***Correspondence:**
Guodu Tang
tangguodu@stu.gxmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *28 October 2021*
Accepted: *03 January 2022*
Published: *03 February 2022*

Citation:
Zhang S, Yang H, Xiang X, Liu L, Huang H and Tang G (2022) THBS2 is Closely Related to the Poor Prognosis and Immune Cell Infiltration of Gastric Cancer. Front. Genet. 13:803460. doi: 10.3389/fgene.2022.803460

INTRODUCTION

Gastric cancer (GC) is the fifth most common malignancy and the third most common cause of death globally (Park et al., 2018). Currently, the treatment of GC includes surgery (Herrera-Almario and Strong, 2016), radiotherapy (Tey et al., 2017), neoadjuvant chemotherapy (Zhandossov et al., 2018), and immunotherapy (Niccolai et al., 2015). Due to the improvement in endoscopic screening technology, in countries with a high incidence of GC, such as Japan and South Korea, more than 50 percent of GC patients can be diagnosed at an early stage (So et al., 2021). However, in most countries with poor screening technology, most patients are diagnosed at an advanced GC stage, which leads to poor treatment outcomes (Choi et al., 2015; Inokuchi et al., 2018). Therefore, biomarkers related to the progression and GC prognosis should be explored to improve the therapeutic effect and reduce mortality owing to GC.

Thrombospondin-2 (THBS2) is a member of the matricellular calcium-binding glycoprotein family, which interacts with growth factors, cell receptors, and extracellular matrix (ECM), and THBS2 plays an important role in cell proliferation, adhesion and apoptosis (Bornstein et al., 2000; Zhuo et al., 2016). It has also been reported that THBS2 may be a serum biomarker in the diagnosis of colon cancer (Wang et al., 2016) and lung cancer (Weng et al., 2016). Bioinformatics analysis suggested that THBS2 might be a potential biomarker for GC (Cao et al., 2018). An *in vitro* study showed that THBS2 silencing inhibited the proliferation, migration and invasion of gastric

TABLE 1 | The clinical characteristic of STAD.

Characteristic	Levels	Overall
n		375
Gender, n (%)	Female	134 (35.7%)
	Male	241 (64.3%)
Age, n (%)	≤ 65	164 (44.2%)
	>65	207 (55.8%)
T stage, n (%)	T1	19 (5.2%)
	T2	80 (21.8%)
	T3	168 (45.8%)
	T4	100 (27.2%)
N stage, n (%)	N0	111 (31.1%)
	N1	97 (27.2%)
	N2	75 (21%)
	N3	74 (20.7%)
M stage, n (%)	M0	330 (93%)
	M1	25 (7%)
Histological type, n (%)	Diffuse Type	63 (16.8%)
	Mucinous Type	19 (5.1%)
	Not Otherwise Specified	207 (55.3%)
	Papillary Type	5 (1.3%)
	Signet Ring Type	11 (2.9%)
	Tubular Type	69 (18.4%)
Pathologic stage, n (%)	Stage I	53 (15.1%)
	Stage II	111 (31.5%)
	Stage III	150 (42.6%)
	Stage IV	38 (10.8%)
Histologic grade, n (%)	G1	10 (2.7%)
	G2	137 (37.4%)
	G3	219 (59.8%)
Residual tumor, n (%)	R0	298 (90.6%)
	R1	15 (4.6%)
	R2	16 (4.9%)
Primary therapy outcome, n (%)	PD	65 (20.5%)
	SD	17 (5.4%)
	PR	4 (1.3%)
	CR	231 (72.9%)
H pylori infection, n (%)	No	145 (89%)
	Yes	18 (11%)
Barrett's esophagus, n (%)	No	193 (92.8%)
	Yes	15 (7.2%)
Anatomic neoplasm subdivision, n (%)	Antrum/Distal	138 (38.2%)
	Cardia/Proximal	48 (13.3%)
	Fundus/Body	130 (36%)
	Gastroesophageal Junction	41 (11.4%)
	Other	4 (1.1%)
Age, median (IQR)		67 (58, 73)

TABLE 2 | Clinical information of 16 patients with gastric cancer for Real-time PCR.

Gender (M/F)	Age (years)	Histological type	Histological grade	Anatomic neoplasm subdivision
M	76	Signet Ring Type	G3	Antrum
M	54	Tubular Type	G3	Antrum
M	73	Tubular Type	G1	Antrum
F	59	Mucinous Type	G3	Fundus/Body
M	49	Papillary Type	G3	Antrum
F	37	Signet Ring Type	G3	Antrum
M	60	Tubular Type	G2	Antrum
M	76	Tubular Type	G1	Antrum
F	48	Signet Ring Type	G3	Cardia
M	65	Papillary Type	G2	Cardia
M	70	Diffuse Type	G3	Antrum
M	82	Tubular Type	G1	Fundus/Body
M	67	Diffuse Type	G2	Cardia
M	67	Tubular Type	G2	Cardia
F	77	Tubular Type	G2	Cardia
F	52	Diffuse Type	G3	Antrum

cancer cells (Ao et al., 2018). The results of Chuanjun Zhuo showed that GC patients with low THBS2 expression had a better prognosis (Zhuo et al., 2016). However, there have also been contrasting reports. It has been reported that the expression of THBS2 is down-regulated in most GC patients, and the higher the expression of THBS2, the better the prognosis of GC patients, but the sample size was too small, only 14 cases (Sun et al., 2014). Therefore, the role of THBS2 in the prognosis of GC needs to be verified further.

Since immune checkpoint blocking therapy is used to treat several types of tumors, some patients have achieved significant clinical responses (Ribas and Wolchok, 2018). In September 2017, pembrolizumab (anti-programmed death 1 [PD1] antibody) was approved to treat GC or gastroesophageal junction cancer; however, the response rate was relatively low (Fashoyin-Aje et al., 2019). The effectiveness of tumor immunotherapy should be based on the premise that effector cells infiltrate into the tumor microenvironment (TME) (Kirkwood et al., 2012). Anti-ctla-4 (T lymphocyte associated antigen 4) monoclonal antibody plays an antitumor role by prolonging T cell stimulation and restoring T cell proliferation (Engelhardt et al., 2006). T cells are the only type of tumor-infiltrating lymphocytes (TILs), and other common TILs cells include macrophages and NK cells. Therefore, understanding the tumor microenvironment and TILs of GC may lay a certain foundation for improving the effect of immunotherapy in GC. However, it has not been reported whether THBS2 can affect the TILs abundance of GC.

In this study, we analyzed the significance of THBS2 in GC in the TCGA database and Gene Expression Omnibus (GEO) database using bioinformatics analysis, including differential expressed genes (DEGs) analysis, functional enrichment analysis, gene set enrichment analysis (GSEA), immune cell infiltration analysis, clinical correlation analysis, and survival analysis. Moreover, a nomogram was also created to predict the overall survival (OS) of patients with GC.

MATERIALS AND METHODS

Data Sources

We downloaded gene expression data and clinical information for GC, which included 32 cancer-adjacent tissues and 375 tumor tissues, from TCGA (https://portal.gdc.cancer.gov/, accessed time: July 31, 2021). Samples with incomplete clinical data were excluded. RNAseq data was converted by log2 in R software for subsequent analysis. The clinical characteristic of GC is presented in **Table 1**. Additionally, two microarray datasets (GSE54129 and GSE13911) containing GC and cancer-adjacent tissues were downloaded from GEO online database (https://www.ncbi.nlm.nih.gov/geo/, accessed time: December 17, 2021).

THBS2 *Differential Expression in Pan-Cancer and GC Tissues*

Log2-converted TPM RNAseq data for cancer-adjacent and tumor tissues (TCGA) were obtained from the UCSC XENA. The differential expression between tumor and non-tumor tissues was tested using Wilcoxon Rank Sum Test and visualized through the box and scatter plots. Additionally, the receiver operating characteristic (ROC) curve was constructed to assess the predictive diagnostic value of THBS2 expression.

Real-Time PCR of THBS2 Expressions in GC and Adjacent Tissues

This study has been approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University and informed consent of all patients. From May to July 2021, cancer and para-cancerous biopsy specimens were incessantly collected from 16 patients diagnosed with GC at the Endoscopy Center of the First Affiliated Hospital of Guangxi Medical University (Nanning, Guangxi). All tissue specimens were immersed in RNA protective solution and quickly transferred to a −80°C refrigerator for preservation. All patients had not

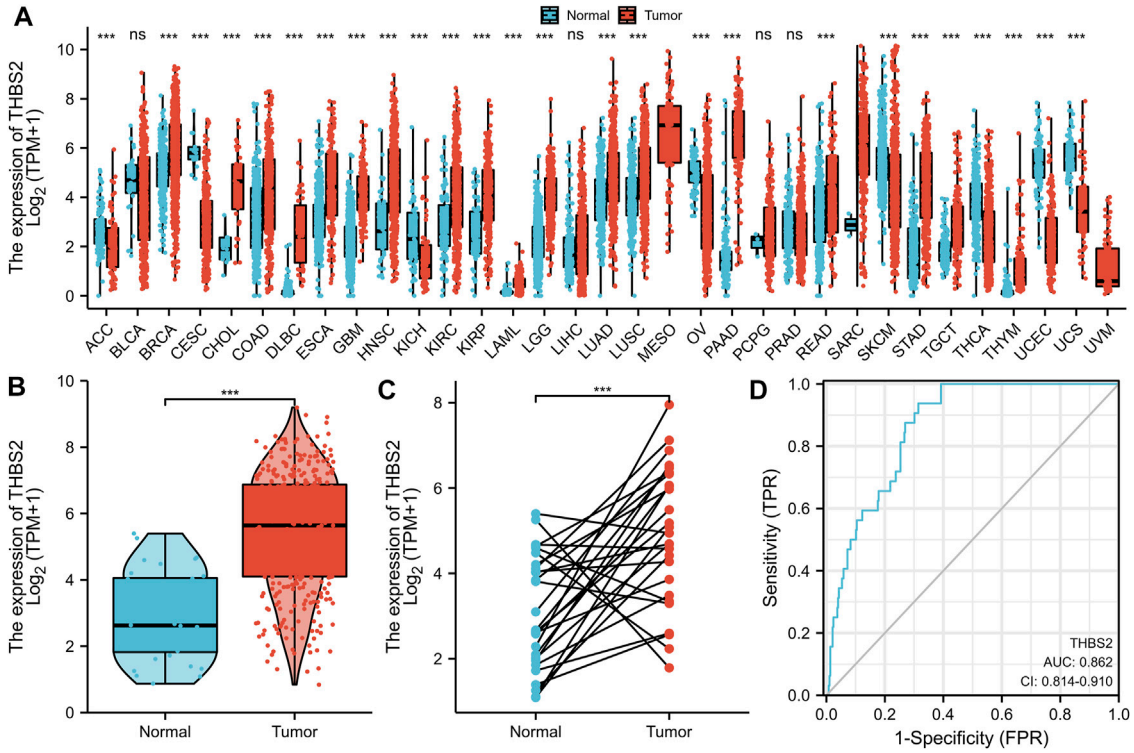


FIGURE 1 | Differential expression of THBS2 in different tumors and THBS2-related differentially expressed genes (DEGs). **(A)** Differential expression of THBS2 of different cancers compared with cancer-adjacent tissues in the TCGA database. **(B,C)** Differential expression of THBS2 in GC. **(D)** ROC curve was used to calculate the predictive diagnostic value of THBS2 expression between GC and para-cancer tissues. Significance marker: NS, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

received any prior treatment for the tumor, including radiation or chemotherapy. Furthermore, patients who were complicated with other known tumors were excluded. The clinical information of these 16 patients is presented in **Table 2**.

RNA Extraction and Quantitative Real-Time PCR

First, total RNA from tissues was extracted using the Trizol reagent (R0016, Beyotime). Then, the first-strand cDNA was synthesized from the total RNA via the PrimeScript™ RT Reagent Kit with gDNA Eraser (RR047A, Takara). After that, the expression of the THBS2 gene was normalized to GAPDH expression. The expression level of the THBS2 gene was computed by using the $2^{-\Delta\Delta}$ Ct way.

Specific primer base sequence:

THBS2-F: 5′-ATCACACGCATCCGTCTCTG-3’.
THBS2-R: 5′-ATCACACGCATCCGTCTCTG-3’.
GAPDH-F: 5′-GTCAGCCGCATCTTCTTT-3’.
GAPDH-R: 5′-CGCCCAATACGACCAAAT-3’.

Immunohistochemistry

We continuously collected tumor and para-tumor tissues from 80 GC patients without other known malignant tumors after surgery in Suqian First People’s Hospital from January 2017 to April 2019. None had a history of chemotherapy or radiation therapy before surgery. All the tissues were modified into tissue chips for further IHC staining. Dewaxing, and hydration were performed before the THBS2 primary antibody (ab112543, Abcam, 1:1000) was added and incubated at 4°C for 12 h. After incubating with HRP labeled linked polymer (KIT-5009, MXB biotechnologies) at 26°C for 40 min, signal detection was performed using DAB (P0202, Beyotime). The expression of THBS2 was measured using ImageJ software and visualized with GraphPad Prism 8.0 after statistical analysis.

Identification of DEGs Between High and low Expression Groups of THBS2

According to the mean value of THBS2 expression, all the samples from TCGA datasets were divided into high and low expression groups. RNA-Seq expression data were processed using the DESeq2 package, and DEGs were defined with a p. adj <0.05 and an absolute logFC >1.5. Detailed gene expression is shown in the volcano map.

Functional Enrichment Analysis of DEGs

In this study, or. Hs.eg.db package was used for ID conversion, while the clusterProfiler package was used for enrichment analysis. The threshold conditions included: p. adj <0.05, and q value < 0.2. Statistical analysis and GSEA visualization are performed by the clusterProfiler package, using C2 collection

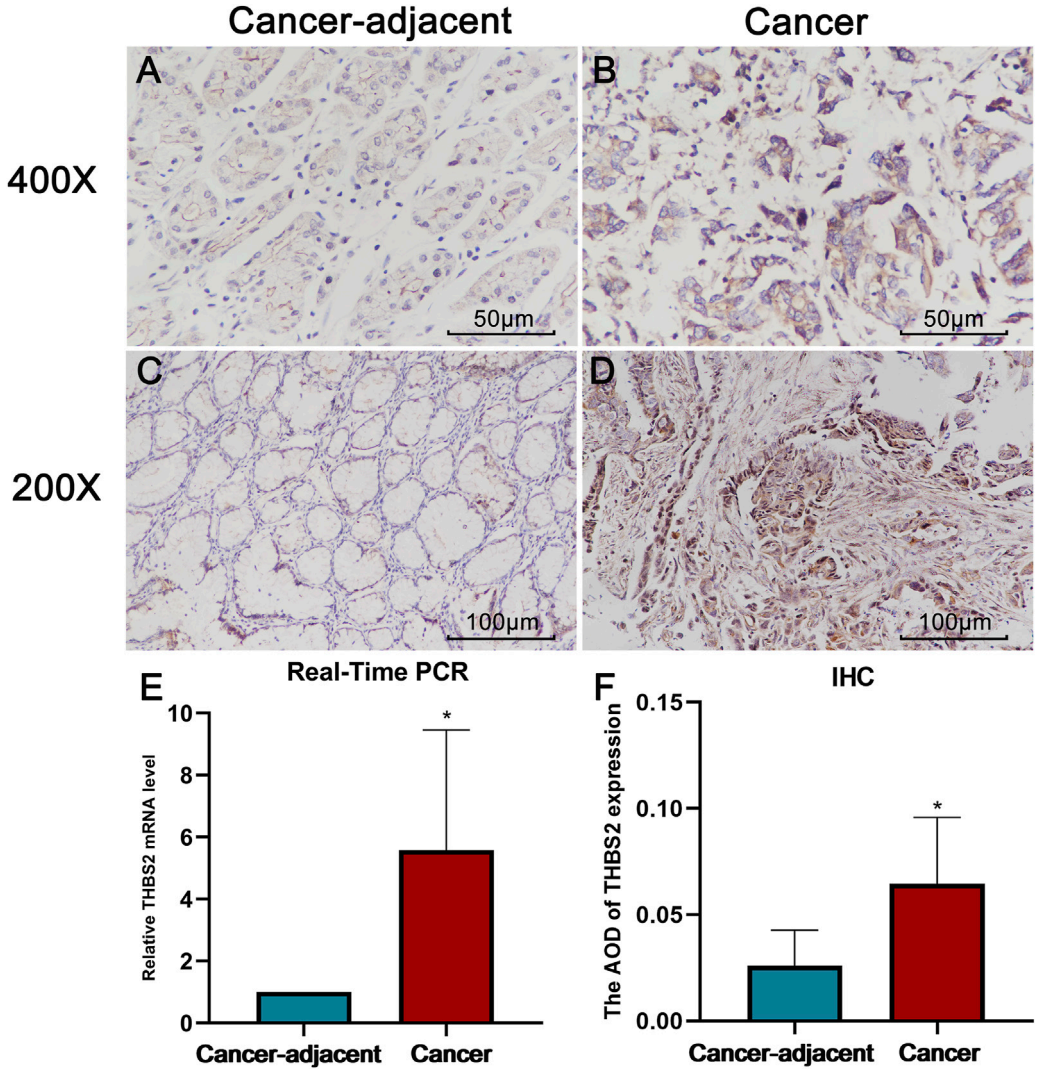


FIGURE 2 | The results of Real-Time PCR and Immunohistochemistry (IHC). **(A)** THBS2 expression in cancer-adjacent tissue (400X). **(B)** THBS2 expression in GC tissue (400X). **(C)** THBS2 expression in cancer-adjacent tissue (200X). **(D)** THBS2 expression in GC tissue (200X). **(E)** Relative THBS2 mRNA level in cancer-adjacent and GC tissues. GC: Gastric cancer. *, $p < 0.05$.

from MSigDB. Permutations with 10,000 times were performed by gene set, and significance was set as an adjusted $P < 0.05$ and false discovery rate (FDR) < 0.25.

Immune Cell Infiltration

The relative tumor infiltration levels of immune cell types were quantified using ssGSEA of GSVA package (Hänzelmann et al., 2013) to interrogate gene expression levels in published signature gene lists (Bindea et al., 2013). To explore the correlation between THBS2 and the immune infiltration levels and the association of immune infiltration with the different expression groups of THBS2, Spearman's correlation and Wilcoxon signed-rank sum test was adopted.

Furthermore, the average optical density (AOD) of THBS2 expression was calculated using ImageJ software based on immunohistochemistry results of 80 GC patients. According to the mean value of AOD of THBS2, all 80 GC patients were divided into high and low expression groups. Finally, 10 cases were randomly selected from the high expression and low expression groups for immunohistochemical staining of macrophages and NK cells. The immunohistochemical procedure was the same as part 2.5, and we used CD56 (MAB-0743, MXB biotechnologies) and CD68 (MAB-0041, MXB biotechnologies) to label NK cells and macrophages, respectively. The proportion of CD56/CD68 positive area was measured using ImageJ software and visualized with GraphPad Prism 8.0 after statistical analysis.

Clinical Correlation Analysis of THBS2 in GC

Wilcoxon signed-rank sum test and logistic regression were used to evaluate the correlation between THBS2 expression and clinicopathological variables. Further, univariate and

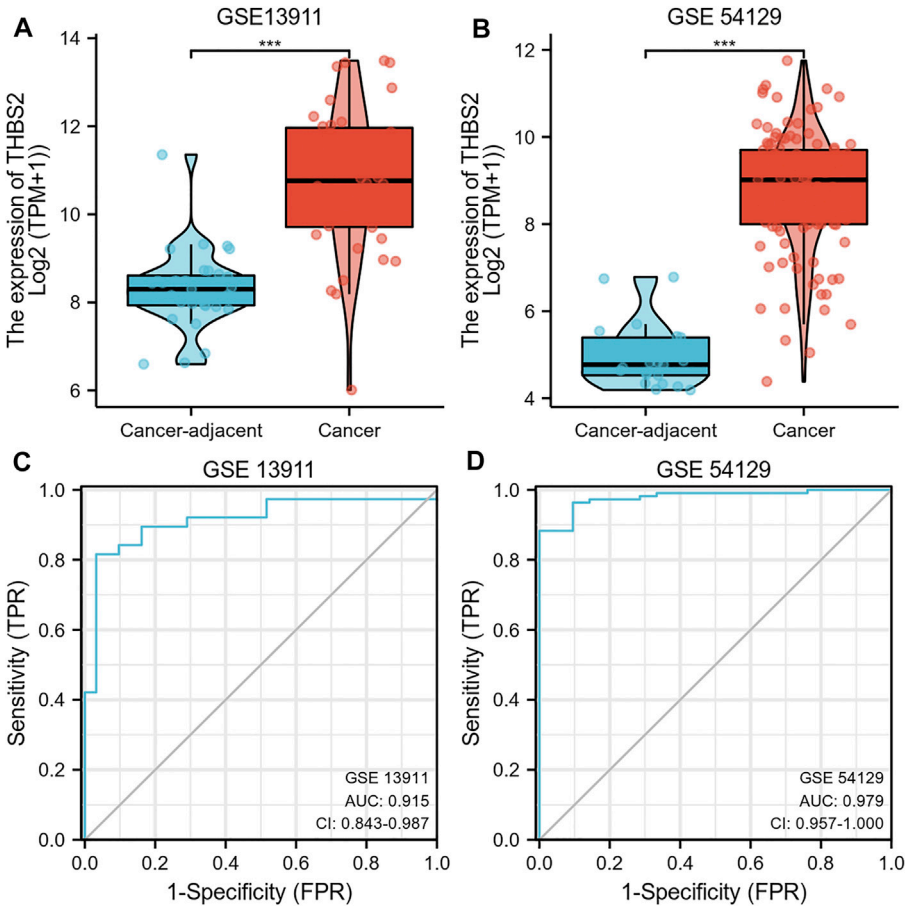


FIGURE 3 | (A) Differential THBS2 mRNA expression between GC and cancer-adjacent tissues in GSE13911. **(B)** Differential THBS2 mRNA expression between GC and cancer-adjacent tissues in GSE54129. **(C)** ROC curve used to calculate the predictive diagnostic value of THBS2 expression between GC and para-cancer tissues (GSE 13911). **(D)** ROC curve used to calculate the predictive diagnostic value of THBS2 expression between GC and para-cancer tissues (GSE54129). TPM: Transcripts Per Kilobase of exon model per Million mapped reads. GC, Gastric cancer; Significance marker: ***, $p < 0.001$.

multivariate Cox regressions were used to compare the effect of THBS2 expression and other clinicopathological variables on survival in patients with GC. Independent factors for GC prognosis were identified using multivariate Cox regression analysis.

Furthermore, we collected clinicopathological data from 80 patients who underwent THBS2 immunohistochemistry to evaluate the relationship between THBS2 expression and clinicopathological variables. Chi-square tests were used to evaluate the relationship between gender, pathological type, residual tumor status, and THBS2 expression. Fisher's exact tests were used to evaluate the relationship between pathologic stage, T stage, N stage, primary treatment outcome, and THBS2 expression. Wilcoxon signed Rank-Sum test was used to evaluate the relationship between age and THBS2 expression.

Construction and Verification of Nomogram

A nomogram used to predict the probability of survival at 1, 3, and 5 years for patients with GC included all independent prognostic factors. The prognostic data were derived from published studies (Liu et al., 2018). Nomogram was formulated using R with the survival and rms package. We used the bootstrap method to calculate the parameters, and the number of repeated calculations for each group of samples was 200. The C-index is used to evaluate the prediction ability of the nomogram; the closer C-index is to 1, the stronger the prediction ability is.

RESULTS

THBS2 Differential Expression in Pan-Cancer and GC Tissues

Significant differential expression was observed in most of the 33 cancers, as illustrated in **Figure 1A**. Secondly, the THBS2 expression in 375 STAD samples and 32 para-cancer samples in the TCGA was compared. There was a significant difference in THBS2 expression between GC tissues and cancer-adjacent tissues ($P < 0.001$) (**Figure 1B**). The THBS2 expression in GC tissues was also significantly higher than that in the corresponding adjacent tissues ($P < 0.001$) (**Figure 1C**).

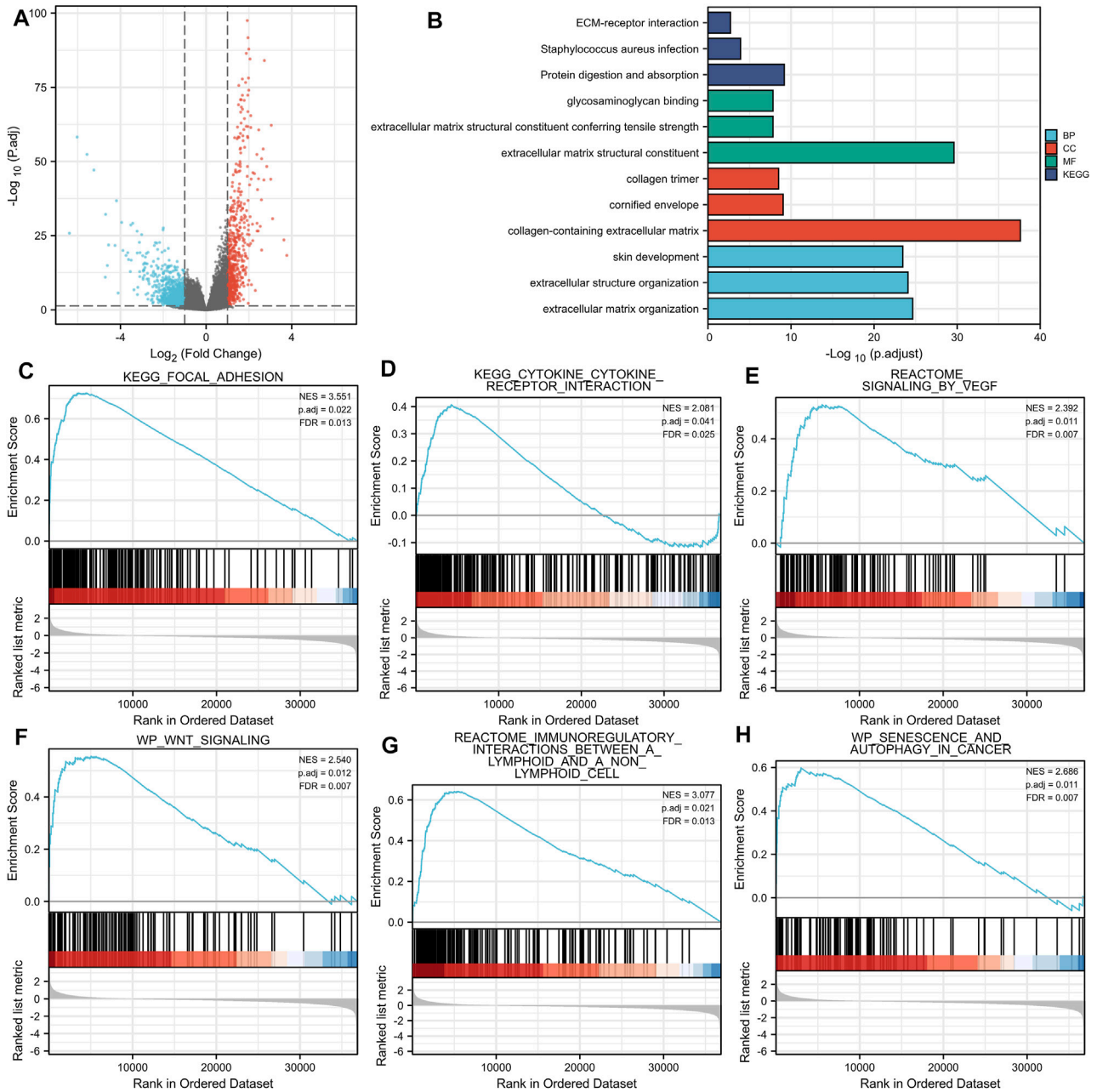


FIGURE 4 | (A) Volcano plots of the DEGs. Blue represents down-regulated DEGs, red represents up-regulated DEGs. **(B)**: The top three items enriched in biological processes (BP), cellular component (CC), molecular function (MF), and Kyoto Encyclopedia of Genes and Genomes (KEGG) of DEGs. **(C–H)**: Enrichment plots from the gene set enrichment analysis (GSEA). NES, normalized enrichment score; p. adj, adjusted *p* value; FDR, false discovery rate.

The area under the curve (AUC) of the ROC curve was used to evaluate the predictive value of THBS2 for the diagnosis of GC. The result of the ROC curve (**Figure 1D**) demonstrated that THBS2 had a high predictive value in distinguishing GC from cancer-adjacent tissues (AUC = 0.864, 95%CI: 0.812–0.915).

The real-time PCR results further verified the reliability of our differential analysis at the transcriptional level (**Figure 2E** Cancer-adjacent vs. Cancer, $P = 0.021$). Moreover, the same results were also observed in IHC (**Figure 2F**, Cancer-adjacent vs. Cancer, $P = 0.010$).

We further analyzed THBS2 mRNA expression in GC and cancer-adjacent tissues RNA sequencing data from GSE13911 and GSE54129. The results showed that the THBS2 mRNA expression level was higher in GC than in cancer-adjacent tissues (**Figure 3A, B**, $p < 0.001$). The result of the ROC curve demonstrated that THBS2 had a high predictive value in distinguishing GC from cancer-adjacent tissues (**Figure 3C**, GSE13911, AUC = 0.915, 95%CI: 0.843–0.987; **Figure 3D**, GSE54129, AUC = 0.979, 95%CI: 0.957–1.000).

TABLE 3 | Details of GO and KEGG enrichment analyses.

Ontology	ID	Description	Gene ratio	Bg ratio	P Value	p.adjust	q value
BP	GO:0030198	extracellular matrix organization	49/329	368/18,670	7.35e-29	2.25e-25	1.97e-25
BP	GO:0043062	extracellular structure organization	51/329	422/18,670	5.45e-28	8.36e-25	7.30e-25
BP	GO:0043588	skin development	50/329	419/18,670	3.46e-27	3.53e-24	3.09e-24
BP	GO:0070268	cornification	29/329	112/18,670	6.02e-26	4.62e-23	4.03e-23
BP	GO:0030216	keratinocyte differentiation	41/329	305/18,670	2.15e-24	1.32e-21	1.15e-21
CC	GO:0062023	collagen-containing extracellular matrix	63/352	406/19,717	9.53e-41	2.47e-38	2.24e-38
CC	GO:0001533	cornified envelope	14/352	65/19,717	6.92e-12	8.96e-10	8.12e-10
CC	GO:0005581	collagen trimer	15/352	87/19,717	3.64e-11	3.15e-09	2.85e-09
CC	GO:0044420	extracellular matrix component	12/352	51/19,717	7.39e-11	4.78e-09	4.34e-09
CC	GO:0005583	fibrillar collagen trimer	7/352	11/19,717	1.69e-10	7.30e-09	6.61e-09
MF	GO:0005201	extracellular matrix structural constituent	39/316	163/17,697	5.72e-33	2.43e-30	2.06e-30
MF	GO:0030020	extracellular matrix structural constituent conferring tensile strength	11/316	41/17,697	9.70e-11	1.50e-08	1.27e-08
MF	GO:0005539	glycosaminoglycan binding	22/316	229/17,697	1.54e-10	1.50e-08	1.27e-08
MF	GO:0005518	collagen binding	13/316	67/17,697	1.57e-10	1.50e-08	1.27e-08
MF	GO:0048018	receptor ligand activity	32/316	482/17,697	1.77e-10	1.50e-08	1.27e-08
KEGG	hsa04974	Protein digestion and absorption	17/147	103/8,076	3.14e-12	6.47e-10	5.92e-10
KEGG	hsa05150	Staphylococcus aureus infection	11/147	96/8,076	1.17e-06	1.21e-04	1.10e-04
KEGG	hsa04512	ECM-receptor interaction	9/147	88/8,076	2.89e-05	0.002	0.002
KEGG	hsa04972	Pancreatic secretion	9/147	102/8,076	9.32e-05	0.004	0.004
KEGG	hsa05146	Amoebiasis	9/147	102/8,076	9.32e-05	0.004	0.004

DEGs Identification and Functional Enrichment Analysis of DEGs

A total of 599 DEGs were identified, including 170 up-regulated and 429 down-regulated DEGs (**Figure 4A**).

Enriched biological processes (BP), cellular components (CC), and molecular function (MF) were used to comprehend the biological functions of DEGs better. The top three items enriched in BP, CC, MF, and KEGG of DEGs were visualized. As shown in **Figure 4B**, THBS2-related genes were involved in KEGG (Protein digestion and absorption, Staphylococcus aureus infection, and ECM-receptor interaction); MF (extracellular matrix structural constituent, extracellular matrix structural constituent conferring tensile strength, and glycosaminoglycan binding); CC (collagen-containing extracellular matrix, cornified envelope, and collagen trimer); and BP (extracellular matrix organization, extracellular structure organization, and skin development). The details of GO and KEGG enrichment analysis results are illustrated in **Table 3**.

GSEA analysis revealed that THBS2-related enrichment pathways were as follows: focal-adhesion, vascular endothelial growth factor (VEGF) signaling, Wnt signaling, immunoregulatory lymphoid, and a non-lymphoid cell, senescence, and autophagy in cancer (**Figures 4C–H**).

THBS2 Expression is Closely Related to the Immune cell Infiltration and Clinicopathological Variables of Gastric Cancer

An analysis based on online data showed that the higher the expression of THBS2, the more macrophages, NK cells, iDC, and eosinophils infiltrate, and the fewer Th17 cells, T helper cells, and NK CD56 bright cells infiltrate in GC tissues. (**Figures 5A–G**). We further verified the differential enrichment of macrophages and NK cells in the tumor and cancer-adjacent tissues by immunohistochemistry. The proportion of positive area of CD56 (NK cells) and CD68 (macrophages) in GC tissues was significantly higher than that in cancer-adjacent tissues (**Figure 6**; CD56, Cancer vs. Cancer-adjacent, $P < 0.0001$; CD68, Cancer vs. Cancer-adjacent, $P < 0.0001$). The proportion of the positive area of CD56 and CD68 in the THBS2 high expression group was significantly higher than that in the low expression group (**Figure 7**; CD56, High vs. Low, $P < 0.01$; CD68, High vs. Low, $P < 0.0001$). Finally, we verified that the proportion of CD56 and CD68 positive area was significantly correlated with THBS2 expression (**Figure 7**; CD56, r = 0.636, $P < 0.01$; CD68, r = 0.466, $P < 0.05$).

There were 375 patients with complete clinicopathologic data in the TCGA database who were included in the cohort. As illustrated in **Figures 8A–D**, the high expression of THBS2 was significantly correlated with the pathological grade (stage I vs. stage II and III & IV, $P < 0.01$), histological grade (G1 & G2 vs. G3, $P < 0.05$), histological type (Diffuse Type vs. Tubular Type, $P < 0.05$), and T stage (T1 vs. T2, T1 vs. T3, T1 vs. T4, $P < 0.001$) of patients. Moreover, there was no significant statistical correlation between the THBS2 expression level and N and M stages (**Figures 8E,F**).

We further analyzed the relationship between THBS2 expression and clinicopathological data of 80 patients who underwent THBS2 immunohistochemistry. The results showed that THBS2 expression was significantly correlated with pathologic stage ($P = 0.044$), T stage (P = 0.003), histologic type ($p < 0.001$), and histological grade ($P = 0.027$) in 80 GC patients who underwent THBS2 immunohistochemistry (**Table 4**).

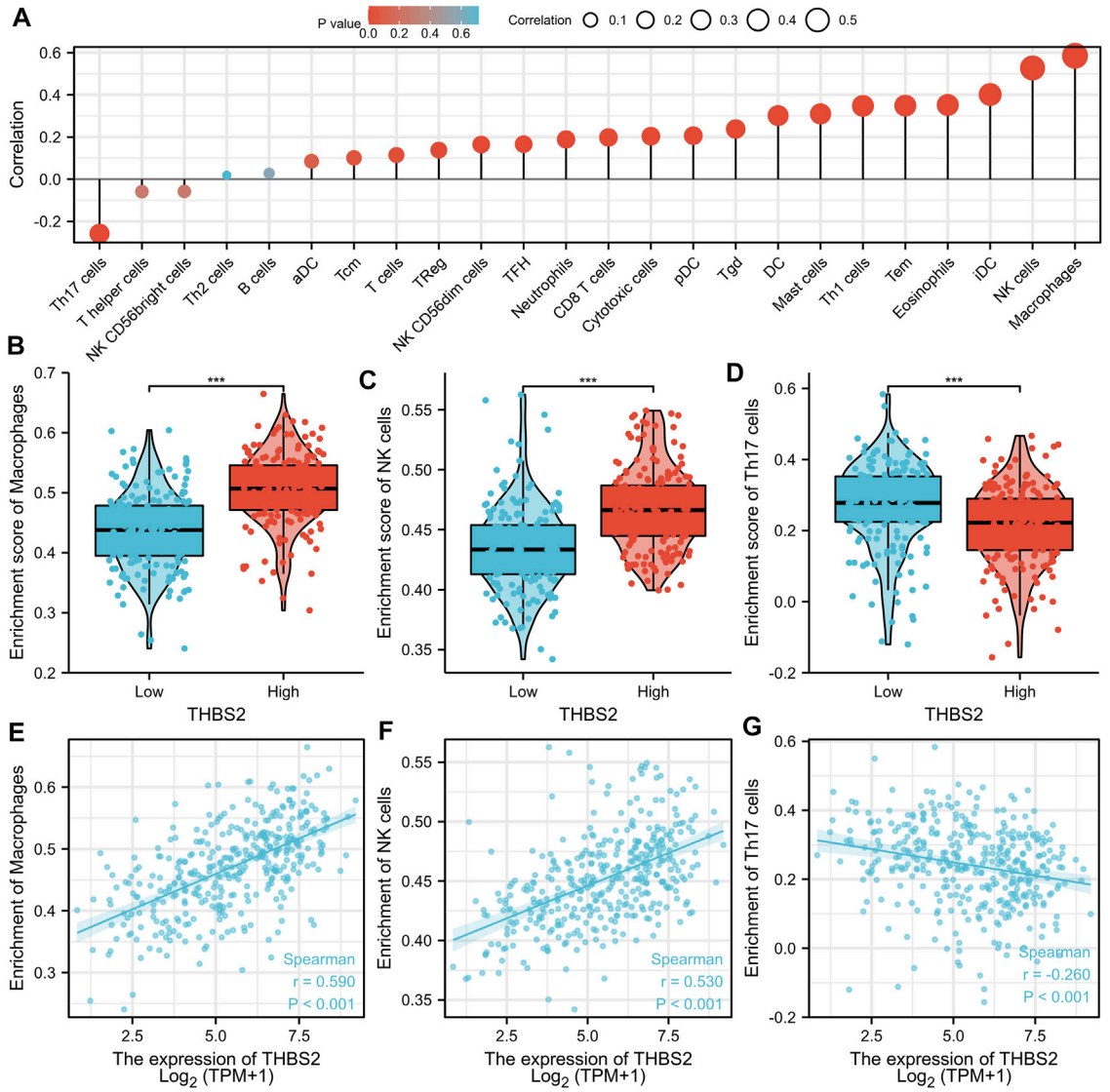


FIGURE 5 | The Correlation Between THBS2 Expression and Immune Infiltration. **(A)** Correlation between the relative abundances of immune cells and THBS2 expression level. The size of dots is positively related to the absolute value of Spearman R. **(B–D)** The difference of immune cells (Macrophages, NK cells, and Th17 cells) between the high and low expression groups based on the median value of THBS2 expression. **(E–G)** The correlation of immune cells (Macrophages, NK cells, and Th17 cells) between the high and low expression groups based on median value of THBS2 expression.

High Expression of THBS2 Was Associated With Poor Prognosis of GC

As can be seen from **Figure 9A**, the high THBS2 expression was significantly correlated with poor OS ($P = 0.003$). Moreover, high THBS2 expression was also associated with worse PFI and DSS; however, the difference was not statistically significant (**Figures 9B,C**).

The subgroup survival analysis showed that the prognosis of GC patients with THBS2 high expression was poor in T3 & T4 (HR = 1.56 [1.08–2.66]), N1 & N2 & N3 (HR = 1.63 [1.12–2.38]), M0 (HR = 1.67 [1.15–2.44]), and stage I and II (HR = 1.84 [1.00–3.39]) subgroups in OS, and stage III and IV (HR = 1.62 [1.00–2.26]) subgroup in PFI (**Figures 9D–I**). However, there was no significant difference among the subgroups of DSS.

A multivariate Cox regression analysis including age, primary therapy outcome, THBS2 expression level, pathologic stage, histologic grade, sex, race, and histological type was performed to eliminate the effect of confounders on univariate Cox regression analysis. Multivariate Cox regression analysis showed that age (>65 vs. ≤65 years, HR [95% CI] = 1.671 [1.127–2.476], $P = 0.011$), primary therapy outcome (CR vs. PD & SD & PR, HR [95% CI] = 4.715 [3.151–7.065]; $P < 0.001$), THBS2 (high vs. low, HR [95% CI] = 1.534 [1.037–2.270], $P = 0.032$), pathologic stage (stage I and II vs. III and IV, HR [95% CI] = 1.518 [1.008–2.285], $p =$

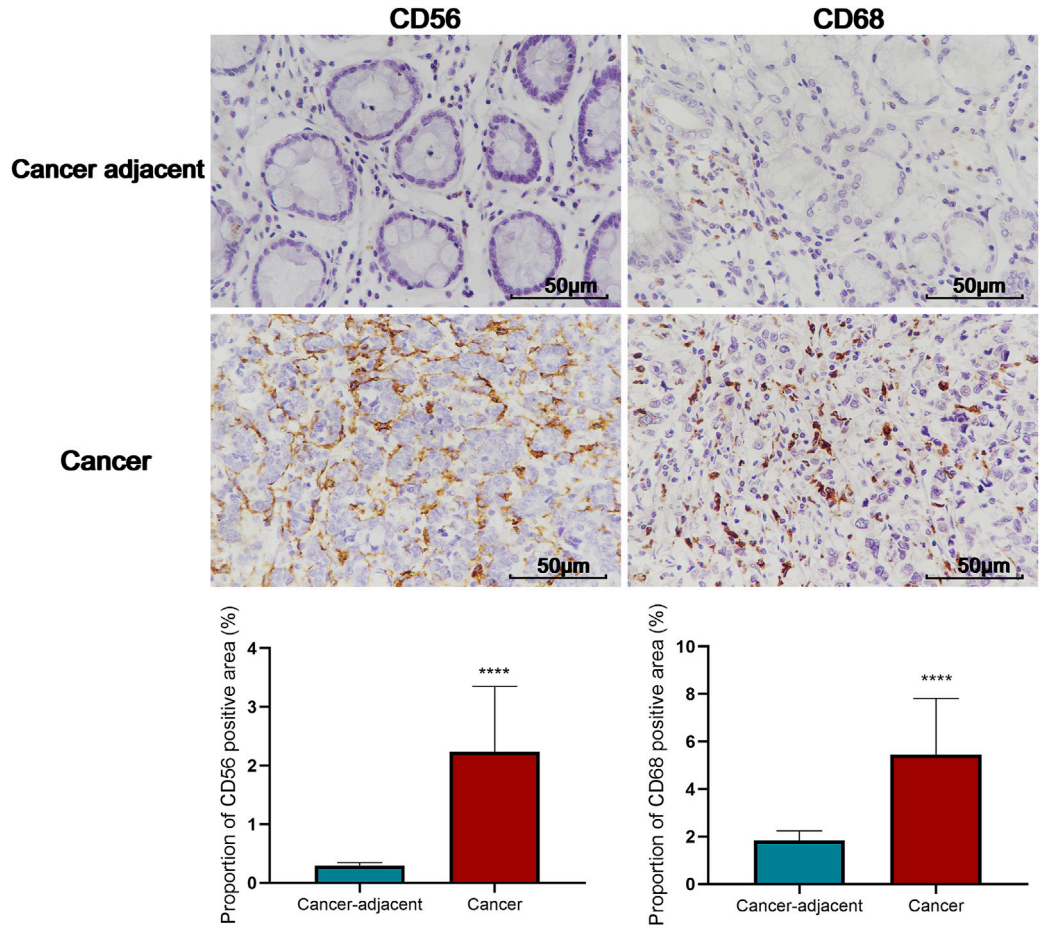


FIGURE 6 | Comparison of the proportion of CD56 and CD68 positive area in tumor and cancer-adjacent tissues (400X). C-D: Comparison of the proportion of CD56 and CD68 positive area in THBS2 high and low expression groups; E-F: Correlation between THBS2 expression and the proportion of CD56/CD68 positive area. ****, $p < 0.0001$.

0.045), and histologic grade (G1 & G2 vs. G3, HR [95% CI] = 1.547 [1.019–2.348], p = 0.040) were significantly correlated with OS in patients with GC (**Table 5**). However, multivariate Cox regression indicated that THBS2 expression levels had no association with poor DSS and DSS PFI (**Tables 6** and **7**).

Construction and Validation of Nomogram

All parameters persisting as independent predictors in the multivariate Cox models of the subgroups, including age, primary therapy outcome, THBS2 expression, pathological stage, and histologic grade, were integrated into the nomogram (**Figure 10**). The nomogram and calibration blots demonstrated that the nomogram-predicted 1-, 3-, and 5-years survival probabilities of GC were similar to the actual probabilities (C-index [95%CI] = 0.725 [0.701–0.750]), indicating that the prediction was in good agreement with the actual survival probability of GC (**Figure 10B**). These results indicated that the prediction ability of the nomogram for the survival probabilities might be clinically applicable also suggests that the nomogram was well-calibrated, with the mean predicted probabilities close to observed probabilities.

DISCUSSION

By analyzing the expression profile of GC patients in the open online database, we found that the expression level of THBS2 in GC tissues was significantly higher than that in cancer-adjacent tissues. There was also a significant difference in THBS2 expression in GC tissues and corresponding adjacent tissues, which was validated using RT-PCR and IHC. Additionally, the ROC curve demonstrated that THBS2 had a potential predictive value in distinguishing GC from cancer-adjacent tissues (AUC = 0.864, 95% CI: 0.812–0.915). This observation suggests that THBS2 may serve as a potential indicator for GC diagnosis.

Totally, 599 DEGs based on THBS2 expression levels were screened, including 170 up-regulated and 429 down-regulated DEGs. The results of GO enrichment analysis showed that the

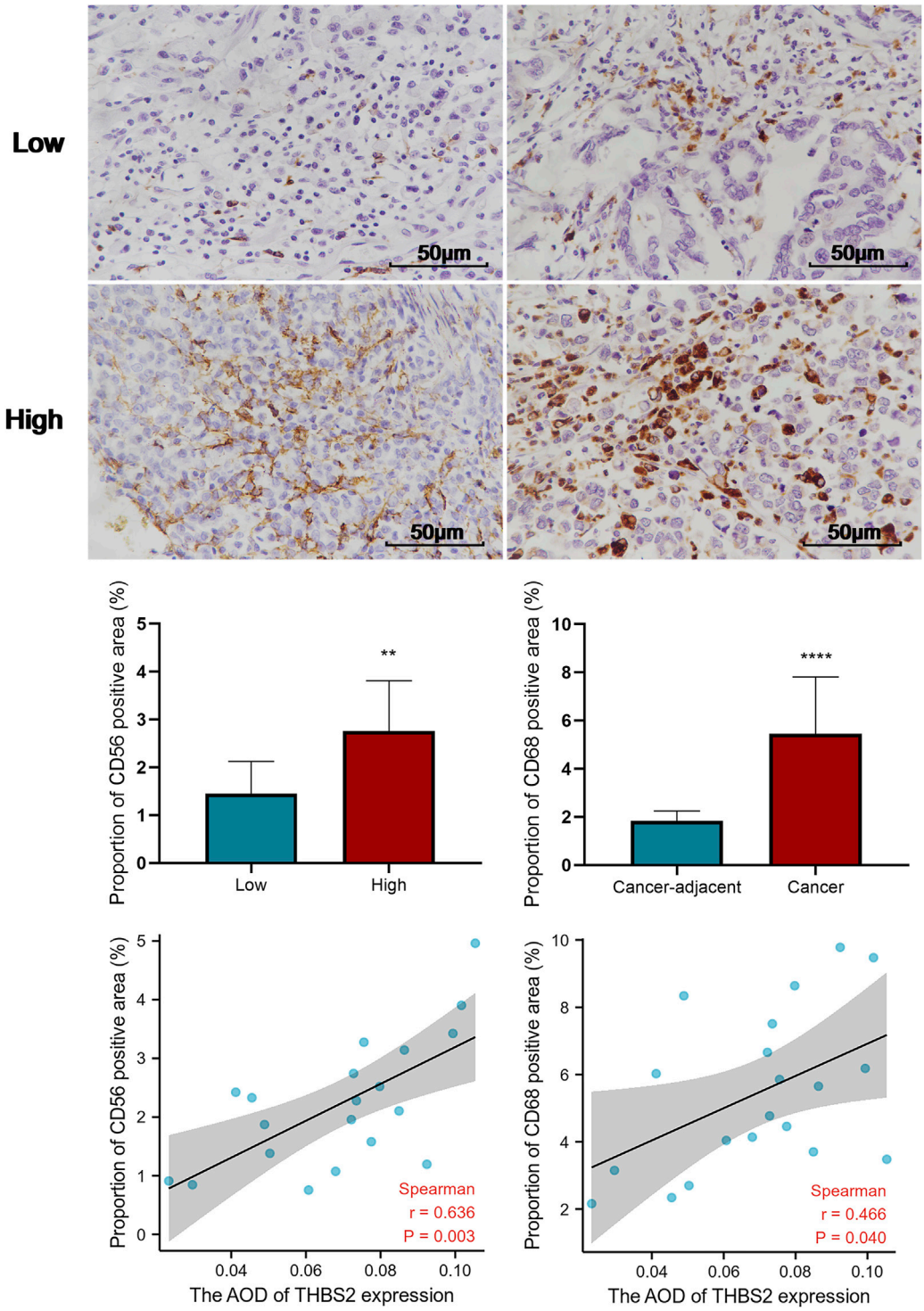


FIGURE 7 | Comparison of the proportion of CD56 and CD68 positive area in THBS2 high and low expression groups, and the correlation between THBS2 expression and the proportion of CD56/CD68 positive area (400X). AOD, average optical density; **, $p < 0.01$; ****, $p < 0.0001$.

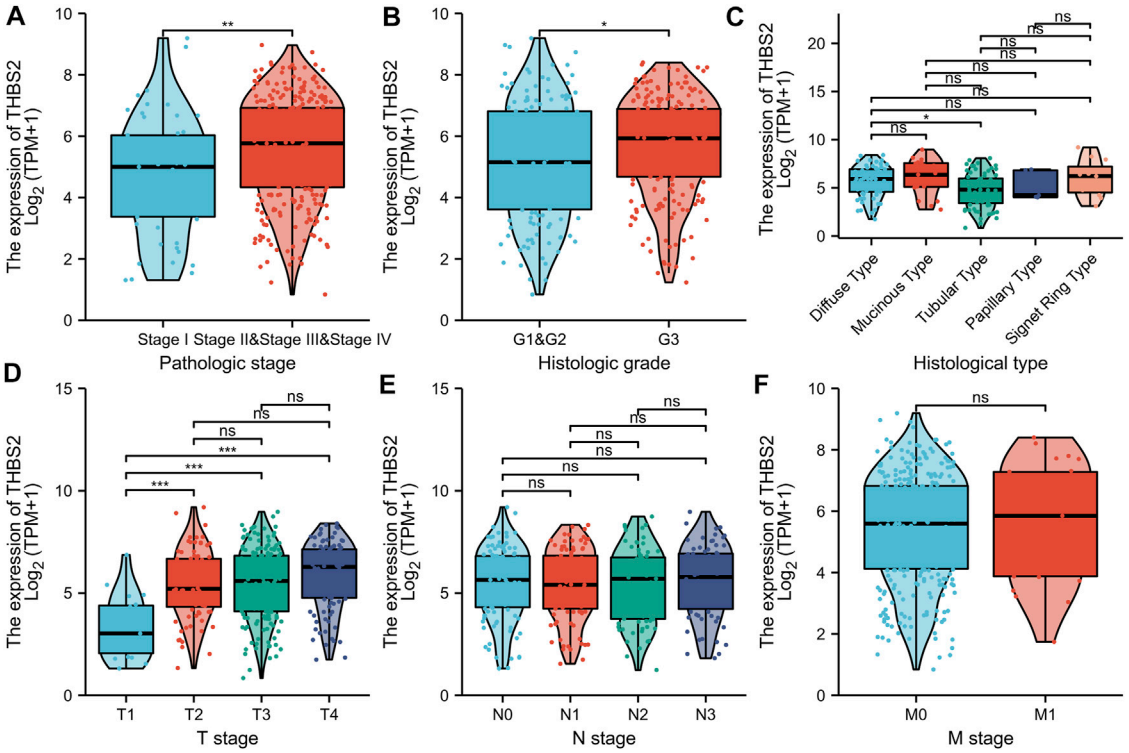


FIGURE 8 | Association with THBS2 expression and clinicopathological characteristics. **(A)** pathologic stage, **(B)** histological grade, **(C)** histologic type, **(D)** T stage **(E)** N stage, and **(F)** M stage in GC patients in TCGA cohort. TCGA, The Cancer Genome Atlas; GC, gastric cancer.

enrichment was primarily related to collagen and extracellular matrix (ECM). In the course of cancer progression, there are apparent changes in ECM. Collagen together with laminin and fibronectin constitutes a key substrate for the growth and migration of cancer cells as a major component of the ECM (Bonnans et al., 2014). Similarly, studies have also demonstrated that the ECM is an important component of all cancer markers that play a crucial role (Andreuzzi et al., 2020). ECM receptor interactions are critical in the tumor microenvironment (Chen et al., 2021), tumorigenesis, and tumor progression (Malik et al., 2015). The ECM protein regulates the metastasis of GC cells through the ITGB4/FAK/SOX2/HIF-1α signaling pathway induced by ECM receptor interaction (Gan et al., 2018). For KEGG pathway enrichment analysis, the DEGs are enriched in ECM-receptor interaction during S. aureus infection. The structural function of ECM is critical to maintaining normal cell activity (Giussani et al., 2019). ECM promotes tumorigenesis and progression by avoiding apoptosis, regulating cell growth, promoting tumor angiogenesis, and acquiring the ability of invasion and metastasis (Pickup et al., 2014; Malik et al., 2015; Poltavets et al., 2018; Eble and Niland, 2019). Signaling pathways associated with protein digestion and absorption are also critical for tumor progression (Mo et al., 2020). Therefore, treatment targeting ECM and tumor angiogenesis may be effective in preventing metastasis and recurrence of GC (Gan et al., 2018). Additionally, GSEA analysis demonstrated that THBS2-related enrichment pathways were as follows: focal-adhesion, VEGF signaling, Wnt signaling, cytokine-cytokine-receptor interaction, immunoregulatory lymphoid and a non-lymphoid cell, senescence, and autophagy in cancer. These pathways are closely related to cell adhesion, immune regulation, tumor autophagy, tumor progression, and tumor angiogenesis, respectively. Adhesion to ECM via specific focal adhesion points is an important step in cancer cell migration and invasion (Paluch et al., 2016; Ringer et al., 2017). VGFR related signaling pathway is important in the molecular pathogenesis of tumor growth and metastasis (Adamis and Shima, 2005). THBS2 may play a crucial role in the occurrence and progression of GC, so it is reasonable to speculate that THBS2 may be a potential therapeutic target for GC.

Previous studies have confirmed that the interaction between tumor-infiltrating immune and malignant cells leads to the immune invasion of tumors, as the immune system plays a dual role by supporting both tumor progression and host defense (Bremnes et al., 2011). This study showed that the abundance of macrophages and NK cells in the tumor microenvironment of GC with high THBS2 expression was higher than that of GC with low THBS2 expression, while the abundance of Th17 cells was lower. Single tumor-initiating cells were detected to recruit polarized M2-like macrophages and assist evasion from immune clearance (Guo

TABLE 4 | The relationship between THBS2 expression and clinicopathological variables in 80 patients underwent immunohistochemistry.

Characteristic	Low	High	p
n	38	42	
Gender (M/F), n (%)			0.634
F	9 (11.2%)	13 (16.2%)	
M	29 (36.2%)	29 (36.2%)	
Pathologic stage, n (%)			**0.044**
I	10 (12.5%)	3 (3.8%)	
II and III	28 (35%)	39 (48.8%)	
T stage, n (%)			**0.003**
T1	11 (13.8%)	2 (2.5%)	
T2	6 (7.5%)	3 (3.8%)	
T3	21 (26.2%)	34 (42.5%)	
T4	0 (0%)	3 (3.8%)	
N stage, n (%)			0.554
N0	11 (13.8%)	10 (12.5%)	
N1	10 (12.5%)	7 (8.8%)	
N2	7 (8.8%)	12 (15%)	
N3	10 (12.5%)	13 (16.2%)	
Histological type, n (%)			< **0.001**
Diffuse Type	5 (6.2%)	23 (28.7%)	
Mucinous Type	1 (1.2%)	5 (6.2%)	
Papillary Type	7 (8.8%)	4 (5%)	
Signet Ring Type	8 (10%)	6 (7.5%)	
Tubular Type	17 (21.2%)	4 (5%)	
Histological grade, n (%)			**0.027**
G1 & G2	23 (28.7%)	14 (17.5%)	
G3	15 (18.8%)	28 (35%)	
Residual tumor, n (%)			0.776
R0	26 (32.5%)	31 (38.8%)	
R1 & R2	12 (15%)	11 (13.8%)	
Primary therapy outcome, n (%)			0.515
CR	25 (31.2%)	34 (42.5%)	
PD	8 (10%)	5 (6.2%)	
PR	2 (2.5%)	1 (1.2%)	
SD	3 (3.8%)	2 (2.5%)	
Anatomic neoplasm subdivision, n (%)			0.151
Antrum	10 (12.5%)	6 (7.5%)	
Cardia	15 (18.8%)	13 (16.2%)	
Fundus/Body	13 (16.2%)	20 (25%)	
other	0 (0%)	3 (3.8%)	
Age (years), meidan (IQR)	63 (57.25, 71.5)	66 (59, 70.75)	0.606

et al., 2017). Moreover, macrophages can promote tumor invasion and metastasis (Doak et al., 2018). In some tumor types such as ovarian, prostate, and colorectal cancer, Th17 cells induce an antitumor immune response by recruiting cytotoxic effector T cells and producing effector cytokines, including interferon-gamma (Kryczek et al., 2009; Tosolini et al., 2011). Therefore, we speculated that the high expression of THBS2 might affect the disease progression and prognosis of patients with GC by inducing changes in the tumor microenvironment.

THBS2, a member of the thrombospondin family of multidomain and secreted multicellular calcium-binding glycoproteins, mediates cell-to-cell and cell-to-matrix interactions (Ng et al., 2021). The expression level of THBS2 in colon cancer was significantly increased, and the higher the expression level of THBS2, the worse the OS of patients (Wang et al., 2016b). The high expression of THBS2 promotes the metastasis of colon cancer and is associated with an advanced clinical stage (Wang et al., 2016b). Our results suggest that the high expression of THBS2 was significantly correlated with the pathological grade (stage I vs. stage II and III & IV, $P < 0.01$), histological grade (G1 & G2 vs. G3, $p < 0.05$), histological type (Diffuse Type vs. Tubular Type, $P < 0.05$), T stage (T1 vs. T2, T1 vs. T3, T1 vs. T4, $P < 0.001$) of patients. The high THBS2 expression was significantly correlated with poor OS. Additionally, the subgroup survival analysis demonstrated that the prognosis of GC patients with THBS2 high expression was poor in T3 & T4, N1 & N2 & N3, M0, and stage I and II subgroups in OS, and stage III and IV subgroups in PFI. This study also indicates that the high expression of THBS2 may accelerate GC progression, leading to a poor prognosis. The higher the expression level of THBS2, the worse the degree of differentiation of GC.

Multivariate Cox regression analysis demonstrated that age, primary therapy outcome, THBS2 expression, pathological

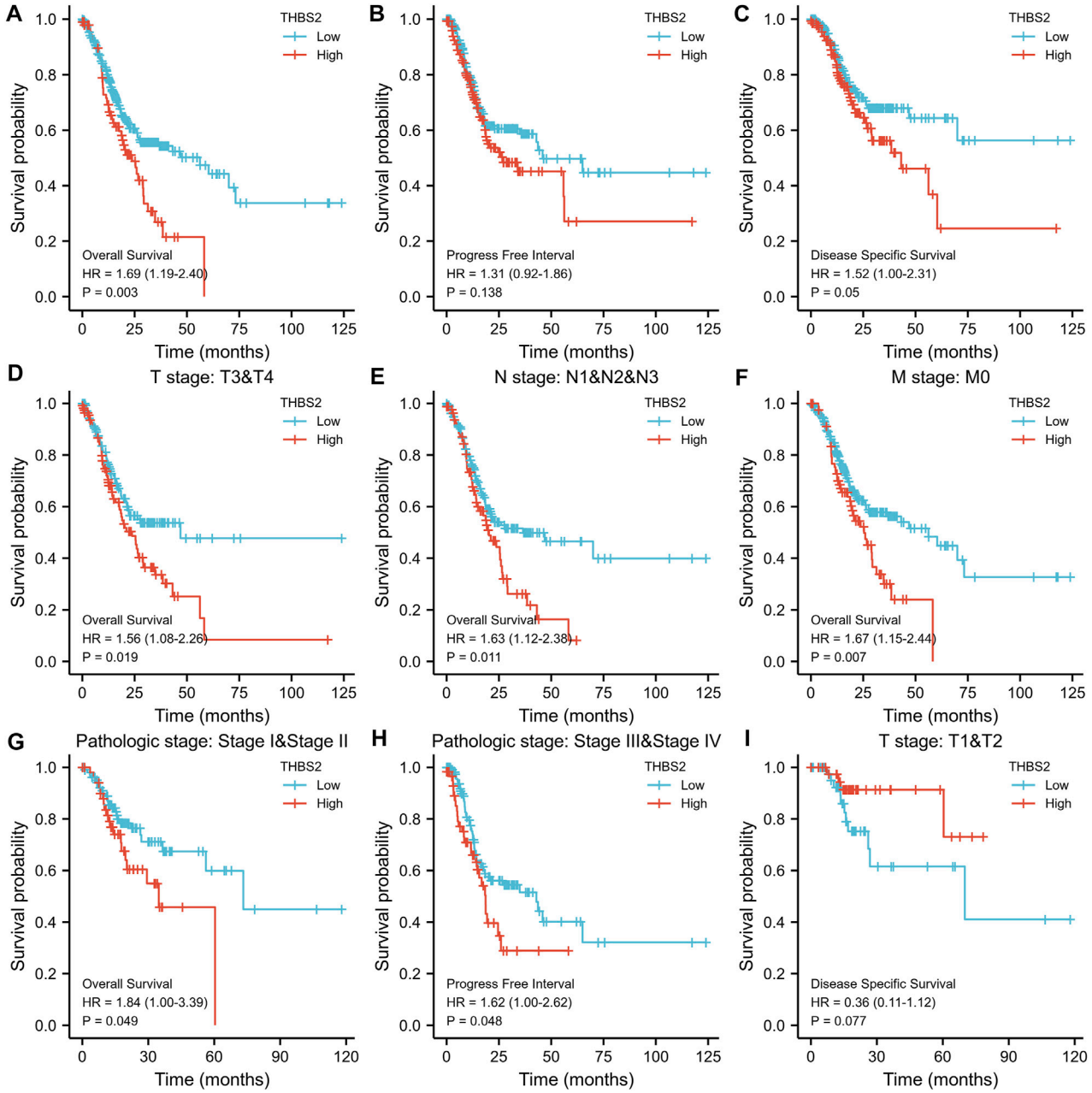


FIGURE 9 | Kaplan-Meier survival curves comparing the high and low expression of THBS2 in GC. **(A–C)** Survival curves of OS, DSS, and PFI between THBS2-high and-low patients with GC. **(D–G)** OS survival curves of T3&T4, N1&N2&N3, M0, and stage IandII subgroups between THBS2-high and -low patients with GC. **(H)** PFI survival curves of stage IIIandIV subgroup between THBS2-high and -low patients with GC. **(I)** DSS survival curves of T1&T2 subgroup between THBS2-high and -low patients with GC. GC, gastric cancer; OS, overall survival; DSS, disease specific survival; PFI, progression free interval.

stage, and histologic grade were independent risk factors associated with poor OS in GC. A nomogram based on the above independent risk factors affecting OS of GC was also established to predict the survival probability of patients with GC. The survival probability of GC patients predicted by the constructed nomogram was similar to the actual survival probability, indicating that the created nomogram could be successfully used to predict the survival of GC patients. The accurate prediction of survival of cancer patients can provide essential information for individualized treatment of cancer patients (Erickson et al., 2014; Gratian et al., 2014). Our nomogram to predict the survival probability of GC has been verified using a calibration blot and should be clinically promoted.

TABLE 5 | Univariate and multivariate regression analysis of Overall Survival (OS) related factors in patients with GC.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P Value	Hazard ratio (95% CI)	P Value
Age	367				
≤7 al	164	Reference			
>65	207	1.620 (1.154–2.276)	**0.005**	1.671 (1.127–2.476)	**0.011**
Primary.therapy.outcome	313				
CR	231	Reference			
PD&SD&PR	86	4.228 (2.905–6.152)	<**0.001**	4.715 (3.151–7.056)	<**0.001**
THBS2	370				
Low	188	Reference			
High	187	1.330 (0.956–1.851)	0.091	1.534 (1.037–2.270)	**0.032**
Pathologic.stage	347				
Stage I&Stage II	164	Reference			
Stage III&Stage IV	188	1.947 (1.358–2.793)	<**0.001**	1.518 (1.008–2.285)	**0.045**
Histologic.grade	361				
G1&G2	147	Reference			
G3	219	1.353 (0.957–1.914)	0.087	1.547 (1.019–2.348)	**0.040**
Gender	370				
Female	134	Reference			
Male	241	1.267 (0.891–1.804)	0.188		
Race	320				
White	238	Reference			
Asian&Black or African American	85	0.801 (0.515–1.247)	0.326		
Histological.type	132				
Diffuse Type	63	Reference			
Tubular Type	69	0.929 (0.534–1.614)	0.793		
Anatomic.neoplasm.subdivision	353				
Cardia/Proximal&Gastroesophageal Junction	89	Reference			
Antrum/Distal&Fundus/Body	268	0.919 (0.628–1.345)	0.663		

TABLE 6 | Univariate and multivariate regression analysis of Disease Specific Survival (DSS) related factors in patients with GC.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P Value	Hazard ratio (95% CI)	P Value
Age	346				
≤6 al	164	Reference			
>65	207	1.211 (0.797–1.840)	0.371		
Primary.therapy.outcome	310				
CR	231	Reference			
PD&SD&PR	86	8.697 (5.439–13.908)	<**0.001**	8.590 (5.270–14.000)	<**0.001**
THBS2	349				
Low	188	Reference			
High	187	1.195 (0.786–1.815)	0.405		
Pathologic.stage	331				
Stage I&Stage II	164	Reference			
Stage III&Stage IV	188	2.146 (1.352–3.404)	**0.001**	1.439 (0.896–2.312)	0.132
Histologic.grade	340				
G1&G2	147	Reference			
G3	219	1.338 (0.862–2.078)	0.194		
Gender	349				
Female	134	Reference			
Male	241	1.573 (0.985–2.514)	0.058	1.228 (0.753–2.002)	0.410
Race	305				
White	238	Reference			
Asian&Black or African American	85	1.097 (0.656–1.836)	0.724		
Histological.type	129				
Diffuse Type	63	Reference			
Tubular Type	69	0.897 (0.480–1.676)	0.734		
Anatomic.neoplasm.subdivision	337				
Cardia/Proximal&Gastroesophageal Junction	89	Reference			
Antrum/Distal&Fundus/Body	268	0.767 (0.485–1.215)	0.259		

TABLE 7 | Univariate and multivariate regression analysis of Progress Free Interval (PFI) related factors in patients with GC.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P Value	Hazard ratio (95% CI)	P Value
Age	369				
≤9 al	164	Reference			
>65	207	0.858 (0.603–1.221)	0.395		
Primary.therapy.outcome	315				
CR	231	Reference			
PD&SD&PR	86	8.041 (5.465–11.832)	**<0.001**	8.704 (5.733–13.215)	**<0.001**
THBS2	372				
Low	188	Reference			
High	187	1.149 (0.807–1.637)	0.441		
Pathologic.stage	349				
Stage I&Stage II	164	Reference			
Stage III&Stage IV	188	1.676 (1.154–2.435)	**0.007**	1.037 (0.696–1.544)	0.858
Histologic.grade	363				
G1&G2	147	Reference			
G3	219	1.540 (1.057–2.245)	**0.025**	1.799 (1.186–2.730)	**0.006**
Gender	372				
Female	134	Reference			
Male	241	1.638 (1.099–2.440)	**0.015**	1.361 (0.893–2.077)	0.152
Race	322				
White	238	Reference			
Asian&Black or African American	85	1.061 (0.688–1.637)	0.787		
Histological.type	132				
Diffuse Type	63	Reference			
Tubular Type	69	0.806 (0.466–1.392)	0.438		
Anatomic.neoplasm.subdivision	354				
Cardia/Proximal&Gastroesophageal Junction	89	Reference			
Antrum/Distal&Fundus/Body	268	0.696 (0.474–1.022)	0.064	0.802 (0.522–1.234)	0.316

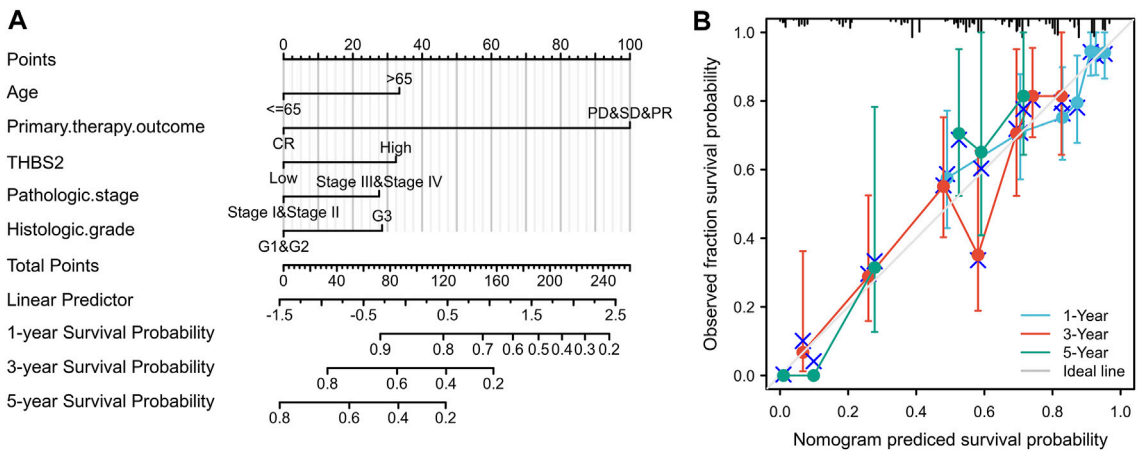


FIGURE 10 | A nomogram and calibration plot to predict GC patients' 1-, 3-, and 5-years OS. **(A)** A nomogram for predicting the 1-, 3-, and 5-years OS for GC patients. **(B)** Calibration plots of the nomogram. GC, gastric cancer; OS, overall survival.

CONCLUSION

Overall, the findings of the current research are summarized below:

Firstly, the THBS2 expression level in GC was significantly higher than that in para-cancer tissues, and THBS2 may be a potential biomarker for GC diagnosis. Secondly, THBS2 may affect the progression and prognosis of GC by changing the tumor microenvironment and may be a potential therapeutic target for GC. Thirdly, histologic grade, primary therapy outcome, age, and THBS2 might be independent risk factors associated with poor OS in GC. Finally, the nomogram may provide more individualized prognostic information for patients with GC. However, neither the TCGA data nor the

hospital data collected were sufficiently large. Therefore, more information should be collected to verify the accuracy of the results.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/http://xena.ucsc.edu/</b&gt.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GT contributed to the conception of the study; SZ performed the experiment and manuscript; HY contributed significantly to analysis and manuscript preparation; XX performed the data analyses and wrote the manuscript; LL and HH helped perform the analysis with constructive discussions.

FUNDING

This research was supported by the National Natural Science Foundation of China (81970558) and Guangxi Natural Science Foundation (2020GXNSFAA259095).

ACKNOWLEDGMENTS

Thanks to Xiaoyang Shi of The First People's Hospital of Suqian for his guidance of experimental technology.

REFERENCES

Adamis, A. P., and Shima, D. T. (2005). The role of vascular endothelial growth factor in ocular health and disease. *Retina* 25 (2), 111–118. doi:10.1097/00006982-200502000-00001

Ao, R., Guan, L., Wang, Y., and Wang, J. N. (2018). Retracted: Silencing of COL1A2, COL6A3, and THBS2 Inhibits Gastric Cancer Cell Proliferation, Migration, and Invasion while Promoting Apoptosis through the PI3k-Akt Signaling Pathway. *J. Cell. Biochem.* 119 (6), 4420–4434. doi:10.1002/jcb.26524

Andreuzzi, E., Capuano, A., Poletto, E., Pivetta, E., Fejza, A., Favero, A., et al. (2020). Role of Extracellular Matrix in Gastrointestinal Cancer-Associated Angiogenesis. *Int J Mol Sci* 21 (10). doi:10.3390/ijms21103686

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Nat Rev Mol Cell Biol* 15 (12), 786–801. doi:10.1038/nrm3904

Bornstein, P., Armstrong, L. C., Hankenson, K. D., Kyriakides, T. R., and Yang, Z. (2000). Thrombospondin 2, a Matricellular Protein with Diverse Functions. *Matrix Biol.* 19 (7), 557–568. doi:10.1016/s0945-053x(00)00104-9

Bremnes, R. M., Al-Shibli, K., Donnem, T., Sirera, R., Al-Saad, S., Andersen, S., et al. (2011). The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer. *J Thorac Oncol* 6 (4), 824–833. doi:10.1097/JTO.0b013e3182037b76

Cao, L., Chen, Y., Zhang, M., Xu, D.-q., Liu, Y., Liu, T., et al. (2018). Identification of Hub Genes and Potential Molecular Mechanisms in Gastric Cancer by Integrated Bioinformatics Analysis. *PeerJ* 6, e5180. doi:10.7717/peerj.5180

Chen, S., Wei, Y., Liu, H., Gong, Y., Zhou, Y., Yang, H., et al. (2021). Analysis of Collagen type X alpha 1 (COL10A1) expression and prognostic significance in gastric cancer based on bioinformatics. *Bioengineered* 12 (1), 127–137. doi:10.1080/21655979.2020.1864912

Choi, K. S., Jun, J. K., Suh, M., Park, B., Noh, D. K., Song, S. H., et al. (2015). Effect of Endoscopy Screening on Stage at Gastric Cancer Diagnosis: Results of the National Cancer Screening Programme in Korea. *Br. J. Cancer* 112 (3), 608–612. doi:10.1038/bjc.2014.608

Doak, G. R., Schwertfeger, K. L., and Wood, D. K. (2018). Distant Relations: Macrophage Functions in the Metastatic Niche. *Trends Cancer* 4 (6), 445–449. doi:10.1016/j.trecan.2018.03.011

Eble, J. A., and Niland, S. (2019). The extracellular matrix in tumor progression and metastasis. *Cell Physiol Biochem* 36 (3), 171–198. doi:10.1007/s10585-019-09966-1

Engelhardt, J. J., Sullivan, T. J., and Allison, J. P. (2006). CTLA-4 Overexpression Inhibits T Cell Responses through a CD28-B7-dependent Mechanism. *J. Immunol.* 177 (2), 1052–1061. doi:10.4049/jimmunol.177.2.1052

Erickson, L. A., Rivera, M., Zhang, J., et al. (2016). Adrenocortical carcinoma: review and update. *Adv Anat Pathol* 21 (3), 151–159. doi:10.1097/pap.0000000000000019

Fashoyin-Aje, L., Donoghue, M., Chen, H., He, K., Veeraraghavan, J., Goldberg, K. B., et al. (2019). FDA Approval Summary: Pembrolizumab for Recurrent Locally Advanced or Metastatic Gastric or Gastroesophageal Junction Adenocarcinoma Expressing PD-L1. *Oncologist* 24 (1), 103–109. doi:10.1634/theoncologist.2018-0221

Gan, L., Meng, J., Xu, M., Liu, M., Qi, Y., Tan, C., et al. (2016). Extracellular matrix protein 1 promotes cell metastasis and glucose metabolism by inducing integrin˜4/FAK/SOX2/HIF-1邂 signaling pathway in gastric cancer . *Oncogene* 37 (6), 744–755. doi:10.1038/onc.2017.363

Gratian, L., Pura, J., Dinan, M., Reed, S., Scheri, R., Roman, S., et al. (2014). Treatment patterns and outcomes for patients with adrenocortical carcinoma associated with hospital case volume in the United States. *Ann Surg Oncol* 21 (11), 3509–3514. doi:10.1245/s10434-014-3931-z

Guo, X., Zhao, Y., Yan, H., Yang, Y., Shen, S., Dai, X., et al. (2017). Single tumor-initiating cells evade immune clearance by recruiting type II macrophages. *Genes Dev* 31 (3), 247–259. doi:10.1101/gad.294348.116

Giussani, M., Triulzi, T., Sozzi, G., and Tagliabue, E. (2019). Tumor Extracellular Matrix Remodeling: New Perspectives as a Circulating Tool in the Diagnosis and Prognosis of Solid Tumors. *Cells* 8 (2). doi:10.3390/cells8020081

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Herrera-Almario, G., and Strong, V. E. (2016). Minimally Invasive Gastric Surgery. *Ann. Surg. Oncol.* 23 (12), 3792–3797. doi:10.1245/s10434-016-5429-3

Inokuchi, M., Nakagawa, M., Tanioka, T., Okuno, K., Gokita, K., and Kojima, K. (2018). Long- and Short-Term Outcomes of Laparoscopic Gastrectomy versus Open Gastrectomy in Patients with Clinically and Pathological Locally

Advanced Gastric Cancer: a Propensity-Score Matching Analysis. *Surg. Endosc.* 32 (2), 735–742. doi:10.1007/s00464-017-5730-7

Kryczek, I., Banerjee, M., Cheng, P., Vatan, L., Szeliga , W., Wei, S., et al. (2009). Phenotype, distribution, generation, and functional and clinical relevance of Th17 cells in the human tumor environments. *Blood* 111 (6), 1141–1149. doi:10.1182/blood-2009-03-208249

Kirkwood, J. M., Butterfield, L. H., Tarhini, A. A., Zarour, H., Kalinski, P., and Ferrone, S. (2012). Immunotherapy of Cancer in 2012. *CA: A Cancer J. Clinicians* 62 (5), 309–335. doi:10.3322/caac.20132

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173 (2), 400–e11. doi:10.1016/j.cell.2018.02.052

Malik, R., Lelkes, P. I., and Cukierman, E. (2016). Biomechanical and biochemical remodeling of stromal extracellular matrix in cancer. *Trends Biotechnol* 33 (4), 230–236. doi:10.1016/j.tibtech.2015.01.004

Mo, L., Wei, B., Liang, R., Yang, Z., Xie, S., Wu, S., et al. (2016). Exploring potential biomarkers for lung adenocarcinoma using LC-MS/MS metabolomics. *J Int Med Res* 48 (4), 300060519897215. doi:10.1177/0300060519897215

Ng, K. Y., Shea, Q. T., Wong, T. L., Luk, S. T, Tong, M., Lo, C. M., et al. (2021). Chemotherapy-Enriched THBS2-Deficient Cancer Stem Cells Drive Hepatocarcinogenesis through Matrix Softness Induced Histone H3 Modifications. *Adv Sci (Weinh)* 8 (5), 2002483. doi:10.1002/advs.202002483

Niccolai, E., Taddei, A., Prisco, D., and Amedei, A. (2015). Gastric Cancer and the Epoch of Immunotherapy Approaches. *Wjg* 21 (19), 5778–5793. doi:10.3748/wjg.v21.i19.5778

Park, J., Forman, D., Waskito, L., Yamaoka, Y., and Crabtree, J. (2018). Epidemiology of *Helicobacter pylori* and CagA-Positive Infections and Global Variations in Gastric Cancer. *Toxins* 10 (4), 163. doi:10.3390/toxins10040163

Paluch, E. K., Aspalter, I. M., and Sixt, M. (2016). Focal Adhesion-Independent Cell Migration. *Annu Rev Cell Dev Biol* 32, 469–490. doi:10.1146/annurev-cellbio-111315-125341

Pickup, M. W., Mouw, J. K., and Weaver, V. M. (2014). The extracellular matrix modulates the hallmarks of cancer . *EMBO Rep* 15 (12), 1243–1253. doi:10.15252/embr.201439246

Poltavets, V., Kochetkova, M., Pitson, S. M., and Samuel, M. S. (2018). The Role of the Extracellular Matrix and Its Molecular and Cellular Regulators in Cancer Cell Plasticity. *Front Oncol* 8, 431. doi:10.3389/fonc.2018.00431

Ribas, A., and Wolchok, J. D. (2018). Cancer Immunotherapy Using Checkpoint Blockade. *Science* 359 (6382), 1350–1355. doi:10.1126/science.aar4060

Ringer, P., Colo, G., F□ssler, R., and Grashoff, C. (2017). Sensing the mechano-chemical properties of the extracellular matrix. *Matrix Biol* 64, 6–16. doi:10.1016/j.matbio.2017.03.004

So, J. B. Y., Kapoor, R., Zhu, F., Koh, C., Zhou, L., Zou, R., et al. (2021). Development and Validation of a Serum microRNA Biomarker Panel for Detecting Gastric Cancer in a High-Risk Population. *Gut* 70 (5), 829–837. doi:10.1136/gutjnl-2020-322065

Sun, R., Wu, J., Chen, Y., Lu, M., Zhang, S., Lu, D., et al. (2014). Down Regulation of Thrombospondin2 Predicts Poor Prognosis in Patients with Gastric Cancer. *Mol. Cancer* 13, 225. doi:10.1186/1476-4598-13-225

Tosolini, M., Kirilovsky, A., Mlecnik, B., Fredriksen, T., Mauger, S., Bindea, G., et al. (2011). Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg, th17) in patients with colorectal cancer. *Cancer Res* 71 (4), 1263–1271. doi:10.1158/0008-5472.Can-10-2907

Tey, J., Soon, Y. Y., Koh, W. Y., Leong, C. N., Choo, B. A., Ho, F., et al. (2017). Palliative Radiotherapy for Gastric Cancer: a Systematic Review and Meta-Analysis. *Oncotarget* 8 (15), 25797–25805. doi:10.18632/oncotarget.15554

Wang, X., Zhang, L., Li, H., Sun, W., Zhang, H., and Lai, M. (2016a). THBS2 Is a Potential Prognostic Biomarker in Colorectal Cancer. *Sci. Rep.* 6, 33366. doi:10.1038/srep33366

Wang, X., Zhang, L., Li, H., Sun, W., Zhang, H., Lai, R., et al. (2016b). THBS2 is a Potential Prognostic Biomarker in Colorectal Cancer. *Sci Rep* 6, 33366. doi:10.1038/srep33366

Weng, T.-Y., Wang, C.-Y., Hung, Y.-H., Chen, W.-C., Chen, Y.-L., and Lai, M.-D. (2016). Differential Expression Pattern of THBS1 and THBS2 in Lung Cancer: Clinical Outcome and a Systematic-Analysis of Microarray Databases. *PLoS One* 11 (8), e0161007. doi:10.1371/journal.pone.0161007

Zhandossov, O., Kaussova, G., and Koten, A. (2018). Combined Treatment for Gastric Cancer: Immunological Approach. *Turk J. Gastroenterol.* 29 (2), 151–156. doi:10.5152/tjg.2018.17398

Zhuo, C., Li, X., Zhuang, H., Tian, S., Cui, H., Jiang, R., et al. (2016). Elevated THBS2, COL1A2, and SPP1 Expression Levels as Predictors of Gastric Cancer Prognosis. *Cell Physiol Biochem* 40 (6), 1316–1324. doi:10.1159/000453184

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Yang, Xiang, Liu, Huang and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 February 2022
doi: 10.3389/fgene.2021.813793

Effects of miR-103a-3p Targeted Regulation of TRIM66 Axis on Docetaxel Resistance and Glycolysis in Prostate Cancer Cells

Qiang Yi, Junfeng Wei and Yangzhou Li*

Department of Urology, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, China

Objective: We aimed to study the expressions of miR-103a-3p and TRIM66 in prostate cancer (PCa) cells, explore the direct target genes of miR-103a-3p, and analyze the effects of miR-103a-3p targeted regulation of the TRIM66 axis on docetaxel (DTX) resistance and glycolysis of PCa cells.

Methods: Human normal prostate cells and PCa cells were used to detect the expressions of miR-103a-3p and TRIM66 and analyze their relationship. DTX-resistant (DR) PCa cells were established and transfected with miR-103a-3p and TRIM66 plasmids. The MTT assay, the plate cloning assay, the wound healing assay, and the Transwell assay were used to detect cell viability, colony formation, cell migration, and cell invasion, respectively. Cell glycolysis was analyzed using a cell glycolysis kit.

Results: The expression of miR-103a-3p was low and that of TRIM66 was high in PCa cells. MiR-103a-3p had a binding site with TRIM66, and the double luciferase report confirmed that they had a targeting relationship. Compared with the PCa group cells, the DTX-resistant group cells showed increased resistance to DTX. The resistance index was 13.33, and the doubling time of the DTX-resistant group cells was significantly longer than that of the PCa group cells. The DTX-resistant group showed more obvious low expression of miR-103a-3p and high expression of TRIM66. After the DTX-resistant group cells were transfected with miR-103a-3p and TRIM66 plasmids, the expression of miR-103a-3p increased significantly and that of TRIM66 decreased significantly. Upregulation of miR-103a-3p and interference with TRIM66 can inhibit the proliferation, metastasis, and glycolysis of DTX-resistant cells.

Conclusion: The expression of miR-103a-3p was downregulated and that of TRIM66 was upregulated in the malignant progression of PCa, especially during DTX resistance. Upregulation of miR-103a-3p and interference with TRIM66 can inhibit DTX resistance and glycolysis of PCa cells. Targeting TRIM66 may provide potential application value in molecular therapy for PCa.

Keywords: prostate cancer, docetaxel resistance, glycolysis, MiR-103a-3p, TRIM66

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Jianhua Xiao,
University of South China, China
Yang Liu,
The First Medical Center of Chinese PLA General Hospital, China
Chengke Luo,
Central South University, China

****Correspondence:***
Qiang Yi
yiqiangwcc@163.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *12 November 2021*
Accepted: *27 December 2021*
Published: *08 February 2022*

Citation:
Yi Q, Wei J and Li Y (2022) Effects of miR-103a-3p Targeted Regulation of TRIM66 Axis on Docetaxel Resistance and Glycolysis in Prostate Cancer Cells. Front. Genet. 12:813793. doi: 10.3389/fgene.2021.813793

INTRODUCTION

Prostate cancer (PCa) is one of the fatal malignant tumors and has become the second leading cause of adult male death. According to the cancer statistics released by China in 2015, PCa accounted for about 2.4% of new cancers in men, including about 60,300 new cases and 26,600 deaths (Prendeville et al., 2017; Davidson et al., 2018). In 2018, PCa comprised about one-fifth of the total number of new cancers in American men, with a total of about 164,649 new cases and 29,430 deaths (Abouhashem and Salah, 2020). About one-third of cases experience disease recurrence, disease progression, or eventually develop into metastatic disease after initial treatment. Unfortunately, in patients with metastatic PCa, the 5-year survival rate decreased to 30%. Although there is increasing evidence that some patients with specific metastatic diseases may benefit from local treatment, such as radiotherapy and surgery (Zhou, 2018; Zupancic et al., 2020), androgen deprivation therapy (ADT) is still the main method for treatment of metastatic PCa (Lu et al., 2021). Cases with PCa and androgen deprivation treatment but still progressing will be diagnosed as castration-resistant PCa (CRPC) (Zhao et al., 2021).

Most metastatic PCa can be effectively treated with docetaxel (DTX). DTX is a drug clinically approved by the United States Food and Drug Administration for the treatment of various metastatic PCa, including androgen-independent PCa and CRPC (Hongo et al., 2021). However, as a first-line treatment for metastatic CRPC, DTX resistance is an important clinical problem. The newly developed new-generation chemotherapeutic drugs for the treatment of DTX-resistant PCa patients are often accompanied by hematotoxicity exceeding their benefits. Although DTX plays a role in the treatment of hormone-refractory PCa, DTX resistance can easily occur due to microtubule mutation and the activation of drug efflux pump after long-term treatment (Bello et al., 2021; Kobayashi et al., 2021). Therefore, studying the chemical resistance mechanism of DTX can improve the efficiency of chemotherapy. Aerobic glycolysis is mainly used by cells to provide energy and materials for their rapid growth, and glycolysis is closely related to the occurrence, development, and metastasis of malignant tumors (Shao et al., 2021). Therefore, the study of cellular glycolysis may be an important target for the treatment of malignant tumors.

MicroRNA (miRNA) is a non-coding RNA with a length of about 18.23 nt naturally generated in cells. It mainly regulates gene expression at the posttranscriptional level. It also inhibits the expression of target genes and affects the function of individual cells by binding to the 3'UTR untranslated region of its target gene messenger RNA (mRNA) (Mafi et al., 2021). However, there are only a few studies on the mechanism of drug resistance and glycolysis of miRNA in PCa. Studies have shown that miR-103a-3p plays a role in a variety of tumors (Huang et al., 2021; Li et al., 2021; Ma et al., 2021), but only a few studies in PCa exist, and its role in chemoresistance and glycolysis is unknown. Therefore, our study will clarify the genetic role of miR-103a-3p in PCa, screen and verify the target genes directly regulated by miR-103a-3p, and analyze the effects of miR-103a-3p regulating its target genes on DTX resistance and glycolysis in PCa, which will lay a certain foundation for the clinical diagnosis and treatment of miR-103a-3p in PCa.

MATERIALS AND METHODS

Material Science

For cells and plasmids, the PCa cell PC-3 was provided by the Cell Bank of the Typical Culture Preservation Center of the Chinese Academy of Sciences. Normal prostate epithelial cell RWPE-1 was provided by the American Type Culture Collection (ATCC). Plasmids overexpressing miR-103a-3p (miR-103a-3p mimic) and TRIM66 interference (si-TRIM66) and their negative controls (mimic-NC and si-NC, respectively) were designed and constructed by the Shanghai Jima Company (Shanghai, China). The cell culture medium (RPMI 1640), fetal bovine serum (FBS), double antibody, defined keratinocyte serum-free medium (K-SFM), RNA extraction kit, MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) kit, and the protein extraction kit were all provided by Kaiji Biology Company. The glucose and lactic acid kits were provided by Lyle Biotech. Transwell cell was purchased from Beijing Mengzhuang Technology Co., Ltd. (Beijing, China) and Lipofectamine 3000 was from Life Technologies (Carlsbad, CA, USA).

METHODS

Cell Culture

PC-3 cells were cultured in RPMI medium containing 10% FBS and 1% double antibody (100%) × penicillin streptomycin solution. Before cell subculture, cell growth was observed under an inverted microscope. Subculture was performed when the cell density reached 80%–90% (do not subculture if the cell state is poor). The normal prostatic epithelial cell RWPE-1 was subcultured in a medium containing 5 ng/ml epidermal growth factor, 50 mg/ml bovine pituitary extract, and 1 × K-SFM cell culture medium of antibiotics in 5% CO_2 at 37°C.

Target Gene Prediction

Based on the research strategy for miRNA, we screened *TRIM66* as the downstream target gene of miR-103a-3p by combining literature search and prediction using the biology website starBase (https://starbase.sysu.edu.cn).

Real-Time Fluorescence Quantitative PCR

A certain amount of Trizol solution was added to PC-3 cells to extract total RNA from the cells. An enzyme-free Eppendorf (EP) tube and gun head and a high-precision pipette gun, among others, were used to extract RNA. Seventy-five percent ethanol solution [anhydrous ethanol and diethylpyrocarbonate (DEPC) water, 3:1] and isopropanol were pre-cooled in advance. The cells were transferred to the enzyme-free EP tube, an appropriate amount of chloroform was added, allowed to stand for 15 min, and then centrifuged. The following parameters were set: 4°C,

TABLE 1 | qRT-PCR primers

Primer sequence	Forward (5′–3′)	Reverse (5′— 3′)
miR-103a-3p	AGCAGCATTGTACAGG GCTATG	CTCTACAGCTATATTGC CAGCCAC
TRIM66	GTATGACCAGAAGAAGTGTGAG	CTGACAGGTTCATGGAAGG
U6	CTCGCTTCGGCAGCACA	AACGCTTCACGAATTTGCGT
GAPDH	CAATAGTGATGACCTGGCCGT	AGAGGGAAATCGTGCGTGAC

12,000 rpm, centrifugation for 20 min. The EP tube was taken out and the upper aqueous phase was sucked out. The volume of the supernatant was calculated and an equal volume of pre-cooled isopropanol was added. After shaking briefly on the oscillator, centrifugation was again done at 4°C, 12,000 rpm, for 15 min, the supernatant was discarded, 1 ml 75% ethanol solution was added, re-centrifuged, and the supernatant discarded. A white RNA precipitation can be seen at the bottom of the EP tube. Of the RNA sample, 2 μl was taken and a spectrophotometer was used to detect the optical density, $OD_{260/280}$. A value between 1.9 and 2.0 denotes high RNA purity and can then be used for subsequent experiments. The remaining RNA was labeled and stored at −80°C and the RNA reverse transcribed into complementary DNA (cDNA). The primers were purchased from Shanghai Gemma Company (Shanghai, China) and the reverse transcription kit purchased from Thermo company. Performed according to the user guide in the fluorescence quantitative kit, U6 and GAPDH were taken as the internal parameters and the relative contents of miR-103a-3p and TRIM66 between PC-3 and RWPE-1 calculated with the $2^{-\Delta\Delta CT}$ method (Wang G et al., 2021). The primer sequences are shown in **Table 1**.

Dual Luciferase Assay

According to the binding site between miR-103a-3p and TRIM66 predicted using starBase (https://starbase.sysu.edu.cn), the wild-type TRIM66 (TRIM66-WT) plasmid and mutant TRIM66 (TRIM66-MUT) plasmid were established. The experiment was carried out with small and medium plasmid extraction kits purchased from Tiangen Biochemical Technology (Beijing, China) according to the manufacturer's instructions and the plasmid solution collected into a centrifuge tube until further use. The state and growth of PC-3 cells were observed. Better quality cells were co-transfected with the miR-103a-3p mimic, mimic-NC, TRIM66-WT, and TRIM66-MUT in 24-well plates. The complete medium was replaced 4–6 h after transfection. The luciferase value could be detected 48 h later. The obtained value is the luciferase value of firefly. Immediately after detection, 50 μl stop solution was added, blown evenly, and the machine for detection immediately started. The obtained value is the luciferase value of Haishen. Results were obtained by dividing the luciferase value of sea kidney by the luciferase value of firefly.

Induction of DTX-resistant Cells

DTX-resistant PCa strains were induced by gradually increasing the concentration of DTX. The initial concentration of DTX was 0.008 μg/L. After continuous culture *in vitro* for 14 months, the drug-resistant strain grew stably to a final concentration of 5 μg/L DTX. DTX-resistant cells were successfully induced.

The Drug Resistance of DTX-resistant Cells Was Detected by MTT

PCa and DTX-resistant cells in the logarithmic growth stage were inoculated on 96-well plates with an inoculation density of 2×10^4 wells. After cell adherence to the wall and growth, different concentrations of DTX were added. The final concentrations of DTX were 100, 50, 20, 10, 5, 2, and 1 μg/L. Three multiple holes were set up for each group of data. After 72 h of drug co-culture, 2 mg/ml MTT 50 μl was added to each well and placed in an incubator at 5% CO_2 and 37°C for 4 h. The culture medium was discarded by turning the plate, 150 μl DMSO was added to each well, and then placed in an incubator at 5% CO_2 and 37°C for 30 min. The absorbance value (OD value) of each hole was measured at 570 nm wavelength of the microplate reader, and the cell survival rate was calculated as follows: [survival rate = (OD value of experimental group/OD value of control group) × 100%]. Taking the drug concentration as the horizontal axis, a concentration utility curve was drawn to determine the half-maximal inhibitory concentration (IC_{50}). The drug resistance index was calculated as follows: IR = IC_{50} of DTX-resistant cells/IC_{50} of PCa cells.

Drawing of the Cell Growth Curve

Logarithmic PCa and DTX-resistant cells were digested by trypsin to prepare 2×10^4 cells/ml suspension, which was inoculated on 24-well plates at 1 ml/well. The cells were counted on days 1, 3, 5, and 7 after inoculation, and three wells were counted each time to draw the growth curve.

DTX-resistant Cell Transfection and Grouping

DTX-resistant cells in the logarithmic growth stage with cell density of about 70%–80% were transfected with Lipofectamine 3000; the miR-103a-3p mimic, si-TRIM66, mimic-NC, and si-NC were also transfected. The following groups were classified: PCa group (no transfection); DTX-resistant group (no transfection); DTX-resistant + mimic-NC group (DTX-resistant group transfected with the miR-103a-3p mimic, negative control); DTX-resistant + miR-103a-3p mimic group (DTX-resistant group transfected with the miR-103a-3p mimic); DTX-resistant + si-NC group (DTX-resistant group with TRIM66

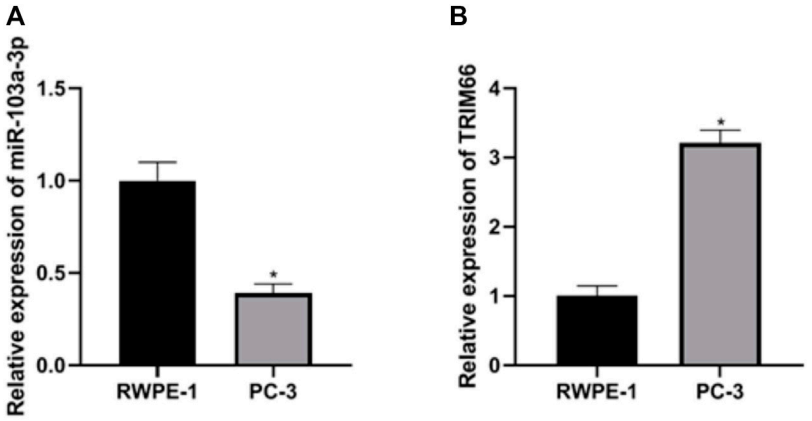


FIGURE 1 | Expressions of miR-103a-3p and TRIM66 in prostate cancer (PCa) cells. *$p < 0.05$ (compared with RWPE-1).

interference, negative control); and DTX-resistant + si-TRIM66 group (DTX-resistant group with TRIM66 interference). Quantitative real-time PCR (qRT-PCR) was used to detect the expressions of miR-103a-3p and TRIM66 in each group.

The Colony Forming Ability of DTX-resistant Cells Was Detected by Cell Cloning Assay

Transfected cells were cultured in a Petri dish with a density of 6×10^2 cells/dish. Cell fluid exchange was performed at 3–4 days and the progress of cell mass size was observed. After the cell mass size reached 50 cells/mass, fixation and staining were performed. For cell fixation, the culture medium was sucked out of the orifice plate, PBS was added, and washing was done two or three times. After sucking out PBS, methanol was added at 500 μl and fixed at room temperature for 15 min. Methanol was then sucked out and 500 μl crystal violet was added for 20 min at room temperature. Then, crystal violet was sucked out, PBS was added, and washing was performed two or three times. Spare photos were kept for further analysis of the results. The number of colonies with more than 50 cells was directly calculated by naked eye or counted using a low-power microscope.

The Migration Ability of DTX-resistant Cells Was Detected by Wound Healing Experiment

The transfected cells of each group were cultured in a Petri dish. When the cell growth density was fused, the cells of each group were scratched along the scale with a 10-μl sterile gun head on an ultra-clean workbench, and then the scratched off-wall floating cells were gently washed with sterile PBS. After 0-h scratch photography, 1% blood serum fresh culture medium was added. After 24 h, the scratch healing rate was calculated as the ratio of the difference between the scratched area at 0 and 24 h to the scratch area at 0 h. The scratch healing rate was calculated to understand the migration and repair of cells in each group.

The Invasion Ability of DTX-resistant Cells Was Detected by the Transwell Assay

Cells in good growth condition were selected and placed in 24-well plates on an ultra-clean sterile worktable, 500 μl of complete culture medium was added into the wells, and then the wells placed in the Transwell chamber with matrix glue evenly spread on the upper layer. After cell digestion, the cells were resuspended in serum-free medium, the final cell density was adjusted to 1×10^5/100 μl, and the cells added into the Transwell chamber. The 24-well plates were placed into the carbon dioxide cell incubator and cultured for 24 h. After cell culture in the Transwell chamber for 24 h, the chamber was taken out and the non-adherent cells in the lower layer of the chamber gently washed away with PBS solution. The cells in the upper layer of the chamber that did not migrate to the lower layer were then gently wiped off with a big cotton swab, placed into the chamber in a 24-well plate containing 500 μl methanol for 15 min, and then the cells transferred into a 24-well plate containing 500 μl crystal violet for staining for 20 min. Finally, the cells were taken out, placed in PBS solution to gently wash away the crystal violet that did not combine with cells, and the residual cells and crystal violet on the upper layer of the cell were again wiped off with a cotton swab. The cells were placed upside down on a super clean sterile workbench, the cells allowed to dry, and then the collected cells placed in a new 24-well plate for photos. Five different visual fields were randomly selected from each chamber to calculate the number of cells passing through.

Detection of Glycolysis

After cell transfection, the culture medium was collected and the cell fragments discarded by centrifugation. The levels of glucose and lactic acid in the culture medium were calculated; together with the cell count, the background value of the culture medium

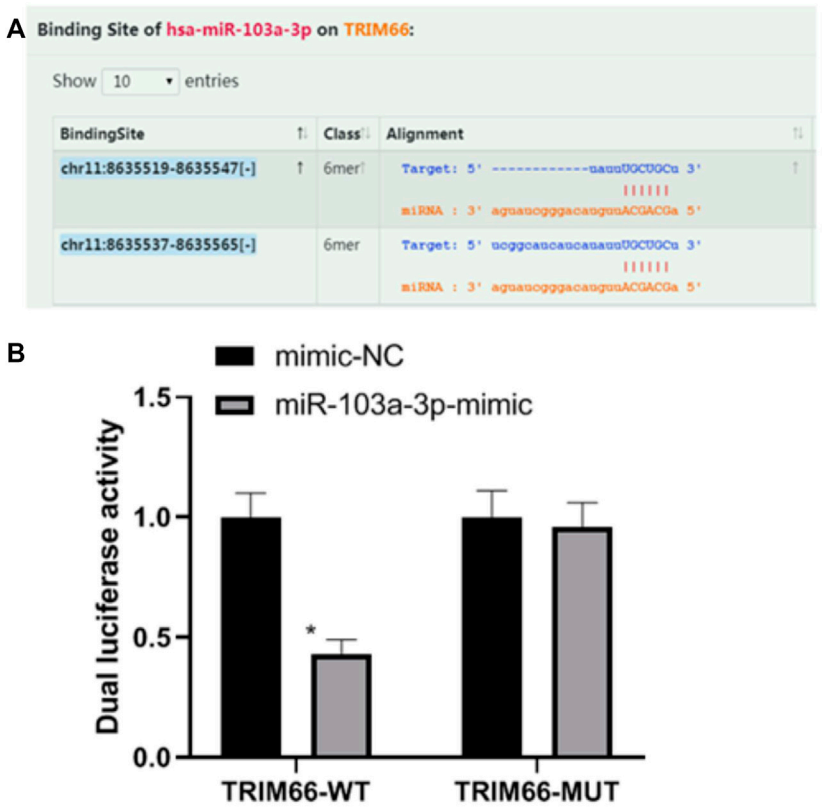


FIGURE 2 | Relationship between miR-103a-3p and TRIM66. *$p < 0.05$ (compared with mimic-NC). Source for **(A):** https://starbase.sysu.edu.cn.

was subtracted from the obtained glucose and lactic acid contents and divided by the number of corresponding cells to obtain the sugar consumption and lactic acid production of each cell. The results were expressed in relative multiples, with the cells in the control group used as the standard.

Statistical Methods

SPSS 19.0 statistical software was used for analysis of the data. Measurement data were expressed as the mean ± SD, and measurement data subject to normal distribution were compared between two groups using *t*-test and paired *t*-test. A $p < 0.05$ indicated that the difference was statistically significant.

RESULTS

Expressions of miR-103a-3p and TRIM66 in PCa Cells

The expressions of miR-103a-3p and TRIM66 in PCa cells were examined in the human normal prostate epithelial cells RWPE-1 and the PCa cell PC-3 using qRT-PCR. The results showed that, compared with RWPE-1 cells, the expression of miR-103a-3p (0.39 ± 0.05) was lower and that of TRIM66 (3.21 ± 0.19) was higher in the PCa cell PC-3 ($p < 0.05$), as shown in **Figures 1A, B**. This suggests that miR-103a-3p and TRIM66 may be involved in the progression of PCa.

Relationship Between miR-103a-3p and TRIM66

Based on the research strategy for miRNA, we screened *TRIM66* as the downstream target gene of miR-103a-3p by combining literature search and prediction on the biology website starBase (https://starbase.sysu.edu.cn). In order to verify the relationship between miR-103a-3p and TRIM66, we performed a dual luciferase reporter assay. The results showed that the luciferase activity of TRIM66-WT transfected with the miR-103a-3p mimic decreased (0.43 ± 0.06, $p < 0.05$), while that of TRIM66-MUT transfected with the miR-103a-3p mimic did not change significantly (0.96 ± 0.10, $p > 0.05$), as shown in **Figures 2A, B**. This suggests that there is a targeting relationship between miR-103a-3p and TRIM66.

Drug Resistance Characteristics and Growth Curve of DTX-resistant Cells

In order to investigate the relationship between miR-103a-3p, TRIM66, and DTX resistance in PCa, DTX-resistant cells were

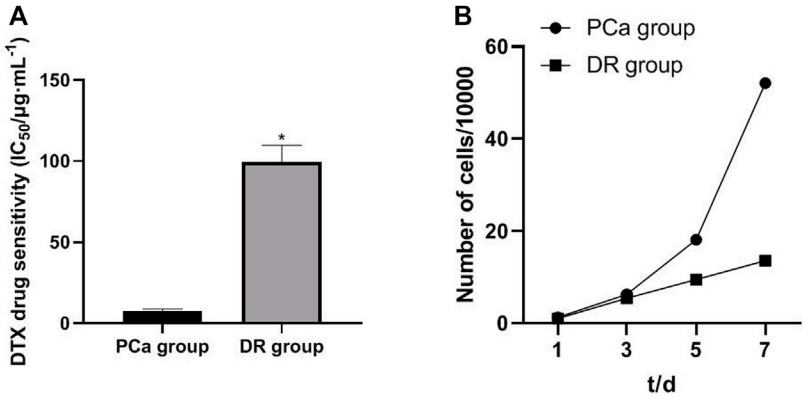


FIGURE 3 | Drug resistance characteristics and growth curve of DTX-resistant cells. *$p < 0.05$ [compared with the prostate cancer (PCa) group].

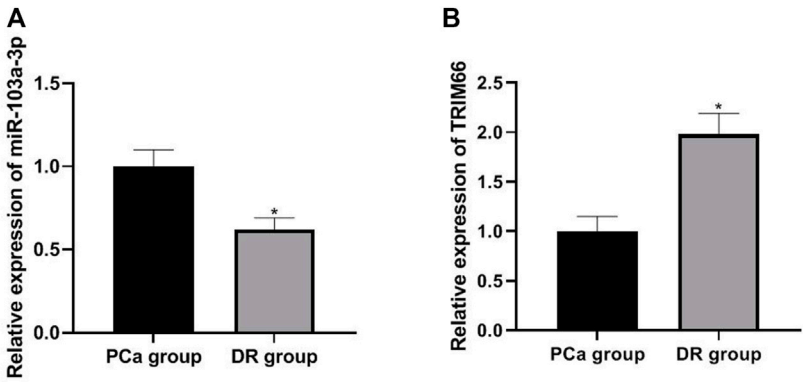


FIGURE 4 | Expressions of miR-103a-3p and TRIM66 in DTX-resistant cells.

established. By detecting the drug resistance of DTX-resistant cells, it was found that, compared with cells in the PCa group, those in the DTX-resistant group showed increased resistance to DTX. The drug resistance index was 13.33, and the doubling time of the DTX-resistant group cells was significantly longer than that of the PCa group cells ($p < 0.05$), as shown in **Figures 3A, B**. As for the relationship between the tumor doubling time and the efficacy of chemotherapy, it is considered that the shorter the tumor doubling time, the more sensitive to chemotherapy drugs and the better the efficacy. On the contrary, the longer the doubling time of the tumor, the lower the sensitivity to chemotherapeutic drugs (Han et al., 2010). Therefore, the results suggest that DTX-resistant cells were successfully induced.

Expressions of miR-103a-3p and TRIM66 in DTX-resistant Cells

The expressions of miR-103a-3p and TRIM66 in DTX-resistant cells were detected in the PCa group and DTX-resistant group cells by qRT PCR. The results showed that the expression of miR-103a-3p (0.62 ± 0.07) was lower and that of TRIM66 (1.98 ± 0.21) was higher in the DTX-resistant group than in the PCa group ($p < 0.05$), as shown in **Figures 4A, B**. This suggests that miR-103a-3p and TRIM66 may be involved in the occurrence of DTX resistance.

Expressions of miR-103a-3p and TRIM66 in DTX-resistant Cells After Transfection

In order to investigate the effects of miR-103a-3p and TRIM66 on DTX resistance, we regulated their expressions in DTX-resistant cells. qRT-PCR showed that there was no significant difference in the expressions of miR-103a-3p between the DTX-resistant group and the DTX-resistant + mimic-NC group ($p > 0.05$). After overexpression of miR-103a-3p, it (3.31 ± 0.25) increased significantly in the DTX-resistant + miR-103a-3p mimic group ($p < 0.05$). There was no significant difference in the expression of TRIM66 between the DTX-resistant group and DTX-resistant + si-NC group ($p > 0.05$). After interference with TRIM66, the expression of TRIM66 (0.41 ± 0.06) decreased significantly in the DTX-resistant + si-TRIM66 group ($p < 0.05$), as shown in **Figures 5A, B**. This suggests that the transfection regulation of DTX-resistant cells was successful. We also found that TRIM66 (0.54 ± 0.06) decreased in the DTX-resistant + miR-103a-3p mimic

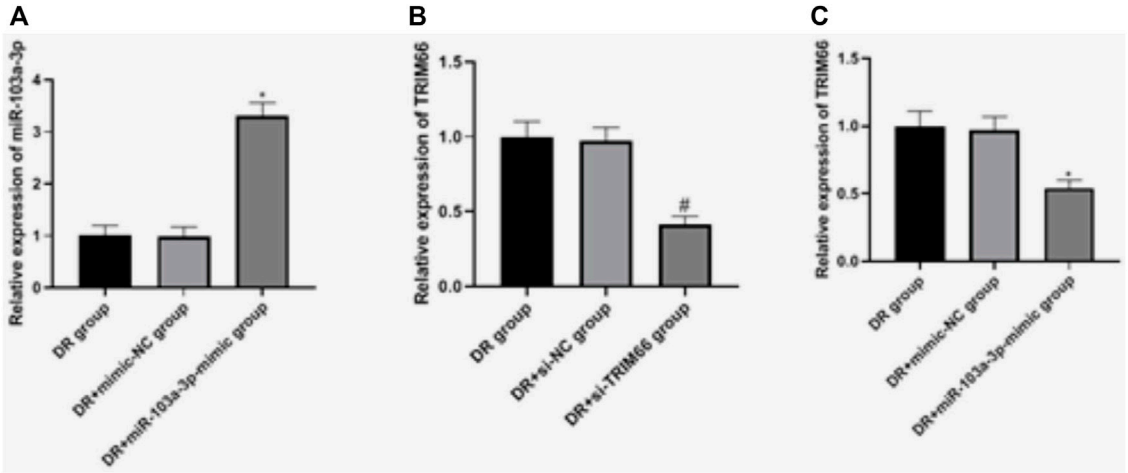


FIGURE 5 | Expressions of miR-103a-3p and TRIM66 in DTX-resistant cells after transfection. *$p < 0.05$ (compared with the DTX-resistant + mimic-NC group); #$p < 0.05$ (compared with the DTX-resistant + si-NC group).

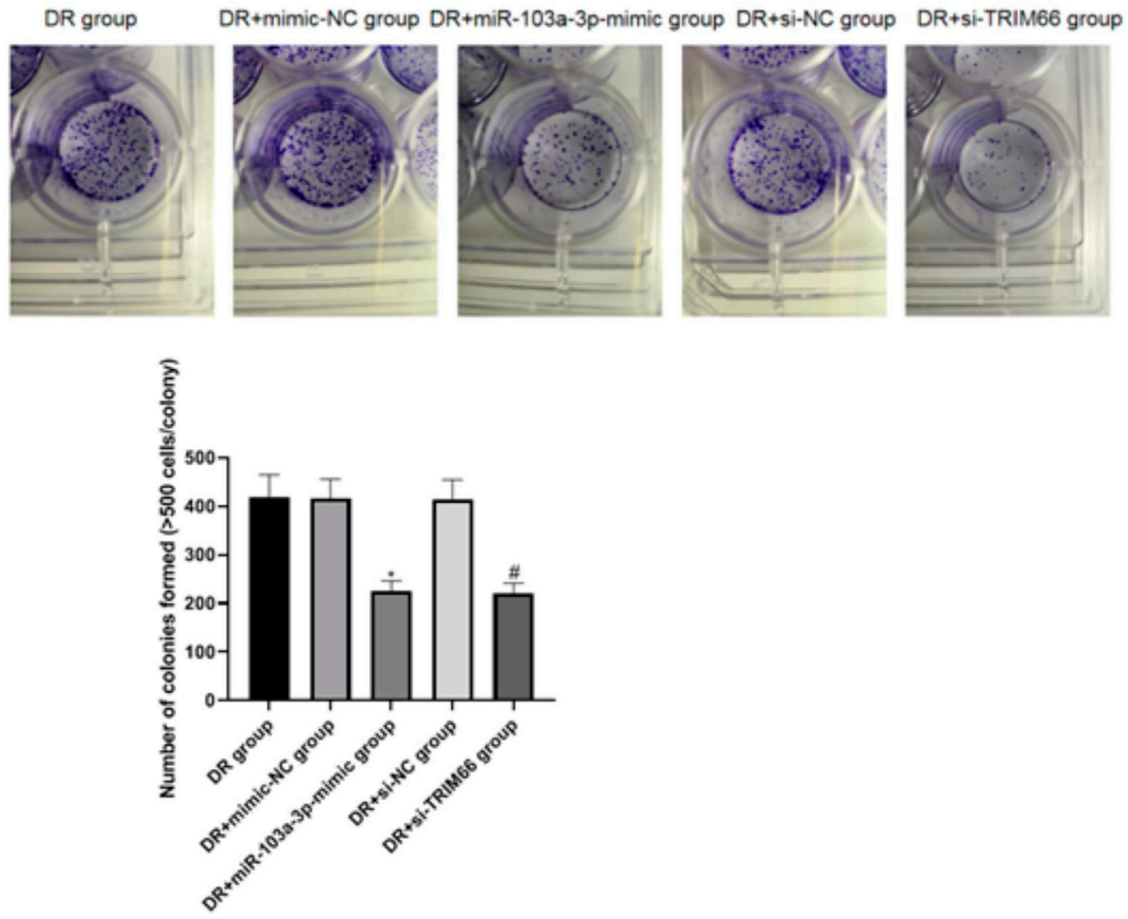


FIGURE 6 | Effect of the upregulation of miR-103a-3p and interference with TRIM66 on the colony formation of DTX-resistant cells. *$p < 0.05$ (compared with the DTX-resistant + mimic-NC group); #$p < 0.05$ (compared with the DTX-resistant + si-NC group).

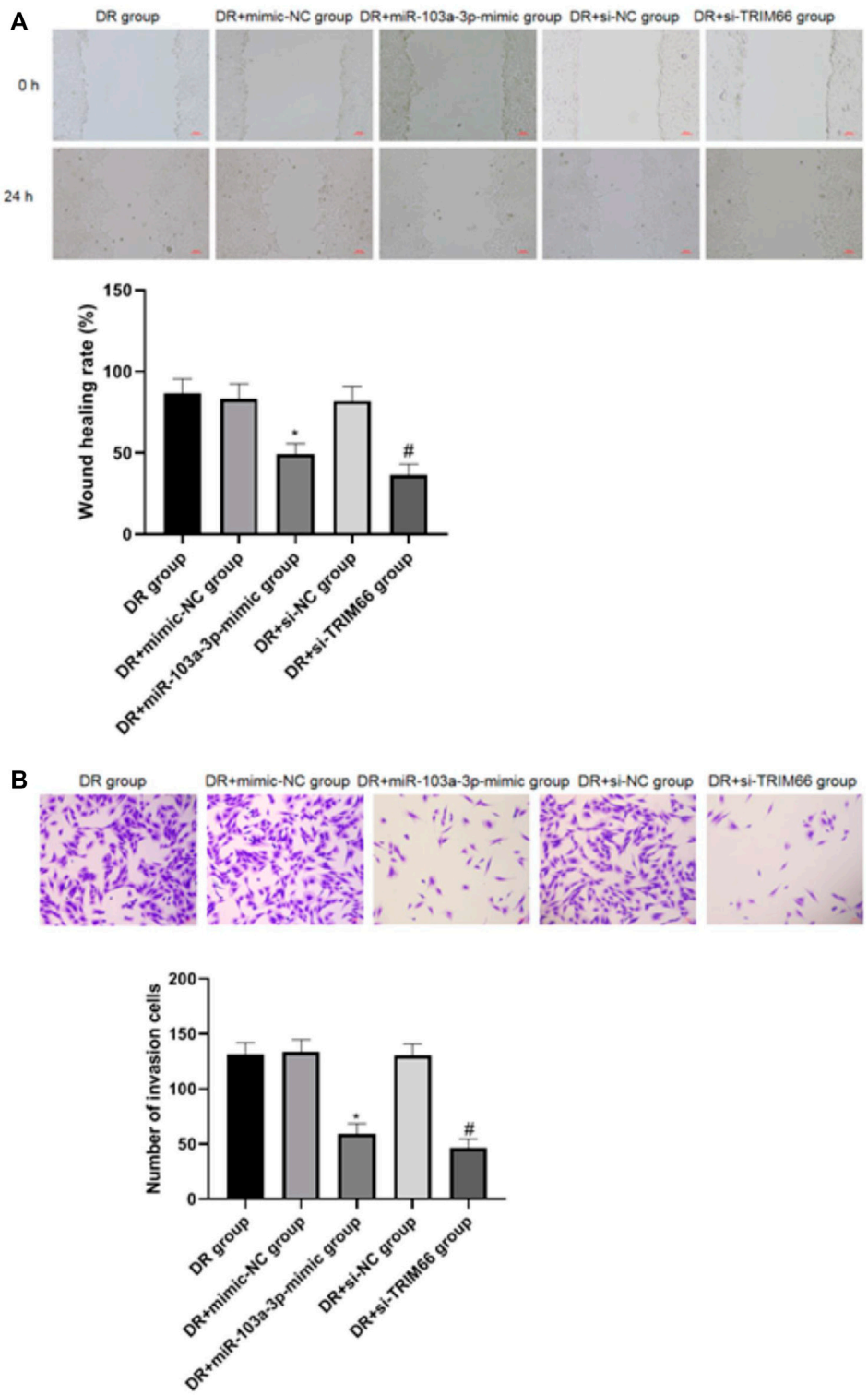


FIGURE 7 | Effect of the upregulation of miR-103a-3p and interference with TRIM66 on DTX-resistant cell metastasis (×100). *$p < 0.05$ (compared with the DTX-resistant + mimic NC group); #$p < 0.05$ (compared with the DTX-resistant + si-NC group).

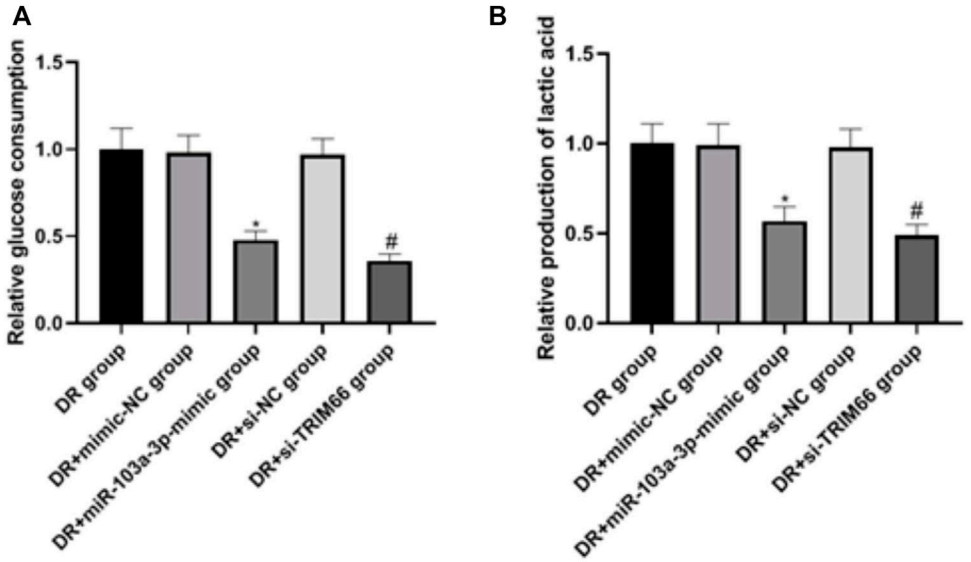


FIGURE 8 | Effect of the upregulation of miR-103a-3p and interference with TRIM66 on the glycolysis of DTX-resistant cells.

group compared with the DTX-resistant + mimic-NC group ($p < 0.05$; **Figure 5C**). The binding sites and targeting relationship between miR-103a-3p and TRIM66 indicated that both can target and regulate the expression of TRIM66.

Effects of the Upregulation of miR-103a-3p and Interference With TRIM66 on the Colony Formation of DTX-resistant Cells

The cell cloning test was used to examine the effect of the upregulation of miR-103a-3p and interference with TRIM66 on the colony formation of DTX-resistant cells. The results showed that there was no significant difference in the number of cell colonies between the DTX-resistant group, DTX-resistant + mimic-NC group, and the DTX-resistant + si-NC group ($p > 0.05$). After overexpression of miR-103a-3p and interference with TRIM66, the colony formation of the DTX-resistant + miR-103a-3p-mimic group (225.17 ± 21.14) and the DTX-resistant + si-TRIM66 group (220.84 ± 20.68) decreased relatively ($p < 0.05$; **Figure 6**). This suggests that the upregulation of miR-103a-3p and interference with TRIM66 can inhibit the colony formation of DTX-resistant cells.

Effects of the Upregulation of miR-103a-3p and Interference With TRIM66 on DTX-resistant Cell Metastasis

The wound healing assay and Transwell test were utilized to investigate the effects of the upregulation of miR-103a-3p and interference with TRIM66 on DTX-resistant cell metastasis. The results showed that there was no significant difference in cell migration and the number of invasive cells between the DTX-resistant group, DTX-resistant + mimic-NC group, and the DTX-resistant + si-NC group ($p > 0.05$). After overexpression of miR-103a-3p and interference with TRIM66, the cell migration ability and the number of invasive cells decreased significantly in the DTX-resistant + miR-103a-3p mimic group (49.38 ± 6.48 and 59.46 ± 9.22, respectively) and the DTX-resistant + si-TRIM66 group (36.49 ± 6.59 and 46.31 ± 8.34, respectively, $p < 0.05$), as shown in **Figures 7A, B**. This suggests that the upregulation of miR-103a-3p and interference with TRIM66 can inhibit DTX-resistant cell metastasis.

Effects of miR-103a-3p and TRIM66 on the Glycolysis of DTX-resistantDTX-resistant Cells

In order to examine the effects of the upregulation of miR-103a-3p and interference with TRIM66 on the glycolysis of DTX-resistant cells, we used a glycolysis kit for related detection. The results showed that there was no significant difference in glucose consumption and lactate production between the DTX-resistant group, the DTX-resistant + mimic-NC group, and the DTX-resistant + si-NC group ($p > 0.05$). After overexpression of miR-103a-3p and interference with TRIM66, glucose consumption and lactate production decreased in the DTX-resistant + miR-103a-3p mimic group (0.48 ± 0.05 and 0.57 ± 0.08, respectively) and the DTX-resistant + si-TRIM66 group (0.36 ± 0.04 and 0.49 ± 0.06, respectively, $p < 0.05$), as shown in **Figures 8A, B**. This suggests that the upregulation of miR-103a-3p and interference with TRIM66 can inhibit the glycolysis of DTX-resistant cells.

DISCUSSION

PCa is a common complex disease with multiple etiologies. It is a common tumor threatening men's health. In China, with the aging of the population, the incidence rate of PCa has been increasing year by year. It has become one of the major threats to the health of Chinese men (Raittinen et al., 2020). In recent years, with the discovery of miRNA, there have been great changes in the classical molecular biology theory. People now have a new understanding of the mechanisms of occurrence and development of malignant tumors. In this process, miRNA has brought a new dawn to PCa research with its important diagnostic and therapeutic value.

Mature miRNA binds to the target gene mRNA through nucleic acid sequence complementarity in order to inhibit the translation of the target gene mRNA or degrade it. Its function involves cell differentiation, proliferation, apoptosis, invasion, and metastasis (Qiu et al., 2021). Some scholars have studied the distribution and biological function of miRNA in mammals and found that it is involved in almost all life processes of organisms, including development events, epidermal morphogenesis, cell differentiation, enzyme activity, hormone secretion, and other regulation processes (Corrêa et al., 2021). Recent studies have shown that tumor tissues have different miRNA expression profiles compared with normal tissues. Comparison and analysis of the abnormal expression of miRNA can help to better understand the diagnosis, treatment, and prognosis of tumors. Therefore, the study of abnormal miRNA expression in PCa can help to discover important miRNA molecules, explain the mechanisms of occurrence and development of PCa, and explore its potential as a therapeutic target for new drugs.

Many studies have reported that miRNAs can target and regulate the expressions of a variety of proto-oncogenes in PCa so as to inhibit its progression, including miR-200b, miR-145, miR-224, and miR-218 (Gurbuz et al., 2021; Zabegina et al., 2021; Chen et al., 2018; Tian et al., 2020). Recent reports have indicated that miR-103 inhibits tumor cell proliferation in PCa by targeting programmed cell death protein 10 (PDCD10), while miR-103a-3p is also low in bladder cancer (Fu et al., 2016; Zhong et al., 2016). On the contrary, other studies have shown that miR-103a-3p is the proto-oncogene of many cancers, which is higher in gastric cancer, breast cancer, and ovarian cancer (Chang et al., 2016; Niu et al., 2019; Wang et al., 2021). In addition, studies have shown that miR-103a-3p is a prognostic biomarker for colon cancer and pleural mesothelioma (Caritg et al., 2016; Cavalleri et al., 2017). These often contrasting results show that the gene fDunction effects of miR-103a-3p in different cancers are usually different. Many reports have found and defined the expression profiles of disease-specific miRNAs in different tumors.

miR-103a and miR-103 are highly homologous, while miR-103a-3p is processed from the 3′-end arm of the miR-103a precursor. Our study found that the relative expression of miR-103a-3p in PCa cells was lower than that in normal prostate cells. This indicates that miR-103a-3p may be involved in the progression of PCa. Through literature search and the starBase website (https://starbase.sysu.edu.cn), we screened the target regulated by miR-103a-3p. We found that TRIM66 was highly expressed in PCa cells. This suggests that TRIM66 may be a pro-oncogene that promotes the malignant progression of PCa. In recent years, with further studies on the mechanism of chemotherapeutic drugs, chemotherapy has become one of the main methods for the clinical treatment of PCa. DTX is a taxane drug. As a first-line chemotherapy drug for the treatment of metastatic castration-resistant PCa, DTX has good clinical efficacy (Wang et al., 2020). However, after DTX treatment, about 95% of patients develop DTX resistance. Chemoresistance is inevitable in tumors that involves a variety of mechanisms, such as reducing the drug concentration of cells, increasing the cell metabolism of drug detoxification proteins, activating the survival signaling pathway, and inhibiting apoptosis (Bodzioch et al., 2021). Therefore, it is of great significance to discover new molecular targets and possible mechanisms of DTX resistance.

In further experimental studies, we established DTX-resistant cells and analyzed the low expression of miR-103a-3p and the high expression of TRIM66 in these cells, the results of which suggest that miR-103a-3p and TRIM66 may be involved in the DTX resistance of PCa. The overexpression of miR-103a-3p was found to inhibit the colony formation, migration, and invasion of DTX-resistantD cells, showing that miR-103a-3p may play the role of a tumor suppressor gene in PCa and reduce DTX resistance. It was found that the expression of TRIM66 in DTX-resistant cells decreased after the overexpression of miR-103a-3p. In addition, after interference with TRIM66, we found that TRIM66 could also inhibit DTX-resistant cells. Previous studies have shown that TRIM66 has an effect on the response to DTX resistance, which is consistent with our results that interference with TRIM66 can consistently inhibit DTX-resistant cell colony formation, migration, and invasion and that miR-103a-3p may inhibit DTX resistance and the biological behavior of cells by targeting TRIM66.

The energy metabolism of malignant tumor cells is also one of the preface and hot spots of current research, but there are still several unexplained mechanisms. Exploring new molecular targets and signal pathways is of great significance for tumor prevention and treatment. Malignant tumor cells mainly use aerobic glycolysis instead of oxidative phosphorylation of normal tissues for glucose metabolism (Warburg effect, discovered by Nobel laureate Warburg in 1920) (Guo et al., 2021). Compared with oxidative phosphorylation, glycolysis is a relatively low-productivity and low-efficiency process, but tumor cells can still quickly obtain energy ATP and growth material to supply their vigorous growth needs (Benny et al., 2020). In our study, it was also found that, after overexpression of miR-103a-3p and interference with TRIM66, the glucose consumption and lactate production of DTX-resistant cells were significantly reduced, significantly inhibiting the glycolytic flow of cells, which would slow down the growth of cancer cells and inhibit tumor.

CONCLUSION

In conclusion, the expression of miR-103a-3p was downregulated and that of TRIM66 was upregulated in the malignant

progression of PCa, and its expression was more obvious in DTX resistance. Upregulation of miR-103a-3p and interference with TRIM66 can inhibit DTX resistance and the glycolysis of PCa cells. Targeting TRIM66 may provide potential application value in molecular therapy for PCa.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

QY was mainly responsible for designing the experiments and writing the article. JW and YL were mainly responsible for performing the experiments and analyzing the data. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.813793/full#supplementary-material

REFERENCES

Abouhashem, N. S., and Salah, S. (2020). Differential Expression of NKX 3.1 and HOXB 13 in Bone Metastases Originating from Prostatic Carcinoma Among the Egyptian Males. *Pathol. - Res. Pract.* 216, 153221. doi:10.1016/j.prp.2020.153221

Bello, T., Paindelli, C., Diaz-Gomez, L. A., Melchiorri, A., Mikos, A. G., Nelson, P. S., et al. (2021). Computational Modeling Identifies Multitargeted Kinase Inhibitors as Effective Therapies for Metastatic, Castration-Resistant Prostate Cancer. *Proc. Natl. Acad. Sci. USA* 118, e2103623118. doi:10.1073/pnas.2103623118

Benny, S., Mishra, R., Manojkumar, M. K., and Aneesh, T. P. (2020). From Warburg Effect to Reverse Warburg Effect; the New Horizons of Anti-cancer Therapy. *Med. Hypotheses* 144, 110216. doi:10.1016/j.mehy.2020.110216

Bodzioch, M., Bajger, P., and Foryś, U. (2021). Angiogenesis and Chemotherapy Resistance: Optimizing Chemotherapy Scheduling Using Mathematical Modeling. *J. Cancer Res. Clin. Oncol.* 147, 2281–2299. doi:10.1007/s00432-021-03657-9

Caritg, O., Navarro, A., Moreno, I., Martínez-Rodenas, F., Cordeiro, A., Muñoz, C., et al. (2016). Identifying High-Risk Stage II Colon Cancer Patients: A Three-MicroRNA-Based Score as a Prognostic Biomarker. *Clin. Colorectal Cancer* 15, e175–e182. doi:10.1016/j.clcc.2016.04.008

Cavalleri, T., Angelici, L., Favero, C., Dioni, L., Mensi, C., Bareggi, C., et al. (2017). Plasmatic Extracellular Vesicle microRNAs in Malignant Pleural Mesothelioma and Asbestos-Exposed Subjects Suggest a 2-miRNA Signature as Potential Biomarker of Disease. *PLoS One* 12, e0176680. doi:10.1371/journal.pone.0176680

Chang, J. T.-H., Wang, F., Chapin, W., and Huang, R. S. (2016). Identification of MicroRNAs as Breast Cancer Prognosis Markers through the Cancer Genome Atlas. *PLoS One* 11, e0168284. doi:10.1371/journal.pone.0168284

Chen, C.-h., Li, S.-x., Xiang, L.-x., Mu, H.-q., Wang, S.-b., and Yu, K.-y. (2018). HIF-1α Induces Immune Escape of Prostate Cancer by Regulating NCR1/NKp46 Signaling through miR-224. *Biochem. Biophysical Res. Commun.* 503, 228–234. doi:10.1016/j.bbrc.2018.06.007

Corrêa, S., Lopes, F. P., Panis, C., Basili, T., Binato, R., and Abdelhay, E. (2021). miRNome Profiling Reveals Shared Features in Breast Cancer Subtypes and Highlights miRNAs that Potentially Regulate MYB and EZH2 Expression. *Front. Oncol.* 11, 710919. doi:10.3389/fonc.2021.710919

Davidsson, S., Andren, O., Ohlson, A.-L., Carlsson, J., Andersson, S.-O., Giunchi, F., et al. (2018). FOXP3+ Regulatory T Cells in normal Prostate Tissue, Postatrophic Hyperplasia, Prostatic Intraepithelial Neoplasia, and Tumor Histological Lesions in Men with and without Prostate Cancer. *Prostate* 78, 40–47. doi:10.1002/pros.23442

Fu, X., Zhang, W., Su, Y., Lu, L., Wang, D., and Wang, H. (2016). MicroRNA-103 Suppresses Tumor Cell Proliferation by Targeting PDCD10 in Prostate Cancer. *Prostate* 76, 543–551. doi:10.1002/pros.23143

Guo, X.-H., Jiang, S.-S., Zhang, L.-L., Hu, J., Edelbek, D., Feng, Y.-Q., et al. (2021). Berberine Exerts its Antineoplastic Effects by Reversing the Warburg Effect via Downregulation of the Akt/mTOR/GLUT1 Signaling Pathway. *Oncol. Rep.* 46, 253. doi:10.3892/or.2021.8204

Gurbuz, V., Sozen, S., Bilen, C., and Konac, E. (2021). miR-148a, miR-152 and miR-200b P-romote P-rostate C-ancer M-etastasis by T-argeting DNMT1 and PTEN E-xpression. *Oncol. Lett.* 22, 805. doi:10.3892/ol.2021.13066

Han, L. Y., Karavasilis, V., Hagen, T. v., Nicum, S., Thomas, K., Harrison, M., et al. (2010). Doubling Time of Serum CA125 Is an Independent Prognostic Factor for Survival in Patients with Ovarian Cancer Relapsing after First-Line Chemotherapy. *Eur. J. Cancer* 46, 1359–1364. doi:10.1016/j.ejca.2010.02.012

Hongo, H., Kosaka, T., Suzuki, Y., and Oya, M. (2021). Discovery of a New Candidate Drug to Overcome Cabazitaxel-Resistant Gene Signature in Castration-Resistant Prostate Cancer by In Silico Screening. *Prostate Cancer Prostatic Dis.* 30. doi:10.1038/s41391-021-00426-0

Huang, J., Lin, F., Xu, C., and Xu, Y. (2021). LINC00662 Facilitates Osteosarcoma Progression via Sponging miR-103a-3p and Regulating SIK2 Expression. *J. Tissue Eng. Regen. Med.* 15, 1082–1091. doi:10.1002/term.3242

Kobayashi, H., Shiota, M., Sato, N., Kobayashi, S., Matsumoto, T., Monji, K., et al. (2021). Differential Prognostic Impact of Complete Blood Count-Related Parameters by Prior Use of Novel Androgen Receptor Pathway Inhibitors in Docetaxel-Treated Castration-Resistant Prostate Cancer Patients. *Anticancer Drugs* 33, e541–e547. doi:10.1097/CAD.0000000000001170

Li, H., Huhe, M., and Lou, J. (2021). MicroRNA-103a-3p Promotes Cell Proliferation and Invasion in Non-small-cell Lung Cancer Cells through Akt Pathway by Targeting PTEN. *Biomed. Res. Int.* 2021, 1–10. doi:10.1155/2021/7590976

Lu, X., Gao, W., Zhang, Y., Wang, T., Gao, H., Chen, Q., et al. (2021). Case Report: Systemic Treatment and Serial Genomic Sequencing of Metastatic Prostate Adenocarcinoma Progressing to Small Cell Carcinoma. *Front. Oncol.* 11, 732071. doi:10.3389/fonc.2021.732071

Ma, G., Chen, J., Wei, T., Wang, J., and Chen, W. (2021). Inhibiting Roles of FOXA2 in Liver Cancer Cell Migration and Invasion by Transcriptionally Suppressing microRNA-103a-3p and Activating the GREM2/LATS2/YAP axis. *Cytotechnology* 73, 523–537. doi:10.1007/s10616-021-00475-2

Mafi, A., Yadegar, N., Salami, M., Salami, R., Vakili, O., and Aghadavod, E. (2021). Circular RNAs; Powerful microRNA Sponges to Overcome Diabetic Nephropathy. *Pathol. - Res. Pract.* 227, 153618. doi:10.1016/j.prp.2021.153618

Niu, Y., Zhang, J., Tong, Y., Li, J., and Liu, B. (2019). Physcion 8-O-β-Glucopyranoside Induced Ferroptosis via Regulating miR-103a-3p/GLS2 axis in Gastric Cancer. *Life Sci.* 237, 116893. doi:10.1016/j.lfs.2019.116893

Prendeville, S., Al-Bozom, I., Compérat, E., Sweet, J., Evans, A. J., Ben-Gashir, M., et al. (2017). Prostate Carcinoma with Amphicrine Features: Further Refining the Spectrum of Neuroendocrine Differentiation in Tumours of Primary Prostatic Origin? *Histopathology* 71, 926–933. doi:10.1111/his.13330

Qiu, K., Song, Y., Rao, Y., Liu, Q., Cheng, D., Pang, W., et al. (2021). Diagnostic and Prognostic Value of MicroRNAs in Metastasis and Recurrence of Head and

Neck Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis. *Front. Oncol.* 11, 711171. doi:10.3389/fonc.2021.711171

Raittinen, P., Niemistö, K., Pennanen, E., Syvälä, H., Auriola, S., Riikonen, J., et al. (2020). Circulatory and Prostatic Tissue Lipidomic Profiles Shifts after High-Dose Atorvastatin Use in Men with Prostate Cancer. *Sci. Rep.* 10, 12016. doi:10.1038/s41598-020-68868-5

Shao, D., Gao, Q., Cheng, Y., Du, D. Y., Wang, S. Y., and Wang, S. X. (2021). Prognostic Value of 18F-Fluorodeoxyglucose PET/CT in the Initial Assessment of Primary Tracheal Malignant Tumor: A Retrospective Study. *Korean J. Radiol.* 22, 425–434. doi:10.3348/kjr.2020.0211

Tian, J., Zhang, H., Mu, L., Wang, M., Li, X., Zhang, X., et al. (2020). The miR-218/GAB2 axis Regulates Proliferation, Invasion and EMT via the PI3K/AKT/GSK-3β Pathway in Prostate Cancer. *Exp. Cel Res.* 394, 112128. doi:10.1016/j.yexcr.2020.112128

Wang, G., Huang, Y., Xie, F., Cheng, X., Chi, J., Lei, Y., et al. (2021). LncRNA MEG3 Promotes Endoplasmic Reticulum Stress and Suppresses Proliferation and Invasion of Colorectal Carcinoma Cells through the MEG3/miR-103a-3p/PDHB ceRNA Pathway. *neo* 68, 362–374. doi:10.4149/neo_2020_200813N858

Wang, S., Gao, G., He, Y., Li, Q., Li, Z., and Tong, G. (2021). Amidation-Modified Apelin-13 Regulates PPARγ and Perilipin to Inhibit Adipogenic Differentiation and Promote Lipolysis. *Bioinorganic Chem. Appl.* 2021, 1–9. doi:10.1155/2021/3594630

Wang, Y., Huang, Z., Chen, C. Z., Liu, C., Evans, C. P., Gao, A. C., et al. (2020). Therapeutic Targeting of MDR1 Expression by RORγ Antagonists Resensitizes Cross-Resistant CRPC to Taxane via Coordinated Induction of Cell Death Programs. *Mol. Cancer Ther.* 19, 364–374. doi:10.1158/1535-7163.MCT-19-0327

Zabegina, L., Nazarova, I., Nikiforova, N., Slyusarenko, M., Sidina, E., Knyazeva, M., et al. (2021). A New Approach for Prostate Cancer Diagnosis by miRNA Profiling of Prostate-Derived Plasma Small Extracellular Vesicles. *Cells* 10, 2372. doi:10.3390/cells10092372

Zhao, Y.-x., Yao, G.-l., Sun, J., Wang, X.-l., Wang, Y., Cai, Q.-q., et al. (2021). Nomogram Incorporating Contrast-Enhanced Ultrasonography Predicting Time to the Development of Castration-Resistant Prostate Cancer. *Clin. Med. Insights Oncol.* 15, 117955492110497. doi:10.1177/11795549211049750

Zhong, Z., Lv, M., and Chen, J. (2016). Screening Differential Circular RNA Expression Profiles Reveals the Regulatory Role of circTCF25-miR-103a-3p/miR-107-CDK6 Pathway in Bladder Carcinoma. *Sci. Rep.* 6, 30919. doi:10.1038/srep30919

Zhou, M. (2018). High-grade Prostatic Intraepithelial Neoplasia, PIN-like Carcinoma, Ductal Carcinoma, and Intraductal Carcinoma of the Prostate. *Mod. Pathol.* 31, 71–79. doi:10.1038/modpathol.2017.138

Zupancic, M., Pospihalj, B., Cerovic, S., Gazic, B., Drev, P., Hocevar, M., et al. (2020). Significance of Nuclear Factor - Kappa Beta Activation on Prostate Needle Biopsy Samples in the Evaluation of Gleason Score 6 Prostatic Carcinoma Indolence. *Radiol. Oncol.* 54, 194–200. doi:10.2478/raon-2020-0019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yi, Wei and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fgene.2022.822966

A Novel Ras--Related Signature Improves Prognostic Capacity in Oesophageal Squamous Cell Carcinoma

Hao-Shuai Yang[1†], Wei Liu[1†], Shao-Yi Zheng[1], He-Yuan Cai[1], Hong-He Luo[1], Yan-Fen Feng[2,3] and Yi-Yan Lei[1]**

[1]Department of Thoracic Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, [3]Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Xiaoxiao Sun,
University of Arizona, United States
Haoyun Lei,
Carnegie Mellon University, United States

***Correspondence:**
Yan-Fen Feng
fengyf@sysucc.org.cn
Yi-Yan Lei
leiyiyan@mail.sysu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 26 November 2021
Accepted: 08 February 2022
Published: 24 February 2022

Citation:
Yang H-S, Liu W, Zheng S-Y, Cai H-Y, Luo H-H, Feng Y-F and Lei Y-Y (2022) A Novel Ras--Related Signature Improves Prognostic Capacity in Oesophageal Squamous Cell Carcinoma. Front. Genet. 13:822966. doi: 10.3389/fgene.2022.822966

Oesophageal squamous cell carcinoma (ESCC) remains a clinically challenging disease with high morbidity rates and poor prognosis. ESCC is also the most common pathological type of oesophageal cancer (EC) in China. Ras-related genes are one of the most frequently mutated gene families in cancer and regulate tumour development and progression. Given this, we investigated the Ras-related gene expression profiles and their values in ESCC prognosis, using data from the Genotype-Tissue Expression (GTEx) and The Cancer Genome Atlas (TCGA) databases. We found that we could identify three distinct oesophageal cancer clusters based on their unique expression profile for 11 differentially expressed Ras-related genes with each of these demonstrating some prognostic value when, evaluated using univariate Cox analysis. We then used multivariate Cox analysis to identify relevant independent prognostic indicators and used these to build a new prognostic prediction model for oesophageal cancer patients using these three Ras-related genes. These evaluations produced an area under the curve (AUC) of 0.932. We found that our Ras-related signatures could also act as independent factors in ESCC prognosis and that patients with low Ras scores showed a higher overall expression levels of various immune checkpoint genes, including TNFSF4, TNFRSF8, TNFRSF9, NRP1, CD28, CD70, CD200, CD276, METTL16, METTL14, ZC3H13, YTHDF3, VIRMA, FTO, and RBM15, as well as a higher CSMD3, FLG, DNAH5, MUC4, PLCO, EYS, and ZNF804B mutation rates, and better sensitivity to drugs such as erlotinib, paclitaxel, and gefitinib. In conclusion, we were able to use the unique expression profiles of several Ras-related genes to produce a novel disease signature which might facilitate improved prognosis in ESCC, providing new insight into both diagnosis and treatment in these cancers.

Keywords: oesophageal squamous cell carcinoma, ras, prognosis, bioinformatics analysis, TCGA, GTEx

INTRODUCTION

Ras proteins are a class of GTPase that function as molecular switches in various signalling pathways regulating cellular proliferation, differentiation, survival, migration, and cytoskeletal dynamism. It has been shown that Ras-related genes are the most commonly mutated gene family in cancer, and Ras activation resulting from mutations in Ras genes or their regulators promotes the development and progression of a variety of cancers. (Pylayeva-Gupta et al., 2011; Simanshu et al., 2017; Moore et al., 2020). Oesophageal carcinoma (EC) is the seventh most common cancer globally with the sixth worst prognosis. Oesophageal squamous cell carcinoma (ESCC) and Oesophageal adenocarcinoma (EAC) are the two major subtypes of EC and are classified based on their pathology, showing obvious differences in incidence, aetiology and clinical characteristics. (Arnold et al., 2015). ESCC is the most common pathological type of EC in China, accounting for approximately 90% of all cases. Most patients are clinically diagnosed at an advanced stage and present with a very poor prognosis, with a 5-years overall survival (OS) of only 18.8%. (Abnet et al., 2018; Arnold et al., 2015). The effect of surgery alone for advanced ESCC is far from satisfactory, because of its high recurrence rates and poor survival. (Bray et al., 2018). Therefore, multidisciplinary treatment is highly recommended to improve prognosis and several recent studies have shown that the Ras signalling pathway plays an important role in the pathogenesis of EC. To the best knowledge of our, few studies have explored the relationship between Ras and ESCC.

Here, we systematically profiled the genomic information from both ESCC and normal samples using their clinical outcomes from the Genotype-Tissue Expression (GTEx) and Cancer Genome Atlas (TCGA) databases. This study was designed to investigate Ras gene expression profiles and their value in ESCC prognosis. We then used this information construct a novel prognostic prediction model for oesophageal cancer patients based on their Ras-related gene profile, providing a new tool for the diagnosis and treatment of ESCC.

METHODS AND MATERIALS

Oesophageal Cancer Dataset Source and Pre-Processing

Transcriptome gene expression, mutation frequency of transcriptome genes, and clinical data from ESCC patients were downloaded from the GTEx (https://commonfund.nih.gov/gtex) and TCGA databases (https://portal.gdc.cancer.gov/). We then controlled for bias by excluding, patients with missing genetic data in the GTEx database and five patients without sufficient clinical follow-up information from the TCGA database. Finally, we downloaded the original "CEL" files for the microarray data from Affymetrix, and applied a robust multiarray averaging method to produce our data set and then downloaded, the "limma" file and completed a robust multiarray averaging method to produce the dataset needed to complete the necessary background adjustment and quantile normalization of this datest in R version 4.0.2. We then downloaded the standardised matrix files directly for microarray data from other platforms and all data were quantile-normalized using a log2-scale transformation to ensure standardisation. Any gene symbols detected using more than one probe were evaluated using their mean expression levels.

Differential Expression Analysis and Enrichment Analysis of Oesophageal Cancer and Oesophageal Tissue

We merged 650 normal oesophageal tissue samples from the GTEx database and 77 oesophageal cancer samples from the TCGA-ESCC dataset using normalisation *via* R package "limma". We then identified the differentially expressed Ras-related genes by comparing the tumour and precancerous tissue using a threshold false discovery rate (FDR) of <0.05, along with |log2 FC (fold-change) | > 2. GO and KEGG enrichment analyses were used to investigate the biological processes implicated by these differentially expressed genes using R package "clusterProfiler" were significance was set at p <0.05.

Identification of Ras-Related Prognostic Genes

We used the data from previous studies to identify 180 critical Ras-related genes based on their KEGG database evaluations, of which 161 were included in the TCGA expression microarray data and thus used in our subsequent analysis. We then narrowed this to 11 genes which were shown to have some prognostic value based on a univariate Cox analysis with a threshold of p <0.05.

Ras-Related Cluster and Clinical Correlation Analysis

We used the K-means algorithm to classify ESCC patients into different clusters based on 11 Ras-related prognostic genes expression (Hartigan and Wong, 1979), and the results showed that K = 3 was the best classification for all 77 TCGA patients in our cohort, producing Clusters 1 (n = 37), 2 (n = 21), and 3 (n = 19). The R package "ConsensusClusterPlus" was then applied to perform the above steps 1,000 times to guarantee the stability of the classification and we also verified the discriminatory power of these clusters using Kaplan-Meier survival analysis (Wilkerson and Hayes, 2010) and performed a correlation analysis with clinical features, which was visualized by R package "pheatmap".

Gene Set Variation Analysis and Functional Annotation

We then used Gene Set Variation Analysis (GSVA) enrichment analysis to investigate differences in these clusters across biological processes, using the "GSVA" R software package. GSVA is a non-parametric, non-supervised method commonly used to estimate changes in the activity of pathways and biological processes in samples from expression datasets. (Hänzelmann

et al., 2013). The "c2.cp.kegg.v6.2.-symbols" gene sets were used to run these GSVA analyses, and were downloaded from the MSigDB database. Statistical significance was set at p <0.05.

Investigation of Tumor Immune Microenvironment and Check Point Genes

We used CIBERSORT (http://cibersort.stanford.edu/) to evaluate the immune infiltration status of each of these three clusters, and the differences in their immune infiltration were then analysed using the Wilcoxon signed-rank test. At the same time, the ESTIMATE algorithm (R package "ESTIMATE") was used to detect the activity of immune and stromal cells and evaluate tumour purity. (Yoshihara et al., 2013). We also performed a systematic search for immune checkpoint blockade gene expression profiles, such as PD-1, PD-L1, and CTLA-4 using the R packages "limma" and "ggpubr".

Establishment and Validation of Ras-Related Risk Assessment Model

To quantify the Ras modification pattern of individual tumours, we constructed a scoring system to assess the risk of ESCC patients, which we called the Ras Score. First, we performed multivariate Cox regression analysis with a threshold of $p < 0.05$ in 11 Ras-related prognostic genes, subsequently performed differential analysis of gene expression between the tumour and precancerous tissue with |log2 FC (fold-change) | > 2, in which three genes (EGFR, RAP1B, and PDGFRA) were identified for model development. The formula used to calculate the Ras score can be described as follows:

$$Ras\ Score = \hat{h}_0(t)\sum_{i=1}^{k}\beta iSi$$

βi was the expression quantity of three genes (EGFR, RAP1B, and PDGFRA) and Si was coefficient of correlation of three genes. We selected the median value as the grouping criterion for the model to differentiate patients into high - and low-score groups.

We then validated this model by splitting the score distribution and survival status dot plots. At the same time, we used Kaplan-Meier survival analysis curves to evaluate differences in the survival statistics for the high and low Ras-score groups. This process was visualised using R packages "survival", "glmnet", "pbapply", "survivalROC", and "survminer". The Ras-score was dichotomised by repeating the test on all possible cutpoints to find the surv-cutpoint function of the maximum rank statistic, and the patients were divided into high and low Ras-score groups according to the median to reduce the calculated batch effect.

ROC Curves

The specificity and sensitivity of the Ras-score were assessed using a receiver operating characteristic (ROC) curve, and the area under the curve (AUC) was quantified using the pROC R package. The AUC for the ROC ranged from 0 to 1, with close to one indicating perfect predictive ability and 0.5 indicating no predictive ability, less than 0.5 indicating worse than random guesses.

Chemotherapeutic Sensitivity Scoring

We then went on to explore the differences in chemotherapeutic sensitivity in the high- and low-score groups using "pRRophetic" in R to predict the half-maximal inhibitory concentration (IC50) of different chemotherapeutic drugs in each patient. This package predicts IC50 by creating statistical models based on drug sensitivity and RNA-Seq data based on the Genomics of Drug Sensitivity in Cancer (GDSC) (www.cancerrxgene.org/) database.

Statistical Analysis

Analysis of the correlation coefficients between Ras and tumour microenvironment (TME)-infiltrating immune cell expression was performed using Spearman's method and distance correlation analysis, respectively. One-way analysis of variance (ANOVA) and the Kruskal–Wallis test were used to compare the differences between three or more groups. (Hazra and Gogtay, 2016).

Univariate and multivariate Cox regression analyses were used to evaluate the correlations between various factors, including Ras-score and clinical characteristics. The log-rank test was used to compare survival differences between different groups and the waterfall function in the "maftools" package was used to visualized the mutation landscape in patients with high and low Ras-score subtypes in the TCGA-ESCC cohort. (Mayakonda et al., 2018). All Statistical p values were set at $p < 0.05$ and all data processing was performed using R version 4.0.2.

RESULTS

Differential Expression of Genes and Functional Enrichment Analysis

A total of 649 normal and 77 ESCC appropriate samples were retrieved from the TCGA and GTEx datasets. Our evaluation identified 4,568 significantly differentially expressed genes between the normal and ESCC samples which are represented in a heat- and volcano map (**Figures 1A,B**), respectively. These genes were then subjected to functional enrichment analysis designed to elucidate the biological functions and pathways of these differentially expressed genes. The GO results showed they were almost all enriched in cell-substrate junctions and focal adhesions (**Figure 1C**) and KEGG analysis showed that they were closely enriched in neurodegeneration-multiple diseases (**Figure 1D**).

Genetic Expression of Ras-Related Genes in ESCC and Functional Enrichment Analysis

Our study cohort was made up of a single TCGA dataset (TCGA-ESCC) which included both the gene expression profile of these samples but also the relevant OS and clinical data needed for this kind of evaluation. We used this data in our univariate Cox

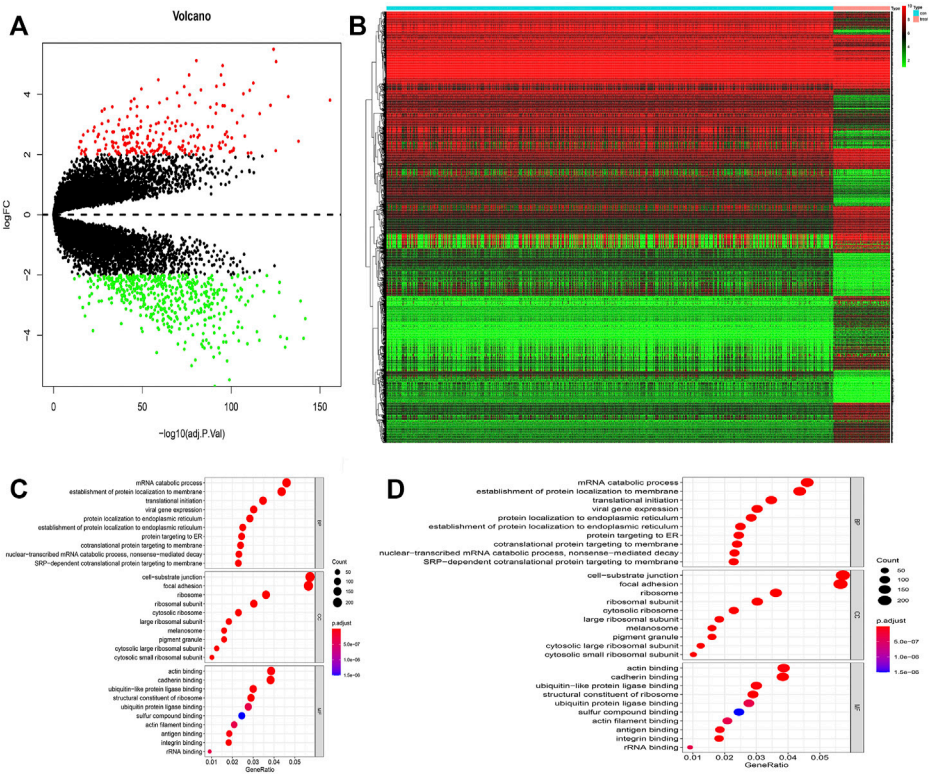


FIGURE 1 | (A) Volcano map of different expressed genes between normal and ESCC samples. **(B)** Heatmap of different expressed genes between normal and ESCC samples. **(C–D)** Bubble plots of GO analyses **(C)** and KEGG analyses **(D)**. The larger bubble indicates the more obvious enrichment.

regression evaluation of 161 Ras-related genes and identified 11 Ras-related genes with probable prognostic value when screened using a p of <0.05 (**Figure 2A**). We then performed functional enrichment analysis to elucidate the biological functions and pathways of 11 prognostic Ras-related genes. GO results showed that ERK1 and ERK2 cascade, cell adhesion molecule binding, and transmembrane receptor protein tyrosine kinase were significantly enriched (**Figure 2B**). Ras signaling pathway, MAPK signaling pathways, Rap1 signaling pathways, and EGFR tyrosine kinase inhibitor resistance were significant KEGG enrichment items (**Figure 2C**).

Ras Modification Patterns Are Predicted by 11 Ras-Related Prognostic Genes

We used the Unsupervised clustering analysis to categorise patients into different Ras-related gene clusters within the TCGA-ESCC cohort based on their expression of each of the 11 Ras-related prognostic genes. The change curve of the consensus cluster cumulative distribution function from $k = 2$ to nine showed that the area under the curve was the largest when $k = 3$ (**Figure 2D**), so we produced three different modification patterns, including Ras-cluster 1 ($n = 37$), Ras-cluster 2 ($n = 21$), and Ras-cluster 3 ($n = 19$) (**Figure 2E**). Prognostic analysis of these three clusters revealed an obviously significant survival advantage for Ras-cluster 3, and the worst survival for Ras-cluster 2 (**Figure 2G**). There was also a significant difference in the Ras gene expression profile among the three clusters (**Figure 2F**). Ras-cluster one was characterised by the increased expression levels of PLCG3, RAP1B, and PAK4, and presented with variable decreases in other Ras-related prognostic genes; Ras-cluster two showed high expression levels of HGF, GNGT2, and CSF1R and Ras-cluster three exhibited a significant increase in the expression of RAC3, PDGFRA, EGFR, PTPN11, and NF1.

Gene Set Variation Analysis and Differences in Immune Characteristics Between Clusters

Given these outcomes we then used GSVA to explore differences in the biological behaviours of these three clusters. These results showed that all three clusters were all using dramatically different immune signalling pathways. Ras-cluster two presented with in enrichment in the pathways associated with full immune activation, including the activation of the chemokine and T cell receptor signalling pathways, cytokine-cytokine receptor interaction, and Toll-like receptor signalling pathways (**Figures 2H–J**).

Thus, we went on to compare the expression of the immune checkpoint genes in these three clusters and found significant

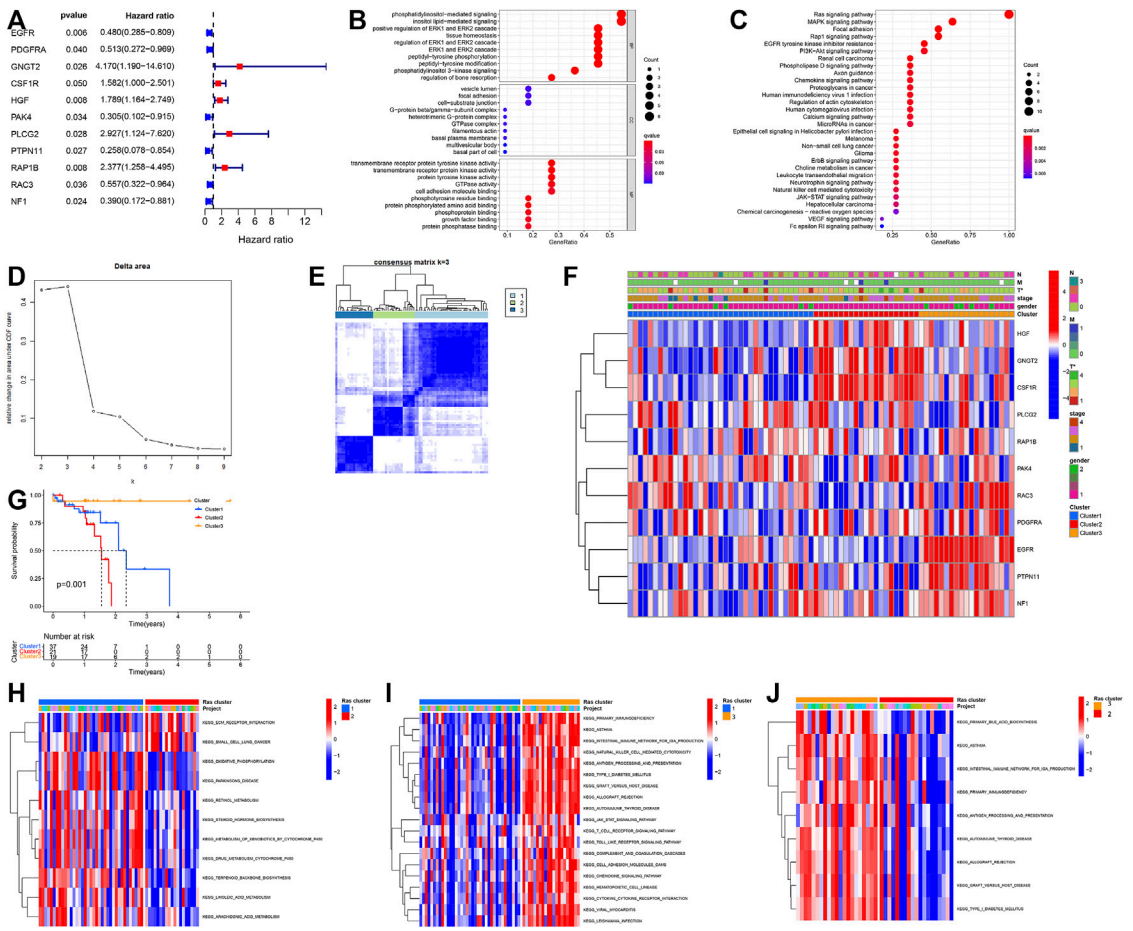


FIGURE 2 | (A) Forest plot of hazard ratios exhibiting the prognostic worth of eleven Ras-related genes. **(B,C)** Bubble plots of GO analyses **(B)** and KEGG analyses **(C)**. **(D,E)** Unsupervised clustering of 11 Ras-related prognostic genes in the TCGA-ESCC cohort. **(F)** Heatmap of Ras-related genes. Blue represents down-regulation and red represents up-regulation of genes. **(G)** Kaplan–Meier survival curve. **(H–J)** GSVA enrichment analysis showing the activation states of biological pathways in distinct Ras modification patterns. The heatmap was used to visualize these biological processes, and red represented activated pathways and blue represented inhibited pathways. **(H)** Ras-cluster1 vs Ras-cluster 3; **(I)** Ras-cluster1 vs Ras-cluster 2; **(J)** Ras-cluster3 vs Ras-cluster two.

differences in PD-1, PD-L1, and CTLA4 expression in each. The expression level of PD-1, PD-L1, and CTLA4 was significantly higher in the Ras-cluster2 (**Figures 3A–C**) as were the number of immune and stromal cells when comparing Ras-cluster two and Ras-clusters one and 3 (**Figures 3D–F**). The purity of the cancer cells in Ras-cluster2 was low and immune cell infiltrations with M1 and M2 macrophages and T cells was significantly higher in Ras-cluster 2. These data were all consistent with the immune checkpoint expression and immune scores for these pathological groups (**Figures 3G–I**).

Reliability of the Novel Three Ras-Associated Gene Signature

We performed multivariate Cox regression analysis and differential analysis of gene expression between the tumour and precancerous tissue on 11 prognostic Ras-related genes, and we got three differently expressed prognostic Ras-related genes (EGFR, RAP1B and PDGFRA) (**Figure 4A**). We then went on to build a Ras-related risk assessment model for predicting the OS of ESCC patients. We did this by constructing a set of scoring systems based on the three most relevant Ras-related prognostic genes (EGFR, RAP1B and PDGFRA) and use these to quantify the Ras modification pattern of individual patients giving them a Ras-score. The formula for calculating the Ras-score for each ESCC patient is as follows: $Ras-score = -0.027390631097796*EGFR - 0.250341831449203*PDGFRA + 0.0669651154761456*RAP1B$.

We then went on to evaluate the value of these Ras-scores in predicting specific patient' outcomes, and these evaluations identified that patients with a low Ras-score experienced significantly better survival rates (**Figure 4B**). We then used ROC analysis to determine the credibility of our model for predicting prognosis. Our ROC AUC values for the one-, two-, and 3-year survival of our TCGA-ESCC cohort were determined to be 0.827, 0.773, and 0.932, respectively, (**Figure 4C**). This same

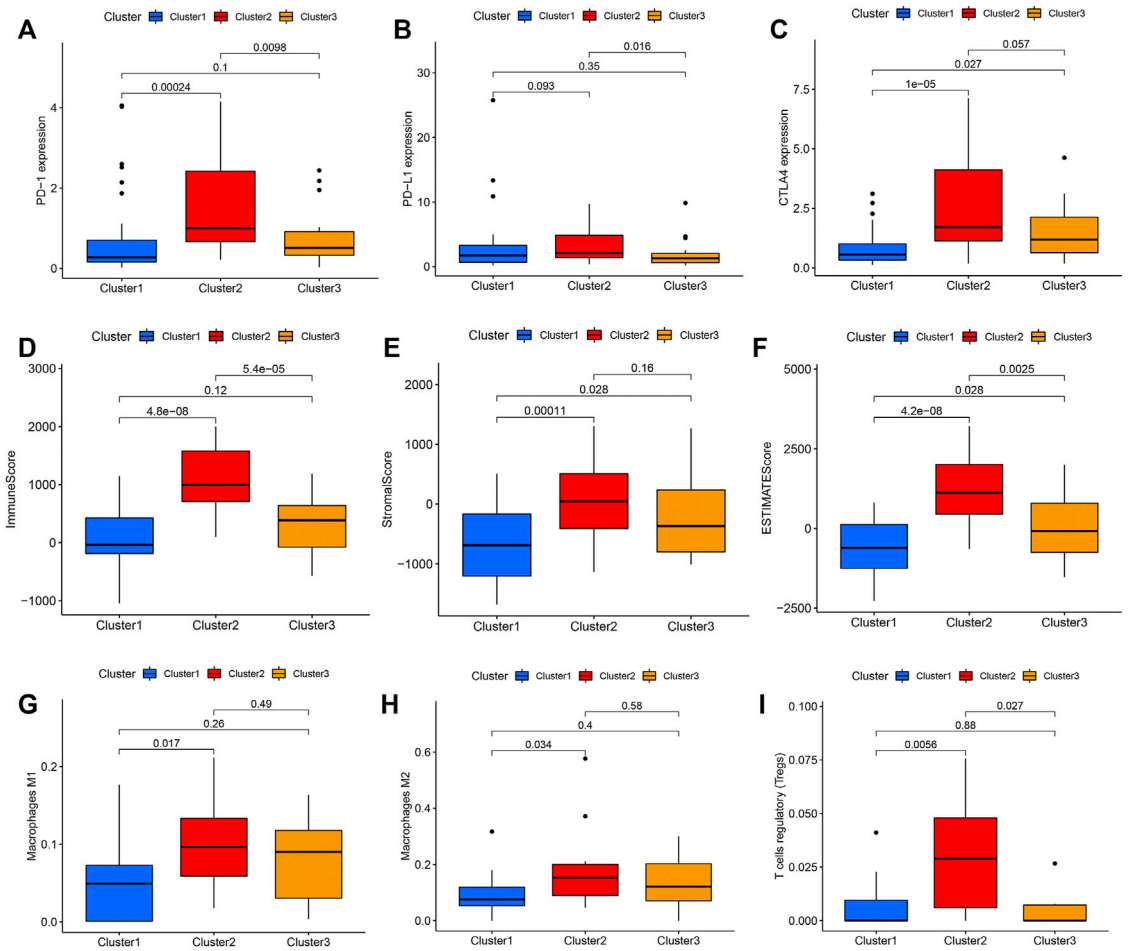


FIGURE 3 | Differential expression of immune checkpoints in Ras-cluster1 vs Ras-cluster 2 vs Ras-cluster three patients. The upper and lower ends of the boxes represented interquartile range of values. The lines in the boxes represented median value, and black dots showed outliers. The asterisks represented the statistical *p* value (ns $P > 0.05$; *$p < 0.05$; **$p < 0.01$): **(A)** PD-1, **(B)** PD-L1, **(C)** CTLA4, **(D)** Immune score, **(E)** Stromal score, **(F)** ESTIMATE score, **(G)** Macrophages M1, **(H)** Macrophages M2, **(I)** T cells regulatory.

analysis was then conducted within the cohort, and the AUC values for stage, T, M, and N in 3-year survival were 0.640, 0.367, 0.545, and 0.624, respectively, (**Figure 4D**).

We then tested whether the Ras-score could serve as an independent prognostic biomarker for ESCC. We demonstrated that stage [p = 0.04, HR = 1.872, 95%CI (1.030–3.402)], T stage [p = 0.923, HR = 1.03 1, 95% CI (0.551–1.804)], N stage [p = 0.007, HR = 2.035, 95% CI (1.210–3.421)], M stage [p = 0.087, HR = 3.007, 95%CI (0.853–10.595)], and Ras-score [p = 0.001, HR = 1.044, 95% CI (1.017–1.071)] showed significant differences in univariate Cox regression analysis (**Figure 4E**), whereas only N stage [p = 0.009, HR = 2.800, 95% CI (1.290–6.082)], M stage [p = 0.042, HR = 46.214, 95% CI (1.145–1864.590)], Ras-score (p <0.001, HR = 1.053, 95% CI (1.025–1.081)] presented as independent prognostic predictors in multivariate Cox regression analysis (**Figure 4F**). Multivariate Cox regression model analysis confirmed that the Ras-score could act as an independent and robust prognostic biomarker for evaluating patient survival in ESCC. We classified the ESCC patients into low- and high-Ras score groups based on the median Ras score (**Figure 4G**) and the predictive performance of our Ras score model for predicting patient OS is shown in **Figure 4H**.

Differences in Mutated and Key Gene Expression Between High and Low Ras-Score Groups

Given these outcomes, we further explored the relationship between the important immune-gene predictors and the two Ras-score groups using differentiation analysis. In this evaluation, we compared the expression levels of several critical immune checkpoint genes, including TNFSF4, TNFRSF8, TNFRSF9, NRP1, CD28, CD70, CD200, CD276, METTL16, METTL14, ZC3H13, YTHDF3, VIRMA, FTO, and RBM15 in each of the Ras-score groups (**Figures 5A,B**) and found that each of these genes were significantly upregulated in the low-Ras-score group when compared to the high-Ras score

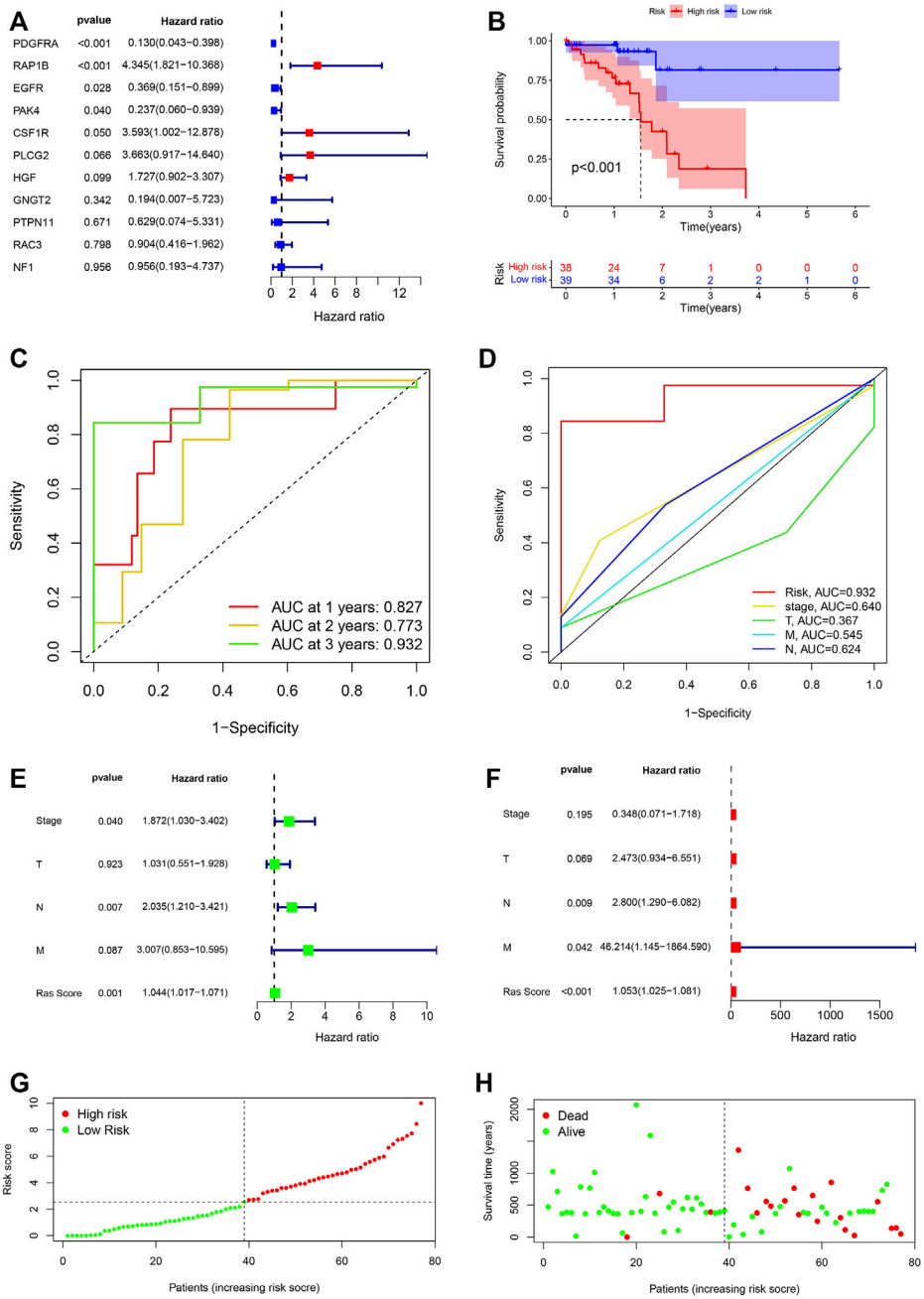


FIGURE 4 | (A) The prognostic Ras-related genes extracted by multivariate Cox regression analysis. **(B)** Kaplan-Meier analysis on high-Ras score and low-Ras score patients. **(C)** Time-dependent ROC curve analyses of Ras score. **(D)** ROC curve analyses of TNM status, stage and Ras score. **(E,F)** Uni- and multi-Cox analyses of clinical factors and risk score with OS. **(G)** Ras score distribution of high-Ras and low-Ras score patients. **(H)** Survival status of high- and low-Ras score patients.

group, while LGALS9 was enriched in the high-Ras-score group.

We also used the gene mutation data from the 77 ESCC patients in our dataset to analyse the difference in gene mutations between these two groups, and the genes with the top 20 mutation rates are shown in **Figures 5C,D**. The results showed that CSMD3, FLG, DNAH5, MUC4, PLCO, EYS, and ZNF804B experienced significantly higher mutation rates in the low-Ras-score group, whereas MUC16, KMT2D, FAT3, and NFE2L2 were highly mutated in the high-Ras-score group.

Correlation Analysis Between Ras-Score and Chemosensitivity

We then evaluated the utility of this score in predicting drug sensitivity in ESCC patients in the high and low-Ras groups.

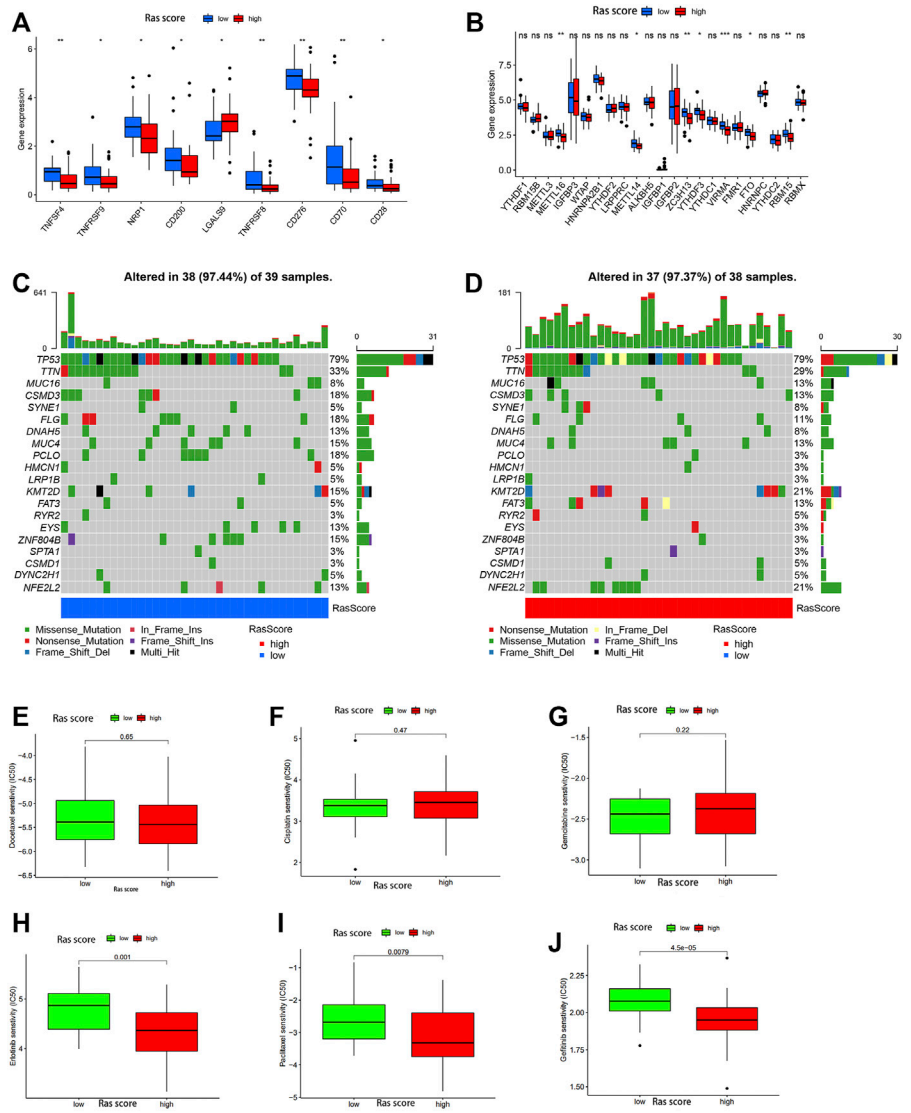


FIGURE 5 | (A,B) The expression of 30 immune checkpoints genes in high-Ras and low-Ras score patients. The upper and lower ends of the boxes represented interquartile range of values. The lines in the boxes represented median value, and black dots showed outliers. The asterisks represented the statistical *p* value (ns $P > 0.05$; *$p < 0.05$; **$p < 0.01$). The one-way ANOVA test was used to test the statistical differences among high-Ras score and low-Ras score patients. **(C,D)** The waterfall plot of tumor somatic mutation established by those with low Ras-score **(C)** and high Ras-score **(D)**. **(E–J)** The sensitivity to drugs in high-Ras and low-Ras score patients: **(E)** Docetaxel, **(F)** Cisplatin, **(G)** Gemcitabine, **(H)** Erlotinib, **(I)** Paclitaxel, **(G)** Gefitinib.

These results showed that there was no significant difference in the sensitivity of these samples to traditional chemotherapeutic drugs such as cisplatin, gemcitabine and docetaxel between the high and low Ras-score groups (**Figures 5E–G**), while paclitaxel had better sensitivity in the high Ras-score group. It is worth noting that targeted drugs such as erlotinib and gefitinib also had better sensitivity in the high Ras-score group (p <0.05) (**Figures 5H–J**).

DISCUSSION

ESCC is a progressive disease with a poor prognosis. At present, clinicians primarily use the TNM staging system to evaluate the prognosis of patients with cancer. However, most studies have revealed heterogeneity in the prognosis of tumours at the same stage. Accumulating evidence suggests that the Ras mutation rate can regulate the process of tumour development and progression, and is also involved in regulating the immune response to these tumours. (Masliah-Planchon et al., 2016; Ryan and Corcoran, 2018; Zhang et al., 2018; Chen et al., 2019; Prior et al., 2020). However, systematic analysis of RAS in ESCC is still rare, and the underlying mechanism remains unclear. (Feng et al., 2018; Li et al., 2019; Feng et al., 2020).

Our initial investigation identified 11 special survival-related Ras-associated genes which could be used to classify three distinct Ras modification patterns. There was a significant difference in

the Ras gene expression profile between each of these three patterns and our evaluations revealed that Ras-cluster three had a particularly significant survival advantage over both Ras-clusters two and 1. Ras-cluster three exhibited significant increases in the expression of RAC3, PDGFRA, EGFR, PTPN11, and NF1. This correlates with the current therapeutic focus on EGFR which includes clinical trials for advanced EC with high HER-2 expression, showing the therapeutic potential of EGFR targets in oesophageal cancer. (Bang et al., 2010; Yan-Ming Yang et al., 2020; Moehler et al., 2020). PDGFRA has demonstrated its potential as a therapeutic target for gastrointestinal stromal tumours (Theiss and Contreras, 2019) and these studies are consistent with the results of our study which suggest that the increased expression of these genes improves the survival potential of patients in Ras-cluster 3.

PD-1, PD-L1, and CTLA4 expression levels were also all increased in Ras-cluster three when compared with the others, while Ras-cluster two had a much higher rate of immune and stromal cell infiltration than either Ras-cluster one or Ras-cluster 3. This included a significant increase in the number of M1 and M2 macrophages and T cells in Ras-cluster2, which was consistent with the immune target expression and immune scores for these patients. Previous studies have shown that tumours with an immunoinflammatory phenotype also exhibit large numbers of stromal and immune cells despite the fact that stromal activation in the TME is thought to be inhibited by increased numbers of T-cells. (Chen and Mellman, 2017). Ras-cluster two was characterised by an immunoinflammatory type, specifically characterised by adaptive immune cell infiltration and immune activation, resulting in likely immune depletion in these patients, which might explain its poor survival potential.

We used the TCGA-ESCC and GTEx-normal sets to identify differentially expressed genes and then specific prognostic Ras-related genes using multiple Cox regression analysis and then used these to build a predictive model for ESCC comprising just three differentially expressed Ras-associated genes.

Kaplan-Meier curve analysis showed that cases in the high Ras-score group were associated with reduced OS when compared with patients in the low Ras-score group. Our ROC values illustrate the solid performance of this prognostic model, with the AUC of these curves for 3-years OS in the TCGA cohort being recorded at 0.932. In addition, we used these ROC curves to compare the predictive efficacy of our proposed model with other clinical features and revealed that our model was more predictive than other prognostic factors for ESCC, such as gender, stage, T stage, N stage and M stage.

It has been shown that KMT2D mutations are associated with increased tumour size and unfavorable prognosis in patients with EC. (Zheng et al., 2021). A phase 3 trial of mantle cell lymphoma showed that mutations in KMT2D mutation were associated with an increased risk of death in these patients. (Ferrero et al., 2020). FAT3 mutations are related to poor prognosis in oesophageal cancer, (Guo et al., 2021), while NFE2L2 mutations were significantly associated with a worse prognosis in ESCC. (Cui et al., 2020). In patients with oesophageal cancer, metabolic reprogramming of the glutathione metabolism, as well as detoxification of ROS by activation of NFE2L2, enhances cancer progression, leading to poor clinical outcomes. (Kitano et al., 2018). These results explain why the prognosis of patients was worse in the high Ras-score group with higher rates of certain gene mutations, such as KMT2D, FAT3, and NFE2L2, further confirming the accuracy and effectiveness of the Ras-score model in predicting patient prognosis. Our results show that the Ras score is a good prognostic index for ESCC when compared to traditional staging. The results that paclitaxel, erlotinib and gefitinib had better sensitivity in the high Ras-score group also show the capacity of Ras score to predict these drugs treatments value in ESCC patients. However, this study might still have its limitations, especially given that the number of training set samples was relatively small (78 samples) and a validation set was not available.

In summary, we identified differentially expressed Ras-related genes that may be involved in ESCC pathogenesis. These genes are of significant value in predicting OS in ESCC patients and may provide new options for individualized therapy. Further studies are necessary to verify the results of our evaluations and future work should include both *in vitro* and *in vivo* validations.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

1) Data curation, Methodology, Investigation, Writing-original draft, Writing-review and editing: H-SY and WL. 2) Writing-original draft and editing: S-YZ and H-YC. 3) Supervision: H-HL. 4) Conceptualization, Writing-review and editing, Supervision: Y-FF and Y-YL.

FUNDING

This work was supported by Grants from the Natural Science Foundation of Guangdong Province (No. 2019A1515011329).

REFERENCES

Abnet, C. C., Arnold, M., and Wei, W.-Q. (2018). Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 154, 360–373. doi:10.1053/j.gastro.2017.08.023

Arnold, M., Soerjomataram, I., Ferlay, J., and Forman, D. (2015). Global Incidence of Oesophageal Cancer by Histological Subtype in 2012. *Gut* 64, 381–387. doi:10.1136/gutjnl-2014-308124

Bang, Y.-J., Van Cutsem, E., Feyereislova, A., Chung, H. C., Shen, L., Sawaki, A., et al. (2010). Trastuzumab in Combination with Chemotherapy versus Chemotherapy Alone for Treatment of HER2-Positive Advanced Gastric or Gastro-Oesophageal junction Cancer (ToGA): a Phase 3, Open-Label, Randomised Controlled Trial. *Lancet* 376, 687–697. doi:10.1016/s0140-6736(10)61121-x

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492

Chen, D. S., and Mellman, I. (2017). Elements of Cancer Immunity and the Cancer-Immune Set point. *Nature* 541, 321–330. doi:10.1038/nature21349

Chen, S., Li, F., Xu, D., Hou, K., Fang, W., and Li, Y. (2019). The Function of RAS Mutation in Cancer and Advances in its Drug Research. *Curr. Pharm. Des.* 25, 1105–1114. doi:10.2174/1381612825666190506122228

Cui, Y., Chen, H., Xi, R., Cui, H., Zhao, Y., Xu, E., et al. (2020). Whole-genome Sequencing of 508 Patients Identifies Key Molecular Features Associated with Poor Prognosis in Esophageal Squamous Cell Carcinoma. *Cell Res.* 30, 902–913. doi:10.1038/s41422-020-0333-6

Feng, Y.-F., Lei, Y.-Y., Lu, J.-B., Xi, S.-Y., Zhang, Y., Huang, Q.-T., et al. (2018). RIT1 Suppresses Esophageal Squamous Cell Carcinoma Growth and Metastasis and Predicts Good Prognosis. *Cell Death Dis.* 9, 1085. doi:10.1038/s41419-018-0979-x

Feng, Y., Yan, S., Huang, Y., Huang, Q., Wang, F., and Lei, Y. (2020). High Expression of RABL6 Promotes Cell Proliferation and Predicts Poor Prognosis in Esophageal Squamous Cell Carcinoma. *BMC Cancer* 20, 602. doi:10.1186/s12885-020-07068-w

Ferrero, S., Rossi, D., Rinaldi, A., Bruscaggin, A., Spina, V., Eskelund, C. W., et al. (2020). KMT2D Mutations and TP53 Disruptions Are Poor Prognostic Biomarkers in Mantle Cell Lymphoma Receiving High-Dose Therapy: a FIL Study. *Haematologica* 105, 1604–1612. doi:10.3324/haematol.2018.214056

Guo, Z., Yan, X., Song, C., Wang, Q., Wang, Y., Liu, X.-P., et al. (2021). FAT3 Mutation Is Associated with Tumor Mutation Burden and Poor Prognosis in Esophageal Cancer. *Front. Oncol.* 11, 603660. doi:10.3389/fonc.2021.603660

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* 28, 100. doi:10.2307/2346830

Hazra, A., and Gogtay, N. (2016). Biostatistics Series Module 3: Comparing Groups: Numerical Variables. *Indian J. Dermatol.* 61, 251–260. doi:10.4103/0019-5154.182416

Kitano, Y., Baba, Y., Nakagawa, S., Miyake, K., Iwatsuki, M., Ishimoto, T., et al. (2018). Nrf2 Promotes Oesophageal Cancer Cell Proliferation via Metabolic Reprogramming and Detoxification of Reactive Oxygen Species. *J. Pathol.* 244, 346–357. doi:10.1002/path.5021

Li, Q., Xu, A., Chu, Y., Chen, T., Li, H., Yao, L., et al. (2019). Rap1A Promotes Esophageal Squamous Cell Carcinoma Metastasis through the AKT Signaling Pathway. *Oncol. Rep.* 42, 1815–1824. doi:10.3892/or.2019.7309

Masliah-Planchon, J., Garinet, S., and Pasmant, E. (2016). RAS-MAPK Pathway Epigenetic Activation in Cancer: miRNAs in Action. *Oncotarget* 7, 38892–38907. doi:10.18632/oncotarget.6476

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Moehler, M., Maderer, A., Thuss-Patience, P. C., Brenner, B., Meiler, J., Ettrich, T. J., et al. (2020). Cisplatin and 5-fluorouracil with or without Epidermal Growth Factor Receptor Inhibition Panitumumab for Patients with Non-resectable, Advanced or Metastatic Oesophageal Squamous Cell Cancer: a Prospective, Open-Label, Randomised Phase III AIO/EORTC Trial (POWER). *Ann. Oncol.* 31, 228–235. doi:10.1016/j.annonc.2019.10.018

Moore, A. R., Rosenberg, S. C., McCormick, F., and Malek, S. (2020). RAS-targeted Therapies: Is the Undruggable Drugged? *Nat. Rev. Drug Discov.* 19, 533–552. doi:10.1038/s41573-020-0068-6

Prior, I. A., Hood, F. E., and Hartley, J. L. (2020). The Frequency of Ras Mutations in Cancer. *Cancer Res.* 80, 2969–2974. doi:10.1158/0008-5472.can-19-3682

Pylayeva-Gupta, Y., Grabocka, E., and Bar-Sagi, D. (2011). RAS Oncogenes: Weaving a Tumorigenic Web. *Nat. Rev. Cancer* 11, 761–774. doi:10.1038/nrc3106

Ryan, M. B., and Corcoran, R. B. (2018). Therapeutic Strategies to Target RAS-Mutant Cancers. *Nat. Rev. Clin. Oncol.* 15, 709–720. doi:10.1038/s41571-018-0105-0

Simanshu, D. K., Nissley, D. V., and McCormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. *Cell* 170, 17–33. doi:10.1016/j.cell.2017.06.009

Theiss, L., and Contreras, C. M. (2019). Gastrointestinal Stromal Tumors of the Stomach and Esophagus. *Surg. Clin. North Am.* 99, 543–553. doi:10.1016/j.suc.2019.02.012

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Yang, Y.-M., Hong, P., Xu, W. W., He, Q.-Y., and Li, B. (2020). Advances in Targeted Therapy for Esophageal Cancer. *Sig. Transduct. Target. Ther.* 5, 229. doi:10.1038/s41392-020-00323-3

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, Z., Ma, M., Hu, R., Xu, B., Zong, L., Wei, H., et al. (2018). RasGRP3, a Ras Guanyl Releasing Protein 3 that Contributes to Malignant Proliferation and Aggressiveness in Human Esophageal Squamous Cell Carcinoma. *Clin. Exp. Pharmacol. Physiol.* 45, 720–728. doi:10.1111/1440-1681.12926

Zheng, B., Song, Z., Chen, Y., and Yan, W. (2021). Genomic Analyses of Metaplastic or Sarcomatoid Carcinomas from Different Organs Revealed Frequent Mutations in KMT2D. *Front. Mol. Biosci.* 8, 688692. doi:10.3389/fmolb.2021.688692

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Liu, Zheng, Cai, Luo, Feng and Lei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fgene.2022.843579

A Comprehensive Pan-Cancer Analysis for Pituitary Tumor-Transforming Gene 1

Siming Gong[1,2†], *Changwu Wu*[2†], *Yingjuan Duan*[3], *Juyu Tang*[1] *and Panfeng Wu*[1*]

[1]Department of Orthopaedics, National Clinical Research Center of Geriatric Disorders, Xiangya Hospital of Central South University, Changsha, China, [2]Institute of Anatomy, University of Leipzig, Leipzig, Germany, [3]Faculty of Chemistry and Mineralogy, University of Leipzig, Leipzig, Germany

Pituitary tumor-transforming gene 1 (PTTG1) encodes a multifunctional protein that is involved in many cellular processes. However, the potential role of PTTG1 in tumor formation and its prognostic function in human pan-cancer is still unknown. The analysis of gene alteration, PTTG1 expression, prognostic function, and PTTG1-related immune analysis in 33 types of tumors was performed based on various databases such as The Cancer Genome Atlas database, the Genotype-Tissue Expression database, and the Human Protein Atlas database. Additionally, PTTG1-related gene enrichment analysis was performed to investigate the potential relationship and possible molecular mechanisms between PTTG1 and tumors. Overexpression of PTTG1 may lead to tumor formation and poor prognosis in various tumors. Consequently, PTTG1 acts as a potential oncogene in most tumors. Additionally, PTTG1 is related to immune infiltration, immune checkpoints, tumor mutational burden, and microsatellite instability. Thus, PTTG1 could be potential biomarker for both prognosis and outcomes of tumor treatment and it could also be a promising target in tumor therapy.

Keywords: PTTG1, prognosis, immune infiltration, enrichment analysis, big data, pan-cancer analysis

OPEN ACCESS

Edited by:
Changwei Li,
Shanghai Jiao Tong University, China

Reviewed by:
Qianqian Song,
Wake Forest School of Medicine, United States
Shakuntala Baichoo,
University of Mauritius, Mauritius

***Correspondence:**
Panfeng Wu
wupanfeng@csu.edu.cn

†*These authors share first authorship.*

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *26 December 2021*
Accepted: *08 February 2022*
Published: *25 February 2022*

Citation:
Gong S, Wu C, Duan Y, Tang J and Wu P (2022) A Comprehensive Pan-Cancer Analysis for Pituitary Tumor-Transforming Gene 1. Front. Genet. 13:843579. doi: 10.3389/fgene.2022.843579

INTRODUCTION

Pituitary tumor-transforming gene 1 (PTTG1) encodes a protein which is involved in many cellular processes (Caporali et al., 2017; Lin et al., 2019a; Zhi et al., 2019). PTTG1, also known as human securin, plays a major role in the regulation of mitosis, especially during sister chromatid separation (Hatcher et al., 2014). Additionally, the PTTG1 protein engages in various other cellular processes, such as DNA damage/repair, apoptosis, and metabolism (Vlotides et al., 2007). According to previous studies, PTTG1 is normally viewed as an oncogene, and high expression of PTTG1 can promote tumorigenesis in many kinds of tumors such as prostate and bladder tumors (Wondergem et al., 2012; Xiang et al., 2017; Cui et al., 2020; Fraune et al., 2020). In particular, PTTG1 is one of the 17 gene signatures that can predict metastasis and prognosis in various tumor types (Ramaswamy et al., 2003). Some previous experiment-based studies have also indicated that PTTG1 is involved in various tumor processes, such as growth and metastasis, by revealing that PTTG1 overexpression is linked to proliferation and invasiveness (Zhang et al., 2014; Yan et al., 2015; Lin et al., 2016; Meng et al., 2020). Interestingly, the downregulation or silencing of PTTG1 could lead to the opposite result (Caporali et al., 2012; Cui et al., 2020; Meng et al., 2020; Zhou et al., 2020).

It has been shown that PTTG1 is involved in various molecular mechanisms such as epithelial-mesenchymal transition (EMT), PI3K/AKT signaling, and mitogen-activated

protein kinase (MAPK) signaling pathways (Lin et al., 2016; Feng et al., 2017; Huang et al., 2018; Hu et al., 2019; Cui et al., 2020). EMT is the process by which epithelial cells may lose their own features, such as polarity and adherence ability, and their cellular migration and invasion ability are enhanced. EMT is a natural procedure in wound repair and inflammation; however, EMT could also be a critical process that could lead to tumor formation (Suarez-Carmona et al., 2017). MAPK, which is a member of the serine-threonine kinase family, is linked to cell proliferation, while PTTG1 is the regulator of mitosis (Fang and Richardson, 2005; Hatcher et al., 2014). Growing evidence has revealed that the MAPK pathway is associated with many processes, such as tumor formation, invasion, and other tumor-related behaving (Burotto et al., 2014; Sun et al., 2019; Lee et al., 2020). In addition, PTTG1 could also be a potential target in tumor therapy for various types of cancer, such as melanoma, breast cancer, and ovarian cancer (Nakachi et al., 2016; Caporali et al., 2017; Meng et al., 2020). PTTG1 is also involved in mitosis and linked to tumor formation, prognosis, and cancer treatment in various tumors. Our previous study showed that the regulator of chromatin condensation 1 (RCC1), which is also involved in the cell cycle, plays a key role in human pan-cancer (Wu et al., 2021). The PTTG1 is linked to many tumors and various of pathways which could be the potential mechanism for tumor formation. Given the important tumor-promoting role of PTTG1 in a variety of tumors, it is necessary to comprehensively understand the potential of PTTG1 as a potential therapeutic target in different tumors through pan-cancer analysis. The aim of this study was to perform a preliminary analysis of PTTG1 expression, mutational information, prognostic value and potential tumor regulatory mechanisms by integrating tumor genome data from multiple public databases.

Specifically, in the present study, the analysis of gene alteration, PTTG1 expression, prognostic function, and PTTG1-related immune analysis in 33 types of tumors was performed based on various databases such as The Cancer Genome Atlas (TCGA) database, the Genotype-Tissue Expression (GTEx) database, and the Human Protein Atlas (HPA) database. Additionally, PTTG1-related gene enrichment was analyzed to investigate the potential relationship and possible molecular mechanisms between PTTG1 and tumors.

MATERIALS AND METHODS

Gene Alteration Analysis

cBioPortal (https://www.cbioportal.org/) is a user-friendly tool that can be used to analyze genetic alterations in different tumors based on public cohort. In this study, the "TCGA Pan Cancer Atlas Studies" was chosen, the name of the target gene "PTTG1" was typed in the input box to query the summary of genetic alterations, including the types of alteration and copy number data. In addition, the mutation site data could be obtained under the "mutation" section of cBioPortal.

The Gene Expression at Transcription and Translation Level

TIMER2.0 tool (http://timer.cistrome.org/) is an easy-to-use online tool that can be used to perform a systemic and comprehensive analysis of gene expression, relation between different genes, and immune-related analysis including the cancer-associated fibroblasts (CAF) and other immune cell infiltration in various types of tumors. In the present study, the analysis of PTTG1 expression at the transcription level in various tumors compared to corresponding normal tissues was analyzed based on the TCGA database. Since the normal tissue was limited in some types of tumors, supplementary analysis for these tumors was performed using the GEPIA2 tool (http://gepia2.cancer-pku.cn/#analysis). Using the "Expression Analysis" module of GEPIA2, the expression of PTTG1 in tumors and corresponding normal tissues was obtained according to the TCGA and GTEx datasets.

The Human Protein Atlas (HPA) (https://www.proteinatlas.org/) dataset is a high-quality database that provides researchers with staining-based expression and RNA sequence data. PTTG1 expression at the translation level could be obtained under the "Pathology" and "Tissue" modules. Additionally, an RNA expression overview could also be obtained from the HPA database.

Survival Analysis

The overall survival (OS) and disease-free survival (DFS) of all involved cancers based on the TCGA dataset could be analyzed using GEPIA2. In this study, the "survival analysis" section of GEPIA2 was used, and the patients were split into two groups at 50–50% cutoff based on PTTG1 expression. Then, the survival map and corresponding Kaplan-Meier curves were obtained together with the p value and hazard ratio (HR). The Kaplan-Meier plotter tool (http://kmplot.com/analysis/) was used to harvest the OS, DFS, post-progression survival (PPS), distant metastasis-free survival (DMFS), and progression-free survival (PFS) according to the Gene Expression Omnibus (GEO) database as supplement analysis. "Autoselect best cutoff" was set in Kaplan-Meier plotter tool and the patient would be separate into 2 separate groups to obtain the survival curves. The log-rank p value, HR, and 95% confidence intervals were also calculated.

Construction of Related Gene Network

The STRING tool (https://string-db.org/) is a website tool used to explore protein-protein interactions (PPIs), which are already known or predicted. The PPI network of PTTG1 was constructed using the STRING tool. We identified 39 proteins that directly interacted with PTTG1, and 39 proteins were identified through experiments. Additionally, the GEPIA2 tool was employed to acquire the top 100 genes related to PTTG1 based on all tumor data from the TCGA cohort. The top five genes correlated with PTTG1 were selected to explore the correlation between them and PTTG1. The "Correlation Analysis" section of GEPIA2 was used to obtain the p values, correlation coefficient values and dot plots. Spearman's correlation tests were also applied to these five genes using the TIMER2.0 tool. After a cross-analyzed between

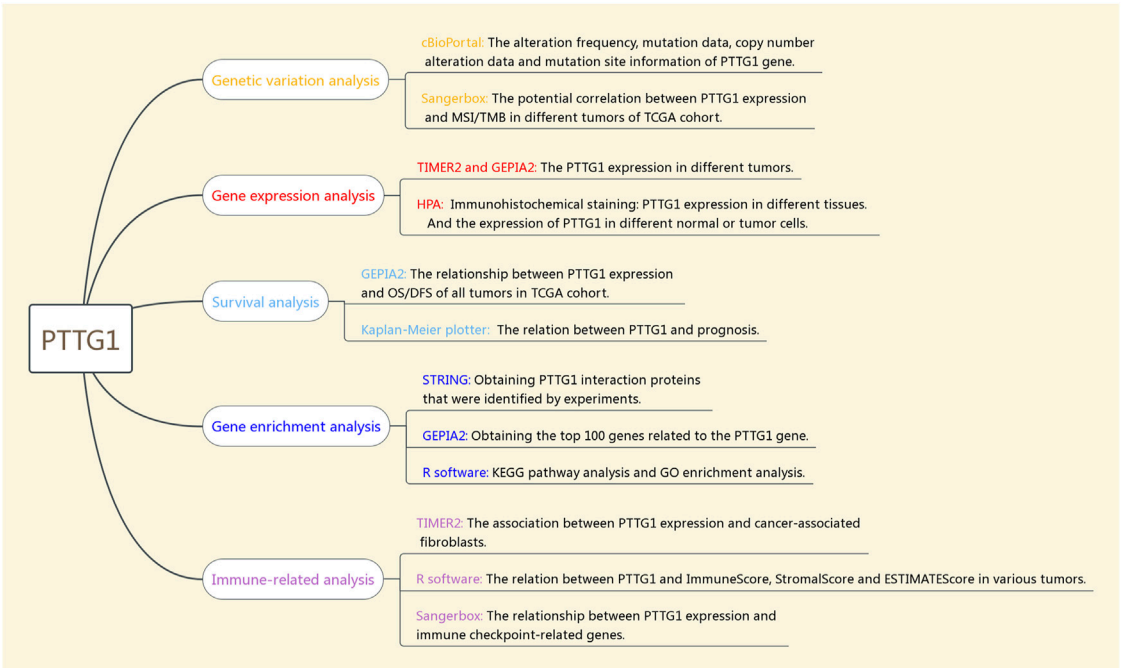


FIGURE 1 | Setup of the integrative and comprehensive Pan-cancer Analysis of PTTG1.

PTTG1-interacted genes and PTTG1-correlated genes, a Venn diagram was acquired. GeneMANIA (https://genemania.org/) was used to identify potential PTTG1 related genes.

KEGG Pathway and GO Enrichment

Based on the combination of PTTG1-interacted genes and PTTG1-correlated genes the R package "clusterProfifiler" was employed to achieve the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Geno Oncology (GO) enrichment analysis including biological process, cellular component, and molecular function.

Immune-Related Analysis

The TIMER2.0 tool was also employed to investigate the correlation between PTTG1 and cancer-associated fibroblasts (CAFs) across diverse tumors based on the TCGA dataset. The "cancer-associated fibroblasts" was chosen in the immune section and the "PTTG1" was input into the Gene expression box. Then, the CAF heat map and corresponding dot plots were obtained.

The Sangerbox tool (http://sangerbox.com/Tool) is a powerful tool that can be used to perform many biological information analyses. In this study, the sangerbox was used to determine the relationship between PTTG1 expression and immune checkpoint, microsatellite instability (MSI), and tumor mutational burden (TMB).

The UCSC website (https://xenabrowser.net/) was used to obtain the pan-cancer data, which included 11,060 samples. In addition, the expression of PTTG1 was analyzed based on the samples after normalization and log2 (x+0.001) transformed. In addition, we selected the gene expression profiles of each tumor and mapped the profiles to GeneSymbol. The R package ESTIMATE was used to calculate the StromalScore, ImmuneScore, and ESTIMATEScore for each tumor (Yoshihara et al., 2013).

RESULTS

The Involved Tumor Types and Genetic Alteration Analysis Data

The tumors that were involved in this study and the corresponding abbreviations were shown in **Supplementary Table S1**. The workflow of the steps for the present study was shown in **Figure 1**. And the number of patients from TCGA were shown in **Supplementary Table S2**. cBioPortal was used to obtain the PTTG1 alteration type and the corresponding frequency. The alteration frequency was 6.26% in kidney renal clear cell carcinoma (KIRC), and most of the frequency was amplified at 6.06% compared to the mutation at 0.2%. Of note, the amplification was the only alteration type in cholangiocarcinoma, sarcoma (SARC), adrenocortical carcinoma (ACC), pancreatic adenocarcinoma (PAAD), thymoma (THYM), cervical squamous cell carcinoma, and thyroid carcinoma (**Figure 2A**). In addition, mutation types, together with the site of PTTG1, were also achieved (**Figure 2B**). There are 31 variants of uncertain/unknown significance (VUS) in PTTG1. Missense mutations were the most common mutations in PTTG1, followed by truncating, splice, and fusion.

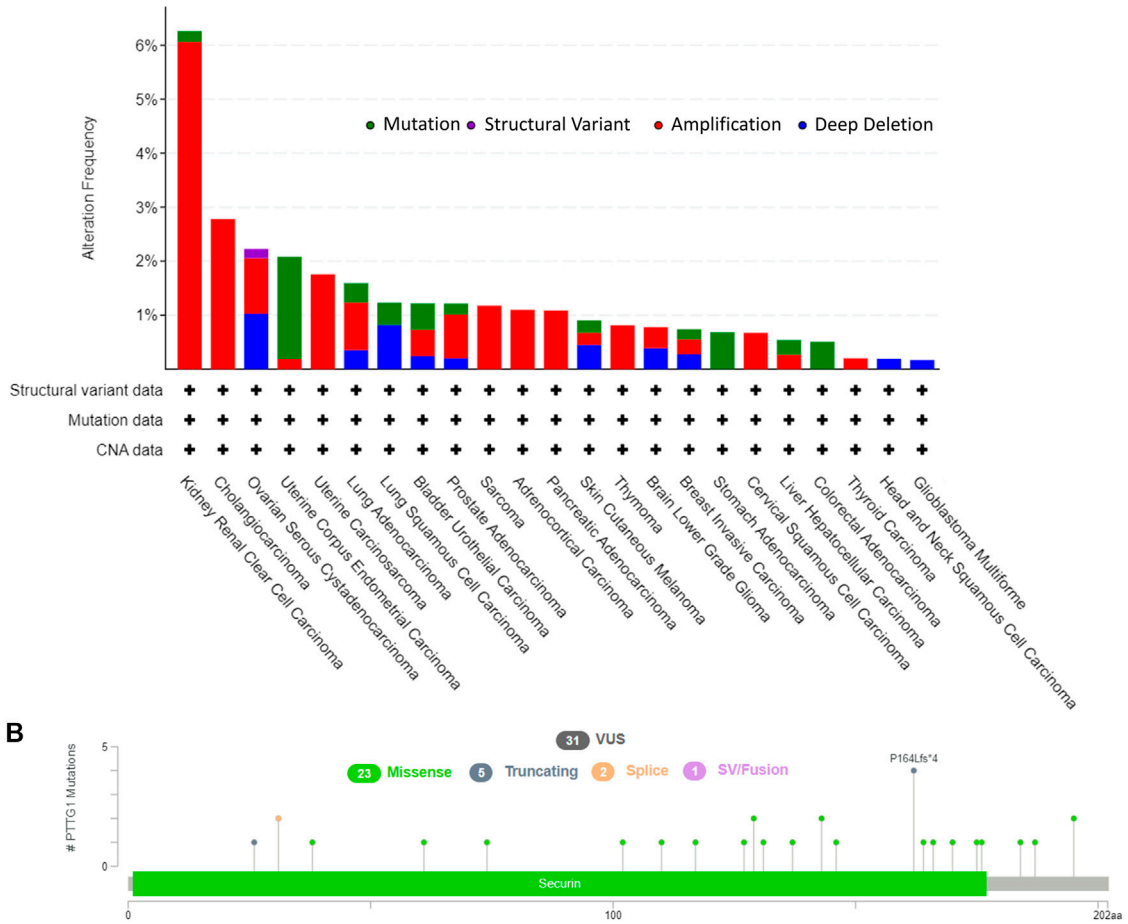


FIGURE 2 | The genetic alteration of PTTG1 in various of tumors. **(A)** The alteration frequency with corelative type was displayed for each tumor. **(B)** The mutation types, sites and case number of PTTG1 genetic alteration were displayed. VUS: variant of uncertain (or unknown) significance.

The Gene Expression Results

First, PTTG1 expression in normal tissues, single cells, and various tumor tissues was analyzed based on a consensus database (**Supplementary Figure S1**). The consensus database is a high-quality database that includes many other databases, such as HPA, GTEx, and FANTOM five cohorts. This shows that PTTG1 expression was enhanced in the bone marrow, lymphoid tissue, and testis with normal tissue specificity (**Supplementary Figure S1A**). Additionally, PTTG1 was overexpressed in some cell types, such as early spermatids and spermatocytes, with RNA single cell type specificity (**Supplementary Figure S1B**). However, the expression of PTTG1 showed low cancer specificity in the RNA cancer category (**Supplementary Figure S1C**). Therefore, PTTG1 expression was at the same level among different cancer categories, but PTTG1 may be overexpressed in some normal tissues or cells compared to other normal tissues.

Using the TIMER2.0 tool, the expression of PTTG1 in diverse tumors and corresponding normal tissues was obtained based on the TCGA database. PTTG1 expression in bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC) were higher than those in normal tissues (**Figure 3A**). However, it could be found that the PTTG1 expression was lower in Thyroid carcinoma (THCA) than the normal tissue ($p < 0.001$).

As the normal samples in some tumor, such as BRCA and PAAD, were not enough to analyze, we did the supplementary analysis. Using the GEPIA2 tool, the supplementary analysis of

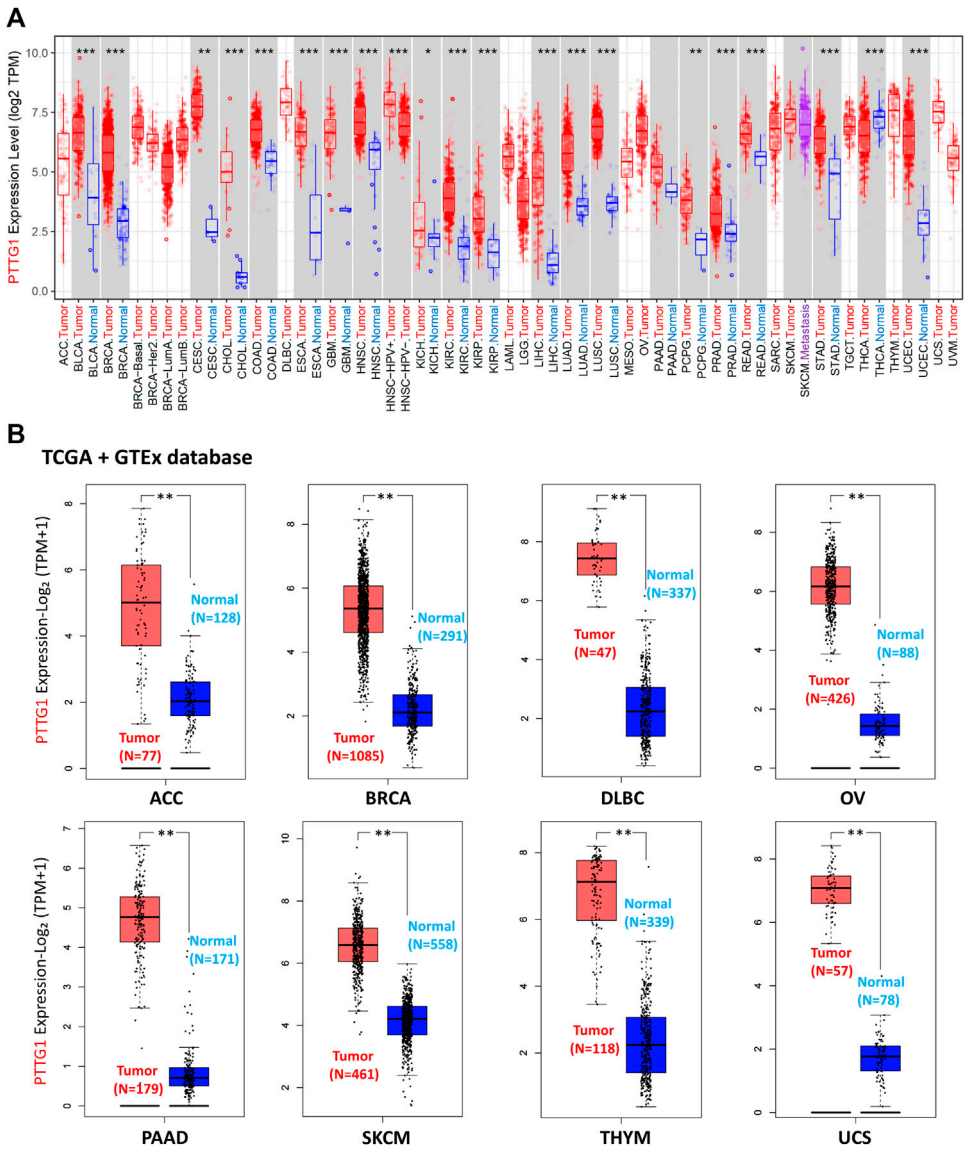


FIGURE 3 | The expression of PTTG1 in different type of tumors at transcription level. **(A)** The PTTG1 expression in different tumors based on TCGA database. **(B)** The expression level of PTTG1 in various type of tumors based on the combination of TCGA and GTEx cohort. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

PTTG1 expression between tumor and normal tissues depended on the combination of TCGA and GTEx datasets was achieved (**Figure 3B** and **Supplementary Figure S2**). PTTG1 expression in ACC, BRCA, lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), ovarian serous cystadenocarcinoma (OV), PAAD, skin cutaneous melanoma (SKCM), THYM, uterine carcinosarcoma (UCS), and sarcoma (SARC) was higher than that in normal tissues (all $p < 0.01$) (**Figure 3B** and **Supplementary Figure S2**). However, PTTG1 expression in acute myeloid leukemia (LAML) and testicular germ cell tumors (TGCT) was lower than that in normal tissues (all $p < 0.01$) (**Supplementary Figure S2**).

Additionally, using the HPA database, immunohistochemical staining of PTTG1 in different tumors and normal tissues was performed. PTTG1 expression at the protein level was higher in various tumors, including BRCA, COAD, and lung adenocarcinoma (LUAD) (**Figure 4**).

Survival Analysis

The data from TCGA database were separated into two groups: the 50% higher expression of PTTG1 group and the 50% lower expression of PTTG1 group. The correlation between the expression of PTTG1 and prognosis across different types of tumors was analyzed. Patients with high expression of PTTG1 were associated with poor overall survival (OS) in ACC, KIRC, KIRP, lower grade glioma (LGG), LIHC, LUAD, mesothelioma (MESO), PAAD, THCA, and uveal melanoma

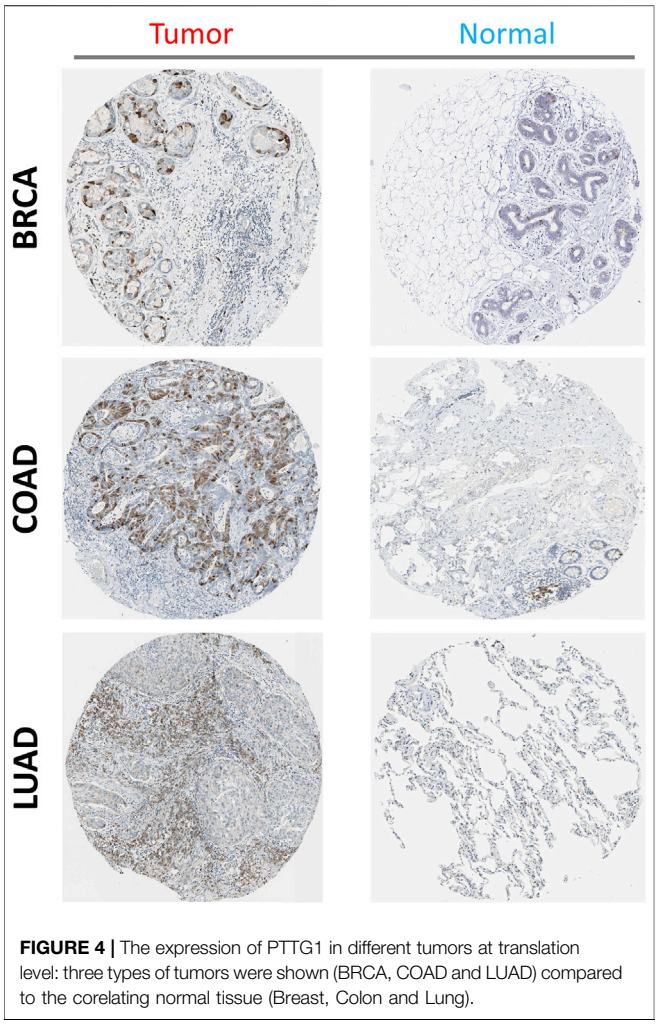


FIGURE 4 | The expression of PTTG1 in different tumors at translation level: three types of tumors were shown (BRCA, COAD and LUAD) compared to the corelating normal tissue (Breast, Colon and Lung).

(UVM) (**Figure 5A**, all $p < 0.05$). Among these types of tumors, the Kaplan–Meier curves of KIRC and LGG are shown. Patients with high expression of PTTG1 were associated with poor disease-free survival (DFS) in the ACC, KIRC, KIRP, LGG, LIHC, MESO, PAAD, PRAD, SARC, and UVM (**Figure 5B**). The Kaplan–Meier curves of LIHC and SARC are shown.

However, no significance could be achieved in other types of tumors if we separated PTTG1 expression as a 50–50% cutoff. Therefore, using the Kaplan–Meier Plotter tool, a further analysis was performed by setting the cutoff as "Autoselect best cutoff". The high expression of PTTG1 was linked to poor prognosis in BLCA, BRCA, CESC, and STAD (**Supplementary Figure S3**). It indicated that PTTG1 may be involved in the malignant progression of a variety of tumors and may serve as a prognostic biomarker.

Gene Enrichment Analysis

To further explore the potential mechanism of PTTG1 in tumor development and adverse clinical outcomes, the protein-protein interaction (PPI) network analysis was performed and PTTG1-interacted and PTTG1-correlated genes were extracted. The PPI network and potential related gene network analyzed by GeneMANIA are shown in **Figure 6A** and **Supplementary Figure S4**. Using the STRING and GEPIA2 tools, 39 PTTG1 interacted genes and 100 PTTG1 correlated genes were identified (**Supplementary Table S3**). The top five PTTG1-correlated genes were kinesin family member C1 (KIFC1), kinesin family member 2C (KIF2C), cell division cycle 20 (CDC20), cyclin B1(CCNB1), and aurora kinase B (AURKB). These five genes were analyzed to explore the relationship between their expression and PTTG1 expression in each type of tumors (**Figure 6B**). Additionally, the relationship between their expression and PTTG1 expression across various tumors was also analyzed (**Figure 6C**). After the cross analysis between the interacting and correlated genes of PTTG1, three genes were identified: cyclin dependent kinase 1 (CDK1), CDC20, and ubiquitin conjugating enzyme E2 C (UBE2C) (**Figure 6D**).

By using the combination of the PTTG1 interacting and correlated gene sets, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Geno Oncology (GO) enrichment analysis was performed. KEGG pathway analysis showed that PTTG1 was linked to cell cycle, oocyte meiosis, human immunodeficiency virus 1 (HIV-1) infection, glioma, and mismatch repair (**Figure 6E**). The GO enrichment analysis showed that PTTG1 was linked to organelle fission, sister chromatid segregation in biological process (BP) terms, chromosome, centromeric region, mitotic spindle in cellular component (CC) terms, protein serine/threonine kinase activity, and damaged DNA binding in molecular function (MF) terms (**Figure 6F**).

Immune-Related Analysis

Four different algorithms, EPIC, MCPCOUNTER, XCELL, and TIDE were used to analyze the relationship between PTGG1 expression and cancer-associated fibroblasts (CAFs). The same tendency could be obtained in most of the algorithms, which would be viewed as reliable. The expression of PTTG1 was positively related to CAF infiltration in KIRP and negatively related to CAF infiltration in various tumors such as BRCA, CESC, and COAD (**Figure 7A**). Typical scatter plots developed by the algorithm are shown in **Figure 7B**. The R package "psych" was used to analyze the relation between PTTG1 and immune infiltration score. Significant differences were observed in most of the involved tumors in all the ImmuneScore, StromalScore, and ESTIMATEScore (**Supplementary Table S4**). PTTG1 was positively correlated with all the ImmuneScore, StromalScore, and ESTIMATEScore in glioma (GBM + LGG) and pan-kidney cohorts (KICH + KIRC + KIRP) (**Figures 7C,D**, all $p < 0.05$).

Immune checkpoint analysis showed that PTTG1 was negatively correlated with most of the checkpoint genes in TGCT, while PTTG1 was positively correlated with immune checkpoints in some other tumors such as PRAD, KIRP, and KIRC (**Figure 8A**, all $p < 0.05$). In addition, the correlation between the expression of PTTG1 and MSI/TMB was also analyzed. PTTG1 expression was positively correlated with MSI in OV, LUSC, PRAD, UCEC, LIHC, SARC, BRCA,

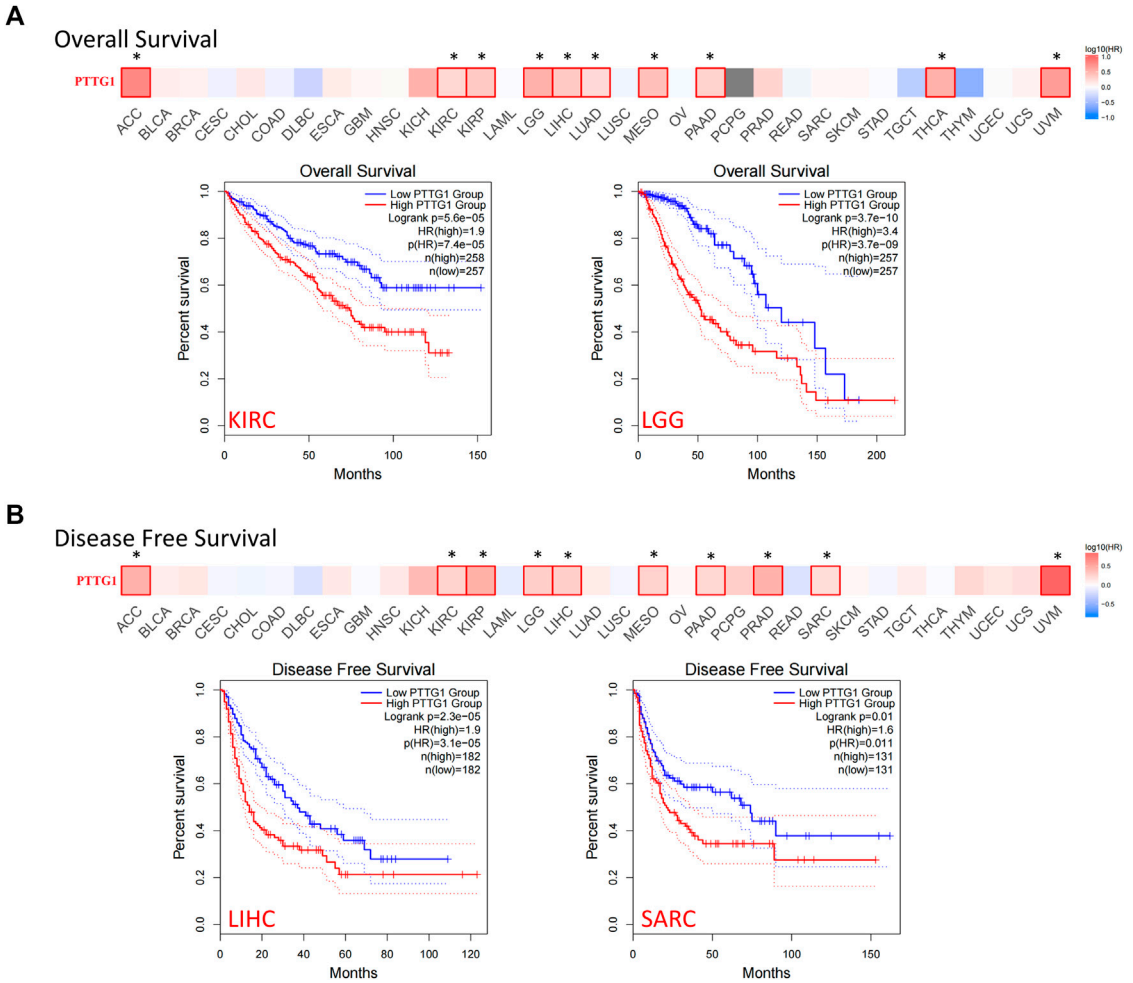


FIGURE 5 | The correlation between PTTG1 expression and survival prognosis of different tumor based on TCGA database. **(A)** The correlation between expression and overall survival (OS) in different tumors. **(B)** The correlation between PTTG1 expression and disease-free survival (DFS) in different tumors. The survival maps and Kaplan–Meier curves of two typical tumors were also shown. Dotted lines: 95% confidence interval. *$p < 0.05$.

COAD, STAD, HNSC, and DLBC (**Figure 8B**, all $p < 0.05$). PTTG1 expression was positively correlated with TMB in GBM, LUAD, PRAD, UCEC, BRCA, COAD, STAD, SKCM, KIRC, LGG, KICH, ACC, and PCPG (**Figure 8B**, all $p < 0.05$).

DISCUSSION

PTTG1 is a crucial gene involved in mitosis, especially during sister chromatid separation (Hatcher et al., 2014). Various research has shown that PTTG1 is linked to tumor formation and prognosis in numerous tumors, such as glioma, KIRC, and LIHC (Wondergem et al., 2012; Lin et al., 2019a; Zhi et al., 2019). But, however, the analysis between expression of PTTG1 and human pan-cancer was still poor. In the present study, using various tools such as TIMER2.0, GEPIA2, HPA, and sangerbox, a comprehensive analysis of PTTG1 was achieved based on the data from TCGA, GEO, and GTEx cohorts. The data were used to explore PTTG1 genetic alterations, gene expression, survival, and immune-related infiltration. In addition, gene enrichment was used to obtain a deeper insight into the potential mechanism of PTTG1 in numerous tumors. Hence, the relation and potential mechanism between PTTG1 and human pan-cancer could be obtained, which would also provide a novel approach for tumor treatment.

Genetic mutation analysis showed that the altered rate was more than 6% in KIRC followed by 2.78% in CHOL (**Figure 2**). In recent years, growing evidence has indicated that mutated-gene-target therapy is a promising approach for tumor therapy (Duffy et al., 2017; Synnott et al., 2017; Kaur et al., 2018). Thus, the PTTG1-tageted gene could also be a potential treatment for tumors such as KIRC and CHOL. PTTG1 expression analysis showed that the expression of PTTG1 was overexpressed in most of the involved tumors, such as BLCA, BRCA, and OV, based on TCGA and GTEx cohorts (**Figure 3**). Hence, our results further support previous studies showing that PTTG1 is an oncogene (Yoon et al., 2012; Noll et al., 2015). Interestingly, the expression of PTTG1 was lower in cancer than corresponding normal tissue in LAML and TGCT

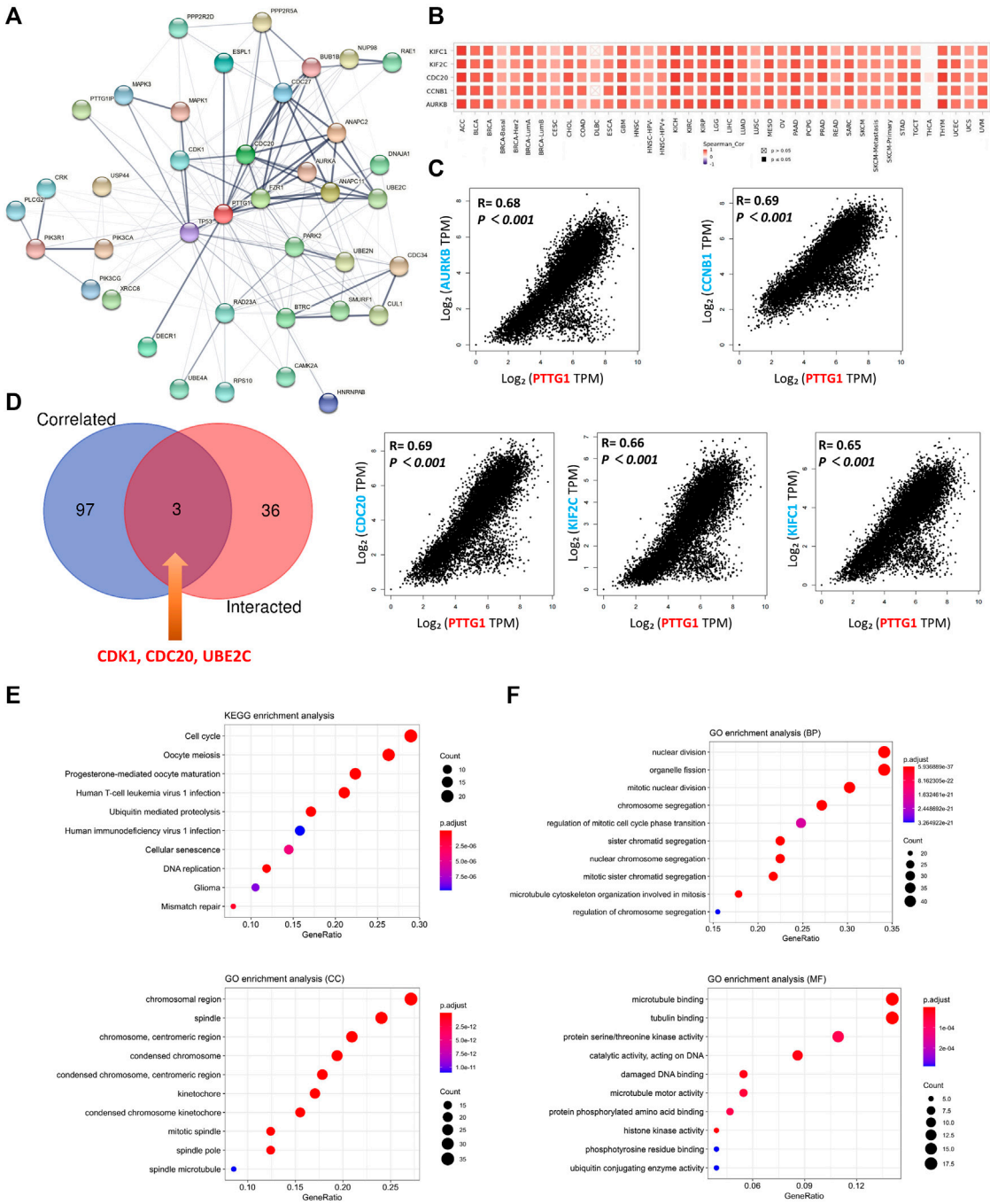


FIGURE 6 | The related gene analysis for PTTG1. **(A)** A protein-protein interaction (PPI) network of 39 PTTG1-interacted experimentally verified proteins were shown. **(B)** The top 100 genes which were linked to the PTTG1 expression were obtained, and the top five genes was shown, named AURKB, CCNB1, CDC20, KIF2C and KIFC1. The heatmap data which displayed the correlation of the top five genes with different tumors were shown. **(C)** The expression correlations between the top five genes and PTTG1 were also displayed. **(D)** The intersection analysis of the PTTG1-correlated and PTTG1-interacted genes was conducted and three genes were obtained, namely CDK1, CDC20 and UBE2C. **(E)** KEGG pathway analysis was obtained according to PTTG1-correlated and interacted genes; **(F)** The GO enrichment in biological process (BP) terms, cellular component (CC) terms, and molecular function (MF) terms.

(**Supplementary Figure S2**), which might mean the PTTG1 would be a "double-edged sword" for the PTTG1 was oncogene in most of tumors but also tumor suppressor gene in tumors such as LAML and TGCT. Although PTTG1 was lower expression in some tumors, it was always associated with poor prognosis. As the OS was normally viewed as the best index for tumor therapy, the DFS also plays a

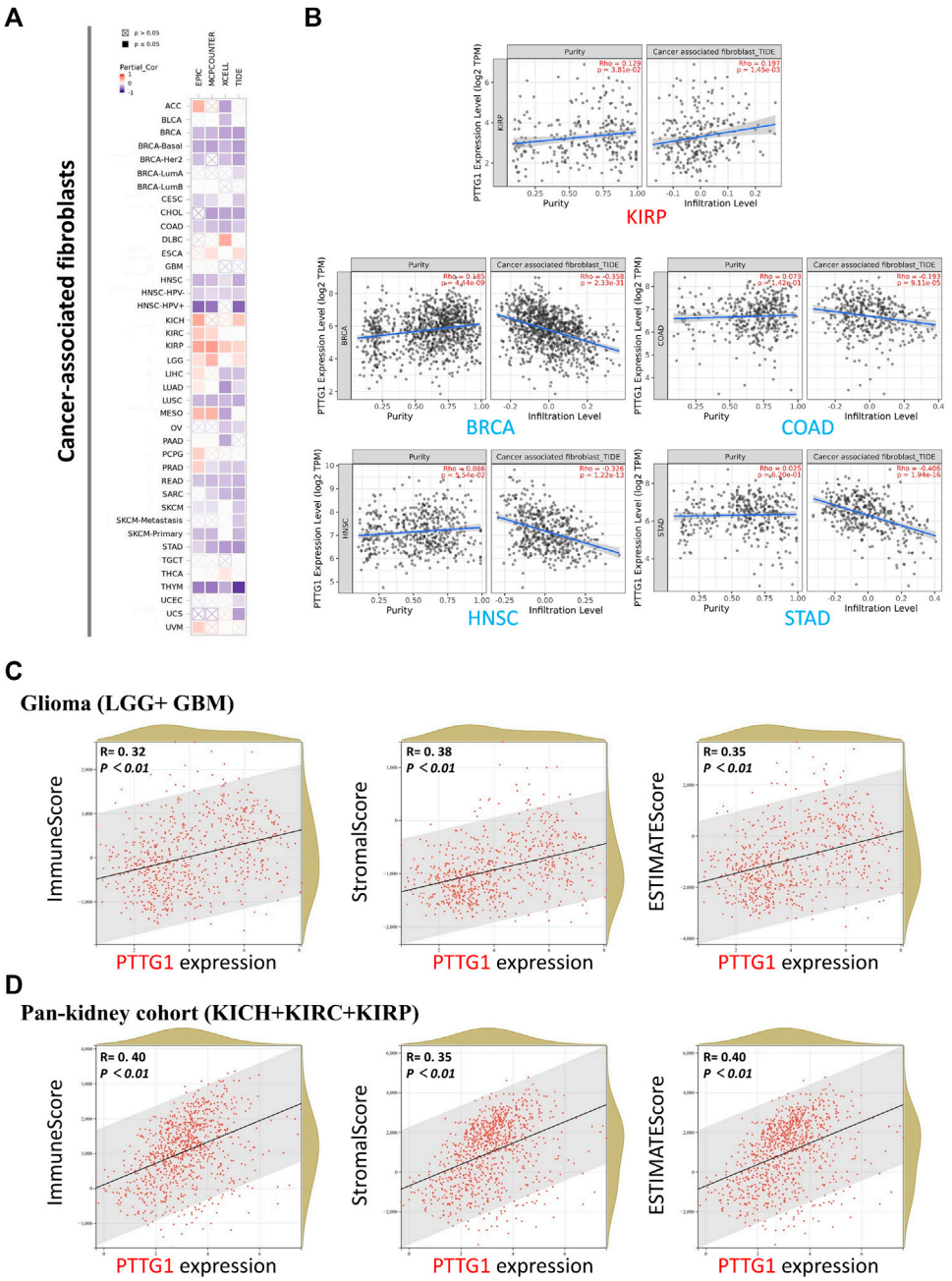


FIGURE 7 | Cancer-associated fibroblasts analysis and immune check point of PTTG1. **(A)** Association between PTTG1 expression and immune infiltration of cancer-associated fibroblasts in different tumors. **(B)** The scatter plots of cancer-associated fibroblasts immune infiltration in different tumors generated based on a certain algorithm. The analysis of correlation between PTTG1 expression and ImmuneScore, StromalScore and ESTIMATEScore was conducted. Two kinds of typical tumors **(C)** glioma (LGG + GBM) **(D)** Pan-kidney cohort (KICH + KIRC + KIRP) were displayed.

crucial role in surgical treatment and radiotherapy for it could be used to evaluate the outcome of the treatment. For the tumors such as ACC, KIRC and LGG, the high expression of PTTG1 was linked to both the OS and DFS. Thus, it makes the PTTG1 be a potential biomarker for these tumors in designing the treatment plan. Previous studies have shown that PTTG1 could be a prognostic biomarker in some tumors, and our results further extend their finding (Repo et al., 2017; Kaur et al., 2018; Romero Arenas et al., 2018; Fraune et al., 2020). Additionally, PTTG1 could also affect the drug sensitivity in some tumors such as BRCA, SKCM, and OV, which provided us with

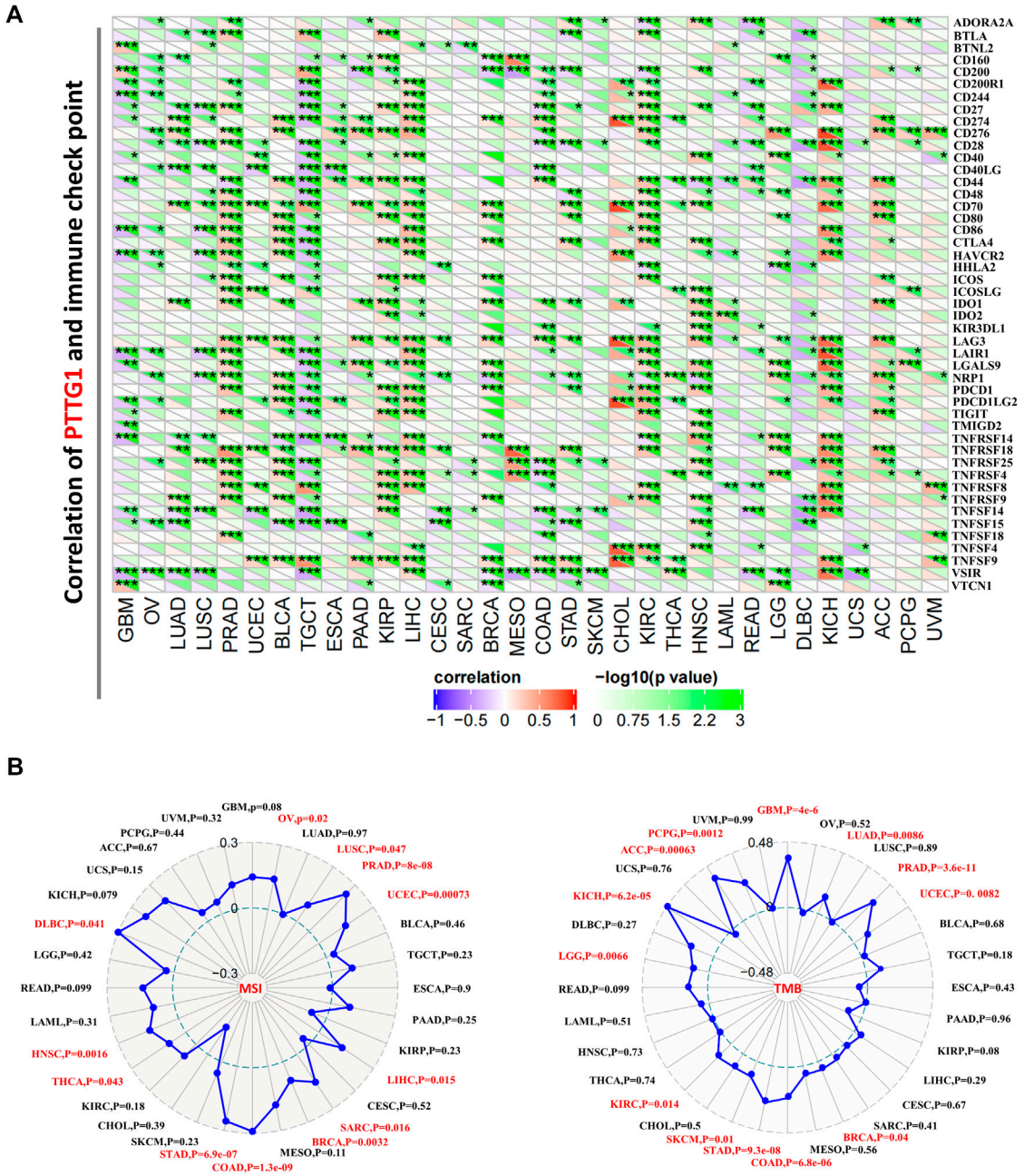


FIGURE 8 | (A) Association between PTTG1 expression and immune checkpoint genes expression in different tumors. **(B)** Correlation between PTTG1 expression and microsatellite instability (MSI) or tumor mutational burden (TMB). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

a novel treatment idea that the PTTG1-target drug or operation could be applied to the patient to enhance the chemotherapy outcomes (Nakachi et al., 2016; Caporali et al., 2017; Meng et al., 2020).

Using the cross analysis of PTTG1 interacting and correlated genes, three common genes were identified: CDK1, CDC20, and UBE2C (**Figure 6D**). The combination of PTTG1 and CDC20 has already been found and is considered to be a hub gene in some tumors such as PRAD, KIRC, and BRCA (Dai et al., 2021; Lin et al., 2019b; Deng et al., 2021). In addition, PTTG1, CDK1, and UBE2C are also viewed as hub genes in COAD (Deng et al., 2021). Moreover, the UBE2C-mediated P53 ubiquitination could be a potential target in LIHC therapy (Zhu et al., 2021). Interestingly, the combination of PTTG1 and the other three genes was not only linked to tumors like LIHC but also to acute type A aortic dissection (Luo et al., 2019; Jiang and Si, 2019). Consequently, there could be a potential mechanism between these four genes, which warrants further exploration. The KEGG pathway analysis indicated that PTTG1 was correlated with the cell cycle, DNA replication, and mismatch repair (**Figure 6E**). This

might reveal a potential mechanism between PTTG1 and tumor formation. EMT is a normal process during wound healing but may also contribute to fibrosis and cancer progression. During this process, epithelial cells may lose their polarity and adherence ability and become invasive (Lamouille et al., 2014). Overexpression of PTTG1 may lead to disorders in the process of cell proliferation, DNA replication, and mismatched DNA repair. Then, normal epithelial cells may become tumor cells and invade. The KEGG pathway also showed that PTTG1 was linked to human T-cell lymphotropic virus type-1 (HTLV-1) infection, and HTLV-1 is a retrovirus that is closely related to adult T-cell leukemia/lymphoma (ATL) (Guerrero et al., 2020). The cell cycle is necessary for the proliferation of HTLV-1, and the tax protein encoded by HTLV-1 may damage the mitotic spindle, resulting in incorrect sister chromatid separation (Levy et al., 2020). Consequently, it might be another potential mechanism between PTTG1/HTLV-1 and tumor formation, such as ALT. GO enrichment showed that PTTG1 was associated with mitosis and protein serine/threonine kinase activity (**Figure 6F**). MAPK, a member of the serine-threonine kinase family, is associated with many tumor processes, such as tumor formation and invasion (Fang and Richardson, 2005; Burotto et al., 2014; Suarez-Carmona et al., 2017; Sun et al., 2019; Lee et al., 2020). Hence, the MAPK pathway might also be a potential mechanism associated with PTTG1 and tumor formation (Noh et al., 2009; Liang et al., 2015; Hu et al., 2019).

CAF has been proven to be closely associated with various tumor procedures and is one of the most important parts of the tumor microenvironment (TME) stroma (**Figure 7A**) (Chen and Song, 2019; Sahai et al., 2020). Our study showed that PTTG1 was positively correlated with CAF in KIRP and LGG but negatively correlated with diverse tumors such as BRCA, COAD, and STAD. Recent research has indicated that CAF-target therapy could shape the TME (Affo et al., 2017; Kubo et al., 2016). Especially, previous studies have shown that the CAF could contribute to the remodeling of extracellular matrix in LIHC and the exosome could be a potential target to reverse chemoresistance (Zhang et al., 2020a; Zhang et al., 2020b). Consequently, CAF-target therapy with exosome could be a promising approach for tumor such as LIHC. Our results provide insight into CAF in various tumors, which could be helpful for further research in CAF and TME. It would be revolutionary changes that researchers have gained deep insight into the immune checkpoint in recent years (Topalian et al., 2016; Abril-Rodriguez and Ribas, 2017; Mezquita et al., 2018). Immune checkpoint inhibitors for cytotoxic T lymphocyte-associated protein-4 (CTLA-4) and programmed cell death protein-1 (PD-1, also known as PDCD1) have achieved great success (Rotte, 2019; Shukla et al., 2018). Our findings revealed that PTTG1 is positively correlated with both CTLA-4 and PD-1 in some tumors such as ACC, KICH, and HNSC, which may indicate that PTTG1 could be a novel biomarker for PD-1 in these tumors (**Figure 8A**). Immune infiltration analysis showed that PTTG1 was positively correlated with ImmuneScore, StromalScore, and ESTIMATEScore in tumors such as glioma and kidney tumors. This means that the overexpression of PTTG1 is linked to high infiltration of immune cells and stromal cells but low tumor purity in these tumors (**Figure 7** and **Supplementary Table S4**). Growing evidence has shown that TMB could be a possible biomarker for tumor sensitivity to immune checkpoint blockade (Chalmers et al., 2017; Chan et al., 2019). The microsatellite instability (MSI) and TMB of PTTG1 could be potential biomarkers for immune checkpoint inhibitors, such as PD-1 or CTLA-4. The PD-1 and CTLA-4 is now regarded as promising target in tumor treatment. Thus, the combination of MSI and TMB would be a novel approach to predict the outcome of treatment for tumors (Goodman et al., 2019; Luchini et al., 2019; Schrock et al., 2019; Jang et al., 2020). In the present study, PTTG1 was positively correlated with TMB, MSI, PD-1, and CTLA-4 in tumors such as PRAD, BRCA, and STAD. Consequently, immune checkpoint inhibitors may have a good outcome in these tumors. However, there are some limitations in this study. The data of RNA sequence and immunohistochemical staining was used to analyze the potential role of PTTG1 in human pan-cancer but the analysis of multi-omics is absence. Previous studies showed that multi-omics analysis including RNA sequencing, proteomic and metabolomic analysis could get a good outcome (Su et al., 2021). Consequently, it would be a promising method to analyze the role of PTTG1 in tumor using the multi-omics approach.

The present study shows the results of the comprehensive pan-cancer analysis data of PTTG1. And it is the first analysis that combined the expression of PTTG1, prognosis, KEGG pathway and GO enrichment analysis, immune-related analysis, immune checkpoint, TMB, and MSI. This study provides novel insights into the potential role of PTTG1 in tumor formation and prognostic function.

CONCLUSION

In summary, our findings show that PTTG1 acts as an oncogene in most of the involved tumors. Overexpression of PTTG1 may lead to a poor prognosis in various tumors. Additionally, PTTG1 is related to immune-related infiltration, immune checkpoints, TMB, and MSI. Consequently, we identified that PTTG1 could be a prognostic biomarker and a novel target for some tumors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization, SG, CW, YD, JT and PW; Data curation, SG, CW and YD; Formal analysis, SG, CW, JT and PW; Writing–original draft, SG, CW and PW; Writing–review and editing, SG, CW and PW.

FUNDING

This research was funded by National Natural Science Foundation of China, grant number 82072194.

ACKNOWLEDGMENTS

The results shown here are in part based upon data generated by TCGA Research Network (https://www.cancer.gov/tcga), GTEx database (https://commonfund.nih.gov/gtex) and HPA database (https://www.proteinatlas.org/humanproteome/pathology).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.843579/full#supplementary-material

REFERENCES

Abril-Rodriguez, G., and Ribas, A. (2017). SnapShot: Immune Checkpoint Inhibitors. *Cancer Cell* 31, 848. doi:10.1016/j.ccell.2017.05.010

Affo, S., Yu, L.-X., and Schwabe, R. F. (2017). The Role of Cancer-Associated Fibroblasts and Fibrosis in Liver Cancer. *Annu. Rev. Pathol. Mech. Dis.* 12, 153–186. doi:10.1146/annurev-pathol-052016-100322

Burotto, M., Chiou, V. L., Lee, J.-M., and Kohn, E. C. (2014). The MAPK Pathway across Different Malignancies: A New Perspective. *Cancer* 120, 3446–3456. doi:10.1002/cncr.28864

Caporali, S., Alvino, E., Lacal, P. M., Ruffini, F., Levati, L., Bonmassar, L., et al. (2017). Targeting the PTTG1 Oncogene Impairs Proliferation and Invasiveness of Melanoma Cells Sensitive or with Acquired Resistance to the BRAF Inhibitor Dabrafenib. *Oncotarget* 8, 113472–113493. doi:10.18632/oncotarget.23052

Caporali, S., Alvino, E., Levati, L., Esposito, A. I., Ciomei, M., Brasca, M. G., et al. (2012). Down-regulation of the PTTG1 Proto-Oncogene Contributes to the Melanoma Suppressive Effects of the Cyclin-dependent Kinase Inhibitor PHA-848125. *Biochem. Pharmacol.* 84, 598–611. doi:10.1016/j.bcp.2012.06.004

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 Human Cancer Genomes Reveals the Landscape of Tumor Mutational burden. *Genome Med.* 9, 34. doi:10.1186/s13073-017-0424-2

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30, 44–56. doi:10.1093/annonc/mdy495

Chen, X., and Song, E. (2019). Turning Foes to Friends: Targeting Cancer-Associated Fibroblasts. *Nat. Rev. Drug Discov.* 18, 99–115. doi:10.1038/s41573-018-0004-1

Cui, L., Ren, T., Zhao, H., Chen, S., Zheng, M., Gao, X., et al. (2020). Suppression of PTTG1 Inhibits Cell Angiogenesis, Migration and Invasion in Glioma Cells. *Med. Oncol.* 37, 73. doi:10.1007/s12032-020-01398-2

Dai, L., Song, Z.-X., Wei, D.-P., Zhang, J.-D., Liang, J.-Q., Wang, B.-B., et al. (2021). CDC20 and PTTG1 Are Important Biomarkers and Potential Therapeutic Targets for Metastatic Prostate Cancer. *Adv. Ther.* 38, 2973–2989. doi:10.1007/s12325-021-01729-3

Deng, P., Tan, M., Zhou, W., Chen, C., Xi, Y., Gao, P., et al. (2021). Bisphenol A Promotes Breast Cancer Cell Proliferation by Driving miR-381-3p-PTTG1-dependent Cell Cycle Progression. *Chemosphere* 268, 129221. doi:10.1016/j.chemosphere.2020.129221

Duffy, M. J., Synnott, N. C., and Crown, J. (2017). Mutant P53 as a Target for Cancer Treatment. *Eur. J. Cancer* 83, 258–265. doi:10.1016/j.ejca.2017.06.023

Fang, J. Y., and Richardson, B. C. (2005). The MAPK Signalling Pathways and Colorectal Cancer. *Lancet Oncol.* 6, 322–327. doi:10.1016/S1470-2045(05)70168-6

Feng, W., Xiaoyan, X., Shenglei, L., Hongtao, L., and Guozhong, J. (2017). PTTG1 Cooperated with GLI1 Leads to Epithelial-Mesenchymal Transition in Esophageal Squamous Cell Cancer. *Oncotarget* 8, 92388–92400. doi:10.18632/oncotarget.21343

Fraune, C., Yehorov, S., Luebke, A. M., Steurer, S., Hube-Magg, C., Büscheck, F., et al. (2020). Upregulation of PTTG1 Is Associated with Poor Prognosis in Prostate Cancer. *Pathol. Int.* 70, 441–451. doi:10.1111/pin.12938

Goodman, A. M., Sokol, E. S., Frampton, G. M., Lippman, S. M., and Kurzrock, R. (2019). Microsatellite-stable Tumors with High Mutational burden Benefit from Immunotherapy. *Cancer Immunol. Res.* 7, 1570–1573. doi:10.1158/2326-6066.CIR-19-0149

Guerrero, C. L. H., Yamashita, Y., Miyara, M., Imaizumi, N., Kato, M., Sakihama, S., et al. (2020). Proteomic Profiling of HTLV-1 Carriers and ATL Patients Reveals sTNFR2 as a Novel Diagnostic Biomarker for Acute ATL. *Blood Adv.* 4, 1062–1071. doi:10.1182/bloodadvances.2019001429

Hatcher, R. J., Dong, J., Liu, S., Bian, G., Contreras, A., Wang, T., et al. (2014). Pttg1/securin Is Required for the Branching Morphogenesis of the Mammary Gland and Suppresses Mammary Tumorigenesis. *Proc. Natl. Acad. Sci.* 111, 1008–1013. doi:10.1073/pnas.1318124111

Hu, Z. G., Zheng, C. W., Su, H. Z., Zeng, Y. L., Lin, C. J., Guo, Z. Y., et al. (2019). MicroRNA-329-mediated PTTG1 Downregulation Inactivates the MAPK Signaling Pathway to Suppress Cell Proliferation and Tumor Growth in Cholangiocarcinoma. *J. Cel. Biochem.* 120, 9964–9978. doi:10.1002/jcb.28279

Huang, J.-l., Cao, S.-w., Ou, Q.-s., Yang, B., Zheng, S.-h., Tang, J., et al. (2018). The Long Non-coding RNA PTTG3P Promotes Cell Growth and Metastasis via Up-Regulating PTTG1 and Activating PI3K/AKT Signaling in Hepatocellular Carcinoma. *Mol. Cancer* 17, 93. doi:10.1186/s12943-018-0841-x

Jang, B.-S., Han, W., and Kim, I. A. (2020). Tumor Mutation burden, Immune Checkpoint Crosstalk and Radiosensitivity in Single-Cell RNA Sequencing Data of Breast Cancer. *Radiother. Oncol.* 142, 202–209. doi:10.1016/j.radonc.2019.11.003

Jiang, T., and Si, L. (2019). Identification of the Molecular Mechanisms Associated with Acute Type a Aortic Dissection through Bioinformatics Methods. *Braz. J. Med. Biol. Res.* 52, e8950. doi:10.1590/1414-431x20198950

Kaur, R. P., Vasudeva, K., Kumar, R., and Munshi, A. (2018). Role of P53 Gene in Breast Cancer: Focus on Mutation Spectrum and Therapeutic Strategies. *Cpd* 24, 3566–3575. doi:10.2174/1381612824666180926095709

Kubo, N., Araki, K., Kuwano, H., and Shirabe, K. (2016). Cancer-associated Fibroblasts in Hepatocellular Carcinoma. *Wjg* 22, 6841–6850. doi:10.3748/wjg.v22.i30.6841

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular Mechanisms of Epithelial-Mesenchymal Transition. *Nat. Rev. Mol. Cel Biol.* 15, 178–196. doi:10.1038/nrm3758

Lee, S., Rauch, J., and Kolch, W. (2020). Targeting MAPK Signaling in Cancer: Mechanisms of Drug Resistance and Sensitivity. *Ijms* 21, 1102. doi:10.3390/ijms21031102

Levy, D., Ferreira, M. C. M. R., Reichert, C. O., de Almeida, L. V., Brocardo, G., Lage, L. A. P. C., et al. (2020). Cell Cycle Changes, DNA Ploidy, and PTTG1 Gene Expression in HTLV-1 Patients. *Front. Microbiol.* 11, 11. doi:10.3389/fmicb.2020.01778

Liang, M., Liu, J., Ji, H., Chen, M., Zhao, Y., Li, S., et al. (2015). A Aconitum Coreanum Polysaccharide Fraction Induces Apoptosis of Hepatocellular Carcinoma (HCC) Cells via Pituitary Tumor Transforming Gene 1 (PTTG1)-Mediated Suppression of the P13K/Akt and Activation of P38 MAPK Signaling Pathway and Displays Antitumor Activity *In Vivo*. *Tumor Biol.* 36, 7085–7091. doi:10.1007/s13277-015-3420-4

Lin, X., Yang, Y., Guo, Y., Liu, H., Jiang, J., Zheng, F., et al. (2019). PTTG1 Is Involved in TNF-α-related Hepatocellular Carcinoma via the Induction of C-myc. *Cancer Med.* 8, 5702–5715. doi:10.1002/cam4.2473

Lin, Y.-H., Tian, Y., Wang, J.-S., Jiang, Y.-G., Luo, Y., and Chen, Y.-T. (2016). Pituitary Tumor-Transforming Gene 1 Regulates Invasion of Prostate Cancer

Cells through MMP13. *Tumor Biol.* 37, 15495–15500. doi:10.1007/s13277-015-3796-1

Lin, Y., Liang, R., Ye, J., Li, Q., Liu, Z., Gao, X., et al. (2019). A Twenty Gene-Based Gene Set Variation Score Reflects the Pathological Progression from Cirrhosis to Hepatocellular Carcinoma. *Aging* 11, 11157–11169. doi:10.18632/aging.102518

Luchini, C., Bibeau, F., Ligtenberg, M. J. L., Singh, N., Nottegar, A., Bosse, T., et al. (2019). ESMO Recommendations on Microsatellite Instability Testing for Immunotherapy in Cancer, and its Relationship with PD-1/pd-L1 Expression and Tumour Mutational burden: A Systematic Review-Based Approach. *Ann. Oncol.* 30, 1232–1243. doi:10.1093/annonc/mdz116

Luo, Y., Shen, D., Chen, L., Wang, G., Liu, X., Qian, K., et al. (2019). Identification of 9 Key Genes and Small Molecule Drugs in clear Cell Renal Cell Carcinoma. *Aging* 11, 6029–6052. doi:10.18632/aging.102161

Meng, C., Zou, Y., Hong, W., Bao, C., and Jia, X. (2020). Estrogen-regulated PTTG1 Promotes Breast Cancer Progression by Regulating Cyclin Kinase Expression. *Mol. Med.* 26, 33. doi:10.1186/s10020-020-00161-7

Mezquita, L., Auclin, E., Ferrara, R., Charrier, M., Remon, J., Planchard, D., et al. (2018). Association of the Lung Immune Prognostic index with Immune Checkpoint Inhibitor Outcomes in Patients with Advanced Non-small Cell Lung Cancer. *JAMA Oncol.* 4, 351–357. doi:10.1001/jamaoncol.2017.4771

Nakachi, I., Helfrich, B., Spillman, M., Mickler, E., Olson, C., Rice, J., et al. (2016). PTTG1 Levels Are Predictive of Saracatinib Sensitivity in Ovarian Cancer Cell Lines. *Clin. Translational Sci.* 9, 293–301. doi:10.1111/cts.12413

Noh, T.-W., Jeong, H. J., Lee, M.-K., Kim, T. S., Kim, S. H., and Lee, E. J. (2009). Predicting Recurrence of Nonfunctioning Pituitary Adenomas. *J. Clin. Endocrinol. Metab.* 94, 4406–4413. doi:10.1210/jc.2009-0471

Noll, J. E., Vandyke, K., Hewett, D. R., Mrozik, K. M., Bala, R. J., Williams, S. A., et al. (2015). PTTG1 Expression Is Associated with Hyperproliferative Disease and Poor Prognosis in Multiple Myeloma. *J. Hematol. Oncol.* 8, 106. doi:10.1186/s13045-015-0209-2

Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003). A Molecular Signature of Metastasis in Primary Solid Tumors. *Nat. Genet.* 33, 49–54. doi:10.1038/ng1060

Repo, H., Gurvits, N., Löyttyniemi, E., Nykänen, M., Lintunen, M., Karra, H., et al. (2017). PTTG1-interacting Protein (PTTG1IP/PBF) Predicts Breast Cancer Survival. *BMC Cancer* 17, 705. doi:10.1186/s12885-017-3694-6

Romero Arenas, M. A., Whitsett, T. G., Aronova, A., Henderson, S. A., LoBello, J., Habra, M. A., et al. (2018). Protein Expression of PTTG1 as a Diagnostic Biomarker in Adrenocortical Carcinoma. *Ann. Surg. Oncol.* 25, 801–807. doi:10.1245/s10434-017-6297-1

Rotte, A. (2019). Combination of CTLA-4 and PD-1 Blockers for Treatment of Cancer. *J. Exp. Clin. Cancer Res.* 38, 255. doi:10.1186/s13046-019-1259-z

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., et al. (2020). A Framework for Advancing Our Understanding of Cancer-Associated Fibroblasts. *Nat. Rev. Cancer* 20, 174–186. doi:10.1038/s41568-019-0238-1

Schrock, A. B., Ouyang, C., Sandhu, J., Sokol, E., Jin, D., Ross, J. S., et al. (2019). Tumor Mutational burden Is Predictive of Response to Immune Checkpoint Inhibitors in MSI-High Metastatic Colorectal Cancer. *Ann. Oncol.* 30, 1096–1103. doi:10.1093/annonc/mdz134

Shukla, S. A., Bachireddy, P., Schilling, B., Galonska, C., Zhan, Q., Bango, C., et al. (2018). Cancer-Germline Antigen Expression Discriminates Clinical Outcome to CTLA-4 Blockade. *Cell* 173, 624–633. doi:10.1016/j.cell.2018.03.026

Su, J., Song, Q., Qasem, S., O'Neill, S., Lee, J., Furdui, C. M., et al. (2021). Multi-Omics Analysis of Brain Metastasis Outcomes Following Craniotomy. *Front. Oncol.* 10, 615472. doi:10.3389/fonc.2020.615472

Suarez-Carmona, M., Lesage, J., Cataldo, D., and Gilles, C. (2017). EMT and Inflammation: Inseparable Actors of Cancer Progression. *Mol. Oncol.* 11, 805–823. doi:10.1002/1878-0261.12095

Sun, M., Zhao, W., Chen, Z., Li, M., Li, S., Wu, B., et al. (2019). Circular RNA CEP128 Promotes Bladder Cancer Progression by Regulating Mir-145-5p/Myd88 via MAPK Signaling Pathway. *Int. J. Cancer* 145, 2170–2181. doi:10.1002/ijc.32311

Synnott, N. C., Murray, A., McGowan, P. M., Kiely, M., Kiely, P. A., O'Donovan, N., et al. (2017). Mutant P53: a Novel Target for the Treatment of Patients with Triple-Negative Breast Cancer? *Int. J. Cancer* 140, 234–246. doi:10.1002/ijc.30425

Topalian, S. L., Taube, J. M., Anders, R. A., and Pardoll, D. M. (2016). Mechanism-driven Biomarkers to Guide Immune Checkpoint Blockade in Cancer Therapy. *Nat. Rev. Cancer* 16, 275–287. doi:10.1038/nrc.2016.36

Vlotides, G., Eigler, T., and Melmed, S. (2007). Pituitary Tumor-Transforming Gene: Physiology and Implications for Tumorigenesis. *Endocr. Rev.* 28, 165–186. doi:10.1210/er.2006-0042

Wondergem, B., Zhang, Z., Huang, D., Ong, C. K., Koeman, J., Hof, D. V. t., et al. (2012). Expression of the PTTG1 Oncogene Is Associated with Aggressive clear Cell Renal Cell Carcinoma. *Cancer Res.* 72, 4361–4371. doi:10.1158/0008-5472.CAN-11-2330

Wu, C., Duan, Y., Gong, S., Kallendrusch, S., Schopow, N., and Osterhoff, G. (2021). Integrative and Comprehensive Pancancer Analysis of Regulator of Chromatin Condensation 1 (RCC1). *Ijms* 22, 7374. doi:10.3390/ijms22147374

Xiang, W., Wu, X., Huang, C., Wang, M., Zhao, X., Luo, G., et al. (2017). PTTG1 Regulated by miR-146a-3p Promotes Bladder Cancer Migration, Invasion, Metastasis and Growth. *Oncotarget* 8, 664–678. doi:10.18632/oncotarget.13507

Yan, H., Wang, W., Dou, C., Tian, F., and Qi, S. (2015). Securin Promotes Migration and Invasion via Matrix Metalloproteinases in Glioma Cells. *Oncol. Lett.* 9, 2895–2901. doi:10.3892/ol.2015.3074

Yoon, C.-H., Kim, M.-J., Lee, H., Kim, R.-K., Lim, E.-J., Yoo, K.-C., et al. (2012). PTTG1 Oncogene Promotes Tumor Malignancy via Epithelial to Mesenchymal Transition and Expansion of Cancer Stem Cell Population. *J. Biol. Chem.* 287, 19516–19527. doi:10.1074/jbc.M111.337428

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, E., Liu, S., Xu, Z., Huang, S., Tan, X., Sun, C., et al. (2014). Pituitary Tumor-Transforming Gene 1 (PTTG1) Is Overexpressed in Oral Squamous Cell Carcinoma (OSCC) and Promotes Migration, Invasion and Epithelial-Mesenchymal Transition (EMT) in SCC15 Cells. *Tumor Biol.* 35, 8801–8811. doi:10.1007/s13277-014-2143-2

Zhang, J., Gu, C., Song, Q., Zhu, M., Xu, Y., Xiao, M., et al. (2020). Identifying Cancer-Associated Fibroblasts as Emerging Targets for Hepatocellular Carcinoma. *Cell Biosci* 10, 127. doi:10.1186/s13578-020-00488-y

Zhang, J., Song, Q., Wu, M., and Zheng, W. (2020). The Emerging Roles of Exosomes in the Chemoresistance of Hepatocellular Carcinoma. *Cmc* 28, 93–109. doi:10.2174/0929867327666200130103206

Zhi, T., Jiang, K., Xu, X., Yu, T., Zhou, F., Wang, Y., et al. (2019). ECT2/PSMD14/PTTG1 axis Promotes the Proliferation of Glioma through Stabilizing E2F1. *Neuro. Oncol.* 21, 462–473. doi:10.1093/neuonc/noy207

Zhou, M., Chen, J., Zhang, H., Liu, H., Yao, H., Wang, X., et al. (2020). KLF10 Inhibits Cell Growth by Regulating PTTG1 in Multiple Myeloma under the Regulation of microRNA-106b-5p. *Int. J. Biol. Sci.* 16, 2063–2071. doi:10.7150/ijbs.45999

Zhu, M., Wu, M., Bian, S., Song, Q., Xiao, M., Huang, H., et al. (2021). DNA Primase Subunit 1 Deteriorated Progression of Hepatocellular Carcinoma by Activating AKT/mTOR Signaling and UBE2C-Mediated P53 Ubiquitination. *Cel Biosci* 11, 42. doi:10.1186/s13578-021-00555-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gong, Wu, Duan, Tang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fgene.2022.808100

KDF1 Promoted Proliferation and Metastasis of Epithelial Ovarian Cancer *via* Wnt/Beta-Catenin Pathway: TCGA-Based Data Mining and Experimental Validation

Changyu Zhu[1,2], *Yilong Liu*[2], *Rongsheng Tong*[1,2]* *and Jianmei Guan*[3]

[1]*Department of Pharmacy, Sichuan Academy of Medical Science and Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China,* [2]*Personalized Drug Therapy Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu, China,* [3]*Central Sterile Supply Department, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, China*

Objectives: It has been reported that keratinocyte differentiation factor 1 (KDF1) was related to proliferation, differentiation, and cell cycle. However, the role of KDF1 has not been reported in ovarian cancer. The present study investigated the function and the potential mechanism of KDF1 in ovarian cancer.

Methods: We evaluated the prognostic value in ovarian cancer based on data from the Cancer Genome Atlas (TCGA) database. The Kruskal–Wallis test, Wilcoxon signed-rank test, and logistic regression were used to evaluate the relationship between KDF1 expression and clinicopathologic features. The Cox regression and the Kaplan–Meier method were adopted to evaluate prognosis-related factors. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) gene enrichment analysis, and Gene Set Enrichment Analysis (GSEA) were performed to identify the key biological process related to KDF1. Then the expression of KDF1 in ovarian cancer tissues was validated by streptavidin–peroxidase (SP) immunohistochemistry. The proliferation and invasion ability of KDF1 were determined by EdU and Transwell assay, respectively, with KDF1 gene silencing and overexpression. The mRNA expression of KDF1 was determined by qPCR. The protein expression of KDF1 was determined using the Western blot.

Methods: By performing differential expression analysis on the ovarian cancer data of the TCGA database, it was found that KDF1 is highly expressed in ovarian cancer patients and associated with poorer overall survival (OS) and progression-free survival (PFS) of ovarian cancer patients. The highly expressed KDF1 may reduce cell adhesion according to GO, KEGG, and GSEA results. After analysis combining the relevant clinical features, we found that the high expression of KDF1 is an independent prognostic factor of ovarian cancer and associated with platinum resistance and tumor metastasis in ovarian cancer. At the same time, the BioGRID database showed that there might be protein–protein interaction between KDF1 and E-cadherin. Then we further validated that the high expression of

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
Zunyi Medical University, China

Reviewed by:
Georges Daoud,
American University of Beirut, Lebanon
Jozsef Dudas,
Innsbruck Medical University, Austria

***Correspondence:**
Rongsheng Tong
tongrs@126.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *03 November 2021*
Accepted: *04 February 2022*
Published: *25 February 2022*

Citation:
Zhu C, Liu Y, Tong R and Guan J (2022) KDF1 Promoted Proliferation and Metastasis of Epithelial Ovarian Cancer via Wnt/Beta-Catenin Pathway: TCGA-Based Data Mining and Experimental Validation. Front. Genet. 13:808100. doi: 10.3389/fgene.2022.808100

KDF1 had a close correlation with the stage and grade of ovarian cancer in ovarian cancer tissue chips. Silencing KDF1 inhibited the proliferation and invasion ability of SKOV3 cells. By contrast, ectopic expression of KDF1 promoted the proliferation and invasion ability of A2780 cells. We also found that KDF1 can interact with E-cadherin and regulate the expression of Wnt5A and β-catenin, hence activating Wnt/β-catenin pathway via *in vitro* and vivo experiments.

Conclusions: Based on the bioinformatics analysis, *in vitro* experiments, and an *in vivo* study, it is indicated that KDF1 played an important role in ovarian cancer progression and might be a therapeutic target for patients with ovarian cancer.

Keywords: TCGA, ovarian cancer, proliferation, KDF1, Wnt

INTRODUCTION

Ovarian cancer (OV) is the only gynecological tumor among the five leading causes of death in women. According to the statistics report, there were 21,750 new cases and 13,940 deaths in the United States in 2020 (Siegel et al., 2020). According to the tumor-initiating cell type, ovarian cancer was divided into three categories: epithelial cancer, interstitial cancer, and germ cell cancer (Nguyen et al., 2019). Epithelial ovarian cancer (EOC) was the most common ovarian malignancy, accounting for more than 90% of all ovarian cancer. Most patients with ovarian cancer are asymptomatic, and the volume of the ovary is relatively small, the anatomical position of which is deep in the pelvic cavity. Patients often cannot observe specific symptoms early, resulting in the difficulty of early screening of the disease. Therefore, more than 70% newly diagnosed patients were at the advanced stage, resulting in a 5-year survival rate of less than 20%.

Based on the epidemiological characteristics of ovarian cancer, early diagnosis and treatment are very important. However, compared to the increasingly advanced detection technology of other tumors, there has been no reliable screening biomarker and therapeutic target for ovarian cancer so far (Menon et al., 2018).

Keratinocyte differentiation factor 1 (*KDF1*), a protein-coding gene containing a domain of unknown function (DUF4656), was first identified as an essential regulator of the proliferation differentiation decision in epidermal progenitor cells. *Kdf1* was found in mice for the first time by the positive genetic method with the function of retarding proliferation through its inhibition of p63.

Shamseldin et al. (2017)had reported that *KDF1* is silenced in a multigenerational family with ectodermal dysplasia. When the *Kdf1* gene was knocked out, the mouse represented the recapitulation of the phenotype (Shamseldin et al., 2017). Recently, a study reported that *KDF1* was a new candidate gene for non-syndromic tooth agenesis (Zeng et al., 2019). Also, *Kdf1* induced epidermal progenitor cell differentiation through interaction with the cell cycle regulator stratifin (Lee et al., 2013). Many recent studies have shown that stratifin is a critical marker in the process of tumor progression (Shiba-Ishii et al., 2019). The genetic interaction between Kdf1 and more widely studied stratifin (Shamseldin et al., 2017) proves that KDF1 may be related to tumor progression.

At the same time, quantitative proteomics was used to study the interaction group of KDF1. Mass spectrometry had identified that KDF1 could specifically bind to IκB kinase α (IKKα) in differentiated keratinocytes and mediate the regulation of deubiquitination on epidermal differentiation (Li et al., 2020). As a mature component of the NF-κB signaling pathway, IKKα also plays a vital role in cancer progression (Liu et al., 2008; Yang et al., 2016).

Although studies have shown that KDF1 is associated with tumor-related molecules, so far, the expression and function of KDF1 in epithelial ovarian cancer have not been reported. Therefore, in this research, we combined bioinformatics and experimental study to investigate the expression pattern and function of KDF1 in epithelial ovarian cancer and explore the value of KDF1 in EOC diagnosis and treatment.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

The HTSeq-FPKM dataset was downloaded from the GDC TCGA Ovarian Cancer queue of UCSC XENA (https://xenabrowser.net/datapages/), which includes 379 ovarian cancer RNA seq-data. Then the clinical data were downloaded from the TCGA database (Blum et al., 2018) (https://portal.gdc.cancer.gov/) until 12 July 2020. Then the TRAIL TOIL RSEM fpkm (n = 7,862) UCSC Toil RNA-seq Recompute dataset was downloaded from the GTEx queue of UCSC XENA. The sequencing data of 88 normal ovarian tissues and 379 ovarian cancer tissues were extracted and analyzed. The different KDF1 expression between OC and non-tumor tissue was also investigated in three RNAseq datasets (GSE12470, GSE18520, and GSE66957), which were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo).

Differentially Expressed Gene Analysis

We used unpaired Student's *t*-test within the DESeq2 R package (4.0.0) (Love et al., 2014) to compare the expression data (HTseq-Counts) between high and low expression groups according to the median *kdf1* expression

level. The thresholds for the DEGs were |log2-fold change (FC)| >2.0 and adjusted $p < 0.05$.

Enrichment Analysis and Protein–Protein Interaction Analysis

ClusterProfiler package in R (4.0.0) (Yu et al., 2012) was used to perform Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis to detect the possible function of KDF1. Samples were divided into high and low expression groups according to the expression level of KDF1 expression. As a computational method, the GSEA determines whether *a priori* defined set of genes has statistical significance and concordant differences in two biological states. We performed GSEA between high- and low-KDF1 groups by GSEA Desktop Application (v4.0.3; Broad Institute, Inc., Cambridge, MA, United States). Additionally, the adjusted P and normalized enrichment score (NES) were utilized to sort the enriched pathways in each phenotype (Subramanian et al., 2005). c2. cp.v7.0. symbols.gmt (Curated) in MSigDB Collections was selected as a reference gene set. Gene sets with a false discovery rate (FDR) < 0.25 and adjusted $p < 0.05$ were considered significantly enriched. The KDF1 interacting protein was predicted by Biogrid and visualized by Cytoscape.

Immunochemistry

This study was approved by the Ethics Committee of Sichuan Provincial People's Hospital. The approval number issued by the Ethics Committee was as follows: Ethic review (fundamental research) No. 109 of 2016. The location and expression of KDF1 in ovarian cancer and normal ovarian tissues were detected by immunohistochemistry. The ovarian cancer tissue chip was purchased from Alenabio with a total of 110 cases. The ovarian cancer tissue chip contains 80 cases of EOC tissue, and the normal ovarian tissue chip includes 30 samples of normal ovarian tissue. The immunohistochemistry kit and DAB kit were purchased from Zsbio. The procedure of the immunohistochemistry experiment refers to that of previous research. The immunohistochemical score (ranging from 0 to 9) was calculated by multiplying the intensity and percentage scores. Staining intensity was graded on a 0–3 scale: 0, absence of staining; 1, weakly stained; 2, moderately stained; and 3, intensely stained. The percentage of positive tumor cells was scored as follows: 0, absence of tumor cells; 1, <33% of tumor cells; 2, 33–66% of tumor cells; and 3, >66% of tumor cells (Mustapar et al., 2020).

Cell Culture, Transfection Procedure, and Reagents

A2780 and SKOV3 cells were cultured in DMEM (Sigma, D5796). The medium included streptomycin and 10% fetal bovine serum. The cells were incubated under 5% CO_2 and 37°C. KDF1 interference vector (called LV3-NC, LV3-shKDF1-1, and LV3-shKDF1-2) and overexpression vector (called LV5-NC and LV5-KDF1) carried by the lentivirus were from Genepharma. The following was the siRNA sequence targeting KDF1. LV3-shKDF1-1: 5′-GUUUGUAAGUAC AAAGGUAA-3'; LV3-shKDF1-2: 5′-GCUGAUGUUCUGUAU CUUAAC-3′ and NC (negative control) siRNA: 5′-UUCUUC GAAGGUGUCACGUTT-3′.

Scratch Assay

The scratch assay was performed as in a previous study (Liang et al., 2007). Cells were cultured in a six-well plate to a confluent monolayer. We used a 10 μL pipette tip to scrape the cell monolayer in a straight line vertically. The debris was then removed by washing the cells thrice with PBS (Boster, Wuhan) and replaced with 2 ml of the original medium. Then the images of cells were captured 48 h after scratch.

EdU Assay

The cell proliferation ability was measured using the EdU experiment. The EdU kit was purchased from GeneCopoeia. The experimental procedures refer to those of the previous literature (Fu et al., 2020).

RT-qPCR

The qPCR steps mainly include RNA extraction and reverse transcription into cDNA and qPCR. The experimental procedures can be referred to those in the previous literature (Han et al., 2017). The primers were synthesized by Genepharma.

Western Blot

The expression of KDF1, GAPDH, E-cadherin, and β-catenin was detected by the Western blot. The primary antibody used in the present research included anti-KDF1 (Abcam, ab224760), anti-GAPDH (Abcam, ab181602), anti-E-cadherin (Abcam, ab40772), anti-Wnt5A (Abcam, ab179824), and anti-β-catenin (Abcam, ab223075). Primary antibodies were diluted in Dilution Buffer (Beyotime, P0256) and incubated overnight at 4°C. We used the gel imaging system to analyze the band density and compare it with the internal control.

Matrigel Invasion Assay

The Matrigel invasion assay was performed to assess cellular invasion ability according to the previous study (Han et al., 2017). 1×10^5 cells were seeded into the upper chamber. After 24 h, the cells on the lower surface of the membrane were fixed with 4% paraformaldehyde and stained with 0.5% crystal solution. Then, the cells were counted and photographed using a microscope.

Co-Immunoprecipitation

Co-immunoprecipitation (Co-IP) was performed as a previous study (Adhikary et al., 2016). Briefly, the cells were soaked in lysis buffer, and specific antibodies were adopted to perform immunoprecipitation. For DNase I co-immunoprecipitation, 500 μg of lysate was digested in DNase I for 1 h at 37°C. The reaction was broken by adding 5 mM EDTA. We used the DNA-free lysate for immunoprecipitation with specific antibodies. After all the reaction ended, we adopted immunoblotting to analyze immunoprecipitants.

Luciferase Reporter Assays

The Dual-Luciferase® Reporter Assay System (Promega) was used to perform luciferase reporter assays. The cells from different

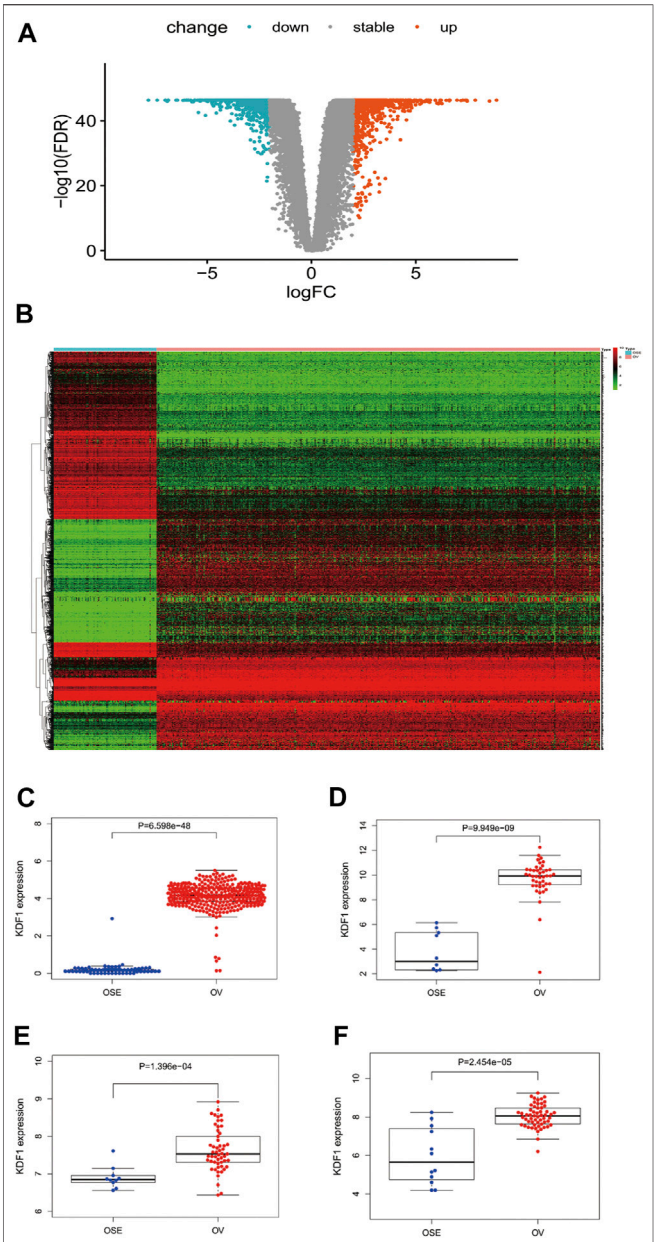


FIGURE 1 | Result of differentially expressed gene analysis. The volcano plot of differentially expressed RNAs in ovarian cancer **(A)**. Heatmap of all differentially expressed genes in ovarian cancer **(B)**. Relationship of expression levels of KDF1 in ovarian cancer tissues and normal ovarian tissues in TCGA **(C)**. Relationship of expression levels of KDF1 in ovarian cancer tissues and normal ovarian tissues was validated by datasets of GSE12470 **(D)**, GSE18520 **(E)**, and GSE66957 **(F)** from the GEO database.

groups were seeded in 24-well plates (2.0×10^5cells per well) and transfected together with a promoter–reporter gene vector and the pRL-TK Renilla luciferase vector. After 48 h of transfection, the cells were harvested and analyzed according to the manufacturer's instructions. The luciferase activities were normalized to the Renilla luciferase activity of the internal control.

In Vivo Tumorigenicity

Four- to six-week-old female BALB/c mice were provided by the Laboratory Animal Centre of Chongqing Medical University (Chongqing, China) and maintained at Sichuan Provincial People's Hospital. The protocols were performed after approval by the Animal Ethics Committee according to issued guidelines. 4×10^7 cells mixed with an equal volume of PBS were injected subcutaneously into the region of the right axilla. Tumor sizes were monitored every 3 days using a vernier caliper, and tumor volumes were calculated using the formula [1/2 × long diameter (cm) × short diameter $(cm)^2$] and expressed in cm^3.

Statistical Analysis

Statistical data acquired from TCGA were merged and processed by R 4.0.0. The Wilcoxon rank-sum test and Wilcoxon signed-rank test were used for comparing the expression levels of KDF1 between OC and the control group. The Kruskal–Wallis test, Wilcoxon rank-sum test, Wilcoxon signed-rank test, and Spearman correlation were used to analyze the relation between KDF1 expression and the grade of clinicopathologic factors. Normal and adjusted Pearson κ^2 test and the Fisher exact test were used to analyze whether the grade of clinicopathologic factors affects KDF1 expression. Univariate Cox regression analysis and multivariate Cox regression analysis were combined to evaluate the prognostic value of KDF1 expression and other clinicopathologic factors on survival. All assays in our study were performed in triplicate. The data of different groups in each assay were compared by using two-sided Student's t-test or analysis of variance (ANOVA). A two-sided $p < 0.05$ was considered significant.

RESULTS

Identification of Differentially Expressed Genes in OC

Based on the cutoff criteria (|logFC| <1.5 and adjusted $p < 0.05$), we used the DESeq2 package in R (Love et al., 2014) to analyze the HTSeq-counts data from TCGA. DEG expressions were illustrated by a heatmap (**Figure 1A**). DEGs included 2008 differentially expressed RNAs (1,017 upregulated and 991 downregulated) (**Figure 1B**). Differential expression analysis between normal and OC groups indicated KDF1 was expressed significantly higher in OV than normal ovarian tissue (**Figure 1C**). KDF1 mRNA expression also exhibited significantly increased in GSE12470 (**Figure 1D**), GSE18520 (**Figure 1E**), and GSE66957 (**Figure 1F**).

Functional Enrichment Analysis of Differentially Expressed Genes and Protein–Protein Interaction Analysis Results

We performed GO and KEGG enrichment analyses of KDF1-associated DEG functions in OC. The GO results displayed that KDF1-associated DEGs had significant regulation on extracellular matrix organization, extracellular structure organization, collagen

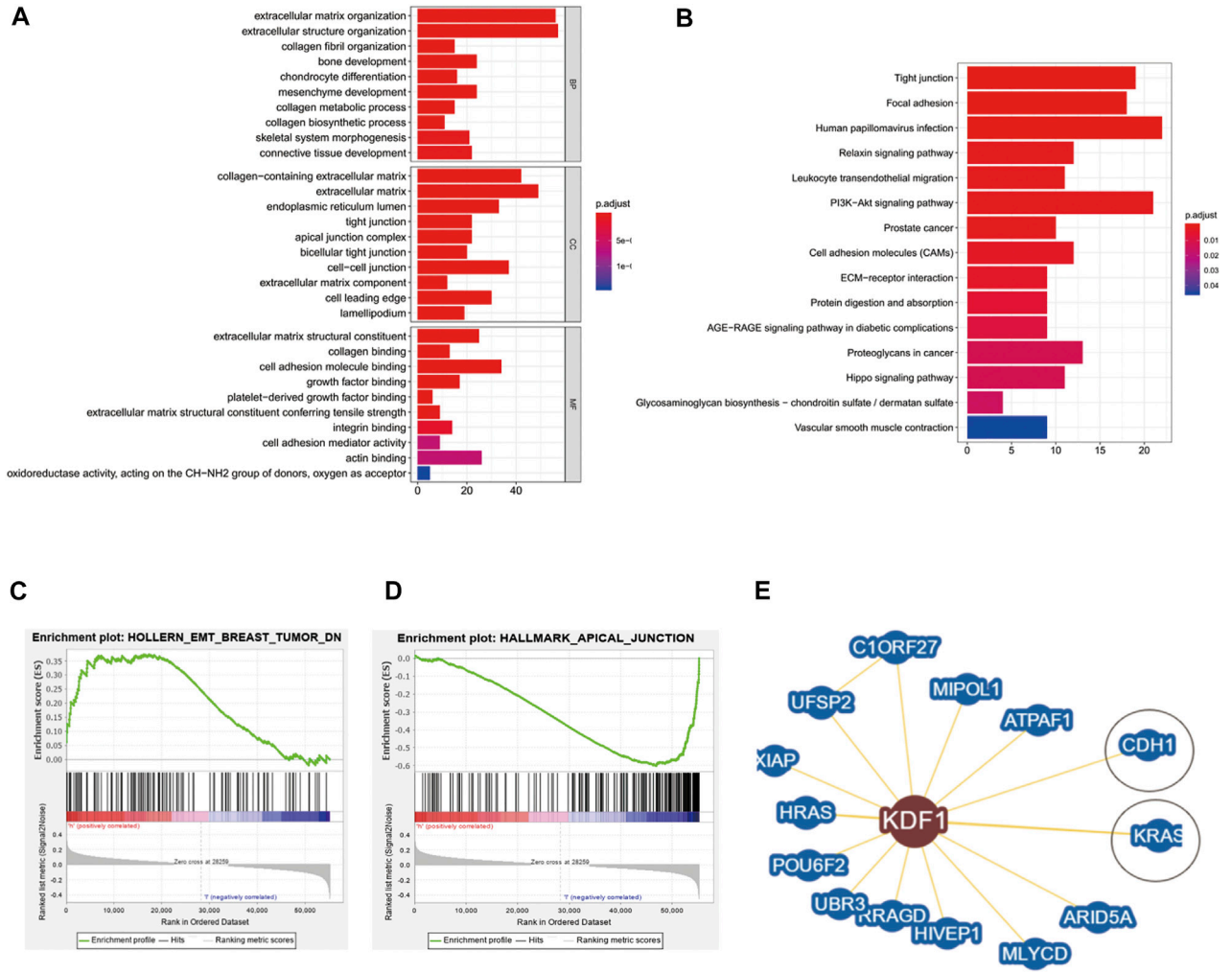


FIGURE 2 | Result of enrichment analysis and protein–protein interaction analysis of KDF1 in ovarian cancer. The result of enrichment analysis from GO **(A)**, KEGG **(B)**, and GSEA **(C)**. The KDF1 interacting protein was predicted by Biogrid **(D)**.

fibril organization, and bone development in the biological process. Moreover, they also related collagen-containing extracellular matrix, extracellular matrix, endoplasmic reticulum lumen, tight junction, apical junction complex, and cell–cell junction of the cellular component. Extracellular matrix structural constituent, collagen binding, cell adhesion molecule binding, growth factor binding, and platelet-derived growth factor binding of molecular functions were also involved in regulating KDF1 interactive genes (**Figures 2A,B**). KDF1-related signaling pathways were identified by GSEA (**Figures 2C,D**). The KDF1 interacting protein predicted by Biogrid included CDH and KRAS (**Figure 2E**).

The Prognostic Value of Keratinocyte Differentiation Factor 1

By analyzing the survival data in the TCGA database, the best cutoff of KDF1 expression was adopted to divide the patients into 331 cases in the high-expression group (1 case missed overall survival data) and 44 cases in the low-expression group. There was a significant difference in the OS–KM curve between high- and low-expression groups ($p = 0.022$) (**Figure 3A**). The same cutoff value was adopted to divide all patients into 332 cases in the high-expression group and 44 cases in the low-expression group. There was a significant difference in the PFS–KM curve between high- and low-expression groups ($p < 0.001$) (**Figure 3B**). KDF1 was also input in the online tool Kaplan–Meier plotter to verify the relationship between the expression of KDF1 and the OS or PFS in OC patients. It was found that the difference of the OS–KM curve and the PFS–KM curve of ovarian cancer patients between the KDF1 high- and low-expression groups was statistically significant ($p = 0.028$) (**Figure 3C**) ($p = 0.0024$) (**Figure 3D**).

When comparing whether there are differences in different clinical features between high- and low-expression groups of KDF1, it is found that KDF1 is significantly correlated with platinum resistance ($p = 0.018$) and new events, including recurrence and progression of disease ($p = 0.037$) (**Table 1**).

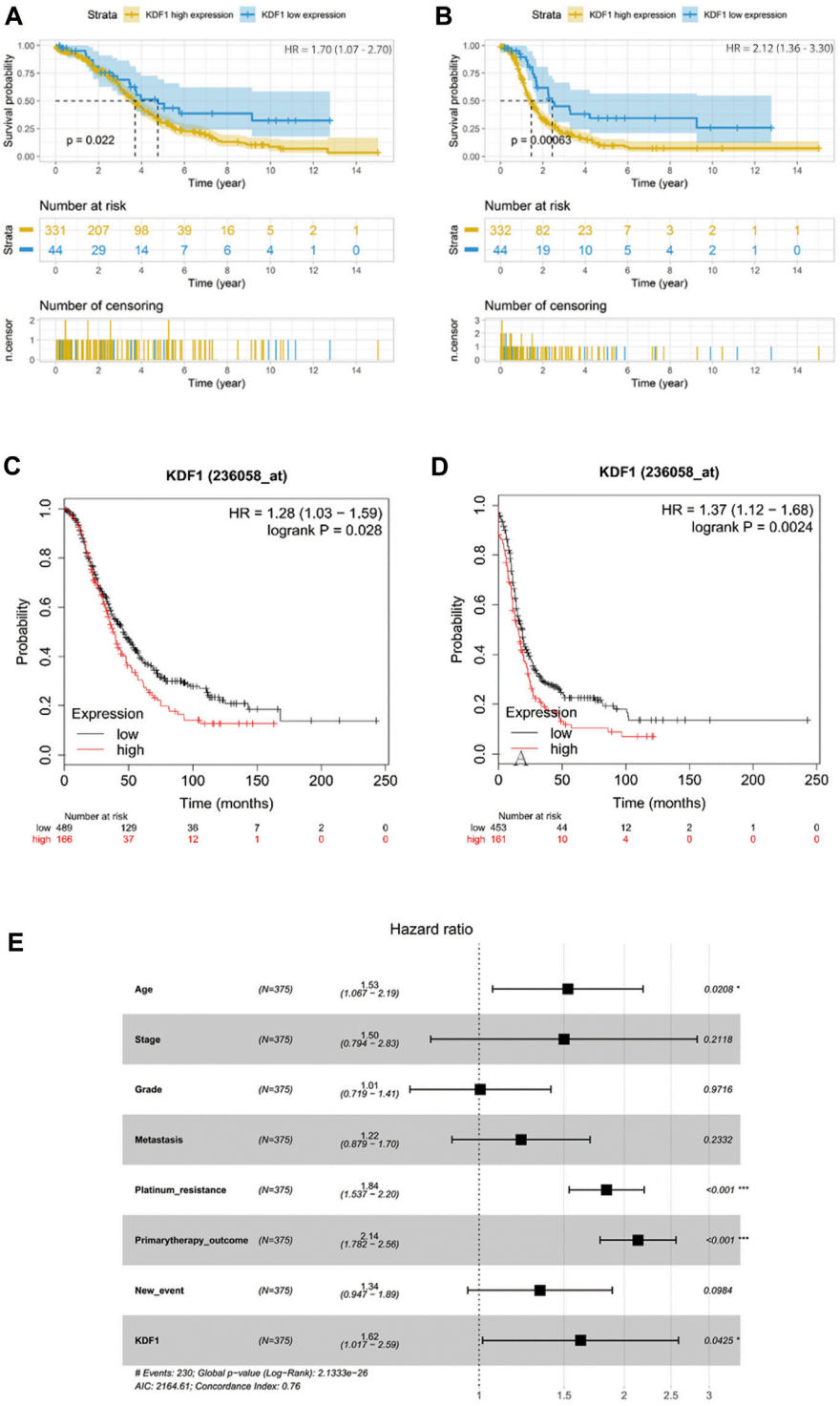


FIGURE 3 | Prognostic value of KDF1 in ovarian cancer. High expression of KDF1 was associated with worse overall survival in TCGA **(A)** and Kaplan–Meier plotter **(C)**. High expression of KDF1 was associated with worse progression-free survival in TCGA **(B)** and Kaplan–Meier plotter **(D)**. Forest plot of the prognostic value of KDF1 in overall survival in TCGA **(E)**.

Univariate Cox regression suggested that stage (p = 0.05), platinum resistance (p < 0.001), and primary treatment outcome (p < 0.001) were related to the prognosis of ovarian cancer, and KDF1 had a trend to reach statistical significance (p = 0.07) (**Table 2**). After removing clinical confounding factors by multivariate Cox regression, the results showed that KDF1 (p = 0.04), platinum

TABLE 1 | Difference of clinical characteristics between high- and low-expression groups of KDF1.

Characteristics		KDF1 expression RNAseq		*p* Value
		Low	High	
Metastasis	No	34	282	0.471
	Yes	8	49	
Stage	Stage I + II	2	21	1
	Stage III + IV	40	310	
Grade	Grade 1 + 2	8	35	0.191
	Grade 3 + 4	34	289	
Platinum_resistance	No	28	182	0.018
	Yes	3	77	
New_event (recurrence or progression of disease)	No	21	109	0.037
	Yes	22	224	
Age	≤50	9	71	0.953
	>50	34	262	
Primarytherapy_outcome	CR	25	186	0.418
	NCR	8	84	

TABLE 2 | Relationship between clinicopathologic parameters and expression of KDF1 in 110 cases of ovarian cancer.

—	No. of patients (*n* = 110)	KDF1 expression Low no. (%)	KDF1 expression High no. (%)	*p* Value —
Characteristics	—	—	—	—
Age(years)	—	—	—	>0.05
<50	53	25	28	—
≥50	57	26	31	—
Normal ovarian	30	28	2	<0.05
Cancer tissues	80	23	57	
FIGO stage	—	—	—	—
I/II	45	19	26	<0.05
III/IV	35	4	31	
Grade	—	—	—	—
1	18	14	4	—
2	25	6	19	—
3	37	3	34	—
Grade 2–3 *versus* 1				<0.05

resistance ($p < 0.001$), and initial treatment outcome ($p < 0.001$) were independent risk factors for the prognosis of ovarian cancer. It is suggested that patients with high KDF1 expression have a higher risk of death (HR = 1.62, 95% CI: 1.017–2.59) than those with low KDF1 expression. The forest plot drawn according to Cox multivariate regression after incorporating relevant clinical features suggests that the high expression of KDF1, platinum resistance, and primary therapy outcome are risk factors for low OS in patients with ovarian cancer (**Figure 3E**).

The Expression of Keratinocyte Differentiation Factor 1 in Ovarian Cancer Tissues and Cell Lines

We first detected the location and expression of KDF1 in ovarian cancer tissues using immunochemistry. KDF1 is mainly located in the cytoplasm. The expression of KDF1 was higher in ovarian cancer tissues than that of normal ovarian tissues (**Figures 4A–D**). The expression of KDF1 in ovarian cancer was relative to grade and stage. The expression of KDF1 was higher in ovarian cancer tissues in advanced stages (stage III/IV) than in those in the early stages (stage I/II; **Table 3**, $p < 0.05$).

Furthermore, the staining intensity correlated with the tumor grade (grades 2–3 *versus* 1, **Table 3**, $p < 0.05$). To investigate the function of KDF1 in ovarian cancer, we screened the mRNA expression of KDF1 in ovarian cancer cell lines. The mRNA expression of SKOV3 and OVCA3WT was higher than that of A2780 (**Figure 4E**). The gene-silencing and gene-overexpressing efficiency were verified by semiquantitative real-time PCR analysis (**Figures 4F,G**). Then the KDF1 protein expression in different groups was examined using the Western blot (**Figures 5A,B**).

Keratinocyte Differentiation Factor 1 Promoted Proliferation, Migration, and Invasion of Ovarian Cancer Cells

We found that the proliferation, migration, and invasion ability of SKOV3 cells were inhibited after the silencing of KDF1. By contrast, the proliferation, migration, and invasion ability of A2780 cells were elevated after ectopic expression of KDF1 (**Figures 5C–E**).

Keratinocyte Differentiation Factor 1 Participated in the Wnt/β-Catenin Pathway in Ovarian Cancer

As predicted, the interaction between KDF1 and E-cadherin was identified by the Co-IP assay. The cells were co-transfected with Flag-E-cadherin, HA-KDF1, and control group was established simultaneously and harvested 24 h later. Anti-HA antibodies pulled the interaction proteins. Then, they were detected by anti-Flag antibodies. The Western blot displayed that Flag bands could not be detected in the cells transfected with Flag-E-cadherin (lane 1) or HA-KDF1 (lane 3) only. However, it can be detected in cells co-transfected with Flag-E-cadherin and HA-KDF1 (lane 2), indicating that the interaction between KDF1 and E-cadherin may exist *in vivo* (**Figure 6A**). So, we guessed that KDF1 may regulate the Wnt/β-catenin pathway. Indeed, TOP-flash luciferase activity indicated the

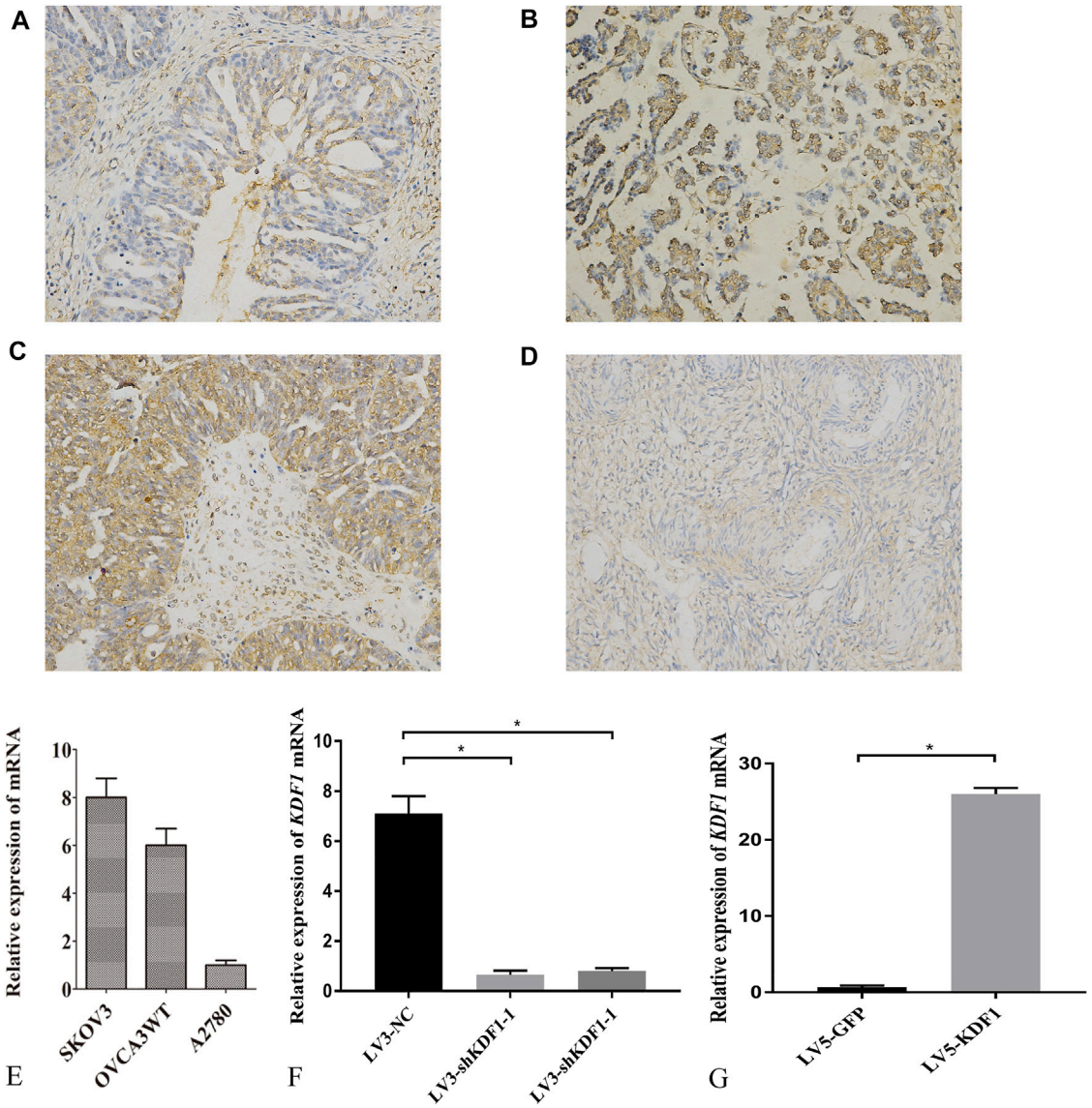


FIGURE 4 | Expression of KDF1 in tissues and cells of ovarian cancer. The expression and location of KDF1 was detected in ovarian cancer tissues **(A–C)** and normal ovarian tissues **(D)**. KDF1 mRNA expression was detected by RT-qPCR **(E–G)**. *indicates $p < 0.05$.

TABLE 3 | Univariate and multivariate analyses of clinicopathologic parameters in patients with ovarian cancer in TCGA-OV.

Characteristics	HR (95% CI) univariate analysis	*p* Value univariate analysis	HR (95% CI) multivariate analysis	*p* Value multivariate analysis
Age	1.329 (0.935–1.891)	0.11	1.528 (1.067–2.188)	0.02
Stage	2.133 (1.002–4.541)	0.05	1.500 (0.794–2.833)	0.21
Grade	1.222 (0.856–1.744)	0.27	1.006 (0.719–1.409)	0.97
Metastasis	1.263 (0.912–1.751)	0.16	1.222 (0.879–1.699)	0.23
Platinum resistance	2.213 (1.866–2.624)	<0.001	1.838 (1.537–2.200)	<0.001
Primary therapy outcome	2.189 (1.867–2.567)	<0.001	2.136 (1.782–2.560)	<0.001
New event	0.963 (0.702–1.321)	0.81	1.338 (0.947–1.888)	0.10
KDF1	1.537 (0.970–2.435)	0.07	1.623 (1.016–2.591)	0.04

HR: hazard ratio, CI: confidence interval.

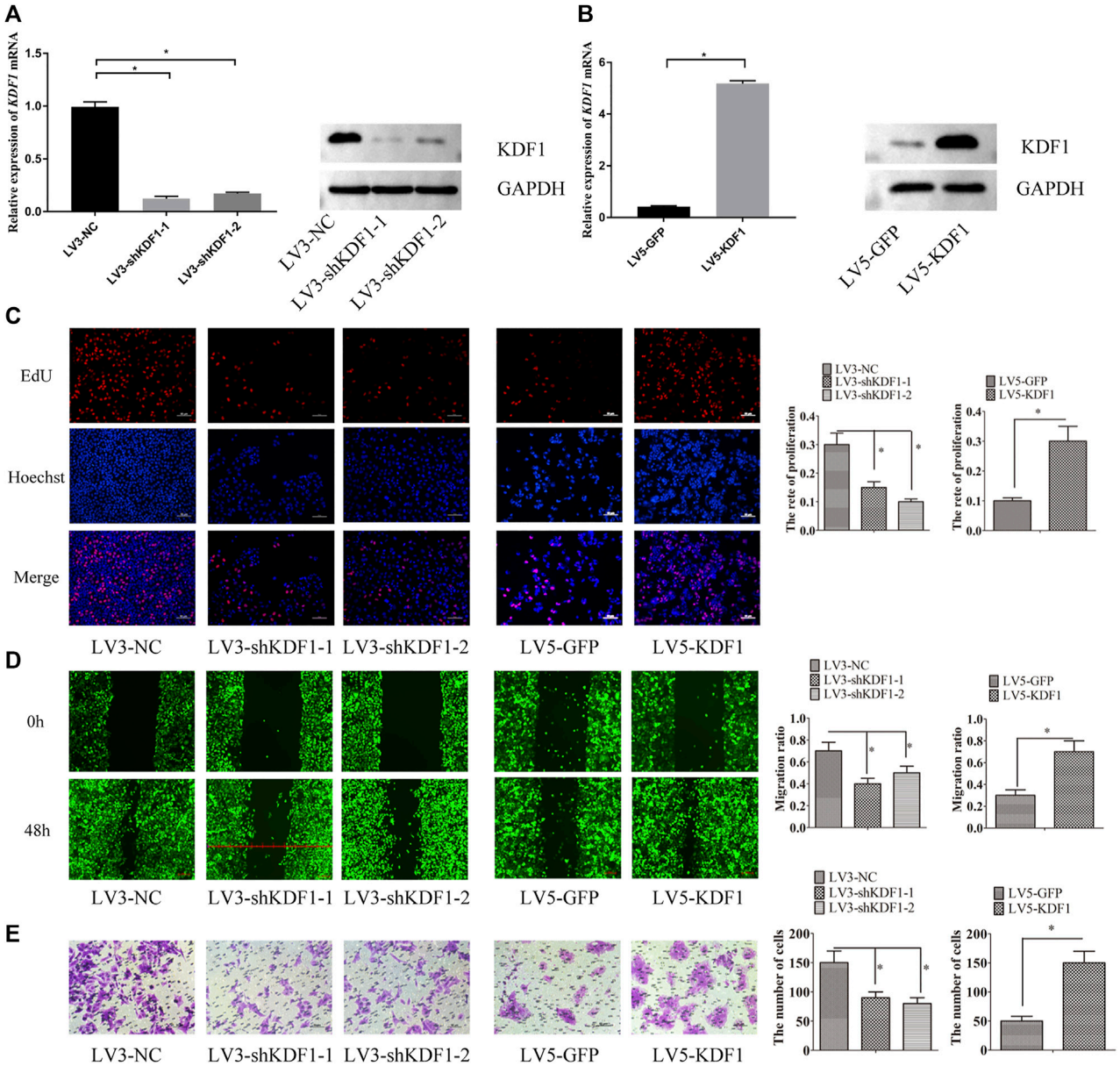


FIGURE 5 | KDF1 promoted proliferation, migration, and invasion of ovarian cancer cells. KDF1 protein expression examined by Western blot **(A,B)**. The cell proliferation ability was detected by EdU assay **(C)**. The cell migration ability was detected by wound scratch assay **(D)**. The cell invasion ability was detected by transwell assay. Error bars represent the standard error **(E)**. *indicates $p < 0.05$.

level of β-catenin and Wnt5A protein was increased in A2780 cells after ectopic expression of KDF1 (**Figures 6B,C**). In contrast, silencing KDF1 decreased TOP-flash luciferase activity, the expression of β-catenin, and Wnt5A protein in SKOV3 cells (**Figures 6B–D**).

Silencing Keratinocyte Differentiation Factor 1 Inhibited SKOV3 Cells Growth *In Vivo*

To investigate the effect of KDF1 on SKOV3 cells' growth *in vivo*, we constructed a nude mouse subcutaneous tumor model. We observed that silencing KDF1 retarded the growth of SKOV3 cells *in vivo* (**Figures 7A–D**). The tumor weight and volume were reduced after silencing KDF1. We also observed that the expression of KDF1, Wnt5a, p-AKT, and β-catenin was decreased after silencing KDF1 (**Figure 7E**).

DISCUSSION

In this study, bioinformatics analysis of sequencing data from TCGA was performed to gain a deeper understanding of the

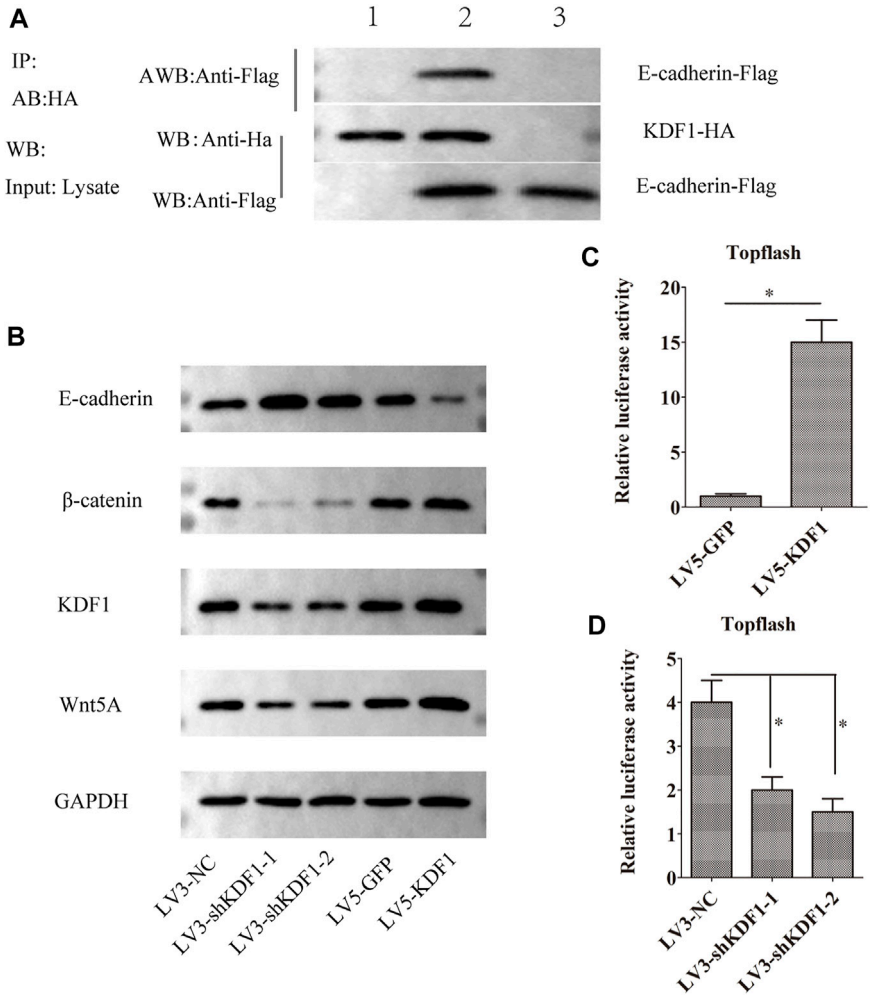


FIGURE 6 | KDF1 participated in the Wnt/β-catenin pathway in ovarian cancer. The Co-IP experiment detected the interaction between KDF1 and E-cadherin protein **(A)**. E-cadherin, Wnt5a, β-catenin, and KDF1 expression was detected using Western blot **(B)**. The regulation of KDF1 on the Wnt pathway was tested using dual luciferase reporter gene experiments **(C,D)**. Error bars represent the standard error. *indicates $p < 0.05$.

potential function of KDF1 in OC and to guide further research in OC. Elevated KDF1 expression in OV was associated with advanced clinical pathologic features (new events including recurrence or progression of the disease), poor prognosis, and survival time, which suggested that KDF1 is a potential prognostic and diagnostic marker deserving further research to validate.

Therefore, we used GO and KEGG databases to analyze the gene function of KDF1 and found that KDF1 was associated with the extracellular matrix, tight junction, and other factors. The composition and structure of the extracellular matrix are regulated to control cell behavior and differentiation. If extracellular matrix dynamics is dysregulated, it will lead to cancer and other diseases (Walker et al., 2018). The tight junction is one of the components of the cell junction complex, which includes tight junction, adhesive junction, and desmosomes. It maintains tissue integrity and promotes cell polarity during epithelial cell–cell junction. The tight junction is the critical intercellular junction to establish the epithelial barrier and maintain epithelial polarity (Otani and Furuse, 2020). Extracellular matrix and tight junction are related to tumor invasion and metastasis.

The enrichment of the GSEA gene revealed that high expression of KDF1 was related to the EMT process of breast cancer and was negatively correlated with the cell apical junction process. EMT is a conserved evolutionary cell development program. It participates in cancer by enhancing cell fluidity, invasiveness, and resistance to apoptosis stimulation, and endows cancer cells with metastatic characteristics (Mittal, 2018).

The PPI network analysis of KDF1 indicated that the essential protein E-cadherin might interact with KDF1. Cadherin is a calcium-dependent cell adhesion protein. E-cadherin was a key in establishing and maintaining polarized and differentiated epithelial cells through intercellular adhesion complexes. It participated in regulating epithelial cell adhesion, migration, and proliferation (Meigs et al., 2002).

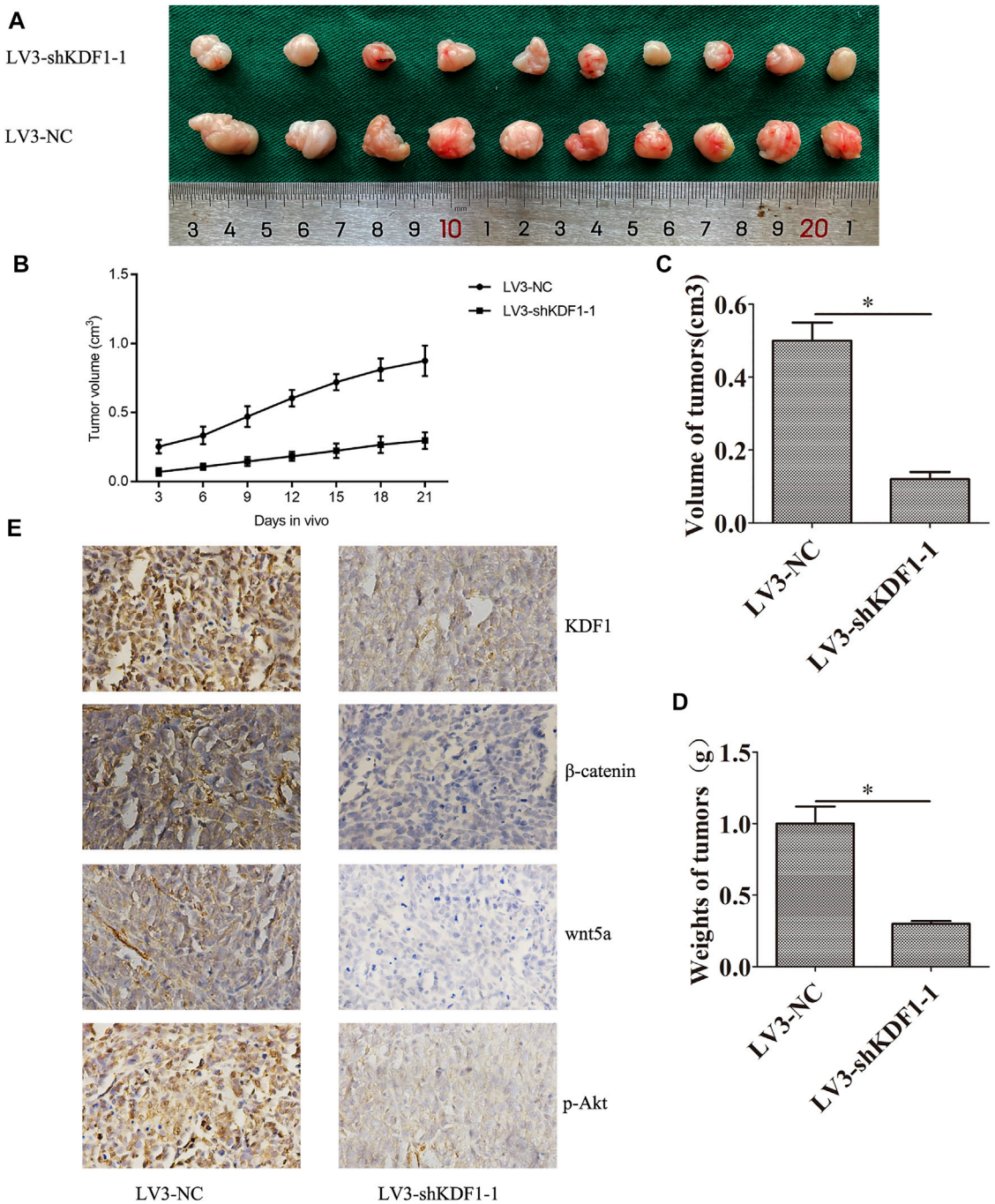


FIGURE 7 | Silencing KDF1 inhibited SKOV3 cells' growth *in vivo*. The picture of xenograft in two groups **(A)**: the growth curve of tumor in two groups **(B)**, the volume of tumor in different groups **(C)**, the weight of tumor in different groups **(D)**, and the expression of KDF1, Wnt5A, p-AKT, and β-catenin was detected by immunohistochemistry **(E)**. Error bars represent the standard error. *indicates $p < 0.05$.

Elevated KDF1 expression in OV was associated with advanced clinical pathologic features (platinum resistance, new events including recurrence and disease progression), poor prognosis, and survival time. Furthermore, in univariate and multivariate Cox regression analyses, we found that KDF1 was an independent prognostic factor after removing confounding factors, which showed a higher predictive value than many other clinical variables. Our results suggested that KDF1 is a potential prognostic and diagnostic marker deserving further clinical validation.

Bioinformatics research led us to see the strong potential of KDF1 in the occurrence, progression, and prognosis of ovarian

cancer, and guided our mechanism research direction into the next stage.

Then, we confirmed that KDF1 is highly expressed in ovarian cancer tissues in the immunohistochemistry assay. High expression of KDF1 was related to tumor stage and histological grade. Our study confirmed that the expression of KDF1 can regulate the phenotype and function of ovarian cancer cells. These effects of KDF1 were further verified by affecting xenograft tumor growth in nude mice. These data indicated that KDF1 was a potential diagnostic marker and therapeutic target for ovarian cancer. Those results suggested that KDF1 is related to cancer progression.

Moreover, we proved that KDF1 could interact with E-cadherin and participate in the Wnt signaling pathway to regulate the EMT process. The activation of the canonical Wnt/β-catenin pathway promoted proliferation and invasion of ovarian cancer (Arend et al., 2013). The Wnt/β-catenin pathway is vital in cell survival and has been implicated in the mechanism of chemoresistance of ovarian cancer (Yamamoto et al., 2019). Research shows that the Wnt/β-catenin pathway is also involved in ovarian tumor angiogenesis (Shoshkes-Carmel et al., 2018) and immune escape (Gregorieff and Clevers, 2005). Wnt activity is related to the grade (Wang et al., 2006), chemoresistance (Chau et al., 2013), and poor prognosis (Jacob et al., 2012; Arend et al., 2013) of patients with ovarian cancer. Meanwhile, the relationship between the Wnt/β-catenin pathway and epithelial-to-mesenchymal transition (EMT) (Arend et al., 2013) has been well documented for a long time.

β-Catenin forms a complex with E-cadherin and has an important role in maintaining epithelial integrity. The deregulation of E-cadherin can accelerate the process of β-catenin entering into the nucleus and activating genes downstream of the pathway (Tafrihi and Nakhaei Sistani, 2017). If the complex is destroyed, it can weaken the adhesion between cells and affect the Wnt signaling pathway (Tian et al., 2011). Wnt5a, which belongs to the Wnt family, is the main regulator of intraperitoneal metastasis and dissemination of ovarian cancer (Asem et al., 2020). In this study, silencing KDF1 can reduce the expression of Wnt5a and β-catenin, while overexpressing KDF1 can increase the expression of Wnt5a and β-catenin. It can be speculated that KDF1 is bound to E-cadherin and played a role in regulating key proteins of the Wnt/β-catenin pathway.

At the same time, dual-luciferase reporter experiment can further determine that the signal pathway reporter is significantly weakened in SKOV3 cells after silencing KDF1, while it is significantly enhanced in A2780 cells after overexpression of KDF1. It was further proved that the expression of KDF1 could affect Wnt/β-catenin pathway activity.

In the immunohistochemistry assay of xenograft tumors in nude mice, compared with the control LV3-NC group, the expression of β-catenin, Wnt5a, and p-Akt in the LV3-KDF1-1 group were decreased. The immunohistochemical results of transplanted tumors proved the results *in vitro* experiments and provided a basis for us to explore the possible interaction between KDF1 and other signal pathways.

We found that KDF1 participated in the Wnt/β-catenin pathway in ovarian cancer by empirical research. Also, we observed that KDF1 directly interacted with E-cadherin. It is suggested that KDF1 and cancer progression, especially through the interaction with E-cadherin, may participate in the Wnt/β-catenin pathway and regulate the EMT process, promoting the proliferation and invasion of OV.

Although our investigation of the relationship between KDF1 and OV helped in further understanding of the vital role of KDF1 in OV, some limitations remained: first, the correlation between KDF1 expression and platinum resistance and other clinical features that needs further investigation. Second, "clinical factors like the details of patients" treatment should be sufficiently considered to clarify the specific role of KDF1 in the development of OV. Third, there are two major shortages in the bioinformatics analysis part of our study. The one is that the treatment information was often inconsistent or even lacking in the public database to clarify the specific role of KDF1 in the development of OV comprehensively; the other is sample size imbalance. We have a smaller number of healthy samples in our control group than that of OV patients in our study; the sample size imbalance may lead to statistical bias. Therefore, future prospective studies are needed to reduce analysis bias.

CONCLUSION

In our study, we combined bioinformatics research and *in vivo* and *in vitro* study to systematically prove that KDF1 was a potential oncogene of ovarian cancer. All studies indicated that KDF1 can be a potential diagnostic marker and therapeutic target for ovarian cancer.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Sichuan Provincial People's Hospital.

AUTHOR CONTRIBUTIONS

RT supervised the whole work; CZ and JG conceived and designed the experiments and wrote the article; CZ, YL, and JG performed the experiments; CZ prepared experimental materials; CZ and YL analyzed the data. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by grants from the National Key Research and Development Program of China (2020YFC2005500) and by grants from the Key Research and Development Program of Science and Technology Department of Sichuan Province (2019YFS0514).

REFERENCES

Adhikary, S., Sanyal, S., Basu, M., Sengupta, I., Sen, S., Srivastava, D. K., et al. (2016). Selective Recognition of H3.1K36 Dimethylation/H4K16 Acetylation Facilitates the Regulation of All-Trans-Retinoic Acid (ATRA)-responsive Genes by Putative Chromatin Reader ZMYND8. *J. Biol. Chem.* 291, 2664–2681. doi:10.1074/jbc.M115.679985

Arend, R. C., Londoño-Joshi, A. I., Straughn, J. M., and Buchsbaum, D. J. (2013). The Wnt/β-Catenin Pathway in Ovarian Cancer: A Review. *Gynecol. Oncol.* 131, 772–779. doi:10.1016/j.ygyno.2013.09.034

Asem, M., Young, A. M., Oyama, C., Claure De La Zerda, A., Liu, Y., Yang, J., et al. (2020). Host Wnt5a Potentiates Microenvironmental Regulation of Ovarian Cancer Metastasis. *Cancer Res.* 80, 1156–1170. doi:10.1158/0008-5472.CAN-19-1601

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Chau, W. K., Ip, C. K., Mak, A. S. C., Lai, H.-C., and Wong, A. S. T. (2013). c-Kit Mediates Chemoresistance and Tumor-Initiating Capacity of Ovarian Cancer Cells through Activation of Wnt/β-Catenin-ATP-Binding Cassette G2 Signaling. *Oncogene* 32, 2767–2781. doi:10.1038/onc.2012.290

Fu, S.-W., Zhang, Y., Li, S., Shi, Z.-Y., Zhao, J., and He, Q.-L. (2020). LncRNA TTN-AS1 Promotes the Progression of Oral Squamous Cell Carcinoma via miR-411-3p/NFAT5 axis. *Cancer Cel Int* 20, 415. doi:10.1186/s12935-020-01378-6

Gregorieff, A., and Clevers, H. (2005). Wnt Signaling in the Intestinal Epithelium: from Endoderm to Cancer. *Genes Dev.* 19, 877–890. doi:10.1101/gad.1295405

Han, D., Li, J., Wang, H., Su, X., Hou, J., Gu, Y., et al. (2017). Circular RNA circMTO1 Acts as the Sponge of microRNA-9 to Suppress Hepatocellular Carcinoma Progression. *Hepatology* 66, 1151–1164. doi:10.1002/hep.29270

Jacob, F., Ukegjini, K., Nixdorf, S., Ford, C. E., Olivier, J., Caduff, R., et al. (2012). Loss of Secreted Frizzled-Related Protein 4 Correlates with an Aggressive Phenotype and Predicts Poor Outcome in Ovarian Cancer Patients. *PLoS One* 7, e31885. doi:10.1371/journal.pone.0031885

Lee, S., Kong, Y., and Weatherbee, S. D. (2013). Forward Genetics Identifies Kdf1/1810019J16Rik as an Essential Regulator of the Proliferation-Differentiation Decision in Epidermal Progenitor Cells. *Develop. Biol.* 383, 201–213. doi:10.1016/j.ydbio.2013.09.022

Li, Y., Tang, L., Yue, J., Gou, X., Lin, A., Weatherbee, S. D., et al. (2020). Regulation of Epidermal Differentiation through KDF 1-mediated Deubiquitination of IKK α. *EMBO Rep.* 21, e48566. doi:10.15252/embr.201948566

Liang, C.-C., Park, A. Y., and Guan, J.-L. (2007). *In Vitro* scratch Assay: a Convenient and Inexpensive Method for Analysis of Cell Migration *In Vitro*. *Nat. Protoc.* 2, 329–333. doi:10.1038/nprot.2007.30

Liu, B., Xia, X., Zhu, F., Park, E., Carbajal, S., Kiguchi, K., et al. (2008). IKKα Is Required to Maintain Skin Homeostasis and Prevent Skin Cancer. *Cancer Cell* 14, 212–225. doi:10.1016/j.ccr.2008.07.017

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Meigs, T. E., Fedor-Chaiken, M., Kaplan, D. D., Brackenbury, R., and Casey, P. J. (2002). Gα12 and Gα13 Negatively Regulate the Adhesive Functions of Cadherin. *J. Biol. Chem.* 277, 24594–24600. doi:10.1074/jbc.M201984200

Menon, U., Karpinskyj, C., and Gentry-Maharaj, A. (2018). Ovarian Cancer Prevention and Screening. *Obstet. Gynecol.* 131, 909–927. doi:10.1097/AOG.0000000000002580

Mittal, V. (2018). Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu. Rev. Pathol. Mech. Dis.* 13, 395–412. doi:10.1146/annurev-pathol-020117-043854

Mustapar, N., Zawawi, M. S. F., and Tuan Sharif, S. E. (2020). The Value of H3K27me3 Immunohistochemistry in Differentiating Malignant Peripheral Nerve Sheath Tumour with its Histologic Mimickers. *Asian Pac. J. Cancer Prev.* 21, 699–705. doi:10.31557/APJCP.2020.21.3.699

Nguyen, V. H. L., Hough, R., Bernaudo, S., and Peng, C. (2019). Wnt/β-catenin Signalling in Ovarian Cancer: Insights into its Hyperactivation and Function in Tumorigenesis. *J. Ovarian Res.* 12, 122. doi:10.1186/s13048-019-0596-z

Otani, T., and Furuse, M. (2020). Tight Junction Structure and Function Revisited. *Trends Cel Biol.* 30, 805–817. doi:10.1016/j.tcb.2020.08.004

Shamseldin, H. E., Khalifa, O., Binamer, Y. M., Almutawa, A., Arold, S. T., Zaidan, H., et al. (2017). KDF1, Encoding Keratinocyte Differentiation Factor 1, Is Mutated in a Multigenerational Family with Ectodermal Dysplasia. *Hum. Genet.* 136, 99–105. doi:10.1007/s00439-016-1741-z

Shiba-Ishii, A., Hong, J., Hirokawa, T., Kim, Y., Nakagawa, T., Sakashita, S., et al. (2019). Stratifin Inhibits SCFFBW7 Formation and Blocks Ubiquitination of Oncoproteins during the Course of Lung Adenocarcinogenesis. *Clin. Cancer Res.* 25, 2809–2820. doi:10.1158/1078-0432.CCR-18-3631

Shoshkes-Carmel, M., Wang, Y. J., Wangensteen, K. J., Tóth, B., Kondo, A., Massasa, E. E., et al. (2018). Subepithelial Telocytes Are an Important Source of Wnts that Supports Intestinal Crypts. *Nature* 557, 242–246. doi:10.1038/s41586-018-0084-4

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Tafrihi, M., and Nakhaei Sistani, R. (2017). E-Cadherin/β-Catenin Complex: A Target for Anticancer and Antimetastasis Plants/Plant-Derived Compounds. *Nutr. Cancer* 69, 702–722. doi:10.1080/01635581.2017.1320415

Tian, X., Liu, Z., Niu, B., Zhang, J., Tan, T. K., Lee, S. R., et al. (2011). E-Cadherin/β-Catenin Complex and the Epithelial Barrier. *J. Biomed. Biotechnol.* 2011, 1–6. doi:10.1155/2011/567305

Walker, C., Mojares, E., and del Río Hernández, A. (2018). Role of Extracellular Matrix in Development and Cancer Progression. *Ijms* 19, 3028. doi:10.3390/ijms19103028

Wang, Y., Hewitt, S. M., Liu, S., Zhou, X., Zhu, H., Zhou, C., et al. (2006). Tissue Microarray Analysis of Human FRAT1 Expression and its Correlation with the Subcellular Localisation of β-catenin in Ovarian Tumours. *Br. J. Cancer* 94, 686–691. doi:10.1038/sj.bjc.6602988

Yamamoto, T. M., McMellen, A., Watson, Z. L., Aguilera, J., Ferguson, R., Nurmemmedov, E., et al. (2019). Activation of Wnt Signaling Promotes Olaparib Resistant Ovarian Cancer. *Mol. Carcinog* 58, 1770–1782. doi:10.1002/mc.23064

Yang, Z., Wang, X.-L., Bai, R., Liu, W.-Y., Li, X., Liu, M., et al. (2016). miR-23a Promotes IKKα Expression but Suppresses ST7L Expression to Contribute to the Malignancy of Epithelial Ovarian Cancer Cells. *Br. J. Cancer* 115, 731–740. doi:10.1038/bjc.2016.244

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zeng, B., Lu, H., Xiao, X., Yu, X., Li, S., Zhu, L., et al. (2019). KDF1 Is a Novel Candidate Gene of Non-syndromic Tooth Agenesis. *Arch. Oral Biol.* 97, 131–136. doi:10.1016/j.archoralbio.2018.10.025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu, Liu, Tong and Guan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fgene.2022.813285

SNRPD1/E/F/G Serve as Potential Prognostic Biomarkers in Lung Adenocarcinoma

Gaohua Liu[1†], *Fuping Li*[2†], *Meichun Chen*[3], *Yang Luo*[1], *Yinhai Dai*[4*] and *Peifeng Hou*[1,5,6*]

[1]Department of Oncology Medicine, Fujian Medical University Union Hospital, Fuzhou, China, [2]Department of Clinical Medicine, Shaanxi University of Chinese Medicine, Xianyang, China, [3]Department of Hematology, Fujian Medical University Union Hospital, Fuzhou, China, [4]Department of Surgical Oncology Medicine, Second Affiliated Hospital of Shaanxi University of Chinese Medicine, Xianyang, China, [5]Fujian Medical University Stem Cell Research Institute, Fuzhou, China, [6]Fujian Key Laboratory of Translational Cancer Medicine, Fuzhou, China

Objectives: Sm proteins (SNRPB/D1/D2/D3/E/F/G), involved in pre-mRNA splicing, were previously reported in the tumorigenesis of several cancers. However, their specific role in lung adenocarcinoma (LUAD) remains obscure. Our study aims to feature abnormal expressions and mutations of genes for Sm proteins and assess their potential as therapeutic targets *via* integrated bioinformatics analysis.

Methods: In this research, we explored the expression pattern and prognostic worth of genes for Sm proteins in LUAD across TCGA, GEO, UALCAN, Oncomine, Metascape, David 6.8, and Kaplan-Meier Plotter, and confirmed its independent prognostic value *via* univariate and multivariate cox regression analysis. Meanwhile, their expression patterns were validated by RT-qPCR. Gene mutations and co-expression of genes for Sm proteins were analyzed by the cBioPortal database. The PPI network for Sm proteins in LUAD was visualized by the STRING and Cytoscape. The correlations between genes for Sm proteins and immune infiltration were analyzed by using the "GSVA" R package.

Results: Sm proteins genes were found upregulated expression in both LUAD tissues and LUAD cell lines. Moreover, highly expressed mRNA levels for Sm proteins were strongly associated with short survival time in LUAD. Genes for Sm proteins were positively connected with the infiltration of Th2 cells, but negatively connected with the infiltration of mast cells, Th1 cells, and NK cells. Importantly, Cox regression analysis showed that high SNRPD1/E/F/G expression were independent risk factors for the overall survival of LUAD.

Conclusion: Our study showed that SNRPD1/E/F/G could independently predict the prognostic outcome of LUAD and was correlated with immune infiltration. Also, this report laid the foundation for additional exploration on the potential treatment target's role of SNRPD1/E/F/G in LUAD.

Keywords: SM proteins, prognostic biomarkers, immune infiltration, target therapy, lung adenocarcinoma

OPEN ACCESS

Edited by:
Changwei Li,
Shanghai Jiao Tong University, China

Reviewed by:
Petar Ozretić,
Rudjer Boskovic Institute, Croatia
Yang Zhang,
Fudan University, China

***Correspondence:**
Yinhai Dai
jx2686@163.com
Peifeng Hou
houpeifeng123@fjmu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *11 November 2021*
Accepted: *14 February 2022*
Published: *03 March 2022*

Citation:
Liu G, Li F, Chen M, Luo Y, Dai Y and Hou P (2022) SNRPD1/E/F/G Serve as Potential Prognostic Biomarkers in Lung Adenocarcinoma. Front. Genet. 13:813285. doi: 10.3389/fgene.2022.813285

Abbreviations: BP, biological processes; CC, cellular components; FP, free progression; GEO, gene expression omnibus; GO, gene ontology; GTEx, genotype-tissue expression project; KEGG, kyoto encyclopedia of genes and genomes; LUAD, lung adenocarcinoma; MF, molecular function; OS, overall survival; PPS, post-progression survival; SLE, systemic lupus erythematosus; ssGSEA, single-sample gene sets enrichment analysis; snRNA, small nuclear RNA; TCGA, the cancer genome atlas.

INTRODUCTION

Lung cancer is the most well-known sort of malignant tumor worldwide and is the major cause of cancer mortality (Sung et al., 2021). Lung adenocarcinoma (LUAD) has been the most common subtype of Non-small cell cancer (NSCLC) (Anonymous, 2015). LUAD is characterized by a lack of early clinical symptoms, a high rate of distant metastasis and drug resistance, which pose serious challenges to clinical treatment. Currently, treatment methods for LUAD mainly include surgical resection, radiotherapy, chemotherapy, immunotherapy, and molecular targeted therapy (Reck and Rabe, 2017). The best treatment for lung cancer is surgical resection in the early stages of lung cancer. However, the early symptoms of the disease are not obvious, easy to be ignored, which leads to a late diagnosis. The treatment of advanced lung adenocarcinoma is limited, and molecular targeted therapy is a promising choice, as well as immunotherapy. However, because of the lack of effective molecular targets, most drugs remain ineffective in the treatment of LUAD patients, of whom the 5-year survival rate is just 15% (Chen et al., 2016). Hence, it is absolutely necessary to recognize effective and dependable biomarkers to determine poor prognoses and direct treatment strategies.

The spliceosome is a ribonucleoprotein (RNP) with a complex ring-shaped structure, which mainly consists of small nuclear ribonucleoproteins (snRNPs), which is involved in splicing the pre rna into mature mRNA. Seven Sm proteins (SNRPB, SNRPD1, SNRPD2, SNRPD3, SNRPE, SNRPF, and SNRPG) and a small eponymous small nuclear RNA (snRNA) compose snRNPs (Chari et al., 2008). Accurate splicing is essential for normal cellular functions like cell proliferation, apoptosis, migration, and invasion. Sm proteins, involved in the formation of anti-Sm antibodies, were significant diagnostic biomarkers in autoimmune diseases, such as systemic lupus erythematosus (SLE). Furthermore, priors works have illustrated that the aberrant expressions of genes for Sm proteins are related to some human cancers, including cervical cancer (Zhu et al., 2020), glioblastoma (Correa et al., 2016), breast cancer (Dai et al., 2021), and hepatocellular carcinoma (Zhan et al., 2020).

Up to now, there have been limited studies investigating the connection between the abnormal expressions of genes for Sm proteins and LUAD. For instance, a report (Liu et al., 2019) showed that SNRPB down-regulation inhibited the growth and metastasis of NSCLC cells *via* RAB26 down-regulation. In the study directed by Valles et al. (2012), SNRPB and SNRPE were proved to be related to poor survival in LUAD patients. However, the precise functional roles of Sm proteins in LUAD are unclear yet. Our study aimed to present abnormal expressions and mutations of genes for Sm proteins and assess their potential as therapeutic targets *via* integrated bioinformatics analysis, which may be helpful to the treatment of LUAD patients.

MATERIALS AND METHODS

Data Source and Flow Chart

The GSE40791 dataset data performed by GPL570 (Zhang et al., 2012) were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (Clough and Barrett, 2016). GSE40791 included 100 non-tumor lung tissues and 94 lung adenocarcinoma tissues (N = 100, T = 94). The TGCA-LUAD data (N = 59, T = 535) were downloaded from the TGCA database (https://portal.gdc.cancer.gov) (Tomczak et al., 2015). A flow chart of this study procedure is shown in **Figure 1**. The characteristics of 535 patients, including their ages, genders, smoking history, TNM stages, pathologic stages, primary therapy outcome, residual tumor and anatomic neoplasm subdivision, are presented in **Supplementary Table S1**.

Comparison of Genes for Sm Proteins Expression Levels Between LUAD and Corresponding Normal Tissue

The Oncomine database (https://www.oncomine.org) (Rhodes et al., 2004) was used to obtain the data of mRNA expressions for Sm proteins in various cancers, and these data were analyzed *via* Student's t-test. And we defined the following threshold: p-value<0.001, fold change >2 and 10% of most highly ranked genes. Furthermore, we compared expression levels of genes for Sm proteins between lung adenocarcinoma samples and normal samples in GSE40791. Moreover, the transcription levels of these genes were validated in the GEPIA 2 database (http://gepia2.cancer-pku.cn) (Tang et al., 2017), including tumor and normal samples from the TCGA and the Genotype-Tissue Expression Project (GTEx) database (https://gtexportal.org/home/) and the TCGA-LUAD data performed by R software.

Cell Culture and RT-qPCR

BEAS-2B cell line (human bronchial epithelial cell line) was purchased from the American Type Culture Collection (Manassas, VA, United States), A549 cell line was purchased from the Type Culture Collection of the Chinese Academy of Sciences, Shanghai, China. Two types of cells were cultured in DMEM with 10% FBS, penicillin (50 U/ml), and streptomycin (50 U/ml). The cells were incubated at 37°C with 5% CO_2.

Total RNA of the above cells was extracted using TRIzol (Thermo, United States), and first-strand complementary DNA (cDNA) synthesis from total RNA was carried out using the GoScript Reverse Transcription System (Promega, United States). RT-qPCR was conducted using an AriaMx Real-Time PCR machine (Agilent Technologies, United States) with TB GreenPremix ExTaq II (Takara Bio, Japan). The primer sequences are shown in **Supplementary Table S2**. RT-qPCR cycle conditions: 3 min at 95°C, 40 cycles of 15 s for 95°C and 60 s for 60°C. Data was normalized to the house-keeping gene GAPDH. The relative gene expression was performed by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

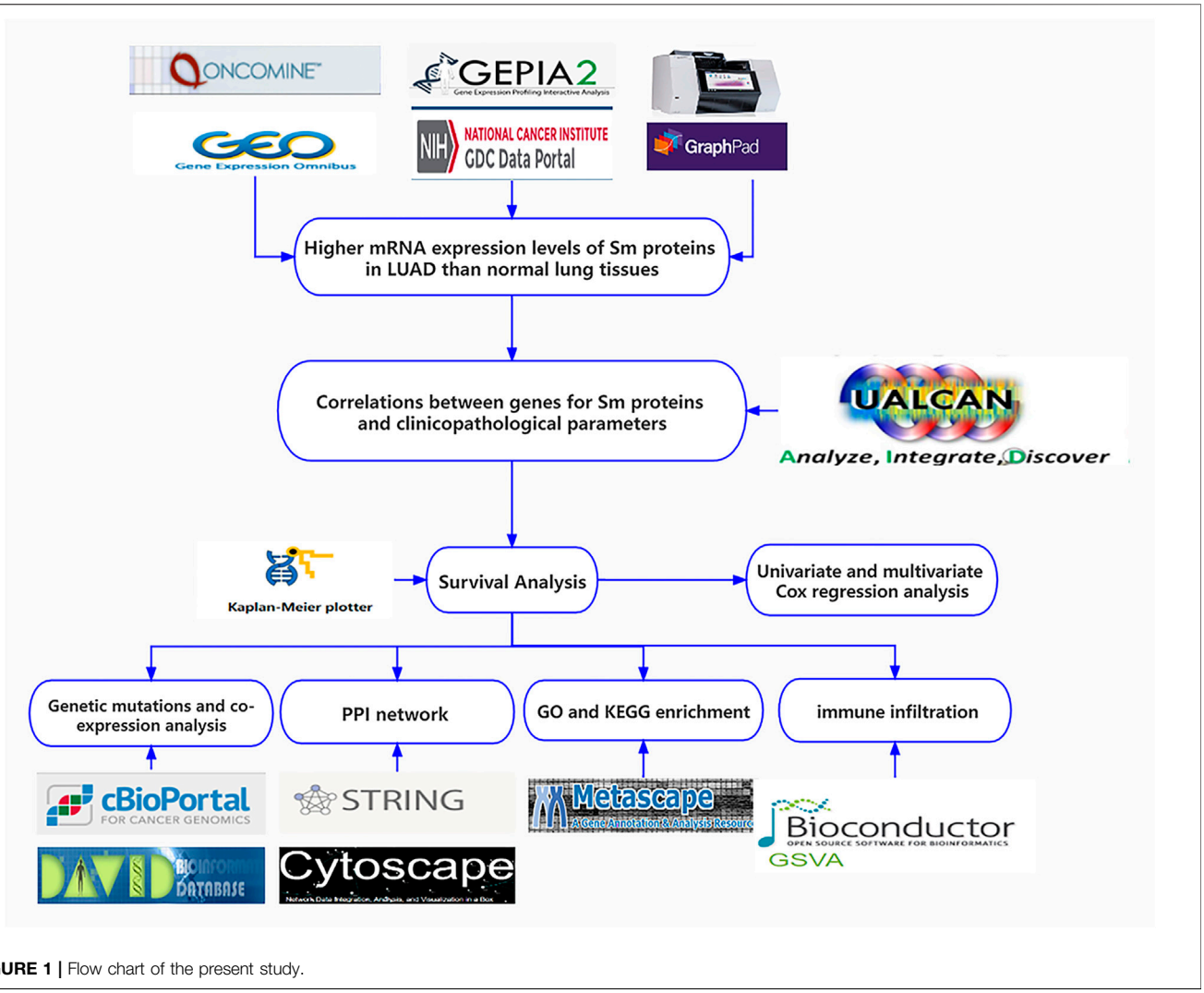


FIGURE 1 | Flow chart of the present study.

Correlations Between Genes for Sm Proteins and Clinicopathological Parameters

In our study, the transcription levels of genes for Sm proteins in TP53 mutation of LUAD patients were analyzed with the lung adenocarcinoma dataset using the UALCAN database (http://ualcan. path.uab.edu) (Chandrashekar et al., 2017). Moreover, we explored the correlation between genes for Sm proteins gene expressions and tumor stages *via* the "Expression DIY module" of GEPIA 2 and selected 50 neighboring genes related to genes for Sm proteins using the "Similar Genes Detection" module.

Survival Analysis

The Kaplan-Meier plotter (http://kmplot.com) (Gyorffy et al., 2013; Nagy et al., 2021), including mRNA expression data and patients' clinical information from GEO and TCGA, is a powerful online tool to further verify the prognostic value of genes in several cancers. The overall survival (OS), free progression (FP), and post-progression survival (PPS) curves of genes for Sm proteins in LUAD were shown *via* the Kaplan-Meier plotter. Meanwhile, we also explored if the dyregulation of genes for Sm proteins had any impacts on the OS of LUAD patients with smoke history by using the Kaplan Meier plotter. In addition, univariate and multivariate cox regression analysis of the TGCA-LUAD data was performed using the "survival" and the "survminer" R package.

Gene Mutations and Co-Expression Analysis

In this study, the lung adenocarcinoma dataset (TCGA, Firehose Legacy), including data from 230 complete samples of 586 patients, was selected and visualized as the map of gene mutations, expression heatmap, and co-expression map of genes for Sm proteins using the cBioPortal database (http://

www. cbioportal.org) (Cerami et al., 2012; Gao et al., 2013). The z-Score threshold was set to ±1.8.

Constructed Protein-Protein Interaction Network and Selected Hub Genes

The tool of STRING (https://string-db.org/) (Szklarczyk et al., 2019) was used to construct the protein-protein interaction (PPI) network between the seven genes for Sm proteins and their 50 frequently neighboring genes. All PPI pairs with a combined score of >0.4 were extracted. Then, we used the "CytoHubba" plugin (v0.1) (Chin et al., 2014) of Cytoscape (v3.8.2) (https://cytoscape.org) (Otasek et al., 2019) to identify hub genes in that PPI network. Furthermore, the online tool of Metascape (http://metascape.org) (Zhou et al., 2019) was used to analyze the MCODE components of that PPI network.

GO and KEGG Enrichment Analyses

Functional enrichment analyses, including gene ontology (GO) analysis comprising cellular component (CC), molecular function (MF), and biological process (BP), and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, were performed *via* the tool of DAVID 6.8 (https://david.ncifcrf.gov) (Huang da et al., 2009). Subsequently, the specific enriched terms for GO and KEGG enrichment analysis were visualized using the "clusterProfiler" package (Yu et al., 2012) in R software. Moreover, we presented the network of enriched terms colored by cluster ID using the online tool of Metascape.

Associations Between Genes for Sm Proteins and Immune Infiltration

Finally, we explore the associations between genes for Sm proteins and 24 immune cell types in TGCA-LUAD. We investigated dendritic cell (DC), activated dendritic cell (aDC), immature dendritic cell (iDC),plasmacytoid dendritic cell (pDC), B cells, $CD8^+$ T cells, Cytotoxic T cells, T cells, T helper cells, T help 1 (Th1) cells, Th17 cells, Th2 cells, T central memory (Tcm), T effector memory (Tem), T follicular helper (Tfh), T gamma delta (Tgd), regulatory T Cell (Treg), eosinophils, macrophages, mast cells, neutrophils, natural killer (NK) cells, NK $CD56^{bright}$ cells and NK $CD56^{dim}$ cells (Bindea et al., 2013), and the above associations was performed using single-sample Gene Sets Enrichment Analysis (ssGSEA) algorithm of R package "GSVA" (Hanzelmann et al., 2013) and lollipop charts were produced using R package "ggplot2."

Statistical Analysis

One-way Analysis of Variance (ANOVA) in the GEPIA 2 database, the log-rank test in Kaplan-Meier survival analysis, and the Cox proportional risk models in univariate and multivariate analyses, were used in statistical analysis. R software (v4.0.2) (http://www.r-project.org) was used for analysis in this study. RT-qPCR data was performed in GraphPad Prism (v9.0.2) (San Diego, CA, United States) and presented as the mean ± S.D. Student's T-test was used for statistical analyses between the data pairs where appropriate. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$ were considered to represent statistical significance.

RESULTS

The mRNA Levels of Sm Proteins Were Significantly Increased in Lung Cancer Tissues and LUAD Cell Lines

From **Figure 2A** and **Table 1**, we could see that the mRNA levels of SNRPB/D1/D2/E/G were higher in lung cancer than in non-cancer tissues. The results were also consistent with the mRNA levels of Sm proteins were upregulated in LUAD compared to normal lung tissues (**Figure 2B**). Moreover, the relative mRNA level of all Sm proteins in A549 cell line is higher than the relative mRNA level of them in normal lung cell lines, BEAS-2B (**Figure 2C**). The transcription levels of genes for Sm proteins considerably increased in data of GEPIA 2 (TCGA-LUAD & GTEx), TCGA-LUAD data and pair LUAD data (**Figures 3A–C**) match the above results.

The Expressions of Genes for Sm Proteins Were Related to Clinicopathological Parameters of LUAD Patients

We research the correlations between the expressions of genes for Sm proteins, TP53 mutation and cancer stages *via* the UALCAN database and the GEPIA2 database. From **Figure 4A**, we could find that the mRNA expressions of genes for Sm proteins in LUAD with TP53 mutation were higher than those without TP53 mutation (**Figure 4A**). Moreover, as can be seen in **Figure 4B**, the mRNA levels of SNRPE varied significantly across different LUAD tumor stages (F value = 3.68, Pr (>F)<0.05), whereas SNRPB/D1/D2/D3/F/G had no significant association with LUAD stages (all Pr (>F) > 0.05).

The Upregulation of Genes for Sm Proteins is Related to Poor Survival Outcomes in LUAD Patients

Using Kaplan-Meier Plotter, we assessed the prognostic values of genes for Sm proteins in LUAD patients. From **Figure 5A**, we could find out high expressions of genes for Sm proteins were associated with short OS (all $p < 0.05$). In addition, high mRNA level of SNRPB/D1/D2/F (HR > 1, $p < 0.05$) were correlated with poor FP and PPS (all $p < 0.05$). However, highly expressed SNRPD3 was not obviously related to the FP of LUAD patients ($p > 0.05$).

As is well known, smoking is a common risk factor leading to lung cancer (Jassem et al., 2009). Hence, we explored if the upregulation of genes for Sm proteins had any impacts on the overall survival of LUAD patients. From **Figure 5B**, in the smoking group, the high expression of SNRPB significantly

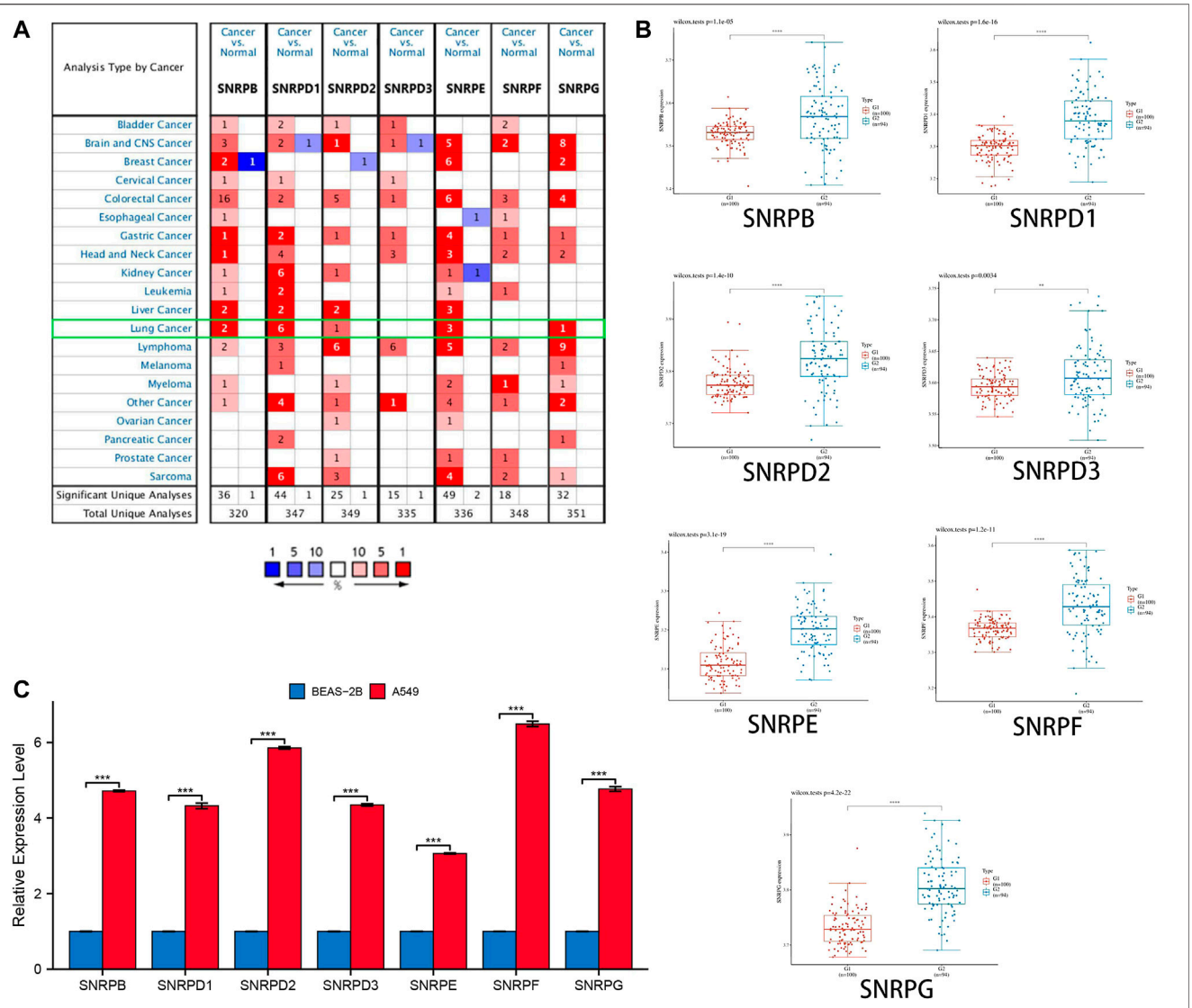


FIGURE 2 | The transcription levels of genes for Sm proteins in lung cancer tissues and lung cell lines. **(A)** The genes for Sm proteins in different human cancers (ONCOMINE). **(B)** Box plots of genes for Sm proteins expression between lung cancer and normal tissues in GSE40791. **(C)** Bar charts of mRNA level of genes for Sm proteins between BEAS-2B and A549 cell line. Students T-test was performed to assess the statistical significance. A p value of <0.05 was regarded as statistically significant.

reduced the OS of LUAD patients. However, in the non-smoking group, the expression of SNRPB had no statistical significance for the overall survival of patients with lung adenocarcinoma. This indicates that SNRPB and smoking factors together could promote the development of lung adenocarcinoma. However, the upregulation of SNRPD1 (HR: 2.36, $p < 0.05$ vs HR: 1.63, $p < 0.05$), SNRPD2 (HR: 3.29, $p < 0.05$ vs HR: 2.52, $p < 0.05$), and SNRPF (HR: 3.27, $p < 0.05$ vs HR: 2.18, $p < 0.05$) had worsening impacts on the OS of LUAD patients without smoking history. Noteworthy, only high expression of SNRPE (HR: 1.28, $p < 0.05$) was correlated to poor OS of lung squamous cell carcinoma (LUSC) (**Supplementary Figure S1**). More research is still needed to interpret potential mechanisms of the above-described events.

Moreover, we attempted to evaluate the independent prognostic values of genes for Sm proteins for OS of LUAD patients. The microarray data and clinical information (**Supplementary Table S1**) of 535 LUAD patients were obtained from TCGA for Cox regression analysis. Univariate Cox regression analysis demonstrated that T, N, M stage, pathologic stage, SNRPD2, SNRPE, and SNRPF were highly correlated with the short OS of LUAD patients (all $p < 0.05$; **Supplementary Table S2**). In multivariate Cox regression analysis, SNRPD1, SNRPE, SNRPF, and SNRPG were observed to have significant associations with poor OS of

TABLE 1 | The mRNA levels of Sm proteins were significantly higher in lung cancer than in normal lung tissues (ONCOMINE).

	Lung cancer VS normal lung tissue	Fold change	*p*-value	T-test	References
SNRPB	Small cell lung carcinoma	3.338	1.21E-4	6.278	Garber et al. (2001)
	Squamous cell lung carcinoma	2.621	1.52E-4	5.135	Garber et al. (2001)
SNRPD1	Large cell lung carcinoma	2.357	9.91E-5	7.079	Garber et al. (2001)
	Small cell lung carcinoma	2.223	1.40E-4	6.131	Garber et al. (2001)
	Squamous cell lung carcinoma	3.502	5.49E-4	7.549	Yamagata et al. (2003)
	Squamous cell lung carcinoma	2.194	7.89E-17	13.403	Hou et al. (2010)
	Large cell lung carcinoma	2.526	5.01E-8	8.052	Hou et al. (2010)
	Squamous cell lung carcinoma	2.478	9.30E-4	5.237	Wachi et al. (2005)
SNRPD2	Squamous cell lung carcinoma	2.046	8.33E-4	5.359	Wachi et al. (2005)
SNRPD3	NA	NA	NA	NA	NA
SNRPE	Large cell lung carcinoma	2.347	7.58E-9	9.113	Hou et al. (2010)
	Squamous cell lung carcinoma	2.146	2.53E-4	7.002	Wachi et al. (2005)
	Lung carcinoid	2.329	2.86E-4	3.974	Bhattacharjee et al. (2001)
SNRPF	NA	NA	NA	NA	NA
SNRPG	Squamous cell lung carcinoma	2.176	1.80E-5	8.278	Wachi et al. (2005)

NA: not available.

LUAD patients (all $p < 0.05$; **Supplementary Figure S2**). Taken together, SNRPD1/E/F/G could be considered as independent prognostic factors for poor OS in LUAD.

Gene Mutations and Co-Expression Analysis of Genes for Sm Proteins in LUAD Patients

Using the cBioPortal database, we explored the mutation rate and co-expression of Sm proteins. As shown in **Figures 6A–C**, gene amplification was the most common type of genes for Sm proteins mutations (**Figure 6A**). 22% (51/230) of LUAD patients were found to have more than one gene mutation. SNRPB, SNRPD1, SNRPD2, SNRPD3, SNRPE, SNRPF, and SNRPG were altered in 2.6%, 4%, 3%, 4%, 8%, 4%, and 3% of the 230 LUAD patients (**Figure 6B**). Likewise, the Expression Heatmap also displayed the degree of genes for Sm proteins mutations (**Figure 6C**). SNRPE had the highest mutation rate among genes for Sm proteins in LUAD. Meantime, we performed co-expression analysis for Sm proteins. Significant and positive correlations were observed among genes for Sm proteins in **Figure 6D**, including SNRPB with SNRPD1/D2/E/F/G, SNRPD1 with SNRPB/D2/D3/E/F/G, SNRPD2 with SNRPB/D1/E/F/G, SNRPD3 with SNRPD1/E/F/G, SNRPE with SNRPB/D1/D2/D3/F/G, SNRPF with SNRPB/D1/D2/D3/E/G, and SNRPG with SNRPB/D1/D2/E/F (**Figure 6D**).

Constructed PPI Network and Selected Hub Genes

To explore the possible genes for Sm proteins protein-mediated biological pathways in lung adenocarcinoma, we constructed a protein interaction network based on the seven Sm proteins (**Figure 6E**) and their 50 frequently neighboring proteins (**Figure 6F**) *via* the STRING database. Meantime, we used Cytoscape software to find out these hub genes related to genes for Sm proteins-mediated biological pathways. As was shown in **Figure 6G**, eight hub genes, including PCNA, NUF2, UBE2T, CCDC59, ERCC1, CDCA3, QPCTC, and TROAP, were tightly correlated with the alterations of genes for Sm proteins. Furthermore, we extracted the three most meaningful MCODE components from the PPI network between Sm proteins and their 50 frequently neighboring proteins. As was shown in **Figure 6H**, Sm core complex, spliceosome, U2-SnRNP, chromosome segregation, chromosome, centromeric region, cell division, and nucleotide excision repair were related to biological function.

GO and KEGG Enrichment Analyses in LUAD Patients

Afterward, genes for Sm proteins and their 50 neighboring genes were analyzed *via* the tool of DAVID 6.8 for GO and KEGG functional enrichment analysis. The results were shown in **Figure 7** and **Supplementary Figure S3**. Biological processes (BP) included spliceosomal snRNP assembly, histone mRNA metabolic process, negative regulation of nuclear division, nuclear division, and mitotic spindle assembly checkpoint (**Figure 7A**). Cellular components (CC) suggested that SNRP genes existed mainly in methylosome, U4 snRNP, U2-type catalytic step 2 spliceosome, SMN-Sm protein complex, U1 snRNP, U2 snRNP, and U12-type spliceosomal complex (**Figure 7A**). Molecular function (MF) indicated that SNRP genes were related to the structural constituents of catalytic activity, acting on DNA, ribonucleoprotein complex binding, DNA N-glycosylase activity, histone kinase activity, and ubiquitin-protein transferase regulator activity (**Figure 7A**). KEGG pathway enrichment analysis suggested that spliceosome, base excision repair, nucleotide excision repair, cell cycle, dna replication, and mismatch repair were significantly associated with the alterations of sm proteins. The complex process of mRNA/RNA splicing was shown in **Supplementary Figure S4**. In addition, the KEGG chord plot

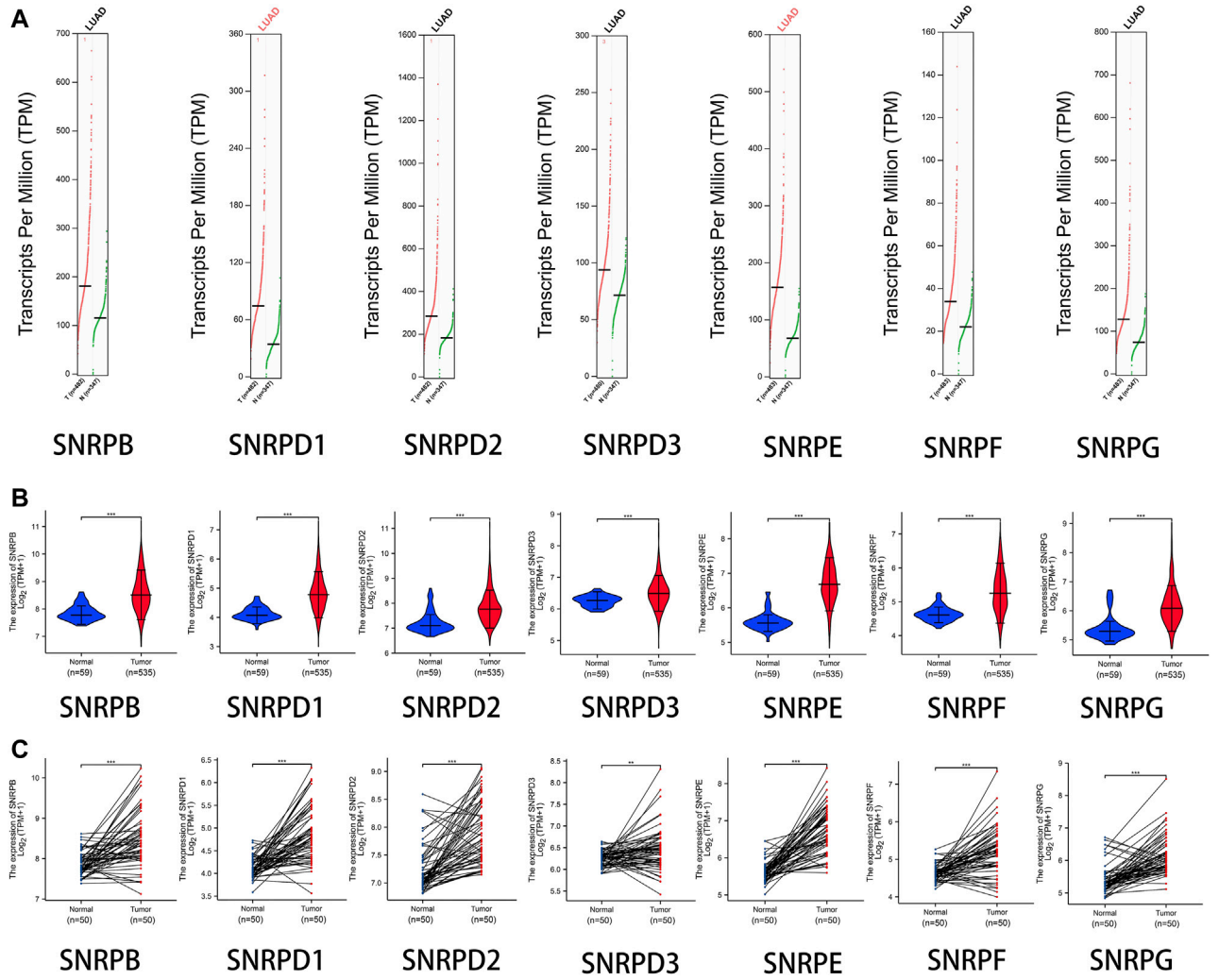


FIGURE 3 | The transcription levels of genes for Sm proteins in LUAD samples and normal lung samples. **(A)** The transcription levels of genes for Sm proteins in data of GEPIA 2 (TCGA-LUAD & GTEx). **(B)** Violin charts of mRNA level for Sm proteins in data of TCGA-LUAD. **(C)** The mRNA expressions of genes for Sm proteins in pair data of TCGA-LUAD. Students T-test was performed to assess the statistical significance. A p value of < 0.05 was regarded as statistically significant.

showed that PCNA was related to 5 vital KEGG pathways. Thus, PCNA may be a very critical gene (**Figure 7B**). To better understand the association among different enriched terms, the KEGG network of enriched terms, colored by cluster ID, was constructed *via* the online tool of Metascape (**Figure 7C**).

Associations Between Genes for Sm Proteins and Immune Infiltration

High genes for Sm proteins expression was considerably related to abundance infiltration of immune cell in the tumor immune microenvironment (**Figure 8**). Thus, further research of genes for Sm proteins in 24 immune cell populations manifested that genes for Sm proteins had positive correlations to Th2 cells, Tgd cells, but an adverse correlation to Th1 cells, mast cells, and NK cells.

DISCUSSION

Lung cancer has been the major cause of tumor occurrence and mortality around the world (Sung et al., 2021). LUAD accounts for half part of lung cancer with an extremely low survival rate (Chen et al., 2016). Although the research on the biological characteristics of lung adenocarcinoma has made some progress in the past few decades, the tumor progression mechanism of lung adenocarcinoma is still unclear. Through literature review and online database bioinformatics analysis and experimental verification, this study verified that genes for Sm proteins are highly expressed in lung adenocarcinoma and have a significant impact on the prognosis of lung cancer. SnRNPs, mainly responsible for splicing the pre-RNA into mRNA, are composed of seven genes for Sm proteins and a small eponymous snRNA (Chari et al., 2008). Accurate splicing is essential to ensure

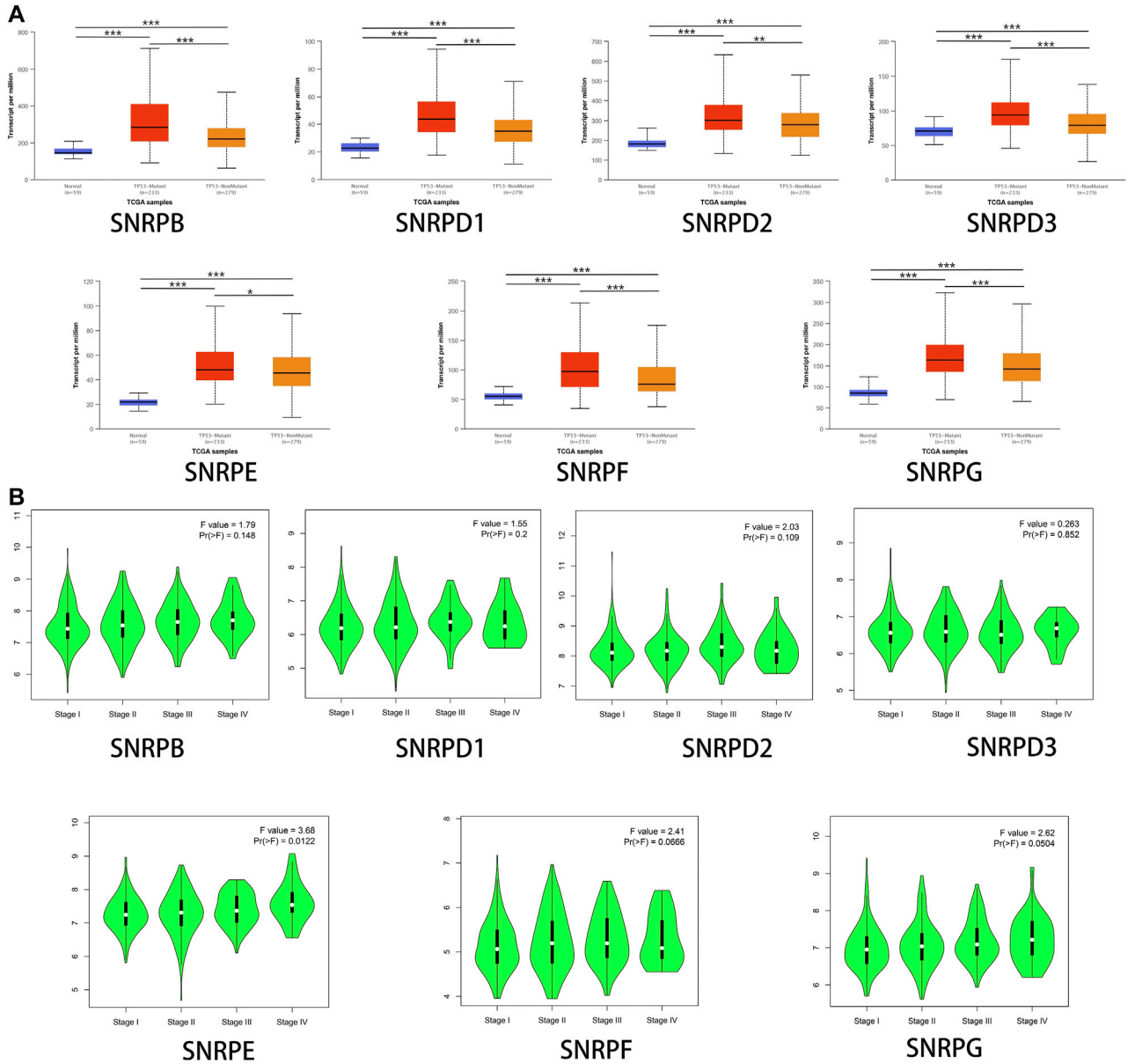


FIGURE 4 | The associations between genes for Sm proteins expression levels and clinicopathological parameters of LUAD patients. **(A)** Correlations between genes for Sm proteins expression levels and TP53 mutation in LUAD (UALCAN). **(B)** Association between mRNA expression of genes for Sm proteins and different tumor stages (GEPIA 2). ANOVA was performed to assess the statistical significance of the variations. F-value indicates the statistical value of the F test; Pr (>F) indicates p value. A p value of <0.05 was regarded as statistically significant.

normal cellular function like cell proliferation, apoptosis, migration, and invasion.

This study covered many aspects, including differential expression analysis, association with clinicopathological parameters, survival analysis, gene mutations, co-expression, functional enrichment analysis, and correlation with immune infiltration, revealing the roles of SNRPB/D1/D2/D3/E/F/G in LUAD.

Previous study reported that SNRPB promotes the tumor formation of NSCLC by regulating RAB26 and SNRPB may predict response to cisplatin-based chemotherapy for NSCLC patients (Liu et al., 2019; Liu et al., 2021). Here, the mRNA expression of SNRPB was found to be upregulated in LUAD compared to non-cancer tissues. Also, SNRPB is highly expressed in cervical cancer (Zhu et al., 2020), glioma (Correa et al., 2016), and hepatocellular cancer (Zhan et al., 2020), which is a prognostic factor. The mechanistic study also reported that overexpressed SNRPB played a carcinogenic role in the progression of LUAD and was mediated by c-Myc (Peng et al., 2020). Moreover, SNRPB was highly correlated with

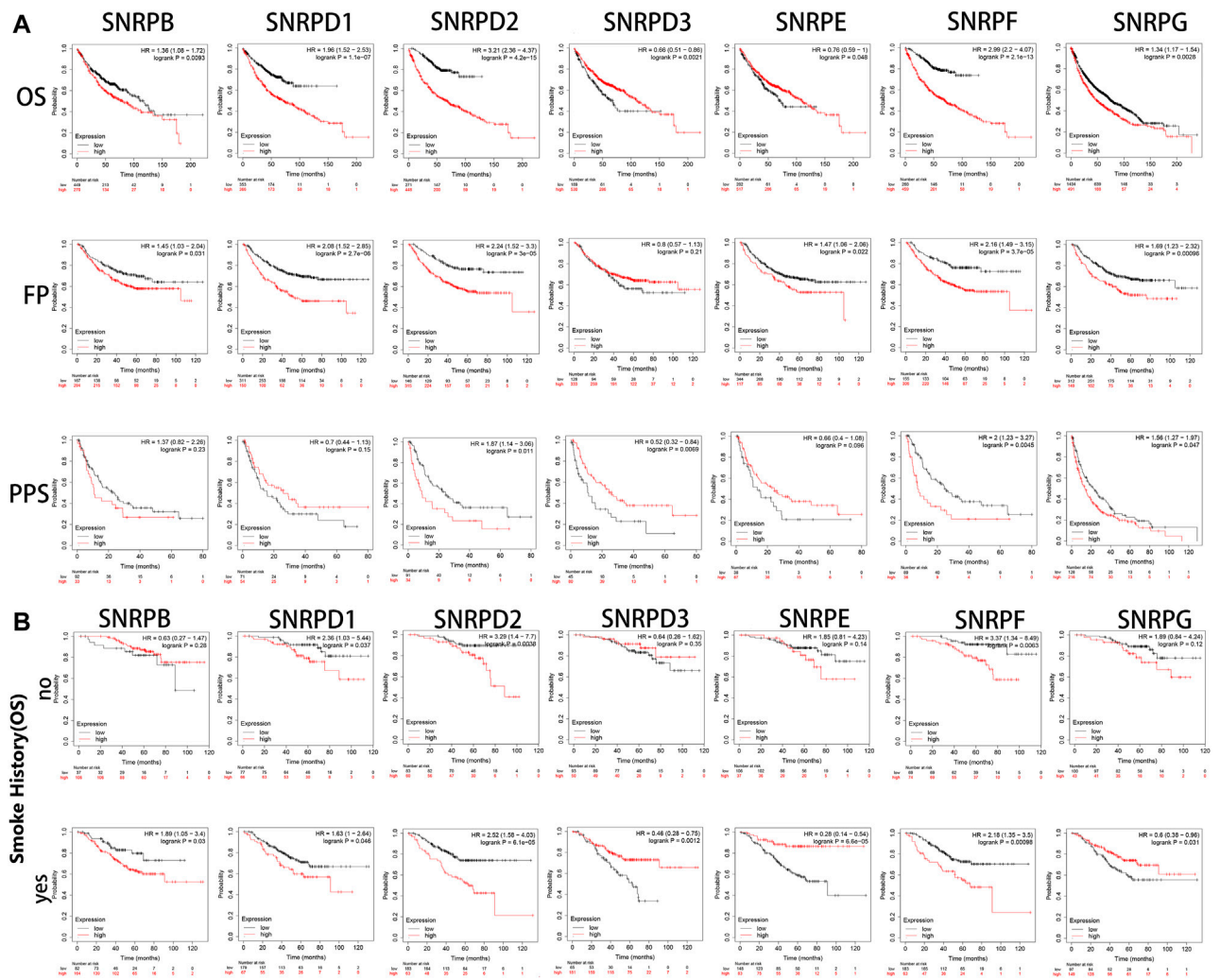


FIGURE 5 | The survival curves of LUAD patients with genes for Sm proteins and impacts of smoking on the survival of these patients (Kaplan-Meier Plotter). **(A)** The Kaplan-Meier survival curves were used to evaluate the survival of LUAD patients with Sm proteins. **(B)** The effects of smoking in the OS of LUAD patients with Sm proteins. The log-rank test was performed in Kaplan-Meier survival analysis. A p value of <0.05 was regarded as statistically significant. HR: hazard ratio.

TP53 mutation, tumor grades, and cancer stages. Interestingly, our study revealed that SNRPB up-regulation was associated with poor OS in LUAD patients, especially in patients with smoking history. Taken together, our study revealed that SNRPB could promoted tumor progression in LUAD.

It has been generally accepted that SNRPD1 is related to SLE (Riemekasten et al., 1998). However, only a few studies have reported the role of SNRPD1 in human cancers. SNRPD1 down-regulation was related to poorer survival in patients with ovarian cancer (Bao et al., 2020). SNRPD1 up-regulation contributed to breast cancer cell proliferation (Dai et al., 2021). Here, we found that SNRPD1 expression was significantly up-regulated in LUAD compared to normal tissues or normal cell lines. Furthermore, SNRPD1 expression was correlated with TP53 mutation, tumor grades, and cancer stages. Overexpression of SNRPD1 could be relevant to the poor OS in LUAD patients. Multivariate Cox regression analysis indicated that SNRPD1 was an independent prognostic factor for poorer OS of LUAD patients. Taken together, all of the above findings support the opinion that SNRPD1 plays an essential role in the carcinogenic effect of LUAD.

SNRPD2 is found closely associated with several cancers, including triple-negative breast cancer (TNBC) (Koedoot et al., 2021) and hepatocellular carcinoma. Interestingly, SNRPD2 and SNRPG were not only major pathogenic genes of Alzheimer's disease but also bridge genes (Tao et al., 2020). However, it is not yet clear to us the specific role that SNRPD2 plays in LUAD. Here, we report SNRPD2 was found overexpressed in LUAD. Additionally, high SNRPD2 expression was correlated with TP53 mutation, tumor grades, and cancer stages.

SNRPD3 (smD3), like SNRPD2, also plays a vital role in TNBC. Moreover, several studies have shown that SNRPD3

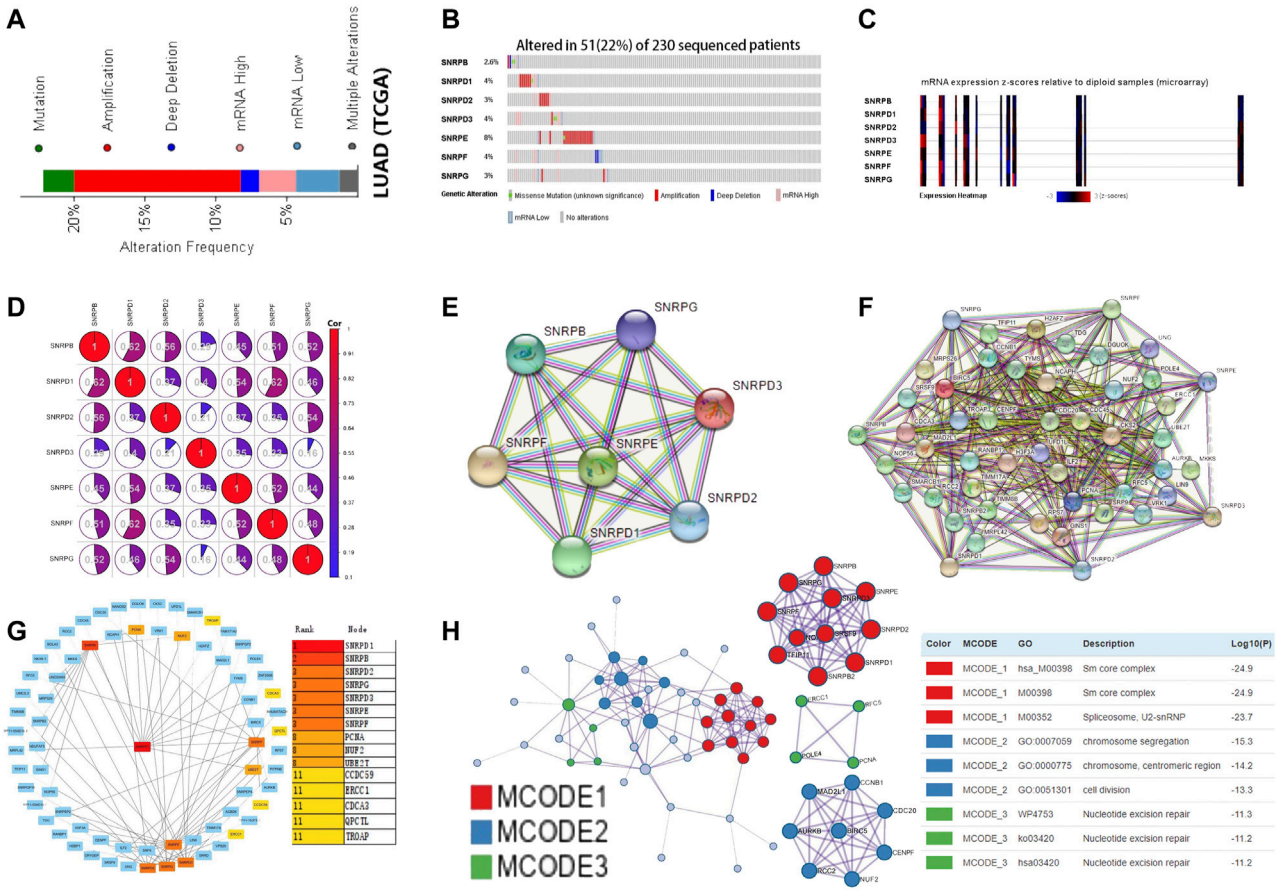


FIGURE 6 | Gene mutations, co-expression, and PPI network for genes for Sm proteins in LUAD (cBioPortal, STRING, Cytoscape, Metascape). **(A–C)** Gene mutations and expression heatmap of genes for Sm proteins in LUAD patients (cBioPortal). **(D)** The co-expression network for genes for Sm proteins (cBioPortal). **(E)** PPI network based on the seven genes for Sm proteins (STRING). **(F)** PPI network based on genes for Sm proteins and their 50 neighbouring proteins (STRING). **(G)** Eight hub genes, including PCNA, NUF2, UBE2T, CCDC59, ERCC1, CDCA3, QPCTC, and TROAP, are shown in the PPI network (Cytoscape). **(H)** Protein-protein interaction network and MCODE components identified in different expressed genes for Sm proteins and 50 most frequently altered neighboring genes (Metascape).

expression is relevant to breast cancer (Koedoot et al., 2021) and NSCLC (Blijlevens et al., 2020). For instance, it was investigated that silencing SNRPD3 was able to promote TP53 expression and kill NSCLC cells effectively (Olst et al., 2017). In this study, higher SNRPD3 expression was found in LUAD. Moreover, SNRPD3 expression was associated with TP53 mutation, tumor grades, and cancer stages. Multivariate Cox regression analysis suggested a strong association between overexpressed SNRPD3 with the poor OS of LUAD patients, which seemed congruous with the oncogenic role of SNRPD3.

SNRPE protein is a core component of Sm proteins, which has been reported in some malignancies, including bladder cancer (Tapak et al., 2015), prostate cancer (Anchi et al., 2012), hepatocellular carcinoma (Jia et al., 2011), and non-small cell lung cancer (Valles et al., 2012). As an example, SNRPE was a deregulated RNA metabolism-related genes of LUAD by bioinformatics analysis (Valles et al., 2012). Similarly, in our report, higher SNRPE expression was found in LUAD than in normal lung tissues. Furthermore, SNRPE expression was relevant to TP53 mutation and tumor grades. Overexpressed SNRPE was an independent prognostic factor for poorer OS in LUAD patients by cox analysis. Multivariate Cox regression analysis indicated that SNRPE was an independent prognostic factor for poorer OS of LUAD patients, suggesting that SNRPG played a carcinogenic role in LUAD. Taken together, our study revealed that SNRPE could promoted oncogenesis in LUAD.

SNRPF was a novel biomarker of colorectal cancer *via* complex gene interaction networks (Sun et al., 2016). However, the role of SNRPF was rarely reported in LUAD. In our study, highly expressed SNRPF was found in LUAD compared to non-cancer tissues. Moreover, abnormal SNRPF expression could be relevant to TP53 mutation, tumor grades, and cancer stages. Multivariate Cox regression analysis indicated that SNRPDF was an independent prognostic factor for poorer OS of LUAD patients. Given the above results, our study showed that SNRPF could play a carcinogenic role in LUAD.

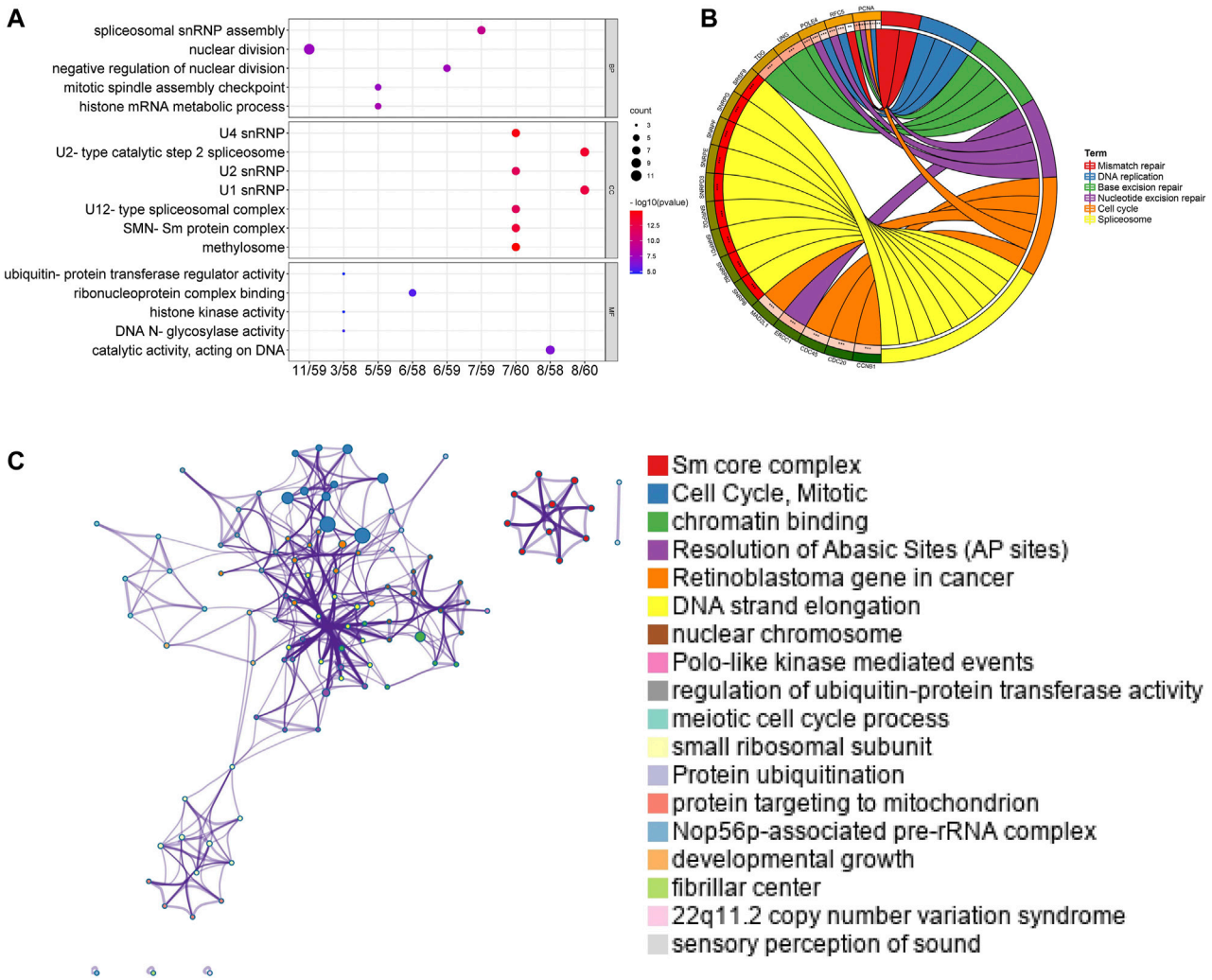


FIGURE 7 | Functional and enrichment analysis for Sm proteins genes and their 50 neighboring genes in LUAD (DAVID 6.8, Metascape). **(A)** GO enrichment analysis of genes for Sm proteins and their 50 neighboring genes (DAVID 6.8). **(B)** Chord diagram of KEGG analyses results. **(C)** Network of seven enriched terms, colored by cluster ID (Metascape).

Downregulated SNRPG could inhabit glioblastoma cells proliferation by p53 signaling pathway (Lan et al., 2020). In this paper, higher mRNA expression of SNRPG was explored in LUAD. Furthermore, SNRPD1 expression could also be relevant to TP53 mutation, tumor grades, and cancer stages. Multivariate Cox regression analysis indicated that SNRPG was an independent prognostic factor for poorer OS of LUAD patients, suggesting that SNRPG played a carcinogenic role in LUAD.

Notably, in this paper, we assessed the relation between genes for Sm proteins and 22 immune infiltration cells in LUAD. genes for Sm proteins gene expressions were positively related to the infiltration of Th2 cells and negatively related to the infiltration of Th1 cells, mast cells, and NK cells. Furthermore, prior studies also suggested that spliceosome was tightly associated with the immune microenvironment. Nevertheless, further studies are still needed to develop immunosuppressants of individual genes for Sm proteins members and apply them to the diagnostic of LUAD.

In this work, there are a few limitations that need to be recognized. Firstly, the data used to assess the prognostic worth of genes for Sm proteins in LUAD patients were mainly from the TGCA database and GEO database. Although the TGCA and GEO sequencing data were experimentally confirmed, additional large sample studies on LUAD patients from other databases are necessary to validate our results. Secondly, Cox analysis indicated that the expressions of SNRPD1//E/F/G were correlation to shorter OS in LUAD. Consequently, SNRPD1/E/F/G were considered as independent prognostic factors and potential prognostic biomarkers for LUAD. Finally, further experiments in cells and animal models are needed to elucidate the underlying mechanisms of how genes for Sm proteins play roles in LUAD.

Taken together, our work illustrates that SNRPD1/E/F/G could independently predict the prognostic outcome of LUAD

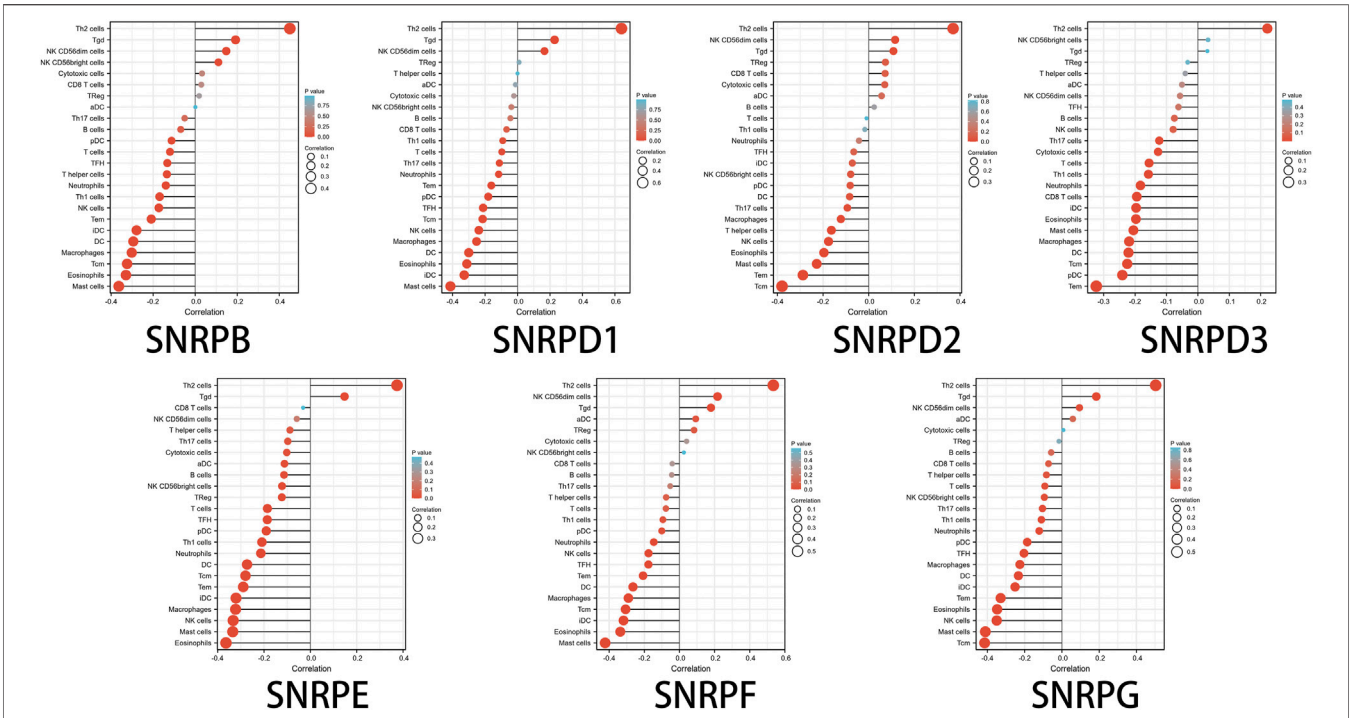


FIGURE 8 | The lollipop charts of the associations between 24 immune cell types and Sm proteins.

and was correlated with immune infiltration. Our findings laid the foundation for further exploration on the promising therapy target's role of Sm proteins in LUAD.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

GL, FL, and MC contributed to the conception of the study; GL and YL performed cell culture and RT-qPCR experiment; GL, FL, and MC contributed significantly to analysis and manuscript preparation; GL, FL, and MC performed the data analyses and wrote the manuscript; GL, FL, MC, YD, and PH helped perform the analysis with constructive discussions, YD and PH revised the manuscript.

ACKNOWLEDGMENTS

We thank all the members for participating in this study. Also we thank the Fujian Medical University Union Hospital for article fee reimbursement.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.813285/full#supplementary-material.

REFERENCES

Anchi, T., Tamura, K., Furihata, M., Satake, H., Sakoda, H., Kawada, C., et al. (2012). SNRPE Is Involved in Cell Proliferation and Progression of High-Grade Prostate Cancer through the Regulation of Androgen Receptor Expression. *Oncol. Lett.* 3, 264–268. doi:10.3892/ol.2011.505

Anonymous (2015). Non-small-cell Lung Cancer. *Nat. Rev. Dis. Primers.* 1, 15048. doi:10.1038/nrdp.2015.48

Bao, M., Zhang, L., and Hu, Y. (2020). Novel Gene Signatures for Prognosis Prediction in Ovarian Cancer. *J. Cel. Mol. Med.* 24, 9972–9984. doi:10.1111/jcmm.15601

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., et al. (2001). Classification of Human Lung Carcinomas by mRNA Expression Profiling Reveals Distinct Adenocarcinoma Subclasses. *Proc. Natl. Acad. Sci.* 98, 13790–13795. doi:10.1073/pnas.191502998

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Blijlevens, M., Komor, M. A., Sciarrillo, R., Smit, E. F., Fijneman, R. J. A., and Van Beusechem, V. W. (2020). Silencing Core Spliceosome Sm Gene Expression Induces a Cytotoxic Splicing Switch in the Proteasome Subunit Beta 3 mRNA in Non-small Cell Lung Cancer Cells. *Int. J. Mol. Sci.* 21, 4192. doi:10.3390/ijms21124192

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.cd-12-0095

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., et al. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19, 649–658. doi:10.1016/j.neo.2017.05.002

Chari, A., Golas, M. M., Klingenhäger, M., Neuenkirchen, N., Sander, B., Englbrecht, C., et al. (2008). An Assembly Chaperone Collaborates with the SMN Complex to Generate Spliceosomal SnRNPs. *Cell* 135, 497–509. doi:10.1016/j.cell.2008.09.020

Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer Statistics in China, 2015. *CA: A Cancer J. Clinicians* 66, 115–132. doi:10.3322/caac.21338

Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 88 (Suppl. l), S11. doi:10.1186/1752-0509-8-s4-s11

Clough, E., and Barrett, T. (2016). The Gene Expression Omnibus Database. *Methods Mol. Biol.* 1418, 93–110. doi:10.1007/978-1-4939-3578-9_5

Correa, B. R., De Araujo, P. R., Qiao, M., Burns, S. C., Chen, C., Schlegel, R., et al. (2016). Functional Genomics Analyses of RNA-Binding Proteins Reveal the Splicing Regulator SNRPB as an Oncogenic Candidate in Glioblastoma. *Genome Biol.* 17, 125. doi:10.1186/s13059-016-0990-4

Dai, X., Yu, L., Chen, X., and Zhang, J. (2021). SNRPD1 Confers Diagnostic and Therapeutic Values on Breast Cancers through Cell Cycle Regulation. *Cancer Cel Int.* 21, 229. doi:10.1186/s12935-021-01932-w

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6, l1. doi:10.1126/scisignal.2004088

Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., et al. (2001). Diversity of Gene Expression in Adenocarcinoma of the Lung. *Proc. Natl. Acad. Sci.* 98, 13784–13789. doi:10.1073/pnas.241500798

Győrffy, B., Surowiak, P., Budczies, J., and Lánczky, A. (2013). Online Survival Analysis Software to Assess the Prognostic Value of Biomarkers Using Transcriptomic Data in Non-small-cell Lung Cancer. *PLoS ONE* 8, e82241. doi:10.1371/journal.pone.0082241

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Hou, J., Aerts, J., Den Hamer, B., Van Ijcken, W., Den Bakker, M., Riegman, P., et al. (2010). Gene Expression-Based Classification of Non-small Cell Lung Carcinomas and Survival Prediction. *PLoS ONE* 5, e10312. doi:10.1371/journal.pone.0010312

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211

Jassem, E., Szymanowska, A., Siemińska, A., and Jassem, J. (2009). Smoking and Lung Cancer. *Pneumonol. Alergol. Pol.* 77, 469–473.

Jia, D., Wei, L., Guo, W., Zha, R., Bao, M., Chen, Z., et al. (2011). Genome-wide Copy Number Analyses Identified Novel Cancer Genes in Hepatocellular Carcinoma. *Hepatology* 54, 1227–1236. doi:10.1002/hep.24495

Koedoot, E., Van Steijn, E., Vermeer, M., González-Prieto, R., Vertegaal, A. C. O., Martens, J. W. M., et al. (2021). Splicing Factors Control Triple-Negative Breast Cancer Cell Mitosis through SUN2 Interaction and Sororin Intron Retention. *J. Exp. Clin. Cancer Res.* 40, 82. doi:10.1186/s13046-021-01863-4

Lan, Y., Lou, J., Hu, J., Yu, Z., Lyu, W., and Zhang, B. (2020). Downregulation of SNRPG Induces Cell Cycle Arrest and Sensitizes Human Glioblastoma Cells to Temozolomide by Targeting Myc through a P53-dependent Signaling Pathway. *Cancer Biol. Med.* 17, 112–131. doi:10.20892/j.issn.2095-3941.2019.0164

Liu, N., Chen, A., Feng, N., Liu, X., and Zhang, L. (2021). SNRPB Is a Mediator for Cellular Response to Cisplatin in Non-small-cell Lung Cancer. *Med. Oncol.* 38, 57. doi:10.1007/s12032-021-01502-0

Liu, N., Wu, Z., Chen, A., Wang, Y., Cai, D., Zheng, J., et al. (2019). SNRPB Promotes the Tumorigenic Potential of NSCLC in Part by Regulating RAB26. *Cell Death Dis.* 10, 667. doi:10.1038/s41419-019-1929-y

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Nagy, Á., Munkácsy, G., and Győrffy, B. (2021). Pancancer Survival Analysis of Cancer Hallmark Genes. *Sci. Rep.* 11, 6047. doi:10.1038/s41598-021-84787-5

Olst, E. S.-V., Blijlevens, M., De Menezes, R. X., Van Der Meulen-Muileman, I. H., Smit, E. F., and Van Beusechem, V. W. (2017). A Genome-wide siRNA Screen for Regulators of Tumor Suppressor P53 Activity in Human Non-small Cell Lung Cancer Cells Identifies Components of the RNA Splicing Machinery as Targets for Anticancer Treatment. *Mol. Oncol.* 11, 534–551. doi:10.1002/1878-0261.12052

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape Automation: Empowering Workflow-Based Network Analysis. *Genome Biol.* 20, 185. doi:10.1186/s13059-019-1758-4

Peng, N., Li, J., He, J., Shi, X., Huang, H., Mo, Y., et al. (2020). c-Myc-mediated SNRPB Upregulation Functions as an Oncogene in Hepatocellular Carcinoma. *Cell Biol. Int.* 44, 1103–1111. doi:10.1002/cbin.11307

Reck, M., and Rabe, K. F. (2017). Precision Diagnosis and Treatment for Advanced Non-small-cell Lung Cancer. *N. Engl. J. Med.* 377, 849–861. doi:10.1056/nejmra1703413

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., et al. (2004). ONCOMINE: a Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* 6, 1–6. doi:10.1016/s1476-5586(04)80047-2

Riemekasten, G., Marell, J., Trebeljahr, G., Klein, R., Hausdorf, G., Häupl, T., et al. (1998). A Novel Epitope on the C-Terminus of SmD1 Is Recognized by the Majority of Sera from Patients with Systemic Lupus Erythematosus. *J. Clin. Invest.* 102, 754–763. doi:10.1172/jci2749

Sun, X., Zhang, H., Luo, L., Zhong, K., Ma, Y., Fan, L., et al. (2016). Comparative Proteomic Profiling Identifies Potential Prognostic Factors for Human clear Cell Renal Cell Carcinoma. *Oncol. Rep.* 36, 3131–3138. doi:10.3892/or.2016.5159

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45, W98–W102. doi:10.1093/nar/gkx247

Tao, Y., Han, Y., Yu, L., Wang, Q., Leng, S. X., and Zhang, H. (2020). The Predicted Key Molecules, Functions, and Pathways that Bridge Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). *Front. Neurol.* 11, 233. doi:10.3389/fneur.2020.00233

Tapak, L., Saidijam, M., Sadeghifar, M., Poorolajal, J., and Mahjub, H. (2015). Competing Risks Data Analysis with High-Dimensional Covariates: an Application in Bladder Cancer. *Genomics, Proteomics & Bioinformatics* 13, 169–176. doi:10.1016/j.gpb.2015.04.001

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *Contemp. Oncol. (Pozn)* 19, A68–A77. doi:10.5114/wo.2014.47136

Valles, I., Pajares, M. J., Segura, V., Guruceaga, E., Gomez-Roman, J., Blanco, D., et al. (2012). Identification of Novel Deregulated RNA Metabolism-Related Genes in Non-small Cell Lung Cancer. *PLoS ONE* 7, e42086. doi:10.1371/journal.pone.0042086

Wachi, S., Yoneda, K., and Wu, R. (2005). Interactome-transcriptome Analysis Reveals the High Centrality of Genes Differentially Expressed in Lung Cancer Tissues. *Bioinformatics* 21, 4205–4208. doi:10.1093/bioinformatics/bti688

Yamagata, N., Shyr, Y., Yanagisawa, K., Edgerton, M., Dang, T. P., Gonzalez, A., et al. (2003). A Training-Testing Approach to the Molecular Classification of Resected Non-small Cell Lung Cancer. *Clin. Cancer Res.* 9, 4695–4704.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zhan, Y.-T., Li, L., Zeng, T.-T., Zhou, N.-N., Guan, X.-Y., and Li, Y. (2020). SNRPB-mediated RNA Splicing Drives Tumor Cell Proliferation and Stemness in Hepatocellular Carcinoma. *Aging* 13, 537–554. doi:10.18632/aging.202164

Zhang, Y., Foreman, O., Wigle, D. A., Kosari, F., Vasmatzis, G., Salisbury, J. L., et al. (2012). USP44 Regulates Centrosome Positioning to Prevent Aneuploidy and Suppress Tumorigenesis. *J. Clin. Invest.* 122, 4362–4374. doi:10.1172/jci63084

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10, 1523. doi:10.1038/s41467-019-09234-6

Zhu, L., Zhang, X., and Sun, Z. (2020). SNRPB Promotes Cervical Cancer Progression through Repressing P53 Expression. *Biomed. Pharmacother.* 125, 109948. doi:10.1016/j.biopha.2020.109948

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Li, Chen, Luo, Dai and Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fgene.2022.842351

ASF1B is a Promising Prognostic Biomarker and Correlates With Immunotherapy Efficacy in Hepatocellular Carcinoma

*Shirong Zhang[1†], Longwen Xu[1†], Jinteng Feng[1†], Deli Tan[1], Yue Zhu[1], Jia Hou[1], Wenyuan Li[1], Kejia Lv[1], Wenjuan Wang[1], Lili Jiang[1], Min Jiao[1] and Hui Guo[1,2,3]**

[1]Department of Medical Oncology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2]Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education of China, Xi'an Jiaotong University, Xi'an, China, [3]Bioinspired Engineering and Biomechanics Center (BEBC), Xi'an Jiaotong University, Xi'an, China

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Anna Lewinska,
University of Rzeszow, Poland
Qianqian Song,
Wake Forest School of Medicine, United States
Wen-Tsan Chang,
Kaohsiung Medical University Hospital, Taiwan

***Correspondence:**
Hui Guo
guohuihappy97@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 23 December 2021
Accepted: 22 February 2022
Published: 10 March 2022

Citation:
Zhang S, Xu L, Feng J, Tan D, Zhu Y, Hou J, Li W, Lv K, Wang W, Jiang L, Jiao M and Guo H (2022) ASF1B is a Promising Prognostic Biomarker and Correlates With Immunotherapy Efficacy in Hepatocellular Carcinoma. Front. Genet. 13:842351. doi: 10.3389/fgene.2022.842351

Background: Anti-silencing function 1B (ASF1B), a histone H3-H4 chaperone, is crucial for S-phase progression and cell proliferation. Recent studies have shown that ASF1B may be used as a new proliferation marker for cancer prognosis. However, the prognostic value and effect of ASF1B on tumor cells and the immune microenvironment in hepatocellular carcinoma (HCC) remain unclear.

Methods: We analyzed the expression of ASF1B and its prognostic value using The Cancer Genome Atlas (TCGA) database (as a training set) and other databases, and we validated the findings by immunohistochemistry in our clinical database, containing 141 HCC patients (as a validation set). Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were performed to probe the tumor-associated biological processes of ASF1B in HCC. The interrelationships between ASF1B expression and tumor immunological characteristics were analyzed by multiple databases. The Imvigor210 cohort was retrieved to assess the ability of ASF1B to predict immunotherapy efficacy.

Results: ASF1B was highly expressed in tumor tissue compared to paracancerous tissue. High ASF1B expression was associated with worse overall survival (OS) and progression-free survival (PFS) in the training set ($p = 0.005$, $p < 0.001$) and validation set ($p < 0.001$, $p < 0.001$). Multivariate analysis revealed that ASF1B was an independent prognostic factor associated with OS and PFS. GSEA and GSVA suggested that ASF1B was involved in tumor-associated biological processes, including the cell cycle, DNA replication, base excision repair, mismatch repair, RNA degradation, ubiquitin-mediated proteolysis, and nucleotide excision repair. Further analysis revealed that the levels of ASF1B were positively correlated with the immune cells infiltration of B cells, $CD8^+$ T cells, $CD4^+$ T cells, neutrophils, and dendritic cells. However, ASF1B was positively correlated with Treg cell infiltration and inhibitory immune checkpoints in exhausted T cells. Patients who received anti-PD-L1 immunotherapy with high ASF1B expression had a higher objective response.

Conclusion: The ASF1B level is an independent prognostic factor and may serve as a potential immunotherapeutic target.

Keywords: Hepatocellular carcinoma, ASF1B, prognosis, immune microenvironment, cancer immunotherapy

INTRODUCTION

Hepatocellular carcinoma (HCC), the major type of primary liver cancer, is estimated to be the fourth most common cause of cancer-related death, which exacts a heavy disease burden worldwide (Llovet et al., 2021). China is one of the highest burden areas for HCC due to the high prevalence of chronic hepatitis B virus infection, which is the most prominent risk factor for HCC development (Yu et al., 2014; Akinyemiju et al., 2017; Liu et al., 2018). Although advancements in the clinical management of HCC have improved patient survival, the prognosis of HCC remains poor due to the high rate of frequent recurrence and intrahepatic metastasis. Recently, many studies have demonstrated that immune checkpoint inhibitors (ICIs), including atezolizumab, nivolumab, and pembrolizumab, are revolutionizing cancer therapy in HCC by inducing durable antitumor responses and overall survival benefits (El-Khoueiry et al., 2017; Zhu et al., 2018; Finn et al., 2020; Yau et al., 2020). However, only a minority of patients achieve this transcendent, durable benefit from ICIs. Thus, identifying a biomarker to predict prognosis and response to ICIs for HCC treatment is urgently required.

The tumor immune microenvironment (TIME) is a crucial factor for the progression of HCC and for the response to immunotherapy (Cariani and Missale, 2019). Over the past decade, emerging evidences have revealed crosstalk between the tumor cell cycle and the TIME (Petroni et al., 2020). Disrupting cell cycle progression through targeting cell cycle regulators modulates the expression of checkpoint molecules and influences immune cell populations of the TIME, ultimately improving the efficacy of ICIs (Goel et al., 2017; Deng et al., 2018; Greten and Korangy, 2018; Zhou et al., 2018; Teh and Aplin, 2019). For example, programmed cell death-ligand 1 (PD-L1) expression fluctuates during the cell cycle, and inhibition of cyclin-dependent kinase 4 and 6 (CDK4/6) increases the level of PD-L1 through multiple pathways (Goel et al., 2017; Zhang H et al., 2018; Jin et al., 2019). In addition, cell cycle-related kinase inhibition reduces the accumulation of myeloid-derived suppressor cells (MDSCs) and improves antitumor immunity (Zhou et al., 2018). Inhibition of CDK4/6 markedly suppresses the proliferation of $CD4^+$ $FOXP3^+$ regulatory T cells (Tregs) and enhances $CD8^+$ T cell activity (Goel et al., 2017; Deng et al., 2018). CDK7 inhibition induces a significant increase in the percentage of total T cells, natural killer cells, dendritic cells, monocytes, and neutrophils (Zhang H. et al., 2020). Therefore, targeting cell cycle regulatory proteins may be a breakthrough point to sensitize tumors to immunotherapy.

In all eukaryotes, DNA and histones are precisely organized into chromatin, and assembly and disassembly processes are vital processes during the cell cycle. Histone chaperones are involved in all aspects of histone dynamics, particularly promoting specific chromatin assembly pathways throughout cellular life (De Koning et al., 2007; Eitoku et al., 2008). Anti-silencing function 1 (ASF1), which is the most conserved H3-H4 chaperone, is crucial for S-phase progression, and it has been implicated in gene replication, transcription, and DNA repair. ASF1 exists in two paralogs, termed ASF1A and ASF1B (Natsume et al., 2007). Although conservation core-binding domain for histones H3-H4 in these paralogs suggests common properties, in fact they are not functionally equivalent. ASF1A contributes mainly to DNA repair and cell senescence, while ASF1B is preferentially involved in cell cycle progression and cell proliferation (Corpet et al., 2011; Jiangqiao et al., 2019). In addition to affecting the tumor cell intrinsic features, it is reported that ASF1B can also have an impact on the tumor microenvironment by promoting the infiltration of immune cells (Zhan et al., 2021). It has also been reported that high ASF1B expression is closely related to poor outcomes of patients with renal cell cancer, cervical cancer, and breast cancer (Corpet et al., 2011; Han et al., 2018; Jiangqiao et al., 2019; Liu et al., 2020). Therefore, ASF1B is gaining attention as a new diagnostic and prognostic biomarker as well as a therapeutic target for these cancers. However, there are no comprehensive reports about the expression and prognostic value of ASF1B and its correlation with tumor immunity in HCC.

In the present study, we systematically and comprehensively analyzed the expression of ASF1B from online public databases and our medical center's databases and then assessed its correlation with clinicopathological factors and patients prognosis using data from The Cancer Genome Atlas database (TCGA) and the database from our center. We further explored the potential associations of ASF1B with tumor-infiltrating immune cells and the efficacy of immunotherapy using multiple online public databases.

MATERIALS AND METHODS

Patients and Sample Collection

In total, 141 patients were enrolled from January 2013 to December 2015 as a validation set to validate the findings from public databases. Sixty-one of these patients had paired tumors and adjacent tissue samples. All patients in this cohort underwent R0 resection, and the pathological diagnosis was confirmed as HCC identified by the pathology department at the First Affiliated Hospital of Xi'an Jiaotong University (Xi'an, China). The patients' primary characteristics are listed in **Table 1**. All patients were followed up until December 2020. The use of human tissues in this study was approved by the Research Ethics Committees.

Immunohistochemistry and Evaluation of Immunostaining

Immunohistochemistry was performed as described previously (Zhang et al., 2019). The ASF1B antibody (human, diluted 1:200, ab235358, Abcam, Cambridge, MA) was applied. The tumor cells in which nuclei were stained dark brown under light microscopy were considered positive. The staining intensity was evaluated with the following scoring system: 0 point represented no staining; 1 point represented weak staining intensity; 2 point represented moderate staining intensity; and 3 point represented strong staining intensity. Additionally, the percentage of stained

TABLE 1 | Demographics and clinicopathological characteristics of the HCC patients in TCGA database and our Center.

Features	Training set (*n* = 371)	Validation set (*n* = 141)
Age [years, median (range)]	59.4 (16–90)	51.0 (27–80)
Sex (Female/Male) (%)	121/250 (32.6/67.4)	22/119 (15.6/84.4)
HBV (Positive/Negative) (%)	—	123/18 (87.2/12.8)
AFP (ng/ml, <400/≥400/unknown) (%)	213/65/93 (57.4/17.5/25.1)	90/51 (63.8/36.2)
Tumor number (Single/Multinodular) (%)	—	98/43 (69.5/30.5)
Tumor differentiation (Edmondson-Steiner grade) (I/II/III/IV/unknown) (%)	55/177/122/12/5 (14.8/47.7/32.9/3.2/1.3)	7/91/43/0/0 (5.0/64.5/30.5/0/0)
Tumor diameter (cm, mean ± SD)	—	7.27 ± 7.26
Vascular invasion (yes/no/unknown) (%)	206/109/56 (55.5/29.4/15.1)	93/48 (66.0/34.0)
T classification (T1/T2/T3/T4/unknown) (%)	181/94/80/13/3 (48.8/25.3/21.6/3.5/0.8)	63/56/19/3 (44.0/39.7/13.5/2.1)
N classification (N0/N1/unknown) (%)	252/4/115 (67.9/1.1/31.0)	135/6 (95.7/4.3)
M classification (M0/M1/unknown) (%)	266/4/101 (71.7/1.1/27.2)	—
TNM stage (I/II/III/IV/unknown) (%)	171/86/85/5/24 (46.1/23.2/22.9/1.3/6.5)	60/55/21/5 (42.6/39.0/14.9/3.5)
ASF1B expression (High/Low) (%)	185/186 (49.9/50.1)	45/96 (31.9/68.1)

HBV, hepatitis B virus; AFP, alpha fetoprotein; TNM, tumor node metastasis.

tumor cells was assessed as follows: 0 point indicated 0%; 1 point indicated less than 25%; 2 point indicated 25–50%; and 3 point indicated more than 50%. The final score was equal to the multiplication of the above two scores. A score of 0–3 point represented a low expression level of ASF1B, while a score greater than 3 point indicated high expression (Liu et al., 2020).

Cell Culture, RNA Interference and Transfection

Hep3B and Huh7 were a gift from JuSeog Lee (MD Anderson Cancer Center, Houston, TX). All cells were maintained in high glucose Dulbecco's modified Eagle's medium (Hyclone, Logan, UT) with 10% FBS (Hyclone, Logan, UT) and 100 U/ml penicillin and streptomycin. Cell lines were incubated with 95% humidified air and 5% CO_2 at 37°C. The human ASF1B-target small interfering (si)RNAs (siRNA1: 5′-3′ CAGGCGGGAAUGUUA GUUATT, 3′-5′ UAACUAACAUUCCCGCCUGTT; siRNA2: 5′-3′ CAUGUUGCCUUUCCUGUCATT, 3′-5′ UGACAG GAAAGGCAACAUGTT) were applied to construct ASF1B knockdown HCC lines. In this process Lipofectamine 2000 (Invitrogen, MA) was used according to it's protocol.

MTT Analysis and Colony Formation Assay

HCC cells (5,000 per well) were seeded into 96-well plates. Relative cell numbers were quantified per day using the 3-(4,5-dimethylthia-zol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay. Absorbance was measured at 492 nm. For colony formation assays, HCC cells were seeded in the 6-well plates at a density of 500/well. After 14 days, Each group were fixed in 4% paraformaldehyde for 20 min and stained with 0.1% crystal violet solution for imaging and counting.

Western Blotting Analysis

At first, protein was extracted from LO2 normal liver cells and 5 HCC cell lines (Hep3B, SMMC7721, MHCC97L, MHCC97H, and Huh7). Then Western blotting was operated in accordance with standard protocols. The following primary antibodies were applied: anti-β-actin antibody (human, diluted 1:10,000, 60004-1-Ig, Proteintech Group, United States) and anti-ASF1B antibody (diluted 1:1,000). The ASF1B detection was repeated at least three times. The band intensities were measured using ImageJ (Bethesda, MD, United States).

Collection of Sample Information From TCGA

Clinical information and transcriptomic data of 375 HCC patients and 49 adjacent cancer samples were downloaded from TCGA data portal (https://portal.gdc.cancer.gov/). Cases with insufficient or missing data were filtered out, and 371 patients data were adopted as a training set for survival analysis. We compared the overall survival (OS) and progression-free survival (PFS) of HCC patients divided by the ASF1B median expression value. The clinical characteristics of the included patients were summarized in **Table 1**.

HCCDB Data Analysis

The Integrative Molecular Database of Hepatocellular Carcinoma (HCCDB) database (http://lifeome.net/database/hccdb) was used to explore ASF1B mRNA expression in HCC and adjacent noncarcinoma tissues. The HCCDB database contains 15 public datasets that cover 3,917 samples (Lian et al., 2018).

Heterogeneity

Profile of somatic mutation data of the TCGA-LIHC cohort were obtained from the GDC portal on February 13, 2022. With the advantage of R "maftools" package, we visualized the MAF files of simple nucleotide variation which processed by the workflow type of varScan2 variant aggregation and masking. The tumor mutation burden (TMB) and the mutant-allele tumor heterogeneity (MATH) score of tumor samples in the TCGA-LIHC dataset were also computed via the "maftools" package.

Gene Set Enrichment Analysis and Gene Set Variation Analysis

Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were used to investigate the mechanisms of ASF1B in HCC (Hu et al., 2021; Lin et al., 2021). In the present study, the enrichment scores (ES) of "c2.cp.kegg.v6.2.symbols.gmt" gene sets from the Molecular

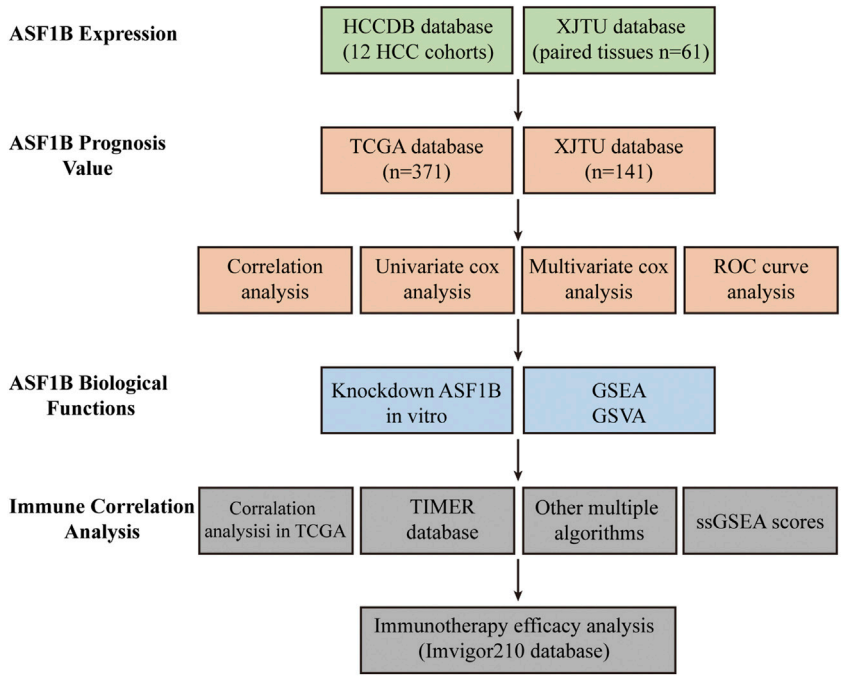


FIGURE 1 | Flow chart of the study design.

Signatures Database (MSigDB) in each group were counted by GSEA software (4.1.0) and reflected the degree to which a given gene set was represented in a ranked list of genes. A nominal p value of <0.05 and a FDR q-value of <0.25 were used as the cutoff criteria. In addition, GSVA was performed to further analyze the difference in pathways between the subtypes of ASF1B using the GSVA package (Subramanian et al., 2005; Hu et al., 2021).

Immune Cell Infiltration Analysis

The tumor immune estimation resource (TIMER) (https://cistrome.shinyapps.io/timer/) was used to analyze immune infiltration in HCC (Li et al., 2017). We used this tool to explore the correlation between ASF1B and the abundances of six immune infiltrates (B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells) and gene markers of infiltrating immune cells in HCC. Furthermore, other algorithms (including XCELL, QUANTISEQ, EPIC, CIBERSORT-ABS and CIBERSORT) were utilized to quantify the infiltrating immune cells and to verify the TIMER results.

Single-Sample Gene Set Enrichment Analysis Analysis

The single-sample gene set enrichment analysis (ssGSEA) was utilized to calculated the infiltrating score of 16 immune cells and the activity of 13 immune-related pathways in the "gsva" R package (Lei et al., 2020), which aided in determining the activity of immune cells and immune pathways of each sample.

Statistical Analyses

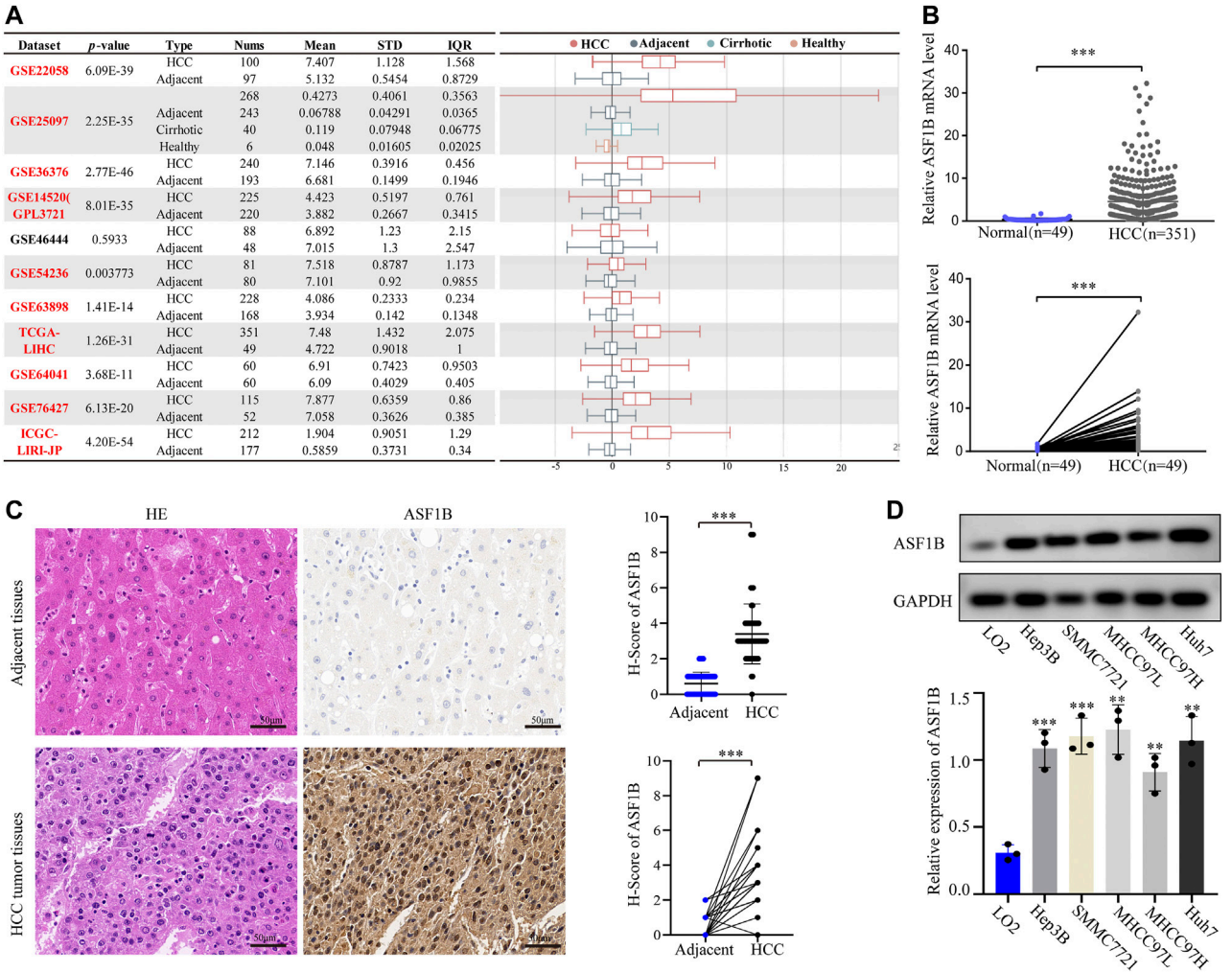
In this study, unpaired t test and one-way analysis of variance were used to calculate the comparison of continuous variables. The chi-square test was used to process the categorical variables. Kaplan-Meier survival curves was used to display the OS and PFS. Univariate and multivariate regression analyses were used to identify independent prognostic factors. The correlation between genes was detected with Pearson analysis. All the tests were two-sided, and $p < 0.05$ was considered statistically significant. SPSS 24.0 (SPSS Inc., Chicago, IL, United States) was used to perform the statistical analyses.

RESULTS

Anti-Silencing Function 1B has Higher Expression in Hepatocellular Carcinoma Tissue Than in Adjacent Noncancerous Tissues

A flowchart of this study is illustrated in **Figure 1**. An analysis of 12 HCC cohorts using the HCCDB database showed that ASF1B mRNA expression was remarkably increased in HCC relative to adjacent noncancerous tissues or cirrhotic tissues in 11 HCC cohorts (**Figure 2A**). Consistently, comparison of ASF1B mRNA expression in cancer and matched paracancerous tissues from TCGA database demonstrated the upregulation of ASF1B in cancer (**Figure 2B**).

To validate the above findings, HCC cancer tissues and corresponding adjacent tissues from 61 patients were tested by



Dataset	*p*-value	Type	Nums	Mean	STD	IQR
GSE22058	6.09E-39	HCC	100	7.407	1.128	1.568
		Adjacent	97	5.132	0.5454	0.8729
GSE25097	2.25E-35		268	0.4273	0.4061	0.3563
		Adjacent	243	0.06788	0.04291	0.0365
		Cirrhotic	40	0.119	0.07948	0.06775
		Healthy	6	0.048	0.01605	0.02025
GSE36376	2.77E-46	HCC	240	7.146	0.3916	0.456
		Adjacent	193	6.681	0.1499	0.1946
GSE14520(GPL3721	8.01E-35	HCC	225	4.423	0.5197	0.761
		Adjacent	220	3.882	0.2667	0.3415
GSE46444	0.5933	HCC	88	6.892	1.23	2.15
		Adjacent	48	7.015	1.3	2.547
GSE54236	0.003773	HCC	81	7.518	0.8787	1.173
		Adjacent	80	7.101	0.92	0.9855
GSE63898	1.41E-14	HCC	228	4.086	0.2333	0.234
		Adjacent	168	3.934	0.142	0.1348
TCGA-LIHC	1.26E-31	HCC	351	7.48	1.432	2.075
		Adjacent	49	4.722	0.9018	1
GSE64041	3.68E-11	HCC	60	6.91	0.7423	0.9503
		Adjacent	60	6.09	0.4029	0.405
GSE76427	6.13E-20	HCC	115	7.877	0.6359	0.86
		Adjacent	52	7.058	0.3626	0.385
ICGC-LIRI-JP	4.20E-54	HCC	212	1.904	0.9051	1.29
		Adjacent	177	0.5859	0.3731	0.34

FIGURE 2 | ASF1B mRNA and protein expression in HCC. **(A)** Chart and plot of ASF1B expression in HCC and matched adjacent noncancerous tissues based on HCCDB database; **(B)** Significant differences of ASF1B mRNA level in HCC and the adjacent noncancerous tissues in TCGA database; **(C)** HE images of HCC tumor tissues and adjacent noncancerous tissues (left); IHC images and H-score of the ASF1B proteins expression in HCC tumor tissues and adjacent noncancerous tissues (right). Magnification = ×400. The positive staining appears brown; **(D)** The protein expression of ASF1B in normal liver cell and HCC cells by western blot. $^{**}p < 0.01$; $^{***}p < 0.001$.

IHC. The protein expression of ASF1B in cancer tissues was higher than that in noncancerous tissues (**Figure 2C**). Western blotting was then performed to detect ASF1B protein expression in one normal liver cell line (LO2) and five HCC cancer cell lines (Hep3B, SMMC7721, MHCC97L, MHCC97H, and Huh7). Compared to normal hepatocytes, ASF1B was remarkably increased in the HCC cell lines (**Figure 2D**). Together, these data illustrated that ASF1B has higher expression in HCC tissue than in adjacent noncancerous tissues.

High Anti-Silencing Function 1B Expression is Associated With Poor Prognosis in Hepatocellular Carcinoma

To explore the prognostic value of ASF1B in HCC, we enrolled two cohorts. The training set was obtained from TCGA database and contained 371 individuals, and the validation set contained 141 HCC patients from our center who had pathology and complete follow-up information. The relationships between ASF1B expression and the clinical features of HCC patients are listed in **Table 2** and **Supplementary Figure S1**. In the training set, a high ASF1B mRNA level was correlated with older patients ($p < 0.001$), males ($p = 0.018$), poorer differentiation ($p < 0.001$), higher levels of AFP ($p < 0.001$), more advanced T stage ($p = 0.001$), and TNM stage ($p = 0.001$). In the validation set, a high ASF1B protein level was correlated with poorer differentiation ($p = 0.038$), more advanced T stage ($p = 0.013$), and TNM stage ($p = 0.002$). The Kaplan-Meier survival curves revealed that patients with high ASF1B expression had poorer OS and PFS than patients with low ASF1B expression in the training set ($p = 0.005$, $p < 0.001$; **Figure 3A**) and the validation set ($p < 0.001$, $p < 0.001$; **Figures 4A,B**).

TABLE 2 | The relationships between ASF1B expression and patients' clinical features in TCGA database and our Center.

Features	Training set (*n* = 371)			Validation set (*n* = 141)		
	Low	High	*p* value	Low	High	*p* value
Age			**<0.001**			0.159
≥65 years	93	56		20	5	
<65 years	93	128		76	40	
Sex			**0.018**			0.992
Male	136	114		81	38	
Female	50	71		15	7	
HBV						0.497
Positive	—	—	—	85	38	
Negative	—	—	—	11	7	
AFP (ng/ml)			**<0.001**			0.917
≥400	20	45		35	16	
<400	122	91		61	29	
Tumor number						0.616
Single	—	—	—	68	30	
Multinodular	—	—	—	28	15	
Tumor differentiation			**<0.001**			**0.038**
I + II	139	93		72	26	
III + IV	44	90		24	19	
Vascular invasion			0.260			0.903
Yes	52	57		33	15	
No	112	94		63	30	
T classification			**0.001**			**0.013**
T1 + T2	150	125		86	33	
T3 + T4	33	60		10	22	
N classification			0.624			0.384
N0	121	131		93	42	
N1	1	3		3	3	
TNM stage			**0.001**			**0.002**
I + II	142	115		85	30	
III + IV	31	59		11	15	

HBV, hepatitis B virus; AFP, alpha fetoprotein; TNM, tumor node metastasis.The bold values means p < *0.05.*

Anti-Silencing Function 1B Expression is an Independent Prognostic Factor for Hepatocellular Carcinoma Patients

Besides, univariate and multivariate Cox regression analyses were performed to investigate whether ASF1B has a clinically independent prognostic value. Univariate Cox regression analysis revealed the association between OS and T stage ($p < 0.001$), M stage ($p = 0.014$), TNM stage ($p < 0.001$), and ASF1B expression ($p < 0.001$) as well as the association between PFS and T stage ($p < 0.001$), TNM stage ($p < 0.001$), and ASF1B expression ($p < 0.001$) in the training set (**Figure 3B**). In the validation set, more clinical factors were included for analysis. The factors significantly associated with OS were tumor size (5 cm) ($p = 0.014$), tumor number ($p = 0.024$), vascular invasion ($p = 0.018$), T stage ($p < 0.001$), N stage ($p = 0.033$), TNM stage ($p < 0.001$), and ASF1B expression ($p < 0.001$). The factors significantly associated with PFS were tumor size (5 cm) ($p = 0.001$), vascular invasion ($p = 0.046$), T stage ($p < 0.001$), TNM stage ($p < 0.001$), and ASF1B expression ($p < 0.001$) in the validation set (**Figure 4C**).

The variables demonstrating significance for the prognosis of HCC patients were included in the multivariate analysis. The analysis showed that ASF1B expression was an independent prognostic factor that was associated with OS ($p = 0.003$, HR = 1.428) and PFS ($p = 0.001$, HR = 1.350) in the training set. In addition, the validation set verified that ASF1B expression was indeed an independent risk factor for patient prognosis (OS: $p = 0.001$, HR = 1.993; PFS: $p = 0.008$, HR = 1.869) (**Figures 3C**, **4D**). Further, the receiver operating characteristic (ROC) curve showed that ASF1b expression had good predictive efficacy in the training set and the validation set (OS: 0.684, 0.697; PFS: 0.673, 0.703) (**Supplementary Figure S2**). Thus, these findings suggested that ASF1B may serve as a valuable predictive factor for HCC patients.

Analysis of the Biological Functions of Anti-Silencing Function 1B in Hepatocellular Carcinoma

Previous studies have reported that ASF1B functions as an oncogene to promote tumor growth by participating in the cell cycle (Corpet et al., 2011; Han et al., 2018; Jiangqiao et al., 2019; Liu et al., 2020). Given the clinical significance of ASF1B, we further analyzed the effects of ASF1B on the biological behaviors of the HCC cells. HCC cell lines with downregulation of ASF1B were constructed (**Figure 5A**). Inhibition of ASF1B significantly inhibited the proliferation and colony formation of Hep3B and Huh7 (**Figures 5B,C**). Then GSEA and GSVA were conducted to explore the biological functions of ASF1B in HCC. GSEA suggested that the ASF1B high-expression phenotype was mainly involved in the cell cycle in HCC. Apart from this, ASF1B was also related to DNA replication, the spliceosome, base excision repair, oocyte meiosis, homologous recombination, mismatch repair, RNA degradation, ubiquitin-mediated proteolysis, and nucleotide excision repair (**Table 3**). In addition, GSVA confirmed that cell cycle pathways were significantly differentially expressed between the high ASF1B group and the low ASF1B group in HCC. The heatmap shown in **Figure 5D** displays the differential expression of specific signaling pathways. Overall, these results demonstrated that ASF1B is mainly correlated with dysregulation of the cell cycle process, which may promote proliferation and be involved in the poor prognosis of HCC patients.

Gene Mutation in Anti-Silencing Function 1B-High and Anti-Silencing Function 1B-Low Expression Group

To explore the difference of genomics between the ASF1B-highand ASF1B-low subsets, we examined somatic mutation from the TCGA-LIHC datasets. **Figure 6A** showed that the top-5 highest mutation prevalence genes in the ASF1B-high group were TP53 (32%), TTN (21%), MUC16 (17%), CTNNB1 (16%), PCLO (9%), whereas those in the ASF1B-low group (**Figure 6B**) were CTNNB1 (32%), TTN (22%), MUC16 (12%), PCLO (11%), and TP53 (9%). We identified two genes (TP53, RB1) were highly mutated in the ASF1B-high group than in ASF1B-low group, while CTNNB1 exhibited higher mutation frequency in ASF1B-low group (**Figure 6C**). Although the TMB was not significantly different between ASF1B-low and ASF1B-high groups, patients from high-ASF1B group possessed

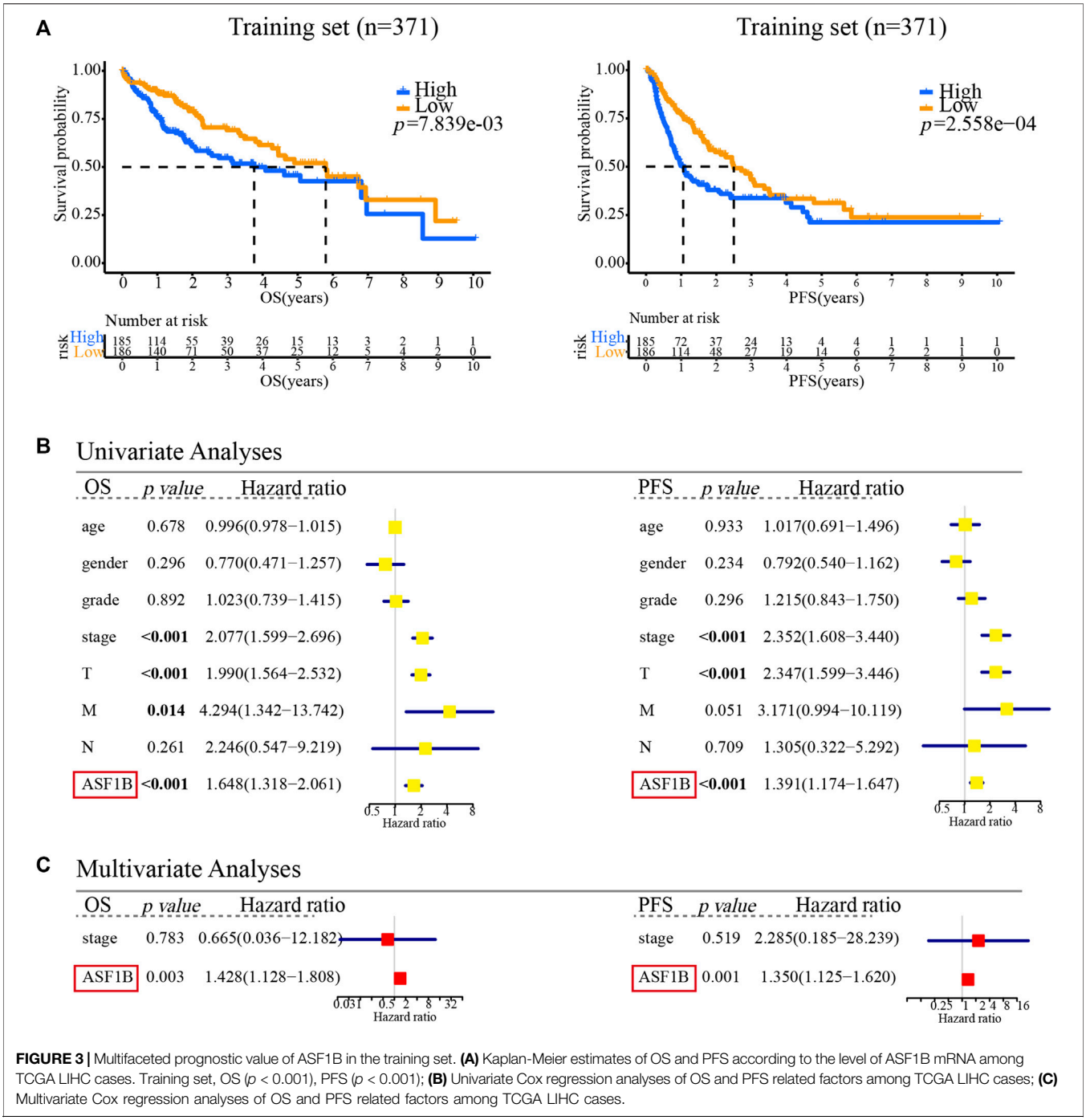


FIGURE 3 | Multifaceted prognostic value of ASF1B in the training set. **(A)** Kaplan-Meier estimates of OS and PFS according to the level of ASF1B mRNA among TCGA LIHC cases. Training set, OS ($p < 0.001$), PFS ($p < 0.001$); **(B)** Univariate Cox regression analyses of OS and PFS related factors among TCGA LIHC cases; **(C)** Multivariate Cox regression analyses of OS and PFS related factors among TCGA LIHC cases.

obviously elevated MATH scores, suggesting a higher level of tumor heterogeneity in this group (**Figure 6D**, $p = 0.025$).

Correlation Between Anti-Silencing Function 1B Expression and Immune Cell Infiltration in Hepatocellular Carcinoma

Emerging preclinical and clinical evidence has revealed that blocking the tumor cell cycle not only inhibits the proliferation of cancer cells but also mediates a wide range of immunomodulatory effects that involve both malignant and nonmalignant components of the TIME, thereby suggesting novel immunotherapeutic avenues (Petroni et al., 2020). Thus, we explored the correlation between ASF1B expression and immune cell infiltration. First, we examined the relationship of ASF1B expression with 15 core cell-cycle modulators and 10 immune checkpoints. According to the Pearson correlation analysis, the ASF1B expression level was significantly

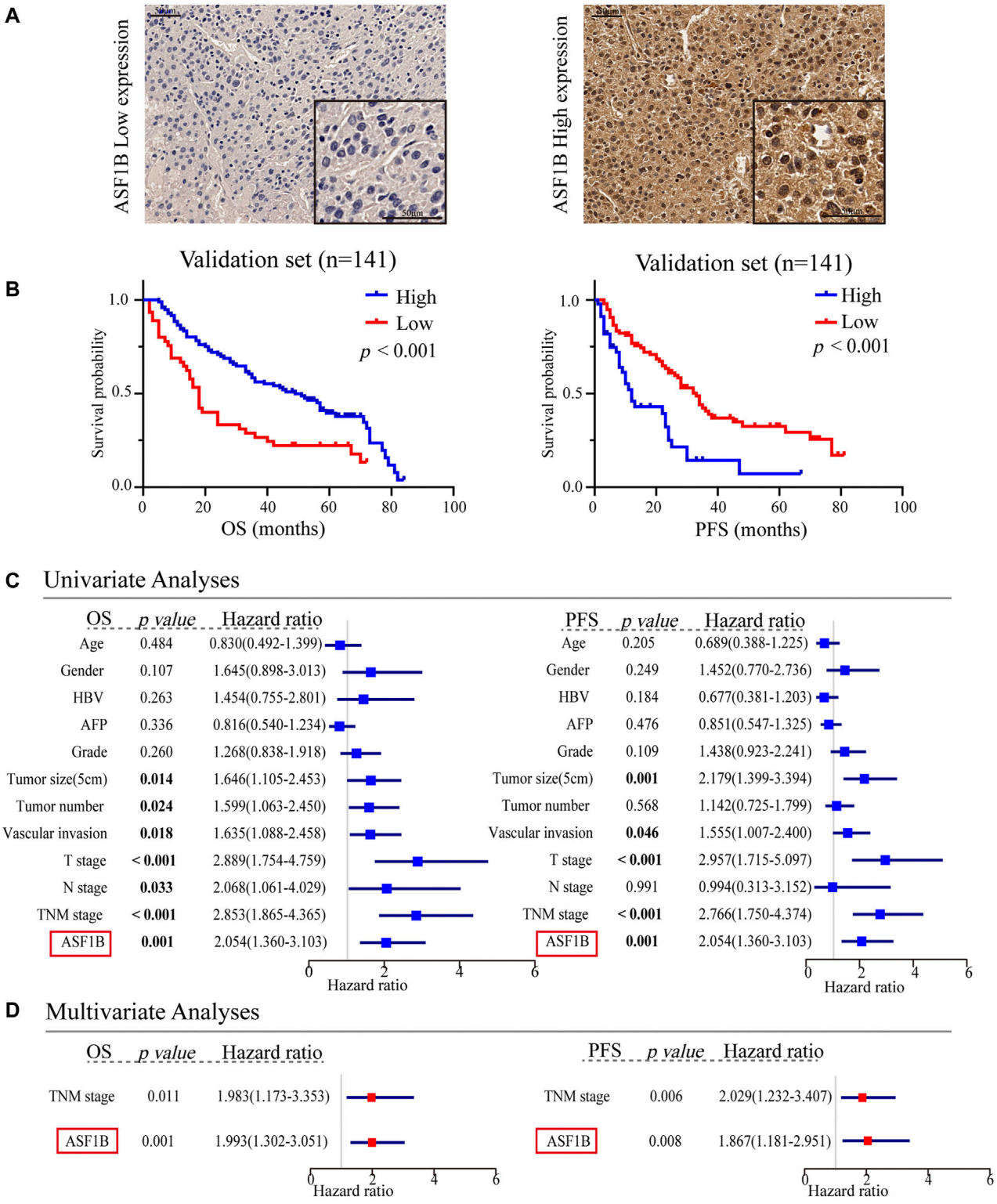


FIGURE 4 | Multifaceted prognostic value of ASF1B in the validation set. **(A)** Representative microphotographs of ASF1B staining in high and low expression groups; **(B)** Kaplan-Meier estimates of OS and PFS according to the level of ASF1B protein expression through IHC in the patients with HCC in our center. Validation set, OS ($p < 0.001$), PFS ($p < 0.001$); **(C)** Univariate Cox regression analyses of OS and PFS related factors among our center HCC cases; **(D)** Multivariate Cox regression analyses of OS and PFS related factors among our center HCC cases.

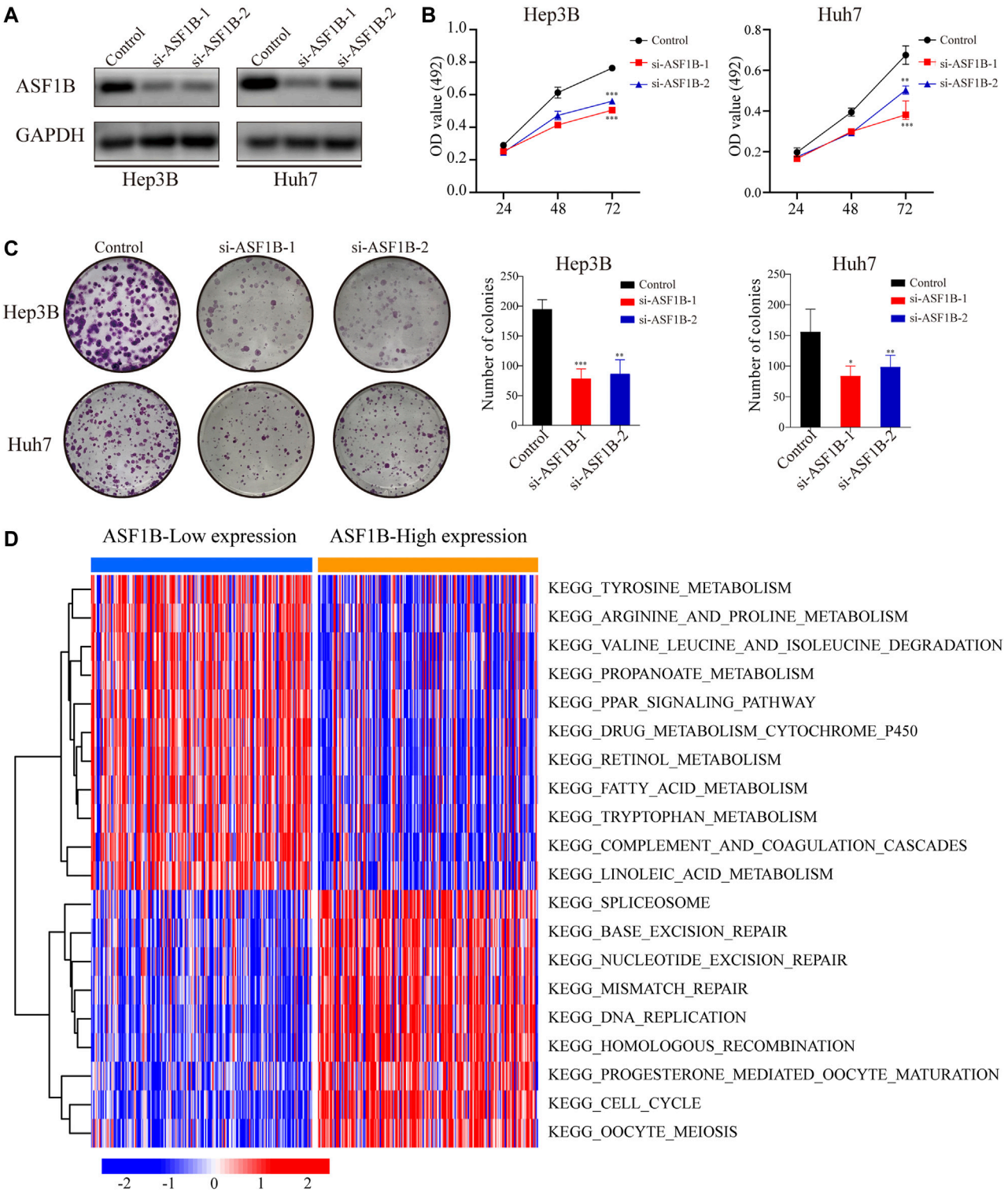


FIGURE 5 | Downregulation of ASF1B inhibits the proliferation of HCC cells. **(A)** The protein levels of ASF1B were detected using Western blotting after transfection of two independent sets of ASF1B-targeted siRNAs in Hep3B and Huh7; The proliferation of HCC cells was detected by **(B)** MTT assay and **(C)** the colony formation assay; **(D)** Functional enrichment analysis of ASF1B-correlated genes in HCC through GSVA. The red node represents up-regulation and the blue node represents down-regulation. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 3 | Gene sets enriched in the ASF1B high-expression phenotype in HCC.

Gene set name	NES	NOM *p*-value	FDR q-value
KEGG_CELL_CYCLE	2.2	<0.001	0.004
KEGG_SPLICEOSOME	2.16	<0.001	0.005
KEGG_BASE_EXCISION_REPAIR	2.1	<0.001	0.007
KEGG_DNA_REPLICATION	2.04	<0.001	0.013
KEGG_OOCYTE_MEIOSIS	2.01	<0.001	0.017
KEGG_HOMOLOGOUS_RECOMBINATION	1.97	<0.001	0.02
KEGG_MISMATCH_REPAIR	1.96	<0.001	0.024
KEGG_RNA_DEGRADATION	1.91	<0.001	0.035
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	1.89	<0.001	0.041
KEGG_NUCLEOTIDE_EXCISION_REPAIR	1.89	<0.001	0.038
KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	1.88	<0.001	0.047

NES, Normalized Enrichment Score; FDR, false discovery rate.

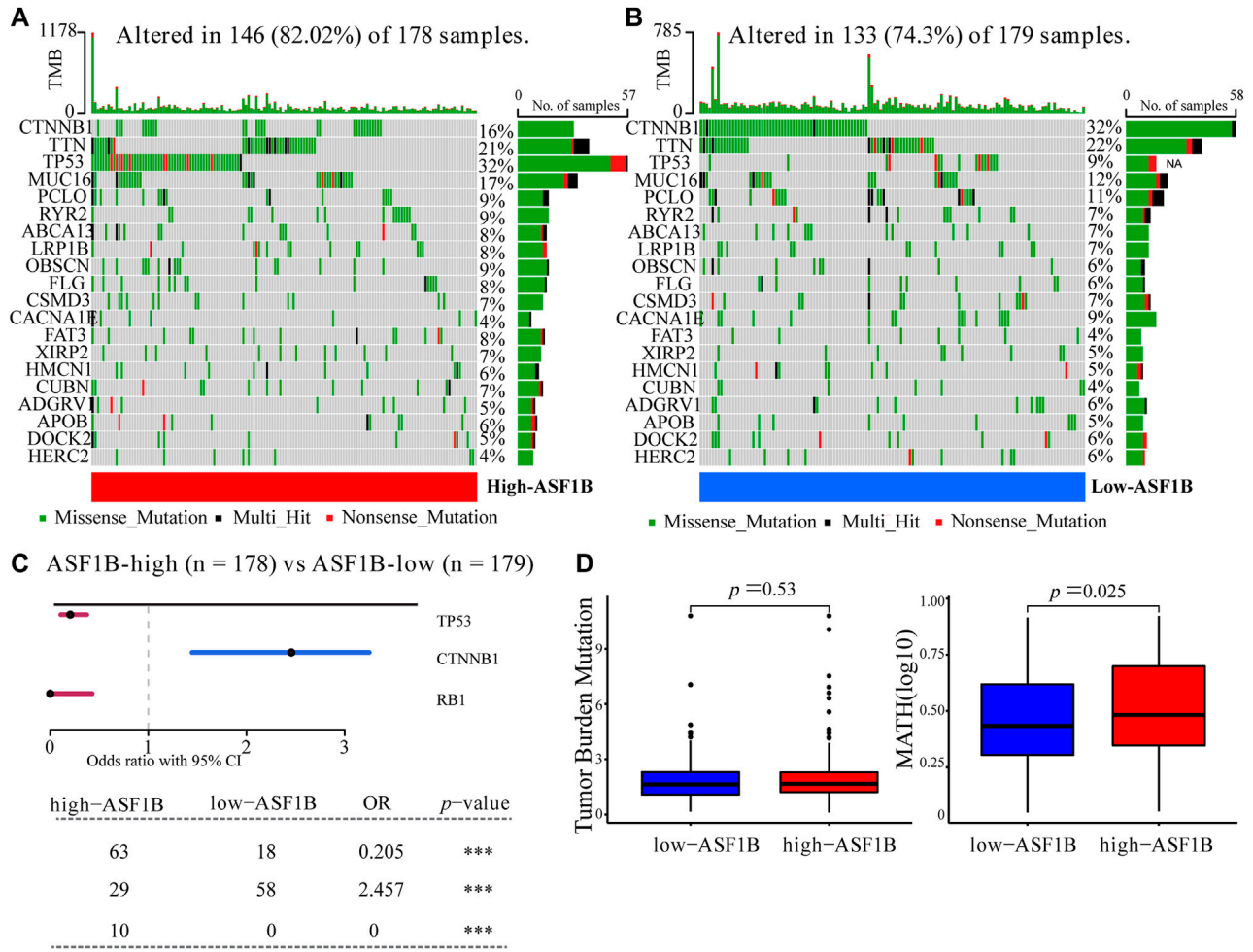


FIGURE 6 | Somatic mutation between ASF1B-high and ASF1B-low expression group. Oncoplots displayed the top-20 mutated genes in the **(A)** high-ASF1B and **(B)** low-ASF1B group of the TCGA-LIHC dataset. **(C)** Forest plot revealed three mutated genes with statistic significant between the ASF1B-high and ASF1B-low group. **(D)** The comparison of TMB (left) and MATH (right) scores in the ASF1B-high and ASF1B-low group. ***$p < 0.001$.

correlated with the expression of cell-cycle modulators, including CDK1, CDK2, CDK4, CDK6, CCND1 and so on. ASF1B expression was also significantly associated with immune checkpoints expression such as CD274, CTLA4 and PDCD1, etc. Meantime, cell cycle regulatory molecules were positively related to immune checkpoint molecules, further suggesting the

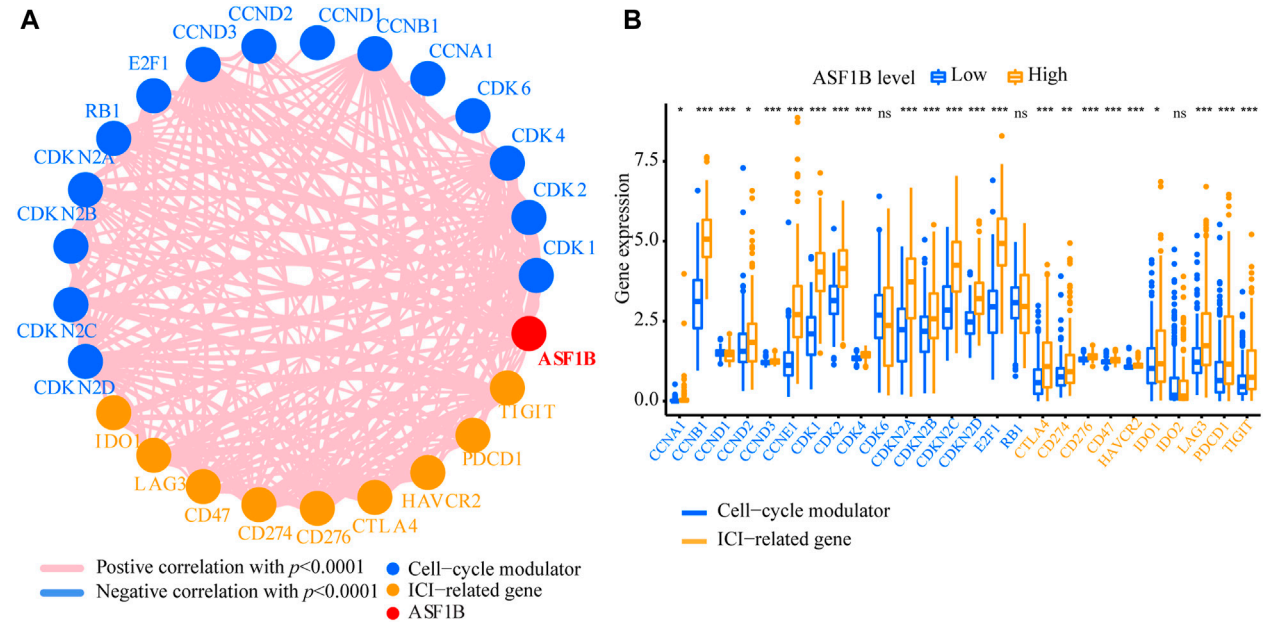


FIGURE 7 | The relationship of cell-cycle modulators and immune checkpoints between ASF1B-high and ASF1B-low groups. **(A)** Correlation of ASF1B, cell-cycle modulators, and immune checkpoints. **(B)** The comparison of cell-cycle modulators and immune checkpoints between ASF1B-high and ASF1B-low groups. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

close link between cell cycle and tumor immunity (**Figure 7A**). Additionally, results of Wilcoxon test showed that 13 cell-cycle modulators and nine immune checkpoints were significantly increased in the ASF1B-high group (**Figure 7B**). Above results validated our hypothesis that ASF1B, as an important cell cycle regulator, may affect the HCC tumor immunity.

Next, we explored the correlation between ASF1B expression and immune cell infiltration. In the TIMER database, the ASF1B expression level was significantly and positively correlated with immune cell infiltration, including $CD4^+$ T cells, $CD8^+$ T cells, B cells, macrophages, neutrophils, and dendritic cells (DCs) in HCC (**Supplementary Figure S3**). The xCell database analysis showed that ASF1B was positively related to $CD4^+$ Th1 cells ($r = 0.332$, $p < 0.001$), $CD4^+$ Th2 cells ($r = 0.642$, $p < 0.001$), and NK cells ($r = 0.352$, $p < 0.001$) infiltration. The QUANTISEQ results indicated that ASF1B was positively related to B cells ($r = 0.311$, $p < 0.001$), $CD8^+$ T cells ($r = 0.331$, $p < 0.001$) and Treg cells ($r = 0.305$, $p < 0.001$) infiltration (**Figure 8A**; **Supplementary Table S1**). However, there were significant differences in the infiltration degrees of B cells, dendritic cells, neutrophils, macrophages, and Thf cells, but not T cells, between the high and low ASF1B expression groups according to ssGSEA (**Figure 8B**). Compared to the low expression group, the related functions or pathways of type I IFN response and type II IFN response were lower in the high ASF1B expression group, while the related functions or pathways of the MHC class I were higher in the high expression group. These results suggested that ASF1B is closely associated with the degree of immune cell infiltration in HCC.

To validate the above results, the correlation between ASF1B and the immune marker sets of immune cells in HCC was explored using the TIMER database. The correlation was adjusted for tumor purity, which influences the immune infiltration analysis. As shown in **Figure 9** and **Table 4**, ASF1B expression was positively correlated with most of the immune marker sets. Interestingly, ASF1B was significantly positively correlated with inhibitory immune checkpoints in exhausted T cells, including PD-1 ($r = 0.428$, $p = 9.06e\text{-}17$), CTLA-4 ($r = 0.421$, $p = 3.09e\text{-}16$), LAG3 ($r = 0.388$, $p = 7.33e\text{-}14$), TIM-3 ($r = 0.421$, $p = 3.09e\text{-}16$), and TIGIT ($r = 0.406$, $p = 4.34e\text{-}15$). In addition, ASF1B also correlated with the CCR8, STAT5b, and TGFB1 gene markers in Tregs as well as the CD11b and CD15 gene markers in neutrophils. The above results suggested that ASF1B may indicate an immunosuppressive tumor microenvironment with increasing immune suppressive cells and dysfunctional and exhausted T cells even though it may cause increased T cell and other immune cells infiltration, which could influence the patient outcome in HCC. The hypothesis figure for the correlation between ASF1B expression and immune cell distribution is illustrated in **Supplementary Figure S4**.

Anti-Silencing Function 1B Predicts Immunotherapy Efficacy

Because ASF1B was associated with immune cell infiltration in tumors and positively correlated with immune checkpoint coinhibitory molecules, we further explored the relationship between ASF1B and the efficacy of immunotherapy, and we also identified the value of ASF1B as a predictive marker. A transcriptome dataset (Imvigor210) of the treatment response data of patients who underwent anti-PD-L1 immunotherapy

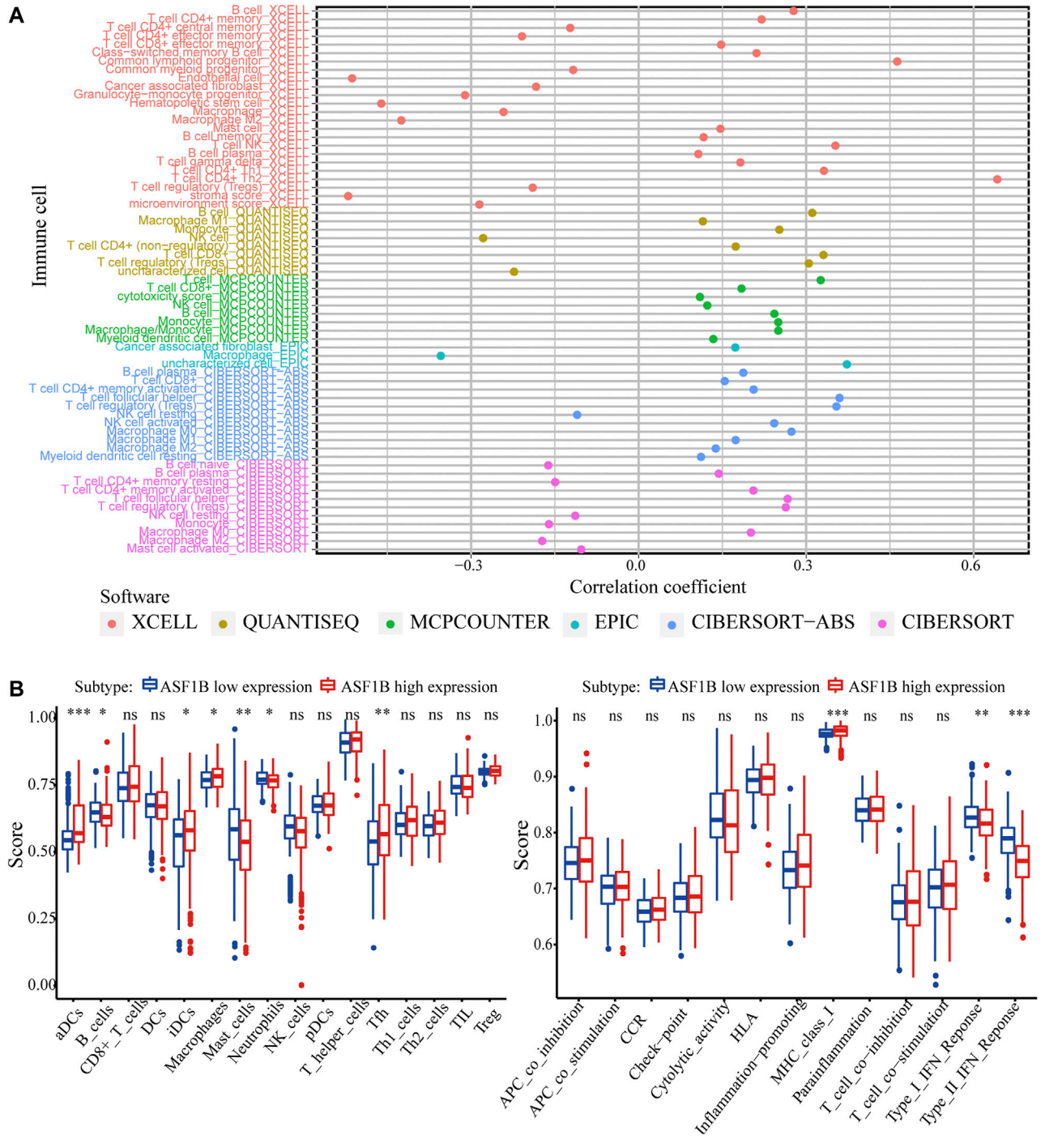


FIGURE 8 | ASF1B expression is related to immune cell infiltration. **(A)** The correlation between immune cell infiltration and ASF1B expression in HCC; **(B)** Comparison of the ssGSEA scores between high and low expression groups in the TCGA cohort. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

was retrieved to assess the ability of ASF1B to predict immunotherapy efficacy (Mariathasan et al., 2018). Kaplan-Meier analysis showed that although there was no significant difference, patients with high ASF1B had a relatively better outcome than patients with low ASF1B ($p = 0.076$) (**Figure 10A**). A higher objective response was observed in patients with high ASF1B expression than in patients with low ASF1B expression (34 vs. 15%, $p < 0.001$) (**Figure 10B**). These results suggested that ASF1B expression identifies patients who will respond to immunotherapy.

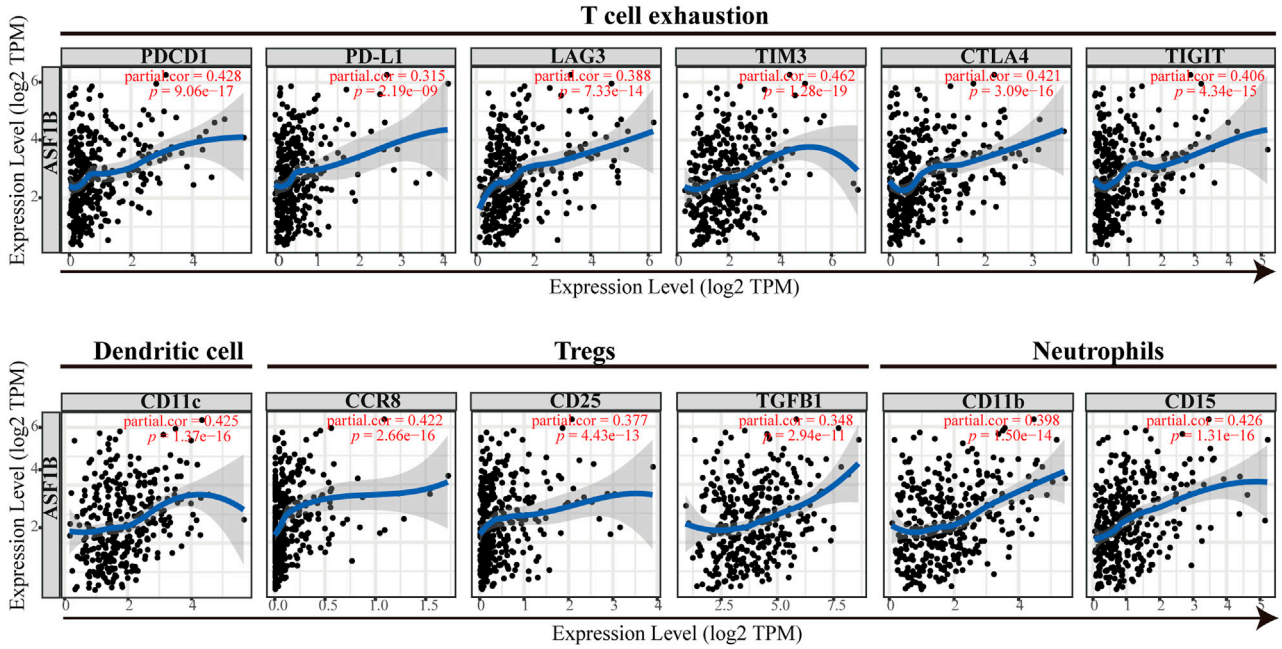


FIGURE 9 | The correlation between the immune marker sets of immune cells and ASF1B expression in HCC.

DISCUSSION

In the present study, by incorporating data from a public databases and our tumor center, we found that ASF1B expression was higher in tumor tissue compared to paracancerous normal tissues and that the high expression of ASF1B predicted poor outcome in HCC patients. Further systematic and comprehensive analysis of public databases revealed that ASF1B was also correlated with immune cell infiltration and activation in the tumor microenvironment of HCC and that it could predict the efficacy of immunotherapy.

The histone H3-H4 chaperone, ASF1, is a key factor for S-phase progression in the cell cycle in various organisms (Schulz and Tyler, 2006; Groth et al., 2007). Previous studies have highlighted the distinct functions of ASF1 isoforms (Abascal et al., 2013). It has been reported that ASF1B, rather than ASF1A, is critical for proliferation and has highly significant prognostic value in breast cancer and cervical cancer (Corpet et al., 2011; Han et al., 2018; Jiangqiao et al., 2019; Liu et al., 2020). ASF1B protein is increased by approximately 5.5-fold in tumor versus normal cells, and it is significantly correlated with the p60, p150, and Ki67 proliferation markers in breast cancer (Corpet et al., 2011). Disrupting ASF1B significantly suppresses tumor growth by G2/S stage cell cycle arrest and stimulates the apoptosis pathway. Furthermore, ASF1B levels, but not ASF1A levels, have a highly significant positive correlation with tumor size, tumor grade, and mitotic cell number. Thus, the expression of ASF1B has high prognostic value in breast cancer (Corpet et al., 2011). However, there exist no systematic and elaborated study on the expression of ASF1B in human HCC tissue and its association with the clinical prognosis of patients. Thus, in the present study, we evaluated ASF1B mRNA levels using different databases, and we evaluated protein levels by immunohistochemical analysis in HCC tumor and paracarcinoma tissues from patients treated at our clinical center. The results showed that the mRNA and protein levels of ASF1B were aberrantly high in HCC, and Western blotting of ASF1B protein levels in a normal liver cell line and tumor cell lines confirmed these results. In the tumor tissue, the expression of ASF1B was associated with tumor grade, T stage, and TNM stage. The survival analysis showed that high expression of ASF1B was closely associated with poor OS and PFS, indicating that ASF1B was an independent prognostic factor for HCC patients. These results were validated by independent samples collected from TCGA database and our center. Biological function analysis revealed that ASF1B was mainly correlated with the cell cycle but also with pathways that affected gene instability and cell proliferation, including DNA replication and gene repair in HCC. These results were consistent with studies addressing other cancers (Corpet et al., 2011; Han et al., 2018; Jiangqiao et al., 2019; Liu et al., 2020). The phosphatidylinositol 3 kinase (PI3K)/protein kinase B (Akt) signaling pathway plays a vital role in cell growth and proliferation, and has been observed to be dysregulated in various cancer types, including HCC (Zhu et al., 2021). It is reported that downregulation of ASF1B significantly decreased the phosphorylation levels of PI3K and protein Akt. Thus the PI3K/Akt pathway may be a possible mechanism underlying the carcinogenic effect of ASF1B (Han et al., 2018). In particular, it was observed that HCC patients in the high ASF1B group had a significantly higher frequency of TP53 and RB1 mutation, which are typical tumor suppressors, and their mutation leads to tumorigenesis and progression in HCC

TABLE 4 | Correlation between ASF1B expression and immune marker sets of immune cells in HCC based data from TIMER database.

Terms	Markers	None		Tumor purity		Terms	Markers	None		Tumor purity	
		R	*p*-value	*R*	*p*-value			*R*	*p*-value	*R*	*p*-value
T cell exhaustion	PDCD1	0.315	5.63E-10	0.428	9.06E-17	TAM	CCL2	0.07	1.81E-01	0.195	2.74E-04
	CD274 (PD-L1)	0.228	8.81E-06	0.315	2.19E-09		CD68	0.204	7.48E-05	0.309	4.66E-09
	CTLA-4	0.299	4.17E-09	0.421	3.09E-16		IL10	0.175	6.97E-04	0.297	1.83E-08
	LAG3	0.334	3.78E-11	0.388	7.33E-14	Natural killer cell	KIR2DL1	−0.012	8.17E-01	−0.029	5.88E-01
	HAVCR2 (TIM3)	0.287	1.81E-08	0.462	1.28E-19		KIR2DL3	0.187	2.84E-04	0.238	7.85E-06
	TIGIT	0.266	1.99E-07	0.406	4.34E-15		KIR3DL1	0.02	6.97E-01	0.04	4.61E-01
	B7-H3	0.541	1.30E-29	0.551	9.32E-29		KIR3DL2	0.09	8.30E-02	0.14	9.01E-03
	B7-H4	0.242	2.29E-06	0.311	3.64E-09		CD56	0.175	7.12E-04	0.271	3.33E-07
	BTLA (CD272)	0.14	6.87E-03	0.28	1.19E-07		CD335 (NCR1)	0.032	5.37E-01	0.098	6.93E-02
	HVEM	0.327	1.03E-10	0.36	5.15E-12	Neutrophils	CD66b (CEACAM8)	0.054	2.97E-01	0.088	1.01E-01
	CD96	0.176	6.49E-04	0.325	6.00E-10		CD11b (ITGAM)	0.29	1.33E-08	0.398	1.50E-14
	CD273 (PD-L2)	0.062	2.35E-01	0.193	3.01E-04		CD15	0.388	9.57E-15	0.426	1.31E-16
	GZMB	0.096	6.51E-02	0.182	6.90E-04	Monocyte	CD14	−0.394	3.07E-15	−0.353	1.56E-11
T cell (general)	CD3D	0.252	8.86E-07	0.38	2.66E-13		CD115 (CSF1R)	0.142	6.23E-03	0.3	1.25E-08
	CD3E	0.164	1.55E-03	0.332	2.54E-10	Dendritic cell	BDCA-1 (CD1C)	0.082	1.13E-01	0.184	5.90E-04
	CD2	0.183	3.98E-04	0.339	1.05E-10		BDCA-3	−0.006	2.07E-01	0.02	7.16E-01
	CD28	0.238	2.73E-06	0.335	1.78E-10		BDCA-4 (NRP1)	0.193	1.79E-04	0.225	2.56E-05
CTL (Cytotoxic T	CD8A	0.197	1.38E-04	0.323	7.60E-10		CD123	0.054	3.02E-01	0.129	1.68E-02
Lymphocytes)	CD8B	0.201	9.72E-05	0.32	1.14E-09		CD11c (ITGAX)	0.283	2.85E-08	0.425	1.37E-16
	CD45	0.206	6.59E-05	0.346	3.82E-11	Th1	T-bet (TBX21)	0.051	3.23E-01	0.169	1.61E-03
	HLA-DPB1	0.175	7.04E-04	0.317	1.82E-09		STAT4	0.242	2.42E-06	0.314	2.42E-09
	HLA-DQB1	0.148	4.39E-03	0.272	2.84E-07		STAT1	0.335	3.76E-11	0.387	8.70E-14
	HLA-DRA	0.172	8.74E-04	0.311	3.60E-09	Th2	GATA3	0.17	1.02E-03	0.318	1.55E-09
	HLA-DPA1	0.142	6.55E-03	0.285	6.83E-08		STAT6	0.127	1.43E-02	0.123	2.26E-02
B cell	CD19	0.269	1.36E-07	0.361	4.57E-12		IL13	0.109	3.60E-02	0.122	2.31E-02
	CD79A	0.156	2.54E-03	0.289	4.46E-08	Tfh	BCL6	0.127	1.47E-02	0.135	1.23E-02
	CD79B	1.148	4.36E-03	0.247	3.45E-06		IL21	0.141	6.53E-03	0.158	5.39E-04
	CD22	0.155	2.77E-03	0.244	4.40E-06	Th17	STAT3	0.13	1.21E-02	0.183	6.39E-04
	CIITA	0.251	9.31E-05	0.39	5.54E-14		IL17A	0.031	5.47E-01	0.046	3.90E-01
M1 Macrophage	INOS (NOS2)	0.001	9.87E-01	0.013	8.16E-01		RORγt	−0.157	2.45E-01	−0.231	1.50E-05
	IRF5	0.404	5.79E-16	0.413	1.11E-15	Treg	FOXP3	0.147	4.57E-03	0.233	1.25E-05
	IL6	0.01	9.86E-01	0.018	2.80E-02		CD25	0.234	5.00E-06	0.377	4.43E-13
	CD80	0.327	1.01E-10	0.459	2.31E-19		CCR8	0.319	3.23E-10	0.422	2.66E-16
	CD64 (FCGR1A)	0.294	7.72E-09	0.461	1.35E-19		STAT5B	0.225	1.24E-05	0.211	8.10E-05
	PTGS2	0.056	2.84E-01	0.189	4.02E-04		TGFB1	0.243	2.21E-06	0.348	2.94E-11
M2 Macrophage	CD163	0.039	4.51E-01	0.16	2.93E-03						
	CD206(MRC1)	-0.091	8.04E-02	−0.009	8.40E-01						
	VSIG4	0.073	1.61E-01	0.199	1.95E-04						
	MS4A4A	0.057	2.72E-01	0.2	1.79E-04						

Purity, correlation adjusted by purity; None, correlation without adjustment; TAM: tumor-associated macrophage; Th: T helper cell; Tfh: Follicular helper T cell; Treg: regulatory T cell; Cor: R *value of Spearman's correlation.*

(Buendia, 2000; Khemlina et al., 2017). In addition, although the comparison of TMB was not different between the two subsets, MATH analysis revealed that the high-ASF1B group exhibited higher abundances of tumor heterogeneity, which is generally an indicator for poor clinical outcomes in multiple malignances, including HCC (Mroz et al., 2015; Ma et al., 2019; McDonald et al., 2019). However, the exact molecular events leading to cancer proliferation and poor prognosis in HCC have not yet been well elucidated, and further research is required. Altogether, ASF1B functions as an oncogene and is a potential diagnostic and prognostic biomarker in HCC.

Cell cycle dysregulation and immune escape are hallmarks of malignant tumors (Hanahan and Weinberg, 2011). Although neither phenomena is a single driver of tumor evolution, each stands for an important axis of potential therapeutic intervention. Surprisingly, several studies in recent years have revealed that these axes have crosstalk, and targeting cell cycle regulatory molecules improves the efficacy of tumor immunotherapy, thereby improving the response rate of immunotherapy (Goel et al., 2017; Deng et al., 2018; Greten and Korangy, 2018; Zhang H. et al., 2018; Zhou et al., 2018; Jin et al., 2019). Inhibition of CDK4/6 enhances tumor antigen presentation and effector T cell infiltration and activation, and it markedly suppresses the proliferation of regulatory T cells (Goel et al., 2017; Deng et al., 2018). Inhibition of cyclin-dependent kinase 9 (CDK9) leads to an increase in immune cell infiltration in the TIME, including $CD45^+$ immune cells, $CD3^+$ T cells, and activated dendritic cells (Zhang J. et al., 2018). It has been reported that ASF1B binds to CDK9 and inhibits its proteasome-mediated ubiquitination and degradation leading to the CDK9 protein

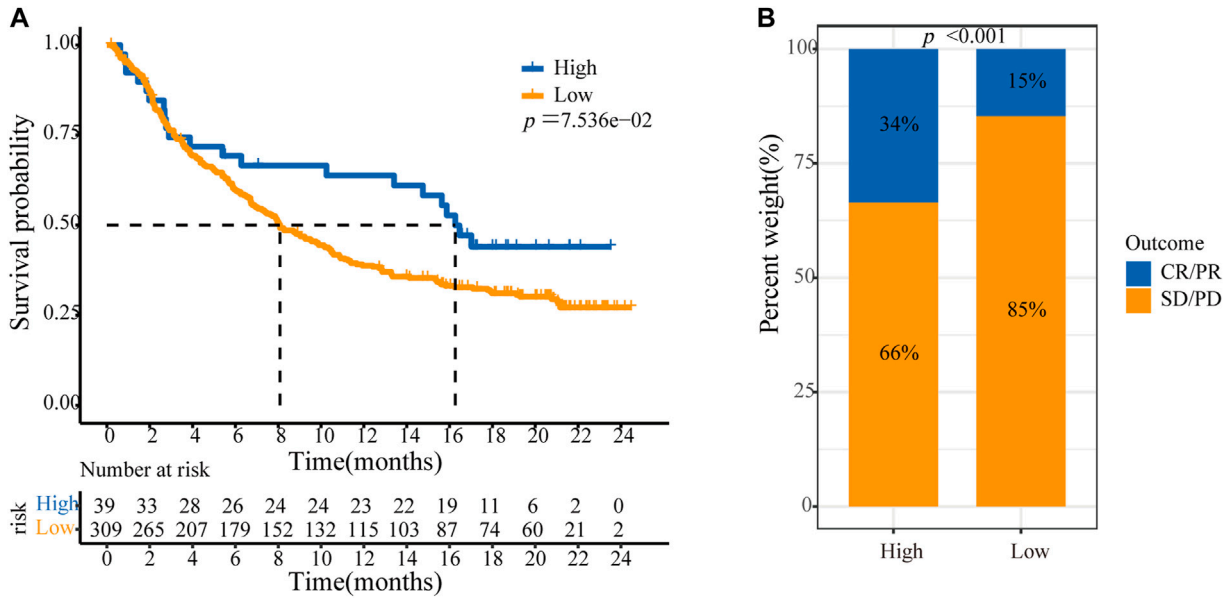


FIGURE 10 | ASF1B predicts immunotherapy efficacy. **(A)** Kaplan–Meier analysis of OS according to the level of ASF1B expression among patients using immunotherapy; **(B)** The immunotherapy efficacy in ASF1B high and low expression groups. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.

stabilization (Liu et al., 2020). Thus, as a cell cycle regulator molecule, we comprehensively explored the relationship between ASF1B and the degree of immune cell infiltration in HCC using multiple databases including TCGA, TIMER, XCELL, QUANTISEQ, EPIC, CIBERSORT-ABS, CIBERSORT and ssGSEA. In the present study, we found that the ASF1B expression level was significantly correlated with the expression of cell-cycle modulators, which was accordance to previously published studies (Corpet et al., 2011; Liu et al., 2020). ASF1B expression was also significantly associated with immune checkpoints expression such as CD274, CTLA4, PDCD1 and so on. Meantime, we also found that cell cycle regulatory molecules were positively related to immune checkpoint molecules, further suggesting the close link between cell cycle and tumor immunity. Additionally, 13 cell-cycle modulators and nine immune checkpoints were significantly increased in the ASF1B-high group in HCC. We also found that ASF1B mRNA levels were positively correlated with the immune cells infiltration degree, such as $CD4^+$ T cells, $CD8^+$ T cells, B cells, neutrophils, macrophages, and DCs in HCC which was also confirmed by Zhan's study. Furthermore, we explored the correlations of ASF1B expression with genetic markers of various immune cells after tumor purity adjusting. The results were in line with the above findings and showed that ASF1B expression was correlated with most immune cell markers in HCC. Zhan et al. (2021) has reported that downregulation of ASF1B inhibited the expression of CD86, CD8, STAT1, STAT4, CD68, and PD1 in HCC cells which further supported our results. However, we found that ASF1B was positively related to Treg cell infiltration and significantly associated with increased Treg markers, including CCR8, STAT5B, TGF-B1, and CD25. In addition, ASF1B was significantly positively correlated with T cell exhaustion markers (PD-1, CTLA4, LAG3, TIM-3, and TIGIT) in HCC. Moreover, the high ASF1B expression group correlated with impaired antitumor immunity, including the low activity of type II IFN response and type I IFN response, which support cytotoxic T lymphocytes by stimulating the maturation of dendritic cells and enhancing their capacity to process and present antigens (Miar et al., 2020). These results provided evidences that even though ASF1B positively correlates with tumor-infiltrating immune cells, such as $CD4^+$ T cells and $CD8^+$ T cells, infiltrating T cells in the tumor microenvironment are inactivated and exhausted. In addition, although TMB was not different between the ASF1B-high group and -low group, MATH analysis revealed that the high ASF1B expression exhibited higher abundances of tumor heterogeneity, which is generally associated with impaired immunity in multiple malignances (McDonald et al., 2019). Considered together, these data suggest that high ASF1B values might be correlated with immunosuppression in HCC. Previous studies have indicated that infiltrating immune cells and immune checkpoint expression in tumor sites influence prognosis and the response rate of immunotherapy (Calderaro et al., 2016; Kim et al., 2018). We found that patients with high ASF1B expression had higher patient response rates and achieved survival benefits from immune checkpoint inhibitors. The molecular mechanisms connecting cell cycle and immune surveillance, two of the most central processes in tumor biology, is still unclear. It is reported that inhibition of cell cycle regulators enhances tumour antigen presentation through stimulating production of type III interferons and finally increasing MHC class I molecules (Goel et al., 2017). And the CDK4/6 inhibitors promote PD-L1 expression by prevent the proteasome-mediated degradation inducing by Cullin3 SPOP E3 ligase (Zhang J. et al., 2018). In addition, the cell cycle block imposed on tumour cells is associated with cellular senescence, which promotes the secretion of cytokines and chemokines such as

CCL5, CXCL9 and CXCL10, increase the infiltration and activation of T cells (Vilgelm et al., 2016). Exosomes, as the bridge between cells in tumor microenvironment, modulate the immune response (Zhang et al., 2021). Tumor microenvironment associated exosomes have the potential to serve as a breakthrough to clarify the connection between the cell cycle and immune surveillance. In addition to immune cells, other non-immune components in the tumor immune microenvironment can also affect tumor immunity, for example fibroblasts, myofibroblasts, endothelial cells, and extracellular matrix (Zhang J. et al., 2020; Guo et al., 2021). How ASF1B affects non immune cells remains to be explored. Therefore, as an important cell cycle regulator, how ASF1B modulates the TIME still needs to be verified by further extensive basic experiments.

The present study systematically reported the clinical significance of ASF1B and its effect on the TIME in HCC, but it also had several limitations. First, although we included data from our center to validate the expression and prognostic value of ASF1B, our results present limited evidence about the mechanism by which ASF1B upregulation affects the outcome of patients and the TIME with HCC. But we are currently conducting basic experiments *in vivo* and vitro to investigate how the aberrant elevation of ASF1B influences tumor behavior and the TIME in HCC. Second, the clinical data on immunotherapy administered to patients with HCC were unclear, which prevented us from performing a more detailed analysis. Thus, more clinical information needs to be collected on HCC patients applying immunotherapy to confirm these results.Third, integrating genomic, transcriptomic, proteomic, metabolomic and epigenomic datasets via multi-omics analysis could derive a deeper understanding of the development and progression of cancer (Sathyanarayanan et al., 2020; Su et al., 2020). More multi-omics data are needed to provide a comprehensive landscape of ASF1B roles in HCC.

CONCLUSION

In summary, this study revealed that the expression of ASF1B was elevated in HCC tissues, and it could be a potential biomarker to predict HCC patient prognosis. In addition, high ASF1B expression is related to immunosuppressive tumor microenvironment characteristics with high expression of T cell exhaustive markers in HCC. However, high ASF1B expression predicts survival benefits from immune checkpoint inhibitors. Thus, these findings may be helpful in the management of patients in clinical practice as well as develop novel treatment strategies in the future to improve the sensitivity of immunotherapy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committees of the First Affiliated Hospital of Xi'an Jiaotong University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SZ and HG conceived the idea and were the grant recipients in this project. LX and JF performed bioinformatic analysis. SZ, DT, and YZ performed experimental validation. JH, WL, and KL downloaded and organized the clinical and gene expression data. WW, LJ, and MJ critically revised the article for essential intellectual content and administrative support. All authors participated in the manuscript preparation, proofreading and submission.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (82072701), the Clinical Research Award of the First Affiliated Hospital of Xi'an Jiaotong University, China (No. XJTU1AF-CRF- 2019-001), and Xi'an Jiaotong University Free Exploration and Innovation Program (Student Category) (No. sxzy022021011).

ACKNOWLEDGMENTS

All authors would like to express our sincere thanks for sharing the data from TCGA database and other databases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.842351/full#supplementary-material

REFERENCES

Abascal, F., Corpet, A., Gurard-Levin, Z. A., Juan, D., Ochsenbein, F., Rico, D., et al. (2013). Subfunctionalization via Adaptive Evolution Influenced by Genomic Context: the Case of Histone Chaperones ASF1a and ASF1b. *Mol. Biol. Evol.* 30 (8), 1853–1866. doi:10.1093/molbev/mst086

Akinyemiju, T., Akinyemiju, T., Abera, S., Ahmed, M., Alam, N., Alemayohu, M. A., et al. (2017). The Burden of Primary Liver Cancer and Underlying Etiologies from 1990 to 2015 at the Global, Regional, and National Level: Results from the Global Burden of Disease Study 2015. *JAMA Oncol.* 3 (12), 1683–1691. doi:10.1001/jamaoncol.2017.3055

Annick Buendia, M. (2000). Genetics of Hepatocellular Carcinoma. *Semin. Cancer Biol.* 10, 185–200. doi:10.1006/scbi.2000.0319

Calderaro, J., Rousseau, B., Amaddeo, G., Mercey, M., Charpy, C., Costentin, C., et al. (2016). Programmed Death Ligand 1 Expression in Hepatocellular Carcinoma: Relationship with Clinical and Pathological Features. *Hepatology* 64 (6), 2038–2046. doi:10.1002/hep.28710

Cariani, E., and Missale, G. (2019). Immune Landscape of Hepatocellular Carcinoma Microenvironment: Implications for Prognosis and Therapeutic Applications. *Liver Int.* 39 (9), 1608–1621. doi:10.1111/liv.14192

Corpet, A., De Koning, L., Toedling, J., Savignoni, A., Berger, F., Lemaître, C., et al. (2011). Asf1b, the Necessary Asf1 Isoform for Proliferation, is Predictive of Outcome in Breast Cancer. *EMBO J.* 30 (3), 480–493. doi:10.1038/emboj.2010.335

De Koning, L., Corpet, A., Haber, J. E., and Almouzni, G. (2007). Histone Chaperones: an Escort Network Regulating Histone Traffic. *Nat. Struct. Mol. Biol.* 14 (11), 997–1007. doi:10.1038/nsmb1318

Deng, J., Wang, E. S., Jenkins, R. W., Li, S., Dries, R., Yates, K., et al. (2018). CDK4/6 Inhibition Augments Antitumor Immunity by Enhancing T-Cell Activation. *Cancer Discov.* 8 (2), 216–233. doi:10.1158/2159-8290.cd-17-0915

Eitoku, M., Sato, L., Senda, T., and Horikoshi, M. (2008). Histone Chaperones: 30 Years from Isolation to Elucidation of the Mechanisms of Nucleosome Assembly and Disassembly. *Cell. Mol. Life Sci.* 65 (3), 414–444. doi:10.1007/s00018-007-7305-6

El-Khoueiry, A. B., Sangro, B., Yau, T., Crocenzi, T. S., Kudo, M., Hsu, C., et al. (2017). Nivolumab in Patients with Advanced Hepatocellular Carcinoma (CheckMate 040): An Open-Label, Non-comparative, Phase 1/2 Dose Escalation and Expansion Trial. *Lancet* 389 (10088), 2492–2502. doi:10.1016/s0140-6736(17)31046-2

Finn, R. S., Qin, S., Ikeda, M., Galle, P. R., Ducreux, M., Kim, T.-Y., et al. (2020). Atezolizumab Plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N. Engl. J. Med.* 382 (20), 1894–1905. doi:10.1056/nejmoa1915745

Goel, S., DeCristo, M. J., Watt, A. C., BrinJones, H., Sceneay, J., Li, B. B., et al. (2017). CDK4/6 Inhibition Triggers Anti-Tumour Immunity. *Nature* 548 (7668), 471–475. doi:10.1038/nature23465

Greten, T. F., and Korangy, F. (2018). CDK20 Inhibition and Immune Checkpoint Blockade: Bringing Cancer Biology and Tumour Immunology Together to Develop Novel Treatment Options for HCC. *Gut* 67 (5), 783–784. doi:10.1136/gutjnl-2017-315091

Groth, A., Corpet, A., Cook, A. J. L., Roche, D., Bartek, J., Lukas, J., et al. (2007). Regulation of Replication fork Progression through Histone Supply and Demand. *Science* 318 (5858), 1928–1931. doi:10.1126/science.1148992

Guo, H., Zhang, T., Yu, Y., and Xu, F. (2021). Cancer Physical Hallmarks as New Targets for Improved Immunotherapy. *Trends Cell Biol.* 31, 520–524. doi:10.1016/j.tcb.2021.03.011

Han, G., Zhang, X., Liu, P., Yu, Q., Li, Z., Yu, Q., et al. (2018). Knockdown of Anti-silencing Function 1B Histone Chaperone Induces Cell Apoptosis via Repressing PI3K/Akt Pathway in Prostate Cancer. *Int. J. Oncol.* 53 (5), 2056–2066. doi:10.3892/ijo.2018.4526

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

Hu, X., Zhu, H., Zhang, X., He, X., and Xu, X. (2021). Comprehensive Analysis of pan-cancer Reveals Potential of ASF1B as a Prognostic and Immunological Biomarker. *Cancer Med.* 10 (19), 6897–6916. doi:10.1002/cam4.4203

Jiangqiao, Z., Tao, Q., Zhongbao, C., Xiaoxiong, M., Long, Z., Jilin, Z., et al. (2019). Anti-silencing Function 1B Histone Chaperone Promotes Cell Proliferation and Migration via Activation of the AKT Pathway in clear Cell Renal Cell Carcinoma. *Biochem. Biophys. Res. Commun.* 511 (1), 165–172. doi:10.1016/j.bbrc.2019.02.060

Jin, X., Ding, D., Yan, Y., Li, H., Wang, B., Ma, L., et al. (2019). Phosphorylated RB Promotes Cancer Immunity by Inhibiting NF-κB Activation and PD-L1 Expression. *Mol. Cell* 73 (1), 22–35.e6. doi:10.1016/j.molcel.2018.10.034

Khemlina, G., Ikeda, S., and Kurzrock, R. (2017). The Biology of Hepatocellular Carcinoma: Implications for Genomic and Immune Therapies. *Mol. Cancer* 16, 149. doi:10.1186/s12943-017-0712-x

Kim, H.-D., Song, G.-W., Park, S., Jung, M. K., Kim, M. H., Kang, H. J., et al. (2018). Association between Expression Level of PD1 by Tumor-Infiltrating CD8+ T Cells and Features of Hepatocellular Carcinoma. *Gastroenterology* 155 (6), 1936–1950.e17. doi:10.1053/j.gastro.2018.08.030

Lei, Z., Yu, S., Ding, Y., Liang, J., Halifu, Y., Xiang, F., et al. (2020). Identification of Key Genes and Pathways Involved in Vitiligo Development Based on Integrated Analysis. *Medicine* 99 (31), e21297. doi:10.1097/md.0000000000021297

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.can-17-0307

Lian, Q., Wang, S., Zhang, G., Wang, D., Luo, G., Tang, J., et al. (2018). HCCDB: A Database of Hepatocellular Carcinoma Expression Atlas. *Genomics Proteomics Bioinformatics* 16 (4), 269–275. doi:10.1016/j.gpb.2018.07.003

Lin, R., Fogarty, C. E., Ma, B., Li, H., Ni, G., Liu, X., et al. (2021). Identification of Ferroptosis Genes in Immune Infiltration and Prognosis in Thyroid Papillary Carcinoma Using Network Analysis. *BMC Genomics* 22 (1), 576. doi:10.1186/s12864-021-07895-6

Liu, X., Song, J., Zhang, Y., Wang, H., Sun, H., Feng, X., et al. (2020). ASF1B Promotes Cervical Cancer Progression through Stabilization of CDK9. *Cell Death Dis.* 11 (8), 705. doi:10.1038/s41419-020-02872-5

Liu, Z., Yang, Q., Shi, O., Ye, W., Chen, X., and Zhang, T. (2018). The Epidemiology of Hepatitis B and Hepatitis C Infections in China from 2004 to 2014: An Observational Population-Based Study. *J. Viral Hepat.* 25 (12), 1543–1554. doi:10.1111/jvh.12938

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., et al. (2021). Hepatocellular Carcinoma. *Nat. Rev. Dis. Primers* 7 (1), 6. doi:10.1038/s41572-020-00240-3

Ma, L., Hernandez, M. O., Zhao, Y., Mehta, M., Tran, B., Kelly, M., et al. (2019). Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer cell* 36, 418–430.e6. doi:10.1016/j.ccell.2019.08.007

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554, 544–548. doi:10.1038/nature25501

McDonald, K.-A., Kawaguchi, T., Qi, Q., Peng, X., Asaoka, M., Young, J., et al. (2019). Tumor Heterogeneity Correlates with Less Immune Response and Worse Survival in Breast Cancer Patients. *Ann. Surg. Oncol.* 26, 2191–2199. doi:10.1245/s10434-019-07338-3

Miar, A., Arnaiz, E., Bridges, E., Beedie, S., Cribbs, A. P., Downes, D. J., et al. (2020). Hypoxia Induces Transcriptional and Translational Downregulation of the Type I IFN Pathway in Multiple Cancer Cell Types. *Cancer Res.* 80, 5245–5256. doi:10.1158/0008-5472.CAN-19-2306

Mroz, E. A., Tward, A. M., Hammon, R. J., Ren, Y., and Rocco, J. W. (2015). Intra-Tumor Genetic Heterogeneity and Mortality in Head and Neck Cancer: Analysis of Data from the Cancer Genome Atlas. *PLoS Med.* 12, e1001786. doi:10.1371/journal.pmed.1001786

Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., and Senda, T. (2007). Structure and Function of the Histone Chaperone CIA/ASF1 Complexed with Histones H3 and H4. *Nature* 446 (7133), 338–341. doi:10.1038/nature05613

Petroni, G., Formenti, S. C., Chen-Kiang, S., and Galluzzi, L. (2020). Immunomodulation by Anticancer Cell Cycle Inhibitors. *Nat. Rev. Immunol.* 20 (11), 669–679. doi:10.1038/s41577-020-0300-y

Sathyanarayanan, A., Gupta, R., Thompson, E. W., Nyholt, D. R., Bauer, D. C., and Nagaraj, S. H. (2020). A Comparative Study of Multi-Omics Integration Tools for Cancer Driver Gene Identification and Tumour Subtyping. *Brief. Bioinformatics* 21, 1920–1936. doi:10.1093/bib/bbz121

Schulz, L. L., and Tyler, J. K. (2006). The Histone Chaperone ASF1 Localizes to Active DNA Replication forks to Mediate Efficient DNA Replication. *FASEB j.* 20 (3), 488–490. doi:10.1096/fj.05-5020fje

Su, J., Song, Q., Qasem, S., O'Neill, S., Lee, J., et al. (2020). Multi-Omics Analysis of Brain Metastasis Outcomes Following Craniotomy. *Front. Oncol.* 10, 615472. doi:10.3389/fonc.2020.615472

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Teh, J. L. F., and Aplin, A. E. (2019). Arrested Developments: CDK4/6 Inhibitor Resistance and Alterations in the Tumor Immune Microenvironment. *Clin. Cancer Res.* 25 (3), 921–927. doi:10.1158/1078-0432.ccr-18-1967

Vilgelm, A. E., Johnson, C. A., Prasad, N., Yang, J., Chen, S.-C., Ayers, G. D., et al. (2016). Connecting the Dots: Therapy-Induced Senescence and a Tumor-Suppressive Immune Microenvironment. *J. Natl. Cancer Inst.* 108, djv406. doi:10.1093/jnci/djv406

Yau, T., Kang, Y.-K., Kim, T.-Y., El-Khoueiry, A. B., Santoro, A., Sangro, B., et al. (2020). Efficacy and Safety of Nivolumab Plus Ipilimumab in Patients with Advanced Hepatocellular Carcinoma Previously Treated with Sorafenib. *JAMA Oncol.* 6 (11), e204564. doi:10.1001/jamaoncol.2020.4564

Yu, R., Fan, R., and Hou, J. (2014). Chronic Hepatitis B Virus Infection: Epidemiology, Prevention, and Treatment in China. *Front. Med.* 8 (2), 135–144. doi:10.1007/s11684-014-0331-5

Zhan, T., Gao, X., Wang, G., Li, F., Shen, J., Lu, C., et al. (2021). Construction of Novel lncRNA-miRNA-mRNA Network Associated With Recurrence and Identification of Immune-Related Potential Regulatory Axis in Hepatocellular Carcinoma. *Front. Oncol.* 11, 626663. doi:10.3389/fonc.2021.626663

Zhang, H., Christensen, C. L., Dries, R., Oser, M. G., Deng, J., Diskin, B., et al. (2020). CDK7 Inhibition Potentiates Genome Instability Triggering Anti-tumor Immunity in Small Cell Lung Cancer. *Cancer Cell* 37 (1), 37–54.e9. doi:10.1016/j.ccell.2019.11.003

Zhang, H., Pandey, S., Travers, M., Sun, H., Morton, G., Madzo, J., et al. (2018). Targeting CDK9 Reactivates Epigenetically Silenced Genes in Cancer. *Cell* 175 (5), 1244–1258.e26. doi:10.1016/j.cell.2018.09.051

Zhang, J., Bu, X., Wang, H., Zhu, Y., Geng, Y., Nihira, N. T., et al. (2018). Cyclin D-CDK4 Kinase Destabilizes PD-L1 via Cullin 3-SPOP to Control Cancer Immune Surveillance. *Nature* 553 (7686), 91–95. doi:10.1038/nature25015

Zhang, J., Gu, C., Song, Q., Zhu, M., Xu, Y., Xiao, M., et al. (2020). Identifying Cancer-Associated Fibroblasts as Emerging Targets for Hepatocellular Carcinoma. *Cell Biosci.* 10, 127. doi:10.1186/s13578-020-00488-y

Zhang, J., Song, Q., Wu, M., and Zheng, W. (2020). The Emerging Roles of Exosomes in the Chemoresistance of Hepatocellular Carcinoma. *Curr. Med. Chem.* 28, 93–109. doi:10.2174/0929867327666200130103206

Zhang, S.-R., Li, H., Wang, W.-Q., Jin, W., Xu, J.-Z., Xu, H.-X., et al. (2019). Arpin Downregulation is Associated with Poor Prognosis in Pancreatic Ductal Adenocarcinoma. *Eur. J. Surg. Oncol.* 45 (5), 769–775. doi:10.1016/j.ejso.2018.10.539

Zhou, J., Liu, M., Sun, H., Feng, Y., Xu, L., Chan, A. W. H., et al. (2018). Hepatoma-Intrinsic CCRK Inhibition Diminishes Myeloid-Derived Suppressor Cell Immunosuppression and Enhances Immune-Checkpoint Blockade Efficacy. *Gut* 67 (5), 931–944. doi:10.1136/gutjnl-2017-314032

Zhu, A. X., Finn, R. S., Edeline, J., Cattan, S., Ogasawara, S., Palmer, D., et al. (2018). Pembrolizumab in Patients with Advanced Hepatocellular Carcinoma Previously Treated with Sorafenib (KEYNOTE-224): A Non-Randomised, Open-Label Phase 2 Trial. *Lancet Oncol.* 19 (7), 940–952. doi:10.1016/s1470-2045(18)30351-6

Zhu, M., Wu, M., Bian, S., Song, Q., Xiao, M., Huang, H., et al. (2021). DNA Primase Subunit 1 Deteriorated Progression of Hepatocellular Carcinoma by Activating AKT/mTOR Signaling and UBE2C-Mediated P53 Ubiquitination. *Cell Biosci.* 11, 42. doi:10.1186/s13578-021-00555-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Xu, Feng, Tan, Zhu, Hou, Li, Lv, Wang, Jiang, Jiao and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 March 2022
doi: 10.3389/fgene.2022.841712

One Ferroptosis-Related Gene-Pair Signature Serves as an Original Prognostic Biomarker in Lung Adenocarcinoma

Lei Li[1,2] *and Buhai Wang*[1]*

[1]Department of Oncology, Northern Jiangsu People's Hospital, Yangzhou, China, [2]Graduate School, Dalian Medical University, Dalian, China

Lung adenocarcinoma is the most common histological subtype of lung cancer which causes the largest number of deaths worldwide. Exploring reliable prognostic biomarkers based on biological behaviors and molecular mechanisms is essential for predicting prognosis and individualized treatment strategies. Ferroptosis is a recently discovered type of regulated cell death. We downloaded ferroptosis-related genes from the literature and collected transcriptome profiles of lung adenocarcinoma from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) to construct ferroptosis-related gene-pair matrixes. Then, we performed the least absolute shrinkage and selection operator regression to build our prognostic ferroptosis-related gene-pair index (FRGPI) in TCGA training matrix. Our study validated FRGPI through ROC curves, Kaplan–Meier methods, and Cox hazard analyses in TCGA and GEO cohorts. The optimal cut-off 0.081 stratified patients into low- and high-FRGPI groups. Also, the low-FRGPI group had a significantly better prognosis than the high-FRGPI group. For further study, we analyzed differentially expressed ferroptosis-related genes between high- and low-FRGPI groups. Gene set enrichment analysis (GSEA) enrichment maps indicated that "cell cycle," "DNA replication," "proteasome," and "the p53 signaling pathway" were significantly enriched in the high-FRGPI group. The high-FRGPI group also presented higher infiltration of M1 macrophages. Meanwhile, there were few differences in adaptive immune responses between high- and low-FRGPI groups. In conclusion, FRGPI was an independent prognostic biomarker which might be beneficial for guiding individualized tumor therapy.

Keywords: ferroptosis, lung adenocarcinoma, gene pair, prognostic marker, Cox model

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Olga Papadodima,
National Hellenic Research Foundation, Greece
Travis Steele Johnson,
Indiana University Bloomington, United States

***Correspondence:**
Buhai Wang
wbhself@sina.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *22 December 2021*
Accepted: *14 February 2022*
Published: *16 March 2022*

Citation:
Li L and Wang B (2022) One Ferroptosis-Related Gene-Pair Signature Serves as an Original Prognostic Biomarker in Lung Adenocarcinoma. Front. Genet. 13:841712. doi: 10.3389/fgene.2022.841712

INTRODUCTION

Lung cancer is the second most frequent cancer and causes the largest number of deaths worldwide, acccounting to 11.4 percent of new cases and 18 percent of cancer-related deaths in 2020 (Ferlay et al., 2020). Among the subtypes of lung cancer, non-small-cell lung cancer (NSCLC) accounts for the largest part and occupies 85 percent of lung cancer cases. Specifically, lung adenocarcinoma of NSCLC, representing 60 percent, is the most common histological

subtype (Arbour and Riely, 2019). Smoking is the acknowledged main risk factor for lung cancer, nevertheless, more possibly leading to squamous carcinoma than adenocarcinoma.

On the basis of histological types, clinical stages, and genetic alterations, integrative treatments are essential for NSCLC. Surgery is the first choice for localized stage I/II/IIIA/IIIB ($T_3N_2M_0$) NSCLC. Radiotherapy achieves curative intents for people who are not eligible for surgery and is helpful for symptomatic relief in smaller doses. Platinum-based chemotherapy, a traditional non-surgical treatment, is still a choice of first line in advanced NSCLC. Due to increasing molecular targets investigated, targeted therapy is a preferred treatment for stage IV NSCLC, especially adenocarcinoma. Meanwhile, immune checkpoint inhibitors are newly regarded as second-line therapy in advanced NSCLC (Zhai et al., 2020).

Localized, regional, and metastatic NSCLC, respectively, represent 63%, 35%, and 7% 5-year survival rate (Howlader et al., 2020, based on November 2019 SEER data submission, posted on the SEER website, April 2020). More accurate and noninvasive prognostic biomarkers are needed. Exploring prognostic biomarkers based on biological and molecular mechanisms is essential for individualized treatment strategies.

Ferroptosis is a newly found type of regulated cell death (RCD), distinct from apoptosis, necroptosis, and pyroptosis (Dixon et al., 2012). Ferroptosis is a reactive oxygen species (ROS)-inducing cell death form and exhibits two main biochemical processes, ferrous iron accumulation and lipid peroxidation (Homma et al., 2019). Excessive ferrous iron (Fe^{2+}), with hydrogen peroxide, generates hydroxyl radicals through the Fenton reaction and then reacts with polyunsaturated fatty acids (PUFAs) to induce lipid peroxidation. Lipid peroxidation ultimately causes membrane oxidative damage to accomplish ferroptosis.

Ferroptosis occurs through two typical pathways, the transporter-dependent pathway and the enzyme-regulated pathway (Tang et al., 2021). System X_c^-, composed of solute carrier family 7 member 11 (SLC7A11) and solute carrier family 3 member 2 (SLC3A2), uptakes cystine to sustain glutathione (GSH) production. Glutathione peroxidase 4 (GPX4) acts as a ferroptosis repressor and reduces lipid peroxidation while converting GSH to oxidized glutathione (GSSG).

Recently, there is growing evidence that oncology patients benefit from triggering ferroptosis of cancer cells during traditional treatments (Dixon et al., 2012; Wu et al., 2020). Classic ferroptosis inducers (FINs), such as erastin, sorafenib, cisplatin, RSL3, and FIN56, inhibit SLC7A11 activity, deplete GSH, or inhibit GPX4 activity to promote ferroptosis. Triggering ferroptosis in cancer shows the drug-resistance reversal effect and the synergistic sensitization effect with chemotherapy, target therapy, radiotherapy, and immunotherapy (Wu et al., 2020).

Several ferroptosis-related gene prognostic models were detected in multiple types of cancer. In our research, based on patients with lung adenocarcinoma, we focused on molecular mechanisms and signaling pathways of ferroptosis. Utilizing ferroptosis-related genes to build ferroptosis-related gene pairs (FRGPs) instead of single genes, we finally constructed an individualized prognostic signature biomarker of lung adenocarcinoma. On this basis, we stratified the risk of lung adenocarcinoma patients to predict prognoses and explore therapies.

MATERIALS AND METHODS

Data Acquisition and Processing

We collected transcriptome profiles of lung adenocarcinoma (LUAD) available in the TCGA database (https://portal.gdc.cancer.gov/) on 6 July 2021.

Preprocessed and aligned RNA-Seq samples were downloaded by selecting HTSeq-Counts as the workflow type on the portal. Clinical and pathological information related to the TCGA–LUAD cohort was retrieved from the cBioportal website (https://www.cbioportal.org) with the "cdgsr" package (Cerami et al., 2012; Jacobsen, 2015).

Meanwhile, we collected two microarray datasets and corresponding clinical information, including GSE68465 and GSE72094, from Gene Expression Omnibus using the "GEOquery" package (https://www.ncbi.nlm.nih.gov/geo/) (Davis and Meltzer, 2007). These two affymetrix microarrays were preprocessed using the RMA method (R package "affy") (Gautier et al., 2004).

Removing samples without overall survival (OS) information or with an OS time of 0 and converting the TNM stage to AJCC staging groups, the TCGA–LUAD cohort (N = 306) was used as the training cohort, whereas GSE68465 (N = 441) and GSE72094 (N = 398) were used as the validation cohorts (**Supplementary Table S1**).

The specific data processing and research flow are shown in **Figure 1**.

Construction of the Training Matrix

One ferroptosis-related gene set was downloaded from FerrDb (http://www.zhounan.org/ferrdb/index.html), the world's first database of ferroptosis regulators, markers, and associations (Zhou and Bao, 2020). There were 108 drivers annotated as genes that promoted ferroptosis, 69 suppressors that prevented ferroptosis, and 111 markers that indicated the occurrence of ferroptosis. Removing multi-annotated genes and selecting ferroptosis-related genes measured by using all the three cohorts, 209 genes were included in the ferroptosis-related gene set (**Supplementary Figure S1**).

The mean expression value of replicated genes was calculated. We defined the combination of two ferroptosis-related genes (FRG-1 and FRG-2) as a ferroptosis-related gene pair (FRGP). In every specific sample, 209 FRGs were compared by the gene expression level with each other to build 21736 FRGPs and generate a score for each FRGP. An FRGP score of 1 was assigned if FRG-1 was less than FRG-2. Otherwise, the FRGP score was 0. This gene-pair-based approach calculated the FRGP score based totally on the gene expression value of each individual sample and could be applied without normalization.

To reduce biases and be meaningful for subsequent analyses, some FRGPs counting 0s or 1s in more than 80% samples were

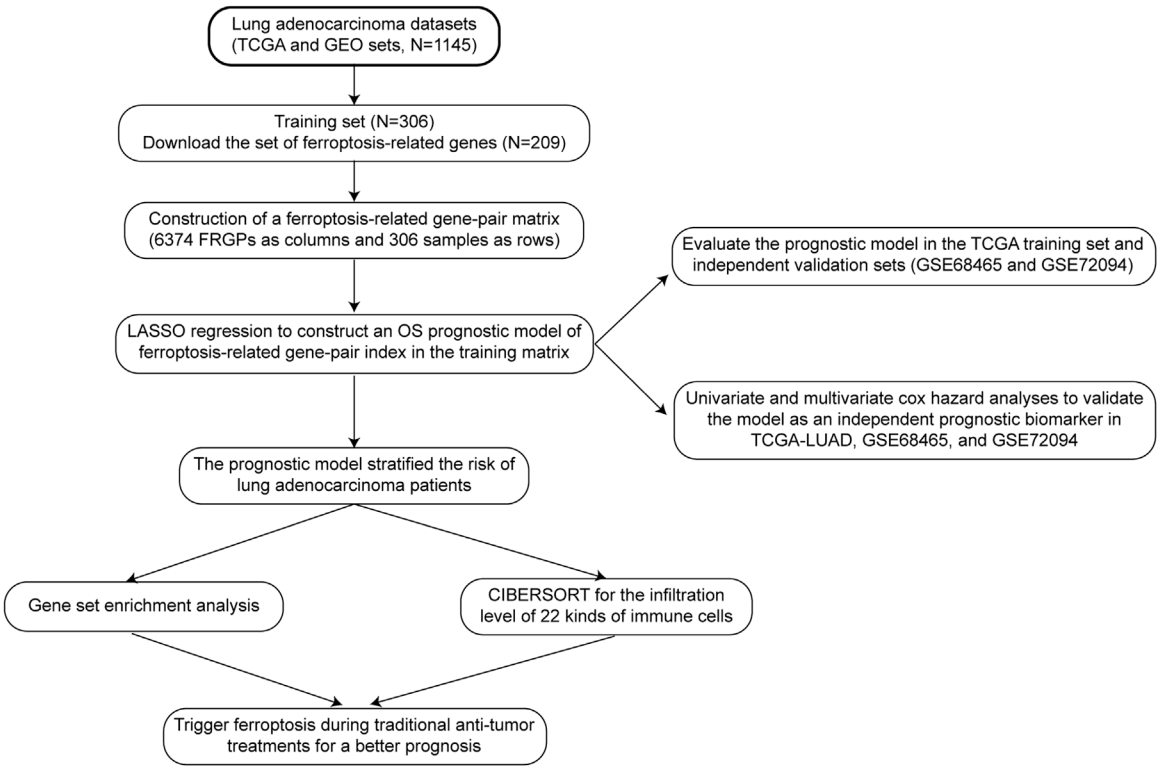


FIGURE 1 | Data processing and research process.

filtered out. Finally, we got 6374 FRGPs as columns and 306 samples as rows to form the training matrix.

Prognostic Ferroptosis-Related Gene-Pair Index Signature Construction

The least absolute shrinkage and selection operator (LASSO) is a statistical method to reduce data dimensionality. We applied the LASSO regression operation with the R package "glmnet" and "survival" in the TCGA FRGP matrix to construct a prognostic ferroptosis-related gene-pair index (FRGPI)(Simon et al., 2011). Insignificant variables whose coefficients became zero and any collinear variables were removed. 10-fold cross validation (CV) divided data into ten equal parts, nine parts as the training set and the remaining one part as the validation part. When partial likelihood deviance was the smallest, we got the minimum of lambda and nine gene pairs as our best FRGPI model. The prognostic index signature is expressed as FRGPI risk score = $\sum n_i$ ($FRGP_i$ *$coef_i$) ($i = 1,2,3 \ldots\ldots 9$, $n = 9$, where n is the number of FRGPs, FRGPi is the score (0 or 1) of the ith FRGP, and coefi is the regression coefficient of the ith FRGP).

Validation of FRGPI as a Prognostic Biomarker

First, we used the R package "survivalROC" to draw ROC curves and calculated the AUC values in the training and validation cohorts. AUC values greater than 0.5 and closer to 1 indicated the prognostic ability of FRGPI.

Second, we used the Kaplan–Meier method to compare survival outcomes between high- and low-FRGPI risk score groups in training and validation cohorts. The optimal cut-off value, determined based on the best balance of sensitivity and specificity to achieve the best AUC in the training cohort, was investigated using the ROC curves with the R package "survivalROC," and "survminer" (Weiss et al., 2003).

Validation of FRGPI as an Independent Prognostic Factor

After verification of the FRGPI significantly stratifying patients into low- and high-risk groups, we performed univariate and multivariate Cox hazard analyses to validate FRGPI as an independent prognostic factor. The hazard ratio (HR) in survival analyses less than 1 meant that the presence of the factor was protective, whereas the hazard ratio more than 1 was harmful.

Analysis of Differentially Expressed Ferroptosis-Related Genes and *MKI67* Between High- and Low-FRGPI Groups

After constructing and validating FRGPI, we compared the expression of ferroptosis-related genes between high- and low-

FRGPI groups in the TCGA cohort and GEO validation cohorts, using the Wilcoxon rank-sum test and the reshape2 package (Wickham, 2007). Meanwhile, we drew the boxplots for visualization using the ggplot2 package (Wickham, 2009).In addition, we compared *MKI67* expression between high- and low-FRGPI groups in the three cohorts. *MKI67* encodes a nuclear protein Ki-67, which is a commonly used marker for cell proliferation.

Gene Set Enrichment Analysis for Kyoto Encyclopedia of Genes and Genomes

Gene set enrichment analysis (GSEA) determines whether the gene sets, not single genes, present differences between different biological status groups and verifies that the gene sets are enriched in one specific clinical group (Subramanian et al., 2005). The gene sets are predefined by previous experiments and function annotations.

We chose gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway which is a collection of pathway maps representing molecular interactions, reactions, and relation networks (Kanehisa et al., 2016).

We performed GSEA–KEGG analyses and drew enrichment plots in the TCGA cohort and GEO validation cohorts, using "c2.cp.kegg.v7.4.entrez.gmt" and the R package "clusterProfiler," and "ggplot2," defining p value < 0.05 and q value < 0.05 as the filtering criteria (Yu et al., 2012).

Calculation of the Infiltration Level of 22 Kinds of Immune Cells

Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) is a deconvolution method for characterizing the cell composition of complex tissues from gene expression profiles (Newman et al., 2015). We made 22 kinds of immune cells as the target characterizing composition. We operated the "CIBERSORT" algorithm with the "Leukocyte signature matrix" (Newman et al., 2015). Based on the composition of 22 kinds of immune cells, we screened out what kind of immune cells differently infiltrated between high- and low-FRGPI groups.

Statistical Analysis

All statistical analyses were based on R Programming Language software (Rx64 3.3.3). The online website "www.genome.jp/kegg/" was used for GSEA–KEGG gene sets. The website "cibersortx" offered thoughts for CIBERSORT analyses in R software.

RESULTS

Prognostic Ferroptosis-Related Gene-Pair Index Construction

First, we downloaded one ferroptosis-related gene set including 209 genes. Two genes as one couple, any couple of the 209 genes could be created. So, 209 genes formed 21736 FRGPs in every individual sample. To calculate each FRGP in every sample, if the FRG-1 value was less than the FRG-2 value, the FRGP score was 1. Otherwise, the FRGP score was 0. Then, we filtered out FRGPs with constant values (0 or 1) in TCGA or GEO datasets. Finally, we got 6374 FRGPs in every individual sample. In TCGA training cohort, we used 306 samples as rows and 6374 FRGP scores as columns to make up the training FRGP matrix.

The training FRGP matrix was used for evaluating the relationship between FRGPs and overall survival rates applied using the LASSO regression operation. The lambda and coef diagrams of FRGPs (**Figure 2A**) were plotted using the LASSO algorithm. With the increase in the lambda value, the coefficients of some FRGPs decreased to zero, which meant that the scores of these FRGPs did not affect the model. We then used 10-fold CV to calculate the partial likelihood deviance of the model (**Figure 2B**). The minimum deviance exported the best model. The best model included nine gene pairs. Finally, we constructed a prognostic FRGPI signature with nine FRGPs and the corresponding coef values (**Figure 2C**).

Verification of FRGPI as a Prognostic Biomarker

The time-dependent ROC curves with AUC values of the training and validation cohorts are all presented in **Figures 2D,F,G**. All the AUC values were more than 0.5 and even greater than 0.7 in training and validation cohorts, indicating that FRGPI had a favorable prognostic ability.

In the training cohort, the AUC value reached 0.773 for 1606 days. We defined the optimal cut-off 0.081 in the curve with the best AUC and "1606" as the time point (**Figure 2E**). 0.081 was used as a cut-off for FRGPI to stratify patients into the low- or high-FRGPI risk score group.

We then performed Kaplan–Meier curves between high- and low-FRGPI groups. All three curves showed that the low-FRGPI group had a significantly better prognosis than the high-FRGPI group ($p < 0.01$, **Figures 3A–C**).

To further validate FRGPI as a prognostic biomarker, the low-FRGPI group also had significantly better prognoses than the high-FRGPI group for 2 and 5 years ($p < 0.05$, **Supplementary Figures S2A–F**). In the early-stage LUAD, all the Kaplan–Meier curves showed that the low-FRGPI group had a significantly better prognosis ($p < 0.01$, **Supplementary Figures S2G–I**). Additionally, ever-smokers with low FRGPI scores owned a better prognosis ($p < 0.05$, **Supplementary Figure S3**). Overall, FRGPI successfully stratified the risk of LUAD patients in all the training and validation sets.

Validation of FRGPI as an Independent Prognostic Factor

Univariate regression results showed that FRGPI was statistically significant in the training and validation cohorts ($p < 0.01$, **Figures 3D–F**). The multivariate regression results showed that FRGPI was an independent prognostic factor in all three cohorts ($p < 0.05$, **Figures 3G–I**). All univariate and multivariate results revealed that the high-FRGPI group matched with a worse

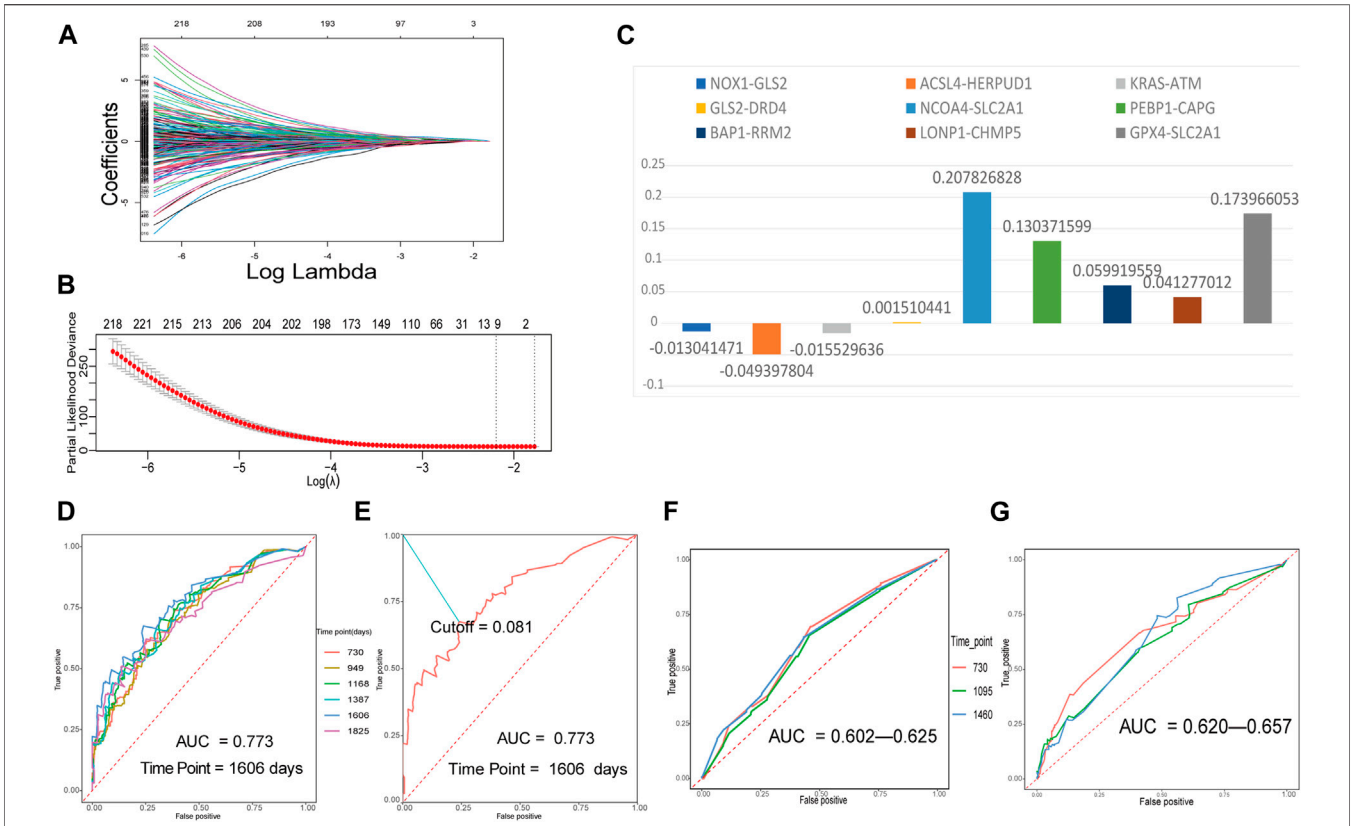


FIGURE 2 | Building a ferroptosis-related gene-pair index (FRGPI) in the training set and verifying FRGPI in the ROC curves. **(A)** The diagram of lambda and coef. **(B)** Performing 10-fold CV to calculate the partial likelihood deviance corresponding to different models. The deviance was the smallest when nine gene pairs were included. The minimum of lambda was 0.1115. **(C)** Our prognostic FRGPI was made up of nine FRGPs and corresponding coef values. **(D)** The time-dependent ROC curves of the training set. **(E)** We defined the optimal cut-off 0.081 in the training set curve with the best AUC and "1606" as the time point. **(F,G)** Time-dependent ROC curves of validation cohorts.

prognosis, with the HR and 95% confidence interval HR of FRGPI more than 1.

In the multivariate regression, it was shown that our FRGPI could be as good as stage III vs. stage I in stratifying patients. Patients whose cancer has progressed to stage III, especially stage IIIB, could hardly get radical surgical therapy. The median PFS of these patients is about 10 months (Ryan et al., 2019).

Analysis of Differentially Expressed Ferroptosis-Related Genes and *MKI67* Between High- and Low-FRGPI Groups

Our FRGPI included 16 ferroptosis-related genes. In addition, SLC7A11 as the main target of FINs was added. We compared the expression of 17 ferroptosis-related genes and *MKI67* between high- and low-FRGPI groups ($p < 0.05$, **Figures 4A–C**). In training and validation cohorts at the same time, solute carrier family 2 member 1 (*SLC2A1*), gelsolin-like actin-capping protein (*CAPG*), ribonucleotide reductase regulatory subunit M2 (*RRM2*), *SLC7A11*, and *MKI67* were significantly up-regulated in the high-FRGPI group. On the contrary, we found *GLS2* and phosphatidylethanolamine-binding protein 1 (*PEBP1*) were down-regulated in the high-FRGPI group in all three cohorts. The high-FRGPI group was marked with significantly higher *MKI67* expression and exhibited higher cancer proliferation potential.

GSEA Based on High- and Low-FRGPI Groups

FRGPI separated patients into high and low groups. The differences of enriched GSEA–KEGG pathways between the two groups are shown in **Figures 4D–F**. There was no enriched KEGG pathway in the low-FRGPI group in the training and validation cohorts at the same time. Conversely, we found that "KEGG CELL CYCLE," "KEGG DNA REPLICATION," "KEGG PROTEASOME," and "KEGG P53 SIGNALING PATHWAY" were significantly enriched in the high-FRGPI group in all three cohorts ($p < 0.05$, **Figures 4D–F**). We inferred that the consistently enriched pathways in the high-FRGPI group in part played important roles in the worse prognosis.

High-FRGPI Group Presented Higher M1 Macrophage Infiltration

The differences of infiltration levels of 22 kinds of immune cells are shown in **Figures 4G–I**. We found that M1 macrophages were

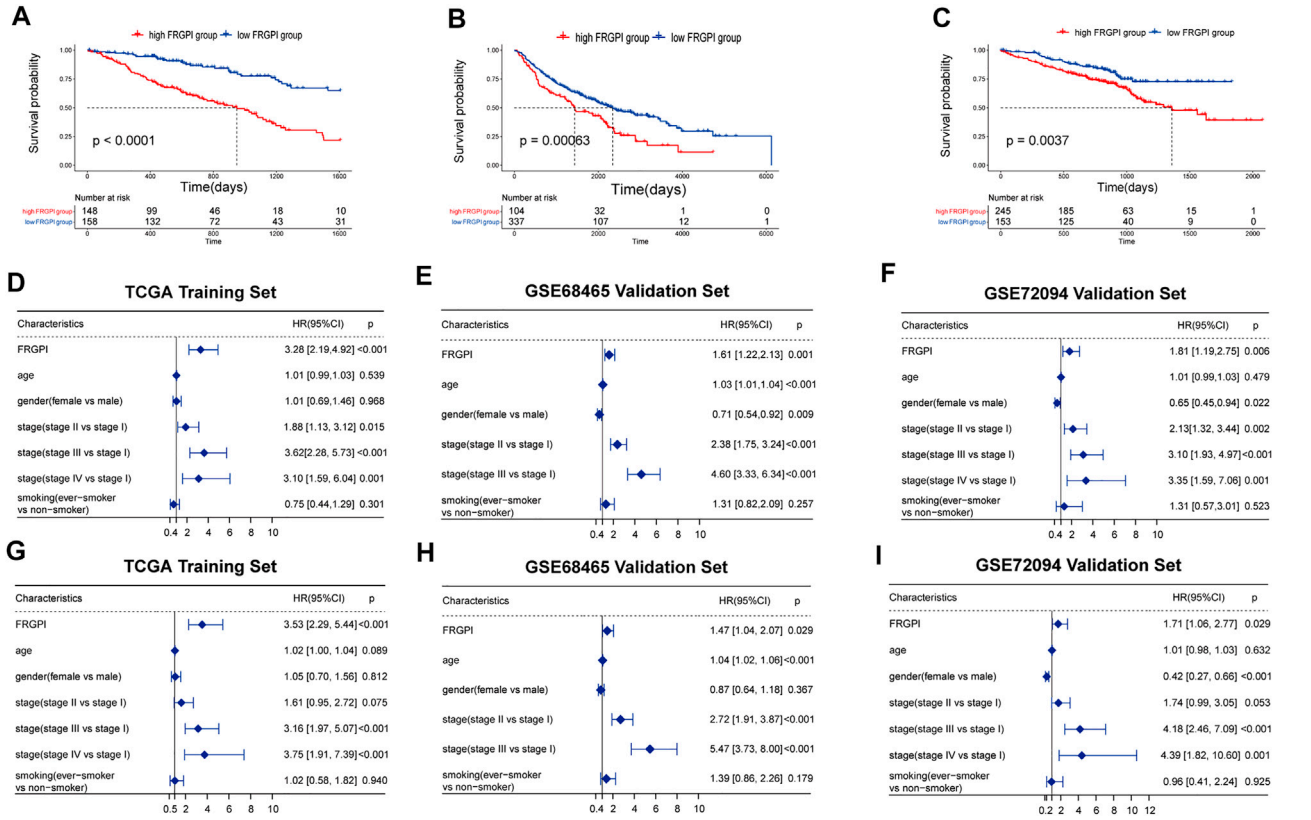


FIGURE 3 | Verification of FRGPI as an independent prognostic biomarker. **(A–C)** Kaplan–Meier curves between high- and low-risk FRGPI groups in training and validation sets. **(D–F)** Univariate Cox analyses of the three cohorts. **(G–I)** Multivariate Cox analyses of the three cohorts.

significantly up-regulated in the high-FRGPI group in all three cohorts; however, activated $CD4^+$ cells, $CD8^+$ cells, dendritic cells, plasma cells, and natural killer cells did not present different infiltration levels. There were few differences in adaptive immune responses between high- and low-FRGPI groups.

DISCUSSION

We built one prognostic model in lung adenocarcinoma consisting of nine ferroptosis-related gene pairs. The 9 gene pairs include 16 individual genes which participate in multiple crucial molecular mechanisms of ferroptosis and tumorigenesis. Then, we found that *SLC2A1*, *PEBP1*, *CAPG*, *RRM2*, *SLC7A11*, and *GLS2* differentially expressed between high- and low-FRGPI groups in all three datasets. Down-regulation of *SLC2A1* can suppress the progression of lung adenocarcinoma (Wang et al., 2017). PEBP1 binds to ALOX15, which is essential for ferroptosis, to promote lipid peroxidation and induce ferroptosis (Wenzel et al., 2017). CAPG and RRM2, could inhibit ferroptosis after stimulation of erastin (Zhang et al., 2019).

Both glutaminase 1 (GLS1, kidney type) and glutaminase 2 (GLS2, liver type) catalyze the conversion of glutamine into glutamate. Nevertheless, only GLS2 is involved in the up-regulation of ferroptosis by inhibiting the production of GPX4 and promoting downstream lipid ROS manufacture (Lukey et al., 2019; Tang et al., 2021). Meanwhile, increased nuclear translocation of GLS2 has been reported to stop the cell cycle at the G2/M stage to prevent proliferation (El-Deiry, 2016). Overexpression of *GLS2* in human lung, liver, and colon cancer cells has been proved to induce significant inhibitions in tumor growth and proliferation (Suzuki et al., 2010). Therefore, as shown in our KEGG results, "KEGG CELL CYCLE" was not enriched in the group in which *GLS2* was up-regulated. Overall, up-regulation of *GLS2* in the low-FRGPI group might be associated with promoting cancer ferroptosis and preventing tumor proliferation.

The GSEA–KEGG results revealed that the four pathways, "KEGG CELL CYCLE," "KEGG DNA REPLICATION," "KEGG PROTEASOME," and "KEGG P53 SIGNALING PATHWAY," were enriched in the high-FRGPI group. "Cell cycle" and "DNA replication" gene sets were associated with cell proliferation and cancer aggressiveness. At the same time, the high-FRGPI group was also marked with higher *MKI67* expression. The high-FRGPI group suffered more risk of tumor progression and might have a worse prognosis.

As previously observed, a high infiltration level of M1 macrophages might be associated with a better survival outcome in NSCLC patients (Ma et al., 2010). Contradictorily, in our study, the high-M1/FRGPI group demonstrated a worse

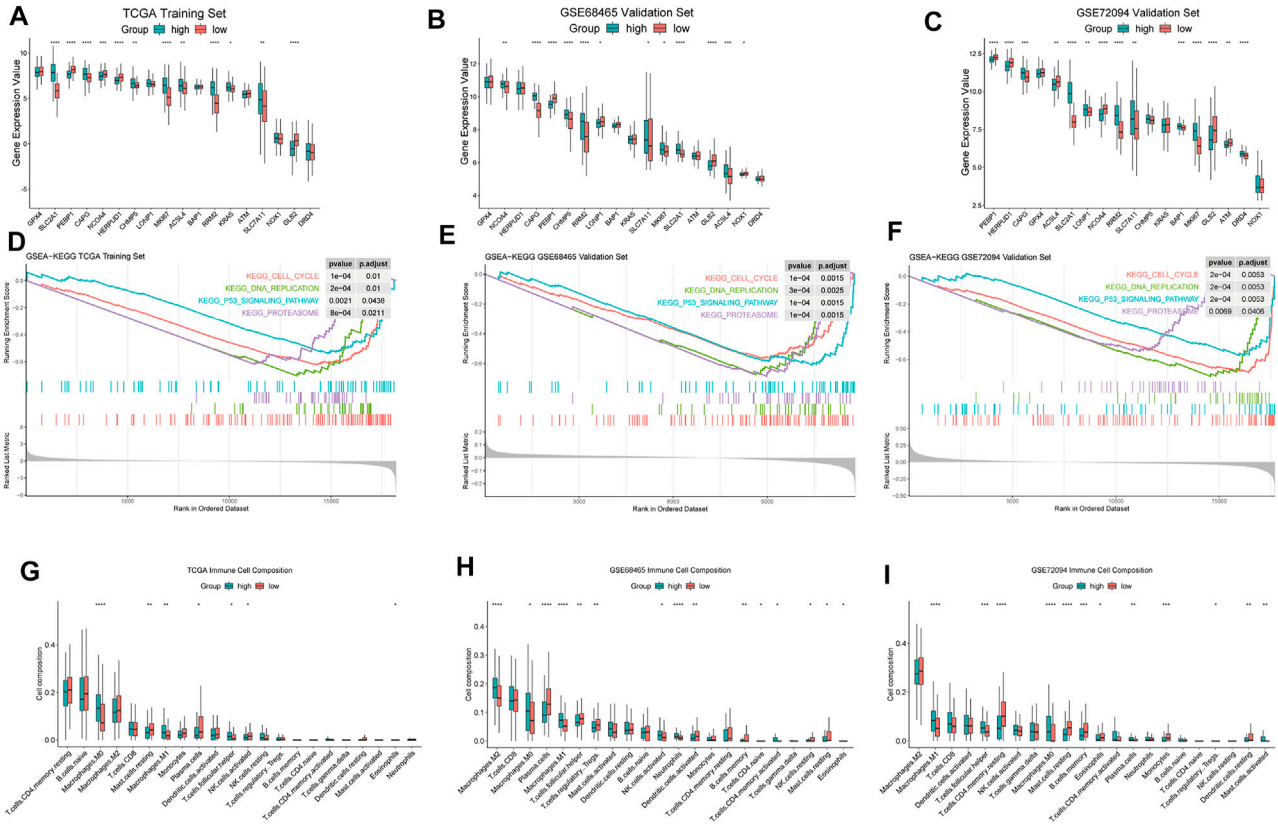


FIGURE 4 | Differences of biological characteristics between high- and low-FRGPI groups. We used the following convention for symbols indicating statistical significance: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, and ****: $p < 0.0001$. **(A–C)** ferroptosis-related genes and *MKI67* for the expression level comparisons in the three cohorts. **(D–F)** Enriched GSEA–KEGG pathways in the three cohorts. **(G–I)** Infiltration levels of 22 kinds of immune cells in the three cohorts.

outcome in all three datasets. This, at least, was not a coincidence. Recent reports in breast cancer showed that "M1" high tumors were definitely associated with more aggressive clinical features (Lu and Ma, 2020; Oshi et al., 2020). In our research, we also detected that the high-M1/FRGPI group presented a higher *MKI67* expression and enriched "cell cycle" and "DNA replication" gene sets, which meant that the high-M1/FRGPI group might have more aggressive cancer cells and advanced cell proliferation. Meanwhile, our CIBERSORT results revealed that the high-M1/FRGPI group did not present favorable immune activities to fight with aggressive cancer cells. Overall, the anti-cancer tumor immune microenvironment could not counterbalance the biologically aggressive features of the high-M1/FRGPI group, possibly leading to the worse survival outcome of the high-M1/FRGPI group.

We detected that *SLC7A11* was up-regulated in the high-FRGPI group. Cystine transporter xCT encoded by *SLC7A11* exports intracellular glutamate and imports extracellular cystine for glutathione biosynthesis and downstream GPX4 to reduce lipid peroxidation. Class I FINs aim at inhibiting SLC7A11 activity to trigger ferroptosis in cancer cells. In addition, class I FINs, such as erastin, synergistically enhance the anti-tumor effect of classical cisplatin chemotherapy in NSCLC (Guo et al., 2018). Superabundant antioxidants in cancer build a huge obstacle to radiotherapy. Class I FINs deplete GPX4 to promote lipid peroxidation and enhance radio sensitivity. In treating lung adenocarcinoma, erastin and x-ray irradiation reinforce each other (Shibata et al., 2019). In brief, the up-regulation of *SLC7A11* in the high-FRGPI group with a worse prognosis indicated one treatment strategy to particularly inhibit SLC7A11 and activate ferroptosis. The strategy focused on using class I FINs to inhibit SLC7A11, induce ferroptosis, and synergistically work together with traditional chemotherapy and radiotherapy for the high-FRGPI group to gain a better prognosis.

In conclusion, we built a robust gene-pair prognostic model of lung adenocarcinoma on the basis of ferroptosis mechanisms. We applied this model to stratify patients into low- and high-FRGPI groups. Moreover, we explored the differences of the biological pathways and tumor immune microenvironment between high- and low-FRGPI groups. Previously, based on candidate prognostic genes, Li et al. (2019), Al-Dherasi et al. (2021), and Liu et al. (2021) identified prognostic signatures in lung adenocarcinoma. These models included 16 genes, 7 genes,

and 4 genes, respectively. Liu et al. (2020) identified a 14-gene signature in lung adenocarcinoma based on differentially expressed genes between tumor and normal tissues. However, this signature was short of validations in independent datasets. Liang et al. (2021) also established one 7-gene prognostic model based on ferroptosis-related genes in lung adenocarcinoma. The AUC of 5-year survival was 0.709 in Liang et al.'s training cohort. Our signature utilized gene pairs to overcome technical problems regarding the comparison between different datasets. The AUC of our signature was 0.773 in the training cohort. In addition to the KEGG pathway enrichment analysis in Liang et al.'s model, we focused on the differentially expressed genes and tumor immune microenvironment. We supposed SLC7A11 as the target of FINs might have something to do with outcomes and guide us to trigger ferroptosis during traditional treatments for better prognoses.

Although we verified FRGPI as an independent prognostic biomarker, our study exposed limitations. Our FRGPI was based on large-scale network datasets and lacked additional local patient data. We used ferroptosis-related gene pairs as FRGPI's components to avoid normalization, but still faced complex intrinsic and extrinsic interference factors which might affect FRGPI accuracy. Meanwhile, the specific functions and biological pathways of genes and gene pairs in FRGPI need further investigation. Moreover, prospective studies and experiments are required for further validations of FRGPI and careful considerations of FRGPI for individual therapies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LL and BW were responsible for the study design. LL was involved in the analysis, interpretation of the data, and manuscript drafting. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of the patients and researchers involved in TCGA and GEO. They thank Subei People's Hospital of Jiangsu Province for the support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.841712/full#supplementary-material

REFERENCES

Al-Dherasi, A., Huang, Q.-T., Liao, Y., Al-Mosaib, S., Hua, R., Wang, Y., et al. (2021). A Seven-Gene Prognostic Signature Predicts Overall Survival of Patients with Lung Adenocarcinoma (LUAD). *Cancer Cel. Int.* 21 (1), 294. doi:10.1186/s12935-021-01975-z

Arbour, K. C., and Riely, G. J. (2019). Systemic Therapy for Locally Advanced and Metastatic Non-small Cell Lung Cancer. *JAMA* 322 (8), 764–774. doi:10.1001/jama.2019.11058

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.Cd-12-0095

Davis, S., and Meltzer, P. S. (2007). GEOquery: a Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14), 1846–1847. doi:10.1093/bioinformatics/btm254

Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., et al. (2012). Ferroptosis: an Iron-dependent Form of Nonapoptotic Cell Death. *Cell* 149 (5), 1060–1072. doi:10.1016/j.cell.2012.03.042

El-Deiry, W. S. (2016). p21(WAF1) Mediates Cell-Cycle Inhibition, Relevant to Cancer Suppression and Therapy. *Cancer Res.* 76 (18), 5189–5191. doi:10.1158/0008-5472.Can-16-2055

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., et al. (2020). *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer. Available at: https://gco.iarc.fr/today.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics* 20 (3), 307–315. doi:10.1093/bioinformatics/btg405

Guo, J., Xu, B., Han, Q., Zhou, H., Xia, Y., Gong, C., et al. (2018). Ferroptosis: A Novel Anti-tumor Action for Cisplatin. *Cancer Res. Treat.* 50 (2), 445–460. doi:10.4143/crt.2016.572

Homma, T., Kobayashi, S., and Fujii, J. (2019). Induction of Ferroptosis by Singlet Oxygen Generated from Naphthalene Endoperoxide. *Biochem. Biophysical Res. Commun.* 518 (3), 519–525. doi:10.1016/j.bbrc.2019.08.073

N. Howlader, A. M. Noone, M. Krapcho, D. Miller, A. Brest, M. Yu, et al. (2020) (Editors). (based on November 2019 SEER data submission, posted to the SEER web site, April 2020) *SEER Cancer Statistics Review, 1975-2017* (Bethesda, MD: National Cancer Institute). Available: https://seer.cancer.gov/csr/1975_2017/.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2016). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi:10.1093/nar/gkw1092

Jacobsen, A. (2015). cgdsr: R-Based API for Accessing the MSKCC Cancer Genomics Data Server (CGDS). R Package Version, 1 (5). Available at: https://CRAN.R-project.org/package=cgdsr (Accessed August 15, 2021).

Li, Y., Ge, D., Gu, J., Xu, F., Zhu, Q., and Lu, C. (2019). A Large Cohort Study Identifying a Novel Prognosis Prediction Model for Lung Adenocarcinoma through Machine Learning Strategies. *BMC Cancer* 19 (1), 886. doi:10.1186/s12885-019-6101-7

Liang, M., Chen, M., Zhong, Y., Singh, S., and Singh, S. (2021). Construction of a Prognostic Model in Lung Adenocarcinoma Based on Ferroptosis-Related Genes. *Front. Genet.* 12, 739520. doi:10.3389/fgene.2021.739520

Liu, J., Hou, S., Wang, J., Chai, Z., Hong, X., Zhao, T., et al. (2020). A Prognostic 14-Gene Expression Signature for Lung Adenocarcinoma: A Study Based on TCGA Data Mining. *Oxid Med. Cel. Longev.* 2020, 8847226. doi:10.1155/2020/8847226

Liu, L., He, H., Peng, Y., Yang, Z., and Gao, S. (2021). A Four-Gene Prognostic Signature for Predicting the Overall Survival of Patients with Lung Adenocarcinoma. *PeerJ* 9, e11911. doi:10.7717/peerj.11911

Lu, J., and Ma, L. (2020). The Role of Tumor-Associated Macrophages in the Development, Metastasis and Treatment of Breast Cancer. *Pathol. - Res. Pract.* 216 (9), 153085. doi:10.1016/j.prp.2020.153085

Lukey, M. J., Cluntun, A. A., Katt, W. P., Lin, M.-c. J., Druso, J. E., Ramachandran, S., et al. (2019). Liver-Type Glutaminase GLS2 Is a Druggable Metabolic Node

in Luminal-Subtype Breast Cancer. *Cel Rep.* 29 (1), 76–88. doi:10.1016/j.celrep.2019.08.076

Ma, J., Liu, L., Che, G., Yu, N., Dai, F., and You, Z. (2010). The M1 Form of Tumor-Associated Macrophages in Non-small Cell Lung Cancer Is Positively Associated with Survival Time. *BMC Cancer* 10 (1), 112. doi:10.1186/1471-2407-10-112

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Oshi, M., Tokumaru, Y., Asaoka, M., Yan, L., Satyananda, V., Matsuyama, R., et al. (2020). M1 Macrophage and M1/M2 Ratio Defined by Transcriptomic Signatures Resemble Only Part of Their Conventional Clinical Characteristics in Breast Cancer. *Sci. Rep.* 10 (1), 16554. doi:10.1038/s41598-020-73624-w

Ryan, K. J., Skinner, K. E., Fernandes, A. W., Punekar, R. S., Pavilack, M., Walker, M. S., et al. (2019). Real-world Outcomes in Patients with Unresected Stage III Non-small Cell Lung Cancer. *Med. Oncol.* 36 (3), 24. doi:10.1007/s12032-019-1249-1

Shibata, Y., Yasui, H., Higashikawa, K., Miyamoto, N., and Kuge, Y. (2019). Erastin, a Ferroptosis-Inducing Agent, Sensitized Cancer Cells to X-ray Irradiation via Glutathione Starvation *In Vitro* and *In Vivo*. *PLoS One* 14, e0225931. doi:10.1371/journal.pone.0225931

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* 39 (5), 1–13. doi:10.18637/jss.v039.i05

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Suzuki, S., Tanaka, T., Poyurovsky, M. V., Nagano, H., Mayama, T., Ohkubo, S., et al. (2010). Phosphate-activated Glutaminase (GLS2), a P53-Inducible Regulator of Glutamine Metabolism and Reactive Oxygen Species. *Proc. Natl. Acad. Sci.* 107, 7461–7466. doi:10.1073/pnas.1002459107

Tang, D., Chen, X., Kang, R., and Kroemer, G. (2021). Ferroptosis: Molecular Mechanisms and Health Implications. *Cell Res.* 31 (2), 107–125. doi:10.1038/s41422-020-00441-1

Wang, Y., Shi, S., Ding, Y., Wang, Z., Liu, S., Yang, J., et al. (2017). Metabolic Reprogramming Induced by Inhibition of SLC2A1 Suppresses Tumor Progression in Lung Adenocarcinoma. *Int. J. Clin. Exp. Pathol.* 10 (11), 10759–10769.

Weiss, H. L., Niwas, S., Grizzle, W. E., and Piyathilake, C. (2003). Receiver Operating Characteristic (ROC) to Determine Cut-Off Points of Biomarkers in Lung Cancer Patients. *Dis. Markers* 19 (6), 273–278. doi:10.1155/2004/856026

Wenzel, S. E., Tyurina, Y. Y., Zhao, J., St. Croix, C. M., Dar, H. H., Mao, G., et al. (2017). PEBP1 Wardens Ferroptosis by Enabling Lipoxygenase Generation of Lipid Death Signals. *Cell* 171 (3), 628–641. doi:10.1016/j.cell.2017.09.044

Wickham, H.. (2009). *ggplot2-Elegant Graphics for Data Analysis (use R!)*. New York: Springer

Wickham, H. (2007). Reshaping Data with the Reshape Package. *J. Stat. Softw.* 21 (12), 1–20. doi:10.18637/jss.v021.i12

Wu, Y., Yu, C., Luo, M., Cen, C., Qiu, J., Zhang, S., et al. (2020). Ferroptosis in Cancer Treatment: Another Way to Rome. *Front. Oncol.* 10, 571127. doi:10.3389/fonc.2020.571127

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhai, X., Jing, X., Li, J., Tian, Y., Xu, S., Wang, M., et al. (2020). Clinical Outcomes for PD-1 Inhibitor Plus Chemotherapy as Second-Line or Later Therapy Compared to PD-1/pd-L1 Inhibitor Alone in Advanced Non-small-cell Lung Cancer. *Front. Oncol.* 10, 556275. doi:10.3389/fonc.2020.556275

Zhang, X., Du, L., Qiao, Y., Zhang, X., Zheng, W., Wu, Q., et al. (2019). Ferroptosis Is Governed by Differential Regulation of Transcription in Liver Cancer. *Redox Biol.* 24, 101211. doi:10.1016/j.redox.2019.101211

Zhou, N., and Bao, J. (2020). FerrDb: a Manually Curated Resource for Regulators and Markers of Ferroptosis and Ferroptosis-Disease Associations. *Database* 2020, baaa021. doi:10.1093/database/baaa021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fgene.2022.838869

Identification of lncRNA/circRNA-miRNA-mRNA ceRNA Network as Biomarkers for Hepatocellular Carcinoma

*Shanshan Chen, Yongchao Zhang, Xiaoyan Ding and Wei Li**

Cancer Center, Beijing Ditan Hospital, Capital Medical University, Beijing, China

Background: Hepatocellular carcinoma (HCC) accounts for the majority of liver cancer, with the incidence and mortality rates increasing every year. Despite the improvement of clinical management, substantial challenges remain due to its high recurrence rates and short survival period. This study aimed to identify potential diagnostic and prognostic biomarkers in HCC through bioinformatic analysis.

Methods: Datasets from GEO and TCGA databases were used for the bioinformatic analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were carried out by WebGestalt website and clusterProfiler package of R. The STRING database and Cytoscape software were used to establish the protein-protein interaction (PPI) network. The GEPIA website was used to perform expression analyses of the genes. The miRDB, miRWalk, and TargetScan were employed to predict miRNAs and the expression levels of the predicted miRNAs were explored *via* OncomiR database. LncRNAs were predicted in the StarBase and LncBase while circRNA prediction was performed by the circBank. ROC curve analysis and Kaplan-Meier (KM) survival analysis were performed to evaluate the diagnostic and prognostic value of the gene expression, respectively.

Results: A total of 327 upregulated and 422 downregulated overlapping DEGs were identified between HCC tissues and noncancerous liver tissues. The PPI network was constructed with 89 nodes and 178 edges and eight hub genes were selected to predict upstream miRNAs and ceRNAs. A lncRNA/circRNA-miRNA-mRNA network was successfully constructed based on the ceRNA hypothesis, including five lncRNAs (DLGAP1-AS1, GAS5, LINC00665, TYMSOS, and ZFAS1), six circRNAs (hsa_circ_0003209, hsa_circ_0008128, hsa_circ_0020396, hsa_circ_0030051, hsa_circ_0034049, and hsa_circ_0082333), eight miRNAs (hsa-miR-150-5p, hsa-miR-19b-3p, hsa-miR-23b-3p, hsa-miR-26a-5p, hsa-miR-651-5p, hsa-miR-10a-5p, hsa-miR-214-5p and hsa-miR-486-5p), and five mRNAs (CDC6, GINS1, MCM4, MCM6, and MCM7). The ceRNA network can promote HCC progression *via* cell cycle, DNA replication, and other pathways. Clinical diagnostic and survival analyses demonstrated that the ZFAS1/hsa-miR-150-5p/GINS1 ceRNA regulatory axis had a high diagnostic and prognostic value.

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

Reviewed by:
Wei Lan,
Guangxi University, China
Sayed Haidar Abbas Raza,
Northwest A&F University, China

****Correspondence:***
Wei Li
weili8989@ccmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *18 December 2021*
Accepted: *24 February 2022*
Published: *21 March 2022*

Citation:
Chen S, Zhang Y, Ding X and Li W (2022) Identification of lncRNA/circRNA-miRNA-mRNA ceRNA Network as Biomarkers for Hepatocellular Carcinoma. Front. Genet. 13:838869. doi: 10.3389/fgene.2022.838869

Conclusion: These results revealed that cell cycle and DNA replication pathway could be potential pathways to participate in HCC development. The ceRNA network is expected to provide potential biomarkers and therapeutic targets for HCC management, especially the ZFAS1/hsa-miR-150-5p/GINS1 regulatory axis.

Keywords: bioinformatics analysis, hepatocellular carcinoma, competing endogenous RNA network, biomarker, diagnosis, prognosis

INTRODUCTION

As one of the most common malignant tumors of digestive system, hepatocellular carcinoma (HCC) is characterized by diverse etiology, high incidence, and poor prognosis. According to statistics, both its incidence and mortality rates are the highest in Asia (Singal et al., 2020). Due to the lack of obvious symptoms, the majority of HCC patients are diagnosed at advanced stages, with limited treatment options. Despite progress has been made in diagnosis and therapy during the last decades, the recurrence and metastasis rates remain high (Ghouri et al., 2017). To confront the current situation, it is crucial to develop new strategies for screening and monitoring of HCC. Alpha-fetoprotein (AFP) is currently the most widely used biomarker for HCC diagnosis, but its sensitivity and specificity are still not satisfactory (Daniele et al., 2004). In recent years, the development of various sequencing platforms and bioinformatics technologies has facilitated the identification of many novel biomarkers. Public databases have provided us with rich and diverse data resources, and we can further improve our understanding of HCC by integrating data from different sources. Although several biomarkers have been reported relating to the diagnosis and prognosis of HCC, such as alpha-fetoprotein lens culinaris agglutin-3 (AFP-L3), des-γ-carboxy prothrombin (DCP), glypican-3 (GPC3), and so on, their practical applications are yet to be evaluated (Piñero et al., 2020). With the vigorous development of bioinformatics, many biomarkers have been identified to be associated with the development of HCC, but few of them have been proven to be of practical use. The number of reliable tumor biomarkers that can be used for the early detection and prognostic assessment is still small in clinical practice. Thus, identification of novel potential biomarkers is warranted, which may contribute to update of diagnostic techniques and improvement of therapeutic efficacy.

It is well known that non-coding RNAs (ncRNAs) account for the vast majority of the human transcriptome, mainly including microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs) (Chan and Tay, 2018). MicroRNA (miRNA) is a class of evolutionarily conserved non-coding RNA (ncRNA) with 18–25 nucleotides in length (Kloosterman and Plasterk, 2006). It participates in a series of physiological and pathological processes *via* mediating the post-transcriptional regulation of target genes (Krol et al., 2010). Aberrant expression of numerous miRNAs has been linked to cancer initiation and progression (Lee and Dutta, 2009). Long non-coding RNA (lncRNA) is known as a type of ncRNA whose length exceeds 200 nucleotides. Studies have revealed that lncRNAs play an important role in cancer development through a variety of mechanisms (Bhan et al., 2017). Circular RNA (circRNA) is a newly discovered ncRNA with a closed-loop structure. Compared with the traditional linear RNA, circRNA is more resistant to RNA exonuclease, without terminal 5′ caps and 3′ polyadenylated tails, and thus more stable (Jeck and Sharpless, 2014). CircRNAs have been confirmed to exert effects on regulating cellular metabolism in cancer (Yu T et al., 2019). In recent years, increasing numbers of researchers have dedicated themselves to exploring the biological functions of ncRNAs. Various computational methods have also been developed for the prediction of potential associations between ncRNAs and disease, which is of critical importance for the identification of biomarkers (Lan et al., 2020; Chen et al., 2021).

The competing endogenous RNA (ceRNA) hypothesis was first put forward by Salmena et al., in 2011 (Salmena et al., 2011). CeRNA is a class of ncRNA that can competitively bind shared miRNAs and cross-regulate each other at the post-transcription level. In the cytoplasm, lncRNA and circRNA can serve as miRNA sponges by common miRNA response elements (MREs) and indirectly regulate the downstream target genes (Taulli et al., 2013; Yao et al., 2019). This ceRNA-based regulatory mechanism has been discovered in multiple cancers. For example, lncRNA HOTAIR regulated HER2 expression through competition for miR-331-3p, thereby facilitating tumor development (Liu et al., 2014). LncRNA H19 was reported to exerted oncogenic functions in gallbladder cancer *via* modulating miR-342-3p and FOXM1 (Wang et al., 2016). In addition, circRNA ciRS-7 could act as the "super sponge" of microRNA-7 (miR-7) and inhibits the activity of miR-7 (Peng et al., 2015). Accumulating evidence has indicated that ceRNA regulation network serves a role in biological processes of HCC development, such as proliferation, metastasis, epithelial to mesenchymal transition (EMT), and chemotherapy resistance (Wu et al., 2018; Liu Z et al., 2019; Huang et al., 2020; Song et al., 2020).

The ceRNA network provides new perspectives for improving diagnosis and treatment for HCC. Even though several ceRNAs have been found associated with HCC progression (Bai et al., 2019; Guo et al., 2019; Wang et al., 2019; Yu J et al., 2019), our current understanding of ceRNA regulatory network in HCC is still very limited. Further exploration is needed to unravel unknown functions and mechanisms of related ceRNA. Sequencing data used in this study were collected from public databases. Starting from the differentially expressed messenger RNAs (mRNAs), we inversely predicted the targeted miRNAs and their relevant lncRNAs and circRNAs, and then constructed a comprehensive ceRNA network. Bioinformatics tools were applied to analyze and discuss the crucial pathways as well as

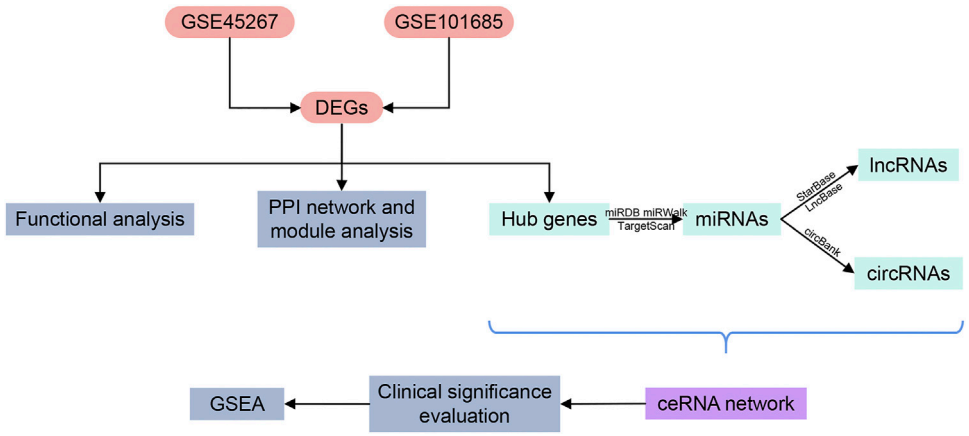


FIGURE 1 | Flowchart of the research design. DEGs, differentially expressed genes; PPI, protein-protein interaction; miRNAs, microRNAs; lncRNAs, long non-coding RNAs; circRNAs, circular RNAs; ceRNA, competing endogenous RNA; GSEA, gene set enrichment analysis.

the diagnostic performance and prognostic value of the key genes. LncRNAs, circRNAs, miRNAs and targeted mRNAs engaged in the ceRNA network may become potential diagnostic biomarkers and therapeutic targets for HCC. The research process is shown in **Figure 1**.

MATERIALS AND METHODS

Screening of Differentially Expressed Genes

Gene Expression Omnibus (GEO) database[1] was searched to obtain appropriate gene expression datasets. The GSE45267 dataset (containing 48 HCC samples and 39 non-cancerous samples) and the GSE101685 dataset (containing 24 HCC samples and eight non-cancerous samples) were selected. Differentially expressed genes (DEGs) between HCC tissue and normal liver tissue were identified *via* GEO2R online tools (Barrett et al., 2013). The adjusted p-value < 0.01 and the value of log-fold change $|\log FC| \geq 1$ were set as DEGs cutoff criteria. The visualization of the DEGs was shown on the heat maps and volcano plots, performed by ComplexHeatmap (Gu et al., 2016) and ggplot2 packages, respectively. The VennDiagram (Chen and Boutros, 2011) package of R was used to screen the common DEGs and construct the Venn diagram.

Functional Enrichment Analysis

WEB-based GEne SeT AnaLysis Toolkit (WebGestalt)[2] (Liao et al., 2019) is a powerful online tool for functional enrichment analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were carried out by WebGestalt for the upregulated and downregulated DEGs, respectively. GO enrichment analyses include biological process (BP), cellular component (CC), and molecular function (MF). False discovery rate (FDR) < 0.05 was considered statistically significant. The ggplot2 package of R was adopted to visualize the results of functional enrichment analyses.

Protein-Protein Interaction Network Construction and Module Analysis

The STRING database[3] (Szklarczyk et al., 2021) was used to obtain the interaction relationships among DEGs. The minimum required interaction score was set to high confidence (0.7). Then the PPI network of DEGs was constructed by Cytoscape 3.7.2 (Shannon et al., 2003). Molecular Complex Detection (MCODE), a plugin in Cytoscape, was applied to screen significant modules in the network. The advanced options were set as degree cutoff = 2, haircut, node score cutoff = 0.2, k-core = 2, and max depth = 100. Subsequently, the KEGG pathway enrichment analyses of genes in the key modules were performed by the clusterProfiler (Yu et al., 2012) package of R. The p-value of less than 0.05 was regarded as significant.

Identification and Verification of Hub Genes

Another plugin in Cytoscape, cytoHubba (Chin et al., 2014), was used to identify hub genes in the network. The top ten nodes ranked by the MCC algorithm were considered as hub genes. The KEGG pathway enrichment analysis was also performed on the hub genes *via* the clusterProfiler package of R. Gene Expression Profiling Interactive Analysis (GEPIA) database[4] is an online website that can provide customizable analyses based on TCGA and GTEx data (Tang et al., 2017). This database was used to perform expression analyses of the hub genes. The $|\text{Log2FC}|$ cutoff was set to 1, and the p-value cutoff was set to 0.01. The KEGG pathway enrichment analysis of hub genes was also conducted by the clusterProfiler package of R.

[1]https://www.ncbi.nlm.nih.gov/geo/
[2]http://www.webgestalt.org
[3]https://string-db.org/
[4]http://gepia.cancer-pku.cn/

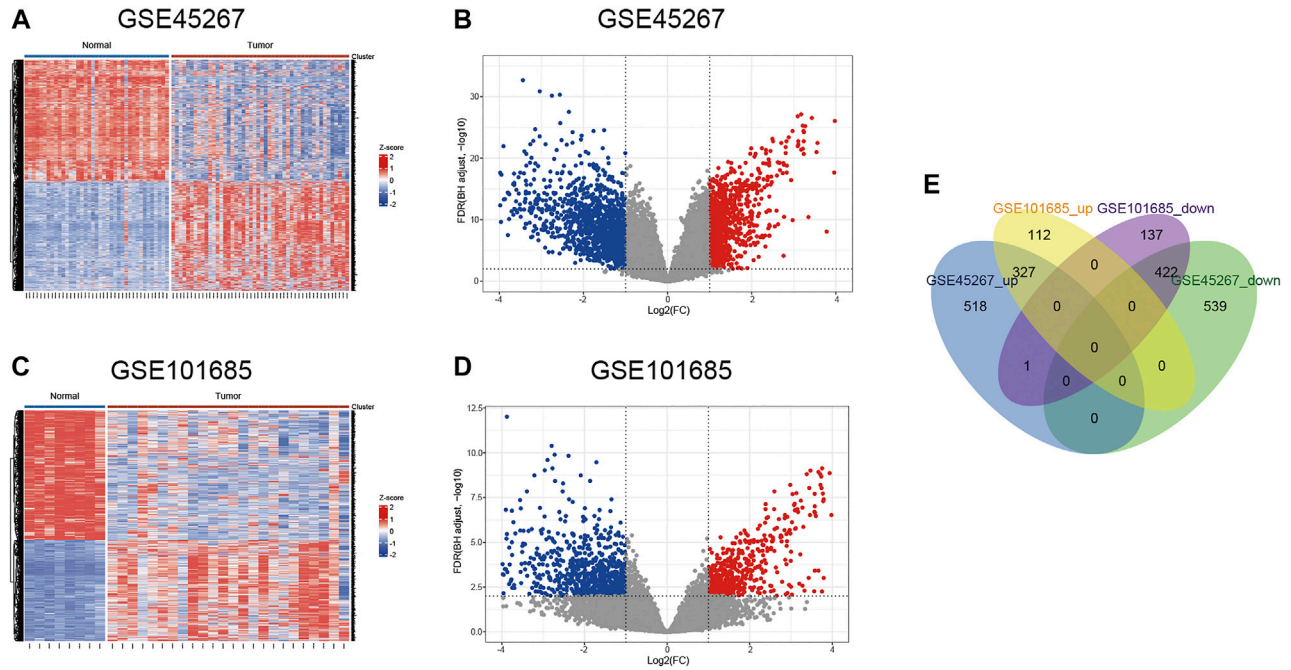


FIGURE 2 | Identification of differentially expressed genes. **(A,B)** Heat map and volcano plot of the GSE45267 dataset. **(C,D)** Heat map and volcano plot of the GSE101685 dataset. **(E)** Venn diagram of the common upregulated and downregulated DEGs. DEGs, differentially expressed genes.

Identification of miRNAs

Three databases, miRDB[5] (Chen and Wang, 2020), miRWalk[6] (Sticht et al., 2018), and TargetScan[7] (Agarwal et al., 2015) were applied to predict the upstream miRNAs for the hub genes based on the regulatory associations. The VennDiagram package of R was used to obtain the intersection between the predicted sets, which enhanced the reliability of the final results. Then Cytoscape software was used to construct a miRNA-mRNA network. The expression levels of the predicted miRNAs were explored *via* OncomiR[8] (Wong et al., 2018) database. p-value < 0.05 was considered to be statistically significant.

Identification of lncRNAs and circRNAs and ceRNA Network Construction

The ceRNA network was constructed based on the interaction relationships among lncRNAs, circRNAs, miRNAs, and mRNAs. To be noted, the ceRNA hypothesis suggests that the expression level of ceRNA should be negatively correlated with miRNA expression and positively correlated with mRNA expression. Therefore, we integrate the predicted relationships and the corresponding expression data to obtain more reliable results. Potential lncRNAs interacted with miRNAs were predicted by the intersection of StarBase[9] (Li J. H et al., 2014) and LncBase[10] (Paraskevopoulou et al., 2016) databases. The GEPIA database was then used to obtain the expression data of the targeted lncRNAs and screen out differentially expressed lncRNAs. CircRNA prediction was performed by the circBank[11] (Liu M et al., 2019) database. The GSE97332 dataset and the GSE164803 dataset were selected to screen the common differentially expressed circRNAs. The identification of final predicted circRNAs was based on the intersection of the predicted group and the differentially expressed group. $p < 0.05$ was considered statistically significant. Ultimately, the lncRNA/circRNA-miRNA-mRNA network was visualized by the ggalluvial package of R and the Cytoscape software.

Diagnostic and Prognostic Analysis of Key Genes

To assess the clinical significance of the key genes in the ceRNA network, we performed diagnostic and prognostic analyses. The expression profiles of mRNA, miRNA, and lncRNA between HCC samples and normal samples were collected from TCGA[12] and the expression profiles of circRNA were obtained from the GSE97332 dataset and the GSE164803 dataset. The pROC (Robin et al., 2011) package of R was utilized to assess the diagnostic value of the genes *via* performing the receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC) ≥ 0.7 was considered to indicate good discriminatory performance.

[5]http://mirdb.org/
[6]http://mirwalk.umm.uni-heidelberg.de/
[7]http://www.targetscan.org/vert_72/
[8]http://www.oncomir.org/
[9]http://starbase.sysu.edu.cn/
[10]http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2/index-predicted
[11]http://www.circbank.cn/
[12]https://portal.gdc.cancer.gov/

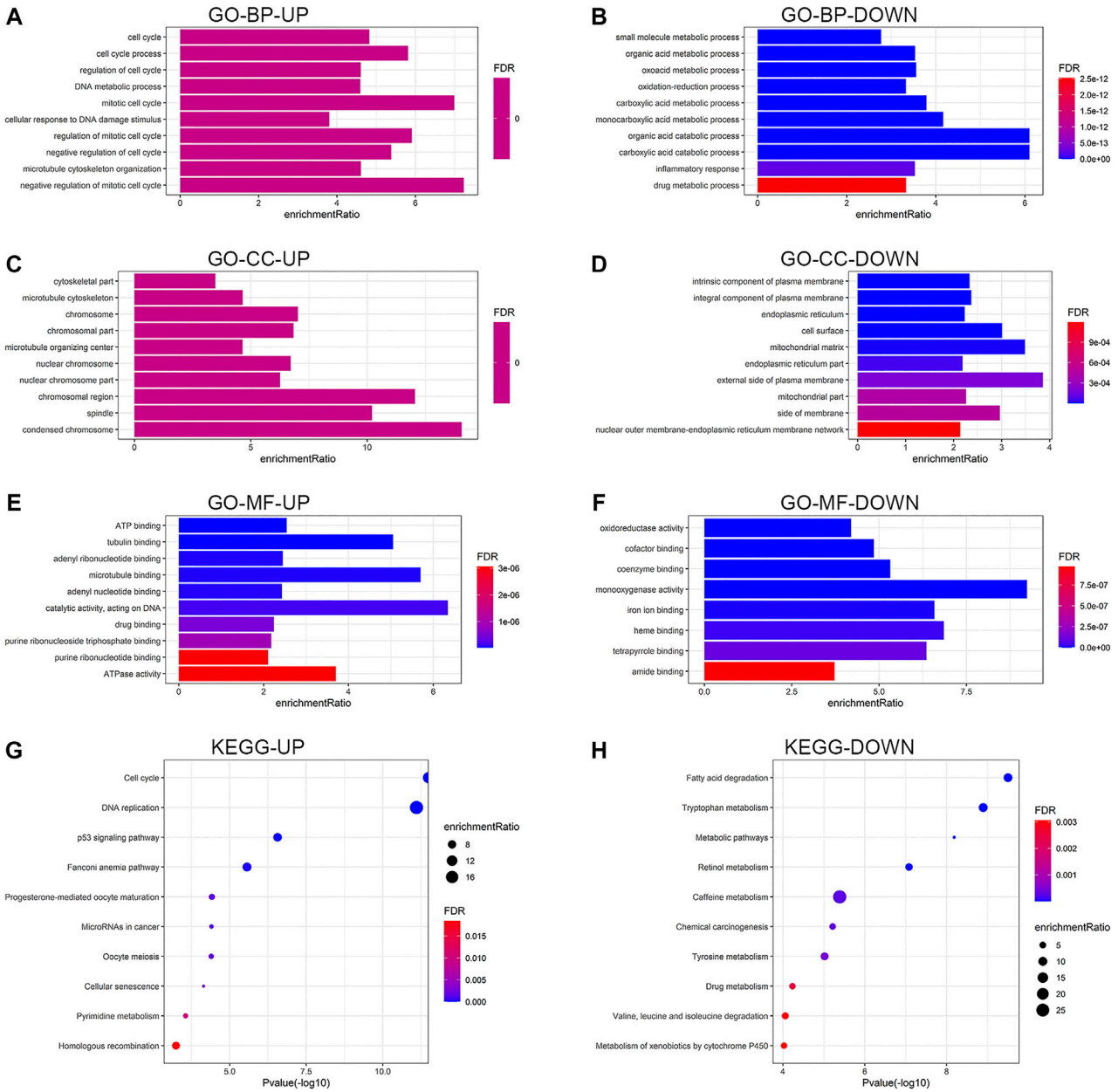


FIGURE 3 | Functional enrichment analysis of the upregulated and downregulated DEGs. Results of GO enrichment analysis of the DEGs, including BP **(A,B)**, CC **(C,D)**, and MF **(E,F)**. **(G,H)** Results of KEGG pathway analysis of the DEGs. DEGs, differentially expressed genes; GO, Gene Ontology; BP, biological process; CC, cellular component, MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes; UP, upregulated; DOWN, downregulated; FDR, false discovery rate.

The Kaplan-Meier (KM) survival analysis was performed to evaluate the prognostic value of the gene expression. The GEPIA online website provided functions to analyze the correlation between gene expression of mRNA and lncRNA and survival of HCC patients. The OncoLnc[13] was employed to explore the relationship of miRNA expression with HCC prognosis. Values of $p < 0.05$ were considered significant.

Gene Set Enrichment Analysis

Data from TCGA was divided into low and high expression groups according to the median value of core genes with great clinical significance. Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was conducted by GSEA software to investigate important pathways associated with the selected

[13]http://www.oncolnc.org/

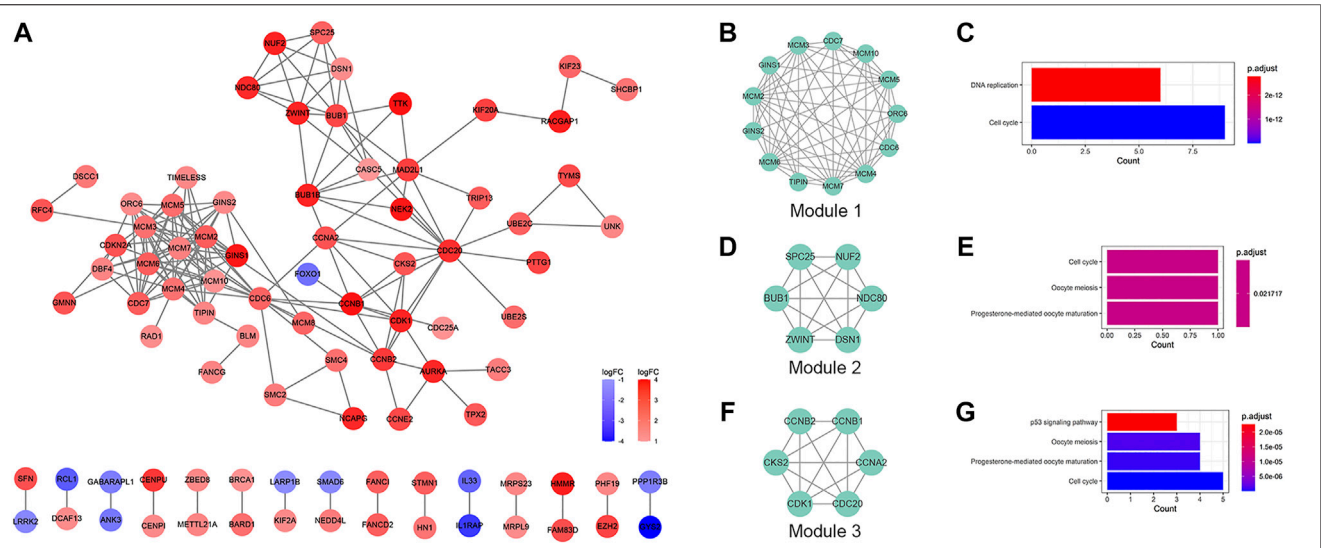


FIGURE 4 | PPI network and module analysis. **(A)** PPI network of the DEGs. Red represents the upregulated DEGs, and blue represents the downregulated DEGs. **(B,D,F)** The selected modules of the PPI network. **(C,E,G)** KEGG pathway analysis of the modules. PPI, protein-protein interaction; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes.

mRNA. Significant pathways were identified by requiring the false discovery rate (FDR) < 0.05.

RESULTS

Identification of Differentially Expressed Genes Between Hepatocellular Carcinoma Tissues and Noncancerous Liver Tissues

According to the pre-set parameters, 1807 DEGs were screened from the GSE45267 dataset, including 846 upregulated and 961 downregulated genes (**Figures 2A,B**). There were 999 DEGs obtained from the GSE101685 dataset, containing 439 upregulated and 560 downregulated genes (**Figures 2C,D**). As shown in **Figure 2E**, a total of 327 upregulated and 422 downregulated overlapping DEGs were identified between HCC tissues and noncancerous liver tissues.

Functional Enrichment Analysis of Upregulated and Downregulated Differentially Expressed Genes

Functional enrichment analysis was performed on the common DEGs. **Figures 3A–H** show the enriched GO functions and KEGG pathways for the upregulated and downregulated DEGs. The upregulated DEGs were mainly enriched in cell cycle, DNA metabolic process, cellular response to DNA damage stimulus, and microtubule cytoskeleton organization in the BP category; condensed chromosome, chromosomal region, and spindle in the cellular component category; catalytic activity, microtubule binding, tubulin binding, and ATPase activity in the molecular function category (**Figures 3A,C,E**). For the downregulated DEGs, the enriched GO terms were inflammatory response, oxidation-reduction process, and various metabolic processes in the BP category; external side of plasma membrane, mitochondrial matrix, cell surface, and side of membrane in the cellular component category; oxidoreductase activity, monooxygenase activity, heme binding, iron ion binding, and tetrapyrrole binding in the molecular function category (**Figures 3B,D,F**). As shown in **Figures 3G,H**, the enriched KEGG pathways for the upregulated DEGs mainly included cell cycle, DNA replication, and p53 signaling pathway, while the downregulated DEGs were highly related to various metabolic pathways. Detailed enrichment results are listed in **Supplementary Table S1**.

TABLE 1 | The top ten genes ranked as hub genes by the MCC algorithm.

Rank	Name	Score
1	MCM2	17570
2	MCM7	17544
3	MCM4	17521
4	MCM6	17281
5	MCM5	16584
6	MCM3	15985
7	MCM10	10104
8	CDC7	6024
9	CDC6	5777
10	GINS1	5760

Protein-Protein Interaction Network, Molecular Complex Detection Analysis and Hub Gene Identification

Based on the STRING database and the Cytoscape software, a PPI network was constructed, consisting of 89 nodes and 178 edges (**Figure 4A**). As indicated in **Figure 4A**, the majority of the genes that met the filter settings in STRING database were

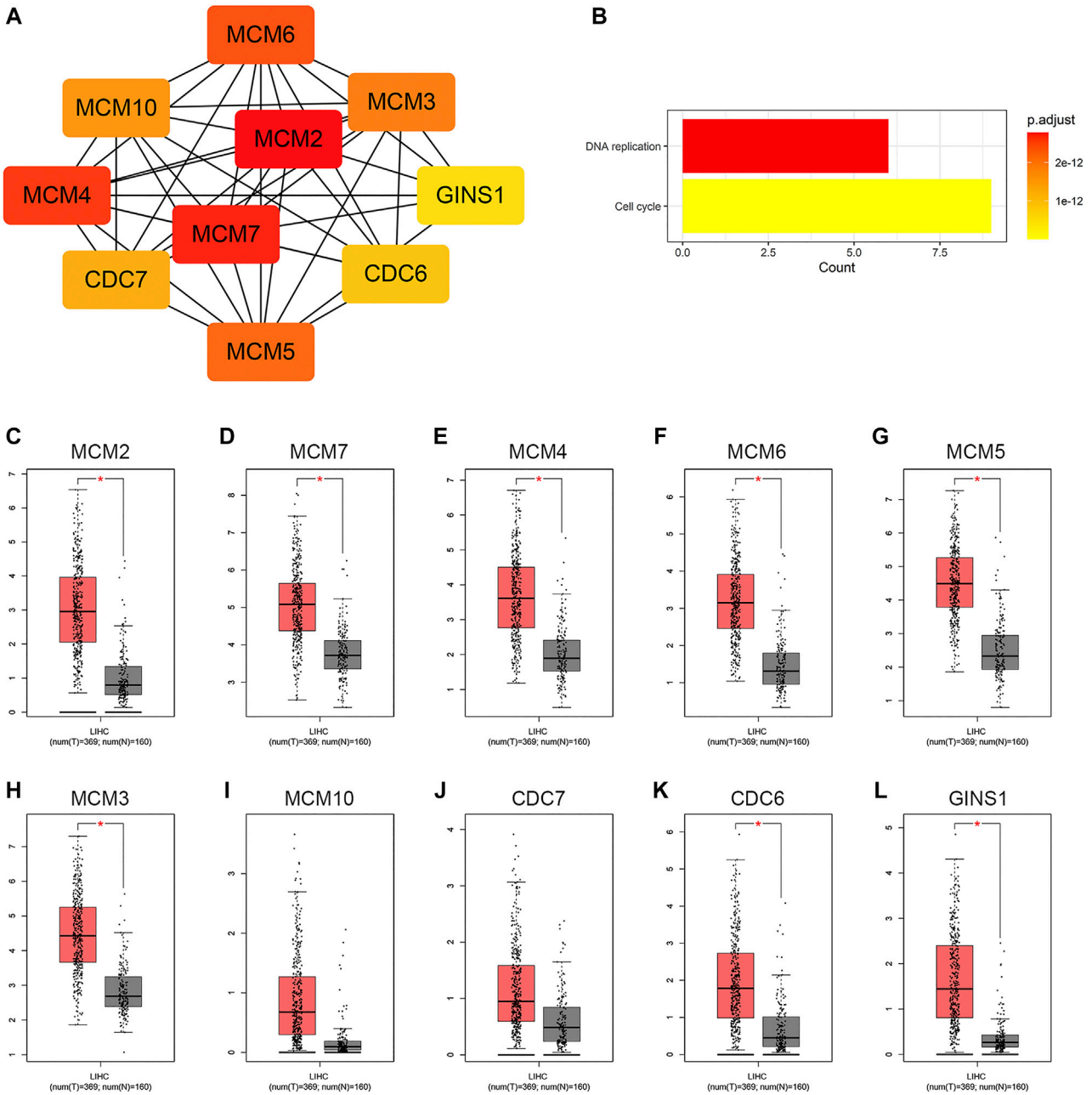


FIGURE 5 | Identification of hub genes in the PPI network. **(A)** Hub gene network. **(B)** KEGG pathway analysis of the hub genes. **(C–L)** Differential expression analysis of the hub genes in GEPIA. Tumor tissue is shown in red, and normal tissue is shown in gray. PPI, protein-protein interaction; KEGG, Kyoto Encyclopedia of Genes and Genomes; GEPIA, Gene Expression Profiling Interactive Analysis; LIHC, liver hepatocellular carcinoma; T, tumor; N, normal.

upregulated. There were three modules selected after MCODE analysis (**Figures 4B,D,F**). With the highest score of 10.167, module 1 was significantly enriched in cell cycle and DNA replication (**Figure 4C**). Furthermore, module 2 was primarily associated with cell cycle, oocyte meiosis, and progesterone-mediated oocyte maturation (**Figure 4E**). Similarly, the most significant enrichment pathways of module 3 were cell cycle, oocyte meiosis, p53 signaling pathway, and progesterone-mediated oocyte maturation (**Figure 4G**). As important components of module 1, the ten genes (MCM2, MCM7, MCM4, MCM6, MCM5, MCM3, MCM10, CDC7, CDC6, and GINS1) were also ranked as hub genes of the whole

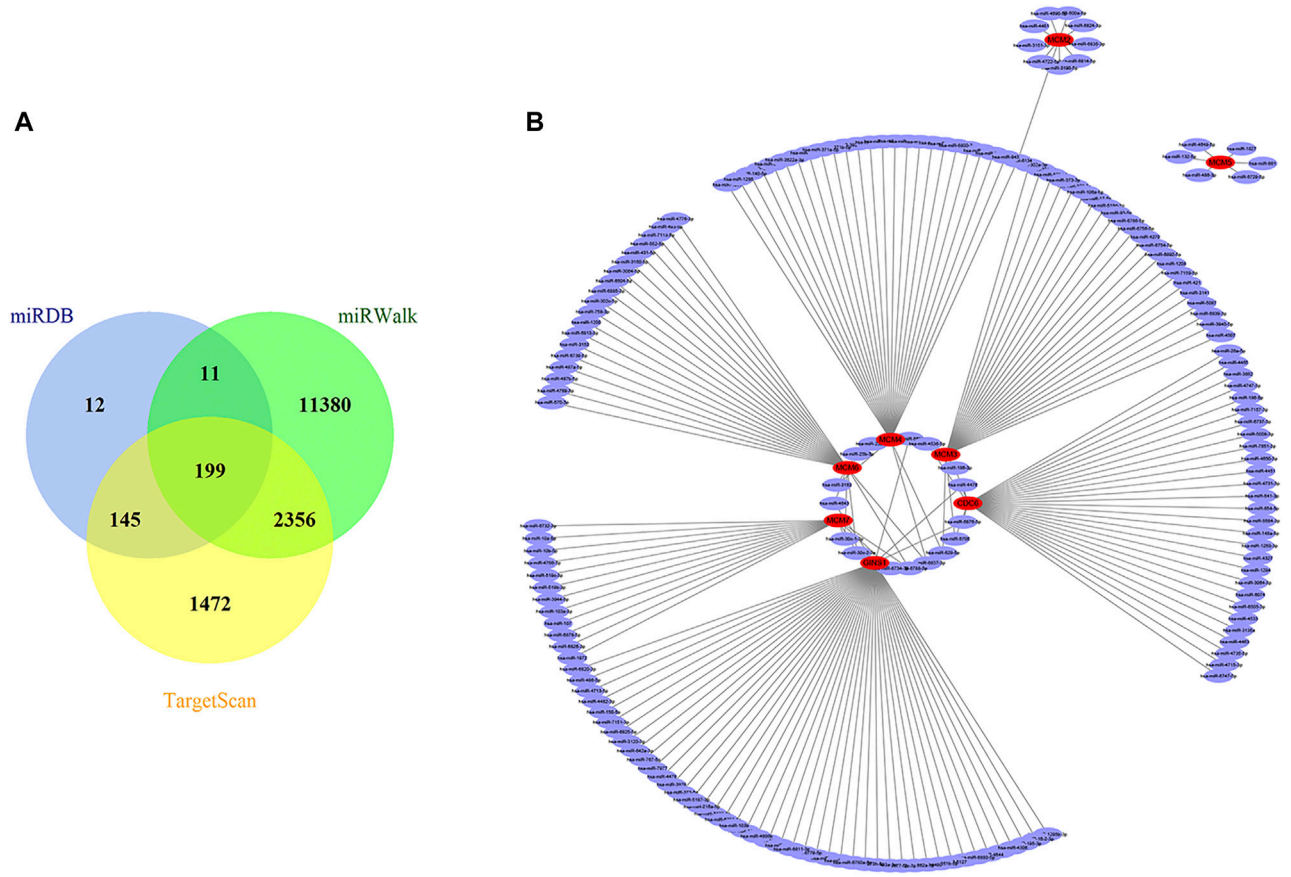


FIGURE 6 | Prediction of targeted miRNAs. **(A)** Venn diagram of the common predicted miRNAs. **(B)** The miRNA-mRNA network constructed based on the predicted relationship. Red represents the mRNAs, and purple represents the miRNAs. MiRNAs, microRNAs; mRNAs, messenger RNAs.

TABLE 2 | Differential expression analysis of the predicted miRNAs in OncomiR.

miRNA name	Cancer abbreviation	T-test *p*-value	Tumor Log2 mean expression	Normal Log2 mean expression
hsa-miR-1258	LIHC	4.44E-24	3.13E-21	0.59
hsa-miR-486-5p	LIHC	1.49E-12	6.08E-11	6.48
hsa-miR-30c-2-3p	LIHC	2.25E-10	3.87E-09	5.12
hsa-miR-214-5p	LIHC	7.94E-10	1.19E-08	2.4
hsa-miR-758-3p	LIHC	1.91E-09	2.37E-08	2.58
hsa-miR-10a-5p	LIHC	1.04E-08	1.02E-07	13.44
hsa-miR-30c-1-3p	LIHC	2.40E-08	2.11E-07	0.78
hsa-miR-195-3p	LIHC	2.17E-07	1.47E-06	1.49
hsa-miR-150-5p	LIHC	6.51E-07	4.10E-06	7.72
hsa-miR-26a-5p	LIHC	2.30E-06	1.25E-05	11.39
hsa-miR-23b-3p	LIHC	4.30E-05	1.74E-04	10.52
hsa-miR-19b-3p	LIHC	7.63E-03	1.81E-02	6.64
hsa-miR-148a-5p	LIHC	9.73E-03	2.24E-02	6.48
hsa-miR-651-5p	LIHC	2.89E-02	5.81E-02	1.68
hsa-miR-16-2-3p	LIHC	4.68E-02	8.54E-02	2.61

miRNAs, microRNAs; LIHC, liver hepatocellular carcinoma.

network, enriched in cell cycle and DNA replication (**Table 1**; **Figures 5A,B**). The GEPIA database was used to verify the expression of the hub genes. Eight of the hub genes (MCM2, MCM7, MCM4, MCM6, MCM5, MCM3, CDC6, and GINS1) were confirmed to have significant differential expression according to the results (**Figures 5C–L**).

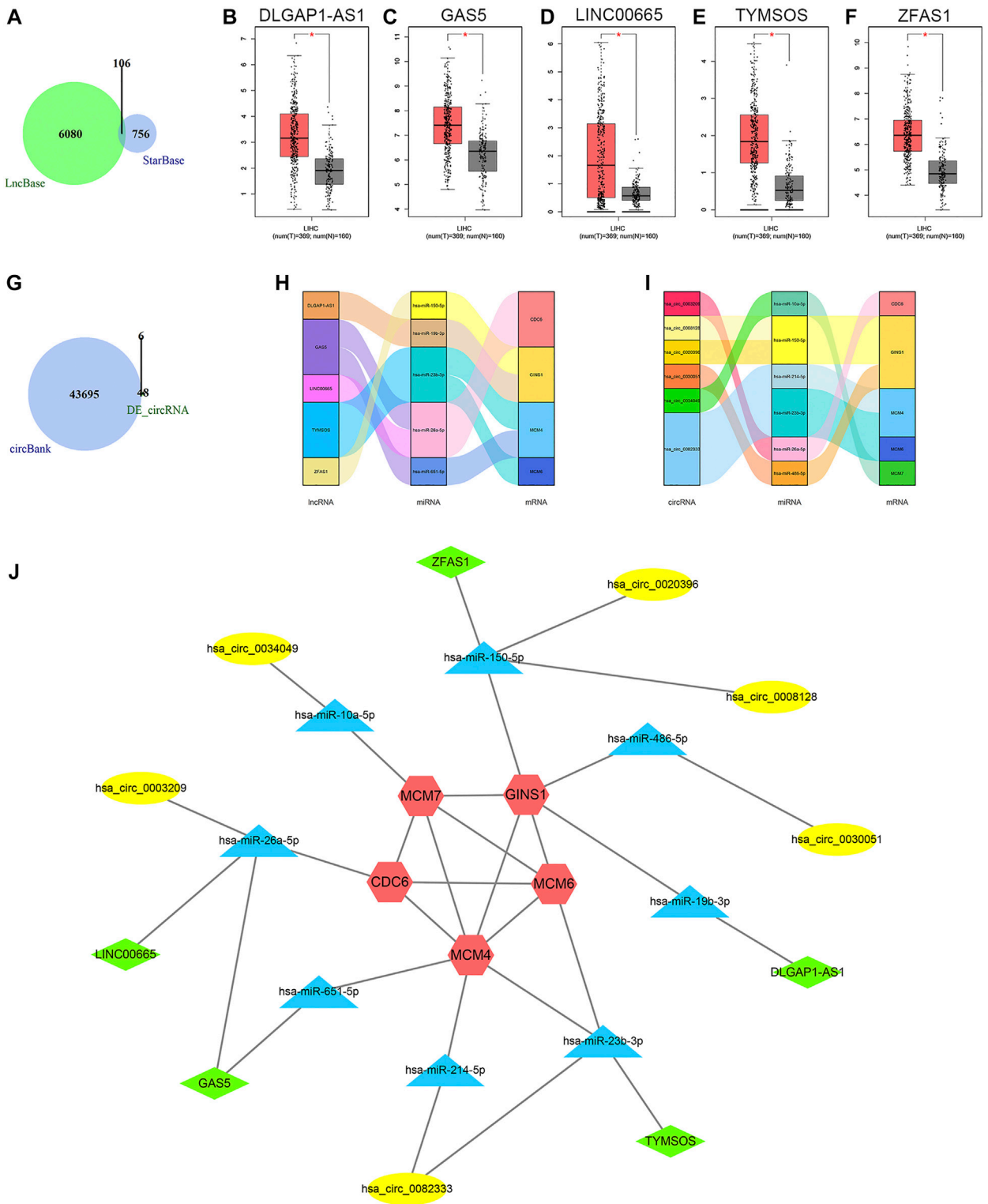


FIGURE 7 | Results of ceRNA prediction and ceRNA network construction. **(A)** Venn diagram of the common predicted lncRNAs. **(B–F)** Differential expression analysis of the lncRNAs in GEPIA. Tumor tissue is shown in red, and normal tissue is shown in gray. **(G)** Venn diagram of the predicted circRNAs and differentially expressed circRNAs. Sankey diagrams of the lncRNA-miRNA-mRNA regulatory relationship **(H)** and circRNA-miRNA-mRNA regulatory relationship **(I)**. **(J)** The ceRNA network. Red represents the mRNAs, blue represents the miRNAs, green represents the lncRNAs, and yellow represents the circRNAs. CeRNA, competing endogenous RNA; lncRNA, long non-coding RNA; GEPIA, Gene Expression Profiling Interactive Analysis; circRNA, circular RNA; miRNA, microRNA; mRNA, messenger RNA; LIHC, liver hepatocellular carcinoma; T, tumor; N, normal; DE_circRNA, differentially expressed circular RNA.

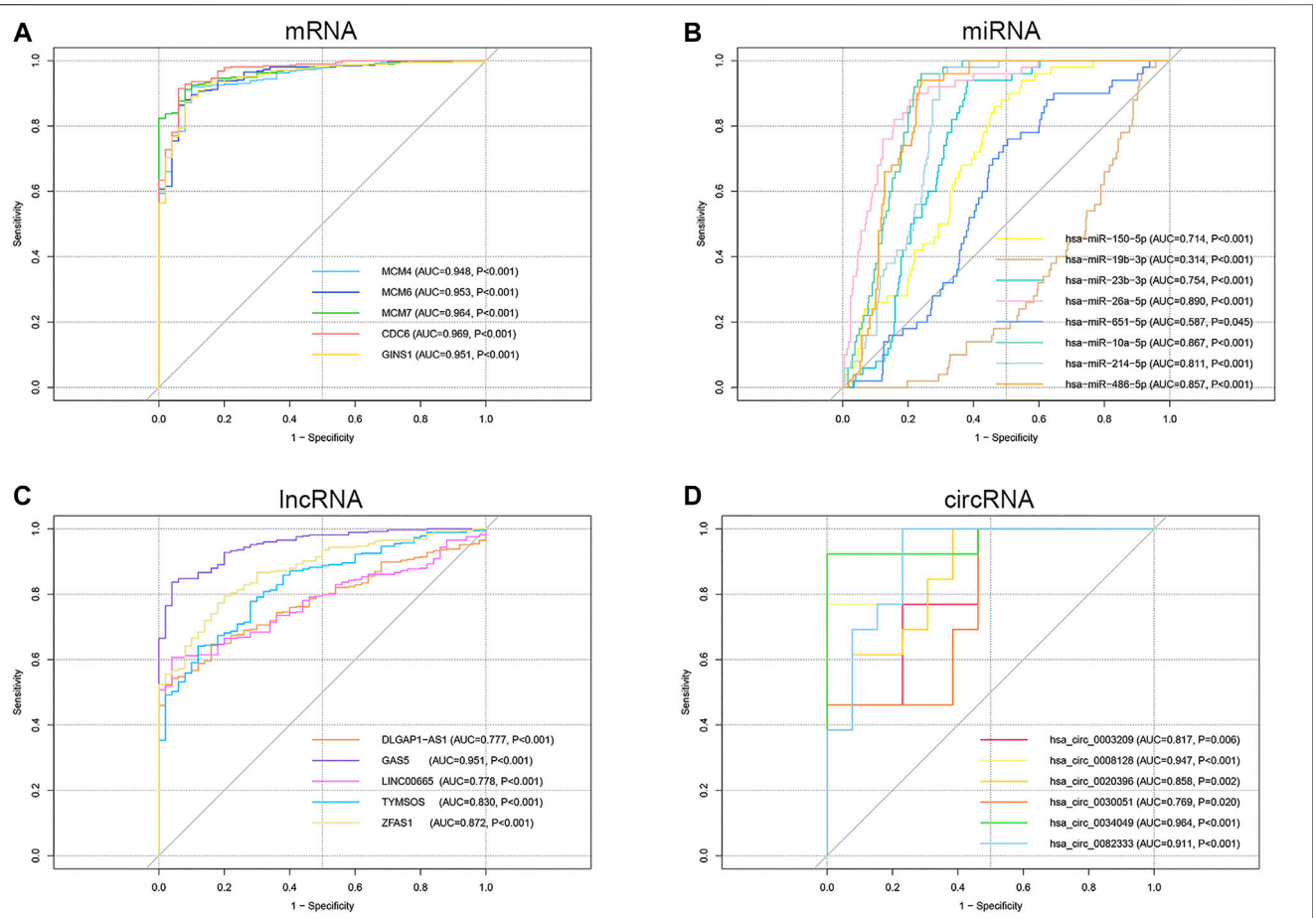


FIGURE 8 | ROC curve analysis of the mRNA **(A)**, miRNA **(B)**, lncRNA **(C)**, and circRNA **(D)** in the ceRNA network. ROC, receiver operating characteristic; mRNA, messenger RNA; miRNA, microRNA; lncRNA, long non-coding RNA; circRNA, circular RNA; ceRNA, competing endogenous RNA; AUC, area under the ROC curve.

Identification of miRNAs, lncRNAs, and circRNAs, and Construction of a ceRNA Network

A total of 199 overlapped upstream miRNAs related to the above eight key genes were predicted by searching miRDB, miRWalk, and TargetScan databases (**Figure 6A**). According to the corresponding relationship, a miRNA-mRNA network was established as shown in **Figure 6B**. The expression data of these predicted miRNAs were obtained in OncomiR database. It was found that only 15 of them had significant differential expression (**Table 2**).

On the basis of the aforementioned miRNAs, 106 common lncRNA-miRNA regulatory pairs were obtained using StarBase and LncBase (**Figure 7A**). Five eligible lncRNAs (DLGAP1-AS1, GAS5, LINC00665, TYMSOS, and ZFAS1) were screened out after the integration with the expression data from GEPIA (**Figures 7B–F**), corresponding to five miRNAs (hsa-miR-150-5p, hsa-miR-19b-3p, hsa-miR-23b-3p, hsa-miR-26a-5p, and hsa-miR-651-5p) and four mRNAs (CDC6, GINS1, MCM4, and MCM6). The circBank database revealed 43701 circRNA-mRNA interactions. Then we selected six circRNAs (hsa_circ_0003209, hsa_circ_0008128, hsa_circ_0020396, hsa_circ_0030051, hsa_circ_0034049, and hsa_circ_0082333) by integrating the microarray data from the GSE97332 dataset and the GSE164803 dataset (**Figure 7G**), corresponding to six miRNAs (hsa-miR-10a-5p, hsa-miR-150-5p, hsa-miR-214-5p, hsa-miR-23b-3p, hsa-miR-26a-5p, and hsa-miR-486-5p) and five mRNAs (CDC6, GINS1, MCM4, MCM6, and MCM7). The lncRNA/circRNA-miRNA-mRNA regulatory relationships were shown as the Sankey diagrams in **Figure 7H** and **Figure 7I**. A total of five lncRNAs, six circRNAs, eight miRNAs, and five mRNAs constituted the ceRNA network (**Figure 7J**).

Diagnostic and Prognostic Value of Key Genes in Hepatocellular Carcinoma

ROC curve analysis was performed on the key genes in the ceRNA network (**Supplementary Tables S2–S5**). The results indicated that all the mRNAs, lncRNAs, circRNAs, and six of the miRNAs had good diagnostic value (AUC>0.7, $p < 0.05$), excluding two miRNAs hsa-miR-19b-3p (AUC = 0.314, $p < 0.001$) and hsa-miR-651-5p (AUC = 0.587, $p = 0.045$) (**Figures 8A–D**). Kaplan-Meier survival analysis demonstrated that the expression levels of all the mRNAs, one miRNA (hsa-miR-150-5p), and one lncRNA (ZFAS1) were significantly

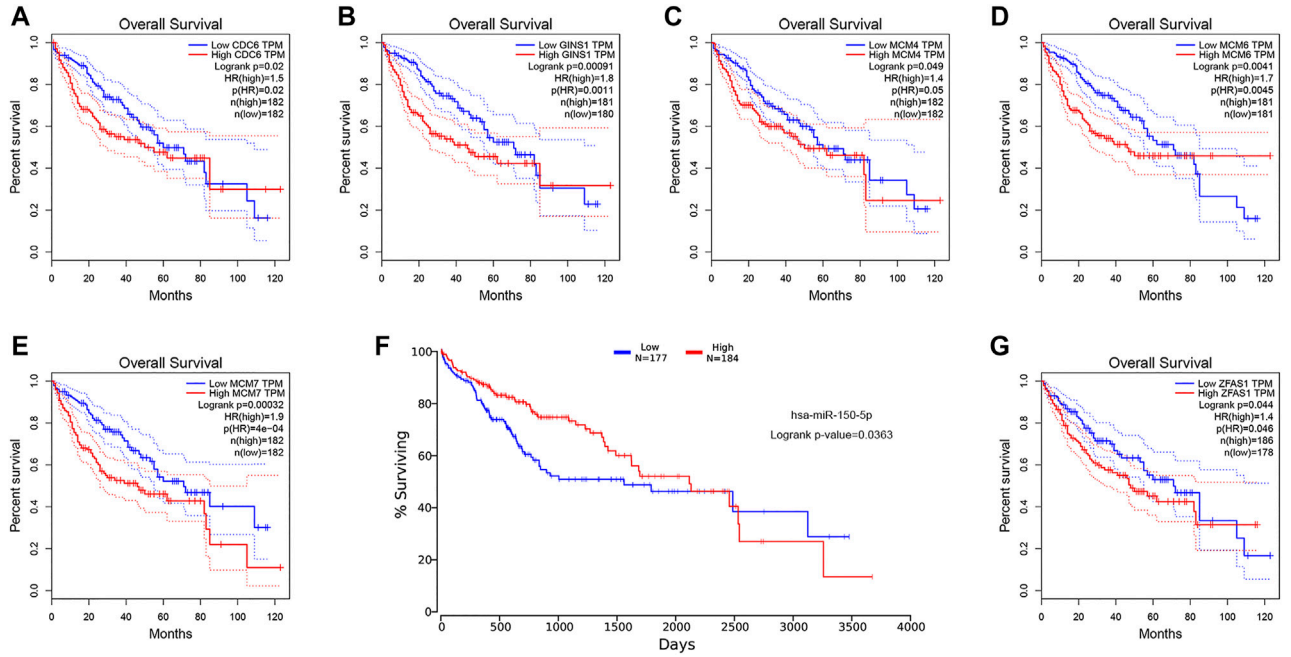


FIGURE 9 | KM survival analysis of the mRNA **(A–E)**, miRNA hsa-miR-150-5p **(F)**, and lncRNA ZFAS1 **(G)**. KM, Kaplan-Meier; mRNA, messenger RNA; miRNA, microRNA; lncRNA, long non-coding RNA; TPM, transcripts per million; HR, hazard ratio.

correlated with poor prognosis in HCC patients ($p < 0.05$) (**Figures 9A–G**). It was noticed that ZFAS1, hsa-miR-150-5p, and GINS1 formed a lncRNA-miRNA-mRNA axis that carried both diagnostic and prognostic significance.

Gene Set Enrichment Analysis Analysis of GINS1 and Construction of a Conceptual Map

GSEA analysis showed that the core gene GINS1 was remarkably related to pathways contributing to HCC development and progression, such as cell cycle, DNA replication, p53 signaling pathway, mTOR signaling pathway, Notch signaling pathway, Wnt signaling pathway, and so on (**Figure 10A**). Based on our results, lncRNA ZFAS1 could sponge hsa-miR-150-5p and upregulate the expression of GINS1 in the cytoplasm of HCC cell. Shown in **Figure 10B** was our final conceptual map. The ZFAS1/hsa-miR-150-5p/GINS1 axis might directly or indirectly impact the HCC development through the pathways in **Figure 10A**.

DISCUSSION

Amounting evidence has shown that ceRNA might play a role in cancer initiation and progression (Qi et al., 2015). In the ceRNA hypothesis (Salmena et al., 2011), the ability of ceRNA to competitively bind to miRNA can influence tumorigenesis and cancer progression *via* regulating mRNA expression. To date, several ceRNAs have been identified to have a role in HCC (Xu G et al., 2020). However, there are still many ceRNAs of potential significance that have yet to be identified and require further exploration. Through bioinformatics analysis, this study attempted to establish a lncRNA/circRNA-miRNA-mRNA network holding biological functions in HCC. Integration of various databases would help achieve more reliable results. Based on the ceRNA hypothesis, a ceRNA network was successfully constructed *via* stepwise reverse prediction from mRNA to lncRNA/circRNA. Our results are expected to provide valuable guidance for HCC management.

The minichromosome maintenance (MCM) family is mainly known for their involvement in DNA replication (Maiorano et al., 2006). Given that DNA replication is a crucial pathway in tumor development, members of the MCM family are implied to be closely related to cancer development as well (Neves and Kwok, 2017; Yu et al., 2020). The overexpression of MCMs in various cancer tissues has been demonstrated by multiple studies, and is generally connected with poor prognostic features (Giaginis et al., 2011; Peng et al., 2016; Liu et al., 2017; Issac et al., 2019). Through functional enrichment analysis, we found that the hub genes with MCMs predominating were mainly involved in cell cycle and DNA replication pathways. There is evidence suggesting that high expression of MCM4 is correlated with clinicopathological variables and prognosis of HCC and silencing MCM4 can suppress the tumorigenicity of hepatoma cells (Xu et al., 2021). Liu et al. found that knockdown of MCM6 in Huh7 cells could cause a delay in S/G2-phase progression through down-regulating the cell cycle checkpoint (Liu et al., 2018). In addition, it has been demonstrated that MCM7 promotes cancer progression through cyclin D1-dependent signaling (Qu et al., 2017). The above studies have displayed the potential of MCMs as biomarkers to engage in HCC management, which is consistent

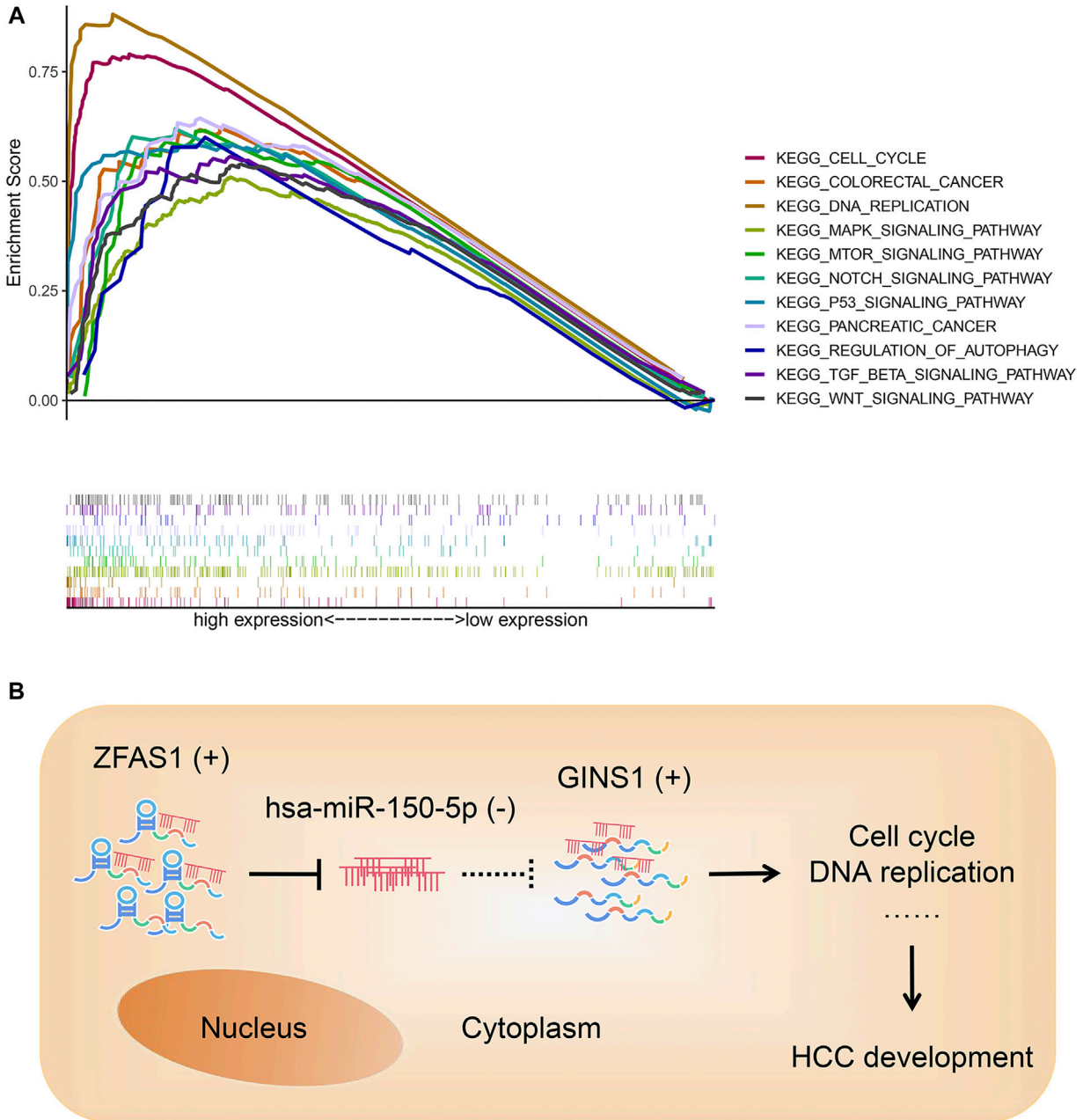


FIGURE 10 | GSEA analysis of GINS1 **(A)** and conceptual map of the mechanism of ZFAS1/hsa-miR-150-5p/GINS1 axis **(B)**. In the cytoplasm of HCC cell, lncRNA ZFAS1 could act as a sponge to bind and negatively regulate the expression of miRNA hsa-miR-150-5p. Then miR-150-5p-mediated suppression of target mRNA was relieved and GINS1 continued to exert an oncogenic role in the development of HCC. GSEA, gene set enrichment analysis; HCC, hepatocellular carcinoma; lncRNA, long non-coding RNA; miRNA, microRNA; mRNA, messenger RNA; KEGG, Kyoto Encyclopedia of Genes and Genomes; (+), upregulated; (–), downregulated.

with the results observed in this study. Known as a molecular switch, CDC6 is considered to have a transcriptional effect on E-Cadherin and subsequently affect EMT (Sideridou et al., 2011; Petrakis et al., 2012). In cancer cells, aberrant expression of CDC6 is involved in proliferation and tumor growth by modulating cell cycle (Lim and Townsend, 2020). Xu et al. revealed that CDC6 was regulated by miR-215-5p to involve in the proliferation of HCC (Xu H et al., 2020). They also found that CDC6 was negatively associated with overall and disease-free survival in HCC patients. In this study, CDC6 was modified by hsa-miR-26a-5p, which has been reported to have an effect on cell proliferation, migration, and invasion in digestive malignancies (Tian et al., 2019; Li H. H et al., 2020; Li Y et al., 2020; Zhang et al., 2020). The GINS complex, a component of the DNA replication machinery, usually participates in DNA replication through interactions with MCMs and other components (Labib and

Gambus, 2007; Yoshimochi et al., 2008). GINS1 is a subunit of the GINS complex. Previous studies have demonstrated that GINS1 is upregulated in tumor samples and correlated with poor prognosis (Bu et al., 2020a; Bu et al., 2020b; Li H et al., 2021). These studies are consistent with our findings. The five mRNAs (CDC6, GINS1, MCM4, MCM6, and MCM7) in the ceRNA network were upregulated in HCC tissues and of great value in HCC diagnosis and prognosis.

MicroRNA (miRNA), a type of single-stranded nonencoding RNA, is widely involved in tumor generation and development (Krol et al., 2010). MiRNAs regulate gene expression primarily by degrading mRNA or inhibiting its translation (Bartel, 2009). In this ceRNA network, eight miRNAs (hsa-miR-150-5p, hsa-miR-19b-3p, hsa-miR-23b-3p, hsa-miR-26a-5p, hsa-miR-651-5p, hsa-miR-10a-5p, hsa-miR-214-5p and hsa-miR-486-5p) were found underexpressed in HCC samples. They linked the ceRNAs to the target mRNAs. Several miRNAs of them have been reported to have a place in proliferation, migration, and invasion of HCC cells (Li T. et al., 2014; Pang et al., 2018; He et al., 2019; Wu et al., 2019; Zhu et al., 2020; Hayashi et al., 2021). Similarly, there have been researches that explored the functions of the five lncRNAs (DLGAP1-AS1, GAS5, LINC00665, TYMSOS, and ZFAS1) on tumorigenesis and progression of HCC or other tumors (Zhang et al., 2019; Zhou et al., 2019; Ding et al., 2020; Lin et al., 2020; Gu et al., 2021). Their expression levels were highly relevant to tumor growth, invasion, and metastasis. With the development of high-throughput sequencing and emergence of bioinformatic methods, studies have increasingly revealed the important roles of circRNAs in various tumors (Lei et al., 2020; Li R et al., 2020). Their unique covalent closed-loop structures make them advantageous for clinical application. Due to the short development period, abundant circRNAs and their functions in HCC need further excavation. Our study reversely predicted six circRNAs (hsa_circ_0003209, hsa_circ_0008128, hsa_circ_0020396, hsa_circ_0030051, hsa_circ_0034049, and hsa_circ_0082333) and all of them had reliable diagnostic value in HCC.

Notably, our study discovered a lncRNA-miRNA-mRNA axis of great clinical significance, namely ZFAS1/hsa-miR-150-5p/GINS1 ceRNA axis. Their expression levels and predicted interactions in HCC are in line with the ceRNA hypothesis. LncRNA ZFAS1 may contribute to tumorigenesis of HCC by sponging hsa-miR-150-5p and regulating the expression of the target mRNA GINS1. Moreover, the diagnosis and prognosis analysis of each gene showed favourable outcomes ($p < 0.05$). Our GSEA analysis found that the target mRNA GINS1 was significantly associated with cell cycle, DNA replication, p53 signaling pathway, mTOR signaling pathway, Notch signaling pathway, Wnt signaling pathway, etc. The majority of these pathways have been confirmed to be involved in HCC development (Greenbaum, 2004; Giovannini et al., 2016; Dimri and Satyanarayana, 2020). The previous study revealed that ZFAS1 acted as an oncogene in HCC progression by binding miR-150 and abrogating its tumor-suppressive function (Li et al., 2015). Besides, *in vitro* experiments have verified that miR-150-5p inhibition significantly promotes hepatoma cell migration and invasion (Li T et al., 2014). GINS1 was reported to be associated with tumor grades and poor survival of HCC patients (Li S et al., 2021). Furthermore, cell cycle, cell proliferation assay, and *in vivo* animal model experiment indicated that knocking down GINS1 induced in G1/S phase cell cycle arrest and decreased HCC cells proliferation (Li S et al., 2021). These studies provide excellent support for our studies.

CONCLUSION

In summary, through integrating data from a variety of databases, we successfully constructed a ceRNA network containing five lncRNAs, six circRNAs, eight miRNAs, and five mRNAs. The ceRNA network can promote HCC progression *via* cell cycle, DNA replication, and other pathways. Clinical diagnostic and survival analyses demonstrated that the ZFAS1/hsa-miR-150-5p/GINS1 ceRNA regulatory axis had a high diagnostic and prognostic value. Our findings are expected to provide potential biomarkers and therapeutic targets for HCC management. Nevertheless, our study also presents some limitations. Results in this study are based on bioinformatic predictions and further experiments and clinical practice are warranted. Additionally, since circRNAs has not been fully studied in HCC, data of circRNAs from the available datasets are limited and insufficient for prognostic analysis. We expect more research with larger sample sizes to expand and refine our conclusions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GEO database (https://www.ncbi.nlm.nih.gov/geo/) and TCGA database (https://portal.gdc.cancer.gov/).

AUTHOR CONTRIBUTIONS

SC and WL contributed to the study conception and design. YZ and XD were involved in the material collection and data analysis. SC performed the visualization of results and wrote the initial draft. WL reviewed and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study is supported by Capital's Funds for Health Improvement and Research (CFH 2020-2-2175) and Beijing Talents Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.838869/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting Effective microRNA Target Sites in Mammalian mRNAs. *Elife* 4. doi:10.7554/eLife.05005

Bai, Y., Long, J., Liu, Z., Lin, J., Huang, H., Wang, D., et al. (2019). Comprehensive Analysis of a ceRNA Network Reveals Potential Prognostic Cytoplasmic lncRNAs Involved in HCC Progression. *J. Cel Physiol* 234 (10), 18837–18848. doi:10.1002/jcp.28522

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Uupdate. *Nucleic Acids Res.* 41, D991–D995. Database issue. doi:10.1093/nar/gks1193

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell* 136 (2), 215–233. doi:10.1016/j.cell.2009.01.002

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res.* 77 (15), 3965–3981. doi:10.1158/0008-5472.Can-16-2634

Bu, F., Zhu, X., Yi, X., Luo, C., Lin, K., Zhu, J., et al. (2020). Expression Profile of GINS Complex Predicts the Prognosis of Pancreatic Cancer Patients. *Onco Targets Ther.* 13, 11433–11444. doi:10.2147/ott.S275649

Bu, F., Zhu, X., Zhu, J., Liu, Z., Wu, T., Luo, C., et al. (2020). Bioinformatics Analysis Identifies a Novel Role of GINS1 Gene in Colorectal Cancer. *Cancer Manag. Res.* 12, 11677–11687. doi:10.2147/cmar.S279165

Chan, J. J., and Tay, Y. (2018). Noncoding RNA:RNA Regulatory Networks in Cancer. *Int. J. Mol. Sci.* 19 (5), 1310. doi:10.3390/ijms19051310

Chen, H., and Boutros, P. C. (2011). VennDiagram: a Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinformatics* 12, 35. doi:10.1186/1471-2105-12-35

Chen, X., Sun, L. G., and Zhao, Y. (2021). NCMCMDA: miRNA-Disease Association Prediction through Neighborhood Constraint Matrix Completion. *Brief Bioinform* 22 (1), 485–496. doi:10.1093/bib/bbz159

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48 (D1), D127–d131. doi:10.1093/nar/gkz757

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 (Suppl. 4), S11. doi:10.1186/1752-0509-8-s4-s11

Daniele, B., Bencivenga, A., Megna, A. S., and Tinessa, V. (2004). Alpha-fetoprotein and Ultrasonography Screening for Hepatocellular Carcinoma. *Gastroenterology* 127 (5 Suppl. 1), S108–S112. doi:10.1053/j.gastro.2004.09.023

Dimri, M., and Satyanarayana, A. (2020). Molecular Signaling Pathways and Therapeutic Targets in Hepatocellular Carcinoma. *Cancers (Basel)* 12 (2), 491. doi:10.3390/cancers12020491

Ding, J., Zhao, J., Huan, L., Liu, Y., Qiao, Y., Wang, Z., et al. (2020). Inflammation-Induced Long Intergenic Noncoding RNA (LINC00665) Increases Malignancy through Activating the Double-Stranded RNA-Activated Protein Kinase/Nuclear Factor Kappa B Pathway in Hepatocellular Carcinoma. *Hepatology* 72 (5), 1666–1681. doi:10.1002/hep.31195

Ghouri, Y. A., Mian, I., and Rowe, J. H. (2017). Review of Hepatocellular Carcinoma: Epidemiology, Etiology, and Carcinogenesis. *J. Carcinog* 16, 1. doi:10.4103/jcar.JCar_9_16

Giaginis, C., Giagini, A., Tsourouflis, G., Gatzidou, E., Agapitos, E., Kouraklis, G., et al. (2011). MCM-2 and MCM-5 Expression in Gastric Adenocarcinoma: Clinical Significance and Comparison with Ki-67 Proliferative Marker. *Dig. Dis. Sci.* 56 (3), 777–785. doi:10.1007/s10620-010-1348-5

Giovannini, C., Bolondi, L., and Gramantieri, L. (2016). Targeting Notch3 in Hepatocellular Carcinoma: Molecular Mechanisms and Therapeutic Perspectives. *Int. J. Mol. Sci.* 18 (1), 56. doi:10.3390/ijms18010056

Greenbaum, L. E. (2004). Cell Cycle Regulation and Hepatocarcinogenesis. *Cancer Biol. Ther.* 3 (12), 1200–1207. doi:10.4161/cbt.3.12.1392

Gu, Y., Wan, C., Zhou, G., Zhu, J., Shi, Z., and Zhuang, Z. (2021). TYMSOS Drives the Proliferation, Migration, and Invasion of Gastric Cancer Cells by Regulating ZNF703 *via* Sponging miR-4739. *Cell Biol Int* 45 (8), 1710–1719. doi:10.1002/cbin.11610

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* 32 (18), 2847–2849. doi:10.1093/bioinformatics/btw313

Guo, D., Li, Y., Chen, Y., Zhang, D., Wang, X., Lu, G., et al. (2019). DANCR Promotes HCC Progression and Regulates EMT by Sponging miR-27a-3p *via* ROCK1/LIMK1/COFILIN1 Pathway. *Cell Prolif* 52 (4), e12628. doi:10.1111/cpr.12628

Hayashi, M., Yamada, S., Kurimoto, K., Tanabe, H., Hirabayashi, S., Sonohara, F., et al. (2021). miR-23b-3p Plays an Oncogenic Role in Hepatocellular Carcinoma. *Ann. Surg. Oncol.* 28 (6), 3416–3426. doi:10.1245/s10434-020-09283-y

He, J., Xiao, B., Li, X., He, Y., Li, L., and Sun, Z. (2019). MiR-486-5p Suppresses Proliferation and Migration of Hepatocellular Carcinoma Cells through Downregulation of the E3 Ubiquitin Ligase CBL. *Biomed. Res. Int.* 2019, 2732057. doi:10.1155/2019/2732057

Huang, G., Liang, M., Liu, H., Huang, J., Li, P., Wang, C., et al. (2020). CircRNA hsa_circRNA_104348 Promotes Hepatocellular Carcinoma Progression through Modulating miR-187-3p/RTKN2 axis and Activating Wnt/β-Catenin Pathway. *Cell Death Dis* 11 (12), 1065. doi:10.1038/s41419-020-03276-1

Issac, M. S. M., Yousef, E., Tahir, M. R., and Gaboury, L. A. (2019). MCM2, MCM4, and MCM6 in Breast Cancer: Clinical Utility in Diagnosis and Prognosis. *Neoplasia* 21 (10), 1015–1035. doi:10.1016/j.neo.2019.07.011

Jeck, W. R., and Sharpless, N. E. (2014). Detecting and Characterizing Circular RNAs. *Nat. Biotechnol.* 32 (5), 453–461. doi:10.1038/nbt.2890

Kloosterman, W. P., and Plasterk, R. H. (2006). The Diverse Functions of microRNAs in Animal Development and Disease. *Dev. Cel* 11 (4), 441–450. doi:10.1016/j.devcel.2006.09.009

Krol, J., Loedige, I., and Filipowicz, W. (2010). The Widespread Regulation of microRNA Biogenesis, Function and Decay. *Nat. Rev. Genet.* 11 (9), 597–610. doi:10.1038/nrg2843

Labib, K., and Gambus, A. (2007). A Key Role for the GINS Complex at DNA Replication forks. *Trends Cel Biol* 17 (6), 271–278. doi:10.1016/j.tcb.2007.04.002

Lan, W., Lai, D., Chen, Q., Wu, X., Chen, B., Liu, J., et al. (2020). LDICDL: LncRNA-Disease Association Identification Based on Collaborative Deep Learning. *Ieee/acm Trans. Comput. Biol. Bioinform*. doi:10.1109/tcbb.2020.3034910

Lee, Y. S., and Dutta, A. (2009). MicroRNAs in Cancer. *Annu. Rev. Pathol.* 4, 199–227. doi:10.1146/annurev.pathol.4.110807.092222

Lei, M., Zheng, G., Ning, Q., Zheng, J., and Dong, D. (2020). Translation and Functional Roles of Circular RNAs in Human Cancer. *Mol. Cancer* 19 (1), 30. doi:10.1186/s12943-020-1135-7

Li, H., Cao, Y., Ma, J., Luo, L., and Ma, B. (2021). Expression and Prognosis Analysis of GINS Subunits in Human Breast Cancer. *Medicine (Baltimore)* 100 (11), e24827. doi:10.1097/md.0000000000024827

Li, H. H., Wang, J. D., Wang, W., Wang, H. F., and Lv, J. Q. (2020). Effect of miR-26a-5p on Gastric Cancer Cell Proliferation, Migration and Invasion by Targeting COL10A1. *Eur. Rev. Med. Pharmacol. Sci.* 24 (3), 1186–1194. doi:10.26355/eurrev_202002_20170

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucleic Acids Res.* 42, D92–D97. Database issue. doi:10.1093/nar/gkt1248

Li, R., Jiang, J., Shi, H., Qian, H., Zhang, X., and Xu, W. (2020). CircRNA: a Rising star in Gastric Cancer. *Cell Mol Life Sci* 77 (9), 1661–1680. doi:10.1007/s00018-019-03345-5

Li, S., Wu, L., Zhang, H., Liu, X., Wang, Z., Dong, B., et al. (2021). GINS1 Induced Sorafenib Resistance by Promoting Cancer Stem Properties in Human Hepatocellular Cancer Cells. *Front Cel Dev Biol* 9, 711894. doi:10.3389/fcell.2021.711894

Li, T., Xie, J., Shen, C., Cheng, D., Shi, Y., Wu, Z., et al. (2015). Amplification of Long Noncoding RNA ZFAS1 Promotes Metastasis in Hepatocellular Carcinoma. *Cancer Res.* 75 (15), 3181–3191. doi:10.1158/0008-5472.Can-14-3721

Li, T., Xie, J., Shen, C., Cheng, D., Shi, Y., Wu, Z., et al. (2014). miR-150-5p Inhibits Hepatoma Cell Migration and Invasion by Targeting MMP14. *PLoS One* 9 (12), e115577. doi:10.1371/journal.pone.0115577

Li, Y., Wang, P., Wu, L. L., Yan, J., Pang, X. Y., and Liu, S. J. (2020). miR-26a-5p Inhibit Gastric Cancer Cell Proliferation and Invasion through Mediated Wnt5a. *Onco Targets Ther.* 13, 2537–2550. doi:10.2147/ott.S241199

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs. *Nucleic Acids Res.* 47 (W1), W199–w205. doi:10.1093/nar/gkz401

Lim, N., and Townsend, P. A. (2020). Cdc6 as a Novel Target in Cancer: Oncogenic Potential, Senescence and Subcellular Localisation. *Int. J. Cancer* 147 (6), 1528–1534. doi:10.1002/ijc.32900

Lin, Y., Jian, Z., Jin, H., Wei, X., Zou, X., Guan, R., et al. (2020). Long Non-coding RNA DLGAP1-AS1 Facilitates Tumorigenesis and Epithelial-Mesenchymal Transition in Hepatocellular Carcinoma *via* the Feedback Loop of miR-26a/b-5p/IL-6/JAK2/STAT3 and Wnt/β-Catenin Pathway. *Cel Death Dis* 11 (1), 34. doi:10.1038/s41419-019-2188-7

Liu, M., Wang, Q., Shen, J., Yang, B. B., and Ding, X. (2019). Circbank: a Comprehensive Database for circRNA with Standard Nomenclature. *RNA Biol.* 16 (7), 899–905. doi:10.1080/15476286.2019.1600395

Liu, X. H., Sun, M., Nie, F. Q., Ge, Y. B., Zhang, E. B., Yin, D. D., et al. (2014). Lnc RNA HOTAIR Functions as a Competing Endogenous RNA to Regulate HER2 Expression by Sponging miR-331-3p in Gastric Cancer. *Mol. Cancer* 13, 92. doi:10.1186/1476-4598-13-92

Liu, Y. Z., Wang, B. S., Jiang, Y. Y., Cao, J., Hao, J. J., Zhang, Y., et al. (2017). MCMs Expression in Lung Cancer: Implication of Prognostic Significance. *J. Cancer* 8 (18), 3641–3647. doi:10.7150/jca.20777

Liu, Z., Li, J., Chen, J., Shan, Q., Dai, H., Xie, H., et al. (2018). MCM Family in HCC: MCM6 Indicates Adverse Tumor Features and Poor Outcomes and Promotes S/G2 Cell Cycle Progression. *BMC Cancer* 18 (1), 200. doi:10.1186/s12885-018-4056-8

Liu, Z., Wang, Y., Wang, L., Yao, B., Sun, L., Liu, R., et al. (2019). Long Non-coding RNA AGAP2-AS1, Functioning as a Competitive Endogenous RNA, Upregulates ANXA11 Expression by Sponging miR-16-5p and Promotes Proliferation and Metastasis in Hepatocellular Carcinoma. *J. Exp. Clin. Cancer Res.* 38 (1), 194. doi:10.1186/s13046-019-1188-x

Maiorano, D., Lutzmann, M., and Méchali, M. (2006). MCM Proteins and DNA Replication. *Curr. Opin. Cel Biol* 18 (2), 130–136. doi:10.1016/j.ceb.2006.02.006

Neves, H., and Kwok, H. F. (2017). In Sickness and in Health: The many Roles of the Minichromosome Maintenance Proteins. *Biochim. Biophys. Acta Rev. Cancer* 1868 (1), 295–308. doi:10.1016/j.bbcan.2017.06.001

Pang, J., Li, Z., Wang, G., Li, N., Gao, Y., and Wang, S. (2018). miR-214-5p Targets KLF5 and Suppresses Proliferation of Human Hepatocellular Carcinoma Cells. *J. Cel Biochem.* doi:10.1002/jcb.27498

Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., et al. (2016). DIANA-LncBase V2: Indexing microRNA Targets on Non-coding Transcripts. *Nucleic Acids Res.* 44 (D1), D231–D238. doi:10.1093/nar/gkv1270

Peng, L., Yuan, X. Q., and Li, G. C. (2015). The Emerging Landscape of Circular RNA ciRS-7 in Cancer (Review). *Oncol. Rep.* 33 (6), 2669–2674. doi:10.3892/or.2015.3904

Peng, Y. P., Zhu, Y., Yin, L. D., Zhang, J. J., Guo, S., Fu, Y., et al. (2016). The Expression and Prognostic Roles of MCMs in Pancreatic Cancer. *PLoS One* 11 (10), e0164150. doi:10.1371/journal.pone.0164150

Petrakis, T. G., Vougas, K., and Gorgoulis, V. G. (2012). Cdc6: a Multi-Functional Molecular Switch with Critical Role in Carcinogenesis. *Transcription* 3 (3), 124–129. doi:10.4161/trns.20301

Piñero, F., Dirchwolf, M., and Pessôa, M. G. (2020). Biomarkers in Hepatocellular Carcinoma: Diagnosis, Prognosis and Treatment Response Assessment. *Cells* 9 (6), 1370. doi:10.3390/cells9061370

Qi, X., Zhang, D. H., Wu, N., Xiao, J. H., Wang, X., and Ma, W. (2015). ceRNA in Cancer: Possible Functions and Clinical Implications. *J. Med. Genet.* 52 (10), 710–718. doi:10.1136/jmedgenet-2015-103334

Qu, K., Wang, Z., Fan, H., Li, J., Liu, J., Li, P., et al. (2017). MCM7 Promotes Cancer Progression through Cyclin D1-dependent Signaling and Serves as a Prognostic Marker for Patients with Hepatocellular Carcinoma. *Cel Death Dis* 8 (2), e2603. doi:10.1038/cddis.2016.352

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA Hypothesis: the Rosetta Stone of a Hidden RNA Language? *Cell* 146 (3), 353–358. doi:10.1016/j.cell.2011.07.014

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Sideridou, M., Zakopoulou, R., Evangelou, K., Liontos, M., Kotsinas, A., Rampakakis, E., et al. (2011). Cdc6 Expression Represses E-Cadherin Transcription and Activates Adjacent Replication Origins. *J. Cel Biol* 195 (7), 1123–1140. doi:10.1083/jcb.201108121

Singal, A. G., Lampertico, P., and Nahon, P. (2020). Epidemiology and Surveillance for Hepatocellular Carcinoma: New Trends. *J. Hepatol.* 72 (2), 250–261. doi:10.1016/j.jhep.2019.08.025

Song, L. N., Qiao, G. L., Yu, J., Yang, C. M., Chen, Y., Deng, Z. F., et al. (2020). Hsa_circ_0003998 Promotes Epithelial to Mesenchymal Transition of Hepatocellular Carcinoma by Sponging miR-143-3p and PCBP1. *J. Exp. Clin. Cancer Res.* 39 (1), 114. doi:10.1186/s13046-020-01576-0

Sticht, C., De La Torre, C., Parveen, A., and Gretz, N. (2018). miRWalk: An Online Resource for Prediction of microRNA Binding Sites. *PLoS One* 13 (10), e0206239. doi:10.1371/journal.pone.0206239

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U S A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49 (D1), D605–d612. doi:10.1093/nar/gkaa1074

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–w102. doi:10.1093/nar/gkx247

Taulli, R., Loretelli, C., and Pandolfi, P. P. (2013). From Pseudo-ceRNAs to Circ-ceRNAs: a Tale of Cross-Talk and Competition. *Nat. Struct. Mol. Biol.* 20 (5), 541–543. doi:10.1038/nsmb.2580

Tian, L., Zhao, Z. F., Xie, L., and Zhu, J. P. (2019). Taurine Up-Regulated 1 Accelerates Tumorigenesis of colon Cancer by Regulating miR-26a-5p/MMP14/p38 MAPK/Hsp27 axis *In Vitro* and *In Vivo*. *Life Sci.* 239, 117035. doi:10.1016/j.lfs.2019.117035

Wang, S. H., Ma, F., Tang, Z. H., Wu, X. C., Cai, Q., Zhang, M. D., et al. (2016). Long Non-coding RNA H19 Regulates FOXM1 Expression by Competitively Binding Endogenous miR-342-3p in Gallbladder Cancer. *J. Exp. Clin. Cancer Res.* 35 (1), 160. doi:10.1186/s13046-016-0436-6

Wang, Y. G., Wang, T., Ding, M., Xiang, S. H., Shi, M., and Zhai, B. (2019). hsa_circ_0091570 Acts as a ceRNA to Suppress Hepatocellular Cancer Progression by Sponging Hsa-miR-1307. *Cancer Lett.* 460, 128–138. doi:10.1016/j.canlet.2019.06.007

Wong, N. W., Chen, Y., Chen, S., and Wang, X. (2018). OncomiR: an Online Resource for Exploring Pan-Cancer microRNA Dysregulation. *Bioinformatics* 34 (4), 713–715. doi:10.1093/bioinformatics/btx627

Wu, L., Pan, C., Wei, X., Shi, Y., Zheng, J., Lin, X., et al. (2018). lncRNA KRAL Reverses 5-fluorouracil Resistance in Hepatocellular Carcinoma Cells by Acting as a ceRNA against miR-141. *Cell Commun Signal* 16 (1), 47. doi:10.1186/s12964-018-0260-z

Wu, Y., Zhou, Y., Huan, L., Xu, L., Shen, M., Huang, S., et al. (2019). LncRNA MIR22HG Inhibits Growth, Migration and Invasion through Regulating the miR-10a-5p/NCOR2 axis in Hepatocellular Carcinoma Cells. *Cancer Sci.* 110 (3), 973–984. doi:10.1111/cas.13950

Xu, G., Xu, W. Y., Xiao, Y., Jin, B., Du, S. D., Mao, Y. L., et al. (2020). The Emerging Roles of Non-coding Competing Endogenous RNA in Hepatocellular Carcinoma. *Cancer Cel Int* 20, 496. doi:10.1186/s12935-020-01581-5

Xu, H., Huang, J., Hua, S., Liang, L., He, X., Zhan, M., et al. (2020). Interactome Analysis of Gene Expression Profiles Identifies CDC6 as a Potential Therapeutic Target Modified by miR-215-5p in Hepatocellular Carcinoma. *Int. J. Med. Sci.* 17 (18), 2926–2940. doi:10.7150/ijms.51145

Xu, Y., Yang, X., Si, T., Yu, H., Li, Y., Xing, W., et al. (2021). MCM4 in Human Hepatocellular Carcinoma: a Potent Prognostic Factor Associated with Cell Proliferation. *Biosci. Trends* 15 (2), 100–106. doi:10.5582/bst.2021.01016

Yao, R. W., Wang, Y., and Chen, L. L. (2019). Cellular Functions of Long Noncoding RNAs. *Nat. Cel Biol* 21 (5), 542–551. doi:10.1038/s41556-019-0311-8

Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F., and Ishino, Y. (2008). The GINS Complex from Pyrococcus Furiosus Stimulates the MCM Helicase Activity. *J. Biol. Chem.* 283 (3), 1601–1609. doi:10.1074/jbc.M707654200

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, J., Yang, M., Zhou, B., Luo, J., Zhang, Z., Zhang, W., et al. (2019). CircRNA-104718 Acts as Competing Endogenous RNA and Promotes Hepatocellular Carcinoma Progression through microRNA-218-5p/TXNDC5 Signaling Pathway. *Clin. Sci. (Lond)* 133 (13), 1487–1503. doi:10.1042/cs20190394

Yu, S., Wang, G., Shi, Y., Xu, H., Zheng, Y., and Chen, Y. (2020). MCMs in Cancer: Prognostic Potential and Mechanisms. *Anal. Cel Pathol (Amst)* 2020, 3750294. doi:10.1155/2020/3750294

Yu, T., Wang, Y., Fan, Y., Fang, N., Wang, T., Xu, T., et al. (2019). CircRNAs in Cancer Metabolism: a Review. *J. Hematol. Oncol.* 12 (1), 90. doi:10.1186/s13045-019-0776-8

Zhang, F., Yang, C., Xing, Z., Liu, P., Zhang, B., Ma, X., et al. (2019). LncRNA GAS5-Mediated miR-1323 Promotes Tumor Progression by Targeting TP53INP1 in Hepatocellular Carcinoma. *Onco Targets Ther.* 12, 4013–4023. doi:10.2147/ott.S209439

Zhang, X. M., Wang, J., Liu, Z. L., Liu, H., Cheng, Y. F., and Wang, T. (2020). LINC00657/miR-26a-5p/CKS2 ceRNA Network Promotes the Growth of Esophageal Cancer Cells *via* the MDM2/p53/Bcl2/Bax Pathway. *Biosci. Rep.* 40 (6), BSR20200525. doi:10.1042/bsr20200525

Zhou, H. L., Zhou, Y. F., and Feng, Z. T. (2019). Long Noncoding RNA ZFAS1 Promotes Hepatocellular Carcinoma Proliferation by Epigenetically Repressing miR-193a-3p. *Eur. Rev. Med. Pharmacol. Sci.* 23 (22), 9840–9847. doi:10.26355/eurrev_201911_19547

Zhu, W. J., Yan, Y., Zhang, J. W., Tang, Y. D., and Han, B. (2020). Effect and Mechanism of miR-26a-5p on Proliferation and Apoptosis of Hepatocellular Carcinoma Cells. *Cancer Manag. Res.* 12, 3013–3022. doi:10.2147/cmar.S237752

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Zhang, Ding and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 01 April 2022
doi: 10.3389/fgene.2022.834731

A Novel Signature Integrated of Immunoglobulin, Glycosylation and Anti-Viral Genes to Predict Prognosis for Breast Cancer

Shengshan Xu[1,2†], Yuchen Liu[2†], Hansu Ma[2], Shuo Fang[3], Shoupeng Wei[2], Xiaoping Li[4], Zhuming Lu[1], Youbin Zheng[5], Tong Liu[6], Xiaojian Zhu[2], Dongming Xu[7*] and Yihang Pan[2*]*

[1]Department of Thoracic Surgery, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, China, [2]Scientific Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, [3]Department of Oncology, The Seventh Affiliated Hospital, Sun Yat-Sen University, Shenzhen, China, [4]Department of Breast, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, China, [5]Department of Radiology, Jiangmen Wuyi Hospital of Traditional Chinese Medicine, Jiangmen, China, [6]Department of Thoracic Surgery, First Hospital of Jilin University, Changchun, China, [7]Department of Neurosurgery, The Country Hospital of Qianguo, Songyuan, China

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Udo Jeschke,
Ludwig Maximilian University of Munich, Germany
Zhu Wang,
Shenzhen Longhua People's Hospital, China

***Correspondence:**
Yuchen Liu
liuych83@mail.sysu.edu.cn
Dongming Xu
xudm0807@163.com
Yihang Pan
panyih@mail.sysu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 13 December 2021
Accepted: 14 March 2022
Published: 01 April 2022

Citation:
Xu S, Liu Y, Ma H, Fang S, Wei S, Li X, Lu Z, Zheng Y, Liu T, Zhu X, Xu D and Pan Y (2022) A Novel Signature Integrated of Immunoglobulin, Glycosylation and Anti-Viral Genes to Predict Prognosis for Breast Cancer. Front. Genet. 13:834731. doi: 10.3389/fgene.2022.834731

Background: Aberrant glycosylation is significantly related to the occurrence, progression, metastasis, and drug resistance of tumors. It is essential to identify glycosylation and related genes with prognostic value for breast cancer.

Objective: We aimed to construct and validate a prognostic model based on glycosylation and related genes, and further investigate its prognosis values in validation set and external independent cohorts.

Materials and Methods: The transcriptome and clinical data of breast cancer patients were downloaded from The Cancer Genome Atlas (TCGA, $n = 1072$), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, $n = 1451$), and GSE2741 ($n = 120$). Glycosylation-related genes were downloaded from the Genecards website. Differentially expressed glycosylation-related geneswere identified by comparing the tumor tissues with the adjacent tissues. The TCGA data were randomly divided into training set and validation set in a 1:1 ratio for further analysis. The glycosylation risk-scoring prognosis model was constructed by univariate and multivariate Cox regression analysis, followed by confirmation in TCGA validation, METABRIC, and GEO datasets. Gene set enrichment analysis (GSEA) and Gene ontology analysis for identifying the affected pathways in the high- and low-risk groups were performed.

Results: We attained 1072 breast cancer samples from the TCGA database and 786 glycosylation genes from the Genecards website. A signature contains immunoglobulin, glycosylation and anti-viral related genes was constructed to separate BRCA patients into two risk groups. Low-risk patients had better overall survival than high-risk patients ($p < 0.001$). A nomogram was constructed with risk scores and clinical characteristics. The area under time-dependent ROC curve reached 0.764 at 1 year, 0.744 at 3 years, and 0.765 at 5 years in the training set. Subgroup analysis showed differences in OS between the high- and low-risk patients in different subgroups. Moreover, the risk score was

confirmed as an independent prognostic indicator of BRCA patients and was potentially correlated with immunotherapy response and drug sensitivity.

Conclusion: We identified a novel signature integrated of immunoglobulin (IGHA2), glycosylation-related (SLC35A2) and anti-viral gene (BST2) that was an independent prognostic indicator for BRCA patients. The risk-scoring model could be used for predicting prognosis and immunotherapy in BRCA, thus providing a powerful instrument for combating BRCA.

Keywords: breast cancer, gene signature, prognosis mode, immunotharapy, glycosylation

INTRODUCTION

Globally, breast cancer is the most common cancer among women, and is the leading cause of cancer deaths among women aged 20 to 59 (R. L. Siegel et al., 2021). The cumulative risk of breast cancer is about 5% in women, and the risk of death is 1.4% (E. Hadadi et al., 2020). In recent years, the incidence of breast cancer has continued to increase by about 0.5% annually (R. L. Siegel et al., 2021), which seriously affects women's health and quality of life. Breast cancer is a highly heterogeneous disease. The main treatments for breast cancer include systemic therapy (chemotherapy, endocrine therapy, targeted therapy, and immunotherapy) and local treatment (surgery and radiotherapy). Breast cancer molecular subtypes were essential indicators for treatment and prognosis. Currently, the majority of patients diagnosed with a specific breast cancer subtype receive the same treatment, even though it has been repeatedly proven that they should adopt differential strategies (S. A. Eccles et al., 2013; S. Fallahpour et al., 2017; W. J. Gradishar et al., 2020). Triple-negative breast cancer (TNBC) is a type of breast cancer that lacks expression of human epidermal growth factor receptor 2 (HER2), progesterone receptor, and estrogen receptor (G. Bianchini et al., 2016). Typically, the prognosis for women with TNBC following metastatic recurrence is much poorer than other subtypes (M. Smid et al., 2008). It is a heterogeneous disease representing about 15% of total breast cancer incidents, which is difficult to treat as lack of available targeted therapies. Chemotherapy remains to be the preferred systemic treatment for TNBC (D. P. Silver et al., 2010), and particularly in those carrying BRCA1 mutations (J. Collignon et al., 2016). This demonstrates the reliable predictive biomarkers are necessary for precise diagnosis and individualized treatment for breast cancer patients, and the precision medicine progress have been fueled by the continuous development of new sequencing and computational technologies.

Genetic and epigenetic alterations are considered the primary causes of cancer development, and the downstream phenotypic changes at the protein level are amongst the driving forces (A. Peixoto et al., 2019). Glycosylation is the most common and complicated post-translational modification for membrane-bound proteins. More than 50–70% of proteins in the circulation are glycosylated, which play important roles in various cellular activities, such as cell growth, differentiation, transformation, and adhesion (Ohtsubo and Marth, 2006). Aberrant glycosylation has been identified as a hallmark of cancer and intimately correlated with cancer occurrence, progression, metastasis, tumor recurrence, and drug resistance (Pinho and Reis, 2015; Rao et al., 2017; A. Chakraborty et al., 2018; Cui et al., 2018). There was also a correlation between glycosylation and antitumor immunity. For example, Freire et al. demonstrated that Tn glycosylation of the MUC6 protein strongly affects the immunogenicity of its B and T cell, and might enable immune escape of tumor cells (T. Freire et al., 2011). Bone marrow stromal antigen 2 (*BST2*) was a type II transmembrane protein, also known as tethered protein, HM1.24 orCD317. BST2 homodimer promotes cancer cell adhesion and enhances cancer cell survival and growth by enhancing proteasomal degradation of pro-apoptotic proteins (Mahauad-Fernandez and Okeoma, 2017). Sayeed et al. indicated that aberrant BST2 overexpression promoted the disappearance of TGFβ-mediated tumor-suppressive effects in breast cancer as a consequence and the ensuing loss of the differentiation program (A. Sayeed et al., 2013). Mahauad Fernandez and Okeoma suggested that BST-2 targets breast cancer cells that are resistant to anoikis via the GRB2/ERK/BIM/Cas3 pathway. Almost all breast tumors express *BST2* to a certain level, and the high expression level of *BST2* was related to progressive malignancies (W. D. Mahauad-Fernandez et al., 2014). Several studies have shown that BST2 regulates the occurrence of gastric cancer, oral squamous cell carcinoma, lung cancer and is involved in tumor metastasis and invasion (W. Wang et al., 2009a; K. H. Fang et al., 2014; W. Liu et al., 2018; W. Liu et al., 2020). Immunoglobulin heavy constant alpha 2 (*IGHA2*), located on chromosome 14, expressed in breast cancer cells and upregulated in advanced breast tumor tissues by comparison with early tumors, was involved in the early stage of the tumor microenvironment remodeling and has been identified as a marker of regional metastasis in lymph nodes. Suki Kang et al. reported that IGHA2 might protect the cells against physiological stresses during the neoplastic process and promote tumor growth in the advanced stages of cancer (S. Kang et al., 2012). Exploring the role of glycosylation, immunoglobulin, and anti-viral in BRCA and related molecules will help us investigate strategies to combat for BRCA. Currently, there is still no study exploring the prognosis value of glycosylation, immunoglobulin, and anti-viral -related genes. Therefore, this study aims to establish a prognostic model for breast cancer based on glycosylation-related genes and evaluate it from multiple dimensions.

MATERIALS AND METHODS

Data Source and Processing

Breast cancer datasets were downloaded from the Molecular Taxonomy of Breast Cancer International Consortium

(METABRIC, http://www.cbioportal.org), The Cancer Genome Atlas (TCGA, https://www.cancer.gov), and the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Patients who met the following selection criteria were included: 1) histologically diagnosed with malignant breast cancer; 2) available RNA expression data; and 3) available OS data. After screening, this study included 1,451 patients from METABRIC, 1,072 patients from TCGA, and 120 patients from GSE2741. Glycosylation-related genes were downloaded from the website of Genecards (https://www.genecards.org). A total of 786 glycosylation genes were analyzed by comparing the tumor tissues with the adjacent tissues to obtain differentially expressed glycosylation-related genes (DEGRGs). A p value less than 0.05 was considered statistically significant.

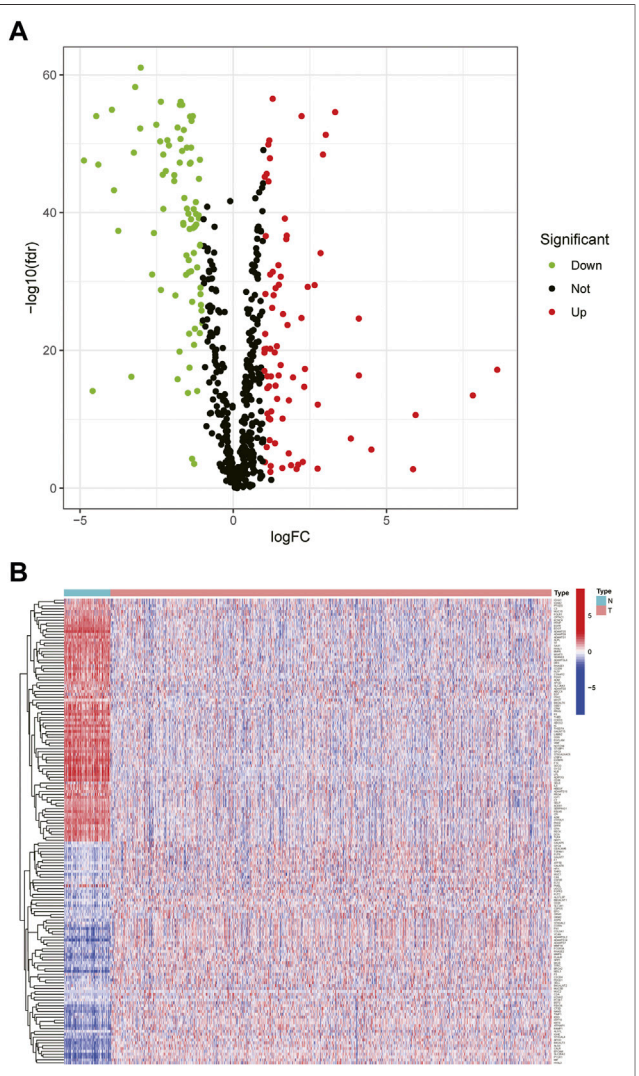


FIGURE 1 | Differential expression of glycosylation-related genes **(A)** The volcano plot showed the up-regulated and down-regulated differential genes in breast tumor tissues compared with adjacent tissues (t-test, Adjust. $p < 0.05$) **(B)** The heat map showed that 163 differential glycosylation-related genes were expressed in tumor tissues and adjacent tissues.

Construction and Validation of the Risk-Scoring Model

The TCGA BRCA data were randomly assigned into training set and validation set according to the ratio of 1:1. In order to determine the survival-related glycosylation genes, we performed univariate Cox regression, in which $p < 0.05$ was set as a cut-off criterion. Subsequently, multivariate Cox regression was performed to construct a prognostic risk-scoring model, in which the risk score for each patient was calculated according to the following formula:

$$RiskScore = [(0.9138 \times Expression\ value\ of\ SLC35A2) + [(-0.2483) \times Expression\ value\ of\ BST2] + [(-0.1002) \times Expression\ value\ of\ IGHA2]]$$

All BRCA patients were assigned to the high- and low-risk groups according to the median risk score in the training set. The difference in OS between these two groups was investigated by the log-rank test and Kaplan-Meier survival analysis. In addition, the distributions of survival status, OS, and risk score in the training set were also plotted.

Independent Prognostic Analysis

To evaluate the relationship between clinicopathological factors and risk scores on survival time, we used the "Survival" R package to perform univariate and multivariate Cox regression. The time-dependent receiver operator characteristic (ROC) curve was drawn, and the R package "timeROC" was applied to determine the prognostic performance of either clinicopathological factors or risk

TABLE 1 | The association between risk score and patients' clinical features in the training set.

Variables	Training Set (n = 432)		Validation Set (n = 429)		*p*-Value
	No.	%	No.	%	
Age					0.986
≤65	319	73.8	318	74.1	
>65	113	26.2	111	25.9	
Stage					0.870
I	72	16.7	79	18.4	
II	254	58.8	251	58.5	
III	97	25.5	92	21.5	
IV	9	2	7	1.6	
T stage					0.622
T1	111	25.7	115	26.8	
T2	253	58.6	258	60.1	
T3	51	11.8	45	10.5	
T4	17	3.9	11	2.6	
N stage					0.504
N0	218	50.5	201	46.9	
N1	141	32.6	147	34.2	
N2	51	11.8	50	11.7	
N3	22	5.1	31	7.2	
M stage					0.812
M0	423	97.9	422	98.4	
M1	9	2.1	7	1.6	

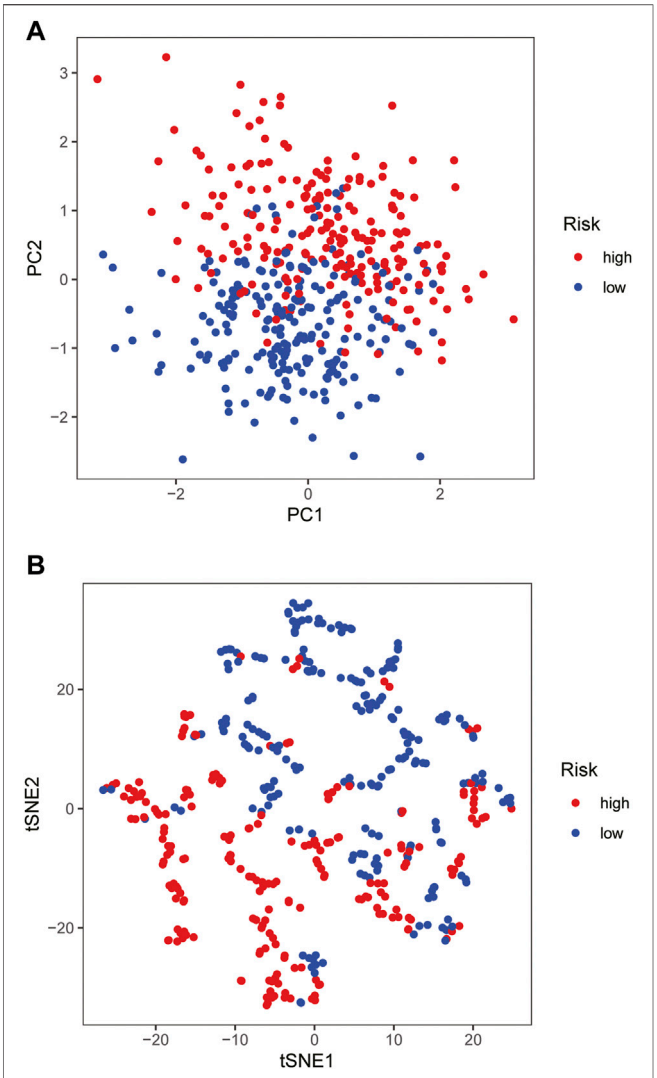


FIGURE 2 | Construction of risk-scoring model **(A,B)** PCA and t-SNE based on the expression profile of the 3 selected signature genes separated different risk groups.

scores on survival time. $p < 0.05$ was considered statistically significant.

Subgroup Analysis

We further evaluated the model's predictive ability by stratifying patients into various subgroups. These variables include age (≤65 and >65 years), tumor stage (I-II and III-IV), T stage (T1-2 and T3-4), N stage (N0 and N1-3), M stage (M0 and M1), estrogen receptor (ER) state (positive and negative), human epidermal growth factor receptor (HER-2) state (positive and negative), triple-negative breast cancer (yes and no). To further confirm the value of the risk score in TNBC and HER-2 positive subgroups, the TNBC and HER-2 positive samples from METABRIC database were used for re-verification. Through clinical survival analysis, the predictive ability of the risk-scoring model in various clinical subgroups was clarified. $p < 0.05$ was considered statistically significant.

Exploration of the Value of the Risk-Scoring Model in Clinical Utility

To further improve the practical value of the risk-scoring model, a nomogram was constructed by integrating age, tumor stage, T stage, N stage, M stage, and risk score to predict the OS of patients at 1, 3, and 5 years. In addition, the C-index was used to measure the accuracy of the nomogram, and the calibration curve was drawn to evaluate the calibration of the model. The ROC curves of various clinical characteristics were drawn, and the AUC was calculated to judge the performance of the prognosis model. In addition, decision curve analysis (DCA) was used to estimate the maximum clinical benefit by logistic regression analysis.

Functional Enrichment Analysis

To explore the difference molecular pathways underlying survival prognosis between the high- and low-risk groups, we used the **"clusterProfiler" R package** to perform Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Bubble Diagram visualized the most important pathways in KEGG and each GO category. Gene Enrichment Score Analysis (GSEA) was also applied to determine signaling pathways regulated in patients of the high- and low-risk groups. $p < 0.05$ and FC > 1.2 (or FC < 0.83) were set as the cut-off values.

Analysis of Tumor Immune Microenvironment

The tumor microenvironment score of each single BRCA patient was estimated using the ESTIMATE algorithm. CIBERSORT algorithm was utilized to evaluate the proportion of 22 human immune cell subsets in the high- and low-risk groups. The "GSVA" R package was applied to perform single-sample gene set enrichment analysis (ssGSEA) to quantify the GSVA scores of the 13 immune signatures. The difference between the expression levels of immune checkpoints in the high- and low-risk groups was further evaluated. p-value < 0.05 was considered statistically significant.

Tumor Mutation Burden Analysis

After obtaining the somatic mutation data in the TCGA BRCA dataset, the "maftools" R package was applied to analyze the tumor mutation burden (TMB) of the training set, TMB value was calculated, and a waterfall chart was drawn. Then it would be assessed whether the TMB scores were related to the risk scores and patient survival probability. We found the median value of TMB and divided TCGA BRCA dataset into high TMB group and low TMB group. Patients were stratified into the new groups by integrating TMB and risk scores, and Kaplan-Meier survival analysis was used. The "Survival" and the "survminer" R packages were used for joint survival analysis. The CBIOPORTAL database (https://www.cbioportal.org/) was used to analyze the mutations profiles in the high-and low-risk groups. Finally, the protein domains where the mutations were located were clarified. $p < 0.05$ was considered statistically significant.

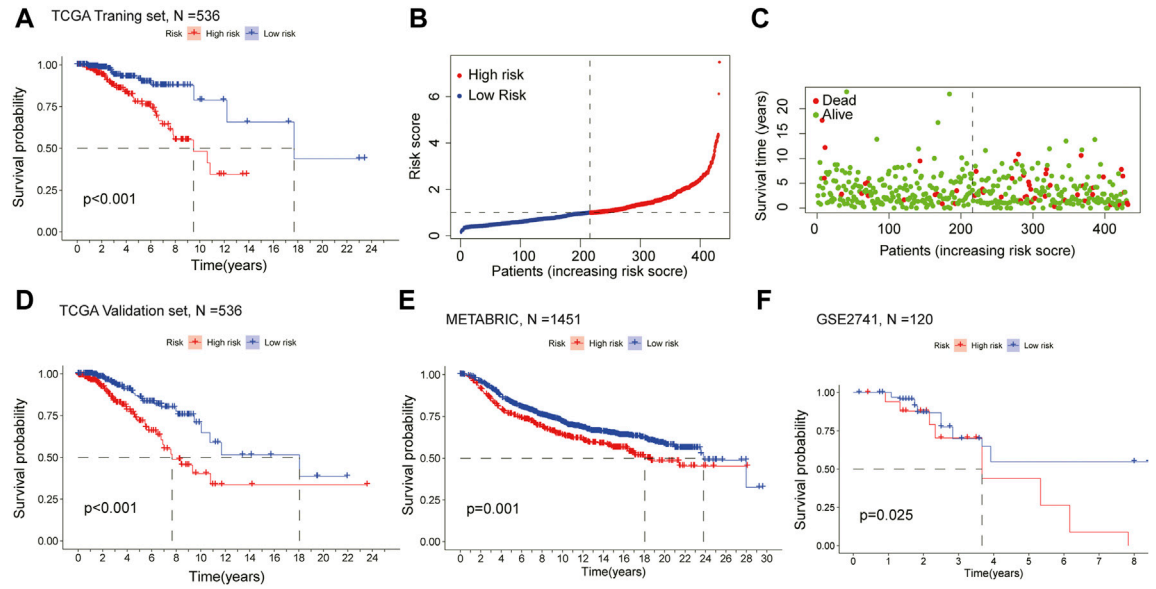


FIGURE 3 | Assessment of the prognostic model in mRNA expression database **(A)** Survival curve of the high- and low-risk groups in the training set (n = 536) (log rank test, $p < 0.05$) **(B)** The distribution of the risk scores in the training set **(C)** The distributions of survival status, OS, and risk score in the training set **(D)** Kaplan-Meier survival curve of validation set (n = 536) comparing the high- and low-risk groups **(E)** Kaplan-Meier survival curve of the high- and low-risk groups in the independent test set METABRIC (n = 1451) **(F)** Kaplan-Meier survival curve analysis in the independent test set GSE2741 (n = 120) comparing two risk groups. All the K-M survival analyses use log-rank tests to determine significant differences between two groups, $p < 0.05$.

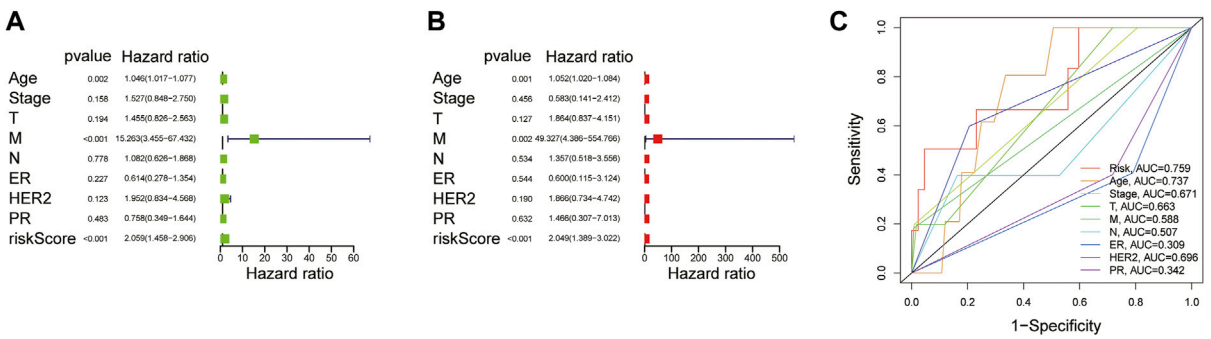


FIGURE 4 | Explore the prognostic value of risk score and clinical features **(A)** Univariate Cox regression analysis of breast cancer patients **(B)** Multivariate Cox regression analysis of breast cancer patients (Wald test, $p < 0.05$) **(C)** The AUC for risk model scores and clinical features according to the ROC curves. Clinical features: Age, ER, PR, HER-2, TNM stage, T, N, and M.

Relationship Between Risk Scores and Immunotherapy

IMvigor210, a phase II trial of atezolizumab (MPDL3280A) in platinum-treated locally advanced or metastatic urothelial carcinoma. It was downloaded to assess the correlation between risk scores and immunotherapy response. The tumor immune dysfunction and exclusion (TIDE) score was calculated online (http://tide.dfci. harvard. edu/) to assess the immune checkpoint inhibitor response between the high- and low-risk groups (Z. Lin et al., 2021). We downloaded RNA-seq and compound activity: DTP NCI-60 through the CellMiner database (https://discover.nci.nih.gov/cellminer) and excluded FDA status as empty or clinical trial data to explore the relationship between the expression of genes and drug sensitivity. The three-dimensional structures of drugs were obtained through the PubChem database (https://pubchem.ncbi.nlm.nih.gov). $p < 0.05$ was considered statistically significant.

RESULTS

Identification of Differential Expression of Glycosylation-Related Genes in Breast Cancer

We compiled the gene expression data of breast cancer from the TCGA database and finally got 1,072 tumor tissues and 99

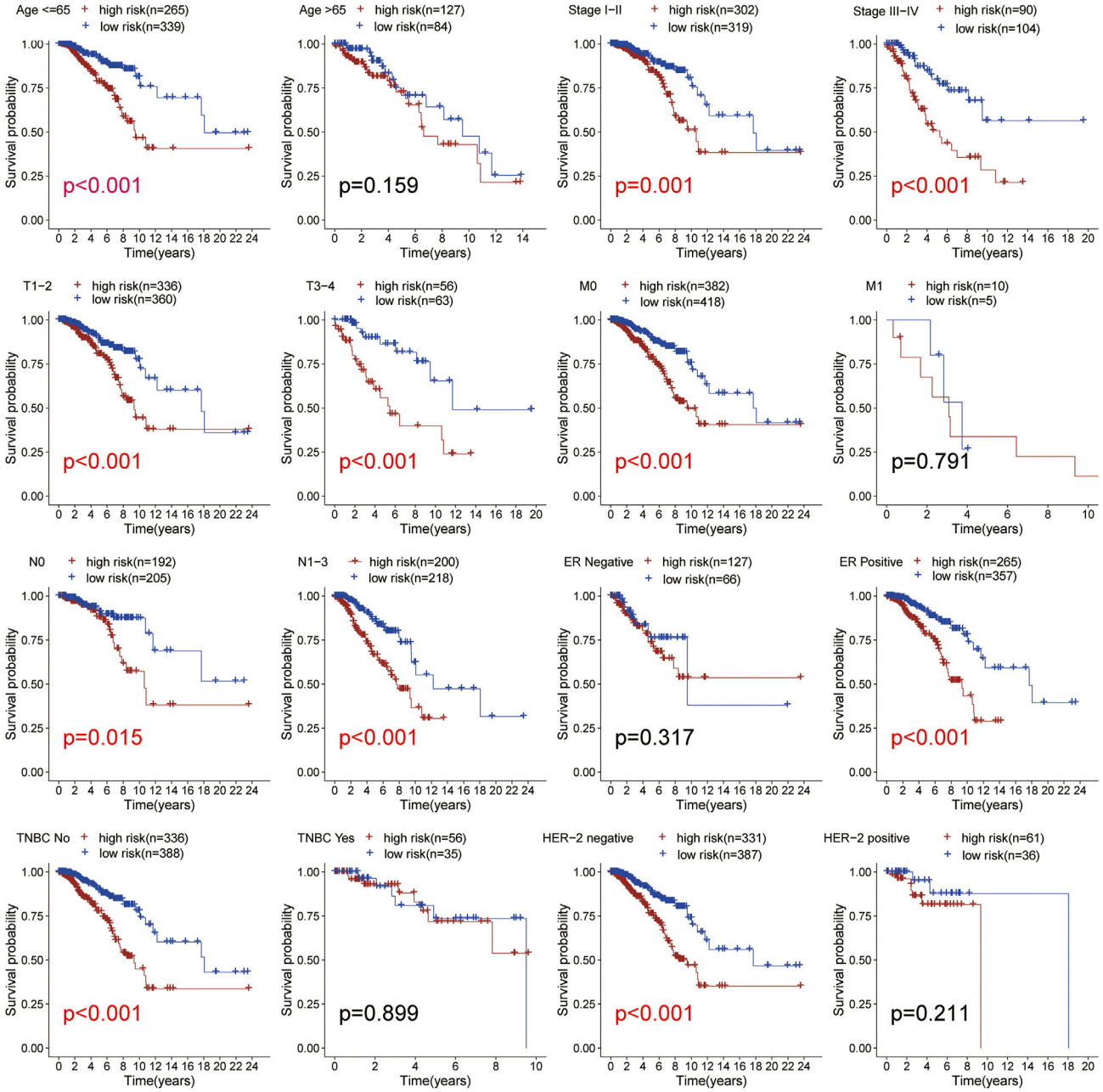


FIGURE 5 | Subgroup analysis. Kaplan-Meier survival curve analysis of patients with high-vs low-risk scores in different subgroups including age, TNM stage, ER status, HER-2 status, TNBC status. All the K-M survival analyses use log-rank tests to determine significant differences in subgroups.

adjacent tissue samples. Around 786 glycosylation-related genes were obtained by the Genecards website with a correlation >2.0 as the cut-off value. Based on standard cut-off values for fold-change in gene expression (|log(FC)|> 1) and false discovery rate (FDR <0.05), the breast cancer tissues had 163 DEGRGs, with 86 down-regulated genes and 77 up-regulated genes (**Figures 1A,B**).

Construction of a Risk-Scoring Model

The samples screened from the TCGA database were randomly assigned to the training set for the construction of the model and the validation set for accuracy estimation according to the ratio of 1:1. The clinical features of all patients are shown in detail in **Table 1**. There were no statistically significant differences in clinical features between patients in the training set and validation set. Based on the training data set, the prognosis-related glycosylation genes were screened by univariate Cox regression ($p < 0.05$), including two low-risk genes [hazard ratio (HR) < 1] and one high-risk gene [risk ratio (HR) > 1]. Then, multivariate Cox analysis was performed to screen three genes related to the glycosylation with prognostic significance

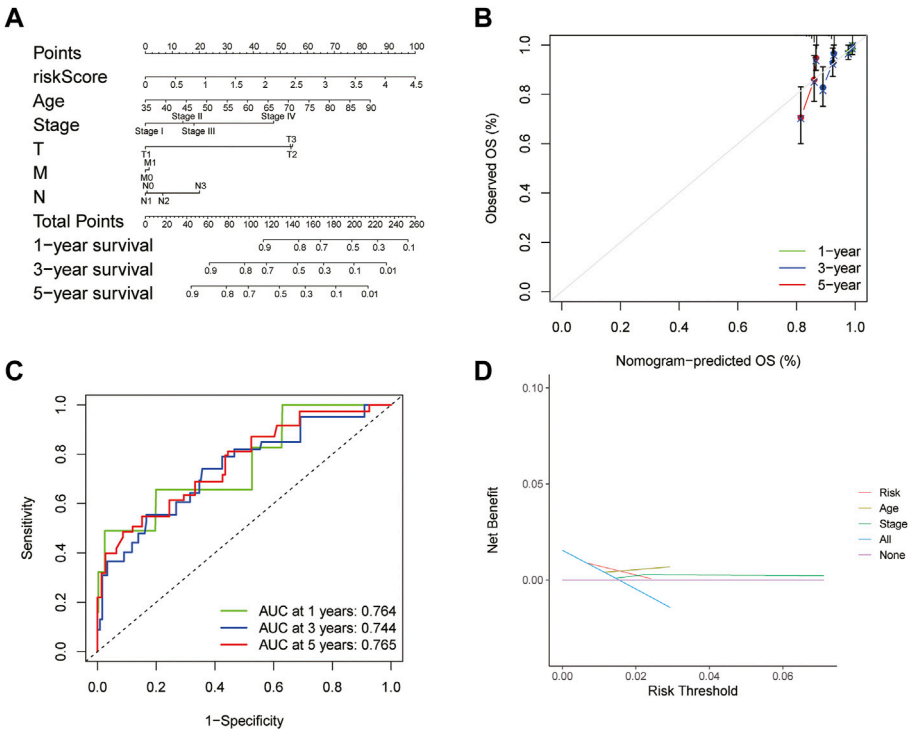


FIGURE 6 | Construction and evaluation of nomogram **(A)** Nomogram for predicting the 1-, 3-, and 5-year survival rates based on the risk score **(B)** Calibration curve of the nomogram **(C)** ROC curve analysis evaluated the prediction performance of nomogram **(D)** Decision curves of "risk", "age", "stage", "all" and "None" models.

($p < 0.05$), namely *BST2, IGHA2,* and *SLC35A2*. These three genes were used to construct a risk-scoring model. According to the risk score formula and the median risk score, the patients with breast cancer were divided into the high- and low-risk groups. The PCA and t-SNE showed that the high- and low-risk groups had different distribution directions, suggesting that the risk-scoring model could clearly divide BRCA patients into two groups (**Figures 2A,B**).

Evaluation and Validation of the Aberrant Glycosylation-Related Risk-Scoring Model

Kaplan-Meier survival analysis showed that the OS of the high-risk group was lower than that of the low-risk group in the training cohort ($p < 0.001$) (**Figure 3A**), demonstrating the excellent predictive value of the risk-scoring model in the training set. The risk curves showed the survival status and risk scores of each breast cancer sample, which was calculated and ranked based on the signature model (**Figure 3B**). The scatter plot represented the OS status of BRCA patients according to the risk score, suggesting that the higher the risk scores were, the higher the number of death was (**Figure 3C**). Similarly, with Kaplan-Meier survival analysis using the validation cohort and external independent test sets, the OS rate of the high-risk group was lower than that of the low-risk group ($p < 0.05$), confirming that the risk-scoring model had a robust prognostic value (**Figures 3D–F**).

Explore the Independent Prognostic Factors of Breast Cancer

To evaluate whether the established risk-scoring model is an independent prognostic factor for breast cancer, univariate and multivariate Cox regression was performed. The HR of the risk scores and 95% CI were 2.059 and 1.458–2.906 ($p < 0.001$), respectively, in univariate Cox regression analysis (**Figure 4A**), which were 2.049 and 1.389–3.022 ($p < 0.001$) in multivariate Cox regression analysis, respectively (**Figure 4B**). The result demonstrated that the risk score was a significant prognostic factor independent of multiple clinicopathological parameters such as the expression level of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor (HER-2) as well as the M stage, N stage, T stage, tumor stage. Additionally, compared with other clinicopathological factors, the AUC of the risk score for 1-year OS shown by ROC analysis reached 0.759, which was superior to other clinicopathological variables (**Figure 4C**). In summary, it can be concluded that the aberrant glycosylation-related risk scoring model is a significant independent prognostic factor for BRCA patients.

Subgroup Analysis

We used the TCGA BRCA samples to verify the relationship between the risk score and the prognosis of clinical features and emphasize the molecular heterogeneity of BRCA. After comparing a statistical difference in OS between the two risk

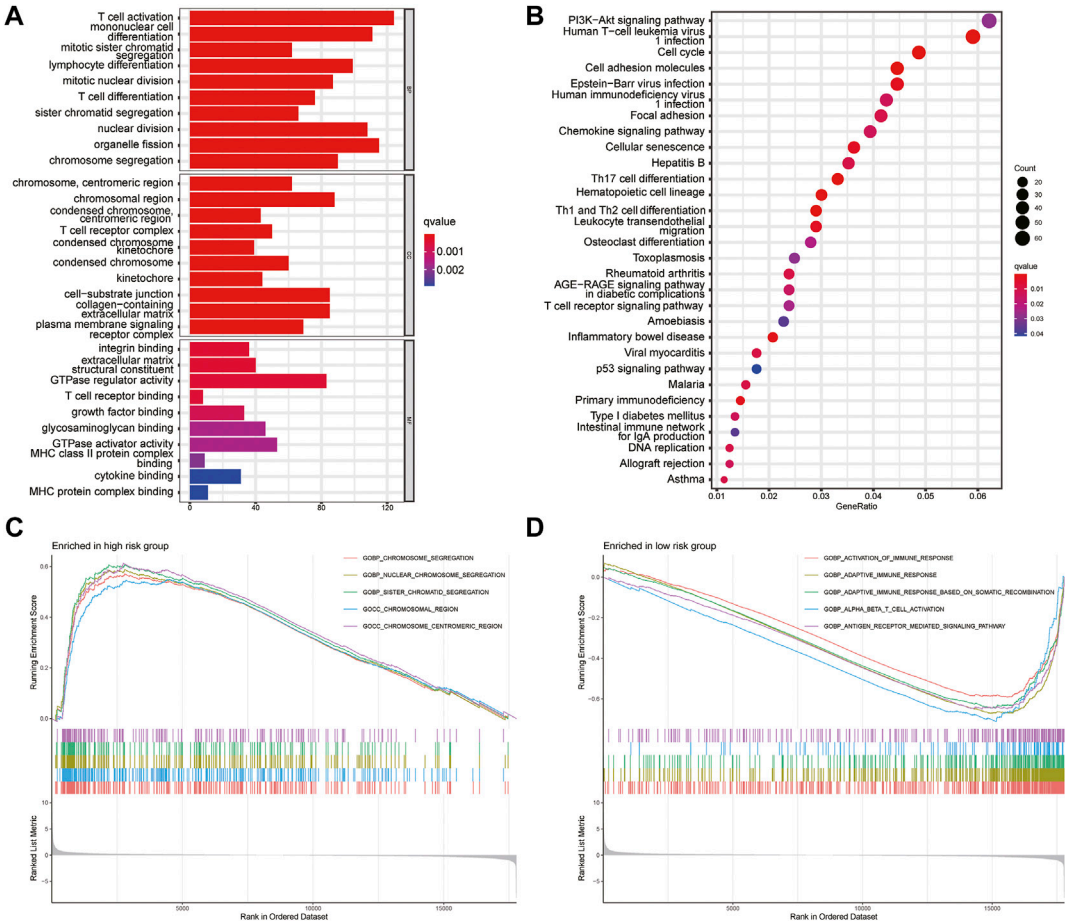


FIGURE 7 | Gene set enrichment analysis **(A)** GO analysis the first 10 items about the enrichment of BP, CC and MF were shown in the bar chart **(B)** The top 30 terms of KEGG pathways enrichment were displayed in the bubble chart **(C,D)** Gene set and function enrichment analysis of differentially expressed genes between the high-risk group low-risk group. $p < 0.05$ and $FC > 1.2$ (or $FC < 0.83$) were set as the cut-off values.

groups with Kaplan Meier survival curve, the results showed that the OS of the high-risk group was lower than that of the low-risk group among age ≤65, ER positive, HER-2 negative, M0, N0, N1-N3, stage I-II, stage III-IV, T1-2, T3-4, and non-TNBC subgroups ($p < 0.05$; **Figure 5**). In addition, the low-risk group had higher OS in the TNBC and HER-2 positive subgroups from METABRIC (**Supplementary Figure S1A,S2**). The results suggested that the risk score is closely related to the clinical features of BRCA and can be used as an effective auxiliary tool to predict the BRCA prognosis.

Clinical Evaluation Ability of the Risk-Scoring Model

A predictive nomogram based on the integration of risk score, pathological stage, and age integration was established in the TCGA cohort (**Figure 6A**). The C-index of the nomogram was 0.676 in the TCGA cohort, indicating that the nomogram had a good predictive performance. The calibration curve analysis verified that the nomogram was reliable and accurate, which demonstrated that the predictive probability of 1-, 3- and 5-year OS was ideally consistent with actual observation (**Figure 6B**). With the analysis of the ROC curve, the AUC value was 0.764, 0.744, and 0.765 for 1-, 3-, and 5-year OS, respectively, indicating that the risk-scoring model had excellent predictive accuracy (**Figure 6C**). In addition, DCA analysis was performed to evaluate the predictive value of the nomogram in clinical decision-making (**Figure 6D**). The above all indicated that the risk-scoring model and nomogram had high reliability.

Gene Set and Function Enrichment Analysis

To analyze the pathways related to the risk-scoring model, GO and KEGG enrichment analysis were performed. GO analysis showed that the differential genes between the high- and low-risk groups were enriched in T cell activation, T cell receptor binding, integrin binding, and nuclear division (**Figure 7A**). KEGG enrichment showed the differential genes between two groups related to PI3K-Akt signaling pathways, cell adhesion molecules, and leukocyte differentiation (**Figure 7B**). GSEA was further used to investigate the key signaling pathways in different risk groups (**Figures 7C,D**). The results revealed that chromosome segregation was enriched in the high-risk group while T cell

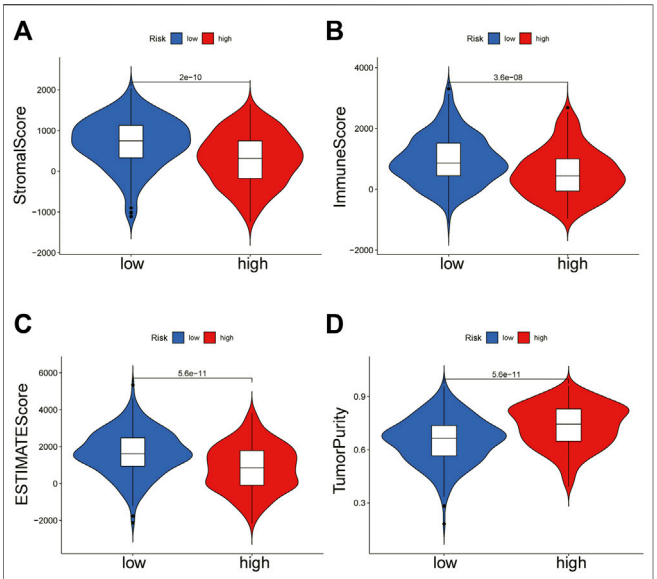


FIGURE 8 | ESTIMATE algorithm calculates immune score. The violin chart showed the comparison of stromal scores **(A)**, immune scores **(B)**, ESTIMATE scores **(C)**, and tumor purity **(D)** between the high-risk group and the low-risk group (Wilcoxon test, $p < 0.05$).

activation, immune response, antigen receptor-mediated signaling pathway were enriched in the low-risk group. These findings explained the poor survival in the high-risk group and may help us gain insight into the implication of aberrant glycosylation-related signature.

Tumor Immune Microenvironment of BRCA

To further explain the difference in survival between the two groups, we explored the relationship between glycosylation and the tumor immune microenvironment. Stromal scores, immune scores, and ESTIMATE scores were evaluated by the ESTIMATE package, all of which were higher in the low-risk group ($p < 0.001$; **Figures 8A–C**). In contrast, patients in the high-risk group were associated with significantly higher tumor purity ($p < 0.001$) (**Figure 8D**). After the proportion of 22 immune cell types was calculated by CIBERSORT, there were significant differences in the infiltration scores of nine immune cells between the two groups (**Figure 9A**), including B cell naive, plasma cell, T cell CD8 +, T cell CD4 + memory resting, T cell gamma delta, Macrophage M0, Macrophage M2, mast cell resting, and T cell follicular helper ($p < 0.05$). Subsequently, we used Spearman Correlation Analysis to explore the relationship between risk score and immune cell infiltration. The results showed that the low-risk group had a higher infiltration level of B cells naive, mast cells resting, plasma cells, T cells CD4 memory resting, T cells CD8, and T cells gamma delta ($p < 0.05$), while higher macrophages M0, macrophages M2 and T cells follicular helper were found in the high-risk group ($p < 0.05$) (**Supplementary Figure S3**). The ssGSEA was performed to quantify the enrichment scores of 13 immune cell-related functions between the two risk groups. The results showed that the scores of APC co-inhibition, CCR, Checkpoint, cytolytic activity, HLA, Inflammation-promoting, Para inflammation, T cells co-stimulation, Type I IFN response, Type II IFN response of patients in the low-risk group were higher ($p < 0.01$, **Figure 9B**), indicating that the low-risk group had higher immune infiltration than the high-risk group did. We further explored the immune checkpoints, and the result showed that the distribution of immune checkpoint-related molecule expression was significantly different between the high- and low-risk groups. The difference analysis confirmed that the expression of 22 immune promoting checkpoints (TNFRSF4, SELP, TLR4, CD40, ENTPD1, CXCL9, TNFRSF18, PRF1, CD28, TNFRSF14, ICAM1, CD40LG, ICOS, CD27, IL12A, IFNG, GZMA, ITGB2, BTN3A2, CCL5, CX3CL1, BTN3A1) in the low-risk group was higher, indicating that the prognosis of the low-risk group was better than that of the high-risk group ($p < 0.05$) (**Figure 9C**), that may provide the potential targets of immunotherapy for BRCA patients. In conclusion, the low-risk group had higher Tumor infiltrating lymphocytes (TILs), of which the prognosis was better, compared with the high-risk group.

Genomic Mutation Analysis

Through the downloaded somatic mutation data, the mutation frequency of the high-risk group and the low-risk group was calculated, and the waterfall chart was drawn to confirm the difference in the distribution of somatic mutations. It found that 169 of 198 (85.35%) BRCA samples in the high-risk group and 165 of 191 (86.39%) BRCA samples in the low-risk group displayed genetic mutations, and missense mutation was the most common variant classification. Moreover, in the high-risk group, PIK3CA had high genetic alterations (25%), which was just junior to the genetic alterations of TP53 (43%). In the low-risk group, PIK3CA had the most genetic alterations (41%) (**Figures 10A,B**). We also found that the TMB of patients in the high-risk group was significantly higher than that in the low-risk group ($p < 0.001$, **Figure 10C**), indicating that BRCA patients in the high-risk groups may derive good outcomes from immune checkpoint inhibitor treatments. The combined survival analysis showed that the prognosis of low-risk and low TMB patients was significantly better than that of high-risk and high TMB patients ($p < 0.01$, **Figure 10D**). Mutations of the genes for constructing the risk-scoring model and the genes with higher mutation frequency in the high-risk group showed that mutations in *SLC35A2* and *BST2* were mainly related to gene amplification, while mutations in *IGHA2*, *TP53*, and *TTN* were mainly related to missense mutations. By the CBIOPORTAL database, the domain where the mutations were located was defined, including *SLC35A2* mutations were localized in the nucleotide sugar transporter domain, *IGHA2* mutations were enriched in the immunoglobulin domain, the majority of *TP53* mutations were centralized in the *P53* DNA binding domain, and *TTN* mutations were localized in the immunoglobulin I-set domain, fibronectin type III domain, protein kinase domain, titin Z, and PPAK motif (**Figure 10E**).

Analysis of Immunotherapy Response and Drug Sensitivity

We explored the OS difference after immunotherapy between the high- and low-risk group in the phase II trial, in which

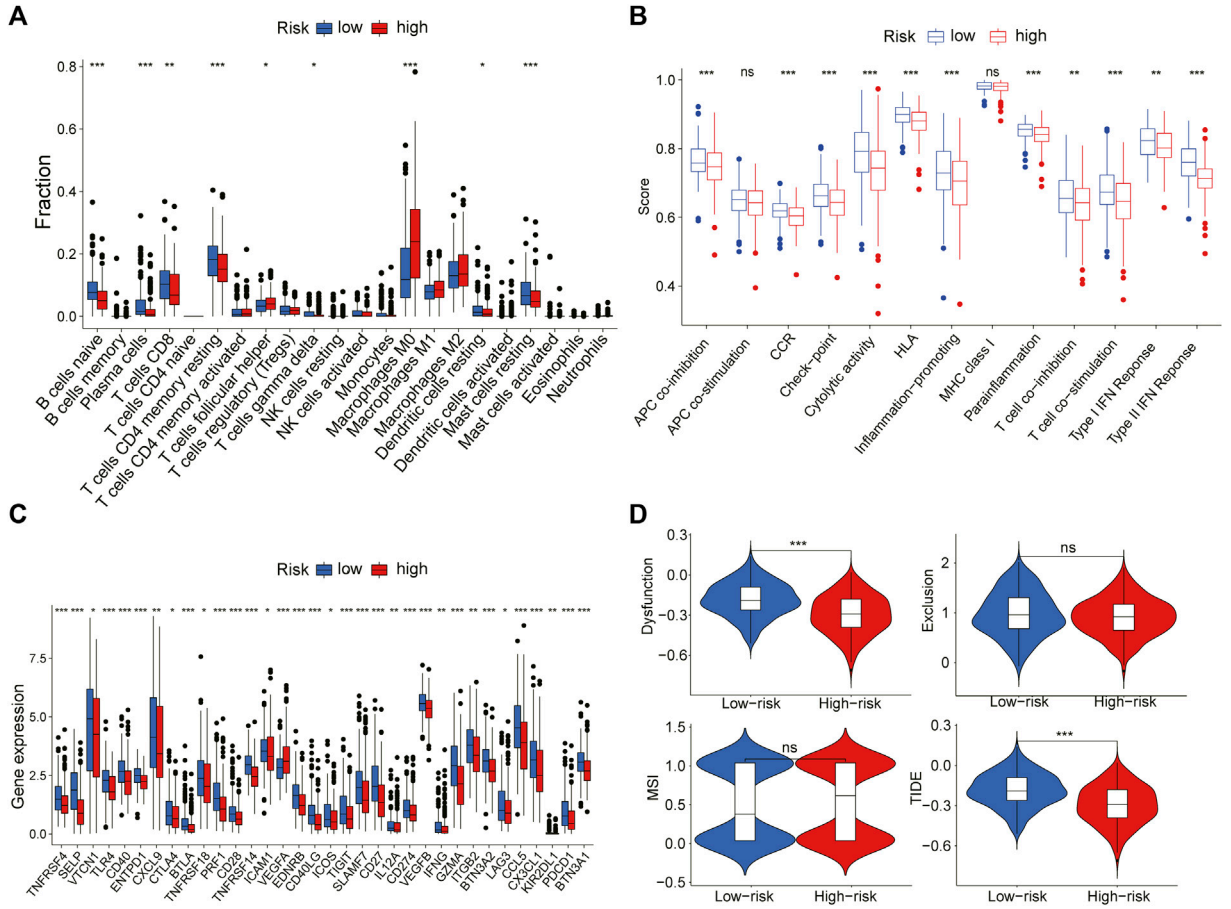


FIGURE 9 | The relation between risk scoring model and the immune microenvironment **(A)** The plot showed the estimated proportion of 22 immune cell in the high- and low-risk group **(B)** Box plot of differences in immune related ssGSEA scores between the two groups **(C)** Box plot of differences in the expression levels of immune related including immune checkpoint genes between the two groups **(D)** The violin plot showed the difference in dysfunction, exclusion, MSI and TIDE signature between two risk groups (Wilcoxon test, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, ns, no significant difference).

atezolizumab was treated with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor210). The patients in the high-risk group had a better effect in receiving immunotherapy ($p < 0.05$, **Figure 11A**). Although there was no significant statistical difference, it was found that the prognosis of the high-risk group was better than that of the low-risk group after immunotherapy ($p = 0.095$, **Figure 11B**). The better response to immunotherapy in the high-risk group may be due to the relatively higher TMB that we had demonstrated before. A similar correlation between TMB and our risk scoring model was identified in TCGA BLCA (**Supplementary Figure S4**). Then, TIDE was further used to assess the potential immunotherapy effect in the high- and low-risk groups. The high-risk group had a lower TIDE score, which represented a lower possibility of immune escape, suggesting the BRCA patients in the high-risk group could benefit more from immune checkpoint inhibitor therapy (**Figure 9D**). Besides, the low-risk group got a higher T-cell dysfunction score (**Figure 9D**). Through further analyzing the drug sensitivity by comparing the expression levels of model genes and drug response data from CELLMINER, we obtained the drugs with the most statistical significance (**Figure 12A**). The expression of *SLC35A2* was positively correlated with the sensitivity of two drugs (Vismodegib and Abiraterone) (HR > 1, $p < 0.05$), in turn, more sensitive in the high-risk group than in the low-risk group. The three-dimensional structures of the two drugs were obtained from the PubChem database, which provided potential guidance for chemotherapy in high-risk BRCA patients (**Figures 12B,C**).

DISCUSSION

Breast cancer is the most common cancer in women, accounting for nearly 25% of all cancer cases in women. It is also the leading cause of cancer deaths among elderly women (Hadadi et al., 2020; Siegel et al., 2021). The 5-year survival rate for patients with metastatic or stage IV breast cancer is 22% (www.cancer.org). The

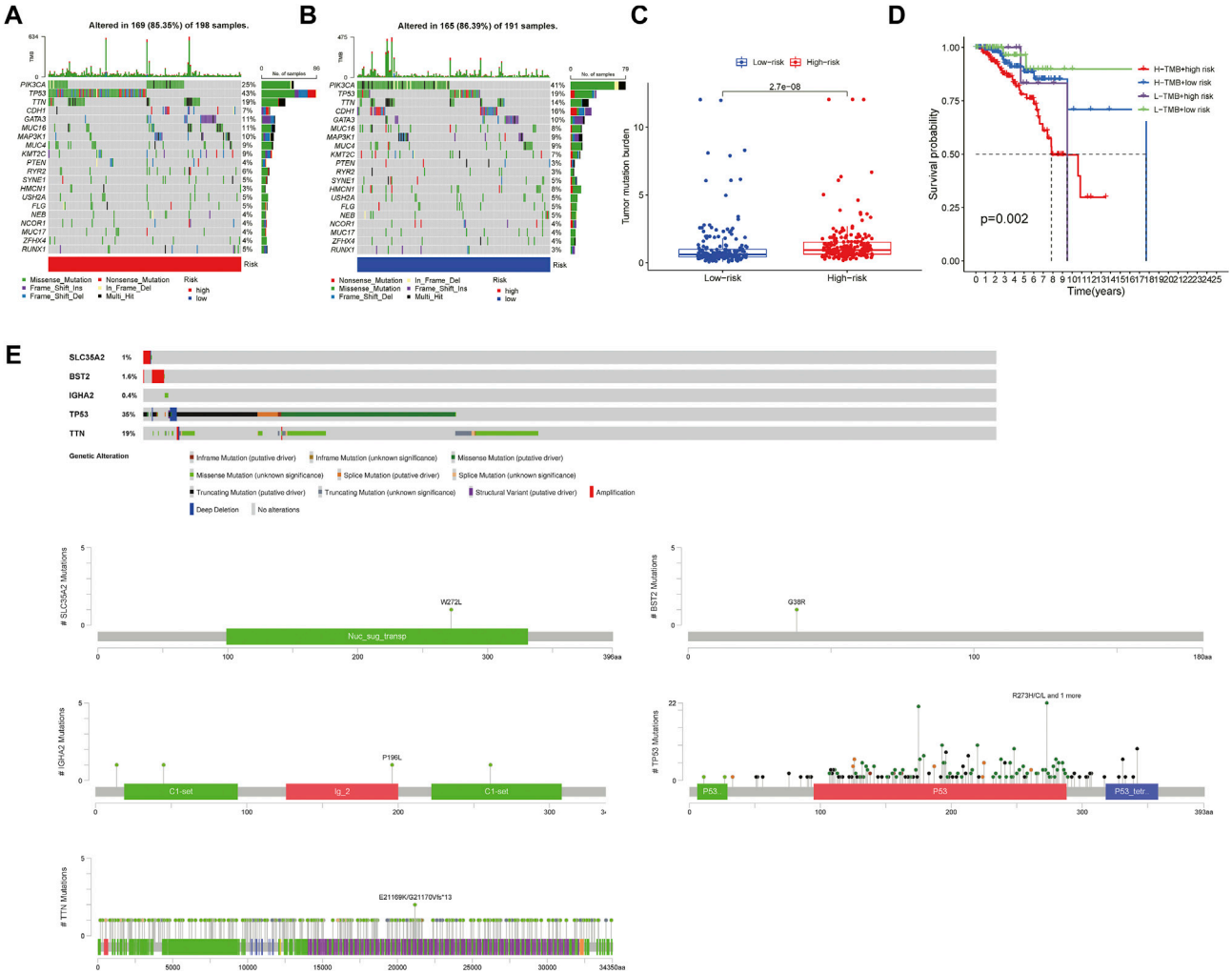


FIGURE 10 | Mutation analysis in the high-risk group and the low-risk group **(A,B)**. The waterfall chart showed the mutation frequency of the high- and low-risk groups **(C)** The box plot reflected the correlation between the risk score and the TMB scores (Wilcoxon test, $p < 0.05$) **(D)** The Survival curve showed the joint effect of TMB score and risk score on the prognosis of patients (log rank test, $p < 0.05$) **(E)** The plot showed the mutation type of key genes and functional domains where the mutations were enriched.

disease has heterogeneity and possesses a diverse mutational landscape, suggesting differences in patients' response to treatments and lack of targeted treatment for patients in specific breast cancer subtypes, which indicates the need to improve the guidance for treatment strategies. In recent years, with the progress of high-throughput sequencing and data analysis, it has become a vital biomedical research tool, which can be used for prognosis prediction, recurrence monitoring, and clinical stratification (Z. Wang et al., 2009b; I. D. Kyrochristos et al., 2019). Therefore, it is urgent to apply this tool to prevent and treat breast cancer. Many shreds of evidence had shown that aberrant glycosylation had multiple effects on cancer's occurrence, progression, invasion, and metastasis (T. D. Rao et al., 2017; J. Cui et al., 2018). Potapenko et al. reported not only that there are significant differences in the expression characteristics of glycosylation-related genes in breast cancer compared to normal breast tissue, but also that glycosylation -related genes show significant differences in expression between breast cancer subtypes and may be associated with patient prognosis and suggested that alterations in glycosylation pathways may occur at different time points in the carcinogenesis process (I. O. Potapenko et al., 2010; I. O. Potapenko et al., 2015). At present, there was still no prognostic model based on glycosylation-related genes for breast cancer. Given the critical impact of glycosylation on tumors, we had developed a prognostic model based on three glycosylation-related genes (*BST2*, *IGHA2*, *SLC35A2*).

Evidence showed that *SLC35A2* belonged to the solute carrier family *SLC35* of human nucleoside sugar transporters, and encoded an X-linked transporter that

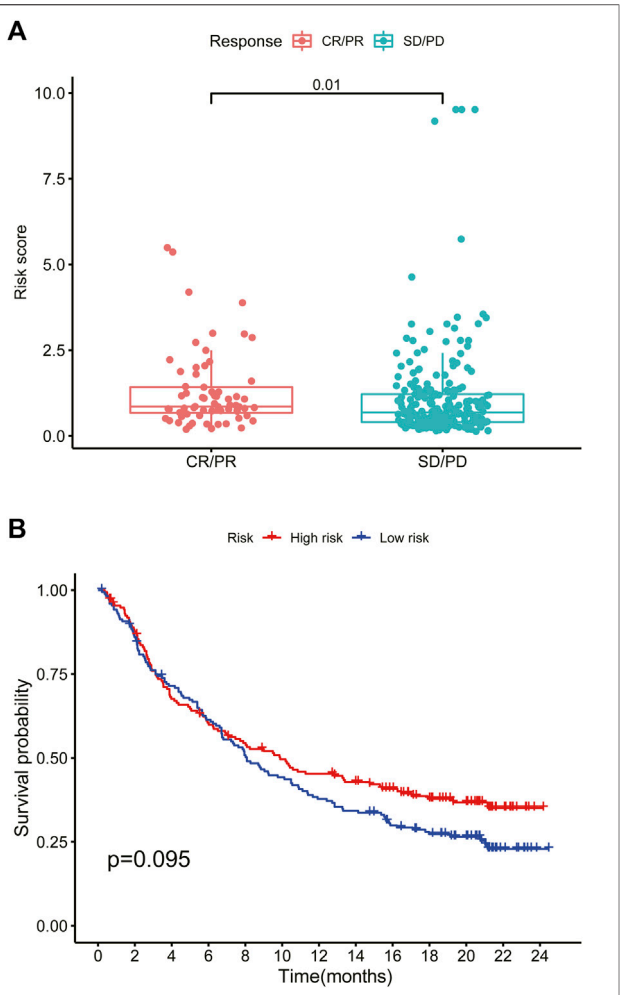


FIGURE 11 | High-risk scores predict immunotherapy response **(A)** Box plot of the effect of the high- and low-risk groups after immunotherapy (Wilcoxon test, $p < 0.05$) **(B)** Survival curve of the high- and low-risk groups after immunotherapy. CR, complete response. PR, partial response. SD, stable disease. PD, progressive disease (log rank test, $p < 0.05$).

transports uridine diphosphate—galactose from the cytoplasm to the lumen of the Golgi apparatus and endoplasmic reticulum. Pathogenic variation was associated with congenital glycosylation disorder characterized by epileptic encephalopathy (D. Quelhas et al., 2021). However, rare research had shown its relationship with cancer. Therefore, exploring the mechanism of *SLC35A2* is in high demand. In our study, the established prognostic model showed that the OS rate of patients in the high-risk group was significantly lower than that in the low-risk group through Kaplan Meier survival analysis ($p < 0.001$). Then, we testified the reliability of the model by using the validation set and external independent test sets. By comparing the risk score with other clinicopathological factors with a ROC curve, it was found that the risk model had a higher prognostic value than other clinicopathological factors. We also constructed a nomogram based on the age, TMN stage, and risk score to improve the accuracy of clinical decision-making. The risk score and nomogram had high reliability by calculating C-index and decision curve analysis. Subsequently, we divided the samples into multiple groups to demonstrate the application of the model in specific categories. The risk model had an excellent predictive value among age ≤65, ER positive, HER-2 negative, M0, N0, N1-N3, Stage I-II, Stage III-IV, T1-2, T3-4, non-TNBC subgroups ($p < 0.05$). In M1 and HER-2 positive subgroups, although there was no significant difference between the high- and low-risk groups, which might be ascribed to the small number of samples and the short follow-up time of patients in the subgroups, the survival time of the low-risk group was higher than that of the high-risk group. There was no difference in the ER negative and TNBC subgroups, resulting from the small sample size. We successfully validated the predictive value of the model in HER-2 positive and TNBC subgroups by METABRIC. Regarding the immunobiology of BRCA, TILs as an important biomarker in predicting the efficacy and outcome of treatment were worth exploring in depth. In breast cancer patients, loss of the anti-HER-2 $CD4^+$ Th1 immune response is independently correlated with disease recurrence (J. Datta et al., 2016). $CD8^+$ TILs, Th1 $CD4^+$ TILs can influence anti-tumor immune response in breast cancer (A. Basu et al., 2019). Several clinical trials have also shown an increased pathological complete response associated with a high density of TILs (Y. Issa-Nummer et al., 2013; C. Denkert et al., 2015). We explored the relationship between model grouping and scores of immune infiltrating cells. The results showed that the low-risk group had a higher infiltrative proportion of B cells naive, Mast cells resting, Plasma cells, T cells CD4 memory resting, T cells CD8, T cells gamma delta, which was in keeping with the previous investigations, suggesting a correlation between glycosylation and TILs. Therefore, impaired anti-tumor immune function may account for the poor prognosis in high-risk patients. Subsequently, we explored the differences in somatic mutations between the high- and the low-risk groups. The results showed that the TMB of the high-risk group was higher, indicating that patients in this group were more likely to receive the benefits of immune checkpoint inhibitor treatment (Steuer and Ramalingam, 2018). Moreover, the high-risk group had a lower TIDE score, which represented a lower possibility of immune escape, suggesting the BRCA patients in the high-risk group could benefit more from immune checkpoint inhibitor therapy (P. Jiang et al., 2018). The combined survival analysis showed that the prognosis of low-risk and low TMB patients was significantly better than that of high-risk and high TMB patients ($p < 0.01$). Then we analyzed chemotherapy based on the key genes. Two drugs (Vismodegib, Abiraterone) were expected to be validated and applied for the high-risk breast cancer patients.

This study also has some limitations. First, as it is a retrospective study derived from public data, it lacks some information such as recurrence time and treatment records. Second, clinical trials are urgently needed to confirm whether inducing glycosylation could improve the efficacy of immunotherapy in human BRCA patients.

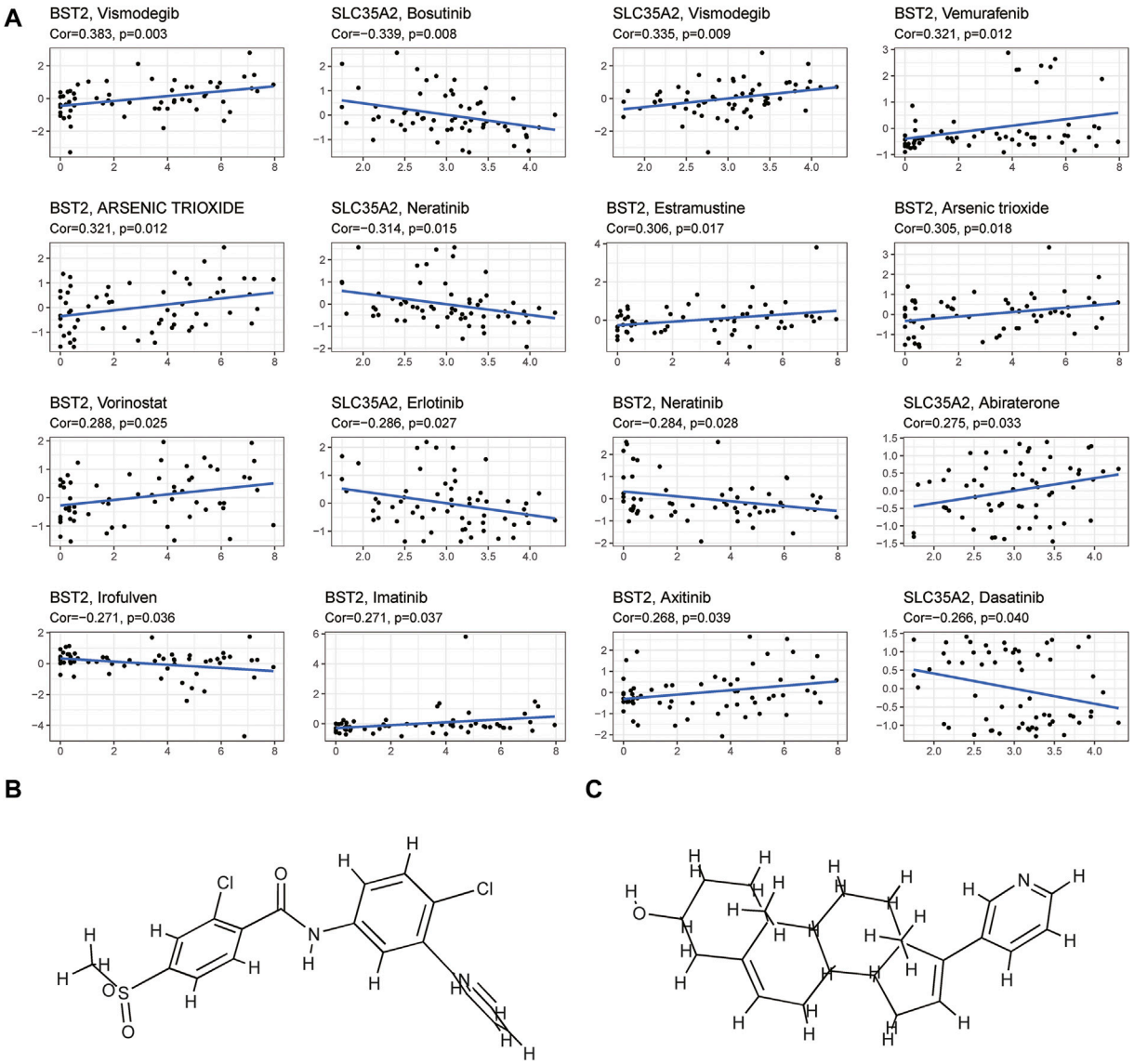


FIGURE 12 | Drug sensitivity analysis **(A)** The plot showed the correlation between the expression levels of key genes in the model and cell sensitivities to certain drugs ($p < 0.05$) **(B,C)** The plot showed the three-dimensional structure of two drugs (Vismodegib and Abiraterone) whose sensitivities were positively correlated with the expression of *SLC35A2*.

Furthermore, the specific molecular mechanism of the gene in the risk-scoring model has not been fully explored.

CONCLUSION

In conclusion, a novel prognostic model integrating glycosylation-related genes was firstly constructed. Through verification in the validation set and external independent test sets, the risk-scoring model has been proved to be an independent prediction model for predicting the prognosis of patients, which was correlated with immunotherapy effect and drug sensitivity. Moreover, we established a prognostic nomogram to predict the OS of patients with BRCA. The novel model might provide insights for predicting the prognosis of BRCA patients and suggestions to guide individual therapeutic strategies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The data could be downloaded at (https://www.cancer.gov; http://www.ncbi.nlm.nih.gov/geo; http://www.cbioportal.org; https://www.genecards.org)and the code used during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

SX and YL performed the data analysis and interpreted the results. SX and YL prepared the draft. YP, SX, and YL designed the research and revised the draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program (2018YFA0902801), Guangdong Provincial Key Laboratory of Digestive Cancer Research (No. 2021B1212040006), Natural Science Foundation of Guangdong Province, China (Grant No. 2020A1515010028), and the Scientific Technology Project of Shenzhen City (JCYJ20210324124003008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.834731/full#supplementary-material

REFERENCES

Basu, A., Ramamoorthi, G., Jia, Y., Faughn, J., Wiener, D., Awshah, S., et al. (2019). Immunotherapy in Breast Cancer: Current Status and Future Directions. *Adv. Cancer Res.* 143, 295–349. doi:10.1016/bs.acr.2019.03.006

Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., and Gianni, L. (2016). Triple-negative Breast Cancer: Challenges and Opportunities of a Heterogeneous Disease. *Nat. Rev. Clin. Oncol.* 13 (11), 674–690. doi:10.1038/nrclinonc.2016.66

Chakraborty, A., Dorsett, K. A., Trummell, H. Q., Yang, E. S., Oliver, P. G., Bonner, J. A., et al. (2018). ST6Gal-I Sialyltransferase Promotes Chemoresistance in Pancreatic Ductal Adenocarcinoma by Abrogating Gemcitabine-Mediated DNA Damage. *J. Biol. Chem.* 293 (3), 984–994. doi:10.1074/jbc.M117.808584

Cui, J., Huang, W., Wu, B., Jin, J., Jing, L., Shi, W.-P., et al. (2018). N- Glycosylation by N -acetylglucosaminyltransferase V Enhances the Interaction of CD147/basigin with Integrin β1 and Promotes HCC Metastasis. *J. Pathol.* 245 (1), 41–52. doi:10.1002/path.5054

Datta, J., Fracol, M., McMillan, M. T., Berk, E., Xu, S., Goodman, N., et al. (2016). Association of Depressed Anti-HER2 T-Helper Type 1 Response with Recurrence in Patients with Completely Treated HER2-Positive Breast Cancer. *JAMA Oncol.* 2 (2), 242–246. doi:10.1001/jamaoncol.2015.5482

Denkert, C., von Minckwitz, G., Brase, J. C., Sinn, B. V., Gade, S., Kronenwett, R., et al. (2015). Tumor-infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy with or without Carboplatin in Human Epidermal Growth Factor Receptor 2-positive and Triple-Negative Primary Breast Cancers. *Jco* 33 (9), 983–991. doi:10.1200/jco.2014.58.1967

Eccles, S. A., Aboagye, E. O., Ali, S., Anderson, A. S., Armes, J., Berditchevski, F., et al. (2013). Critical Research Gaps and Translational Priorities for the Successful Prevention and Treatment of Breast Cancer. *Breast Cancer Res.* 15 (5), R92. doi:10.1186/bcr3493

Fallahpour, S., Navaneelan, T., De, P., and Borgo, A. (2017). Breast Cancer Survival by Molecular Subtype: a Population-Based Analysis of Cancer Registry Data. *CMAJ Open* 5 (3), E734–e739. doi:10.9778/cmajo.20170030

Fang, K.-H., Kao, H.-K., Chi, L.-M., Liang, Y., Liu, S.-C., Hseuh, C., et al. (2014). Overexpression of BST2 Is Associated with Nodal Metastasis and Poorer Prognosis in Oral Cavity Cancer. *The Laryngoscope* 124 (9), E354–E360. doi:10.1002/lary.24700

Freire, T., Lo-Man, R., Bay, S., and Leclerc, C. (2011). Tn Glycosylation of the MUC6 Protein Modulates its Immunogenicity and Promotes the Induction of Th17-Biased T Cell Responses. *J. Biol. Chem.* 286 (10), 7797–7811. doi:10.1074/jbc.M110.209742

Gradishar, W. J., Anderson, B. O., Abraham, J., Aft, R., Agnese, D., Allison, K. H., et al. (2020). Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 18 (4), 452–478. doi:10.6004/jnccn.2020.0016

Hadadi, E., Taylor, W., Li, X.-M., Aslan, Y., Villote, M., Rivière, J., et al. (2020). Chronic Circadian Disruption Modulates Breast Cancer Stemness and Immune Microenvironment to Drive Metastasis in Mice. *Nat. Commun.* 11 (1), 3193. doi:10.1038/s41467-020-16890-6

Issa-Nummer, Y., Darb-Esfahani, S., Loibl, S., Kunz, G., Nekljudova, V., Schrader, I., et al. (2013). Prospective Validation of Immunological Infiltrate for Prediction of Response to Neoadjuvant Chemotherapy in HER2-Negative Breast Cancer - A Substudy of the Neoadjuvant GeparQuinto Trial. *PLoS One* 8 (12), e79775. doi:10.1371/journal.pone.0079775

Jerusalem, G., Collignon, J., Schroeder, H., and Lousberg, L. (2016). Triple-negative Breast Cancer: Treatment Challenges and Solutions. *Bctt* 8, 93–107. doi:10.2147/bctt.s69488

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Kang, S., Maeng, H., Kim, B. G., Qing, G. M., Choi, Y. P., Kim, H. Y., et al. (2012). *In Situ* identification and Localization of IGHA2 in the Breast Tumor Microenvironment by Mass Spectrometry. *J. Proteome Res.* 11 (9), 4567–4574. doi:10.1021/pr3003672

Kyrochristos, I. D., Ziogas, D. E., and Roukos, D. H. (2019). Dynamic Genome and Transcriptional Network-based Biomarkers and Drugs: Precision in Breast Cancer Therapy. *Med. Res. Rev.* 39 (3), 1205–1227. doi:10.1002/med.21549

Lin, Z., Xu, Q., Miao, D., and Yu, F. (2021). An Inflammatory Response-Related Gene Signature Can Impact the Immune Status and Predict the Prognosis of Hepatocellular Carcinoma. *Front. Oncol.* 11, 644416. doi:10.3389/fonc.2021.644416

Liu, W., Cao, Y., Guan, Y., and Zheng, C. (2018). BST2 Promotes Cell Proliferation, Migration and Induces NF-Kb Activation in Gastric Cancer. *Biotechnol. Lett.* 40 (7), 1015–1027. doi:10.1007/s10529-018-2562-z

Liu, W., Li, Y., Feng, S., Guan, Y., and Cao, Y. (2020). MicroRNA-760 Inhibits Cell Viability and Migration through Down-Regulating BST2 in Gastric Cancer. *J. Biochem.* 168 (2), 159–170. doi:10.1093/jb/mvaa031

Mahauad-Fernandez, W. D., DeMali, K. A., Olivier, A. K., and Okeoma, C. M. (2014). Bone Marrow Stromal Antigen 2 Expressed in Cancer Cells Promotes Mammary Tumor Growth and Metastasis. *Breast Cancer Res.* 16 (6), 493. doi:10.1186/s13058-014-0493-8

Mahauad-Fernandez, W. D., and Okeoma, C. M. (2017). Cysteine-linked Dimerization of BST-2 Confers Anoikis Resistance to Breast Cancer Cells by Negating Proapoptotic Activities to Promote Tumor Cell Survival and Growth. *Cell Death Dis* 8 (3), e2687. doi:10.1038/cddis.2017.68

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126 (5), 855–867. doi:10.1016/j.cell.2006.08.019

Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L. L., and Ferreira, J. A. (2019). Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. *Front. Oncol.* 9, 380. doi:10.3389/fonc.2019.00380

Pinho, S. S., and Reis, C. A. (2015). Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* 15 (9), 540–555. doi:10.1038/nrc3982

Potapenko, I. O., Haakensen, V. D., Lüders, T., Helland, Å., Bukholm, I., Sørlie, T., et al. (2010). Glycan Gene Expression Signatures in normal and Malignant Breast Tissue; Possible Role in Diagnosis and Progression. *Mol. Oncol.* 4 (2), 98–118. doi:10.1016/j.molonc.2009.12.001

Potapenko, I. O., Lüders, T., Russnes, H. G., Helland, Å., Sørlie, T., Kristensen, V. N., et al. (2015). Glycan-related Gene Expression Signatures in Breast Cancer Subtypes; Relation to Survival. *Mol. Oncol.* 9 (4), 861–876. doi:10.1016/j.molonc.2014.12.013

Quelhas, D., Correia, J., Jaeken, J., Azevedo, L., Lopes-Marques, M., Bandeira, A., et al. (2021). SLC35A2-CDG: Novel Variant and Review. *Mol. Genet. Metab. Rep.* 26, 100717. doi:10.1016/j.ymgmr.2021.100717

Rao, T. D., Fernández-Tejada, A., Axelrod, A., Rosales, N., Yan, X., Thapi, S., et al. (2017). Antibodies against Specific MUC16 Glycosylation Sites Inhibit Ovarian Cancer Growth. *ACS Chem. Biol.* 12 (8), 2085–2096. doi:10.1021/acschembio.7b00305

Sayeed, A., Luciani-Torres, G., Meng, Z., Bennington, J. L., Moore, D. H., and Dairkee, S. H. (2013). Aberrant Regulation of the BST2 (Tetherin) Promoter Enhances Cell Proliferation and Apoptosis Evasion in High Grade Breast Cancer Cells. *PLoS One* 8 (6), e67191. doi:10.1371/journal.pone.0067191

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

Silver, D. P., Richardson, A. L., Eklund, A. C., Wang, Z. C., Szallasi, Z., Li, Q., et al. (2010). Efficacy of Neoadjuvant Cisplatin in Triple-Negative Breast Cancer. *Jco* 28 (7), 1145–1153. doi:10.1200/jco.2009.22.4725

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A. M., Yu, J., Klijn, J. G. M., et al. (2008). Subtypes of Breast Cancer Show Preferential Site of Relapse. *Cancer Res.* 68 (9), 3108–3114. doi:10.1158/0008-5472.can-07-5644

Steuer, C. E., and Ramalingam, S. S. (2018). Tumor Mutation Burden: Leading Immunotherapy to the Era of Precision Medicine? *Jco* 36 (7), 631–632. doi:10.1200/jco.2017.76.8770

Wang, W., Nishioka, Y., Ozaki, S., Jalili, A., Abe, S., Kakiuchi, S., et al. (2009a). HM1.24 (CD317) Is a Novel Target against Lung Cancer for Immunotherapy Using anti-HM1.24 Antibody. *Cancer Immunol. Immunother.* 58 (6), 967–976. doi:10.1007/s00262-008-0612-4

Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-seq: a Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* 10 (1), 57–63. doi:10.1038/nrg2484

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xu, Liu, Ma, Fang, Wei, Li, Lu, Zheng, Liu, Zhu, Xu and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 06 April 2022
doi: 10.3389/fgene.2022.851390

Identification and Validation of Chromobox Family Members as Potential Prognostic Biomarkers and Therapeutic Targets for Human Esophageal Cancer

Xuefen Fang[1,2], *Junjun Wang*[3], *Jiabing Chen*[1,2], *Mingkai Zhuang*[1,2], *Tingxuan Huang*[1,2], *Zhixin Chen*[1,2], *Yuehong Huang*[1,2], *Biyun Zheng*[1,2,4]* and *Xiaozhong Wang*[1,2]*

[1]Department of Gastroenterology and Fujian Institute of Digestive Disease, Fujian Medical University Union Hospital, Fuzhou, China, [2]Fujian Medical University Cancer Center, Fujian Medical University, Fuzhou, China, [3]Department of Clinical Laboratory, Fujian Provincial Hospital Southern Branch, Fuzhou, China, [4]Department of Endoscopy Center, Fudan University Shanghai Cancer Center, Shanghai, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Naoko Hattori,
National Cancer Center Research Institute, Japan
Lin Shan,
Capital Medical University, China

***Correspondence:**
Biyun Zheng
drzhengby@163.com
Xiaozhong Wang
wangxiaozhong@fjmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 09 January 2022
Accepted: 11 March 2022
Published: 06 April 2022

Citation:
Fang X, Wang J, Chen J, Zhuang M, Huang T, Chen Z, Huang Y, Zheng B and Wang X (2022) Identification and Validation of Chromobox Family Members as Potential Prognostic Biomarkers and Therapeutic Targets for Human Esophageal Cancer. Front. Genet. 13:851390. doi: 10.3389/fgene.2022.851390

Background: Chromobox family proteins (CBXs) are vital components of epigenetic regulation complexes and transcriptionally inhibit target genes by modifying the chromatin. Accumulating evidence indicates that CBXs are involved in the initiation and progression of multiple malignancies. However, the expression, function, and clinical relevance such as the prognostic and diagnostic values of different CBXs in esophageal carcinoma (ESCA) are still unclear.

Methods: We applied Oncomine, TCGA, GEO, GEPIA, UALCAN, Kaplan–Meier plotter, cBioPortal, Metascape, and TIMER to investigate the roles of CBX family members in ESCA. Additionally, quantitative real-time PCR (RT-PCR), western blot, and immunofluorescence were used to verify the expression of CBX family members in ESCA clinical samples.

Results: Compared with normal tissues, the mRNA expression levels of CBX1/3/8 were significantly increased in ESCA, whereas CBX7 mRNA expression was reduced in both the TCGA cohort and GEO cohort. In the TCGA cohort, ROC curves suggested that CBX1/2/3/4/8 had great diagnostic value in ESCA, and the AUCs were above 0.9. Furthermore, upregulation of CBX1/3/8 and downregulation of CBX7 were closely related to the clinicopathological parameters in ESCA patients, such as tumor grades, tumor nodal metastasis status, and TP53 mutation status. The survival analysis indicated that higher CBX1/3/8 mRNA expressions and lower CBX7 expression suggested an unfavorable prognosis in ESCA. High genetic change rate (52%) of CBXs was found in ESCA patients. Functions and pathways of mutations in CBXs and their 50 frequently altered neighbor genes in ESCA patients were investigated; the results showed that DNA repair and DNA replication were correlated to CBX alterations. Moreover, we found a significant correlation between the expression level of CBX family members and the infiltration of immune cells in ESCA. Finally, we verified the expression of CBX family members in clinical samples and found the results were consistent with the databases.

Conclusion: Our study implied that CBX1/3/7/8 are potential targets of precision therapy for ESCA patients and new biomarkers for the prognosis.

Keywords: chromobox (CBX) protein, esophageal cancer, bioinformatics analysis, prognosis, biomarker, immunofluorescence

INTRODUCTION

ESCA is the seventh most frequent cancer and the sixth leading cause of cancer death, which accounts for almost 572,000 new cases and more than 509,000 deaths annually (Bray et al., 2018). Although prominent progress in the diagnosis and treatment of ESCA has been achieved, clinical outcomes for ESCA patients remain dismal, with 15–25% five-year overall survival rate worldwide (Pennathur et al., 2013). Therefore, the management of ESCA patients remains a considerable therapeutic challenge. It is extremely urgent to investigate the underlying mechanism of the carcinogenesis of ESCA, which will provide novel insights into the discovery of innovative therapeutic targets and diagnostic biomarkers.

Increasing evidence has suggested that aberrant epigenetic regulation influences the pathogenesis and progression of ESCA (Cao et al., 2020). The Polycomb group (PcG) complex, involved in the important epigenetic mechanism that regulates gene expression through chromatin mollification, plays principal roles in maintaining stem cell pluripotency and senescence and is implicated in cancer and other diseases (Chan and Morey, 2019). PcG complexes can be assembled into two distinct multi-protein complexes known as Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2), which are associated with gene silencing via histone-modifying activities (Morey and Helin, 2010; Chan and Morey, 2019). As canonical components of PRC1, CBX family members have been shown to control the tumorigenesis and progression of several human malignancies by increasing the tumor stem cells' self-renewal (Klauke et al., 2013; Gil and O'Loghlen, 2014).

Until now, a total of eight members of the CBX family (CBX1/2/3/4/5/6/7/8) have been identified in the human genome, with links mainly to heterochromatin, gene expression, apoptosis, and developmental program (Kaustov et al., 2011; Ma et al., 2014). CBXs are categorized into two types based on their molecular structure: 1) heterochromatin protein 1 (HP1) includes CBX1, CBX3, and CBX5 that share an N-terminal chromosome and a C-terminal chromophore domain; 2) polycomb (Pc) includes CBX2, CBX4, CBX6, CBX7, and CBX8 with a conserved N-terminal chromodomain and a C-terminal polycomb repressor box (Wotton and Merrill, 2007).

Deregulation of CBXs is associated with tumorigenesis of various cancer types and has significant prognostic value. Multi-omics integrative analysis revealed the antagonistic roles of CBX2 and CBX7 in metabolic reprogramming of breast cancer, which could predict patients' outcomes and sensitivity to FDA-approved/investigational drugs (Iqbal et al., 2021). In hepatocellular carcinoma (HCC), aberrant expressions of eight CBXs members were significantly associated with clinical cancer stages and pathological tumor grades (Ning et al., 2018). Higher mRNA expressions of CBX1/2/3/6/8 were related to shorter overall survival (OS) in HCC patients (Ning et al., 2018). Recent reports have indicated that higher mRNA expression of CBX3-6 and lower mRNA expression of CBX7 were significantly associated with poor prognosis and survival rate of gastric cancer patients (Ma et al., 2020). Upregulation of CBX2 in patients with ESCA was intensely related to poor disease-specific survival and recurrence rate (Ueda et al., 2020). Overexpression of CBX8 in ESCA was correlated with cell proliferation and predicted poor prognosis (Zhang et al., 2019). Conversely, it was reported that CBX8 could directly suppress the Snail promoter activity, contributing to inhibiting ESCA metastasis (Wang et al., 2017). CBXs may play both anti-tumor and pro-tumor roles depending on tumor types and cellular context. Therefore, it is necessary to clarify profoundly the distinct functions and prognostic value of CBX family members in ESCA.

To the best of our knowledge, this was the first study conducted to explore the potential oncogene values of CBX family members in ESCA using integration bioinformatics analysis. Due to the rapid development of microarray technology and RNA-sequencing technology in the last decade, RNA and DNA research has taken a great revolution and become an essential component of biomedical research (Sealfon and Chu, 2011). In this regard, we analyzed the expression, clinical parameters, and genetic alterations of different CBX proteins in ESCA patients and predicted their prognostic values, utilizing thousands of gene expression or copy number variations published online. Overall, our results indicated that CBXs serve as effective prognostic biomarkers and potential targets for the research on the clinical intervention of ESCA.

MATERIALS AND METHODS

ONCOMINE

ONCOMINE (www.oncomine.org) is an accessible online cancer microarray database providing powerful, genome-wide expression analysis (Rhodes et al., 2004). In our study, data were obtained to assess the transcriptional expression of CBX proteins between different carcinomas and adjacent normal control tissues ESCA. The difference in transcriptional expression of CBXs in ESCA was determined by Student's t-test. Sufficient fold changes ≥1.5, significant p value < 0.05, and gene rank ≥ the top 10% were set as the threshold.

The Cancer Genome Atlas (TCGA) Database

We obtained the raw counts of RNA-sequencing data and corresponding clinical information of CBX family members of

TCGA ESCA tissue data from the TCGA dataset (http://tcga-data.nic.nih.gov/). RNAseq data in FPKM (Fragments Per kilobase per Million) format and log2 conversion for expression comparison between samples (Vivian et al., 2017). We draw receiver-operating characteristic (ROC) curves by using the "pROC" package. For Kaplan–Meier curves, p values and the hazard ratio (HR) with 95% confidence interval (CI) were generated by log-rank tests and unvaried Cox proportional hazards regression. CBX family member expressions and their correlation with the infiltration abundance of immune cells such as ADC [activated DC], B cells, CD8 T cells, Cytotoxic cells, DC, Eosinophils, and iDC [immature DC] in ESCA were evaluated using Spearman's correlation with TCGA ESCA in the project level 3 HTSeq-RNAseq FPKM format data and clinical data. All analytical methods previously mentioned and R packages were performed using R software version v3.3.6 (Liu et al., 2018). $p < 0.05$ was regarded as statistically significant.

The Gene Expression Omnibus (GEO)

The gene expression microarray of GSE20347 and GSE38129 were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) of the National Center for Biotechnology Information (NCBI) via the GEO 1010query package (Davis and Meltzer, 2007). Altogether, 17 tumors and matched normal adjacent tissue samples were obtained from GSE20347, while 30 ESCC and 30 normal samples were obtained from GSE38129. The detection platform of the above expression microarrays was GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array.

GEPIA

GEPIA (http://gepia.cancer-pku.cn/) is a newly generated web server containing RNA sequence expression data of 9,736 tumors and 8,587 normal samples based on TCGA and the GTEx databases (Tang et al., 2017). We used the "Single Gene Analysis" module to perform differential mRNA expression analysis according to pathological stages, survival analysis, and correlation analysis. "Multiple Gene Comparison" module of GEPIA was utilized to assess the multiple gene comparison analysis of the CBX family, using the "ESCA" dataset. The p value cutoff was 0.05.

UALCAN

UALCAN (http://ualcan.path.uab.edu/analysis.html) is a comprehensive database based on level 3 RNA-seq and clinical data of 31 cancer types from The Cancer Genome Atlas (TCGA) and MET500 cohort data (Chandrashekar et al., 2017). Therefore, it can provide information about the relative transcriptional expression of genes in carcinomas compared with normal samples. Furthermore, the database also presents information on the association of transcriptional expression with relative clinicopathological features. In our study, we employed UALCAN to investigate the CBXs mRNA expressions and their relationship with clinicopathological parameters of ESCA. The significant difference of transcriptional expression between groups was evaluated using two-sample Student's t-test and $p < 0.01$ was regarded as statically significant.

Kaplan–Meier Plotter

The Kaplan–Meier plotter (http://kmplot.com/analysis/) is a tool that can give information on the effects of 54,000 genes on survival in 21 cancer types (Nagy et al., 2018). We made use of the Kaplan–Meier plotter to evaluate the association between the mRNA expression levels of CBX members with OS of ESCA patients. Information about the number-at-risk cases, median values of mRNA expression levels, the hazard ratio (HR), 95% confidence intervals (CIs), and log-rank p value can be accessed at the K-M plotter webpage. A statically significant difference was considered when p value was below 0.05.

cBioPortal

cBioPortal (www.cbioportal.org) is an open-access resource used to visualize and analyze multidimensional cancer genomics datasets (Cerami et al., 2012). Based on the TCGA database, genetic alterations of the CBX gene in patients with ESCA were obtained from cBioPortal. The obtained mRNA expression z-score threshold was 1.8 between the unaltered and altered patients.

Metascape

Metascape (http://metascape.org) is a free, well-maintained, user-friendly tool for gene annotation and gene list enrichment analysis (Zhou et al., 2019). We used Metascape to perform pathway and process enrichment analysis of the CBX genes and neighboring genes closely related to CBXs alteration via the "Custom Analysis" module, for GO and KEGG enrichment as well as protein-protein interaction analyses.

TIMER

TIMER (https://cistrome.shinyapps.io/timer) is a comprehensive resource that could provide systematic analyses with the dataset of 10,897 samples among diverse cancers in the TCGA database (Li et al., 2017). CBX family expression scatter plots and their correlation with the abundance of immune cells such as B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells in ESCA were assessed using Spearman's correlation with TCGA_ESCA datasets. The infiltration abundance for each somatic copy number alterations (SCNA) category was compared to the normal by a two-sided Wilcoxon rank-sum test, and statistical significance was identified as $p < 0.01$.

Clinical Samples

ESCA tissues and the corresponding adjacent normal tissues were collected from patients undergoing surgery at Fujian Union Hospital between April 2020 and April 2021. Patients who received neoadjuvant chemoradiotherapy were excluded. The Ethics Committee of the Fujian Medical University Union Hospital approved for use of all specimens, and all patients provided written informed consent. The tissues were stored in liquid nitrogen until use.

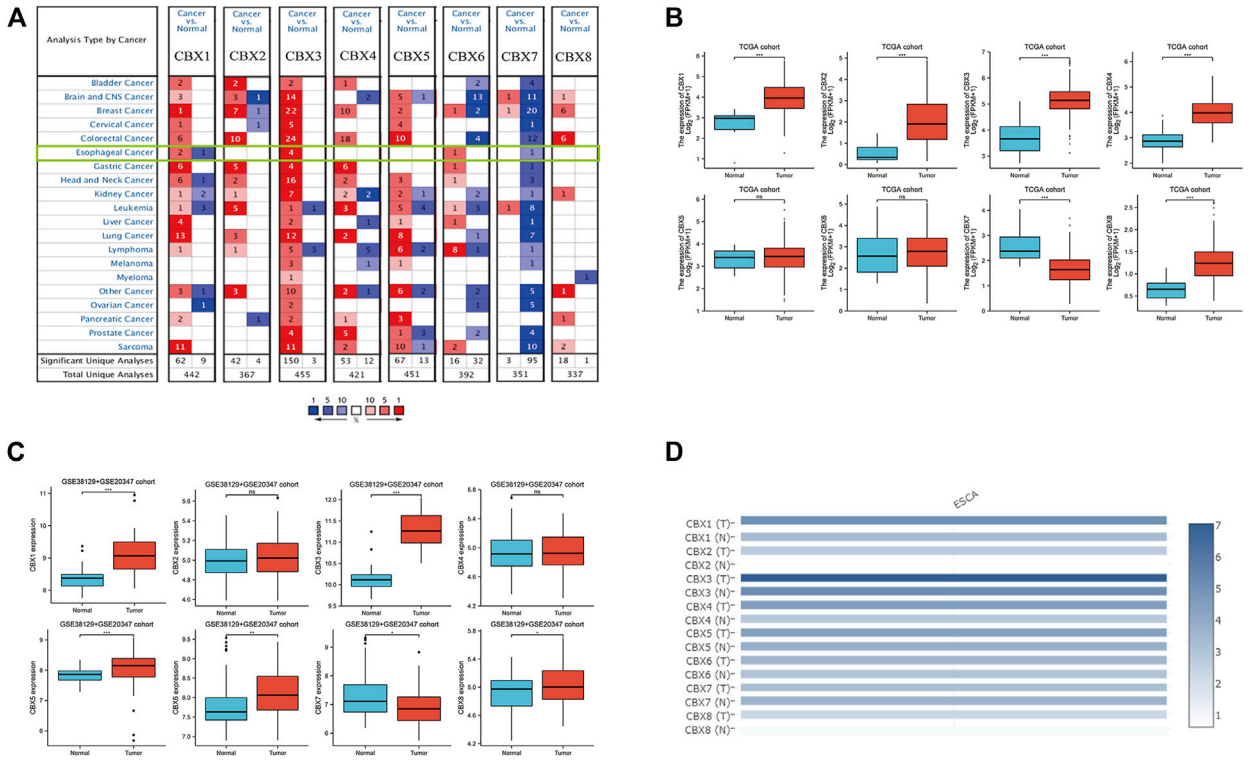


FIGURE 1 | Expression of CBX family members in ESCA and normal. **(A)** mRNA levels of CBX family members in various types of cancer (ONCOMINE). The figure shows the numbers of datasets with statistically significant mRNA overexpression (red) or downregulated expression (blue) of CBXs. Cutoff of *p* value: 0.05, fold change: 1.5, gene rank: 10%, data type: mRNA. **(B)** Transcriptional of CBX family members in ESCA and normal (TCGA). **(C)** Expression of CBX family members in ESCA and adjacent normal (GSE38129 + GSE20347 cohort). The *p* value was set at 0.05. **(D)** Relative mRNA expression level of CBXs in ESCA and normal tissues in GEPIA. Color intensity represents the fold change expression of the genes in the tissue. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

QRT-PCR Analysis

We quantified the level of CBX family members in tumor tissues ($n = 17$) as well as adjacent normal tissues ($n = 17$) obtained from patients with ESCA, using Quantitative real-time PCR. Total RNA from specimens were isolated using Trizol reagent (Invitrogen). Quantitative RT-PCR was performed with SYBR Green Real-Time Mix (Roche) by a 7,500 Real-time PCR according to the manufacturer's protocol (Applied Biosystems). PCR amplification was conducted in the following conditions: 95°C for 10 min, 95°C for 15 s (denature), and 60 °C for 1 min (anneal/extend) for 40 cycles, 95 °C for 15 s, and 60 °C for 1 min, and then 95 °C for 15 s (Mel curve). Primers are listed in **Supplementary Table S1**. The fold change was quantified via $2 - \Delta\Delta Ct$ [$\Delta\Delta Ct = (\Delta Ct$ of genes of interest$) - (\Delta Ct$ of β-actin)].

Western Blotting

We extracted protein from tumor tissues ($n = 12$) and adjacent normal tissues ($n = 12$) tissues in radioimmunoprecipitation assay (RIPA) buffer (Beyotime, Shanghai, China). The protein content was determined by the Bicinchoninic acid (BCA) assay kit (Beyotime, Shanghai, China). Proteins were separated by SDS-PAGE and were visualized using the ECL system (Bio-Rad, Hercules, CA, United States). Primary antibodies used in this study were as follows: CBX1 Polyclonal Antibody (Proteintech, 10241-2 -AP), CBX3 Polyclonal Antibody (Proteintech, 11650-2 -AP), CBX7 Polyclonal Antibody (Proteintech, 26278-1-AP), and CBX8 Monoclonal Antibody (Santa Cruz, sc-374332). Monoclonal Anti-GAPDH antibody was purchased from Cell Signaling Technology (Danvers, MA, United States).

Immunofluorescence Staining

ESCA tumor tissues and adjacent normal tissue were collected after surgery, then were fixed in 10% neutral buffered formalin and embedded in paraffin. Tissue samples were then sectioned at 4um thickness. Sections were deparaffinized, rehydrated, processed for antigen retrieval, blocked, incubated with primary antibody (CBX1:1:500 dilution, CBX3:1:500 dilution, CBX8:1:50 dilution) at 4 °C overnight, followed by incubation corresponding fluorescence-conjugated secondary antibody (1: 500 dilution) for 1 h at room temperature. Then slides were incubated with DAPI (BD Biosciences) for 10 min at RT and mounted with ProLong Gold antifade reagent (Invitrogen). Subsequently, the fluorescence images were captured with confocal microscopy.

TABLE 1 | Significant changes of CBX expression in the transcription level between ESCA and normal esophagus tissues (ONCOMINE).

	Types of ESCA vs. normal	Fold change	*p* value	*t*-test	Ref
CBX1	Esophageal squamous cell carcinoma	1.662	5.56E-13	8.101	Su Esophagus 2 (106)
	Esophageal adenocarcinoma	2.780	7.87E-4	3.720	Hao Esophagus (48)
CBX3	Esophageal squamous cell carcinoma	2.188	3.75E-9	15.587	Su Esophagus 2 (106)
	Esophageal squamous cell carcinoma	2.389	3.21E-10	10.596	Hu Esophagus (34)
	Esophageal adenocarcinoma	2.559	1.20E-4	4.887	Kimchi Esophagus (118)
	Barrett's esophagus	1.629	1.06E-5	5.127	Kim Esophagus (118)
CBX6	Esophageal squamous cell carcinoma	2.101	1.43E-6	6.336	Hu Esophagus (34)
CBX7	Esophageal squamous cell carcinoma	–1.621	5.17 E-8	5.770	Su Esophagus 2 (106)

ESCA: esophagus cancer; CBX: chromobox.

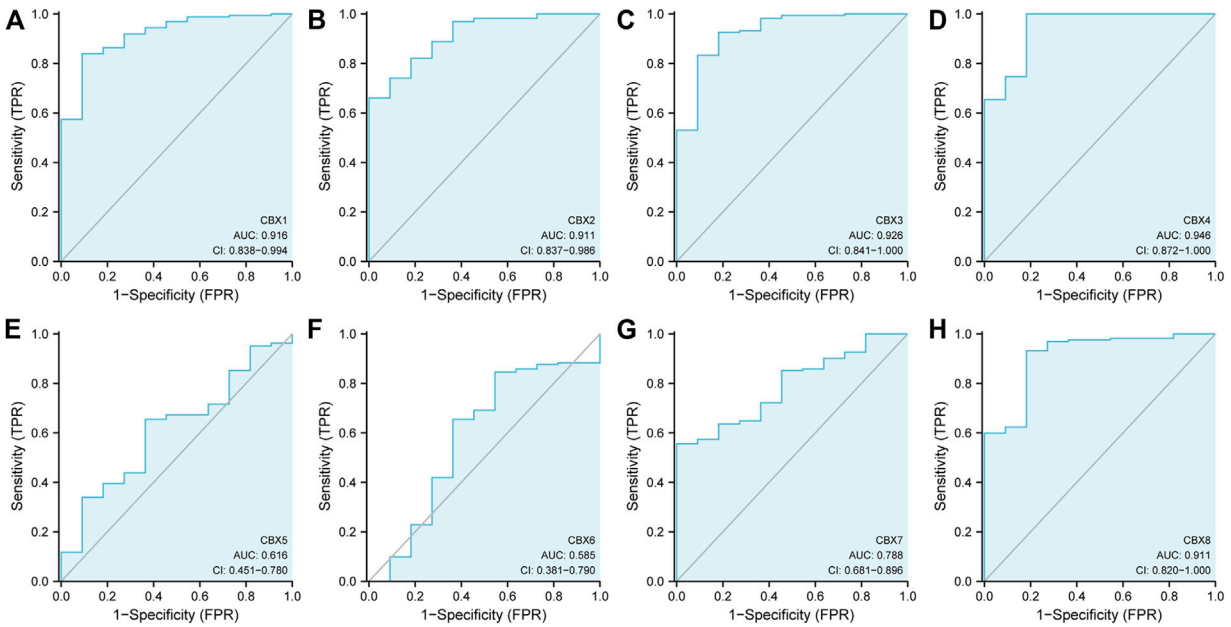


FIGURE 2 | ROC curve analysis of CBXs diagnostic capability in ESCA cancer **(A–H)** (TCGA): Normal versus tumor.

RESULTS

Aberrant Expression and Diagnostic Capability of CBX Family Members in ESCA Patients

The expression of CBX family members were collected using ONCOMINE database (**Figure 1A**). We first investigated the mRNA transcriptional levels of CBXs in ESCA and normal tissues with ONCOMINE. As shown in **Figure 1A** and **Table 1**, the transcriptional levels of CBX1, CBX3, and CBX6 in ESCA were significantly elevated while CBX7 was significantly reduced compared to normal tissues. TCGA ESCA cohort showed that the expression levels of CBX1 ($p < 0.001$), CBX2 ($p < 0.001$), CBX3 ($p < 0.001$), CBX4 ($p < 0.001$), and CBX8 ($p < 0.001$) were higher while the expression level of CBX7 ($p < 0.001$) was lower in ESCA samples compared to paired para-cancerous samples (**Figure 1B**). Then we evaluated the expression levels of CBXs in ESCA and normal tissues with the GEO database (GSE20347 and GSE38129). The transcriptional levels of CBX1 ($p < 0.001$), CBX3 ($p < 0.001$), CBX5 ($p < 0.001$), CBX6 ($p = 0.006$), and CBX8 ($p = 0.049$) in ESCA were significantly elevated, while CBX7 ($p = 0.015$) was decreased in ESCA tissues compared to normal tissues (**Figure 1C**). We also compared the relative expression levels of CBXs in ESCA using GEPIA, the results showed that the relative expression of CBX3 was the highest among all the CBX proteins (**Figure 1D**).

Next, we used the ROC curve of CBX family members to access the diagnostic capability of CBX family members for ESCA by using the TCGA cohort. The results indicated that CBX1, CBX2, CBX3, CBX4 and CBX8 had great diagnostic capability with AUC of 0.916(95% CI: 0.838–0.994),0.911(95% CI: 0.837–0.986),0.926(95% CI: 0.841–1.000),0.946(95% CI:

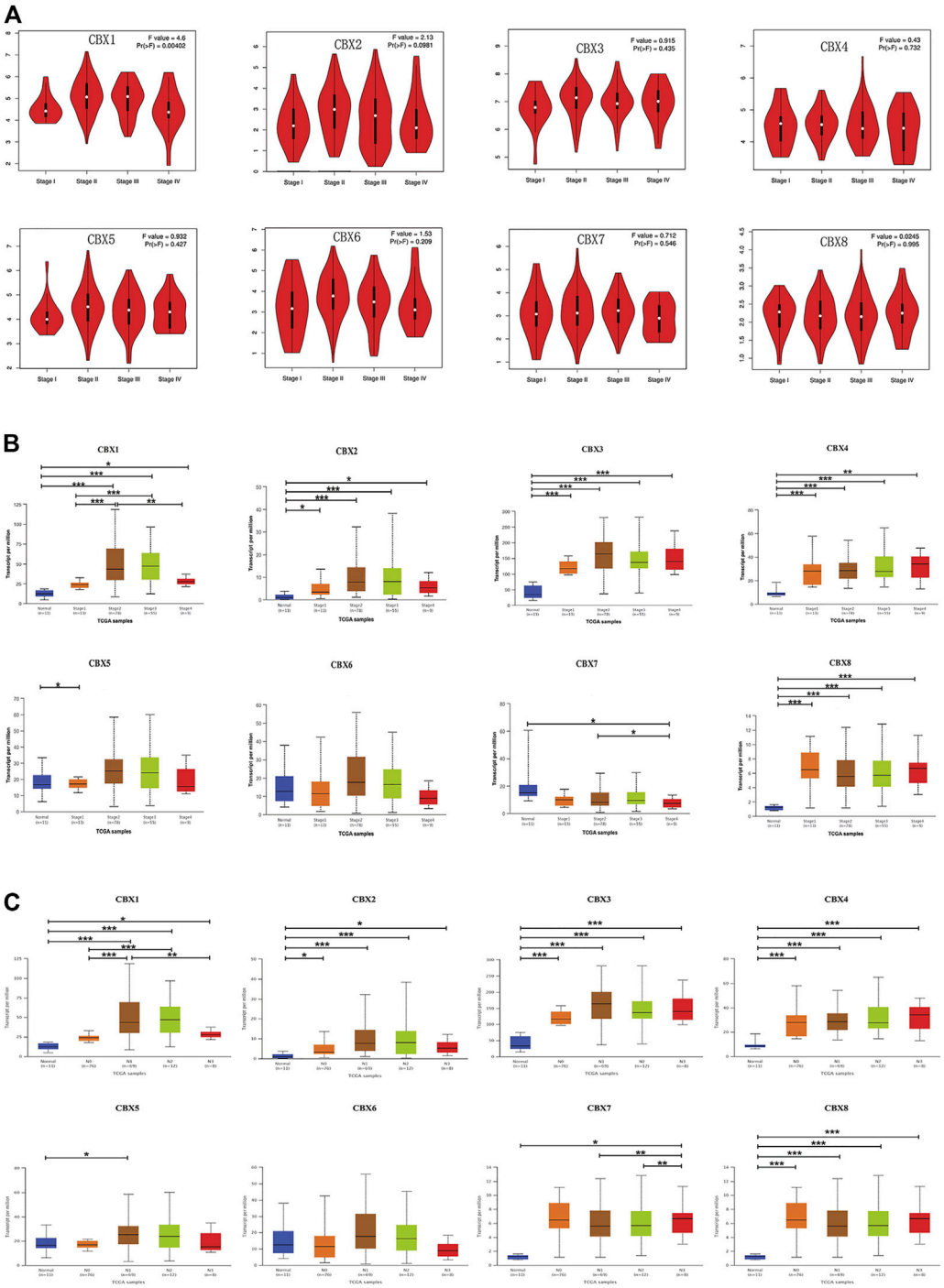


FIGURE 3 | (A) Correlation between different expressed CBXs and the pathological stage of ESCA patients in GEPIA. *$p < 0.05$. **(B)** Relationship between mRNA expression of distinct CBX family members and individual cancer stages of ESCA patients. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(C)** Association of mRNA expression of distinct CBXs family members with tumor nodal metastasis status of ESCA patients. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

0.872–1.000), and 0.911 (95% CI: 0.820–1.000), respectively. CBX7 had moderate diagnostic capability with AUC of 0.788 (95% CI: 0.681–0.896). CBX5 and CBX6 had low diagnostic capability with AUC of 0.616 (95% CI: 0.451–0.780) and 0.515 (95% CI: 0.381–0.790) (**Figure 2**), respectively.

In short, we found CBX1, CBX3, and CBX8 were significantly upregulated in ESCA vs. normal tissues, while CBX7 was downregulated in ESCA. Furthermore, CBX1, CBX2, CBX3, CBX4, and CBX8 had great diagnostic capability to distinguish ESCA from normal tissues.

Association of the mRNA Expression of CBX Family Members With the Clinicopathological Parameters of ESCA Patients

Since the mRNA expression of CBX family members was aberrant in ESCA patients, we next analyzed the relationship between the mRNA expression of different CBX family members and the clinicopathological parameters of ESCA patients. First, we utilized GEPIA to further detect the correlation between the mRNA expression levels of different CBX family members and the pathological stage of ESCA patients. We found a significant correlation between the expression of CBX1 (p = 0.004) and the pathological stage (**Figure 3A**). CBX1 expression was increased in the advanced stage (stage II/III) as compared with those in the early tumor stage (stage I). These data suggest that CBX1 plays a significant role in the tumorigenesis and progression of ESCA.

We also analyzed the relationship between the mRNA expression of different CBX family members and the cancer stages, tumor grades, tumor nodal metastasis status, and TP53 mutation status of ESCA patients by UALCAN (http://ualcan.path.uab.edu). As shown in **Figure 3B**, mRNA expressions of CBX1, CBX2, CBX3, CBX4, and CBX8 were upregulated in ESCA as compared with normal. The expression of CBX1 tended to be higher as the tumor stage increased, which was consistent with the previous findings in the GEPIA database. Nevertheless, the mRNA expression of CBX7 was the highest in normal tissues, and as the cancer stage increased, the mRNA expression of CBX7 tended to decrease. CBX6 mRNA expression had no significant relationship with tumor stages. Similarly, mRNA expressions of CBX1, CBX2, CBX3, CBX4, and CBX8 were significantly related to patients' tumor nodal metastasis status. Patients who were in more advanced tumor nodal metastasis status tended to express higher mRNA of CBX1. However, normal tissues had the highest mRNA expression of CBX7, and the mRNA expression of CBX7 tended to be lower as tumor nodal metastasis status increased (**Figure 3C**). We then assessed the relationship between the different expressions of CBX family members and the tumor grades of ESCA patients. As shown in **Supplementary Figure S1A**, mRNA expressions of CBX1, CBX2, CBX3, CBX4, and CBX8 were significantly associated with tumor grades. The mRNA expression of CBX1, CBX2, CBX3, CBX4, and CBX8 tended to be elevated as the tumor grade increased. However, the mRNA expression of CBX7 tended to be lower in grade 4 compared with grade 3. TP53 mutation has been recognized as the most common event and frequently occurs in ESCA (Yang et al., 2020). Interestingly, as shown in **Supplementary Figure S1B**, CBX family members' expressions were associated with TP53 mutation status in ESCA patients. CBX1, CBX2, CBX3, and CBX8 were upregulated in ESCA patients with TP53 mutation compared with normal tissues and TP53 non-mutation patients.

Taken together, these data suggested that the mRNA overexpression of CBX1, CBX3, and CBX8 were significantly related to tumor grades and patients' tumor nodal metastasis status. CBX1, CBX2, CBX3, and CBX8 were significantly related to TP53 mutation in ESCA patients.

The Prognostic Value of CBX Family Members in ESCA Patients

To evaluate the value of differentially expressed CBX family members in the progression of ESCA, we assessed the correlation between differentially expressed CBX family members and clinical outcomes using GEPIA. Disease-free survival curves (DFS) were presented in **Supplementary Figure S2A**. ESCA patients with lower transcriptional levels of CBX1 (HR = 1.6, p = 0.044) were significantly associated with longer DFS. The value of differentially expressed CBX in the overall survival of ESCA patients was also evaluated (**Supplementary Figure S2B**). CBX family members did not seem to have a significant effect on Overall Survival (OS).

We also analyzed the prognostic values of CBX family members in patients with ESCA using the TCGA database. Also, the result suggested that ESCA patients with higher CBX3 ((HR = 1.78, p = 0.026)) and CBX4 (HR = 1.94, p = 0.011) levels had shorter OS (**Supplementary Figure S3A**). Increased CBX4 (HR = 2.07, p = 0.017) and CBX5 (HR = 1.94, p = 0.03) mRNA levels were remarkably associated with shorter DFS (**Supplementary Figure S3B**). Moreover, increased CBX 1 (HR = 1.71, p = 0.045) mRNA level was significantly correlated with shorter Progress Free Interval (PFI) (**Supplementary Figure S3C**).

Likewise, we used the Kaplan–Meier plotter to analyze the prognostic values of CBX family members in patients with ESCA subgroup analysis (**Figure 4**). Esophageal squamous cell carcinoma patients with the higher transcriptional level of CBX4 (HR = 2.93, p = 0.008) were closely associated with shorter OS (**Figure Figure4A**). The overexpression of CBX3 (HR = 3.12, p = 0.00028) and CBX8 (HR = 2.27, p = 0.035) mRNAs in esophageal adenocarcinoma patients were significantly correlated with shorter OS, whereas the overexpression of CBX7 mRNA (HR = 0.48, p = 0.039) was significantly correlated with longer OS (**Figure 4B**).

Overall, these data demonstrated that increased CBX1/3/4/5/8 and decreased CBX7 were significantly associated with unfavorable clinical outcomes in ESCA patients.

Genetic Alteration, Expression, and Interaction Analyses of CBX Family Members in ESCA Patients

Using cBioPortal, we investigated the genetic alterations of CBX family members and discovered a high alteration frequency (52%) in ESCA patients. CBX3, CBX1, and CBX2 ranked as the top three genes with genetic alterations, which altered in 25, 15, and 15% of the queried ESCA samples, respectively. The primary alteration type in these samples was the enhanced mRNA expression (**Figure 5A**). Besides, we calculated the correlations of CBXs with each other in GEPIA online tool for correlation analysis in ESCA patients and included Pearson's correlation. The results showed that there were co-expression associations between the following CBX proteins: CBX1 positively with CBX2/3/4/5/6/8, CBX2 positively with CBX1/3/4/5/6/8, CBX3 positively with CBX1/2/5/8, CBX4 positively with CBX1/2/5/6/8, CBX5 positively with CBX1/2/3/4/6/7/8, CBX6 positively with CBX1/2/4/5/7/8, CBX7 positively with CBX5/6, and CBX8 positively

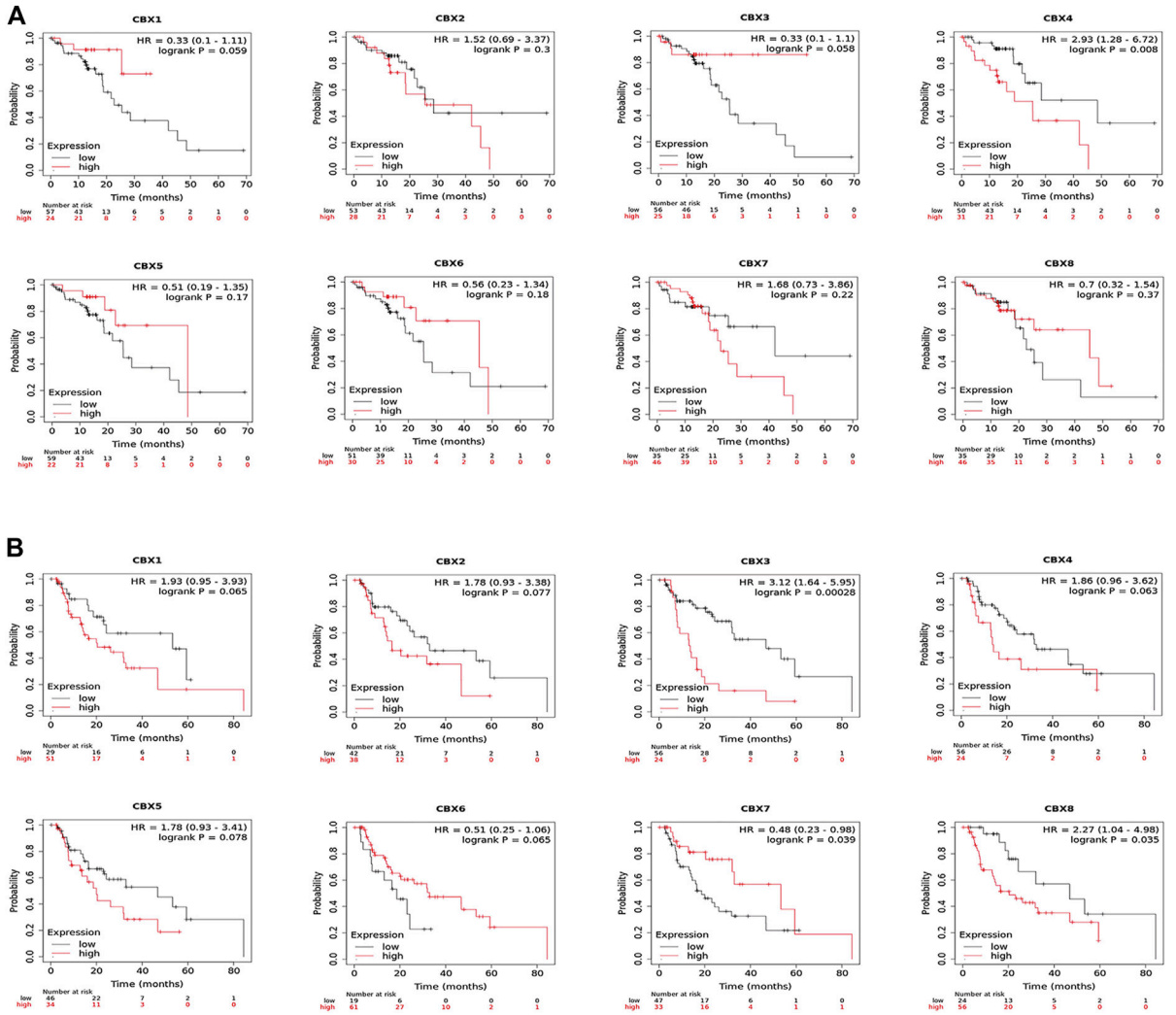


FIGURE 4 | (A) Prognostic value of CBXs in esophageal squamous cell carcinoma (ESCC) patients in the overall survival curve (Kaplan–Meier Plotter). **(B)** Prognostic value of CBX family members in esophageal adenocarcinoma (EAC) patients in the overall survival curve (Kaplan–Meier Plotter).

with CBX1/2/3/4/5/6 (**Figure 5B**). Furthermore, CBX1/2/3/8 were positively co-expressed with TP53 (**Figure 5C**).

Enrichment of CBX Family Members and Their 50 Frequently Altered Neighbor Genes' Ontology in ESCA Patients

Using cBioPortal, we found 50 genes were most associated with each CBX family member. Some genes were positively associated with CBX family members, whereas others were negatively associated with the proteins. We used Metascape for Gene Ontology (GO) analysis of each CBX protein for biological processes, cellular components, and molecular functions. Moreover, functions of CBX family members and their 50 frequently altered neighbor genes were analyzed by Kyoto Encyclopedia of Genes and Genome (KEGG) and protein-protein interaction (PPI) enrichment analyses in Metascape. As shown in **Figure 6A**, biological processes such as GO: 0015931 (nucleobase-containing compound transport), GO:0006260 (DNA replication), GO:0006281 (DNA repair), GO: 0000226 (microtubule cytoskeleton organization), and GO: 0030029 (actin filament-based process) were prominently regulated by the CBX family members mutations in ESCA. Cellular components, including GO: 0098687 (chromosomal region), GO: 0000781 (chromosome, telomeric region), GO: 0090734 (site of DNA damage), GO: 0005635 (nuclear envelope) and GO: 0034399 (nuclear periphery) were remarkably associated with the CBX family members alterations. Besides, CBX family members mutations also significantly affected the molecular functions, such as GO: 0003779 (actin-binding), GO: 0003682 (chromatin binding), GO: 0032138 (single base insertion or deletion binding), GO: 0003697 (single-stranded DNA binding), and GO: 0015932 (nucleobase-containing compound transmembrane transporter activity).

In KEGG analysis, 11 pathways including ko:03430 (mismatch repair), ko:04512 (ECM-receptor interaction), ko:04910(Insulin signaling pathway), ko:04310 (Wnt signaling pathway), and ko:

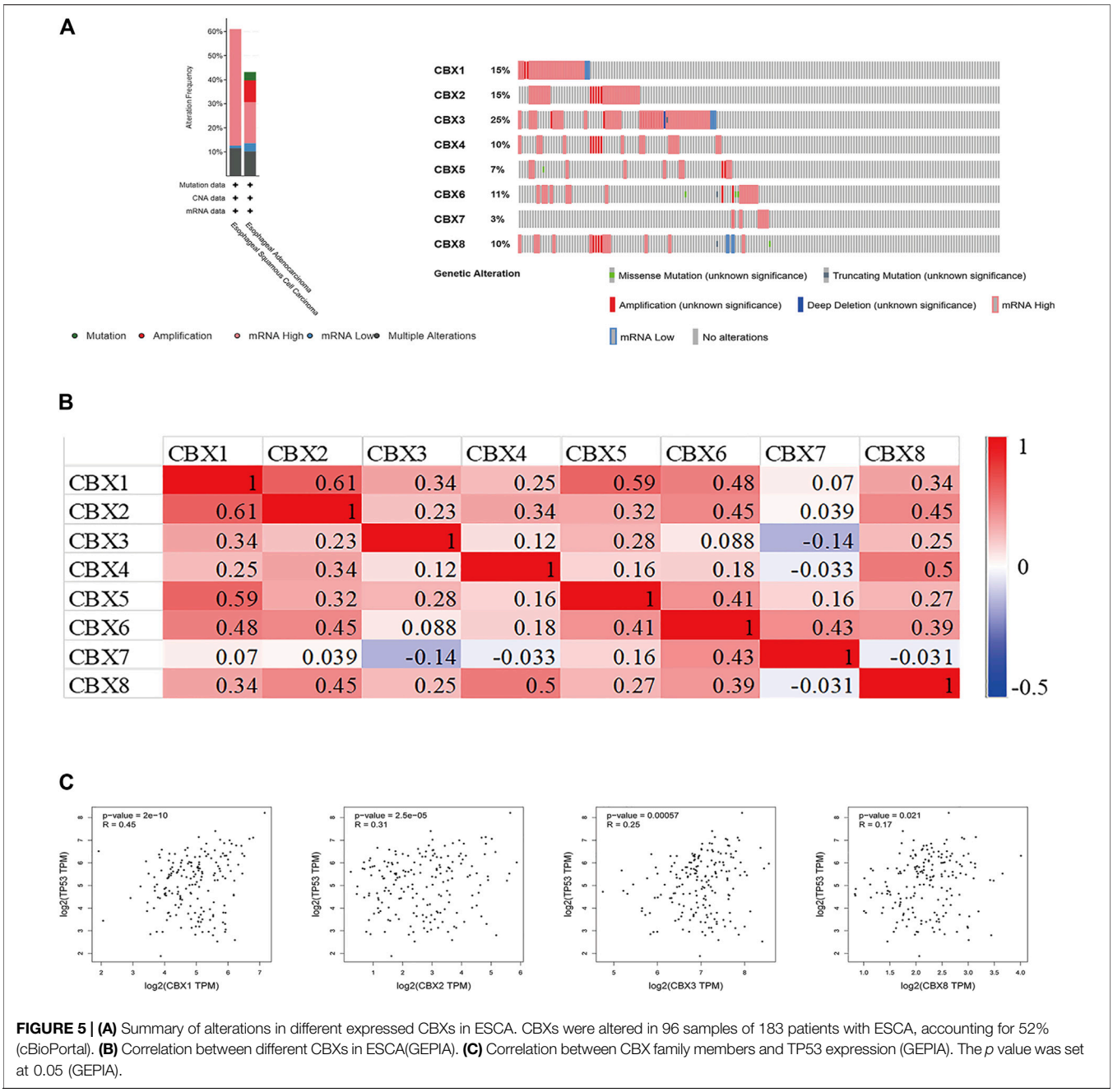
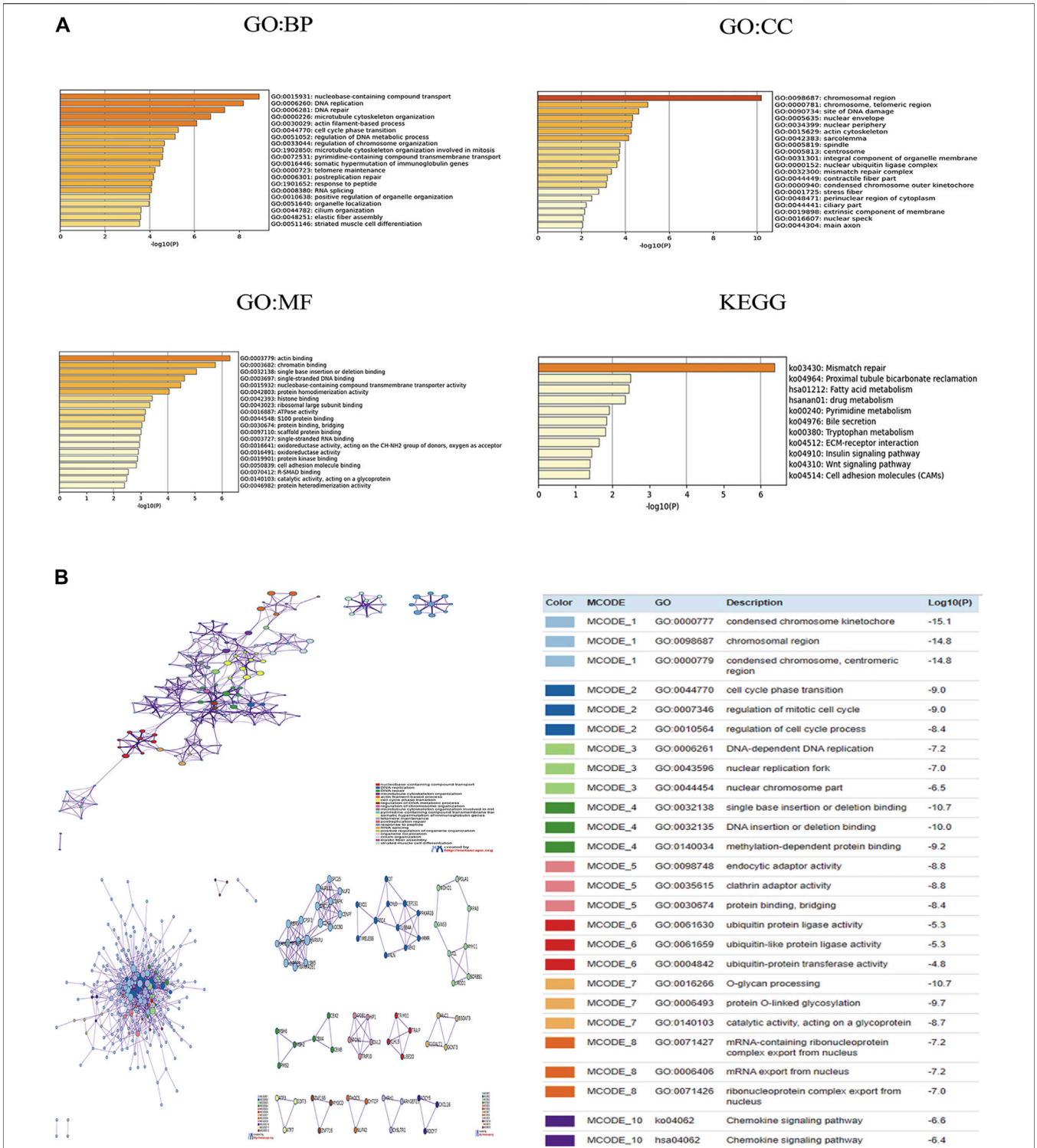


FIGURE 5 | (A) Summary of alterations in different expressed CBXs in ESCA. CBXs were altered in 96 samples of 183 patients with ESCA, accounting for 52% (cBioPortal). **(B)** Correlation between different CBXs in ESCA(GEPIA). **(C)** Correlation between CBX family members and TP53 expression (GEPIA). The *p* value was set at 0.05 (GEPIA).

04514(Cell adhesion molecules (CAMs) were associated with the functions of CBX family members mutations in ESCA (**Figure 6A**). **Figure 6B** presented the PPI network associated with the genes; it was mainly related to cell cycle, gene mutation, DNA replication, and epigenetics.

Immune Cell Infiltration Analysis of CBX Family Members in ESCA

Accumulating evidence indicated that immune cell infiltrations are closely related to tumor progression and clinical outcome. In this study, we also explored the correlation between CBX family members and immune cell infiltration in ESCA by utilizing the TIMER database (**Supplementary Figure S4**). There was a positive correlation between CBX1 expression and the infiltration of macrophages (Cor = 0.171, p = 2.20e-02), and a negative correlation between CBX1 expression and the infiltration of neutrophils (Cor = −0.15, p = 4.45e-02). CBX2 expression was negatively associated with the infiltration of $CD8^+$ T cells (Cor = −0.184, p = 1.35e-02). CBX3 expression was negatively associated with the infiltration of dendritic cells (Cor = −0.249, p = 7.38e-04). Similarly, the expression of CBX5 was positively associated with the infiltration of macrophages (Cor = 0.199, p = 7.27e-03). There was a



Color	MCODE	GO	Description	Log10(P)
	MCODE_1	GO:0000777	condensed chromosome kinetochore	-15.1
	MCODE_1	GO:0098687	chromosomal region	-14.8
	MCODE_1	GO:0000779	condensed chromosome, centromeric region	-14.8
	MCODE_2	GO:0044770	cell cycle phase transition	-9.0
	MCODE_2	GO:0007346	regulation of mitotic cell cycle	-9.0
	MCODE_2	GO:0010564	regulation of cell cycle process	-8.4
	MCODE_3	GO:0006261	DNA-dependent DNA replication	-7.2
	MCODE_3	GO:0043596	nuclear replication fork	-7.0
	MCODE_3	GO:0044454	nuclear chromosome part	-6.5
	MCODE_4	GO:0032138	single base insertion or deletion binding	-10.7
	MCODE_4	GO:0032135	DNA insertion or deletion binding	-10.0
	MCODE_4	GO:0140034	methylation-dependent protein binding	-9.2
	MCODE_5	GO:0098748	endocytic adaptor activity	-8.8
	MCODE_5	GO:0035615	clathrin adaptor activity	-8.8
	MCODE_5	GO:0030674	protein binding, bridging	-8.4
	MCODE_6	GO:0061630	ubiquitin protein ligase activity	-5.3
	MCODE_6	GO:0061659	ubiquitin-like protein ligase activity	-5.3
	MCODE_6	GO:0004842	ubiquitin-protein transferase activity	-4.8
	MCODE_7	GO:0016266	O-glycan processing	-10.7
	MCODE_7	GO:0006493	protein O-linked glycosylation	-9.7
	MCODE_7	GO:0140103	catalytic activity, acting on a glycoprotein	-8.7
	MCODE_8	GO:0071427	mRNA-containing ribonucleoprotein complex export from nucleus	-7.2
	MCODE_8	GO:0006406	mRNA export from nucleus	-7.2
	MCODE_8	GO:0071426	ribonucleoprotein complex export from nucleus	-7.0
	MCODE_10	ko04062	Chemokine signaling pathway	-6.6
	MCODE_10	hsa04062	Chemokine signaling pathway	-6.4

FIGURE 6 | (A) Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genome (KEGG) analysis of CBX genes and similar genes. **(B)** Network of enriched terms of CBX genes and similar genes; Protein-protein interaction (PPI) enrichment analysis of these genes and MCODE (Molecular Complex Detection) components identified in the gene lists. Analyses were conducted in Metascape.

negative correlation between CBX6 expression and the infiltration of $CD8^+$ T cells (Cor = -0.167, p = 2.51e-02), and a positive correlation between CBX6 expression and the infiltration of macrophages (Cor = 0.263, p = 3.62e-04). There was a positive correlation between CBX7 expression and the infiltration of B cells (Cor = 0.254, p = 5.87e-04), $CD4^+$ T cells

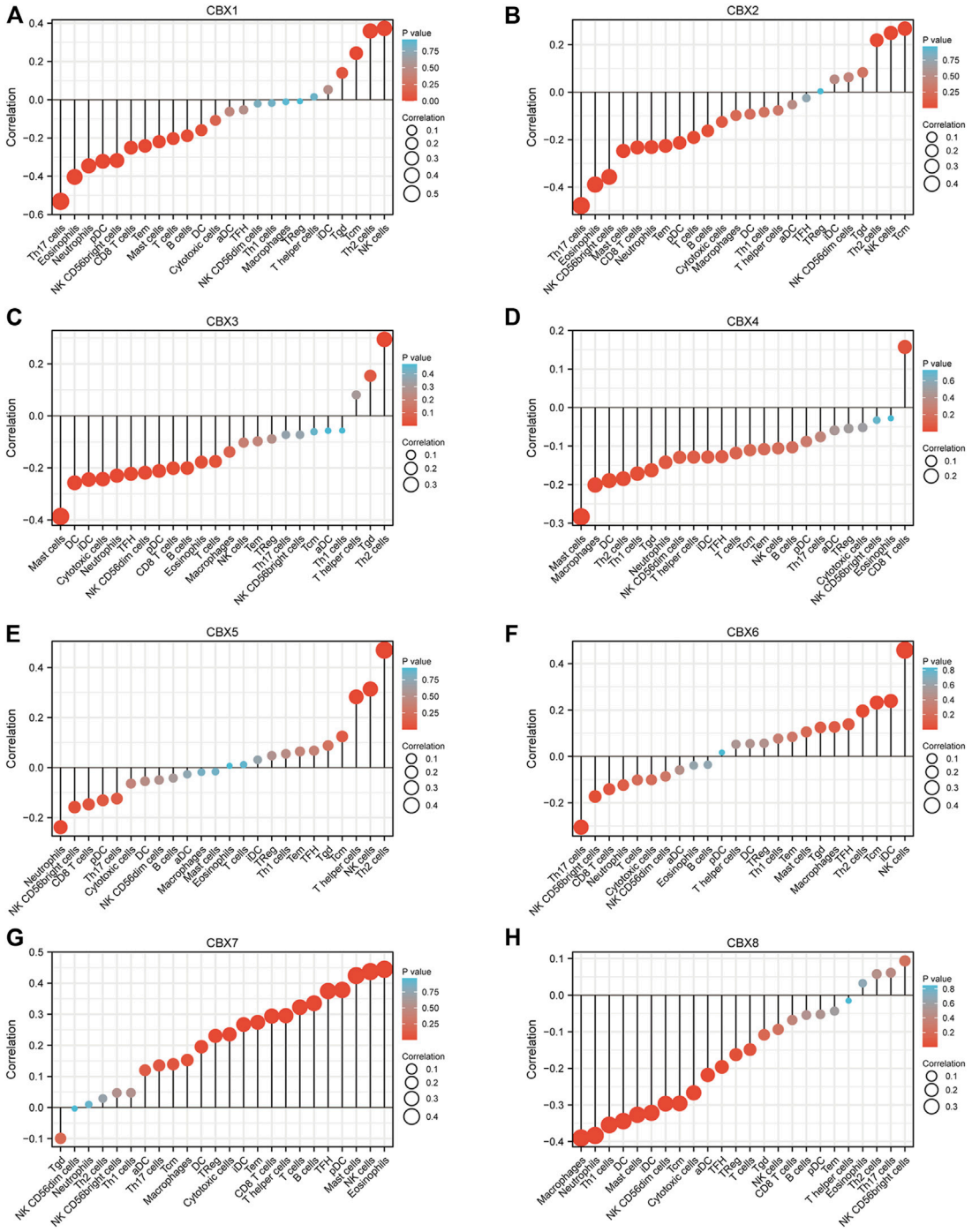


FIGURE 7 | Correlations between differentially expressed CBX family members and immune cell infiltrations (TCGA). Correlations between the abundance of immune cells and the expression of CBX1-8 **(A–H)**.

(Cor = 0.183, p = 1.43e-02) and macrophages (Cor = 0.283, p = 1.20e-04). The expression of CBX8 was positively associated with the infiltration of B cells (Cor = 0.222, p = 2.82e-03), and negatively associated with the infiltration of neutrophils (Cor = −0.164, p = 2.78e-02), and dendritic cells (Cor = −0.375, p = 2.09e-07).

TCGA ESCA project in level 3 HTSeq-FPKM format RNAseq data and clinical data showed that CBX1, CBX2, CBX3, CBX4, and CBX8 were negatively related to the infiltration of most immune cells, while CBX7 was positively associated with the infiltration of most immune cells (**Figure 7**). We also analyzed the relationship between CBX1/3/7/8 expression levels and immune

cell infiltration in ESCA. Patients in the high CBX1 expression group (n = 173, data from TCGA database) presented a decrease in the numbers of infiltrating T cells, B cells, CD8 T cells, neutrophils, and Th17 cells (**Supplementary Figure S5A**). The group with higher expression of CBX3 showed a decrease in immune cell infiltration, including T cells, B cells, cytotoxic cells, CD8 T cells, neutrophils, and Th17 cells (**Supplementary Figure S5B**). The higher CBX7 expression group presented higher infiltration of immune which can kill tumor cells, including T cells, B cells, Cytotoxic cells, CD8 T cells, NK cells, iDC cells, Treg cells, and Th17 cells (**Supplementary Figure S5C**). Patients in the high CBX1 expression group indicated a decrease in the numbers of infiltrating T cells, cytotoxic cells, neutrophils, iDC cells, macrophages cells, and Treg cells (**Supplementary Figure S5D**).

These indicated that CBX family members may interact with immune cell infiltration, and then influence the outcomes of ESCA patients.

Association of CBX Family Members With PRC2 in ESCA Patients

Aberrant epigenetic regulation has been reported to promote the pathogenesis and progression of ESCA. The methylation of lysine 27 on histone H3 (H3K27me3) is a chromatin marker associated with nucleosome condensation and silencing of gene expression (Adema and Colla, 2022). PCR1 and PCR2 played critical roles in establishing and maintaining the H3K27me3 mark. PRC2 comprised three core components (EZH2, SUZ12, and EED). EZH2 has hmtase activity, which is maintained by the presence of SUZ12 and EED (Erokhin et al., 2021). We used TCGA databases to analyze the expression levels of PRC2 components in ESCA. As shown in **Supplementary Figure S6A**, the transcriptional levels of EZH2, SUZ12, and EED were significantly elevated in ESCA tissues compared to normal tissues. We further used TCGA databases to analyze the association of CBX family members with PRC2 components in ESCA patients. The results showed that except CBX7, other CBX family members were positively correlated with PRC2 components, among which CBX1/3/8 had the highest correlation. (**Supplementary Figure S6**).

Validation of CBX Family Members in Clinical Samples

To validate the finding in the mentioned databases and further reveal which CBX members play a crucial role in the progression of ESCA, we used real-time PCR to detect all the mRNA expression of CBX family members using clinical samples. The analysis results showed that CBX1, CBX2, CBX3, CBX4, CBX5, CBX6, and CBX8 were significantly higher in ESCA tissues compared to normal esophageal tissues, which is consistent with the results of bioinformatics analysis mentioned previously (**Figures 8A–F,H**). However, CBX7 was downregulated in tumor tissues in half of collected ESCA patients, although there was no difference in all patients (**Figure 8G**).

We also investigated the protein expression of CBXs in ESCA tissues using western blotting and immunofluorescence staining. Western blotting results showed that the protein expression of CBX1, CBX3, and CBX8 were much higher in ESCA tissues than precancerous tissues (**Figure 9**). Immunofluorescence staining showed that the expression of CBX1, CBX3, and CBX8 were significantly higher in ESCA than adjacent normal tissues, which corroborated the results of western blot assay (**Figure 10**).

DISCUSSION

ESCA is one of the most common and aggressive malignancies worldwide, with dismal clinical outcomes. It is universally acknowledged that ESCA evolution is a long-term, multistep process that begins from basal cell hyperplasia, low-grade dysplasia, high-grade dysplasia, carcinoma *in situ* to invasive carcinoma, and metastasis (Wang et al., 2005; Wei et al., 2015). Over the past couple of decades, genomic and epigenomic factors have been widely illustrated, which take part in the transformation of esophageal squamous precancerous lesions into ESCA (Lin et al., 2018). In addition to cancer genetics, abnormal epigenetic regulation including abnormal DNA methylation, aberrant histone modifications, and alterations of various non-coding RNAs have also been found to take an important part in driving the pathogenesis and progression of ESCA(Cao et al., 2020). Being important components of epigenetic regulation complexes, CBX family members affect the carcinogenesis and development of various cancers, including ESCA, liver cancer, and breast cancer. Despite some members of CBXs proteins having been confirmed to be implicated in ESCA, diverse roles of eight CBX family members in ESCA remain to be fully elucidated. In this study, we used various large databases to explore the role of CBX family members in ESCA with five aspects: expression pattern, clinicopathological parameters, prognostic value, genetic mutation, and immune cell infiltration.

Recently, studies from Li et al. discovered that high CBX1 expression was connected to aggressive types of breast cancers (TNBC phenotype), and the patients were inclined to have P53 mutations and lymph node metastasis (Li X et al., 2020). Prognosis analysis showed that high CBX1 was related to worse distant metastasis-free survival in breast cancer patients (Li X et al., 2020). Noticeably higher mRNA and protein expressions of CBX1 were discovered in hepatocellular carcinoma tissues compared to normal tissues significantly linked with shorter OS (Ning et al., 2018). CBX1 overexpression promoted HCC cell growth and migration by activating the Wnt/β-Catenin signaling pathway, whereas inhibition of CBX1 or knockdown of β-Catenin markedly decreased CBX1-mediated cell proliferation (Yang et al., 2018). The present study found that the mRNA expression of CBX1 in ESCA was higher than that in normal tissues in databases (TCGA and GEO cohort), which were verified by RT-PCR, western blot, and immunofluorescence in our clinical samples. ROC curve

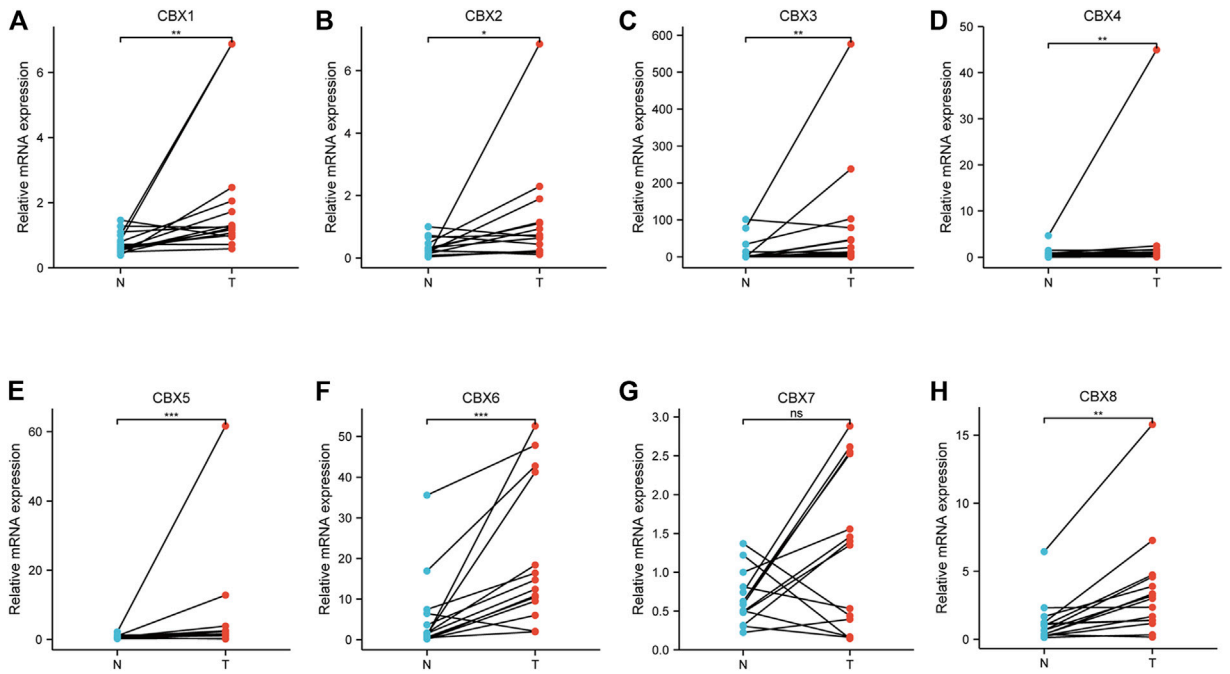


FIGURE 8 | Real-time qPCR validation of CBX family members in 17 ESCA and normal esophageal tissues **(A–H)**. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ analysis by paired t-test.

analyses showed that CBX1 had great diagnostic capability to distinguish ESCA from normal. Also, we found that overexpression of CBX1 in ESCA is significantly correlated with clinical tumor stage, tumor nodal metastasis status, tumor grade, and TP53 mutation status. Furthermore, the overexpression of CBX1 mRNA was markedly related to shorter DFS and PFI, indicating CBX1 took part in the tumorigenesis of ESCA. These results implied that CBX1 is a tumor promoter and biomarker for ESCA patients.

Growing evidence demonstrated that CBX3 deserves attention in the tumorigenesis and treatment of human malignancies. In patients with non-small cell lung cancer (NSCLC), elevated CBX3 expression is associated with poor survival (Chang et al., 2018). Dramatic upregulation of CBX3 had been found in colorectal cancer (CRC) tissues, which are related to unfavorable DFS(Li Q et al., 2020). CBX3 has been identified as a positive regulator of aerobic glycolysis and promotes growth by suppressing fructose-1,6-bisphosphatase 1 in pancreatic cancer (Chen et al., 2018). In glioma patients, CBX3 is dramatically upregulated in tumor tissues and cells, correlated with unfavorable prognosis, and it could regulate the proliferation of glioma U87 cells through CDKNIA (S. Zhao et al., 2019). Moreover, in patients with HCC and breast cancers, CBX3 overexpression promotes the proliferation of tumor cells and is associated with poor prognosis (Liang et al., 2017; Zhong et al., 2019). Herein, consistent with these previous studies, we indicated that CBX3 mRNA was upregulated in ESCA (both TCGA cohort and GEO cohort) and this expression was significantly related to tumor grades, tumor nodal metastasis status, and TP53 mutation status. ROC curve analyses indicated that CBX3 had great diagnostic capability in ESCA. In addition, overexpression of CBX3 was dramatically associated with shorter OS of ESCA patients. Therefore, we postulate that CBX3 takes part in the occurrence and progression of ESCA and might be a prognosis biomarker in ESCA.

Paradoxical roles of CBX7 had been shown in different malignant carcinomas (Pallante et al., 2015). For example, CBX7 has been recently demonstrated to be overexpressed in ovarian cancer and reduced overall survival rates compared with patients not expressing CBX7 (Shinjo et al., 2014). Furthermore, studies have shown that CBX7 could combine with E-box to inhibit tumor proliferation and migration via suppressing TWIST1 function (Li J et al., 2020). Consistently with an oncogenic role, CBX7 controls the lifespan of various human primary cells, also immortalizing mouse fibroblasts through the regulation of the Ink4a/Arf locus (Gil et al., 2004). Conversely, CBX7 plays as a tumor suppressor and is negatively correlated with cancer aggressiveness. CBX7 was downregulated in gastric cancer tissues compared to normal tissues, and this downregulation of CBX7 was closely related to poor OS (Ma et al., 2020). CBX7 reduces the emergence of breast adenocarcinoma by inhibiting the Wnt/β-catenin pathway via upregulation of the Wnt antagonist DKK-1 expression (Kim et al., 2015). In our study, CBX7 was downregulated in ESCA and had moderate diagnostic capability in ESCA. Further studies showed that CBX7 was negatively correlated with tumor stage, tumor grade, and tumor nodal metastasis status. Besides, the downregulation of CBX7 mRNA was markedly correlated with poor OS in ESCA. These findings indicated that CBX7 functions as an anti-cancer effect in ESCA.

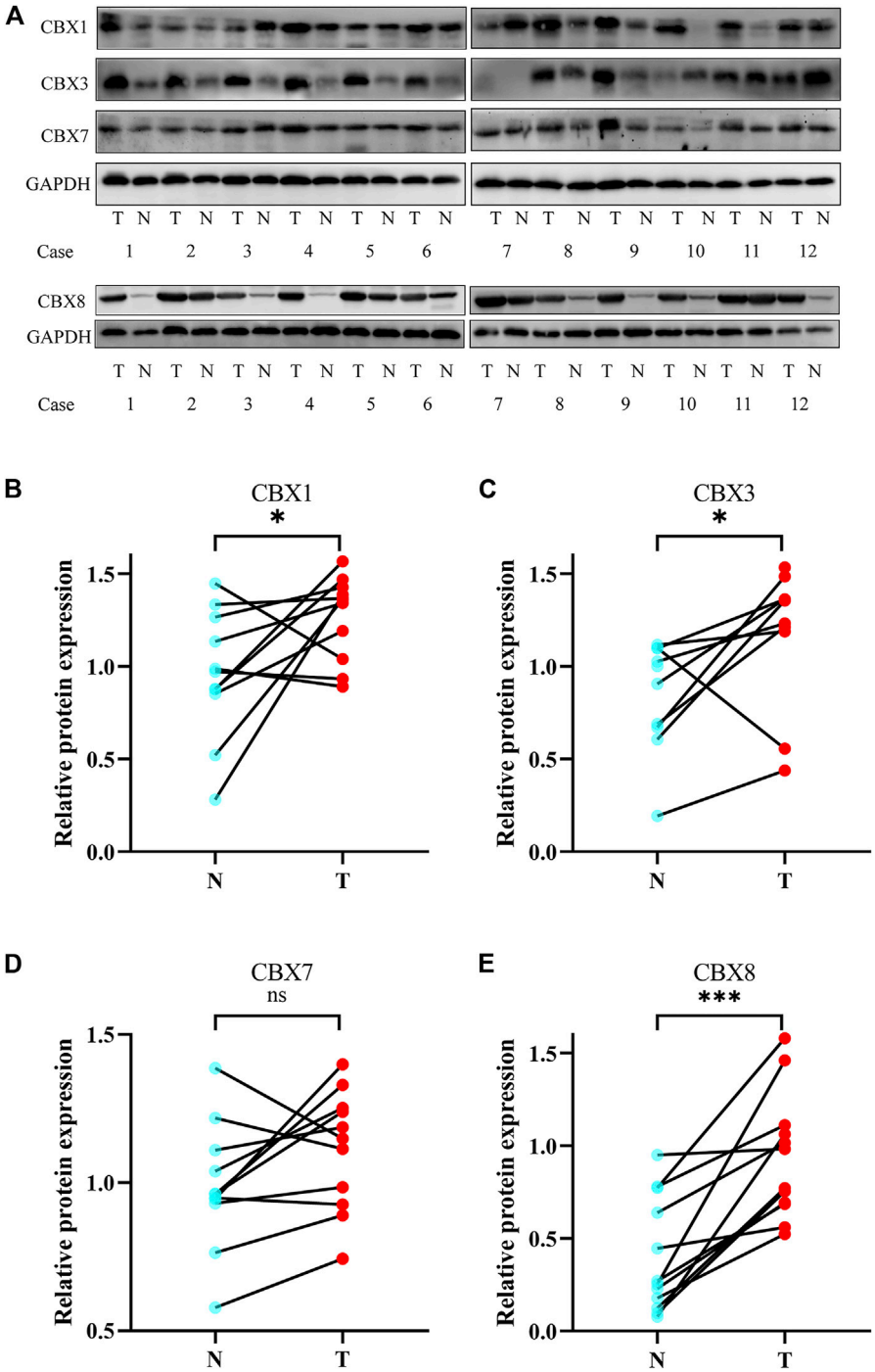


FIGURE 9 | Protein expression levels of CBX1, CBX3, CBX7, and CBX8 in 12 ESCA tissues and adjacent normal tissues (western blot). **(A)** CBX1, CBX3, and CBX8 were upregulated in ESCA tissues compared with adjacent normal tissues, while CBX7 was expressed at similar levels in ESCA and normal esophageal tissues. **(B–E)** Statistical results of western blotting of CBX1, CBX3, CBX7, and CBX8 in ESCA tissues and normal tissues. (*$p < 0.05$, analysis by paired t-test.)

Upregulation of CBX8 had been revealed in HCC tissues and indicated a worse prognosis in patients (Gao et al., 2015). *In vitro* study had shown that the high expression of CBX8 facilitated tumor proliferation and metastasis by stimulating the AKT/β-catenin pathway (Zhang et al., 2018). Likewise, remarkably higher expression of CBX8 was also observed in HCC tissues, which was significantly linked to cancer stages and tumor grades. Furthermore, high expression of CBX8 was dramatically correlated with shorter OS in patients with liver cancer (Ning et al., 2018). CBX8 performs a conflicting role in ESCA: it

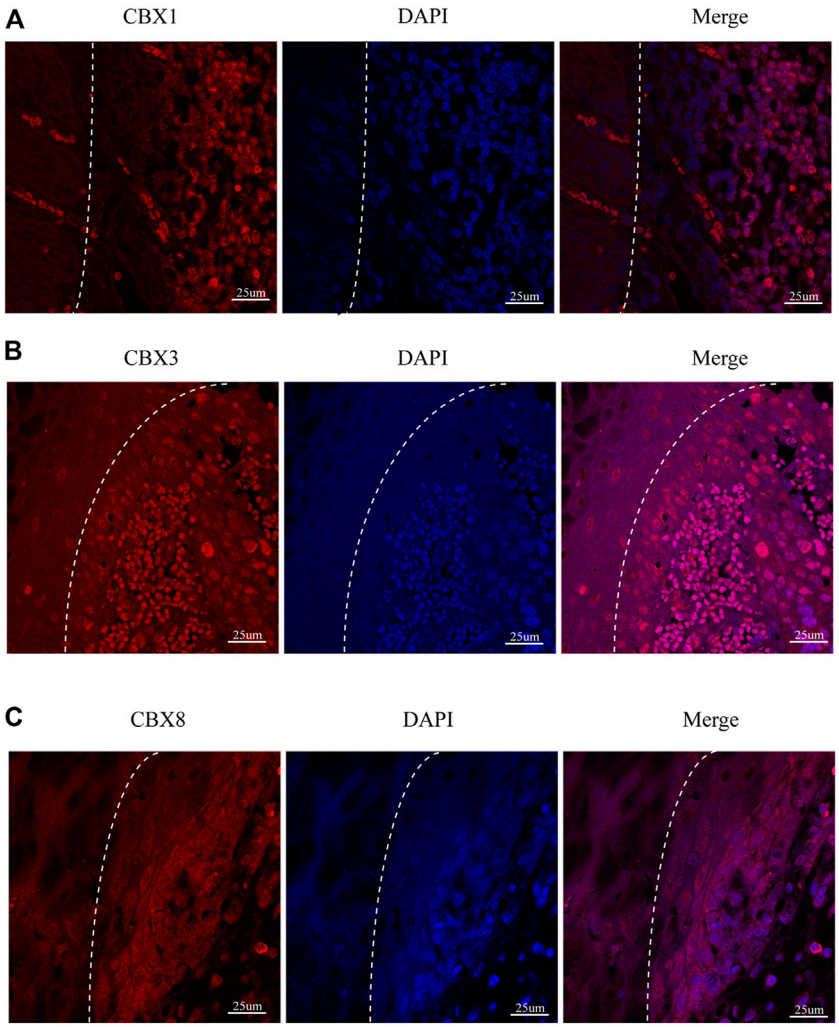


FIGURE 10 | Fluorescence images of paraffin-embedded sections of human ESCA and adjacent tissues with confocal microscopy. **(A)** CBX1 was significantly higher in ESCA tissues than that in precancerous tissues. **(B)** CBX3 was significantly higher in ESCA tissues than that in precancerous tissues. **(C)** CBX8 was significantly higher in ESCA tissues than that in precancerous tissues. The white dashed lines separate the precancerous tissues (on the left side of the figure) from the ESCA tissues (on the right side of the figure). Scale bars = 25 um. (Magnification 63×/1.40 oil).

promotes cell proliferation but inhibits cell migration, invasion, and metastasis (Wang et al., 2017). In this study, CBX8 was upregulated in ESCA, and it was significantly correlated with clinical tumor grade, tumor nodal metastasis status, and TP53 mutation status. CBX8 showed great diagnostic capability in ESCA. Furthermore, the high expression of CBX8 in Esophageal Adenocarcinoma patients was significantly related to poor OS, indicating CBX8 involved in the tumorigenesis of ESCA.

Additionally, the genetic analysis demonstrating high genetic alterations of CBXs were found in ESCA patients, and the most alteration was the high expression of mRNA. There was a mutually cooccurring connection between different CBXs, indicating that CBXs take an antagonistic or synergistic role in the tumorigenesis of ESCA. Then we found 50 genes most associated with each CBX gene by using cBioPortal. These genes were further annotated based on GO enrichment analysis and KEGG pathway enrichment analysis. The results indicated that the roles of these genes are found to be primarily associated with DNA replication and DNA repair, signaling pathways that involved Mismatch repair (MMR), and Wnt signaling pathways. DNA MMR genes play critical roles in retaining genome stability. It is widely known that TP53 is a tumor suppressor gene, implicated in the regulation of cell growth, apoptosis, cell cycle, differentiation, and senescence in ESCA. TP53 gene mutations and protein accumulations are early and frequent events in ESCA (Shimada, 2018). Patients with ESCA and TP53 gene mutations were correlated with poor overall survival compared with patients without TP53 mutations (Fisher et al., 2017). In our study, we found CBX1/2/3/8 was positively co-expressed with TP53. CBX1/2/3/8 were upregulated in ESCA patients with TP53 mutation

compared with normal tissues and TP53 non-mutation patients. CBX family members may promote ESCA development through the P53 pathway.

In recent years, the tumor microenvironment (TME) is becoming increasingly relevant in cancer research (Xu et al., 2021). Immune cells in TME may play tumor-promoting and suppressive roles, thereby influencing the clinical outcome (Yang et al., 2019). The CBX family members have been reported to contribute to the infiltration of immune cells in various cancers (Li Q et al., 2020; Zhou et al., 2021). In our study, we showed that the CBX family members' expression could be significantly related to the infiltration of immune cells in ESCA, indicating that CBXs might also affect the immune status. In particular, CBX1and CBX3 were negatively related to the infiltration of $CD8^+$T cells. $CD8^+$T cells are cytotoxic T lymphocytes which are generally considered as the main component of anti-tumor immunity (Mahmoudi et al., 2021). In various cancers, increased $CD8^+$ T cell infiltrations in the tumor mass are associated with improved patient survival (Zhao et al., 2018). In addition, in this study, we found CBX7 was positively associated with the infiltration of most immune cells, including T cells, B cells, cytotoxic cells, $CD8^+$T cells, NK cells, iDC cells, and Treg cells. These results indicated that CBX7 may play a key role in influencing the immune status of ESCA.

Polycomb group (PcG) proteins including PRC1 and PRC2, are essential epigenetic regulators that maintain transcriptional repression (Vizán et al., 2020). PRC2 consists of three core components [Enhancer of zeste homolog 2 (EZH2), suppressor of zeste 12 (SUZ12), and embryonic ectoderm development (EED)], which medicated histone methyltransferase activity (Zheng et al., 2021). Overexpression of EZH2 in cancer cells results in transcriptional repression through increased H3K27me3 activity. Enzymatic action by EZH2 at target genes requires the binding of SUZ12 and EED. We analyzed the expression levels of PRC2 components and found EZH2, SUZ12, and EED were upregulated in ESCA. Furthermore, CBX family members (especially CBX1/3/8) were positively correlated with PRC2 components. Hence, we believe that the CBX family, as members of PRC1, promote tumorigenesis by interacting with members of PRC2 through establishing and maintaining the H3K27me3 mark.

CONCLUSION

In conclusion, we systematically investigated the various expression and prognostic values of CBX family member genes in ESCA using bioinformatics analyses, and we verified these results in tissue samples. Our findings indicated that CBX1/3/8 are tumor promoters, while CBX7 serves as a tumor suppressor in ESCA. Although molecular mechanism studies are needed to validate our findings, our work provides new insights to improve the accuracy of prognosis and precision therapy for ESCA patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by comprehensive informed consent for the use of resected tissue for research which was obtained from all patients, and the study protocol was approved by the Ethics Committee of Fujian Medical University Union Hospital (No. 2020WSJK066). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XW and BZ designed the study; XF and JW collected the human tissues and conducted the experiments; JC, MZ, TH and ZC performed data mining; XF and JW wrote the manuscript; and YH helped to write the revised manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported in part by the Key Clinical Specialty Discipline Construction Program of Fujian, China (Min Wei Ke Jiao 2012 No. 149) to XW, the Educational Research Project for Young and Middle-aged Teachers in Fujian Province (JAT200125) to XF, the Startup Fund for scientific research from Fujian Medical University (2020QH1154 and No. 2020QH1071), the Young and Middle-aged Talents Training Project of Fujian Provincial Heath commission (2020GGA031) to BZ, and the General Project of Fujian Natural Science Foundation (2020J011005) to BZ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.851390/full#supplementary-material

REFERENCES

Adema, V., and Colla, S. (2022). EZH2 Inhibitors: The Unpacking Revolution. *Cancer Res.* 82 (3), 359–361. doi:10.1158/0008-5472.Can-21-4311

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Cao, W., Lee, H., Wu, W., Zaman, A., McCorkle, S., Yan, M., et al. (2020). Multi-Faceted Epigenetic Dysregulation of Gene Expression Promotes Esophageal Squamous Cell Carcinoma. *Nat. Commun.* 11 (1), 3675. doi:10.1038/s41467-020-17227-z

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Chan, H. L., and Morey, L. (2019). Emerging Roles for Polycomb-Group Proteins in Stem Cells and Cancer. *Trends Biochem. Sci.* 44 (8), 688–700. doi:10.1016/j.tibs.2019.04.005

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., et al. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19 (8), 649–658. doi:10.1016/j.neo.2017.05.002

Chang, S.-C., Lai, Y.-C., Chen, Y.-C., Wang, N.-K., Wang, W.-S., and Lai, J.-I. (2018). CBX3/Heterochromatin Protein 1 Gamma Is Significantly Upregulated in Patients with Non-Small Cell Lung Cancer. *Asia-Pac J. Clin. Oncol.* 14 (5), e283–e288. doi:10.1111/ajco.12820

Chen, L.-Y., Cheng, C.-S., Qu, C., Wang, P., Chen, H., Meng, Z.-Q., et al. (2018). CBX3 Promotes Proliferation and Regulates Glycolysis via Suppressing FBP1 in Pancreatic Cancer. *Biochem. Biophysical Res. Commun.* 500 (3), 691–697. doi:10.1016/j.bbrc.2018.04.137

Davis, S., and Meltzer, P. S. (2007). GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23 (14), 1846–1847. doi:10.1093/bioinformatics/btm254

Erokhin, M., Chetverina, O., Győrffy, B., Tatarskiy, V. V., Mogila, V., Shtil, A. A., et al. (2021). Clinical Correlations of Polycomb Repressive Complex 2 in Different Tumor Types. *Cancers* 13 (13), 3155. doi:10.3390/cancers13133155

Fisher, O. M., Lord, S. J., Falkenback, D., Clemons, N. J., Eslick, G. D., and Lord, R. V. (2017). The Prognostic Value of TP53 Mutations in Oesophageal Adenocarcinoma: A Systematic Review and Meta-Analysis. *Gut* 66 (3), 399–410. doi:10.1136/gutjnl-2015-310888

Gao, S. B., Sun, S. L., Zheng, Q. L., Zhang, L., Zhu, Y., Jin, G. H., et al. (2015). Genetic Alteration and Misexpression of Polycomb Group Genes in Hepatocellular Carcinoma. *Am. J. Cancer Res.* 5 (10), 2969–2979.

Gil, J., Bernard, D., Martínez, D., and Beach, D. (2004). Polycomb CBX7 Has a Unifying Role in Cellular Lifespan. *Nat. Cel Biol* 6 (1), 67–72. doi:10.1038/ncb1077

Gil, J., and O'Loghlen, A. (2014). PRC1 Complex Diversity: Where Is it Taking Us? *Trends Cell Biology* 24 (11), 632–641. doi:10.1016/j.tcb.2014.06.005

Iqbal, M. A., Siddiqui, S., Ur Rehman, A., Siddiqui, F. A., Singh, P., Kumar, B., et al. (2021). Multiomics Integrative Analysis Reveals Antagonistic Roles of CBX2 and CBX7 in Metabolic Reprogramming of Breast Cancer. *Mol. Oncol.* 15, 1450–1465. doi:10.1002/1878-0261.12894

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., et al. (2011). Recognition and Specificity Determinants of the Human Cbx Chromodomains. *J. Biol. Chem.* 286 (1), 521–529. doi:10.1074/jbc.M110.191411

Kim, H. Y., Park, J. H., Won, H. Y., Lee, J. Y., and Kong, G. (2015). CBX7 Inhibits Breast Tumorigenicity through DKK-1-mediated Suppression of the Wnt/β-catenin Pathway. *FASEB J.* 29 (1), 300–313. doi:10.1096/fj.14-253997

Klauke, K., Radulović, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., et al. (2013). Polycomb Cbx Family Members Mediate the Balance between Haematopoietic Stem Cell Self-Renewal and Differentiation. *Nat. Cel Biol* 15 (4), 353–362. doi:10.1038/ncb2701

Li, J., Alvero, A. B., Nuti, S., Tedja, R., Roberts, C. M., Pitruzzello, M., et al. (2020). CBX7 Binds the E-Box to Inhibit TWIST-1 Function and Inhibit Tumorigenicity and Metastatic Potential. *Oncogene* 39 (20), 3965–3979. doi:10.1038/s41388-020-1269-5

Li, Q., Pan, Y., Cao, Z., and Zhao, S. (2020). Comprehensive Analysis of Prognostic Value and Immune Infiltration of Chromobox Family Members in Colorectal Cancer. *Front. Oncol.* 10, 582667. doi:10.3389/fonc.2020.582667

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Li, X., Gou, J., Li, H., and Yang, X. (2020). Bioinformatic Analysis of the Expression and Prognostic Value of Chromobox Family Proteins in Human Breast Cancer. *Sci. Rep.* 10 (1), 17739. doi:10.1038/s41598-020-74792-5

Liang, Y.-K., Lin, H.-Y., Chen, C.-F., and Zeng, D. (2017). Prognostic Values of Distinct CBX Family Members in Breast Cancer. *Oncotarget* 8 (54), 92375–92387. doi:10.18632/oncotarget.21325

Lin, D.-C., Wang, M.-R., and Koeffler, H. P. (2018). Genomic and Epigenomic Aberrations in Esophageal Squamous Cell Carcinoma and Implications for Patients. *Gastroenterology* 154 (2), 374–389. doi:10.1053/j.gastro.2017.06.066

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173 (2), 400–416. doi:10.1016/j.cell.2018.02.052

Ma, R.-G., Zhang, Y., Sun, T.-T., and Cheng, B. (2014). Epigenetic Regulation by Polycomb Group Complexes: Focus on Roles of CBX Proteins. *J. Zhejiang Univ. Sci. B* 15 (5), 412–428. doi:10.1631/jzus.B1400077

Ma, T., Ma, N., Chen, J.-L., Tang, F.-X., Zong, Z., Yu, Z.-M., et al. (2020). Expression and Prognostic Value of Chromobox Family Members in Gastric Cancer. *J. Gastrointest. Oncol.* 11 (5), 983–998. doi:10.21037/jgo-20-223

Mahmoudi, R., Dianat-Moghadam, H., Poorebrahim, M., Siapoush, S., Poortahmasebi, V., Salahlou, R., et al. (2021). Recombinant Immunotoxins Development for HER2-Based Targeted Cancer Therapies. *Cancer Cel Int* 21 (1), 470. doi:10.1186/s12935-021-02182-6

Morey, L., and Helin, K. (2010). Polycomb Group Protein-Mediated Repression of Transcription. *Trends Biochemical Sciences* 35 (6), 323–332. doi:10.1016/j.tibs.2010.02.009

Nagy, Á., Lánczky, A., Menyhárt, O., and Győrffy, B. (2018). Validation of miRNA Prognostic Power in Hepatocellular Carcinoma Using Expression Data of Independent Datasets. *Sci. Rep.* 8 (1), 9227. doi:10.1038/s41598-018-27521-y

Ning, G., Huang, Y.-L., Zhen, L.-M., Xu, W.-X., Jiao, Q., Yang, F.-J., et al. (2018). Transcriptional Expressions of Chromobox 1/2/3/6/8 as Independent Indicators for Survivals in Hepatocellular Carcinoma Patients. *Aging* 10 (11), 3450–3473. doi:10.18632/aging.101658

Pallante, P., Forzati, F., Federico, A., Arra, C., and Fusco, A. (2015). Polycomb Protein Family Member CBX7 Plays a Critical Role in Cancer Progression. *Am. J. Cancer Res.* 5 (5), 1594–1601.

Pennathur, A., Gibson, M. K., Jobe, B. A., and Luketich, J. D. (2013). Oesophageal Carcinoma. *Lancet* 381 (9864), 400–412. doi:10.1016/S0140-6736(12)60643-6

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., et al. (2004). ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* 6 (1), 1–6. doi:10.1016/s1476-5586(04)80047-2

Sealfon, S. C., and Chu, T. T. (2011). RNA and DNA Microarrays. *Methods Mol. Biol.* 671, 3–34. doi:10.1007/978-1-59745-551-0_1

Shimada, H. (2018). p53 Molecular Approach to Diagnosis and Treatment of Esophageal Squamous Cell Carcinoma. *Ann. Gastroenterol. Surg.* 2 (4), 266–273. doi:10.1002/ags3.12179

Shinjo, K., Yamashita, Y., Yamamoto, E., Akatsuka, S., Uno, N., Kamiya, A., et al. (2014). Expression of Chromobox Homolog 7 (CBX7) Is Associated with Poor Prognosis in Ovarian Clear Cell adenocarcinomaviaTRAIL-Induced Apoptotic Pathway Regulation. *Int. J. Cancer* 135 (2), 308–318. doi:10.1002/ijc.28692

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247

Ueda, S., Kanda, M., Sato, Y., Baba, H., Nakamura, S., Sawaki, K., et al. (2020). Chromobox 2 Expression Predicts Prognosis after Curative Resection of Oesophageal Squamous Cell Carcinoma. *Cancer Genom. Proteom.* 17 (4), 391–400. doi:10.21873/cgp.20198

Vivian, J., Rao, A. A., Nothaft, F. A., Ketchum, C., Armstrong, J., Novak, A., et al. (2017). Toil Enables Reproducible, Open Source, Big Biomedical Data Analyses. *Nat. Biotechnol.* 35 (4), 314–316. doi:10.1038/nbt.3772

Vizán, P., Gutiérrez, A., Espejo, I., García-Montolio, M., Lange, M., Carretero, A., et al. (2020). The Polycomb-Associated Factor PHF19 Controls Hematopoietic

Stem Cell State and Differentiation. *Sci. Adv.* 6 (32), eabb2745. doi:10.1126/sciadv.abb2745

Wang, G.-Q., Abnet, C., Shen, Q., Lewin, K., Sun, X., Roth, M., et al. (2005). Histological Precursors of Oesophageal Squamous Cell Carcinoma: Results from a 13 Year Prospective Follow up Study in a High Risk Population. *Gut* 54 (2), 187–192. doi:10.1136/gut.2004.046631

Wang, G., Tang, J., Zhan, W., Zhang, R., Zhang, M., Liao, D., et al. (2017). CBX8 Suppresses Tumor Metastasis via Repressing Snail in Esophageal Squamous Cell Carcinoma. *Theranostics* 7 (14), 3478–3488. doi:10.7150/thno.20717

Wei, W.-Q., Chen, Z.-F., He, Y.-T., Feng, H., Hou, J., Lin, D.-M., et al. (2015). Long-Term Follow-Up of a Community Assignment, One-Time Endoscopic Screening Study of Esophageal Cancer in China. *J. Clin. Oncol.* 33 (17), 1951–1957. doi:10.1200/jco.2014.58.0423

Wotton, D., and Merrill, J. C. (2007). Pc2 and SUMOylation. *Biochem. Soc. Trans.* 35, 1401–1404. doi:10.1042/bst0351401

Xu, T., Yu, S., Zhang, J., and Wu, S. (2021). Dysregulated Tumor-Associated Macrophages in Carcinogenesis, Progression and Targeted Therapy of Gynecological and Breast Cancers. *J. Hematol. Oncol.* 14 (1), 181. doi:10.1186/s13045-021-01198-9

Yang, L., Li, A., Lei, Q., and Zhang, Y. (2019). Tumor-intrinsic Signaling Pathways: Key Roles in the Regulation of the Immunosuppressive Tumor Microenvironment. *J. Hematol. Oncol.* 12 (1), 125. doi:10.1186/s13045-019-0804-8

Yang, L., Shen, C., Pettit, C. J., Li, T., Hu, A. J., Miller, E. D., et al. (2020). Wee1 Kinase Inhibitor AZD1775 Effectively Sensitizes Esophageal Cancer to Radiotherapy. *Clin. Cancer Res.* 26 (14), 3740–3750. doi:10.1158/1078-0432.Ccr-19-3373

Yang, Y.-F., Pan, Y.-H., Tian, Q.-H., Wu, D.-C., and Su, S.-G. (2018). CBX1 Indicates Poor Outcomes and Exerts Oncogenic Activity in Hepatocellular Carcinoma. *Translational Oncol.* 11 (5), 1110–1118. doi:10.1016/j.tranon.2018.07.002

Zhang, C. Z., Chen, S.-L., Wang, C.-H., He, Y.-F., Yang, X., Xie, D., et al. (2018). CBX8 Exhibits Oncogenic Activity via AKT/β-Catenin Activation in Hepatocellular Carcinoma. *Cancer Res.* 78 (1), 51–63. doi:10.1158/0008-5472.Can-17-0700

Zhang, Y., Chen, H., Zhu, H., and Sun, X. (2019). CBX8 Promotes Tumorigenesis and Confers Radioresistance in Esophageal Squamous Cell Carcinoma Cells through Targeting APAF1. *Gene* 711, 143949. doi:10.1016/j.gene.2019.143949

Zhao, J.-J., Zhou, Z.-Q., Wang, P., Chen, C.-L., Liu, Y., Pan, Q.-Z., et al. (2018). Orchestration of Immune Checkpoints in Tumor Immune Contexture and Their Prognostic Significance in Esophageal Squamous Cell Carcinoma. *Cancer Manag. Res.* 10, 6457–6468. doi:10.2147/cmar.S181949

Zhao, S.-P., Wang, F., Yang, M., Wang, X.-Y., Jin, C.-L., Ji, Q.-K., et al. (2019). CBX3 Promotes Glioma U87 Cell Proliferation and Predicts an Unfavorable Prognosis. *J. Neurooncol.* 145 (1), 35–48. doi:10.1007/s11060-019-03286-w

Zheng, Y., Huang, G., Silva, T. C., Yang, Q., Jiang, Y.-Y., Koeffler, H. P., et al. (2021). A Pan-Cancer Analysis of CpG Island Gene Regulation Reveals Extensive Plasticity within Polycomb Target Genes. *Nat. Commun.* 12 (1), 2485. doi:10.1038/s41467-021-22720-0

Zhong, X., Kan, A., Zhang, W., Zhou, J., Zhang, H., Chen, J., et al. (2019). CBX3/HP1γ Promotes Tumor Proliferation and Predicts Poor Survival in Hepatocellular Carcinoma. *Aging* 11 (15), 5483–5497. doi:10.18632/aging.102132

Zhou, J., Chen, Z., Zou, M., Wan, R., Wu, T., Luo, Y., et al. (2021). Prognosis and Immune Infiltration of Chromobox Family Genes in Sarcoma. *Front. Oncol.* 11, 657595. doi:10.3389/fonc.2021.657595

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fang, Wang, Chen, Zhuang, Huang, Chen, Huang, Zheng and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fgene.2022.835355

Prognostic and Diagnostic Values of Semaphorin 5B and Its Correlation With Tumor-Infiltrating Immune Cells in Kidney Renal Clear-Cell Carcinoma

*Junping Ding[1], Shubin Zhao[1], Xianhua Chen[2], Changjun Luo[3], Jinjian Peng[1], Jiantan Zhu[1], Yongqi Shen[4], Zhou Luo[5] and Jianlin Chen[2]**

[1]Departments of Urology of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [2]Departments of Clinical Laboratory, Key Laboratory of Medical Molecular Diagnostics of Liuzhou, Key Laboratory for Nucleic Acid Molecular Diagnosis and Application of Guangxi Health & Wellness Commission, Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [3]Departments of Cardiology of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [4]Departments of Oncology of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China, [5]Departments of Infectious Diseases of Affiliated Liutie Central Hospital of Guangxi Medical University, Liuzhou, China*

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Yingkun Xu,
Chongqing Medical University, China
Gong Cheng,
Huazhong University of Science and Technology, China

***Correspondence:**
Jianlin Chen
cjl801900@163.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *14 December 2021*
Accepted: *11 March 2022*
Published: *11 April 2022*

Citation:
Ding J, Zhao S, Chen X, Luo C, Peng J, Zhu J, Shen Y, Luo Z and Chen J (2022) Prognostic and Diagnostic Values of Semaphorin 5B and Its Correlation With Tumor-Infiltrating Immune Cells in Kidney Renal Clear-Cell Carcinoma. Front. Genet. 13:835355. doi: 10.3389/fgene.2022.835355

Background: Semaphorin 5B (SEMA5B) has been described to be involved in the development and progression of cancer. However, the potential diagnostic and prognosis roles and its correlation with tumor-infiltrating immune cells in KIRC have not been clearly reported yet.

Methods: The mRNA level of SEMA5B was analyzed via the TCGA and GTEx database as well as the CCLE dataset and verified by GSE53757 and GSE40435 datasets. Meanwhile, the protein level of SEMA5B was analyzed by CPTAC and validated by HPA. The diagnostic value of SEMA5B was analyzed according to the TCGA database and validated by GSE53757, GSE46699, and GSE11024 + GSE46699 datasets. Then, the survival analysis was conducted using GEPIA2. R software (v3.6.3) was applied to investigate the relevance between SEMA5B and immune checkpoints and m6A RNA methylation regulator expression. The correlation between SEMA5B and MMRs and DNMT expression and tumor-infiltrating immune cells was explored via TIMER2. Co-expressed genes of SEMA5B were assessed by cBioPortal, and enrichment analysis was conducted by Metascape. The methylation analysis was conducted with MEXPRESS and MethSurv online tools. Gene set enrichment analysis (GSEA) was applied to annotate the biological function of SEMA5B.

Results: SEMA5B was significantly upregulated at both the mRNA and protein levels in KIRC. Further analysis demonstrated that the mRNA expression of SEMA5B was significantly correlated with gender, age, T stage, pathologic stage, and histologic grade. High levels of SEMA5B were found to be a favorable prognostic factor and novel diagnostic biomarker for KIRC. SEMA5B expression was shown to be significantly associated with the abundance of immune cells in KIRC. Also, SEMA5B expression was significantly correlated with the abundance of MMR genes, DNMTs, and m6A regulators in KIRC. Enrichment analysis indicated that the co-expressed genes may involve in crosslinking in the extracellular matrix (ECM). GSEA disclosed that

SYSTEMIC_LUPUS_ERYTHEMATOSUS and NABA_ECM_REGULATORS were prominently enriched in the SEMA5B low-expression phenotype. Finally, the methylation analysis demonstrated a correlation between hypermethylation of the SEMA5B gene and a poor prognosis in KIRC.

Conclusion: Increased SEMA5B expression correlated with immune cell infiltration, which can be served as a favorable prognostic factor and a novel diagnostic biomarker for KIRC.

Keywords: sema5b, KIRC, diagnostic, biomarker, prognosis

INTRODUCTION

Renal cell carcinoma (RCC) is the third most common malignancy in the urological system (Siegel et al., 2018) and ranks in the top 10 most common malignancies worldwide, accounting for nearly 3% of all malignancies (Sigel et al., 2002). Kidney renal clear cell carcinoma (KIRC) is the main histology type of RCC, accounting for 80% of tumors (Shuch et al., 2015). In the past few decades, although significant progress has been made in the treatment of KIRC, the 5-year overall survival rate (OS) for patients with advanced stages is still less than 10% (Hsieh et al., 2017). Currently, surgery is the mainstay of treatment for KIRC, but is associated with a high rate of recurrence and distant metastasis (Chin et al., 2006). Chemotherapy and radiotherapy were the main adjuvant treatment after operation in the past decade; both of them have been proven to exhibit poor prognosis (Ravaud et al., 2016; Gao et al., 2020). In recent years, immunotherapy has shown good clinical effects in the treatment of renal cancer and has become a research hotspot in this field. However, immunotherapy has a low rate of reaction owing to a lack of sensitive and specific biomarkers (Atkins, 2009). Therefore, it is necessary to find more effective and safer therapeutic targets to improve the survival outcome of ccRCC patients.

In 1993, Kolodkin originally identified and characterized that semaphorins are the family of phylogenetically conserved molecules which play some roles in providing attractive and repulsive axon guidance cues during axon growth (Schwab, 1996). A previous study found that the Semaphorin 4D can affect tumor progression by various mechanisms, including modulation of tumor angiogenesis (Basile et al., 2007). Semaphorin 5B (SEMA5B) is a member of semaphorins. The knowledge about SEMA5B is mostly acquired from the studies of mouse retinas and chicken spinal cords, where SEMA5B uses plexina1 and plexina3 signals as rejection cues (Matsuoka et al., 2011; Liu et al., 2014). In recent years, increasing evidence has indicated the pivotal role of SEMA5B involved in the pathogenesis or progression of tumors. SEMA5B has been identified as a candidate gene for changes associated with asbestos exposure in pulmonary disease studies (Mäki-Nevala et al., 2016). WU et al. also found that SEMA5B can be considered as a candidate gene for a susceptibility locus in esophageal cancer (Wu et al., 2014). In addition, SEMA5B acts as a potential prognostic marker in gastric carcinoma (Cao et al., 2021). However, it was not until 2009 that researchers first demonstrated the expression of SEMA5B in human kidney (Cuellar, 2009). A few recent studies disclosed that downregulation SEMA5B expression attenuated the RCC cell viability (Hirota et al., 2006), while increased SEMA5B expression promoted proliferation in HK2 cells (Kundu et al., 2020). Moreover, RT-qPCR confirmed that SEMA5B expression was significantly elevated in ccRCC, and SEMA5B was identified as a target gene of HIF that promoted tumor growth *in vivo* (Kundu et al., 2020). However, to our best knowledge, the potential diagnostic and prognosis roles as well as the underlying mechanism of SEMA5B in KIRC have not been clearly described yet.

In the current study, we carried out a bioinformatics analysis and primarily aimed to comprehensively explore the expression level, diagnostic values, prognostic values, and gene methylation characteristics of SEMA5B in KIRC. Next, the correlation between SEMA5B expression and the immune environment was also explored. Finally, we identified and analyzed the co-expressed genes of SEMA5B in KIRC. This study provides a comprehensive insight into the underlying significance of SEMA5B and would lead to a better understanding of the possible role of SEMA5B in tumor immunology and its clinical value in KIRC.

MATERIALS AND METHODS

Expression Analyses of *SEMA5B*

Based on the TCGA database, we employed GEPIA2 (http://gepia2.cancer-pku.cn/#analysis) to assess the gene expression profiles of SEMA5B among 33 cancer types. Therefore, we investigated the expression level of SEMA5B gene between KIRC and normal tissues in TCGA datasets alone or the combination of TCGA with GTEx databases (https://www.gtexportal.org/home/-index.html). Furthermore, RNA-sequencing and clinical follow-up (overall survival [OS] and disease-free survival [DFS]) data of 539 ccRCC and 72 adjacent normal tissues were downloaded from TCGA, and the expression level of SEMA5B in 72 paired samples was determined. The log2 [TPM (Transcripts per million) +1] or log2 [FPKM (Fragments Per Kilo base per Million) +1] transformed expression data was applied for data analysis. We also explored the expression level of the protein of SEMA5B between primary tumor and normal tissues via the CPTAC (Clinical proteomic tumor analysis consortium) dataset. The immunohistochemical data of SEMA5B protein expression in KIRC and normal tissue was obtained from the Human Protein

Atlas (HPA) (https://www.proteinatlas.org/). The microarray data of GSE53757 and GSE40435 were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and used as validation KIRC datasets, and gene expression profiles were identified using the edgeR package (Smyth, 2010) in R (v3.6.3). The cell line expression matrix of 32 tumors was obtained from the CCLE dataset (https://portals.broadinstitute.org/ccle/about). Gene expression profiles were constructed by the R v4.0.3 software package ggplot2 (v3.3.3) (Ghandi et al., 2019).

Diagnostic, Survival, and Prognosis Analysis

We used the "Survival Analysis" module of GEPIA2 to plot the OS (overall survival) and DFS (disease-free survival) survival curves in KIRC with the "Median" as the expression thresholds for splitting the high-expression and low-expression cohorts. The correlation between SEMA5B expression and OS, RFS (recurrence-free survival) in KIRC were also analyzed by Kaplan–Meier plotter (http://kmplot. com/analysis/) (Lánczky et al., 2016). The hazard ratio (HR) with 95% confidence intervals and log-rank *p*-value were also computed. The diagnostic values of SEMA5B was calculated by the pROC package (Robin et al., 2011) of R version 3.6.3, and the ROC curves were visualized by the ggplot2 package (Wickham, 2016). The prognosis values of SEMA5B and clinical features were calculated by univariate and multivariate Cox regression analysis. The R package "randomForest" was used for random forest regression.

Association Between SEMA5B Expression and Clinical Features

After 539 KIRC cases with complete clinical data obtained, the correlation between clinical features and SEMA5B expression were analyzed on R. Wilcoxon rank sum or Kruskal–Wallis rank sum tests were used for significance tests.

Analyses of Genes Co-expressed With SEMA5B in KIRC

First, we identified the 300 co-expressed genes of SEMA5B by "Co-expression" module of cBioPortal, selected the top six co-expressed genes of SEMA5B according to *p* values, and visualized the correlation analysis in the cBioPortal. Then, we conducted the survival analysis of top six co-expressed genes via the GEPIA2. In addition, we performed the enrichment analysis (Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway) of co-expressed genes of SEMA5B using the Metascape (Zhou et al., 2019).

Association of SEMA5B and Immune Cell Infiltration in KIRC

TIMER2 (https://cistrome.shiny apps. io/timer/) database was used to systematically explore the correlations between SEMA5B expression and the tumor-infiltrating immune cells (TIICs).

Correlation Analysis Between SEMA5B and MMR Genes and DNA Methyltransferase Analysis

TIMER database was used to assess the correlations between SEMA5B expression and DNA mismatch repair (MMR) genes (MLH1, MSH2, MSH6, PMS2, and EPCAM), DNA methyltransferase genes (DNMT1, DNMT2, DNMT3A, and DNMT3B). GEPIA was used to verify the gene correlation analysis in the TIMER database.

Correlation Analysis Between SEMA5B and Immune Checkpoint Gene Analysis

The expression levels of 8 immune checkpoint-related genes (SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3, and PDCD1LG2) were detected, and the R 3.6.3 programming language was used to perform the statistical analysis.

DNA Methylation Analysis

The MEXPRESS tool was used to analyze the SEMA5B methylation corresponding to TCGA–KIRC cohorts. Also, the Pearson correlation analyses between DNA methylation and SEMA5B expression were performed. The MethSurv tool was used to visualize SEMA5B methylation and the Kaplan–Meier based correlation between SEMA5B hyper/hypomethylation and OS.

Correlations Between SEMA5B Expression and m6A RNA Methylation Regulators

Differentially expressed m6A RNA methylation regulators (KIRCs vs normal tissues, high- and low-SEMA5B KIRCs) were analyzed by the Mann–Whitney *U* test method in R (version R 3.6.3). ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Gene Set Enrichment Analysis (GSEA)

GSEA (Subramanian et al., 2005) analysis was used to identify functional and biological pathways between low and high expression of SEMA5B from TCGA gene expression data. The expression level of SEMA5B was served as a phenotype label. Normalized enrichment scores (NES) were acquired by analyzing genes with permutations 1,000 times. The statistical significance of pathways is dependent on normal $p < 0.05$ and false discovery rate (FDR) q<0.05.

RESULTS

The Expression Profiles of SEMA5B Were Upregulated in KIRC

Based on the TCGA database, we explored the expression profiles of SEMA5B in pan-cancers by GEPIA2. As shown in **Figure 1A**, the mRNA expression of SEMA5B was overexpressed in tumors than normal tissues in different types of cancer. Especially, SEMA5B was found to be highly expressed in KIRC than normal tissues (**Figure 1B**, $p < 0.001$) and paired adjacent normal tissues (**Figure 1C**, $p < 0.001$). Including the normal tissue of the GTEx

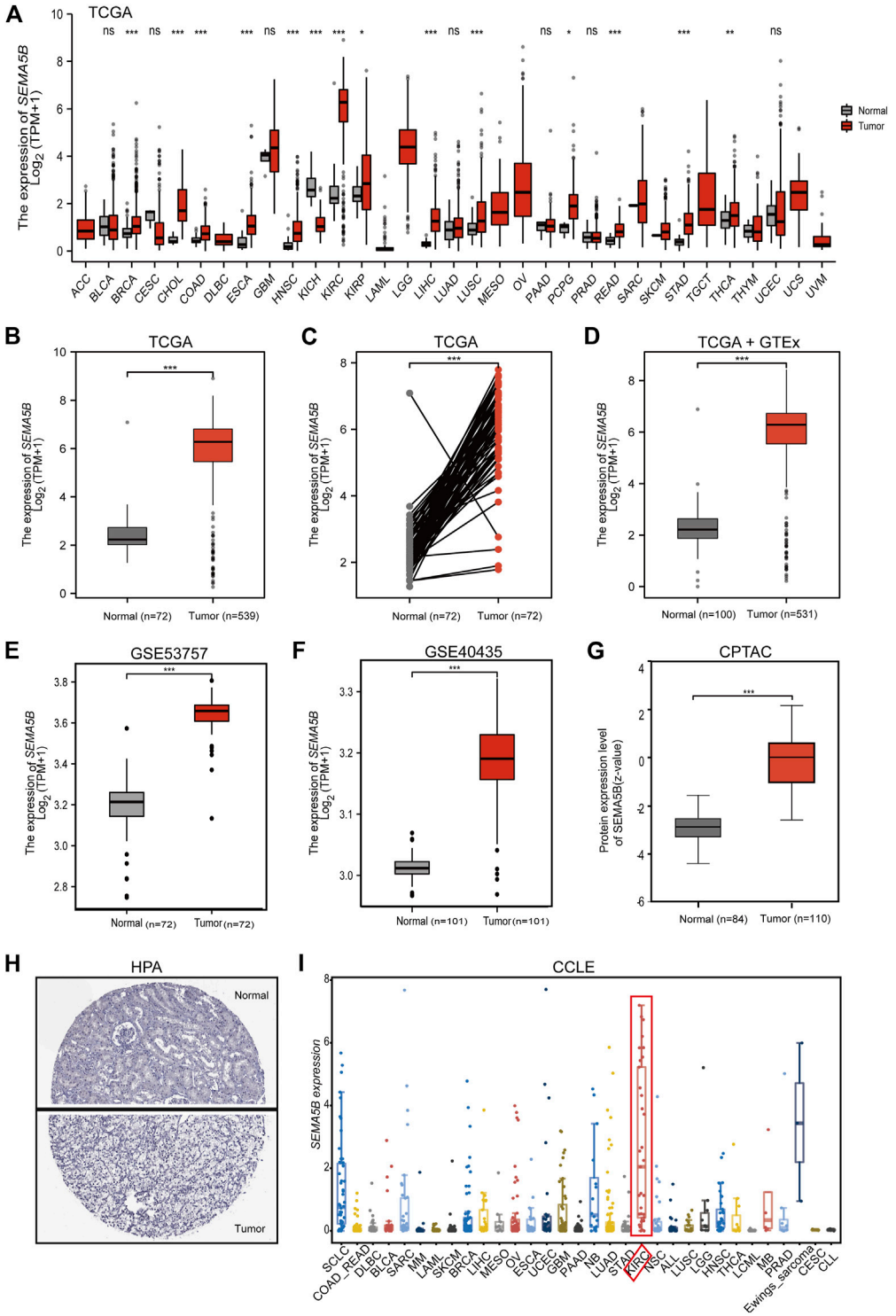


FIGURE 1 | The mRNA and protein level of SEMA5B in the KIRC tissue and KIRC cell lines. **(A)** The comparison of SEMA5B expression between tumor and noncancerous adjacent tissues in different types of cancers based on the TCGA database. **(B)** The expression of SEMA5B in normal and tumor tissues based on the TCGA. **(C)** The expression of SEMA5B in paired tissues based on the TCGA. **(D)** The expression of SEMA5B in normal and tumor tissues based on matching TCGA normal and GTEx data. The expression of SEMA5B in KIRC based on GSE53757 **(E)** and GSE40435 **(F)**. **(G)** Protein expression levels of SEMA5B in normal and KIRC tumor tissues based on the CPTAC dataset. **(H)** Representative immunohistochemistry images of SEMA5B in KIRC and normal tissues derived from the HPA database. **(I)** The mRNA expression levels of SEMA5B in kidney cancer cell lines using the CCLE. ns, $p \geq 0.05$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 1 | Relationship between the clinical features and SEMA5B expression in patients with KIRC.

Characteristic	Low expression of SEMA5B,n (%)	High expression of SEMA5B,*n* (%)	χ^2	P
N	269	270		
Gender			8.75	**0.003**
Female	76 (14.1%)	110 (20.4%)		
Male	193 (35.8%)	160 (29.7%)		
Age			5.61	**0.018**
<=60	120 (22.3%)	149 (27.6%)		
>60	149 (27.6%)	121 (22.4%)		
T stage			9.9	**0.019**
T1	132 (24.5%)	146 (27.1%)		
T2	44 (8.2%)	27 (5%)		
T3	84 (15.6%)	95 (17.6%)		
T4	9 (1.7%)	2 (0.4%)		
N stage			0.48	0.488
N0	121 (47.1%)	120 (46.7%)		
N1	10 (3.9%)	6 (2.3%)		
M stage			1.73	0.188
M0	209 (41.3%)	219 (43.3%)		
M1	45 (8.9%)	33 (6.5%)		
Pathologic stage			8.54	**0.036**
Stage I	127 (23.7%)	145 (27.1%)		
Stage II	36 (6.7%)	23 (4.3%)		
Stage III	55 (10.3%)	68 (12.7%)		
Stage IV	49 (9.1%)	33 (6.2%)		
Histologic grade			12.5	**0.006**
G1	2 (0.4%)	12 (2.3%)		
G2	112 (21.1%)	123 (23.2%)		
G3	101 (19%)	106 (20%)		
G4	47 (8.9%)	28 (5.3%)		

Note: Data in bold indicates p < *0.05.*

dataset as controls for further analysis, SEMA5B was still highly expressed in KIRC than normal tissues (**Figure 1D**, $p < 0.001$). Furthermore, the online Gene Expression Omnibus (GEO) database showed consistent results based on the data from the GSE53757 and GSE40435 datasets (**Figure 1** E, F, all $p < 0.001$). Then, we analyzed the protein expression of SEMA5B from the CPTAC databases. More importantly, the protein expression of SEMA5B was also elevated in tumor samples when compared to adjacent non-tumor samples (**Figure 1G**, $p < 0.001$). Moreover, the IHC staining results from the HPA database further confirmed the significantly higher protein levels of SEMA5B in tumor tissues (**Figure 1H**). We also performed the expression levels of SEMA5B in RCC cell lines using the Cancer Cell Line Encyclopedia (CCLE) of the Broad Institute. Our results showed that SEMA5B was significantly upregulated in RCC cell lines than other cancer types (**Figure 1I**). Collectively, all these data strongly confirmed the high-expression status of SEMA5B in KIRC.

The Expression Level of SEMA5B Was Associated With the Clinicopathological Characteristics of KIRC Patients

Since the function of SEMA5B in KIRC is still unclear, a way to further analyze the relationships between SEMA5B expression and some clinical parameters in KIRC patients is necessary. We separated the patients into high- and low-SEMA5B mRNA-expressing groups based on the median SEMA5B mRNA expression. Remarkably, SEMA5B expression was associated with gender ($p = 0.003$), age ($p = 0.018$), T stage ($p = 0.019$), pathologic stage ($p = 0.036$), and histologic grade ($p = 0.006$) (**Table 1**). However, there was no significant difference in the distribution of clinical N stages ($p = 0.488$) and M stage ($p = 0.188$). We conducted additional analyses to further examine the association of the expression of SEMA5B with clinicopathological parameters in KIRC from TCGA samples. As shown in **Figure 2**, high expression of SEMA5B was significantly correlated with age ($p < 0.05$), gender ($p < 0.01$), T stage ($p < 0.001$), M stage ($p < 0.05$), pathologic stage ($p < 0.001$), and histologic grade ($p < 0.001$). However, high expression of SEMA5B was not significantly correlated with N stage ($p > 0.05$).

Predictive Value of SEMA5B for KIRC Diagnosis and Prognosis

ROC curve is routinely used to evaluate the diagnostic value of the biomarkers in clinics. We then demonstrated the potential for using SEMA5B as a diagnostic biomarker for KIRC; ROC curves and the area under the curve (AUC) were calculated. As shown in **Figure 3B**, the ROC curve for SEMA5B exhibited an area under the curve (AUC) of 0.928 for distinguishing clear cell RCC. As the high expression of SEMA5B had a trend to be associated with pathologic stage in KIRC patients (**Figure 3A**), we hypothesized that it could be a better early diagnostic parameter for KIRC. As the AUCs were 0.926,0.871,0.964, and 0.946, SEMA5B showed significantly high sensitivity and specificity in discriminating power of disease stage for the early-stage and late-stage KIRC samples (**Figures 3C–F**). Thus, the diagnostic capability of SEMA5B was further verified in GSE53757, GSE46699, and GSE11024 + GSE46699 datasets. The AUC values for SEMA5B in discriminating KIRC patients from healthy controls with a AUC of 0.987(95% CI 0.965–1.000), 0.942(95% CI 0.902–0.983), and 0.937 (95% CI 0.897–0.977), respectively. Overall, these results illustrated the potential value of SEMA5B gene as a powerful biomarker for KIRC diagnosis.

Next, we used the K-M analyses of SEMA5B to test its ability to predict clinical outcomes. As can be seen from **Figure Figure3J**-3K, SEMA5B-high group demonstrated significantly better overall survival and disease-specific survival when compared with low group in KIRC (all $p < 0.05$). However, there were no statistical differences in progression-free interval between the high- and low-SEMA5B groups (Figure3L, $p > 0.05$).

Moreover, we performed a Cox regression univariable and multivariable analysis of characteristics potentially associated with clinical outcomes to further evaluate the predictive value of SEMA5B on overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). As shown in **Table 2**, in the univariate Cox regression analysis, pathologic stage, T stage, N stage, M stage, histologic grade, and hemoglobin were the prognostic factors for OS, DSS, and PFI. Low SEMA5B expression was a risk factor for overall survival (HR: 1.611, $p < 0.01$) and disease-specific survival (HR: 1.896, $p < 0.01$), although it did not provide any significant predictive ability for progression-

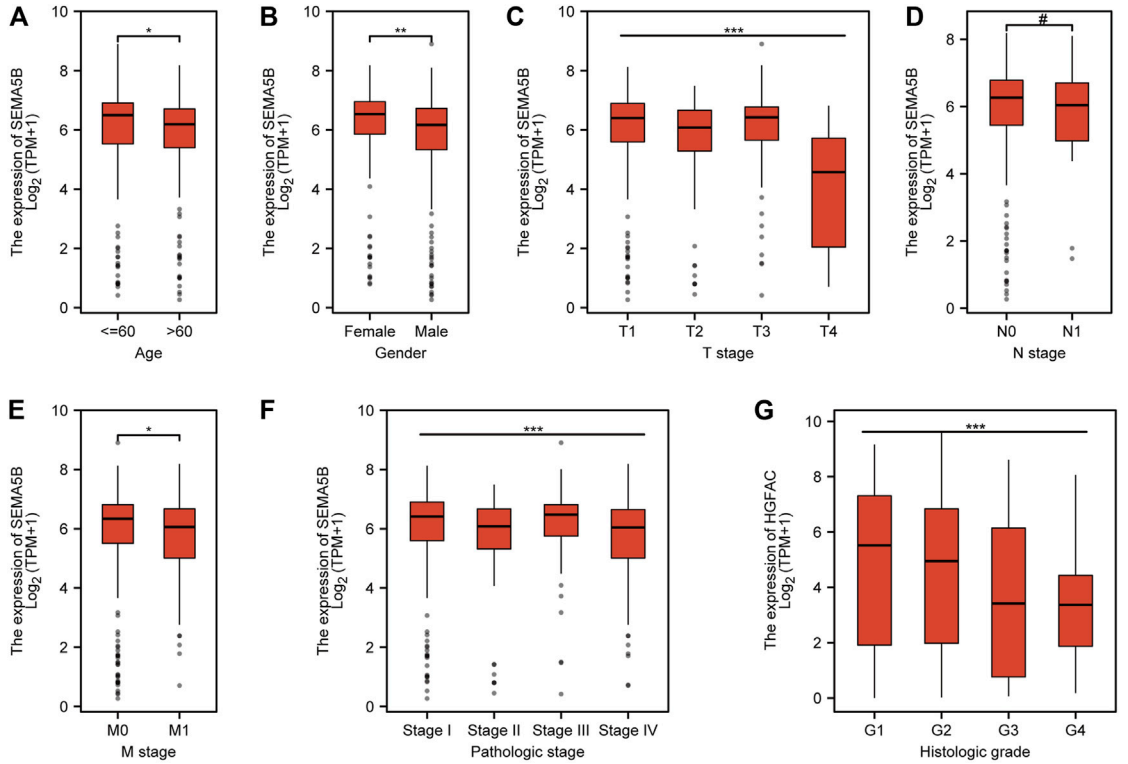


FIGURE 2 | Associations of SEMA5B expression with clinical parameters. **(A)** Age ($p < 0.05$); **(B)** Gender ($p < 0.01$); **(C)** T stage ($p < 0.001$); **(D)** N stage ($p < 0.05$); **(E)** M stage ($p < 0.05$); **(F)** Pathologic stage ($p < 0.001$); and **(G)** histologic grade ($p < 0.001$).

free interval. Notably, the result of multivariate Cox regression revealed that M stage was an independent risk factor for OS (HR: 2.767, $p < 0.001$), DSS (HR: 3.840, $p < 0.001$) and PFI (HR: 6.751, $p < 0.001$).

Analyses of Genes Co-Expressed With SEMA5B in KIRC

Genes with similar expression profiles (co-expressed genes) are often functionally related, and consequently, co-expression analysis is a robust method for gene function prediction. We applied the c-BioPortal web server to identify the top 300 co-expressed genes with SEMA5B in three different studies from TCGA (TCGA, Firehose Legacy; TCGA, Nature 2013; TCGA, PanCancer Atlas). As shown in **Figure 4A**, 156 SEMA5B positively co-expressed genes were obtained, which were duplicate genes in three TCGA studies. The values of Spearman's correlation were shown in **Figure 4B**. The top six co-expressed genes of SEMA5B arranged by adjusted p values were identified. The correlation analysis revealed that SEMA5B was highly positively correlated with protein disulfide-isomerase A5 (PDIA5), acyl-CoA dehydrogenase family member 11 (ACAD11), solute carrier family 25 member 34 (SLC25A34), selenocysteine-specific elongation factor (EEFSEC), intraflagellar transport protein 122 homolog (IFT122), and solute carrier family 23 member 3 (SLC23A3) (**Figures 4C–H**). The survival map indicated that all six genes were prognostic factors for favorable survival of KIRC, among which five genes showed statistically significant relationships with survival (**Figure 4I**). Overall survival analysis further confirmed that high expression levels of PDIA5, ACAD11, SLC25A34, EEFSEC, and SLC23A3 were associated with favorable prognosis of KIRC (**Figure 4J**-N).

To further explore the enrichment function of co-expressed genes of SEMA5B, 156 SEMA5B positively co-expressed genes were selected to conduct the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses through the Metascape. As can be seen from **Figure 5A** and **Table 3**, co-expressed genes of SEMA5B were closely enriched in small molecule catabolic process, transport of small molecules, positive regulation of inflammatory response, and diseases of signal transduction by growth factor receptors and second messengers. KEGG pathway analysis identified enrichment for several KIRC-related pathways, including nuclear receptors meta-pathway, NABA ECM AFFILIATED, and the constitutive and rostane receptor pathway. The networks of enrichment terms of SEMA5B colored by cluster ID and p value were displayed in **Figures 5B,C**. Together,

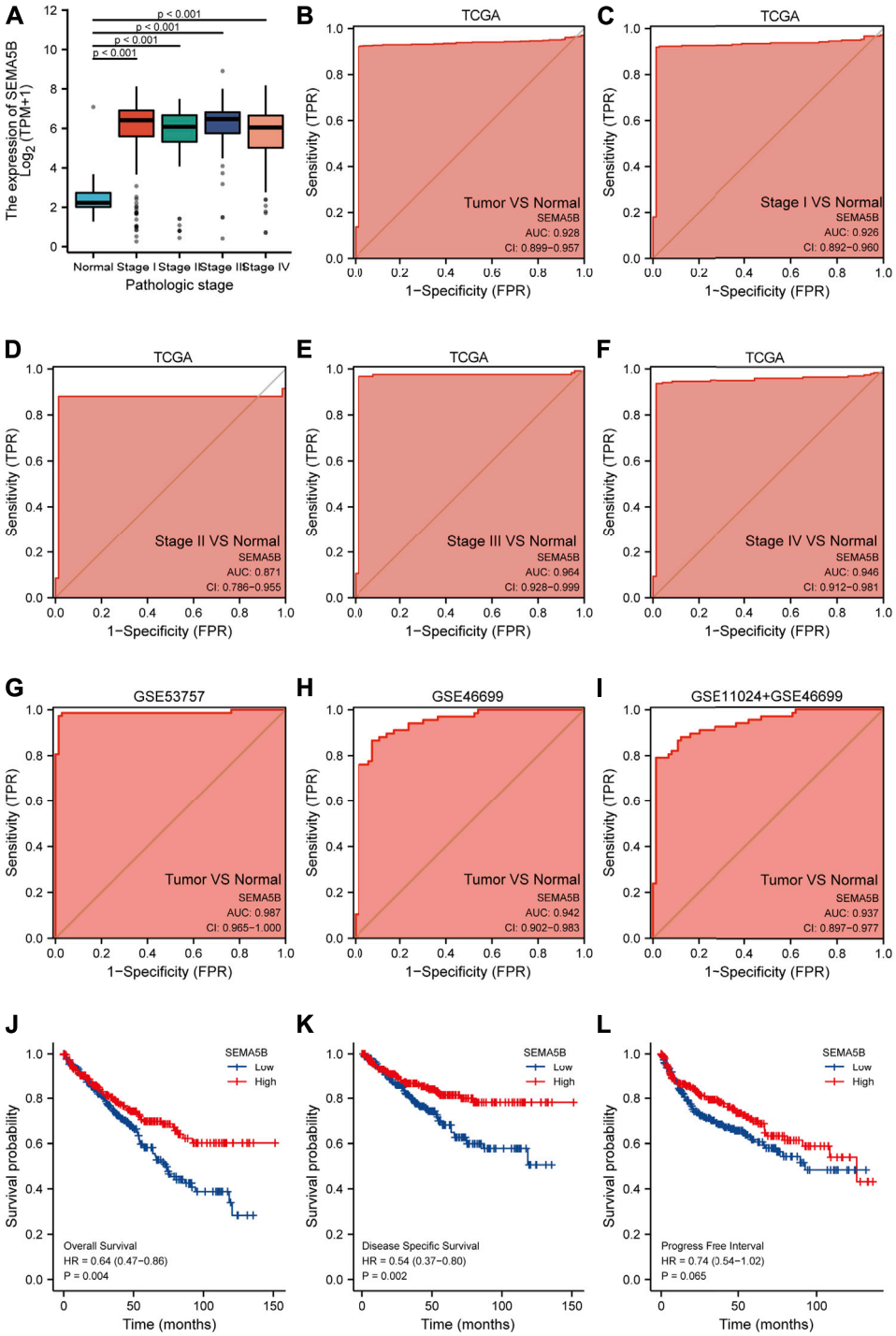


FIGURE 3 | Predictive value of SEMA5B expression for diagnosis and clinical outcomes in KIRC. **(A)** SEMA5B expression is higher in each pathologic stages of KIRC than normal tissue based on the TGCA dataset (all $p < 0.001$); ROC curves of SEMA5B for KIRC cases with overall tumors **(B)**, stage I **(C)**, stage II **(D)**, stage III **(E)**, and stage IV in TCGA datasets; ROC curve in validation set of GSE 53757 **(G)**, GSE46699 **(H)**, and GSE11024 + GSE46699 **(I)** for discriminating overall tumors from normal. Shown are the Kaplan–Meier analyses comparing overall survival **(J)**, disease-specific survival **(K)**, and progression-free interval **(L)** between high- and low-SEMA5B expression groups.

TABLE 2 | Cox regression analysis for clinical outcomes in KIRC patients.

Characteristics	HR for OS (95% CI)		HR for DSS (95% CI)		HR for PFI (95% CI)	
	Univariate	Multivariate	Univariate	Multivariate	Univariate	Multivariate
Gender (male vs. female)	0.930 (0.682–1.268)		1.220 (0.807–1.845)		**1.515*** (1.067–2.151)	1.462 (0.917–2.331)
Age (≤60 vs. >60years)	**0.567***** (0.417–0.770)	0.718 (0.458–1.128)	0.749 (0.513–1.094)		0.784 (0.574–1.071)	
Pathologic stage (stage III–IV vs. stage I–II)	**3.946***** (2.872–5.423)	1.030 (0.365–2.909)	**9.835 ***** (5.925–16.325)	2.146 (0.657–7.004)>	**6.817 ***** (4.770–9.744)	2.273 (0.848–6.093)
T stage (T3–T4 vs. T1–T2)	**3.228 ***** (2.382–4.374)	1.882 (0.754–4.698)	**5.542 ***** (3.652–8.411)	1.798 (0.709–4.557)	**4.522 ***** (3.271–6.253)	1.683 (0.738–3.840)
N stage (N1 vs. N0)	**3.453***** (1.832–6.508)	1.017 (0.439–2.356)	**3.852***** (1.825–8.132)	0.998 (0.422–2.361)	**3.682 ***** (1.891–7.167)	0.755 (0.330–1.723)
M stage (M1 vs. M0)	**4.389***** (3.212–5.999)	**2.767 ***** (1.551–4.936)	**9.108 ***** (6.209–13.361)	**3.840 ***** (2.034–7.249)	**8.968 ***** (6.464–12.442)	6.751 ***** (3.600–12.659)
Histologic grade (G3–G4 vs. G1–G2)	**2.702***** (1.918–3.807)	1.413 (0.831–2.402)	**4.793 ***** (2.889–7.952)	1.475 (0.715–3.043)	**3.646 ***** (2.503–5.310)	1.262 (0.719–2.214)
Hemoglobin (normal vs. abnormal)	**0.417 ***** (0.292–0.595)	0.582 * (0.355–0.955)	**0.399***** (0.254–0.627)	0.718 (0.399–1.291)	**0.572 ***** (0.403–0.811)	0.955 (0.582–1.567)
SEMA5B (low vs. high)	**1.611**** (1.187–2.186)	1.313 (0.829–2.079)	**1.896 **** (1.277–2.816)	1.600 (0.912–2.810)	1.344 (0.982–1.840)	1.341 (0.829–2.171)

Note: Data in bold indicates p < *0.05. HR, hazard ratio; OS, overall survival; DSS, disease-specific survival; PFI, progression-free interval; CI, confidence interval.* *p < *0.05;* **p < *0.01;* ***p < *0.001.*

these data strongly suggest that SEMA5B may play a crucial role in KIRC through crosslinking in the extracellular matrix (ECM).

Correlation Analysis of SEMA5B Expression and Immune Infiltration of Immune Cells in KIRC

High levels of tumor-infiltrating lymphocytes (TIL) have been reported to correlate with favorable prognoses in a variety of solid organ malignancies. The relationship between tumor-infiltrating lymphocytes (TIL) and SEMA5B expression was examined. As shown in **Figure 6**, SEMA5B expression was significantly positive correlated with the level of $CD4^+$ T cells (Rho = 0.209, p = 6.30e-06), B cells (Rho = 0.127, p = 6.45e-03), NK cells (Rho = 0.178, p = 1.21e-04), myeloid dendritic cells (Rho = 0.105, p = 2.38e-02), neutrophils (Rho = 0.347, p = 1.60e-14), and monocytes (Rho = 0.122, p = 8.57e-03). However, there was no significant correlation between the expression level of SEMA5B with tumor purity (Rho = 0.041, p = 3.77e-01) and the level of $CD8^+$ T cells (Rho = 0.025, p = 5.95e-01). Together, these results demonstrate that SEMA5B could promote the recruitment of immune cells in the tumor microenvironment in KIRC.

Immune Checkpoint Gene Analysis

Different cancer types can manipulate immune checkpoint expression to evade immunosurveillance, leading to the development of immune checkpoint blockade as a successful therapeutic strategy for certain cancers. We detected the expression levels of 8 immune checkpoint-related genes (SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3, and PDCD1LG2). As shown in **Figure7A**, there were significantly higher expression of SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3, and PDCD1LG2 in the KIRC tissues compared to the normal tissues (all p < 0.05). SIGLEC15 was found to be higher in the low-SEMA5B expression group than in the high-expression group (**Figure7B**). Correlation analysis also exhibited a negative correlation between SEMA5B expression and that of SIGLEC15 (**Figures 7C,D**). These results suggested that patients with high SEMA5B expression may have better immunotherapy effects and reflect better survival.

SEMA5B Is Correlated With MMR Gene Levels and DNA Methyltransferase Gene Expression in Humans

Previous studies have shown that DNA mismatch repair (MMR) deficiency leads to DNA replication errors (Georgakopoulos-Soares et al., 2020), higher somatic mutations, and tumorigenesis. In order to determine the potential role of SEMA5B in the prognosis of KIRC, we evaluated the association of the expression level of SEMA5B with the levels of five MMR genes. Results revealed that SEMA5B expression exhibited a positive relationship to MMR genes in KIRC (**Table 4**). Recently, DNA methylation has been recognized as an epigenetic modification that can affect gene expression (Tiffen et al., 2020). Also, the status change of the DNA methylation is considered to be an important factor in tumorigenesis. The correlations between SEMA5B and four DNA methyltransferases were further evaluated. Evidently, SEMA5B expression is closely related to the expression of DNMT1, DNMT2, DNMT3A, and DNMT3B in KIRC.

SEMA5B DNA Methylation Analysis

To further address whether SEMA5B expression might be influenced by DNA methylation states in KIRC, we utilized MethSurv tool to visualize the correlation between gene expression and methylation sites. We identified 33 significantly differentially methylated CpG

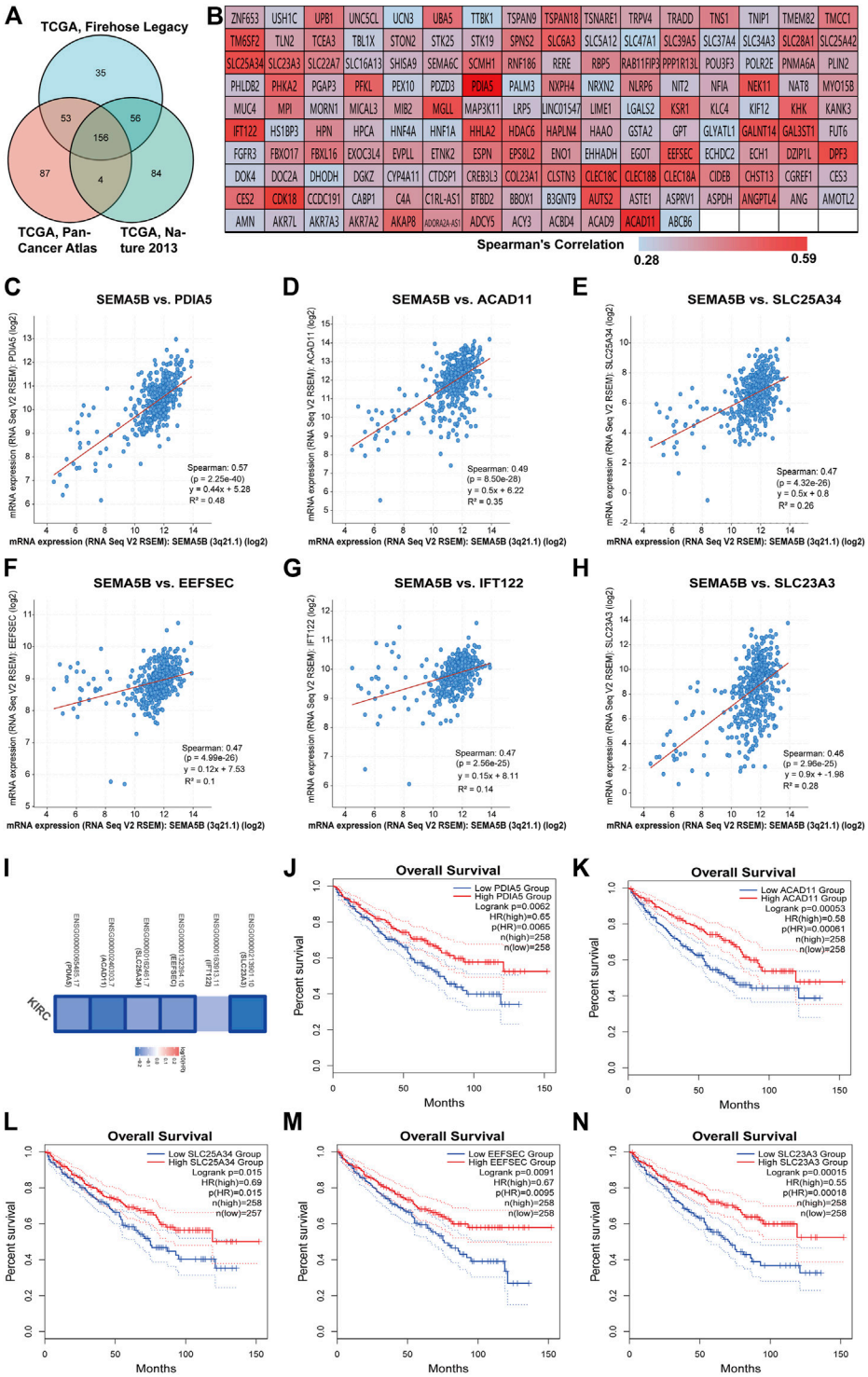


FIGURE 4 | Analyses of genes co-expressed with SEMA5B. **(A)** The Venn diagram of the 156 co-expressed genes in three TCGA studies on KIRC. **(B)** The heat map of the correlation between SEMA5B and its 156 co-expressed genes according to the values of Spearman's rank correlation coefficient. **(C–H)** Scatterplot showing the correlation between SEMA5B and its top six co-expressed genes. **(I)** Survival map of the six co-expressed genes based on the TCGA–KIRC project data was generated with the GEPIA2 online tool. Blue boxes indicate statistically significant genes. **(J–N)** The K-M (Kaplan–Meier) plots of overall survival (OS) shows the difference between the low- and high-expression of **(J)** PDIA5, **(K)** ACAD11, **(L)** SLC25A34, **(M)** EEFSEC, and **(N)** SLC23A3 in KIRC.

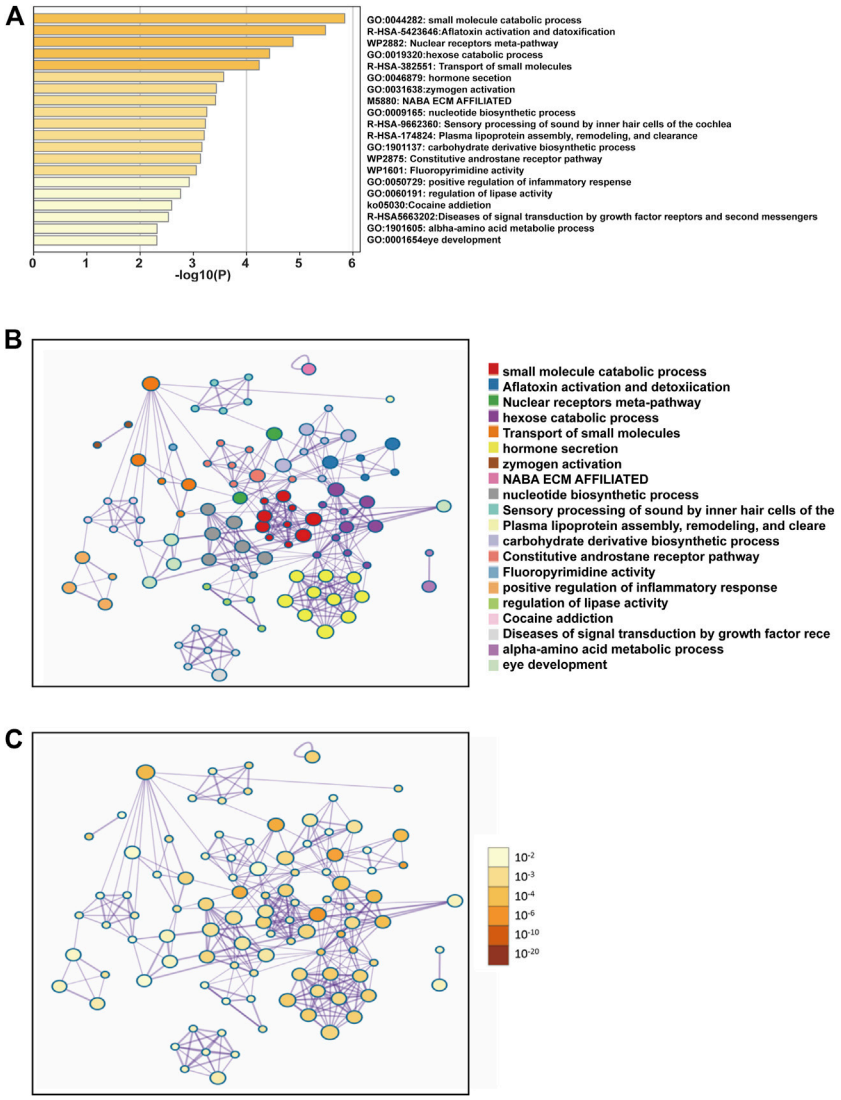


FIGURE 5 | Functional enrichment analysis of the co-expressed genes of SEMA5B. **(A)** Bar graph of enriched terms across co-expressed genes constructed by Metascape, colored by *p*-values. **(B)** Interactive network of the top 20 enriched terms colored by cluster ID. **(C)** Interactive network of the top 20 enriched terms colored by *p*-values.

sites. Notably, methylation in these sites correlated inversely with gene expression: cg06656414, cg21390574, cg09241381, cg22549268, cg20288129, cg26105015, cg11483789, cg02586610, cg21395519, cg17633431, cg15475502, cg04830808, cg11160908, cg00108098, and cg02934082 (**Figure 8A**). Survival analysis revealed that hypermethylation at cg04830808, cg06656414, cg13482010, cg09555073, cg21390574, cg22549268, cg15475502, cg00108098, cg11160908, cg02934082, and cg17633431 in the SEMA5B promoter was correlated with a poor prognosis (**Figures 8B–L**). In addition, as shown in **Figure 8M**, methylation levels in the SEMA5B-promoter region were significantly higher in normal subjects than in KIRC patients. These results were consistent with our previous survival analysis (**Figures 3J, K**), suggesting that the SEMA5B gene methylation in KIRC correlates with the progression of this type of cancer.

Expression of m6A RNA Methylation Regulators in KIRC

Numerous recent studies have shown that m6A modification plays a key role in various types of carcinogenesis (Panneerdoss et al., 2018; Zheng et al., 2021). In current study, the following 19 common m6A regulators were studied: METTL3, METTL14, WTAP, RBM15, RBM15B, ZC3H13, YTHDC1, YTHDC2, YTHDF3, YTHDF1, YTHDF2, HNRNPC, IGF2BP1, IGF2BP2, IGF2BP3, RBMX,

TABLE 3 | Top 20 clusters with their representative enriched terms (one per cluster).

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0044282	GO biological processes	Small-molecule catabolic process	12	7.69	−5.85	−1.48
R-HSA-5423646	Reactome gene sets	Aflatoxin activation and detoxification	4	2.56	−5.48	−1.47
WP2882	WikiPathways	Nuclear receptor meta-pathway	10	6.41	−4.87	−1.11
GO:0019320	GO biological processes	Hexose catabolic process	4	2.56	−4.43	−0.85
R-HSA-382551	Reactome gene sets	Transport of small molecules	14	8.97	−4.24	−0.78
GO:0046879	GO biological processes	Hormone secretion	8	5.13	−3.57	−0.38
GO:0031638	GO biological processes	Zymogen activation	4	2.56	−3.43	−0.38
M5880	Canonical pathways	NABA ECM AFFILIATED	6	3.85	−3.42	−0.38
GO:0009165	GO biological processes	Nucleotide biosynthetic process	7	4.49	−3.25	−0.35
R-HSA-9662360	Reactome gene sets	Sensory processing of sound by inner hair cells of the cochlea	4	2.56	−3.23	−0.35
R-HSA-174824	Reactome gene sets	Plasma lipoprotein assembly, remodeling, and clearance	4	2.56	−3.20	−0.35
GO:1901137	GO biological processes	Carbohydrate derivative biosynthetic process	11	7.05	−3.16	−0.34
WP2875	WikiPathways	Constitutive androstane receptor pathway	3	1.92	−3.13	−0.34
WP1601	WikiPathways	Fluor pyrimidine activity	3	1.92	−3.06	-0.34
GO:0050729	GO biological processes	Positive regulation of inflammatory response	5	3.21	-2.92	−0.25
GO:0060191	GO biological processes	Regulation of lipase activity	4	2.56	−2.76	−0.20
ko05030	KEGG pathway	Cocaine addiction	3	1.92	−2.59	−0.08
R-HSA-5663202	Reactome gene sets	Diseases of signal transduction by growth factor receptors and second messengers	8	5.13	−2.53	−0.05
GO:1901605	GO biological processes	Alpha-amino acid metabolic process	5	3.21	−2.32	0.00
GO:0001654	GO biological processes	Eye development	7	4.49	−2.32	0.00

"Count" is the number of genes in the provided lists with membership in the given ontology term. "%" is the percentage of all of the provided genes that are found in the given ontology term (only input genes with at least one ontology term annotation are included in the calculation). "Log10(P)" is the p *value in log base 10. "Log10(q)" is the multi-test adjusted* p *value in log base 10.*

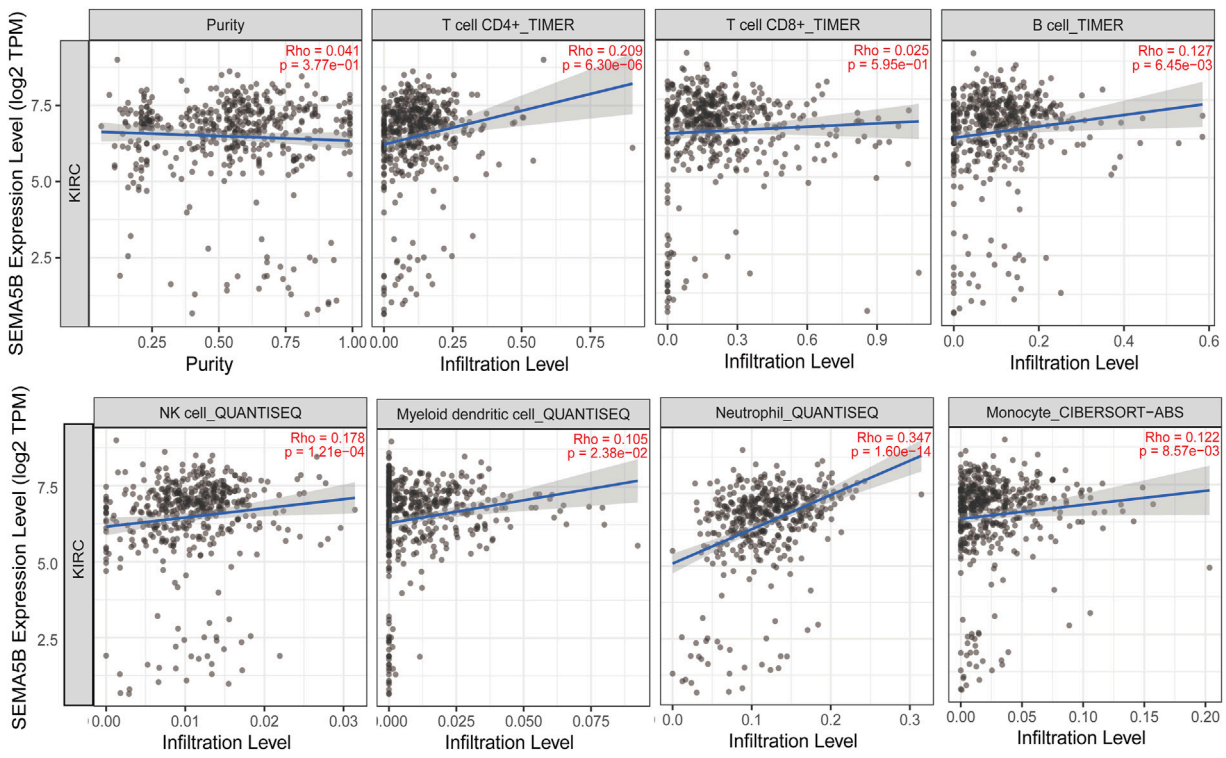


FIGURE 6 | Correlation of SEMA5B expression with immune infiltration level in KIRC. The study was performed. The correlation between the SEMA5B expression and immune cells was conducted in the TIMER database and evaluated by Spearman's correlation.

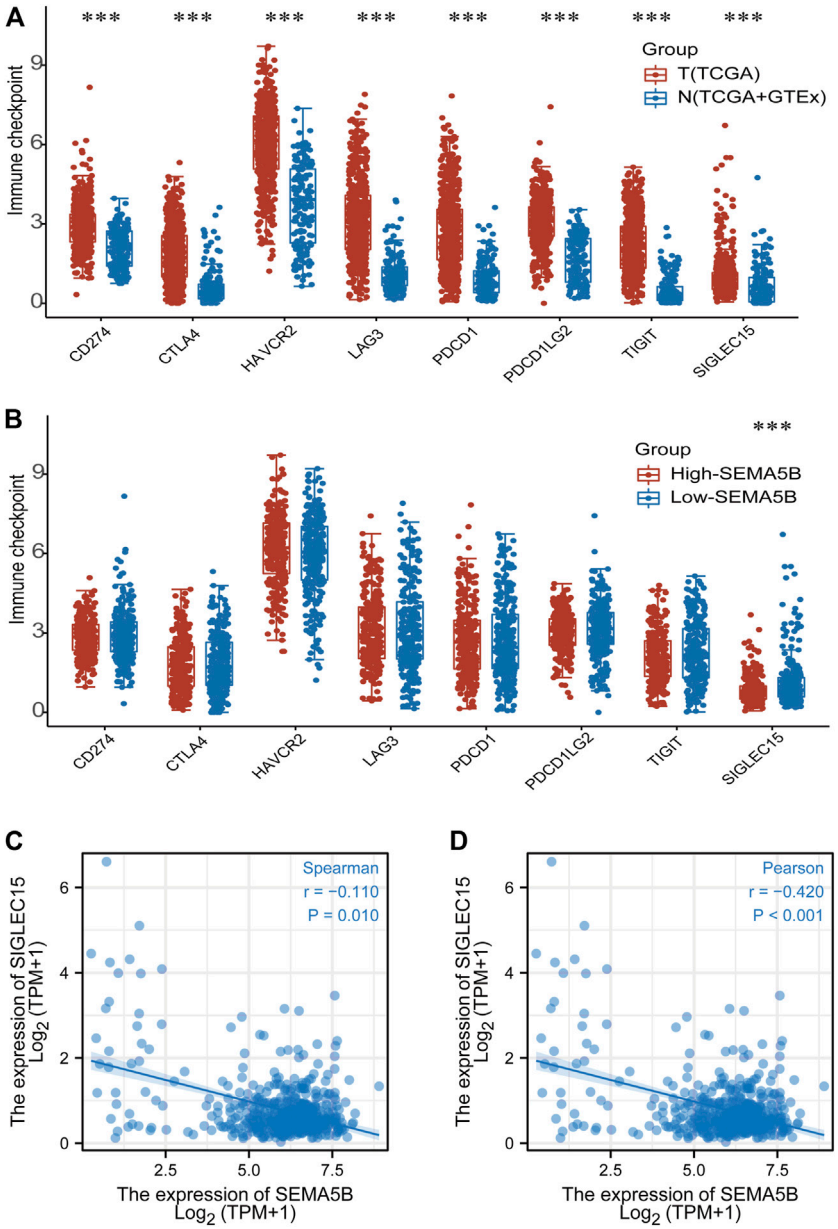


FIGURE 7 | Immune checkpoint gene analysis. **(A)** The expression distribution of checkpoint genes in tumor tissues and normal tissues. **(B)** Expression of checkpoint genes in high- and low-SEMA5B expression groups. Spearman correlation **(C)** and Pearson correlation **(D)** for the relationship between the expression levels of SEMA5B and SIGLEC15. ***$p < 0.001$.

HNRNPA2B1, FTO, and ALKBH5. Based on the TCGA data cohort and the normal tissue of the GTEx dataset, we found that all the 19 m6A regulators were differentially expressed between KIRC tissues and normal tissues, including three downregulated genes (METTL3, IGF2BP2, and HNRNPA2B1) and 16 upregulated genes (METTL14, WTAP, RBM15, RBM15B, ZC3H13, YTHDC1, YTHDC2, YTHDF3, YTHDF1, YTHDF2, HNRNPC, IGF2BP1, IGF2BP3, RBMX, FTO, and ALKBH5) (**Figure 9A**). In order to further study the expression correlation between SEMA5B and m6A regulators, we conducted a comparative analysis on the expression of m6A regulators with high and low SEMA5B expression of KIRC. Results showed that IGF2BP2, IGF2BP3 expression were significantly decreased in high-SEMA5B group than the low-SEMA5B expression group (**Figures 9B–D**). These results suggest that m6A regulators may involve in the expression of SEMA5B in KIRC.

GSEA Analysis of SEMA5B Expression

To find enriched pathways related to SEMA5B and to identify potential mechanism of SEMA5B's biological role in KIRC, we then performed the GSEA analysis (S1). Results of GSEA showed

TABLE 4 | Correlation analysis between SEMA5B and MMRs and DNA methyltransferase genes in TIMER and GEPIA.

Description	Gene markers	TIMER2		GEPIA	
		R	*p* value	R	*p* value
DNA MMRs	EPCAM	0.22	**1.5E-06**	0.21	**1.8e–06**
	MLH1	–0.10	**3.4E-02**	0.022	0.61
	MSH2	0.17	**3.5E-04**	0.19	**1.9e–05**
	MSH6	0.17	**1.7E-04**	0.19	**1.7e–05**
	PMS2	0.16	**4.1E-04**	0.12	**8.3e–04**
DNA methyltransferase genes	DNMT1	0.13	**3.8E-03**	0.17	**1.1e–05**
	DNMT2	0.009	0.846	0.19	**1.5e–05**
	DNMT3A	0.20	**1.3E-05**	0.22	**2.3e–07**
	DNMT3B	0.12	**8.4E-03**	0.16	**3.1e–05**

Note: Data in bold indicates p < *0.05.*

SYSTEMIC_LUPUS_ERYTHEMATOSUS and NABA_ECM_REGULATORS were prominently enriched in the low-SEMA5B expression phenotype.

DISCUSSION

Although early study has reported SEMA5B upregulation in clear cell renal carcinoma (ccRCC) and suggested the role of SEMA5B in tumor growth (Kundu et al., 2020), there is a lack of comprehensive studies on the clinical role of SEMA5B. Therefore, we used bioinformatics analysis of multiple online databases to investigate and obtain more comprehensive insights into the possible roles of SEMA5B in KIRC.

In this study, we observed prominent upregulated mRNA and protein expression of SEMA5B in KIRC and clear cell renal carcinoma cell lines. Decreased SEMA5B expression was correlated with older age, male, higher T stage and N stage, advanced histological grade and pathological stage, and shorter OS and DSS. We also revealed and validated a marked diagnostic capability of SEMA5B for KIRC. The co-expression analysis further indicated that SEMA5B and its co-expressed genes may act as good prognosis markers for KIRC. Furthermore, SEMA5B and its co-expressed genes were specifically enriched in small molecule catabolic process, transport of small molecules, positive regulation of inflammatory response, and diseases of signal transduction by growth factor receptors and second messengers. They may also involve in crosslinking in the extracellular matrix (ECM). Finally, we found that SEMA5B expression was correlated with immune checkpoint, DNA repair and methylation, immune cell infiltration, and various m6A RNA methylation regulators. Thus, our study elucidates the clinical value of SEMA5B in KIRC, complements the results of previous studies (Hirota et al., 2006; Kundu et al., 2020), and provides a different perspective on how SEMA5B affects cancer progression and metastasis as well as provides insights for improving cancer immunotherapy in the future.

Although the effects of SEMA5B were reported more than 30 years ago, the links between SEMA5B and human cancer has remained elusive. In this study, we conducted a comprehensive differential expression analysis of SEMA5B in normal and tumor from multiomics data integration and analysis in specific cancer. Results show that the mRNA expression level of SEMA5B is upregulated in a variety of tumors. Second, we found and confirmed that SEMA5B exhibited a significantly higher mRNA and protein expression profiles in KIRC tumor tissues and cell lines through multiple databases. SEMA5B was reported to upregulate in KIRC and effective downregulation of its expression levels in RCC cells significantly attenuated RCC cell viability (Hirota et al., 2006). Upregulated SEMA5B also significantly promoted proliferation in HK2 cells (Kundu et al., 2020). In summary, it confirms that SEMA5B expression was significantly higher in tumors rather than normal tissues in KIRC. A high expression level of KIRC may involve in the development of KIRC. A recent study has shown that the upregulated SEMA5B was closely related to the prognosis of gastric adenocarcinoma (Cao et al., 2021). In our study, we also found that the expression of SEMA5B was significantly correlated with the OS and DSS of KIRC. Previous research suggested that the upregulated SEMA5B may be a candidate marker for the diagnosis of KIRC (Hirota et al., 2006). In our study, based on the mRNA expression level in TCGA, the ROC curve for SEMA5B discrimination of KIRC diagnosis had an AUC of 0.928. Moreover, we demonstrated that SEMA5B showed strong diagnostic ability at different stages of tumor diagnosis. In addition, we validated our results on the inference of four different datasets in GEO, and found consistent and significant diagnostic efficiency. All of these strongly suggest that SEMA5B is a convincing biomarker for KIRC diagnosis.

Genes which are co-expressed have a higher probability of having related functions than those which are not co-expressed (Farahbod and Pavlidis, 2020). We conducted the co-expression analysis and explored the potential functions and mechanisms involving SEMA5B in KIRC. Our findings demonstrate that SEMA5B was highly positively correlated with the top six co-expressed genes (PDIA5, ACAD11, SLC25A34, EEFSEC, IFT122, and SLC23A3), and five of these genes served as a protective factor for the favorable survival of KIRC. However, some of these genes have been reported to act as poor prognosis in malignancy (Kho et al., 2021; Zhang et al., 2021). To further explore the potential oncogenic mechanism of SEMA5B action, we conducted the GO and KEGG analyses of the co-expressed genes. The enrichment analysis indicated that SEMA5B may be involved in several KIRC-related pathways, including nuclear receptors meta-pathway, NABA ECM AFFILIATED,

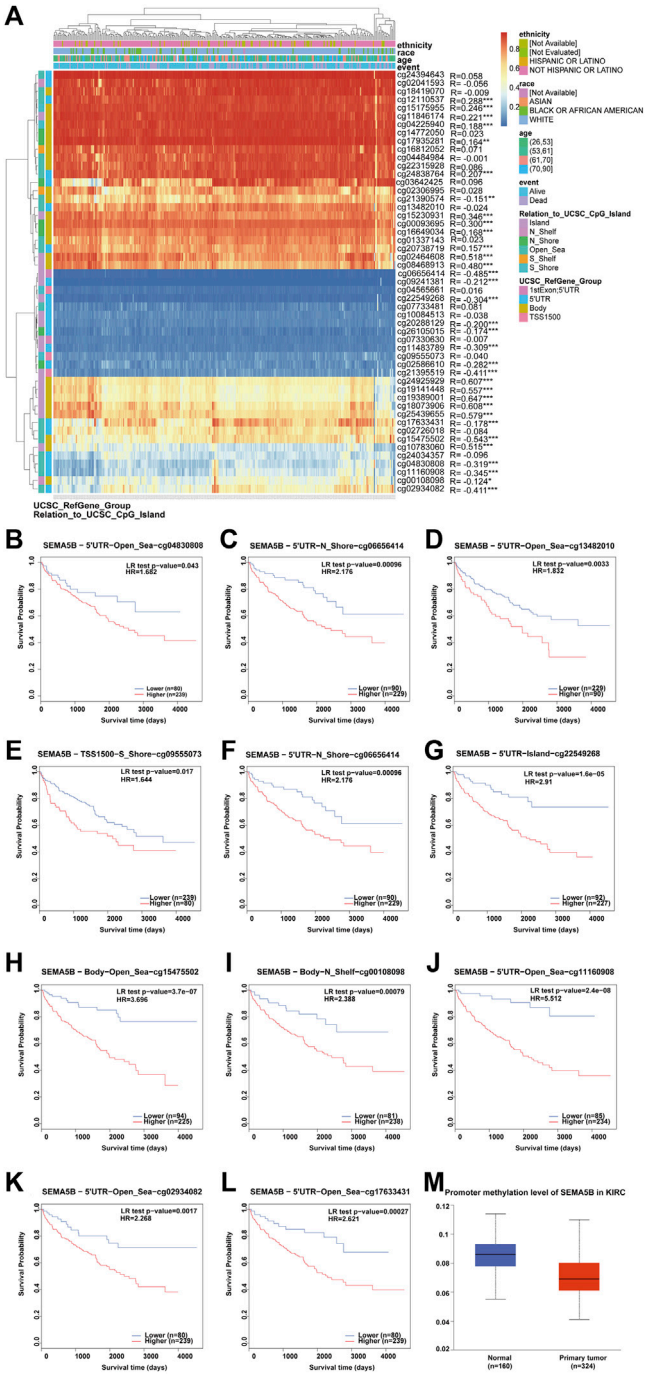


FIGURE 8 | Analysis of SEMA5B methylation in KIRC. **(A)** Heat map of the methylation level of the SEMA5B gene. The correlations between SEMA5B methylation and its expression were analyzed. $p < 0.05$ was considered statistically significant. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. **(B–J)** Survival analysis based on methylation at multiple sites. **(M)** The promoter methylation level of SEMA5B in KIRC.

and constitutive androstane receptor pathway. We further investigated the function of SEMA5B and the probable mechanism underlying the effects of SEMA5B on the progression and metastasis KIRC based on GSEA. As there is little literature on SEMA5B, with the performance of GSEA, we only found that _SYSTEMIC_LUPUS_ERYTHEMATOSUS and NABA_ECM_REGULATORS were significantly enriched in the SEMA5B low-expression phenotype.

Notably, SEMA5B may be involved in small molecule catabolic process, transport of small molecules, positive regulation of inflammatory response, and diseases of signal transduction by growth factor receptors and second

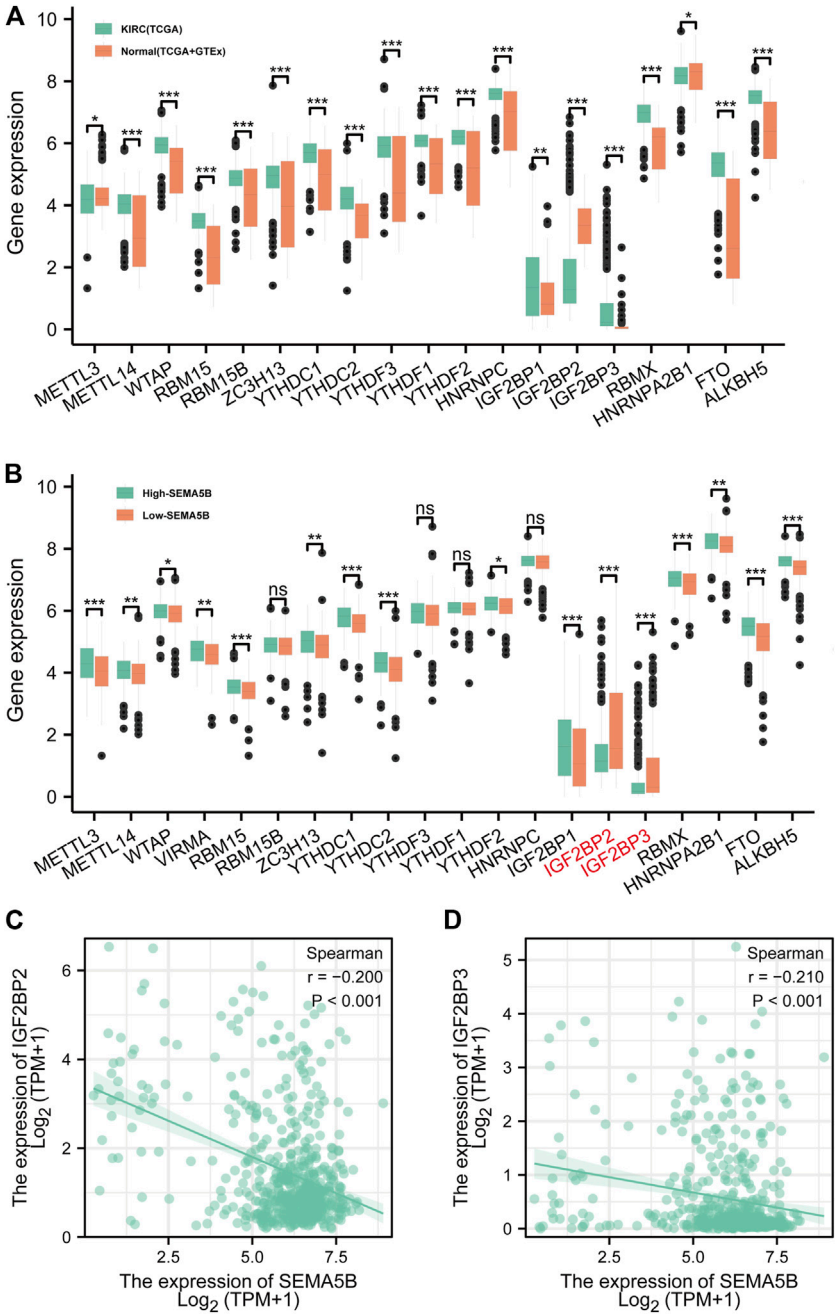


FIGURE 9 | The correlation between the SEMA5B expression and m6A regulators in KIRC. **(A)** m6A regulators expression between KIRC tissues and normal tissues according to the TCGA and GTEx. **(B)** m6A regulators expression between high- and low-SEMA5B expression groups of KIRC. Spearman correlation for the relationship between the expression levels of SEMA5B and IGF2BP2 **(C)**, and IGF2BP3 **(D)**. ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

messengers. It is well accepted that adhesion to ECM is a crucial step in cancer progression; the ECM components such as collagen, fibroblasts, and their associated signaling molecules contribute to tumor cell proliferation, migration, and invasion in various cancers (Brassart-Pasco et al., 2020; Moreira et al., 2020). In brief, those results exhibited that SEMA5B may play a crucial role in KIRC through crosslinking in the extracellular matrix (ECM).

In many cancer types, an immune infiltrate within the tumor is typically associated with a better prognosis and with response to immunotherapy (Robertson et al., 2017). In the present study, the abundance of infiltration by multiple immune cells (such as, $CD4^+$ T cells, B cells, NK cells, Myeloid dendritic cells, neutrophils, and monocytes) was found to significantly positively correlate with the expression of SEMA5B. The role of $CD4^+$ T cells is increasingly being studied that stromal infiltration of $CD4^+$ T cells in cancers is associated

with better OS and disease-specific survival (Soo et al., 2018; Niemeijer et al., 2020). NK cells have powerful antitumor effects. Similarly, NK cell infiltrates in primary colorectal, gastric, and lung cancer proved to be correlated with better patient survival outcomes (Malladi et al., 2016). Tumor-infiltrating dendritic cells are potent antigen-presenting cells (Mathios et al., 2016). Neutrophils are proven to be associated with better prognosis in different cancer types (Qu et al., 2018). Moreover, upon infiltration, monocytes differentiate and promote tumor cell death through an as-yet unknown mechanism (Maximov et al., 2019). Therefore, overexpressed SEMA5B appeared to enhance tumor immunity, synergistically enhance immune cell infiltration of tumor, and further boost the antitumor immune response, and finally inhibit tumor progression.

The importance of immune-checkpoints has been increasingly recognized in tumor immunology. During tumorigenesis, various immune checkpoints are induced to create an immunosuppressive TME for escaping immune surveillance (Quail and Joyce, 2013). We found immune checkpoints genes including SIGLEC15, TIGIT, CD274, HAVCR2, PDCD1, CTLA4, LAG3, and PDCD1LG2 were highly expressed in KIRC. Previous studies have reported that SIGLEC15 as an immune suppressor inhibits T cell proliferation and activation *in vitro* and *in vitro* (Wang et al., 2019). In addition, they found SIGLEC15 deficiency promoted T cell responses, better tumor control, and overall survival in a mouse melanoma model. In this study, we found that SEMA5B levels showed a negative correlation with SIGLEC15 in KIRC, suggesting that SEMA5B may effect of SIGLEC15 on tumor immunity. Combined with the previous findings, we speculated that the elevated SEMA5B may interfere with SIGLEC15 expression through some unknown mechanism, thus affecting the immune effect of cytotoxic T cells and ultimately prolonging the survival of tumor patients. It suggests that SIGLEC15 is highly likely to be a potential therapeutic target of KIRC. However, the specific mechanism still needs further in-depth study.

Previous literature indicates that mutations or defects in the MMR genes (MLH1, MSH2, MSH6, PML2, and EPCAM) can lead to the accumulation of genetic errors, resulting in genomic or microsatellite instability, which contributes to the development of tumors (Armaghany et al., 2012). In addition, the function loss of those genes brings about a mutator phenotype which causes an increasing tumor risk (Jiricny, 2006). Our results show that SEMA5B expression was positively related to the expression of MMRs in KIRC. The relationship between DNA methyltransferases (DNMTs) and tumorigenesis has been widely discussed. Previous studies have shown that DNA methylation is a common epigenetic feature of cancer. (Manoochehri et al., 2020; Yin et al., 2020). In this study, we found a significantly positive correlation between the expression of SEMA5B and DNMTs levels in KIRC. Moreover, we found a correlation between hypermethylation at multiple sites of SEMA5B and a poor prognosis in KIRC. In general, high levels of promoter methylation tend to reduce gene expression or silence the gene (Jones, 2012). Our findings show that the degree of methylation in the SEMA5B promoter region was significantly lower in KIRC compared to normal people. Thus, our findings indicate that gene-related modifications may have little influence on the gene expression of SEMA5B, while factors that reduce the expression of SEMA5B will more or less affect the survival of patients, which further demonstrates that SEMA5B is indeed a protective factor of KIRC.

Although our current study has conducted in-depth exploration of the role of SEMA5B in KIRC, there are still some limitations. First, despite using multiple database data, however, it was not validated in our own clinical samples. Thus, further experimental verifications are necessary. Second, the prognostic value of SEMA5B in KIRC also needs further verification. Third, this study only focuses on the expression, diagnosis, and prognosis of SEMA5B, the mechanisms by which SEMA5B promotes tumor progression and metastases in KIRC need further elucidation.

CONCLUSION

In conclusion, our study indicated that SEMA5B is significantly upregulated and is associated with immune infiltration in KIRC. SEMA5B can serve as a favorable prognostic factor and a novel diagnostic biomarker for KIRC. In addition, SEMA5B may involve in crosslinking in the extracellular matrix (ECM) and correlates with the abundance MMR genes, DNMTs, and m6A regulators in KIRC. Therefore, we hypothesized that SEMA5B may be promising molecular targets for the early diagnosis, a potential prognostic biomarker and targeted therapy of KIRC.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conception and design: JC; provision of study materials or patients: JD, SZ, XC, CL, JP, JZ, and YS; collection and assembly of data: JD, SZ, XC, CL, and ZL; data analysis and interpretation: JD, SZ, and JC; manuscript writing: all authors; and final approval of the manuscript: all authors.

FUNDING

This study was supported by the Guangxi Zhuang Autonomous Region Health and Family Planning Commission: Z20211108.

ACKNOWLEDGMENTS

The authors sincerely thank all participants involved in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.835355/full#supplementary-material

REFERENCES

Armaghany, T., Wilson, J. D., Chu, Q., and Mills, G. (2012). Genetic Alterations in Colorectal Cancer. *Gastrointest. Cancer Res.* 5 (1), 19

Atkins, R. J. T. L. O. (2009). Resistance to Targeted Therapy in Renal-Cell Carcinoma [J]. *Lancet Oncol.* 10, 992. doi:10.1016/S1470-2045(09)70240-2

Basile, J. R., Holmbeck, K., Bugge, T. H., and Gutkind, J. S. (2007). MT1-MMP Controls Tumor-Induced Angiogenesis through the Release of Semaphorin 4D. *J. Biol. Chem.* 282, 6899–6905. doi:10.1074/jbc.M609570200

Brassart-Pasco, S., Brézillon, S., Brassart, B., Ramont, L., Oudart, J.-B., and Monboisse, J. C. (2020). Tumor Microenvironment: Extracellular Matrix Alterations Influence Tumor Progression. *Front. Oncol.* 10, 397. doi:10.3389/fonc.2020.00397

Cao, H., Song, X., and Hong, Y. J. (2021). Increased Expression of SEMA5B in Gastric Adenocarcinoma Predicts Poor Prognosis: a Study Based on TCGA Data. *Zhounghua. Zhong. Liu. Za Zhi* 43 (8), 856–860. doi:10.3760/cma.j.cn112152-20200104-00005

Chin, A. I., Lam, J. S., Figlin, R. A., and Belldegrun, A. S. (2006). Surveillance Strategies for Renal Cell Carcinoma Patients Following Nephrectomy. *Rev. Urol.* 8 (1), 1–7.

Cuellar, L. M. (2009). Identification and Localization of Novel Genes Preferentially Expressed in Human Kidney Glomerulus, *J. Blackwell Publishing Asia [J]* 14(1): 94–104. doi:10.1111/j.1440-1797.2008.01009.x

Farahbod, M., and Pavlidis, P. (2020). Untangling the Effects of Cellular Composition on Coexpression Analysis. *Genome Res.* 30 (6), 849–859. doi:10.1101/gr.256735.119

Gao, H., Li, Y., and Lin, T. (2020). CIP2A Promotes the Survival of Renal clear Cell Carcinoma Caki-2 Cells [J]. *FEBS. Open. Bio.* 11 542. doi:10.1002/2211-5463.12870

Georgakopoulos-Soares, I., Koh, G., Momen, S. E., Jiricny, J., Hemberg, M., and Nik-Zainal, S. (2020). Transcription-coupled Repair and Mismatch Repair Contribute towards Preserving Genome Integrity at Mononucleotide Repeat Tracts. *Nat. Commun.* 11 (1), 1980. doi:10.1038/s41467-020-15901-w

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., et al. (2019). Next-generation Characterization of the Cancer Cell Line Encyclopedia. *Nature* 569 (7757), 503–508. doi:10.1038/s41586-019-1186-3

Hirota, E., Yan, L., Tsunoda, T., Ashida, S., Fujime, M., Shuin, T., et al. (2006). Genome-wide Gene Expression Profiles of clear Cell Renal Cell Carcinoma: Identification of Molecular Targets for Treatment of Renal Cell Carcinoma. *Int. J. Oncol.* 29 (4), 799. doi:10.3892/ijo.29.4.799

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., et al. (2017). Renal Cell Carcinoma. *Nat. Rev. Dis. Primers* 3, 17009. doi:10.1038/nrdp.2017.9

Jiricny, J. (2006). The Multifaceted Mismatch-Repair System. *Nat. Rev. Mol. Cel Biol* 7 (5), 335–346. doi:10.1038/nrm1907

Jones, P. A. (2012). Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and beyond. *Nat. Rev. Genet.* 13 (7), 484–492. doi:10.1038/nrg3230

Kho, P. F., Wang, X., Cuéllar-Partida, G., Dörk, T., Goode, E. L., Lambrechts, D., et al. (2021). Multi-tissue Transcriptome-wide Association Study Identifies Eight Candidate Genes and Tissue-specific Gene Expression Underlying Endometrial Cancer Susceptibility. *Commun. Biol.* 4 (1), 1211. doi:10.1038/s42003-021-02745-3

Kundu, A., Nam, H., Shelar, S., Chandrashekar, D. S., Brinkley, G., Karki, S., et al. (2020). PRDM16 Suppresses HIF-Targeted Gene Expression in Kidney Cancer. *J. Exp. Med.* 217 (6), 1005. doi:10.1084/jem.20191005

Lánczky, A., Nagy, Á., and Bottai, G. (2016). miRpower: a Web-Tool to Validate Survival-Associated miRNAs Utilizing Expression Data from 2178 Breast Cancer Patients. *Breast Cancer Res. Treat.* 160 (3), 439–446.

Liu, R. Q., Wang, W., Legg, A., Abramyan, J., and O'Connor, T. P. (2014). Semaphorin 5B Is a Repellent Cue for Sensory Afferents Projecting into the Developing Spinal Cord. *Devolpment* 141 (9), 1940–1949. doi:10.1242/dev.103630

Mäki-Nevala, S., Sarhadi, V. K., Knuuttila, A., Scheinin, I., Ellonen, P., Lagström, S., et al. (2016). Driver Gene and Novel Mutations in Asbestos-Exposed Lung Adenocarcinoma and Malignant Mesothelioma Detected by Exome Sequencing. *Lung* 194 (1), 125–135. doi:10.1007/s00408-015-9814-7

Malladi, S., Macalinao, D. G., Jin, X., He, L., Basnet, H., Zou, Y., et al. (2016). Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* 165 (1), 45–60. doi:10.1016/j.cell.2016.02.025

Manoochehri, M., Wu, Y., Giese, N. A., Strobel, O., Kutschmann, S., Haller, F., et al. (2020). SST Gene Hypermethylation Acts as a Pan-Cancer Marker for Pancreatic Ductal Adenocarcinoma and Multiple Other Tumors: toward its Use for Blood-Based Diagnosis. *Mol. Oncol.* 14 (6), 1252–1267. doi:10.1002/1878-0261.12684

Mathios, D., Kim, J. E., Mangraviti, A., Phallen, J., Park, C. K., Jackson, C. M., et al. (2016). Anti-PD-1 Antitumor Immunity Is Enhanced by Local and Abrogated by Systemic Chemotherapy in GBM. *Sci. Transl Med.* 8 (370), 370ra180. doi:10.1126/scitranslmed.aag2942

Matsuoka, R. L., Chivatakarn, O., Badea, T. C., Samuels, I. S., Cahill, H., Katayama, K.-i., et al. (2011). Class 5 Transmembrane Semaphorins Control Selective Mammalian Retinal Lamination and Function. *Neuron* 71 (3), 460–473. doi:10.1016/j.neuron.2011.06.009

Maximov, V., Chen, Z., Wei, Y., Robinson, M. H., Herting, C. J., Shanmugam, N. S., et al. (2019). Tumour-associated Macrophages Exhibit Anti-tumoural Properties in Sonic Hedgehog Medulloblastoma. *Nat. Commun.* 10 (1), 2410. doi:10.1038/s41467-019-10458-9

Moreira, A. M., Pereira, J., Melo, S., Fernandes, M. S., Carneiro, P., Seruca, R., et al. (2020). The Extracellular Matrix: An Accomplice in Gastric Cancer Development and Progression. *Cells* 9 (2). doi:10.3390/cells9020394

Niemeijer, A.-L. N., Sahba, S., Smit, E. F., Lissenberg-Witte, B. I., de Langen, A. J., and Thunnissen, E. (2020). Association of Tumour and Stroma PD-1, PD-L1, CD3, CD4 and CD8 Expression with DCB and OS to Nivolumab Treatment in NSCLC Patients Pre-treated with Chemotherapy. *Br. J. Cancer* 123 (3), 392–402. doi:10.1038/s41416-020-0888-5

Panneerdoss, S., Eedunuri, V. K., Yadav, P., Timilsina, S., Rajamanickam, S., Viswanadhapalli, S., et al. (2018). Cross-talk Among Writers, Readers, and Erasers of m6A Regulates Cancer Growth and Progression. *Sci. Adv.* 4, eaar8263. doi:10.1126/sciadv.aar8263

Qu, X., Tang, Y., and Hua, S. (2018). Immunological Approaches towards Cancer and Inflammation: A Cross Talk. *Front. Immunol.* 9, 563. doi:10.3389/fimmu.2018.00563

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental Suppressor and Potential Target for Normalization Cancer Immunotherapy [J]. *Nat. Med.*

Ravaud, A., Motzer, R. J., Pandha, H. S., George, D. J., Pantuck, A. J., Patel, A., et al. (2016). Adjuvant Sunitinib in High-Risk Renal-Cell Carcinoma after Nephrectomy. *N. Engl. J. Med.* 375, 2246–2254. doi:10.1056/NEJMoa1611406

Robertson, A. G., Shih, J., Yau, C., Gibb, E. A., Oba, J., Mungall, K. L., et al. (2017). Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer cell* 32 (2), 204–e15. doi:10.1016/j.ccell.2017.07.003

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12 (1), 77. doi:10.1186/1471-2105-12-77

Schwab, M. E. (1996). Molecules Inhibiting Neurite Growth: A Minireview. *Neurochem. Res.* 21 (7), 755–761. doi:10.1007/bf02532297

Shuch, B., Amin, A., Armstrong, A. J., Eble, J. N., Ficarra, V., Lopez-Beltran, A., et al. (2015). Understanding Pathologic Variants of Renal Cell Carcinoma: Distilling Therapeutic Opportunities from Biologic Complexity. *Eur. Urol.* 67 (1), 85–97. doi:10.1016/j.eururo.2014.04.029

Siegel, R. L., Miller, K. D., and Jemal, A. J. (2018). *Cancer Statistics*. Atlanta: American Cancer Society, 68. doi:10.3322/caac.21442

Sigel, R. K., Ma, J., and Zou, Z. (2002). *Global Cancer Statistics [J]*. Atlanta: American Cancer Society.

Smyth, G. K. J. B. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26 (1), 139, doi:10.1093/bioinformatics/btp616

Soo, R. A., Chen, Z., Yan Teng, R. S., Tan, H.-L., Iacopetta, B., Tai, B. C., et al. (2018). Prognostic Significance of Immune Cells in Non-small Cell Lung Cancer: Meta-Analysis. *Oncotarget* 9 (37), 24801–24820. doi:10.18632/oncotarget.24835

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tiffen, J., Gallagher, S. J., Filipp, F., Gunatilake, D., Emran, A. A., Cullinane, C., et al. (2020). EZH2 Cooperates with DNA Methylation to Downregulate Key Tumor Suppressors and IFN Gene Signatures in Melanoma. *J. Invest. Dermatol.* 140 (12), 2442–2454. doi:10.1016/j.jid.2020.02.042

Wang, J., Sun, J., and Liu, L. N. (2019). Siglec-15 as an Immune Regulation of Tumor Progression and Metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394

Wickham, H. (2016). *ggplot2, Elegant Graphics for Data Analysis*. Berlin: Springer-Verlag. doi:10.1007/978-3-319-24277-4

Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., et al. (2014). Corrigendum: Genome-wide Association Analyses of Esophageal Squamous Cell Carcinoma in Chinese Identify Multiple Susceptibility Loci and Gene-Environment Interactions. *Nat. Genet.* 46 (9), 1040–1041. doi:10.1038/ng0914-1040b

Yin, Y., Che, K., and Hu, J. (2020). Hypermethylation of the RSK4 Promoter Associated with BRAF V600E Promotes Papillary Thyroid Carcinoma [J]. *Cint. J. Oncol.* 26, 1248. doi:10.3892/ijo.2020.4999

Zhang, H., He, J., Dai, Z., Wang, Z., Liang, X., He, F., et al. (2021). PDIA5 Is Correlated with Immune Infiltration and Predicts Poor Prognosis in Gliomas. *Front. Immunol.* 12, 628966. doi:10.3389/fimmu.2021.628966

Zheng, F., Du, F., Qian, H., Zhao, J., Wang, X., Yue, J., et al. (2021). Expression and Clinical Prognostic Value of m6A RNA Methylation Modification in Breast Cancer. *Biomark. Res.* 9 (1), 28. doi:10.1186/s40364-021-00285-w

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ding, Zhao, Chen, Luo, Peng, Zhu, Shen, Luo and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fgene.2022.877086

Identification of a Gene Signature to Aid Treatment Decisions by Integrated Analysis of Mutated Genes Between Primary and Metastatic Prostate Cancer

*Qinyu Li[1], Xueyan Xiao[2], Bingliang Chen[1], Guoda Song[1], Kai Zeng[1] and Jianping Miao[3]**

[1]Department of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Gastroenterology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Department of Geriatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China*

Prostate cancer is one of the most common malignancies in males. Despite the recent development of advanced diagnostic platforms and treatment, patients with metastatic disease still have a poor five-year survival rate. Cancer metastasis is correlated with the characteristics of the tumor microenvironment and is significantly associated with patient prognosis. In this study, we obtained mutated genes with significant differences between primary and metastatic prostate cancer from the COSMIC database. Unsupervised consensus clustering was used based on the 1,051 genes obtained, and two PCa clusters were identified, which exhibited different prognostic outcomes and immune characteristics. Next, we generated a scoring system and evaluated the prognostic value of riskscore and its potential to aid treatment decisions in clinical practice. The riskscore could be applied to predict patients' response to immunotherapy and sensitivity to Docetaxel. In conclusion, this study performed an integrated analysis of mutated genes between primary and metastatic prostate cancer and provides a novel assessment scheme to precisely select treatment strategies.

Keywords: metastasis, mutation, prostate cancer, treatment decision, unsupervised consensus clustering

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Atrayee Bhattacharya,
Dana–Farber Cancer Institute, United States
Neeti Sharma,
Ajeenkya D Y Patil University, India

****Correspondence:***
Jianping Miao
miaojianping888@hotmail.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *16 February 2022*
Accepted: *25 March 2022*
Published: *12 April 2022*

Citation:
Li Q, Xiao X, Chen B, Song G, Zeng K and Miao J (2022) Identification of a Gene Signature to Aid Treatment Decisions by Integrated Analysis of Mutated Genes Between Primary and Metastatic Prostate Cancer. Front. Genet. 13:877086. doi: 10.3389/fgene.2022.877086

INTRODUCTION

Prostate cancer (PCa) is the second most common type of cancer diagnosed in males (Sung et al., 2021). While several patients run an indolent course, most patients present with high-risk localized, locally advanced, or metastatic cancer (Teo et al., 2019). Despite localized prostate cancer exhibiting long-term survival, metastatic prostate cancer remains largely incurable even after intensive treatment (Wang et al., 2018). It has been reported that more than ninety percent of cancer-related deaths result from metastasis, and most prostate cancer patients die from metastasis (Rycaj et al., 2017). Therefore, exploring those genes

Abbreviations: COSMIC, The Catalogue of Somatic Mutations in Cancer; ECM, Extracellular matrix; FC, Fold change; GDSC, Genomics of Drug Sensitivity in Cancer; GEO, Gene Expression Omnibus; GSVA, Gene set variation analysis; PCa, Prostate cancer; PCA, principal components analysis; ROC, Receiver operating characteristic; ssGSEA, Single sample gene set enrichment analysis; TCGA, The Cancer Genome Atlas; TMB, Tumor mutation burden; TME, Tumor microenvironment.

with significant differences between primary and metastatic prostate cancer may help us to predict the prognosis of patients and formulate a more effective treatment regimen.

Metastasis is processed by the mechanisms by which tumor cells invade local tissues, reach the circulation, and colonize distant organs (López-Soto et al., 2017). Recent research has suggested that immune cells can regulate these steps of metastasis by influencing the extracellular matrix (ECM) (Blomberg et al., 2018). ECM remodeling can facilitate metastasis by influencing the architecture of the surrounding tissue to favor tumor cell invasion (Ghajar and Bissell, 2008) or allowing the release and diffusion of the pro-tumoral signaling molecules (Cox and Erler, 2011). Currently, the clinical successes of immunotherapy, such as immune checkpoint blockade, have revolutionized cancer therapeutics, and astonishing clinical responses have been achieved in several types of cancers (van den Bulk et al., 2018). Immunotherapy results in long-term durable remission in some advanced cancer patients (Ganesh et al., 2019). However, a large proportion of patients cannot benefit from checkpoint blockade. Therefore, how to choose suitable treatment is critical in clinical practice, and the development of immunotherapy calls for a better understanding of the influence of immune regulation on metastasis to enhance the treatment efficacy for patients with metastatic disease.

COSMIC, the Catalogue Of Somatic Mutations In Cancer, contains the most detailed and comprehensive materials of somatic mutations in human cancer (Tate et al., 2019). Its latest release includes almost 6 million coding mutations across 1.4 million tumor samples. In this study, we downloaded mutation data and corresponding sample features and performed Chi-square test to identify those genes with significant differences in mutation frequencies between primary and metastatic prostate cancer. Then, we filtered these genes by the univariate Cox regression method, performed unsupervised clustering method and identified two PCa clusters based on these mutated genes. Comprehensive analysis revealed that the two subclasses were significantly different in progression-free survival, characteristics of the immune microenvironment and the expression of immune checkpoint genes. Moreover, we extracted the feature genes to construct a riskscore by principal component analysis and evaluated the prognostic value of the riskscore and its potential to aid treatment decisions in clinical practice.

MATERIALS AND METHODS

Data Collection and Pre-Processing

Transcriptional data (read counts), clinical characteristics and somatic mutation data were acquired from the TCGA database (**Supplementary Table S1**). Next, we downloaded four datasets with the same platform (Affymetrix GPL570) from the GEO database: GSE69223 (Meller et al., 2016) (N = 30), GSE32448 (Derosa et al., 2012) (N = 80), GSE55945 (Arredouani et al., 2009) (N = 21), and GSE46602 (Mortensen et al., 2015) (N = 50). Thereafter, we adjusted the background by using the "RMA" algorithm of the "affy" R package (Gautier et al., 2004) and removed the batch effect by the "ComBat" algorithm of the "sva" package (Leek et al., 2012). Therefore, we can merge the four datasets as the validation cohort. Moreover, GSE21034 (Taylor et al., 2010) (N = 370) was utilized to validate the prognostic value.

Identification of Mutated Genes Between Primary and Metastatic Prostate Cancer

COSMIC is currently the broadest database of mutations in human cancer (Forbes et al., 2017). COSMIC mutation data (Genome Screens) and corresponding sample features were downloaded, and we estimated the frequency of each mutation site. Chi-square test was applied to discover the mutated genes with significant differences between primary and metastatic prostate cancer. Next, we executed prognostic analysis for each gene discovered by univariate Cox regression, and the genes related to prognosis with p-value < 0.05 were extracted for further analysis.

Identification of PCa Subclasses

Unsupervised consensus clustering of the obtained genes was executed by using the k-means algorithm in the "ConsensusClusterPlus" package (Wilkerson and Hayes, 2010), which was repeated 1,000 times to ensure the stability of the classification. Survival differences between the two clusters were visualized by Kaplan–Meier curves. To explore the molecular characteristics of the two clusters. The "c2.cp.kegg.v7.4.symbols" gene set was downloaded from the MSigDB database, and we applied the "GSVA" package (Hänzelmann et al., 2013) to perform the GSVA analysis.

Immune Infiltration Levels Between PCa Subclasses

The "ssGSEA" method was performed to estimate the infiltration degrees of 28 immune cells by using the "GSVA" package. Estimate is commonly used to calculate scores reflecting the infiltration levels of immune cells and stromal cells in the tumor microenvironment by the package "estimate" (Yoshihara et al., 2013). We applied the estimate algorithm to calculate the ImmuneScore and StromalScore of each sample. Additionally, the correlation between the expression of immune checkpoint genes and androgen receptor between the two clusters was estimated.

Generation of Riskscore

The Pearson correlation coefficients of mutated genes with the identified PCa subclasses were estimated, and the signature genes A and B were obtained based on the correlation coefficients. Then, we applied the Boruta algorithm to the positively and negatively correlated genes to select feather genes. Finally, principal component analysis (PCA) was performed to estimate the first principal components of signature genes A and B. We defined the riskscore of each sample as follows:

$$\text{Riskscore} = \sum PC1B - \sum PC1A$$

Correlation Between Clinical Parameters, Immune Infiltration and Riskscore

The difference in the riskscore in patients stratified by clinical parameters was estimated to expound the effect of the riskscore on

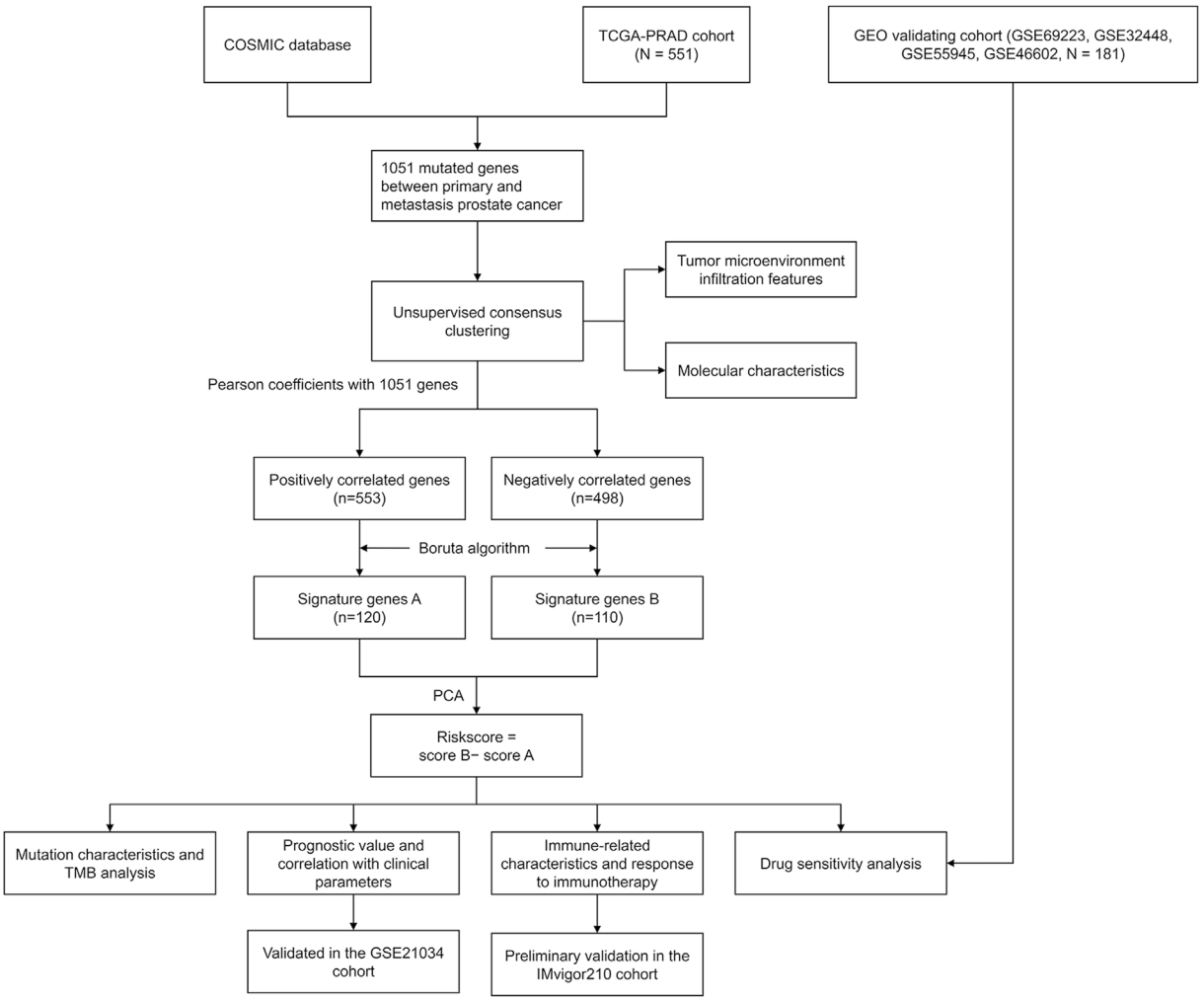


FIGURE 1 | Flow chart of this study.

cancer progression. Moreover, immune cell infiltration, ImmuneScore, StromalScore and the expression of immune checkpoint genes were also assessed between the high- and low-risk groups.

Tumor Mutation Burden Analysis

Tumor mutation burden (TMB) has been demonstrated as a predictive biomarker to identify whether patients with cancer can respond to immune checkpoint inhibitors well (Merino et al., 2020). Here, we explored the correlation between TMB and riskscore. Furthermore, we divided patients into four subgroups based on the median value of riskscore and TMB. Survival differences of the four subgroups were visualized by Kaplan–Meier curves.

Benefits of the Riskscore to Aid Treatment Decisions

Since the comparison of the expression of different immune checkpoint genes between the high- and low-risk groups was performed, here we used an immunotherapeutic cohort (IMvigor210 cohort) as a validation cohort (Mariathasan et al., 2018). We first evaluated the influence of the riskscore on the prognosis of patients treated with immunotherapy. Then, the riskscore of patients with different clinical statuses after treatment were compared. Finally, transcriptional data of tumor cell lines and IC50 values of antitumor drugs from the GDSC database were used to perform the drug sensitivity analysis by using the "pRRophetic" package (Geeleher et al., 2014).

Statistical Analysis

All analyses were performed in RStudio 4.0.4. Correlation analysis was computed by the Spearman method. Student's *t* test and the Wilcoxon test were applied for two-group comparisons. Correspondingly, the Kruskal–Wallis test and one-way ANOVA were used for multiple groups. Statistical significance was defined as p-value < 0.05.

RESULTS

Genetic Alterations in Prostate Cancer

The roadmap of this study is illustrated in **Figure 1**. The top 20 mutated genes for prostate cancer are shown in **Figure 2A**, and

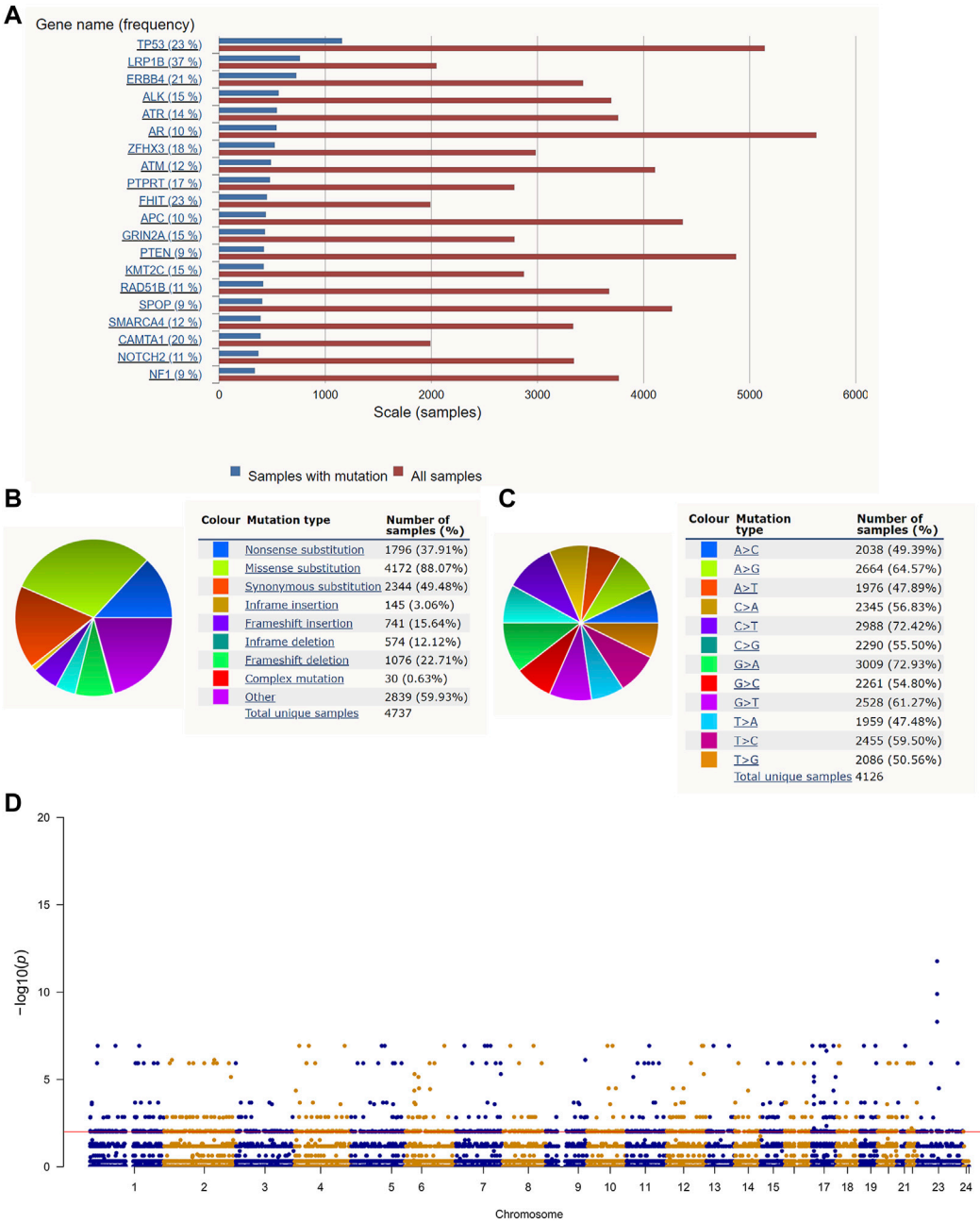


FIGURE 2 | Gene mutation in prostate cancer: **(A)** The top 20 mutated genes in prostate cancer. **(B,C)** The distribution of different types of mutations in prostate cancer. **(D)** The Manhattan plot of mutation sites that have significantly different mutation frequencies between primary and metastatic prostate cancer.

the mutation frequencies are displayed next to the gene name. Furthermore, **Figures 2B,C** shows the distribution of different types of mutations for prostate cancer. Missense substitution (88.07%), synonymous substitution (49.48%) and nonsense substitution (37.91%) were the main types of mutations, and the substitution mutations mainly included G > A (72.93%), C > T (72.42%), A > G (64.57%), and G > T (61.27%). Moreover, the Manhattan plot depicted mutation sites that had significant differences between primary and metastatic prostate cancer (**Figure 2D**).

Identification of PCa Subclasses

We obtained 3,716 mutated genes with significant differences between primary and metastatic prostate cancer by the Chi-square test (**Supplementary Table S2**). Thereafter, we explored the prognostic value of these genes for progression-

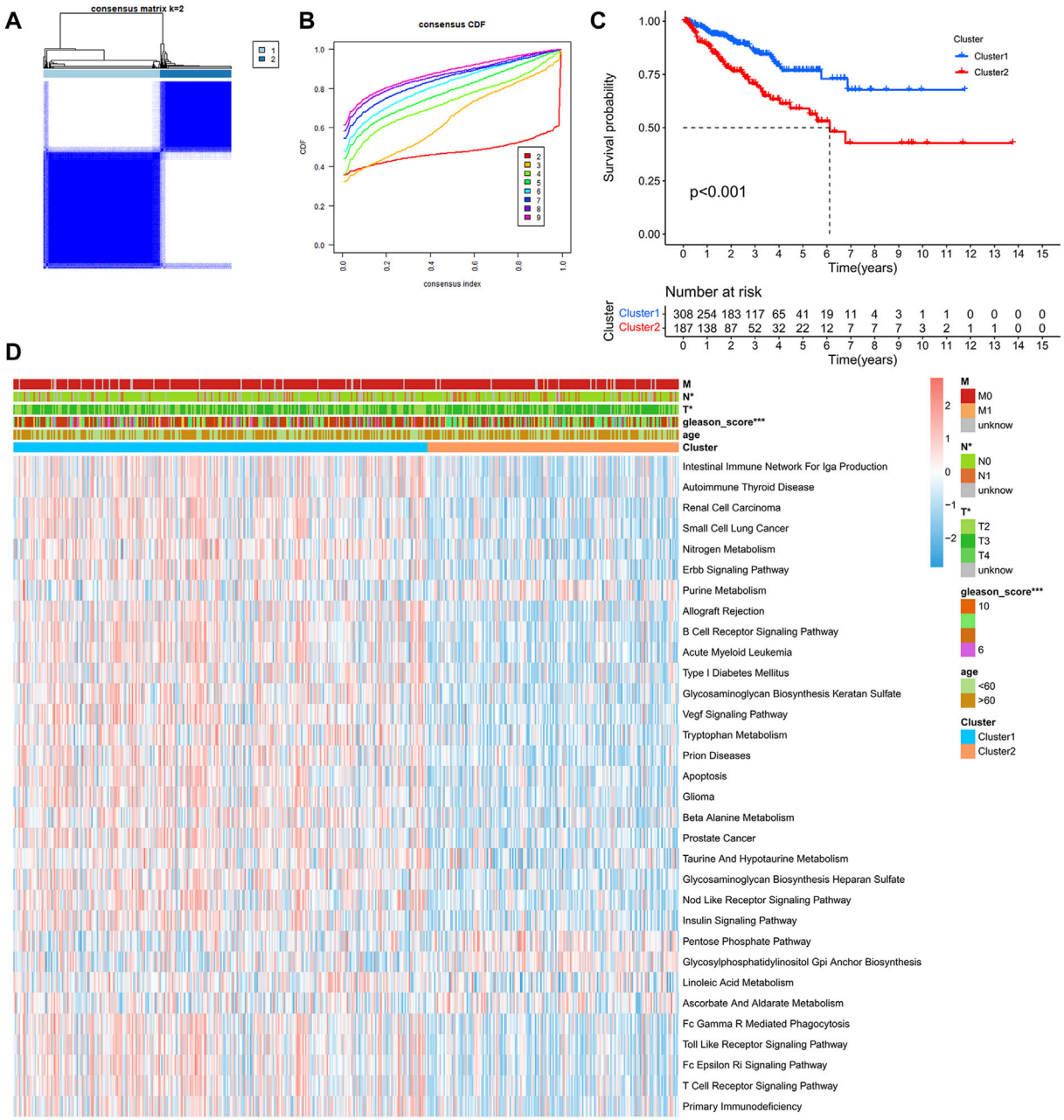


FIGURE 3 | Identification of PCa subclasses by unsupervised consensus clustering: **(A)** Matrix heatmap of the K-means clustering using 1,051 mutated genes between primary and metastatic prostate cancer. **(B)** CDF curve of the clustering result. **(C)** Kaplan–Meier survival curve of PFS between different clusters. **(D)** Heatmap of GSVA enrichment based on KEGG pathways between different clusters. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

free survival (PFS) by the univariate Cox method, and 1,051 genes were extracted (**Supplementary Table S3**) for further analysis.

After comprehensive consideration of CDF curves and delta area, we chose $k = 2$ as the optimal cluster number for the clustering (**Figure 3**). Finally, 308 patients were assigned to Cluster 1, and 187 patients were assigned to Cluster 2. We also applied t-SNE dimension reduction, and the results suggested that the discrimination among subgroups was decent (**Supplementary Figure S1**). Survival curves suggested that Cluster 2 had a significant survival advantage compared with Cluster 1 (**Figure 3C**). Moreover, GSVA analysis and limma analysis (log FC > 0.2, adjusted p-value < 0.05) were performed. Significant differences in pathways related to cancer progression, such as the ERBB signaling pathway and VEGF signaling pathway, and pathways associated with the immune response, such as the B cell receptor signaling pathway and T cell receptor signaling pathway, were observed between the two clusters (**Figure 3D**).

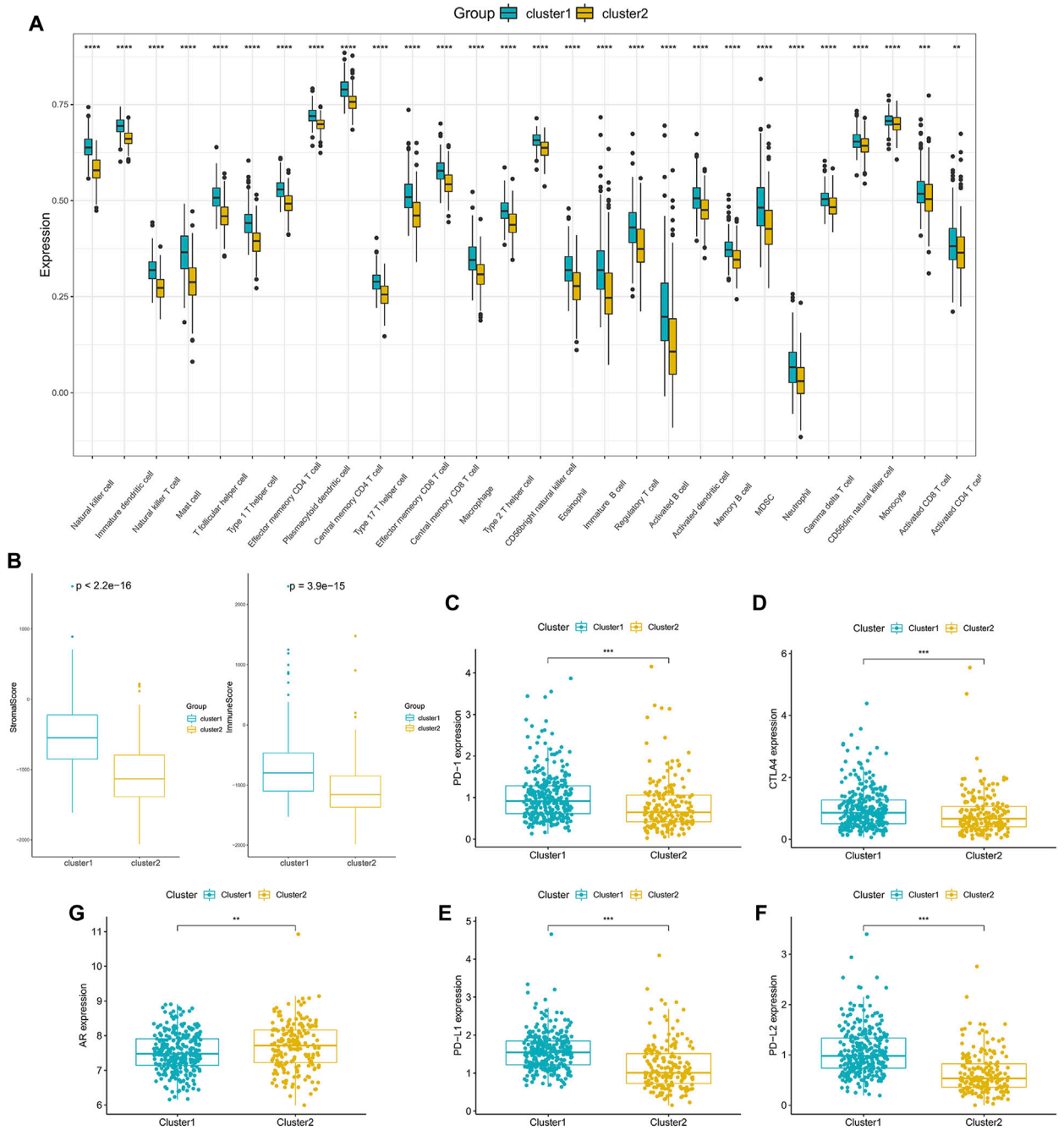


FIGURE 4 | Tumor microenvironment characteristics of different PCa subclasses: **(A)** The proportions of TME cells in the two clusters. **(B)** ImmuneScore and StromalScore of the two subgroups. **(C–F)** The expression of the immune checkpoint genes PD-1 **(C)**, CTLA4 **(D)**, PD-L1 **(E)** and PD-L2 **(F)** between the two clusters. **(G)** The expression of AR in the two clusters. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

In order to delve into the immune-related characteristics of the two clusters, the infiltration levels of immune cells were compared between the two clusters. A significant difference was observed in the infiltration degree of all immune cells, and all immune cells infiltration were lower in Cluster 2 (**Figure 4A**). Moreover, the results indicated that both the StromalScore and ImmuneScore of Cluster 2 were significantly lower compared with Cluster 1 (**Figure 4B**). What's more, the expression of immune checkpoint genes, including CTLA4, PD-1, PD-L1 and PD-L2, appeared to be decreased in Cluster 2 (**Figures 4C–F**). However, the expression of AR was higher in Cluster 2 than in Cluster 1 (**Figure 4G**), which is consistent with the poor survival of Cluster 2.

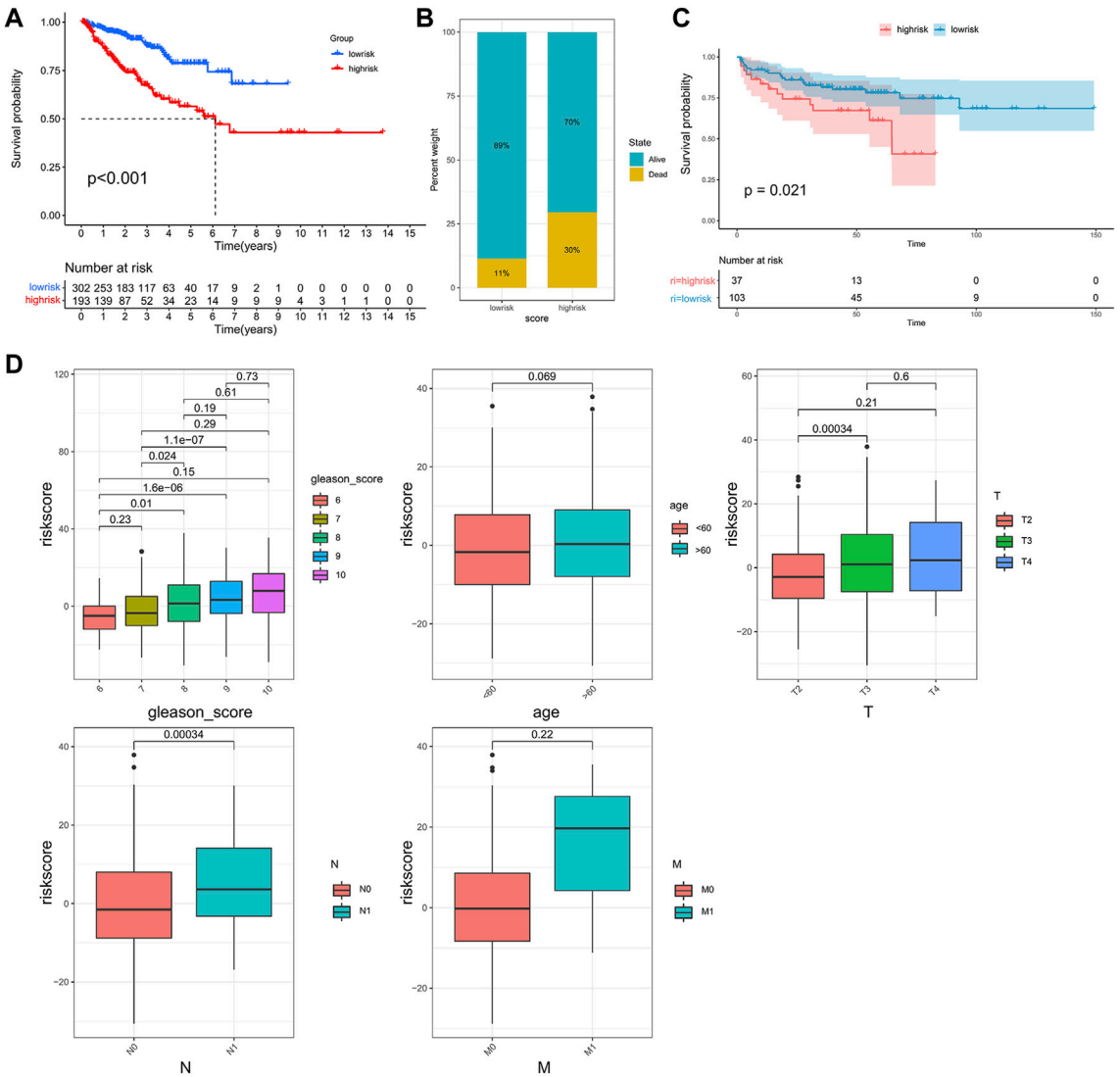


FIGURE 5 | Construction of the riskscore and prognosis analysis. **(A)** Kaplan–Meier survival curve between high- and low-risk subgroups in the TCGA cohort. **(B)** The proportion of the survival rate between the high- and low-risk subgroups. **(C)** Kaplan–Meier survival curve between the high- and low-risk subgroups in the validation cohort, GSE21034. **(D)** The differences in riskscore between patients with different clinicopathological parameters (Gleason score, age, T stage, N stage, M stage).

Construction of the Riskscore For Each Sample and the Prognostic Value

Previous results indicated that the subclass was closely associated with the prognosis and immune infiltration levels of patients. However, this population-based classification cannot be directly used in clinical practice. Therefore, we constructed a scoring system to estimate the riskscore to predict the outcome of the patients and aid in making treatment decisions. After performing the Boruta algorithm, 230 genes that were positively and negatively correlated to the subclass were defined as signature genes A and B (**Supplementary Table S4**). The riskscore was acquired by conducting PCA on each signature gene (**Supplementary Table S5**). Patients were classified into high- and low-risk groups according to the cut-off point gained by the "survminer" package (**Supplementary Figure S2**). The results of survival analysis showed that patients in the high-risk group had lower PFS than those in the low-risk group (**Figure 5A**). As shown in **Figure 5B**, the high-risk group possessed a higher proportion of death. In the validation cohort GSE21034, patients with higher riskscore also showed a significantly shorter median PFS (**Figure 5C**). Moreover, it was observed that the riskscore was elevated in the high-risk clinical group with the progression of tumor (**Figure 5D**), and patients who achieved complete response after treatment had a significantly lower riskscore than other outcomes (**Supplementary Figure S3**).

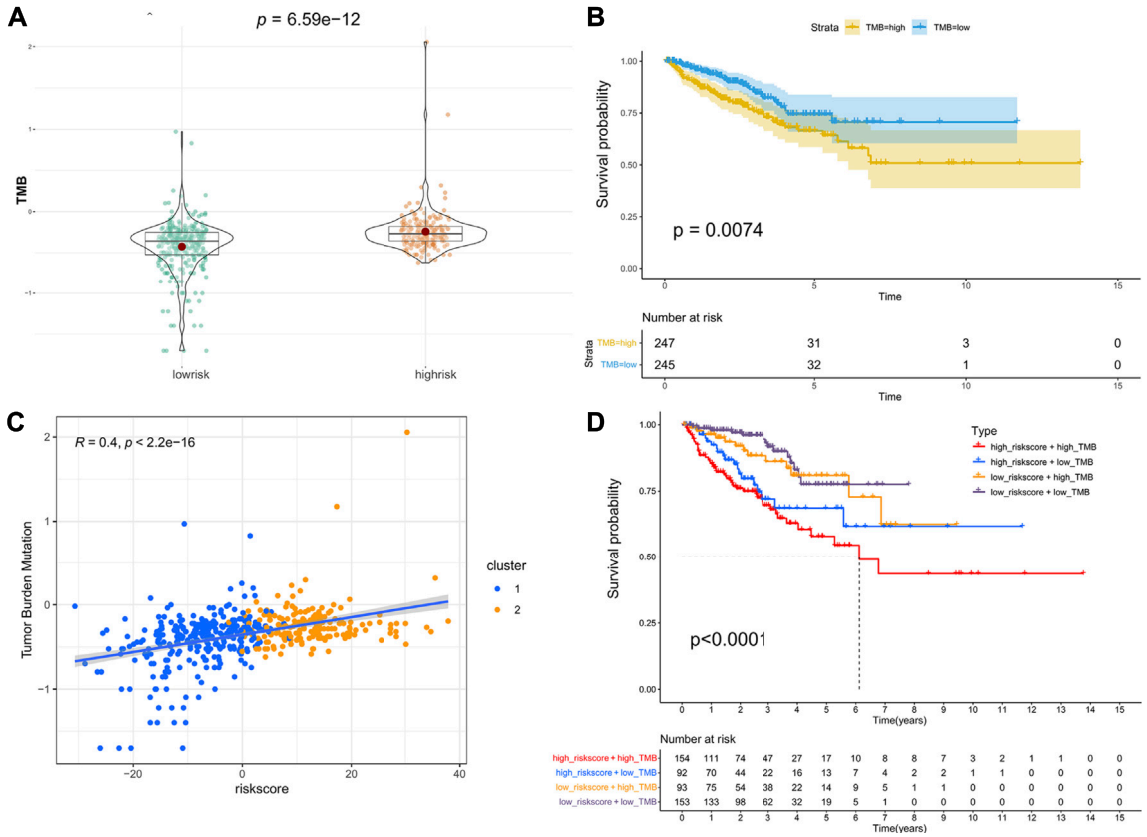


FIGURE 6 | Association between riskscore and TMB: **(A)** The difference in TMB between the high- and low-risk groups. **(B)** Kaplan–Meier survival curves of PFS between the high- and low-TMB groups. **(C)** The correlation of riskscore and TMB. **(D)** Kaplan–Meier survival curves of PFS among the four subgroups stratified by riskscore and TMB.

Relationship Between Riskscore and TMB

TMB is emerging as a potential biomarker to predict the patients response to immune checkpoint inhibitors (Chan et al., 2019). Here, we evaluated the association between the riskscore and TMB. As shown in **Figure 6A**, patients in the high-risk group had a higher TMB than those in the low-risk group, and patients with a higher TMB had lower PFS (**Figure 6B**). Moreover, the correlation analysis indicated that the riskscore was positively associated with TMB (**Figure 6C**). Next, we combined the riskscore and TMB to divide patients into four subgroups. The patients with a high riskscore and high TMB had the shortest median PFS, and patients with a low riskscore and low TMB performed the best prognosis (**Figure 6D**). Finally, the mutation status of genes with high mutation frequencies in the high- and low-risk groups was visualized (**Supplementary Figure S4**).

Correlation Between Immunotherapy Reactivity, Drug Sensitivity and Riskscore

Anticancer immunotherapies involving immune checkpoint inhibitors have emerged as new therapeutic regimens (O'Donnell et al., 2019). The tumor microenvironment (TME) was proven to be tightly linked to tumor progression and metastasis (Brassart-Pasco et al., 2020) and can blunt the therapeutic response, thus affecting the clinical outcome (Wu and Dai, 2017). To further explore the correlations between patients' response to immunotherapy and riskscore, we first compared the immune cell infiltration levels between high- and low-risk groups, and the results indicated that compared with the low-risk patients, the infiltration levels of 28 immune cells in the high-risk patients were significantly downregulated (**Figure 7A**).

Moreover, infiltration estimation for TCGA was downloaded from TIMER 2.0 (Li et al., 2020), which includes several methods, such as XCELL, TIMER, QUANTISEQ, MCPCOUNTER, EPIC, and CIBERSORT. We correlated the immune cell infiltration levels and the riskscore, and the results also suggested that the riskscore was negatively related to most of the immune cells (**Supplementary Figure S5**). Additionally, the StromalScore and ImmuneScore of the high-risk patients were significantly lower than low-risk patients as well (**Figure 7B**). Furthermore, the associations between immune checkpoint inhibitor genes and the riskscore were evaluated. The expression of ICI genes, such as CTLA4, PD-1, PD-L1 and LAG3, were downregulated in the high-risk groups compared with the low-risk groups (**Figure 7C**). Finally, we validated the performance of the riskscore in the IMvigor210 cohort. As shown in **Figure 8A**, the low-risk patients still showed a significant survival advantage compared with high-

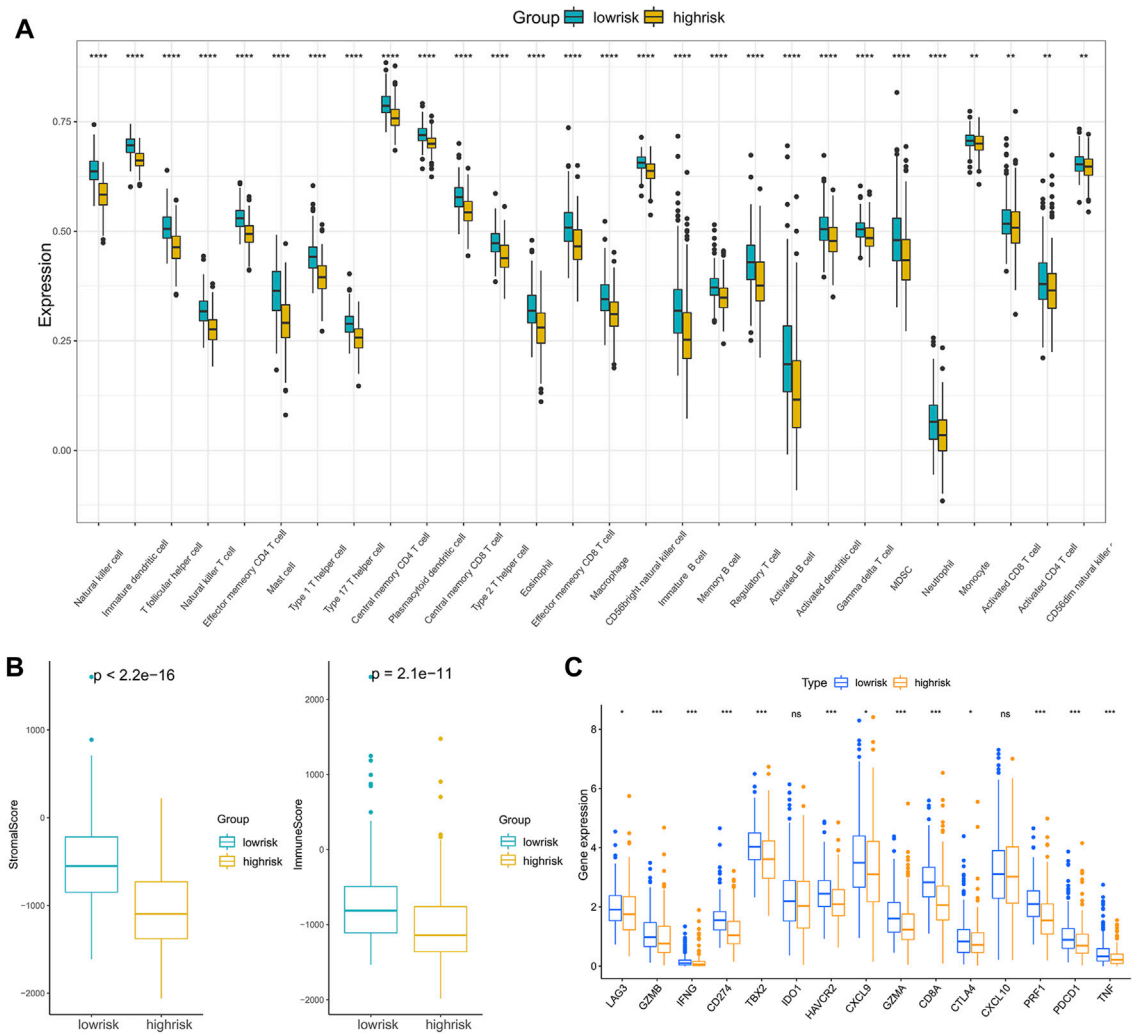


FIGURE 7 | Analysis of the correlation between the riskscore and immune cell infiltration: **(A)** The immune cell infiltration levels between the high- and low-risk subgroups. **(B)** ImmuneScore and StromalScore between the high- and low-risk subgroups. **(C)** The expression of immune checkpoint genes in the high- and low-risk subgroups. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

risk patients. We evaluated the differences in riskscore among patients with different responses to immunotherapy. Patients with complete response (CR) had the lowest riskscore while patients performed progressive disease (PD) had the highest riskscore (**Figure 8B**). These results suggested that patients in the high-risk group may not have a satisfying response to immunotherapy.

Considering that androgen deprivation therapy and chemotherapy still play vital roles in the treatment of prostate cancer. We used the GDSC database to explore the association between the riskscore and drug sensitivity. The results revealed little difference in the predicted IC50 of Bicalutamide between the high- and low-risk groups (**Figures 8C,E**). However, both in the TCGA and GEO cohorts, significant differences were observed in the predicted IC50 of Docetaxel between the high- and low-risk groups (**Figures 8D,F**). The IC50 of Docetaxel was significantly lower in the high-risk group, suggesting that these patients are sensitive to Docetaxel.

DISCUSSION

Prostate cancer affects millions of men all over the world, and accounts for 7% of newly diagnosed cancers in men worldwide (Rebello et al., 2021). While the prognosis of localized PCa has a good 5-years survival rate, the 5-years survival rate of metastatic PCa decreases significantly to only 30% (Siegel et al., 2020). It is well-known that the growth and progression of prostate cancer are significantly influenced by androgen, and androgen deprivation is an effective treatment strategy which is widely used in clinical practice (Marques et al., 2005). However, among patients with metastatic disease, a substantial proportion will develop metastatic castration-resistant prostate cancer (mCRPC), which is not sensitive to androgen deprivation therapy. Therefore, the long-term prognosis for patients with mCRPC is extremely poor (Henríquez et al., 2021). On that account, it is critical to unearth the mechanism of metastasis of prostate

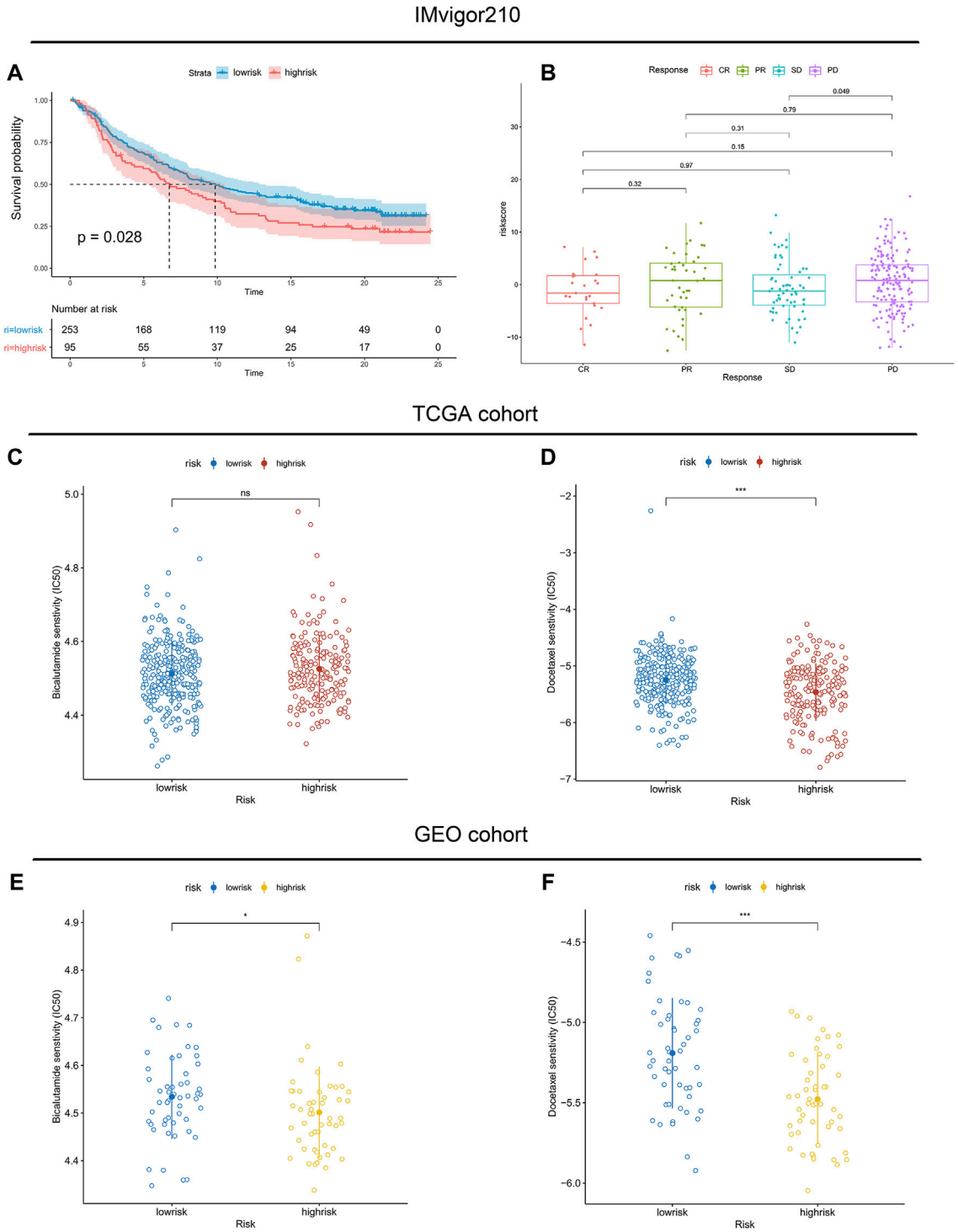


FIGURE 8 | Association between immunotherapy reactivity, drug sensitivity and riskscore. **(A)** Kaplan–Meier survival curve between high- and low-risk subgroups in the IMvigor210 cohort. **(B)** The difference in riskscore among patients with different responses to immunotherapy. **(C,E)** The predicted IC50 values of Bicalutamide in the TCGA cohort and validated GEO cohort. **(D,F)** The predicted IC50 values of Docetaxel in the TCGA cohort and validated GEO cohort.

cancer, which may assist us in predicting the prognosis of patients and forming a desirable therapeutic regimen. Although multitudinous studies have explored the correlation of some specific genes in tumor metastasis, few studies have focused on the overall mutated genes between primary and metastatic prostate cancer. Therefore, we used the COSMIC database,

which contains the most detailed resource of somatic mutations in human cancer, and executed the Chi-square test to obtain mutated genes with significant differences between primary and metastatic prostate cancer.

In this study, unsupervised clustering method was conducted, and two subclasses of PCa were obtained. Then, we comprehensively assessed the two clusters of prostate cancer and explored their biological characteristics. We observed that significant differences existed between the two clusters in some carcinogenic activation signaling pathways, such as the ErbB signaling pathway (Wang, 2017) and Vegf signaling pathway (Apte et al., 2019), and pathways associated with the immune response, such as the B cell receptor signaling pathway and T cell receptor signaling pathway. Moreover, the results indicated that all the immune cell infiltration levels and the expression of immune checkpoint genes were lower in Cluster 2, which was associated with poorer survival. We supposed that the poor prognosis of patients in Cluster 2 was due to tumor immune escape.

Next, we evaluated the riskscore of each patient by using the "Boruta" algorithm and PCA analysis. Its prognostic value was demonstrated both in TCGA and GEO cohorts. Since cancer develops as a result of somatic mutation and clonal selection (Martincorena et al., 2017), herein, we correlated riskscore and TMB and found a significant positive correlation between riskscore and TMB. Moreover, we assessed the mutation status of genes with high mutation frequencies in the high- and low-risk groups. It was observed that the high-risk groups contained more mutated samples and more mutation types.

Androgen deprivation therapy is a standard treatment used in all stages of recurrent prostate cancer. However, patients will develop CRPC eventually (Gamat and McNeel, 2017). In the past, the consensus was that immunotherapy might be ineffective in prostate cancer due to the immunosuppressive microenvironment (Chakravarty et al., 2020). However, with the recent development of advanced molecular diagnostic platforms, immunotherapy has revolutionized the treatment of prostate cancer and is re-emerging as a practicable option for patients, especially for CRPC (Cha et al., 2020). Nevertheless, a key challenge for immunotherapies is that these treatments have serious adverse effects, including autoimmunity and nonspecific inflammation (Riley et al., 2019). Additionally, many patients appear to have innate or acquired resistance to immunotherapies (O'Donnell et al., 2019). Therefore, it is critical to find reliable validated biomarkers to predict the immunotherapy responsiveness of patients. In fact, it is obvious that using a single biomarker to predict benefit from immunotherapy strategies is unstable. Consequently, we extracted 230 feature genes to construct the riskscore. According to the results, all immune cell infiltration levels were higher in the low-risk groups, and immune checkpoint genes used in immune checkpoint blockade therapy, such as PD-1, CTLA4 and PD-L1, were also more highly expressed in the low-risk groups. Therefore, we suppose that patients identified as having a low riskscore may benefit from the therapeutic strategy combining immune checkpoint blockade therapy, while patients are diagnosed with a high riskscore. Since open-access data of prostate cancer cohorts accepting immunotherapy are rare, we used patients in the IMvigor210 cohort for preliminary validation. We observed that patients who reacted as complete response had the lowest riskscore while patients performed progressive disease had the highest riskscore, which is consistent with the trends of expression of immune checkpoints.

Considering it is unrealistic that utilizing immunotherapy alone can dramatically change the outcome of prostate cancer right now, the combination of conventional cytotoxic agents, androgen deprivation therapy and personalized immunotherapy is more appropriate for patients. We used the GDSC database, which is the largest public resource for information on drug sensitivity in cancer cells (Yang et al., 2013), to predict the IC50 values of drugs for treating prostate cancer. We observed that the high-risk group was more sensitive to Docetaxel than the low-risk group. In fact, Docetaxel was the first systemic therapy to demonstrate survival benefit in mCRPC and became the standard of care for mCRPC in 2004 (Teo et al., 2019). It is still recommended as a first-line treatment for mCRPC in the latest EAU guidelines (Cornford et al., 2021). Therefore, it seems reasonable that patients identified with high riskscores had a higher sensitivity to Docetaxel. Bicalutamide is a competitive androgen receptor antagonist that leads to prostate cell apoptosis and the inhibition of prostate cancer growth (Wellington and Keam, 2006). We also evaluated the IC50 values of Bicalutamide in the high- and low-risk groups. However, there were few differences in the predicted IC50 of Bicalutamide between the high- and low-risk groups. Regrettably, limited by the data currently available in the GDSC database, we could not evaluate the IC50 values of Abiraterone and Enzalutamide in this study. As second-generation androgen receptor inhibitors, they have already been recommended by the latest EAU guidelines (Cornford et al., 2021). Numerous studies have demonstrated that the second-generation androgen receptor inhibitor is associated with improved outcomes compared with bicalutamide in CRPC (Penson et al., 2016; Naiki et al., 2021; Ueda et al., 2021; Vaishampayan et al., 2021). Therefore, the drug sensitivity of Abiraterone and Enzalutamide between high- and low-risk groups needs to be further explored. Moreover, due to the lack of available immunotherapy cohorts of prostate cancer, preliminary validation was performed in the IMvigor210 cohort for bladder cancer. The ability of the riskscore to predict the immunotherapy response of patients still needs further validation in immunotherapy cohorts of prostate cancer.

CONCLUSION

In summary, we selected mutated genes with significant differences between primary and metastatic prostate cancer from the COSMIC database and identified two PCa clusters that exhibited different prognostic outcomes and immune characteristics. For a better application in clinical practice, we constructed a scoring system and evaluated the prognostic value of the riskscore and its potential to aid treatment decisions. The riskscore could be applied to predict patients' response to immunotherapy and sensitivity to Docetaxel. The results suggested that immunotherapy may benefit patients in the

low-risk group, while Docetaxel is more effective for patients identified in the high-risk group.

WEBSITE

TCGA database: https://portal.gdc.cancer.gov/; GEO database: https://www.ncbi.nlm.nih.gov/geo/; COSMIC database: https://cancer.sanger.ac.uk/cosmic; MSigDB database: https://www.gsea-msigdb.org/gsea/msigdb/; GDSC database: https://www.cancerrxgene.org/.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

QL proposed the conception and designed the roadmap of this research. JM provided administrative support. XX, BC, GS, and KZ performed the statistical analyses and interpreted the data. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (grant number 81902619) and the National Natural Science Foundation of Hubei Province (2020CFB591).

ACKNOWLEDGMENTS

We thank TCGA (https://portal.gdc.cancer.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/) databases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.877086/full#supplementary-material

REFERENCES

Apte, R. S., Chen, D. S., and Ferrara, N. (2019). VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell* 176, 1248–1264. doi:10.1016/j.cell.2019.01.021

Arredouani, M. S., Lu, B., Bhasin, M., Eljanne, M., Yue, W., Mosquera, J.-M., et al. (2009). Identification of the Transcription Factor Single-Minded Homologue 2 as a Potential Biomarker and Immunotherapy Target in Prostate Cancer. *Clin. Cancer Res.* 15, 5794–5802. doi:10.1158/1078-0432.CCR-09-0911

Blomberg, O. S., Spagnuolo, L., and de Visser, K. E. (2018). Immune Regulation of Metastasis: Mechanistic Insights and Therapeutic Opportunities. *Dis. Model. Mech.* 11, dmm036236. doi:10.1242/dmm.036236

Brassart-Pasco, S., Brézillon, S., Brassart, B., Ramont, L., Oudart, J.-B., and Monboisse, J. C. (2020). Tumor Microenvironment: Extracellular Matrix Alterations Influence Tumor Progression. *Front. Oncol.* 10, 397. doi:10.3389/fonc.2020.00397

Cha, H.-R., Lee, J. H., and Ponnazhagan, S. (2020). Revisiting Immunotherapy: A Focus on Prostate Cancer. *Cancer Res.* 80, 1615–1623. doi:10.1158/0008-5472.CAN-19-2948

Chakravarty, D., Huang, L., Kahn, M., and Tewari, A. K. (2020). Immunotherapy for Metastatic Prostate Cancer: Current and Emerging Treatment Options. *Urol. Clin. North Am.* 47, 487–510. doi:10.1016/j.ucl.2020.07.010

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30, 44–56. doi:10.1093/annonc/mdy495

Cornford, P., van den Bergh, R. C. N., Briers, E., Van den Broeck, T., Cumberbatch, M. G., De Santis, M., et al. (2021). EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. Part II-2020 Update: Treatment of Relapsing and Metastatic Prostate Cancer. *Eur. Urol.* 79, 263–282. doi:10.1016/j.eururo.2020.09.046

Cox, T. R., and Erler, J. T. (2011). Remodeling and Homeostasis of the Extracellular Matrix: Implications for Fibrotic Diseases and Cancer. *Dis. Model. Mech.* 4, 165–178. doi:10.1242/dmm.004077

Derosa, C. A., Furusato, B., Shaheduzzaman, S., Srikantan, V., Wang, Z., Chen, Y., et al. (2012). Elevated Osteonectin/SPARC Expression in Primary Prostate Cancer Predicts Metastatic Progression. *Prostate Cancer Prostatic Dis.* 15, 150–156. doi:10.1038/pcan.2011.61

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., et al. (2017). COSMIC: Somatic Cancer Genetics at High-Resolution. *Nucleic Acids Res.* 45, D777–D783. doi:10.1093/nar/gkw1121

Gamat, M., and McNeel, D. G. (2017). Androgen Deprivation and Immunotherapy for the Treatment of Prostate Cancer. *Endocr. Relat. Cancer* 24, T297–T310. doi:10.1530/ERC-17-0145

Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., et al. (2019). Immunotherapy in Colorectal Cancer: Rationale, Challenges and Potential. *Nat. Rev. Gastroenterol. Hepatol.* 16, 361–375. doi:10.1038/s41575-019-0126-x

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics* 20, 307–315. doi:10.1093/bioinformatics/btg405

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9, e107468. doi:10.1371/journal.pone.0107468

Ghajar, C. M., and Bissell, M. J. (2008). Extracellular Matrix Control of Mammary Gland Morphogenesis and Tumorigenesis: Insights from Imaging. *Histochem. Cel Biol.* 130, 1105–1118. doi:10.1007/s00418-008-0537-1

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Henríquez, I., Roach, M., Morgan, T. M., Bossi, A., Gómez, J. A., Abuchaibe, O., et al. (2021). Current and Emerging Therapies for Metastatic Castration-Resistant Prostate Cancer (mCRPC). *Biomedicines* 9, 1247. doi:10.3390/biomedicines9091247

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28, 882–883. doi:10.1093/bioinformatics/bts034

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48, W509–W514. doi:10.1093/nar/gkaa407

López-Soto, A., Gonzalez, S., Smyth, M. J., and Galluzzi, L. (2017). Control of Metastasis by NK Cells. *Cancer Cell* 32, 135–154. doi:10.1016/j.ccell.2017.06.009

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554, 544–548. doi:10.1038/nature25501

Marques, R. B., Erkens-Schulze, S., de Ridder, C. M., Hermans, K. G., Waltering, K., Visakorpi, T., et al. (2005). Androgen Receptor Modifications in Prostate Cancer Cells upon Long-Termandrogen Ablation and Antiandrogen Treatment. *Int. J. Cancer* 117, 221–229. doi:10.1002/ijc.21201

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., et al. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 171, 1029–1041. doi:10.1016/j.cell.2017.09.042

Meller, S., Meyer, H.-A., Bethan, B., Dietrich, D., Maldonado, S. G., Lein, M., et al. (2016). Integration of Tissue Metabolomics, Transcriptomics and Immunohistochemistry Reveals ERG- and gleason Score-specific Metabolomic Alterations in Prostate Cancer. *Oncotarget* 7, 1421–1438. doi:10.18632/oncotarget.6370

Merino, D. M., McShane, L. M., Fabrizio, D., Funari, V., Chen, S.-J., White, J. R., et al. (2020). Establishing Guidelines to Harmonize Tumor Mutational burden (TMB): In Silico Assessment of Variation in TMB Quantification across Diagnostic Platforms: Phase I of the Friends of Cancer Research TMB Harmonization Project. *J. Immunother. Cancer* 8, e000147. doi:10.1136/jitc-2019-000147

Mortensen, M. M., Høyer, S., Lynnerup, A.-S., Ørntoft, T. F., Sørensen, K. D., Borre, M., et al. (2015). Expression Profiling of Prostate Cancer Tissue Delineates Genes Associated with Recurrence after Prostatectomy. *Sci. Rep.* 5, 16018. doi:10.1038/srep16018

Naiki, T., Takahara, K., Ito, T., Nakane, K., Sugiyama, Y., Koie, T., et al. (2021). Comparison of Clinical Outcomes between Androgen Deprivation Therapy with Up-Front Abiraterone and Bicalutamide for Japanese Patients with LATITUDE High-Risk Prostate Cancer in a Real-World Retrospective Analysis. *Int. J. Clin. Oncol.* 27, 592–601. doi:10.1007/s10147-021-02071-y

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer Immunoediting and Resistance to T Cell-Based Immunotherapy. *Nat. Rev. Clin. Oncol.* 16, 151–167. doi:10.1038/s41571-018-0142-8

Penson, D. F., Armstrong, A. J., Concepcion, R., Agarwal, N., Olsson, C., Karsh, L., et al. (2016). Enzalutamide versus Bicalutamide in Castration-Resistant Prostate Cancer: The STRIVE Trial. *J. Clin. Oncol.* 34, 2098–2106. doi:10.1200/JCO.2015.64.9285

Rebello, R. J., Oing, C., Knudsen, K. E., Loeb, S., Johnson, D. C., Reiter, R. E., et al. (2021). Prostate Cancer. *Nat. Rev. Dis. Primers* 7, 9. doi:10.1038/s41572-020-00243-0

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery Technologies for Cancer Immunotherapy. *Nat. Rev. Drug Discov.* 18, 175–196. doi:10.1038/s41573-018-0006-z

Rycaj, K., Li, H., Zhou, J., Chen, X., and Tang, D. G. (2017). Cellular Determinants and Microenvironmental Regulation of Prostate Cancer Metastasis. *Semin. Cancer Biol.* 44, 83–97. doi:10.1016/j.semcancer.2017.03.009

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 47, D941–D947. doi:10.1093/nar/gky1015

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., et al. (2010). Integrative Genomic Profiling of Human Prostate Cancer. *Cancer Cell* 18, 11–22. doi:10.1016/j.ccr.2010.05.026

Teo, M. Y., Rathkopf, D. E., and Kantoff, P. (2019). Treatment of Advanced Prostate Cancer. *Annu. Rev. Med.* 70, 479–499. doi:10.1146/annurev-med-051517-011947

Ueda, T., Shiraishi, T., Ito, S., Ohashi, M., Matsugasumi, T., Yamada, Y., et al. (2021). Abiraterone Acetate versus Bicalutamide in Combination with Gonadotropin Releasing Hormone Antagonist Therapy for High Risk Metastatic Hormone Sensitive Prostate Cancer. *Sci. Rep.* 11, 10094. doi:10.1038/s41598-021-89609-2

Vaishampayan, U. N., Heilbrun, L. K., Monk, P., Tejwani, S., Sonpavde, G., Hwang, C., et al. (2021). Clinical Efficacy of Enzalutamide vs Bicalutamide Combined with Androgen Deprivation Therapy in Men with Metastatic Hormone-Sensitive Prostate Cancer. *JAMA Netw. Open* 4, e2034633. doi:10.1001/jamanetworkopen.2020.34633

van den Bulk, J., Verdegaal, E. M., and de Miranda, N. F. (2018). Cancer Immunotherapy: Broadening the Scope of Targetable Tumours. *Open Biol.* 8, 180037. doi:10.1098/rsob.180037

Wang, G., Zhao, D., Spring, D. J., and DePinho, R. A. (2018). Genetics and Biology of Prostate Cancer. *Genes Dev.* 32, 1105–1140. doi:10.1101/gad.315739.118

Wang, Z. (2017). ErbB Receptors and Cancer. *Methods Mol. Biol.* 1652, 3–35. doi:10.1007/978-1-4939-7219-7_1

Wellington, K., and Keam, S. J. (2006). Bicalutamide 150mg. *Drugs* 66, 837–850. doi:10.2165/00003495-200666060-00007

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, T., and Dai, Y. (2017). Tumor Microenvironment and Therapeutic Response. *Cancer Lett.* 387, 61–68. doi:10.1016/j.canlet.2016.01.043

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Xiao, Chen, Song, Zeng and Miao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fgene.2022.825261

Construction of a Necroptosis-Related miRNA Signature for Predicting the Prognosis of Patients With Hepatocellular Carcinoma

Tongyu Meng, Qingfeng Wang, Yufeng Yang, Yanling Ren* and Yan Shi**

Liaoning University of Traditional Chinese Medicine, Shenyang, China

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Zhiyun Guo,
Southwest Jiaotong University, China
Zaoqu Liu,
First Affiliated Hospital of Zhengzhou University, China

***Correspondence:**
Yufeng Yang
yyf701@163.com
Yanling Ren
lnutcmren@163.com
Yan Shi
lntcmshiyan@163.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *30 November 2021*
Accepted: *14 March 2022*
Published: *12 April 2022*

Citation:
Meng T, Wang Q, Yang Y, Ren Y and Shi Y (2022) Construction of a Necroptosis-Related miRNA Signature for Predicting the Prognosis of Patients With Hepatocellular Carcinoma. Front. Genet. 13:825261. doi: 10.3389/fgene.2022.825261

Background: Many miRNAs have been demonstrated to be associated with the prognosis of hepatocellular carcinoma (HCC). However, how to combine necroptosis-related miRNAs to achieve the best predictive effect in estimating HCC patient survival has not been explored.

Methods: The mRNA and miRNA expression profile were downloaded from a public database (TCGA-LIHC cohort). Necroptosis-related genes were obtained from previous references, and necroptosis-related miRNAs were identified using Pearson analysis. Subsequently, differential expression miRNAs (DEms) were identified in HCC and paracancer normal samples based on necroptosis-related miRNA expression. The whole set with HCC was randomized into a training set and testing set (1:1). LASSO-Cox regression analysis was used to construct an miRNA signature. Multiple statistical methods were used to validate the clinical benefit of signature in HCC patients, including receiver operator characteristic (ROC) curves, Kaplan–Meier survival analyses, and decision curve analysis (DCA). The downstream target genes of miRNAs were obtained from different online tools, and the potential pathways involved in miRNAs were explored. Finally, we conducted RT-qPCR in SK-HEP-1, THLE-3, and HUH-7 cell lines for miRNAs involved in the signature.

Results: The results showed that a total of eight specific necroptosis-related miRNAs were screened between HCC and adjacent tissues in the training set. Subsequently, based on the aforementioned miRNAs, 5-miRNA signature (miR-139-5p, hsa-miR-326, miR-10b-5p, miR-500a-3p, and miR-592) was generated by LASSO-Cox regression analysis. Multivariate Cox regression analysis showed that the risk scores were independent prognostic indicators in each set. The area under curves (AUCs) of 1 year, 3 years, 5 years, and 7 years were high in each set (AUC >0.7). DCA analysis also revealed that the risk score had a potential benefit than other clinical characteristics. Meanwhile, survival analysis showed that the high-risk group showed low survival probabilities. Moreover, the results of enrichment analysis showed that specific miRNAs were mainly enriched in the cAMP signaling pathway and TNF signaling

pathway. Finally, the results of RT-qPCR were consistent with the prediction results in public databases.

Conclusion: Our study establishes a robust tool based on 5-necroptosis-related miRNAs for the prognostic management of HCC patients.

Keywords: hepatocellular carcinoma, miRNAs, signature, necroptosis, prognosis

INTRODUCTION

By 2025, more than 1 million new cases of hepatocellular carcinoma (HCC) will be diagnosed each year worldwide, a serious situation that will pose a major challenge to global healthcare; more importantly, the 5-year survival rate for HCC patients has decreased by 20% globally and as low as 12% in Asian countries such as Japan. (Llovet et al., 2021). Smoking, drinking, and viral infection are risk factors for the occurrence and prognosis of HCC (El-Serag, 2011). In addition, local recurrence and metastasis reduce the survival rate in HCC patients (Di Sandro et al., 2019). Therefore, intricate etiological factors and heterogeneity of HCC make prognostic prediction challenging. There is an urgent need to develop a new prognostic model considering the limitations of HCC treatment strategies.

Necroptosis is a cell death independent of caspase (Linkermann and Green, 2014) including the following characteristics: incomplete cell membrane, intracellular metabolic abnormalities, and release of inflammatory factors. (Christofferson and Yuan, 2010). It is worth noting that necroptosis plays an important role in the occurrence and development of various diseases such as neurodegenerative diseases (Dionísio et al., 2020), ischemic cardiovascular (Ruan et al., 2019), and cancers (Gong et al., 2019). Interestingly, necroptosis has been shown to play a dual role in cancer. On the one hand, the hub regulators of necroptosis can promote metastasis and progression of cancer (Strilic et al., 2016); on the other hand, necroptosis can also prevent tumor development when apoptosis function in cancer cells is impaired (Long and Ryan, 2012). Meanwhile, necroptosis is regulated by intracellular signaling factors such as tumor necrosis receptor factor (TNFR) (Lalaoui et al., 2015), pattern recognition receptors (PRRs) (Li et al., 2012), and T-cell receptors (TCRs) (Hitomi et al., 2008). At present, relevant reports have revealed the regulation mechanism of necroptosis in HC. Xiang et al. (2021) discovered how CX32 induces necroptosis, and high CX32 expression may be represented as a resistant role to apoptosis inducers. Meanwhile, mosaic mouse models were used to reveal how necroptosis microenvironment directs lineage commitment in liver cancer (Seehawer et al., 2018). In addition, it is reported that some compounds can induce necroptosis to treat cancer. These evidence remind us that necroptosis has a potential application value in the treatment of HCC. In addition, microRNAs (miRNAs) are small molecules encoded by the genomes of eukaryotes, similar to siRNA (Bartel, 2004). There have been many studies on miRNA in HCC, such as the one where Wang et al. (2017a) found that miR-218 can suppress the metastasis and EMT of HCC cells *via* targeting SERBP1. Moreover, different noncoding RNAs also interact with each other in the progression of HCC. Ding et al. (2020) revealed that hsa_circ_0001955 enhances proliferation, migration, and invasion of HCC cells through miR-145-5p/NRAS axis. Surprisingly, a combination of miRNA signatures in predicting HCC survival was elucidated by Li et al. (2020). In addition, enhancers can act as tissue-specific

TABLE 1 | Clinicopathological features of TCGA-LIHC.

Variables	Count	Percentage (%)
Age (mean ± SD)	59.57 ± 13.33	
Status		
Alive	249	66.05
Dead	128	33.95
Gender		
Male	255	67.64
Female	122	32.36
Pathological stage		
Stage I	175	46.42
Stage II	87	23.08
Stage III	3	0.80
Stage IIIA	65	17.24
Stage IIIB	9	2.39
Stage IIIC	9	2.39
Stage IV	2	0.53
Stage IVA	1	0.27
Stage IVB	2	0.53
Unknown	24	6.37
T staging		
T1	185	49.07
T2	93	24.67
T2a	1	0.27
T2b	1	0.27
T3	45	11.94
T3a	29	7.69
T3b	7	1.86
T4	13	3.45
TX	1	0.27
Unknown	2	0.53
N staging		
N0	257	68.17
N1	4	1.06
NX	115	30.50
Unknown	1	0.27
M staging		
M0	272	72.15
M1	4	1.06
MX	101	26.79
Grade		
G1	55	14.59
G2	180	47.75
G3	124	32.89
G4	13	3.45
Unknown	5	1.33

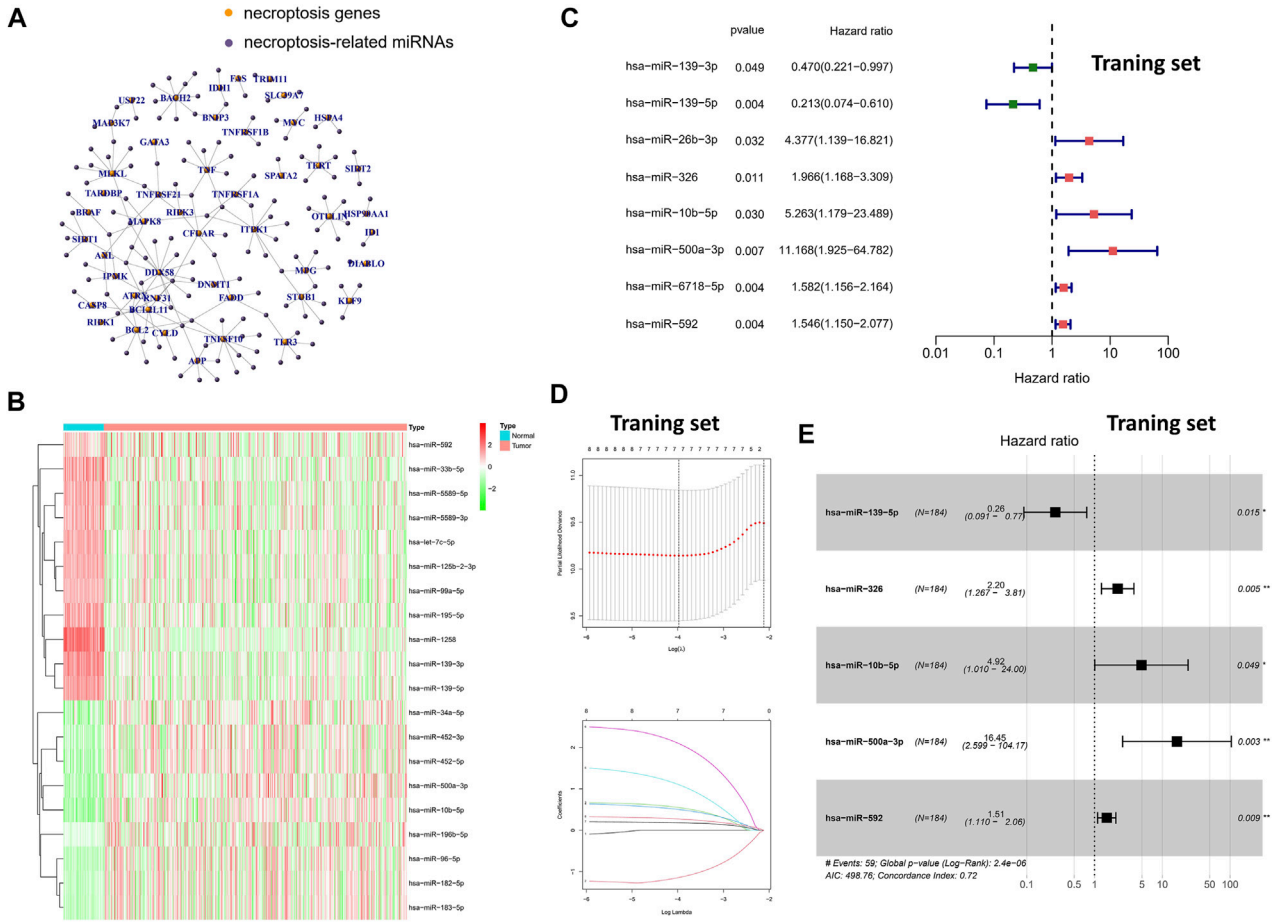


FIGURE 1 | Screening of specific necroptosis-related miRNAs. **(A)** Pearson correlation analysis on 2435 miRNAs and 67 necroptosis-related genes. **(B)** The heatmap of DEms in different tissues samples. **(C)** Univariate Cox regression analysis of 35-miRNAs (only the 8 miRNAs with statistical significance were shown). **(D)** LASSO regression analysis. **(E)** Multivariate Cox regression analysis in 8 prognostic miRNAs.

TABLE 2 | Results of multivariate Cox regression in 5-necroptosis-related miRNAs.

miRNAs	Coef	HR	HR.95L	HR.95H	*p*
hsa-miR-139-5p	−1.3322	0.2639	0.0906	0.7687	0.0146
hsa-miR-326	0.7874	2.1976	1.2674	3.8105	0.0051
hsa-miR-10b-5p	1.5938	4.9225	1.0097	23.9995	0.0486
hsa-miR-500a-3p	2.8005	16.4536	2.5987	104.1749	0.0029
hsa-miR-592	0.4140	1.5128	1.1095	2.0627	0.0089

cis-regulatory elements to positively regulate gene expression by miRNAs (Zhang et al., 2020).

However, no study has systematically used necroptosis-related miRNA to predict the prognosis of HC patients.

In order to solve the aforementioned problem, we downloaded expression data and clinical features of TCGA-LIHC cohort from public database and extracted miRNA data related to necroptosis. Then, a 5-miRNA prognosis signature was constructed by LASOO-Cox regression analysis, and its prognostic ability was verified in different cohorts. The functional enrichment analysis of downstream genes of miRNAs is performed to explore the potential mechanism. Finally, we conducted RT-qPCR in SK-HEP-1, THLE-3, and HUH-7 cell lines for 5-miRNAs involved in the signature.

MATERIALS AND METHODS

Datasets and Pre-Processing

We downloaded the clinical features and miRNA expression of HC patients from The Cancer Genome Atlas (TCGA) database (Wang et al., 2016) *via* the R package "TCGAbiolinks". The miRNA expression profile in TCGA-LIHC includes 50 normal samples and 375 tumor samples (**Table 1**). Based on previous studies (Liu et al., 2021a; Visalli et al., 2018), we obtained 16 miRNAs associated with necroptosis, which were miR-495, miR-331-3p, miR-15a, miR-148a-3p, miR-7-5p, miR-141-3p, miR-425-5p, miR-200a-5p, miR-210, miR-223-3p, miR-500a-3p, miR-181-5p, miR-16-5p, hsa-miR-371-5p, hsa-miR-373, and hsa-miR-543 in detail. In addition, 67 necroptosis-related genes were extracted based on previous bioinformatics research (Hu et al., 2022; Huang et al.,

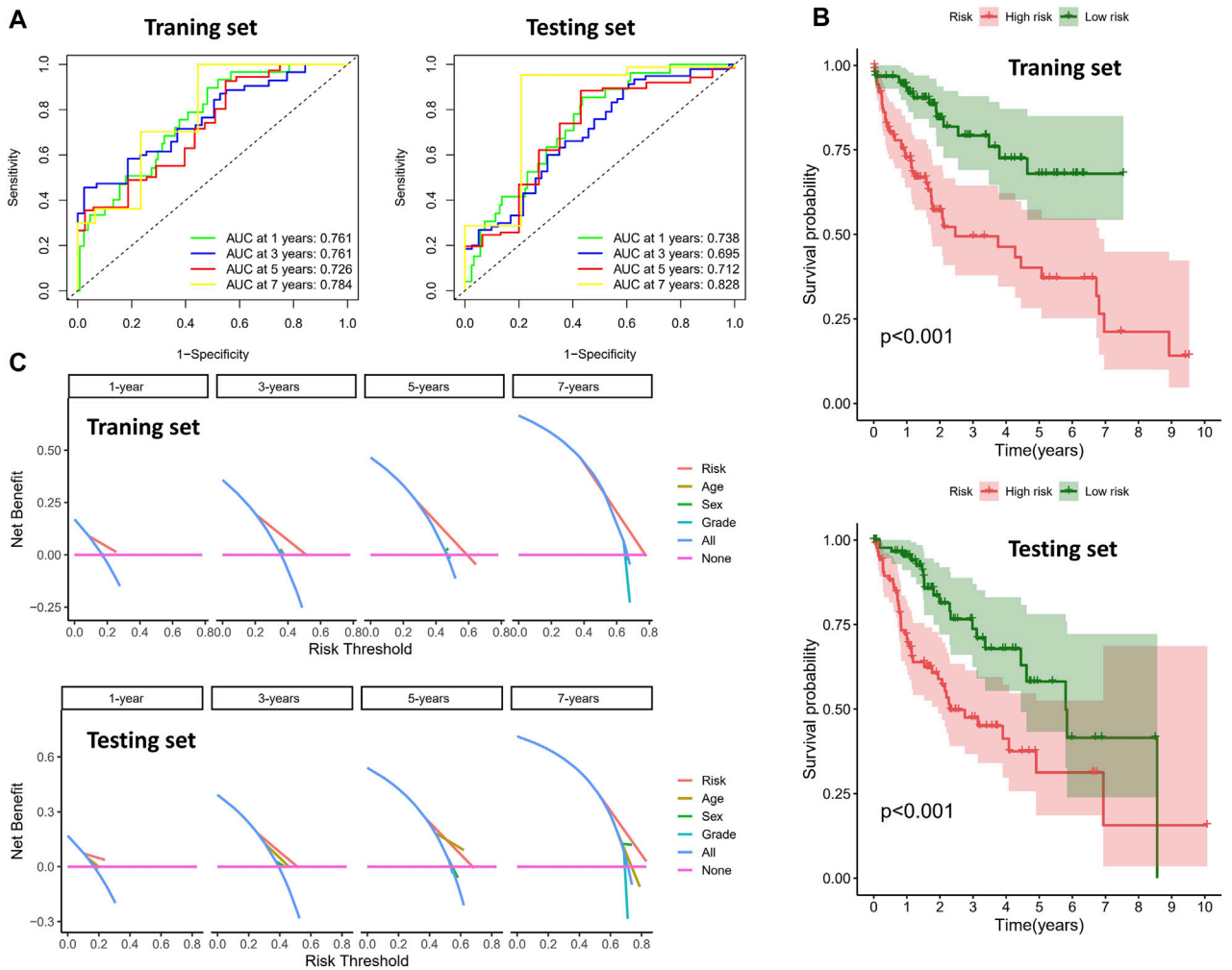


FIGURE 2 | Clinical benefits using novel signature. **(A)** ROC analysis in different sets (the left figure is the training set and the right figure is the testing set). **(B)** Survival analysis of different risk groups (the top figure is the training set and the bottom figure is the testing set). **(C)** DCA analysis in different sets (the top figure is the training set and the bottom figure is the testing set).

2021). We performed Pearson correlation analysis (|cor|>0.15, p < 0.05) on 2435 miRNAs and 67 necroptosis-related genes from the raw data [RNA-seq (log FPKM+1 format), miRNA-seq (log RPM +1 format)]. Ultimately, we annotated 144 necroptosis-related miRNAs in TCGA-LIHC from the references and Pearson correlation analysis.

Calculation of the Risk Score *via* the Necroptosis-Related MicroRNA Signature

The differentially expressed miRNAs (DEms) in normal samples and tumor samples were screened by using "limma" package (Ritchie et al., 2015) in R software. The thresholds in 'limma' package were set to adjusted p-value < 0.05, and |logFCfilter| = 1. At a ratio of 1:1, the whole set was divided into two using the 'caret' package in R software. We performed LASSO-Cox regression analysis in the training set (p < 0.05), and the risk score was calculated as follows:

$$risk\ score = \sum_{i=1}^{n} Coef_i * x_i,$$

where $Coef$ is the coefficient, and x is the expression value of each selected miRNA. LASSO is a popular algorithm which was extensively utilized in medical studies (Liu et al., 2022a), (Liu et al., 2022b), (Liu et al., 2022c), (Liu et al., 2022d). The risk signature for predicting survival was assessed by the AUC value. We calculated the median score of risk score, which is used to select high-risk and low-risk groups in each sets. Kaplan–Meier survival analysis suggested the difference in high-risk and low-risk groups. To better evaluate clinical applications, we calculated the net benefit rates of different clinical characteristics by the DCA curve analysis.

miRNA–mRNA Network

miRDB (Chen and Wang, 2020), TargetScan (Riffo-Campos et al., 2016), and MiTarBase (Huang et al., 2020) online tools were used to predict the downstream target genes of miRNAs

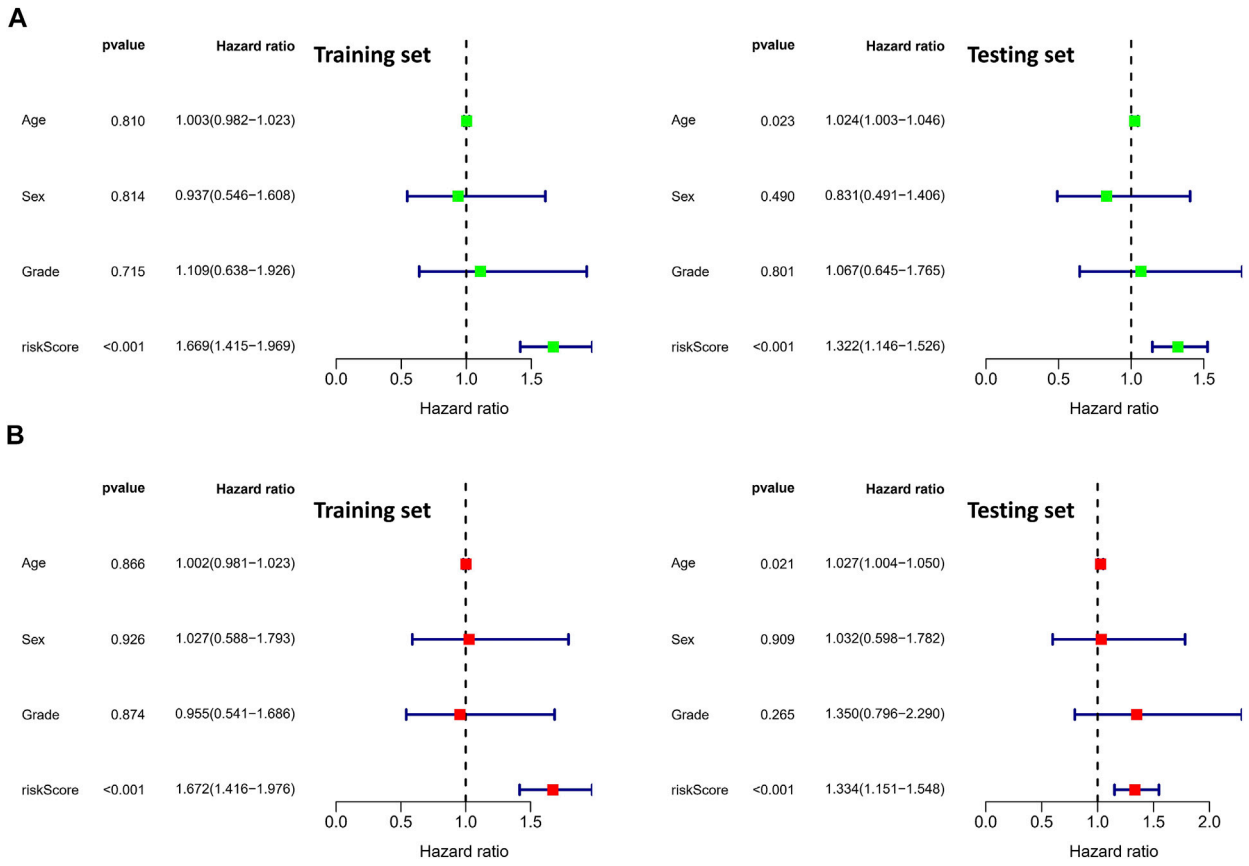


FIGURE 3 | Identification of independent prognostic factors. **(A)** Univariate Cox analysis in clinical features and risk score (the left figure is the training set and the right figure is the testing set). **(B)** Multivariate Cox analysis in clinical features and risk score (the left figure is the training set and the right figure is the testing set).

significantly related to prognosis, and only genes co-existing in three databases could be selected as the ultimate target genes. Finally, Cytoscape software was used to demonstrate the mRNA–target genes network.

Enrichment Analysis

GO enrichment analysis is a commonly used bioinformatics method, which is used to search for comprehensive information of large-scale genetic data. Meanwhile, the KEGG pathway enrichment analysis is widely used to understand biological mechanisms. The enrichment analysis was performed in downstream mRNAs by using "ggplot2" and "clusterProfiler" packages in R software (Yu et al., 2012).

Cell Culture and qRT-PCR

As seen in a previous study (Chen et al., 2020), HCC cell lines HUH-7 and SK-HEP-1 from humans and a normal human liver cell line THLE-3 were purchased from Shanghai Institute of Cell Biology and have been identified by STR genotyping test. These cells in a culture room (5% CO2 and 37°C) were cultured using RPMI-1640 medium with 10% FBS. We used miRcute miRNA Isolation Kit (Zhisheng, Nanjing, China) to isolate total miRNA. For miRNA, miRcute Plus miRNA First-Strand cDNA Synthesis kits (Zhisheng, Nanjing, China) were used for reverse transcription. The second step was completed using miRcute Plus miRNA qPCR Detection Kits (Zhisheng, Nanjing, China). Small RNA RNU6B (U6) (RiboBio, Guangzhou, China) was used as a control for the expression of miRNA (Li et al., 2022a). Primer sequences are summarized in previous studies, including miR-139-5p (Li et al., 2022b), hsa-miR-326 (Wei et al., 2022), miR-10b-5p (Niu et al., 2021), miR-500a-3p (Long et al., 2022), and miR-592 (Paul et al., 2021).

Statistical Analysis

All statistical analyses were performed using the R software (v.4.0.1). Detailed statistical methods about transcriptome data are covered in the bioinformatics method section. For the symbols, ***, **, *, and ns, refer to $p < 0.001$, <0.01, <0.05, and not significant, respectively.

RESULTS

Screening of Specific Necroptosis-Related MicroRNAs in Hepatocellular Carcinoma

Based on previous studies and Pearson correlation analysis, we obtained 144 miRNAs associated with necroptosis (**Figure 1A**). Subsequently, we performed differential expression analysis of

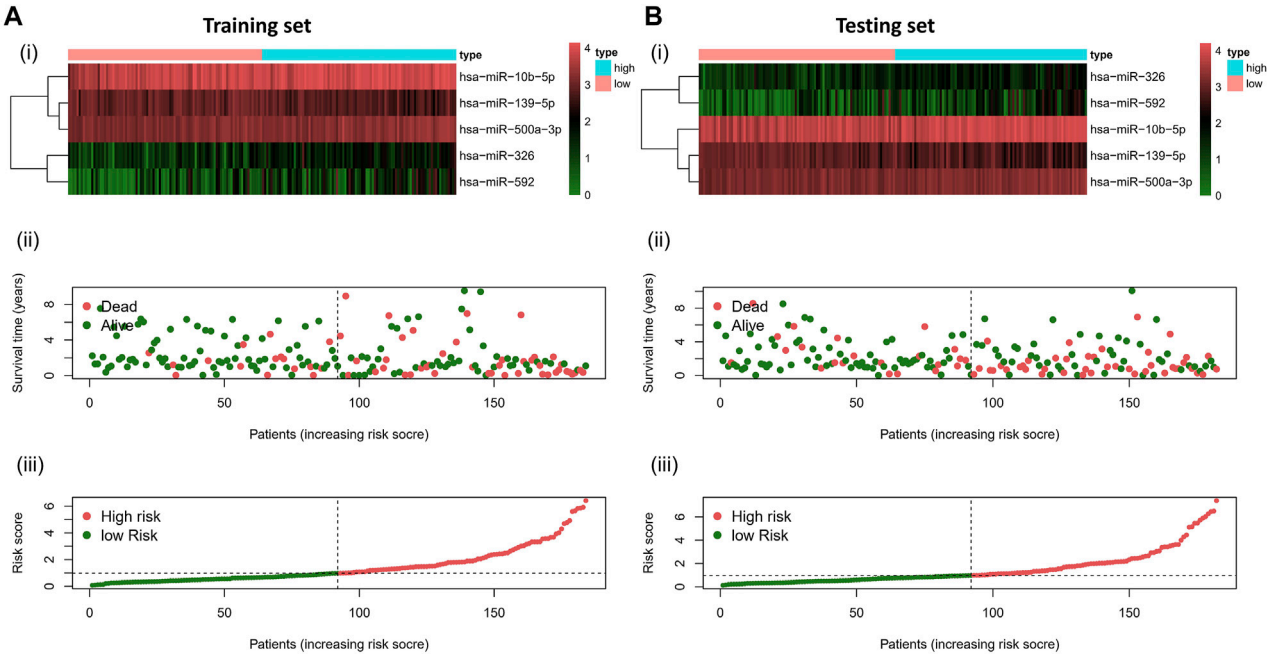


FIGURE 4 | Risk distribution of all patients. **(A)**Training sets **[(A), i]** Heatmap of different risk groups in 5-specific necroptosis-related miRNAs. **[(A), ii]** Risk score scatter plot. **[(A), iii]** Risk score curve plot. **(B)** Testing sets **[(B), i]** Heatmap of different risk group in 5-specific necroptosis-related miRNAs. **[(B), ii]** Risk score scatter plot. **[(A), iii]** Risk score curve plot. Patients are categorized into low-risk (green) and high-risk (red) groups.

the aforementioned miRNAs in the TCGA-LIHC cohort by "limma" package. Finally, a total of 35 necroptosis-related miRNAs were identified in 50 normal samples and 375 tumor samples (**Figure 1B**). Compared with normal tissues, the aforementioned 35-miRNAs were abnormally expressed in HCC tissues, indicating that these miRNAs were worthy of further exploration. Subsequently, univariate cox regression was performed for further screening of 35-miRNAs (**Figure 1C**), and LASSO regression analysis further eliminated the redundant genes in 8 prognostic miRNAs (**Figure 1D**). Finally, multivariate COX regression analysis was performed for the aforementioned eight prognostic miRNAs, and the minimum Akaike information criterion (AIC) value was reached when miR-139-5p, hsa-miR-326, miR-10b-5p, miR-500a-3p, and miR-592 were included in the regression equation (**Figure 1E**).

A Novel Signature Based on 5-Specific Necroptosis-Related miRNAs

We developed a novel signature based on the regression coefficients of 5-specific necroptosis-related miRNAs (**Table 2**), and all patients were divided into high-risk and low-risk groups according to the median value of the risk score in the training set. Hence, the formula of risk score = (−1.3322 × expression level of hsa-miR-139-5p) + (0.7874 × expression level of hsa-miR-326) + (1.5938 × expression level of hsa-miR-10b-5p) + (2.8005 × expression level of hsa-miR-500a-3p) + (−0.4140 × expression level of hsa-miR-592). It is worth mentioning that we also performed Kaplan–Meier analysis and log-rank test on 5-miRNAs in the whole set (**Supplementary Figure S1**). The results showed that low expression of the mir-139-5p group had less possibility of survival ($p < 0.05$). Taken together, our data showed that miR-139-5p, hsa-miR-326, miR-500a-3p, and miR-592 may have potential implications for the survival of HCC patients in the whole set.

Clinical Benefits of Using the Novel Signature in Patients With Hepatocellular Carcinoma

To validate the prognostic value of the risk score, we conducted survival analyses by plotting Kaplan–Meier curves and ROC analysis. As a result, the area under curves (AUCs) of 1 year, 3 years, 5 years, and 7 years were high in the training and testing sets (AUC = 0.761, 0.761, 0.726, and 0.784 in the training set; AUC = 0.738, 0.695, 0.712, and 0.828 in the testing set), as shown in **Figure 2A**. In addition, the survival rate can be separated between the high-risk and low-risk groups, and the high-risk group showed low survival probabilities ($p < 0.001$), as shown in **Figure 2B**. Interestingly, DCA analysis also revealed that the risk score had a net benefit value than other clinical characteristics (**Figure 2C**). To validate whether the risk score, patient age, tumor grade, and gender can function as independent predictors, we conducted univariate (**Figure 3A**) and multivariate Cox regression (**Figure 3B**) analyses for these clinical characteristics in the training set and testing set. The results showed that the risk score was an independent prognosis factor among the four clinical characteristics in multivariate analysis

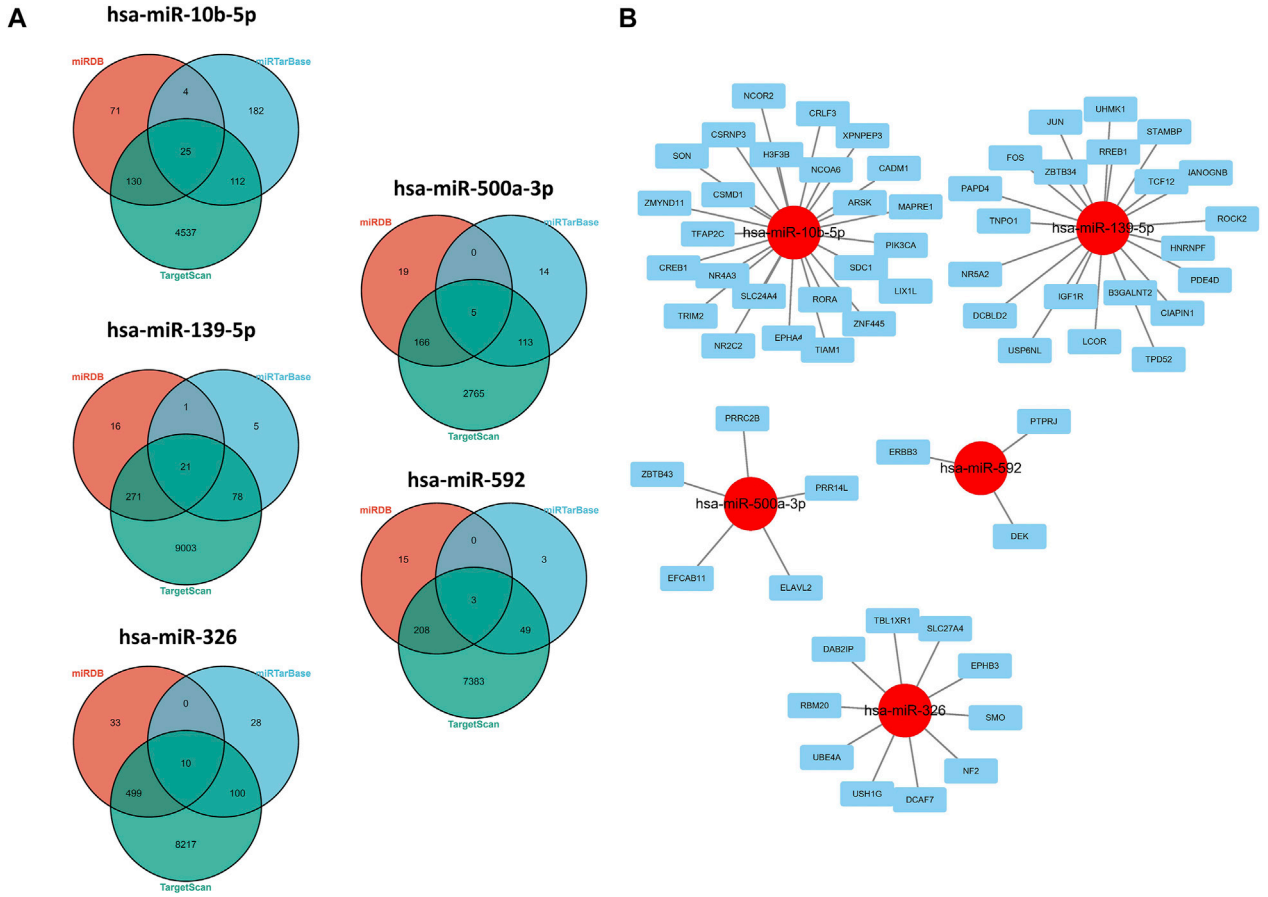


FIGURE 5 | Regulatory network of 5-miRNAs. **(A)** Venn plots in three databases. **(B)** Regulatory network conducted by Cytoscape software.

(HR = 1.672 in training set, HR = 1.334 in testing set). The heatmap showed the expression of 5-specific necroptosis-related miRNAs in different risk groups [**Figures 4Ai, Bi**]. In addition, as the risk score increased, the patient death risk increased and the survival time decreased [**Figure 4Aii,iii, Bii,iii**].

Construction of Regulatory Network

As shown in the results of the aforementioned section, miR-139-5p, hsa-miR-326, miR-10b-5p, miR-500a-3p, and miR-592 may have potential implications for survival of HCC patients, so we explored the potential regulatory axis using miRDB, TargetScan, and MiTarBase online tools. Finally, 21, 10, 25, 5, and 3 potential downstream targeted genes were identified in miR-139-5p, hsa-miR-326, miR-10b-5p, miR-500a-3p, and miR-592, respectively (**Figure 5A**). As shown in **Figure 5B**, we visualized the aforementioned network. More specifically, ARSK, CADM1, EPHA4, H3F3B, MAPRE1, NCOA6, NCOR2, NR2C2, NR4A3, PIK3CA, RORA, CREB1, CRLF3, CSMD1, CSRNP3, SLC24A4, SON, TFAP2C, TIAM1, TRIM2, XPNPEP3, ZMYND11, ZNF445, LIX1L, and SDC1 were the downstream targets of miR-10b-5p. ZBTB34, UHMK1, USP6NL, CIAPIN1, TPD52, IGF-1R, STAMBP, TCF12, DCBLD2, TNPO1, B3GALNT2, NANOGNB, NR5A2, PAPD4, PDE4D, ROCK2, RREB1, JUN, FOS, HNRNPF, and LCOR were the downstream targets of miR-139-5p. UBE4A, USH1G, DAB2IP, DCAF7, EPHB3, NF2, RBM20, SLC27A4, SMO, and TBL1XR1 were the downstream targets of miR-326. ZBTB43, PRR14L, PRRC2B, EFCAB11, and ELAVL2 were the downstream targets of miR-500a-3p. ERBB3, DEK, and PTPRJ were the downstream targets of miR-592.

Gene Enrichment Analysis

In order to further explore the biological functions involved in necroptosis-related miRNAs, we conducted an in-depth enrichment analysis. The results of GO enrichment analysis of the aforementioned targeted genes showed that they were mainly enriched in the regulation of neuron death, DNA-templated transcription, and gliogenesis. (**Figure 6A**). Meanwhile, KEGG enrichment further elucidated possible pathways such as proteoglycans in cancer, cAMP signaling pathway, and TNF signaling pathway. (**Figure 6B**).

qRT-PCR

The expression levels of the 5-miRNAs were determined using the TCGA database and qRT-PCR experiment. As shown in **Figure 7A**, the violin plot showed that tumor tissues had higher expression levels of miR-10b-5p and miR-500a-3p than normal

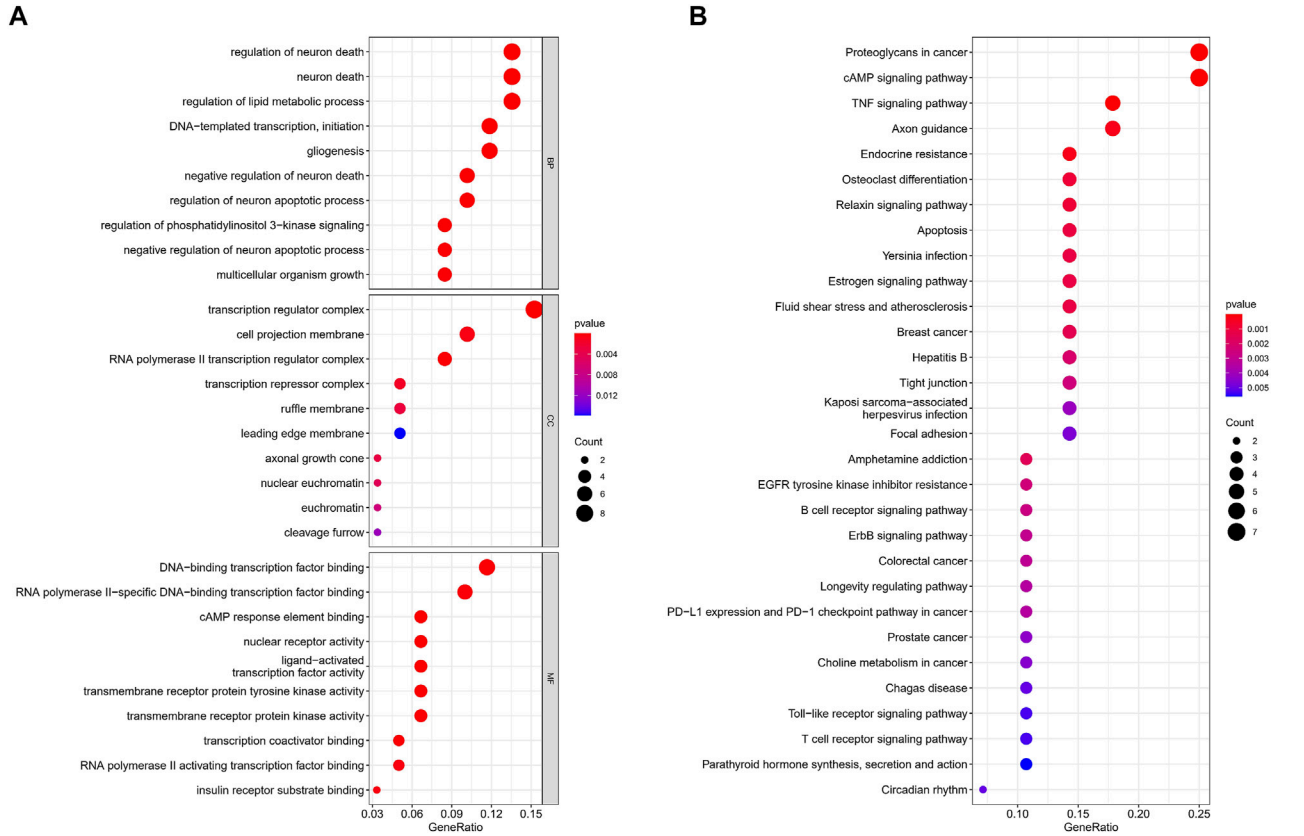


FIGURE 6 | GO and KEGG enrichment analysis. **(A)** GO analysis. **(B)** KEGG analysis.

tissues, while normal samples had higher expression levels of miR-139-5p, hsa-miR-326, and miR-592 compared to HCC samples. Subsequently, HCC cell lines (HUH-7 and SK-HEP-1) and a normal human liver cell line (THLE-3) were used to validate the results of tissue samples. In the validation of miR-139-5p (**Figure 7B**), hsa-miR-326 (**Figure 7C**), miR-10b-5p (**Figure 7D**), and miR-500a-3p (**Figure 7E**), the expression of HUH-7 was higher than that of THLE-3. However, there was no statistical difference in the expression of the aforementioned miRNAs in SK-HEP-1 vs. THLE-3. In the validation of miR-592, no significant differences in the expression of all three cell lines were observed (**Figure 7F**). Overall, most miRNA expressions were at the same level in cell lines and tissues, except for miR-592.

DISCUSSION

Although there are many studies on miRNA expression to predict the prognosis of patients, no study has systematically used necroptosis-related miRNA to predict the prognosis of HCC patients. To our knowledge, our study is the first to explore the clinical application of necroptosis-related miRNAs in an HCC cohort. In this study, a total of 5-miRNA signatures were generated by Cox regression. Subsequently, a series of statistical analyses demonstrated the excellent clinical application for the risk score calculated based on the risk signature. Overall, most miRNA expressions were at the same level in cell lines and tissues, except for miR-592. Moreover, we found that specific miRNAs were mainly enriched in the cAMP signaling pathway, TNF signaling pathway, and Wnt/β-catenin pathway. Fortunately, the role of the aforementioned pathways in the pathogenesis of HCC has been verified in *in vitro* experiments. Liu et al. (2021b) revealed that cellular retinol binding protein-1 inhibits cancer stemness *via* upregulating WIF1 to suppress the Wnt/β-catenin pathway in hepatocellular carcinoma. Meanwhile, Yassin et al. (2022) explored silybum marianum total extract, silymarin and silibinin abate hepatocarcinogenesis, and hepatocellular carcinoma growth *via* modulation of the TNF signaling pathway. In addition, activation of the cAMP signaling pathway in HCC also appeared to be associated with epigenetic modifications, as in Cai et al.'s (2021) study that RBM15-mediated m6A modification might facilitate the progression of HCC *via* the IGF2BP1-YES1-MAPK axis. This seems to suggest that there is cross-talk in the pathways that promote the development of HCC. Thus, this study lays the foundation for our future *in vivo* and *in vitro* experiments.

Most of the miRNAs participating in signatures have been revealed to be associated with cancer progression. In miR-139-5p, it suppressed the proliferation and migration of hepatocellular carcinoma cells by downregulating the expression of ENAH

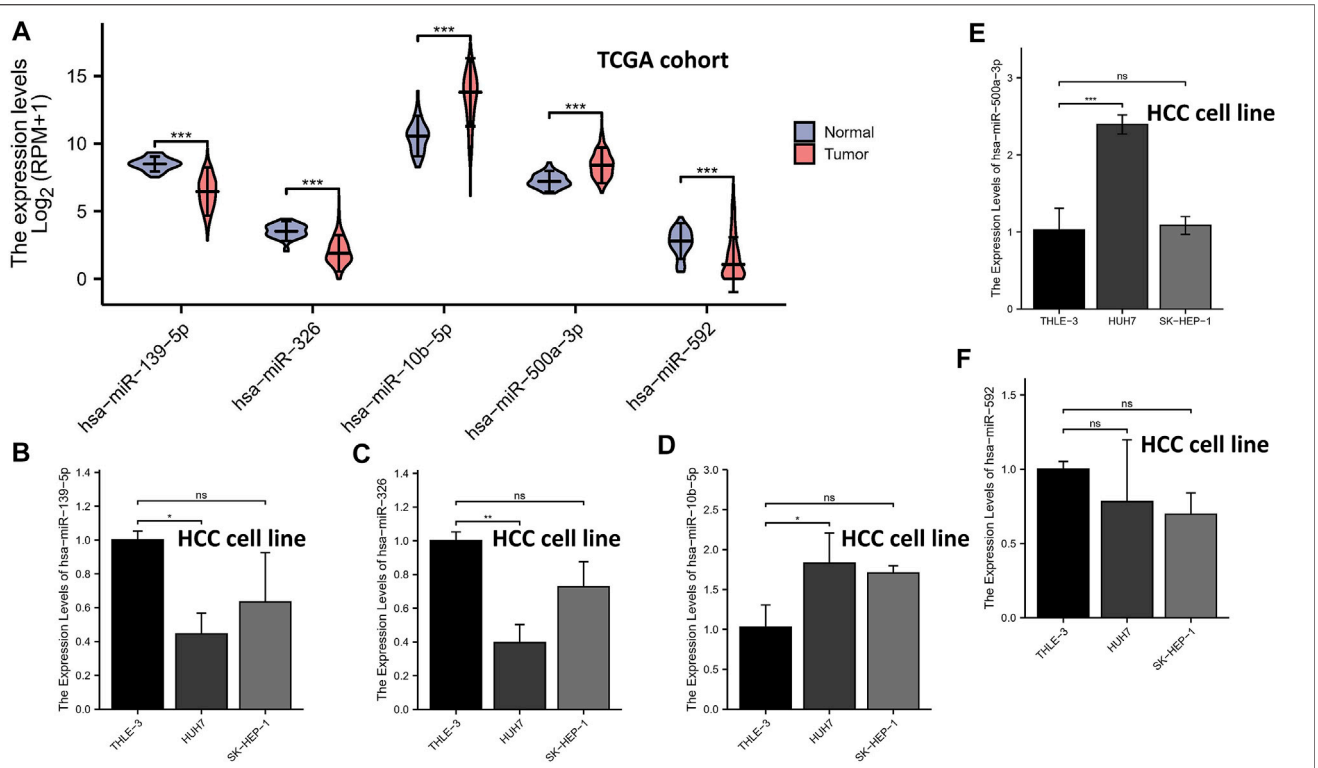


FIGURE 7 | qRT-PCR. The expression levels of the 5-miRNAs in the TCGA database **(A)**. The expression levels of miR-139-5p **(B)**, hsa-miR-326 **(C)**, miR-10b-5p **(D)**, miR-500a-3p **(E)**, and miR-592 **(F)**.

(Zhang et al., 2022). Meanwhile, miR-326 can suppresses the Hippo pathway when combined with PAX8 in trophoblast (Zang et al., 2022). Moreover, Wang and Chen (2022) also revealed that overexpression of miR-362 inhibited the expression of JMJD2A in nasopharyngeal carcinoma, and aberrant miR-362 may be associated with EBV-infection. In miR-10b-5p, it can inhibit tumorigenesis in the gastric cancer xenograft mice model by downregulating Tiam1 (Liu et al., 2021c); therefore, Yan et al. (2021) put forward a novel idea for exosomal miR-10b-5p in gastric cancer. A novel conclusion was that upregulated exosomal miR-10b-5p is involved in fibroblasts in tumor microenvironment. In miR-500a-3p, it has been shown that silence miR-500a-3p may serve as a new therapeutic strategy in the treatment of HCC (Jiang et al., 2017). Currently, there are three studies on the mechanism of miR-592 in HCC, in which the downstream consists of IGF-1R (Wang et al., 2017b), WSB1 (Jia et al., 2016), and DEK (Li et al., 2015). In addition to the miRNAs involved in the model, other miRNAs have also been studied in depth in HCC. In some studies about miR-223-3p, it can be used as a novel noninvasive biomarker for HCV-positive cirrhosis and HCC (Oksuz et al., 2015). Circulating mirR-223-3p can represent novel diagnostic and prognostic markers for HBV-associated HCC patients (Pratedrat et al., 2020). It is interesting to note that miR-331-3p (AUC: 0.832) has better diagnostic performance than AFP (Jin et al., 2019). On the other hand, these evidence suggest that sequencing of blood samples from HCC patients may play a noninvasive role in predicting survival outcomes. The novel mechanism of miR-500a-3p promoting HCC stem cell maintenance suggests that miR-500a-3p may be a novel therapeutic strategy (Jiang et al., 2017). Unfortunately, the critical role of the aforementioned pathways in HCC necroptosis is not well explained at present.

However, this study only used the data from the public database TCGA to construct the model, and there was no condition to collect our data to validate 5-miRNA signature, which was a limitation to our study. In addition, we have no conditions to verify the mechanism of these necroptosis-related miRNAs. In the future, we can use risk signatures based on altered 5-miRNAs, which may improve prognostic prediction and may lead to development of targeted therapy.

CONCLUSION

Comprehensively, our study suggests that necroptosis-related miRNAs is closely associated with the prognosis of HCC, and we established a robust tool for the prognostic management of HCC patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found at: https://portal.gdc.cancer.gov/.

ETHICS STATEMENT

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

AUTHOR CONTRIBUTIONS

TM and QW have contributed equally to this work and share first authorship. TM, QW, YY and YS conceived, designed, and supervised the study. TM and YR performed data analysis and drafted the manuscript. YS and YR helped revise the manuscript. TM, QW and YY collected the data. All the authors reviewed and approved the final manuscript.

FUNDING

This work was supported by the Liaoning Province "Xingliao Talents Plan" youth top-notch talent funding project (grant number XLYC1807145) and the Liaoning Provincial Department of Human Resources and Social Security talents self-help project (grant number 20200512).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.825261/full#supplementary-material

Supplementary Figure S1 | Kaplan–Meier analysis about 5-specific necroptosis-related miRNAs in the whole set. miR-10b-5p **(A)**, miR-139-5p **(B)**, hsa-miR-326 **(C)**, miR-500a-3p **(D)**, and miR-592 **(E)**.

REFERENCES

Bartel, D. P. (2004). MicroRNAs. *Cell* 116 (2), 281–297. doi:10.1016/s0092-8674(04)00045-5

Cai, X., Chen, Y., Man, D., Yang, B., Feng, X., Zhang, D., et al. (2021). RBM15 Promotes Hepatocellular Carcinoma Progression by Regulating N6-Methyladenosine Modification of YES1 mRNA in an IGF2BP1-dependent Manner. *Cell Death Discov.* 7, 315. doi:10.1038/s41420-021-00703-w

Chen, B., Liao, Z., Qi, Y., Zhang, H., Su, C., Liang, H., et al. (2020). miR-631 Inhibits Intrahepatic Metastasis of Hepatocellular Carcinoma by Targeting PTPRE. *Front. Oncol.* 10, 565266. doi:10.3389/fonc.2020.565266

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48 (D1), D127–D131. doi:10.1093/nar/gkz757

Christofferson, D. E., and Yuan, J. (2010). Necroptosis as an Alternative Form of Programmed Cell Death. *Curr. Opin. Cel Biol.* 22 (2), 263–268. doi:10.1016/j.ceb.2009.12.003

Di Sandro, S., Centonze, L., Centonze, L., Pinotti, E., Lauterio, A., De Carlis, R., et al. (2019). Surgical and Oncological Outcomes of Hepatic Resection for BCLC-B Hepatocellular Carcinoma: a Retrospective Multicenter Analysis Among 474 Consecutive Cases. *Updates Surg.* 71 (2), 285–293. doi:10.1007/s13304-019-00649-w

Ding, B., Fan, W., and Lou, W. (2020). hsa_circ_0001955 Enhances *In Vitro* Proliferation, Migration, and Invasion of HCC Cells through miR-145-5p/NRAS Axis. *Mol. Ther. - Nucleic Acids* 22, 445–455. doi:10.1016/j.omtn.2020.09.007

Dionísio, P. A., Amaral, J. D., and Rodrigues, C. M. P. (2020). Molecular Mechanisms of Necroptosis and Relevance for Neurodegenerative Diseases. *Int. Rev. Cel Mol Biol* 353, 31–82. doi:10.1016/bs.ircmb.2019.12.006

El-Serag, H. B. (2011). Hepatocellular Carcinoma. *N. Engl. J. Med.* 365 (12), 1118–1127. doi:10.1056/NEJMra1001683

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The Role of Necroptosis in Cancer Biology and Therapy. *Mol. Cancer* 18. 100. doi:10.1186/s12943-019-1029-8

Hitomi, J., Christofferson, D. E., Ng, A., Yao, J., Degterev, A., Xavier, R. J., et al. (2008). Identification of a Molecular Signaling Network that Regulates a Cellular Necrotic Cell Death Pathway. *Cell* 135 (7), 1311–1323. doi:10.1016/j.cell.2008.10.044

Hu, T., Zhao, X., Zhao, Y., Cheng, J., Xiong, J., and Lu, C. (2022). Identification and Verification of Necroptosis-Related Gene Signature and Associated Regulatory Axis in Breast Cancer. *Front. Genet.* 13, 842218. doi:10.3389/fgene.2022.842218

Huang, H.-Y., Lin, Y.-C. -D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., et al. (2020). miRTarBase 2020: Updates to the Experimentally Validated microRNA-Target Interaction Database. *Nucleic Acids Res.* 48 (D1), D148–D154. doi:10.1093/nar/gkz896

Huang, Y., Zou, Y., Xiong, Q., Zhang, C., Sayagués, J. M., Shelat, V. G., et al. (2021). Development of a Novel Necroptosis-Associated miRNA Risk Signature to Evaluate the Prognosis of colon Cancer Patients. *Ann. Transl Med.* 9 (24), 1800. doi:10.21037/atm-21-6576

Jia, Y.-Y., Zhao, J.-Y., Li, B.-L., Gao, K., Song, Y., Liu, M.-Y., et al. (2016). miR-592/WSB1/HIF-1α axis Inhibits Glycolytic Metabolism to Decrease Hepatocellular Carcinoma Growth. *Oncotarget* 7 (23), 35257–35269. doi:10.18632/oncotarget.9135

Jiang, C., Long, J., Liu, B., Xu, M., Wang, W., Xie, X., et al. (2017). miR-500a-3p Promotes Cancer Stem Cells Properties via STAT3 Pathway in Human Hepatocellular Carcinoma. *J. Exp. Clin. Cancer Res.* 36, 99. doi:10.1186/s13046-017-0568-3

Jin, W., Zhong, N., Wang, L., Yu, J., Yin, F., and Zhang, K. (2019). MiR-331-3p Inhibition of the Hepatocellular Carcinoma (HCC) Bel-7402 Cell Line by Down-Regulation of E2F1. *J. Nanosci Nanotechnol* 19 (9), 5476–5482. doi:10.1166/jnn.2019.16535

Lalaoui, N., Lindqvist, L. M., Sandow, J. J., and Ekert, P. G. (2015). The Molecular Relationships between Apoptosis, Autophagy and Necroptosis. *Semin. Cel Develop. Biol.* 39, 63–69. doi:10.1016/j.semcdb.2015.02.003

Li, J., McQuade, T., Siemer, A. B., Napetschnig, J., Moriwaki, K., Hsiao, Y.-S., et al. (2012). The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis. *Cell* 150 (2), 339–350. doi:10.1016/j.cell.2012.06.019

Li, W., Kong, X., Huang, T., Shen, L., Wu, P., and Chen, Q.-F. (2020). Bioinformatic Analysis and *In Vitro* Validation of a Five-microRNA Signature as a Prognostic Biomarker of Hepatocellular Carcinoma. *Ann. Transl Med.* 8 (21), 1422. doi:10.21037/atm-20-2509

Li, X., Zhang, W., Zhou, L., Yue, D., and Su, X. (2015). MicroRNA-592 Targets DEK Oncogene and Suppresses Cell Growth in the Hepatocellular Carcinoma Cell Line HepG2. *Int. J. Clin. Exp. Pathol.* 8, 12455

Li, X., Ren, Y., Liu, D., Yu, X., and Chen, K. (2022a). Role of miR-100-5p and CDC25A in Breast Carcinoma Cells. *PeerJ* 9, e12263. doi:10.7717/peerj.12263

Li, Y., Sun, Y., Li, Z., Li, S., and Wu, C. (2022b). MiR-139-5p Inhibits the Development of Gastric Cancer through Targeting TPD52. *J. Healthc. Eng.* 2022, 4033373. doi:10.1155/2022/4033373

Linkermann, A., and Green, D. R. (2014). Necroptosis. *N. Engl. J. Med.* 370 (5), 455–465. doi:10.1056/NEJMra1310050

Liu, F., An, X., Zhao, X., Zhang, N., Chen, B., Li, Z., et al. (2021c). MiR-10b-5p Inhibits Tumorigenesis in Gastric Cancer Xenograft Mice Model through Down-Regulating Tiam1. *Exp. Cel Res.* 407 (2), 112810. doi:10.1016/j.yexcr.2021.112810

Liu, X., Shan, W., Li, T., Gao, X., Kong, F., You, H., et al. (2021b). Cellular Retinol Binding Protein-1 Inhibits Cancer Stemness via Upregulating WIF1 to Suppress Wnt/β-Catenin Pathway in Hepatocellular Carcinoma. *BMC Cancer* 21 (1), 1224. doi:10.1186/s12885-021-08967-2

Liu, Y., Chen, Q., Zhu, Y., Wang, T., Ye, L., Han, L., et al. (2021a). Non-coding RNAs in Necroptosis, Pyroptosis and Ferroptosis in Cancer Metastasis. *Cel Death Discov.* 7, 210. doi:10.1038/s41420-021-00596-9

Liu, Z., Guo, C., Dang, Q., Wang, L., Liu, L., Weng, S., et al. (2022b). Integrative Analysis from Multi-center Studies Identities a Consensus Machine Learning-Derived lncRNA Signature for Stage II/III Colorectal Cancer. *EBioMedicine* 75, 103750. doi:10.1016/j.ebiom.2021.103750

Liu, Z., Guo, Y., Yang, X., Chen, C., Fan, D., Wu, X., et al. (2022d). Immune Landscape Refines the Classification of Colorectal Cancer with Heterogeneous Prognosis, Tumor Microenvironment and Distinct Sensitivity to Frontline Therapies. *Front. Cel Dev. Biol.* 9, 784199. doi:10.3389/fcell.2021.784199

Liu, Z., Liu, L., Weng, S., Guo, C., Dang, Q., Xu, H., et al. (2022a). Machine Learning-Based Integration Develops an Immune-Derived lncRNA Signature for Improving Outcomes in Colorectal Cancer. *Nat. Commun.* 13, 816. doi:10.1038/s41467-022-28421-6

Liu, Z., Xu, H., Weng, S., Ren, Y., and Han, X. (2021c). Stemness Refines the Classification of Colorectal Cancer with Stratified Prognosis, Multi-Omics Landscape, Potential Mechanisms, and Treatment Options. *Front. Immunol.* 13, 828330. doi:10.3389/fimmu.2022.828330

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., et al. (2021). Hepatocellular Carcinoma. *Nat. Rev. Dis. Primers* 7. doi:10.1038/s41572-020-00240-3

Long, J., Liu, B., Yao, Z., Weng, H., Li, H., Jiang, C., et al. (2022). miR-500a-3p Is a Potential Prognostic Biomarker in Hepatocellular Carcinoma. *Ijgm* 15: 1891–1899. doi:10.2147/IJGM.S340629

Long, J. S., and Ryan, K. M. (2012). New Frontiers in Promoting Tumour Cell Death: Targeting Apoptosis, Necroptosis and Autophagy. *Oncogene* 31 (49), 5045–5060. doi:10.1038/onc.2012.7

Niu, X., Sun, H., Qiu, F., Liu, J., Yang, T., and Han, W. (2021). miR-10b-5p Suppresses the Proliferation and Invasion of Primary Hepatic Carcinoma Cells by Downregulating EphA2. *Biomed. Res. Int.* 2021, 1382061. doi:10.1155/2021/1382061

Oksuz, Z., Serin, M. S., Kaplan, E., Dogen, A., Tezcan, S., Aslan, G., et al. (2015). Serum microRNAs; miR-30c-5p, miR-223-3p, miR-302c-3p and miR-17-5p Could Be Used as Novel Non-invasive Biomarkers for HCV-Positive Cirrhosis and Hepatocellular Carcinoma. *Mol. Biol. Rep.* 42 (3), 713–720. doi:10.1007/s11033-014-3819-9

Paul, R., Bapat, P., Deogharkar, A., Kazi, S., Singh, S. K. V., Gupta, T., et al. (2021). MiR-592 Activates the mTOR Kinase, ERK1/ERK2 Kinase Signaling and Imparts Neuronal Differentiation Signature Characteristic of Group 4 Medulloblastoma. *Hum. Mol. Genet.* 30 (24), 2524–2525. doi:10.1093/hmg/ddab264

Pratedrat, P., Chuaypen, N., Nimsamer, P., Payungporn, S., Pinjaroen, N., Sirichindakul, B., et al. (2020). Diagnostic and Prognostic Roles of Circulating miRNA-223-3p in Hepatitis B Virus-Related Hepatocellular Carcinoma. *PLoS One* 15 (4), e0232211. doi:10.1371/journal.pone.0232211

Riffo-Campos, Á., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for Sequence-Based miRNA Target Prediction: What to Choose? *Ijms* 17 (12), 1987. doi:10.3390/ijms17121987

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Ruan, Z.-h., Xu, Z.-x., Zhou, X.-y., Zhang, X., and Shang, L. (2019). Implications of Necroptosis for Cardiovascular Diseases. *Curr. Med. Sci.* 39 (4), 513–522. doi:10.1007/s11596-019-2067-6

Seehawer, M., Heinzmann, F., D'Artista, L., Harbig, J., Roux, P.-F., Hoenicke, L., et al. (2018). Necroptosis Microenvironment Directs Lineage Commitment in Liver Cancer *Nature* 562, 69–75. doi:10.1038/s41586-018-0519-y

Strilic, B., Yang, L., Albarrán-Juárez, J., Wachsmuth, L., Han, K., Müller, U. C., et al. (2016). Tumour-cell-induced Endothelial Cell Necroptosis via Death Receptor 6 Promotes Metastasis. *Nature* 536 (7615), 215–218. doi:10.1038/nature19076

Visalli, M., Bartolotta, M., Polito, F., Oteri, R., Barbera, A., Arrigo, R., et al. (2018). miRNA Expression Profiling Regulates Necroptotic Cell Death in Hepatocellular Carcinoma. *Int. J. Oncol.* 53 (2), 771–780. doi:10.3892/ijo.2018.4410

Wang, T., Xu, L., Jia, R., and Wei, J. (2017a). MiR-218 Suppresses the Metastasis and EMT of HCC Cells via Targeting SERBP1. *Acta Biochim. Biophys. Sin (Shanghai).* 49 (5), 383–391. doi:10.1093/abbs/gmx017

Wang, W., Zhang, H., Tang, M., Liu, L., Zhou, Z., Zhang, S., et al. (2017b). MicroRNA-592 Targets IGF-1R to Suppress Cellular Proliferation, Migration and Invasion in Hepatocellular Carcinoma. *Oncol. Lett.* 13 (5), 3522–3528. doi:10.3892/ol.2017.5902

Wang, X., and Chen, P. (2022). Aberrant miR-362-3p Is Associated with EBV-Infection and Prognosis in Nasopharyngeal Carcinoma and Involved in Tumor Progression by Targeting JMJD2A. *Ott* 15, 121. doi:10.2147/OTT.S325100

Wang, Z., Jensen, M. A., and Zenklusen, J. C. (2016). A Practical Guide to the Cancer Genome Atlas (TCGA). *Methods Mol. Biol.* 1418, 111–141. doi:10.1007/978-1-4939-3578-9_6

Wei, B., Wang, Z., Lian, Q., Chi, B., and Ma, S. (2022). hsa_circ_0139402 Promotes Bladder Cancer Progression by Regulating Hsa-miR-326/pax8 Signaling. *Dis. Markers*, 2022, 9899548. doi:10.1155/2022/9899548

Xiang, Y. k., Peng, F. h., Guo, Y. q., Ge, H., Cai, S. y., Fan, L. x., et al. (2021). Connexin32 Activates Necroptosis through Src-mediated Inhibition of Caspase 8 in Hepatocellular Carcinoma. *Cancer Sci.* 112 (9), 3507–3519. doi:10.1111/cas.14994

Yan, T., Wang, X., Wei, G., Li, H., Hao, L., Liu, Y., et al. (2021). Exosomal miR-10b-5p Mediates Cell Communication of Gastric Cancer Cells and Fibroblasts and Facilitates Cell Proliferation. *J. Cancer* 12, 2140–2150. doi:10.7150/jca.47817

Yassin, N. Y. S., AbouZid, S. F., El-Kalaawy, A. M., Ali, T. M., Almehmadi, M. M., and Ahmed, O. M. (2022). Silybum marianum Total Extract, Silymarin and Silibinin Abate Hepatocarcinogenesis and Hepatocellular Carcinoma Growth via Modulation of the HGF/c-Met, Wnt/β-Catenin, and PI3K/Akt/mTOR Signaling Pathways. *Biomed. Pharmacother.* 145, 112409. doi:10.1016/j.biopha.2021.112409

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zang, J., Yan, M., Zhang, Y., Peng, W., Zuo, J., Zhou, H., et al. (2022). MiR-326 Inhibits Trophoblast Growth, Migration, and Invasion by Targeting PAX8 via Hippo Pathway. *Reprod. Biol. Endocrinol.* 20, 38. doi:10.1186/s12958-022-00909-2

Zhang, Y., Li, M., Qiu, Y., Wu, Y., Chen, S., Ni, B., et al. (2022). MiR-139-5p/ENAH Affects Progression of Hepatocellular Carcinoma Cells. *Biochem. Genet.* [Epub ahead of print]. doi:10.1007/s10528-022-10204-9

Zhang, Y., Qian, M., Tang, F., Huang, Q., Wang, W., Li, Y., et al. (2020). Identification and Analysis of P53-Regulated Enhancers in Hepatic Carcinoma. *Front. Bioeng. Biotechnol.* 8, 668, doi:10.3389/fbioe.2020.00668

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Meng, Wang, Yang, Ren and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

BRIEF RESEARCH REPORT
published: 25 April 2022
doi: 10.3389/fgene.2022.872253

Evaluating a Panel of Autoantibodies Against Tumor-Associated Antigens in Human Osteosarcoma

Manli Luo[1,2,3†], Songmei Wu[1,3†], Yan Ma[1], Hong Liang[1], Yage Luo[1], Wentao Gu[1], Lijuan Fan[1,3], Yang Hao[1,3], Haiting Li[1,3] and Linbo Xing[1,3*]*

[1]Luoyang Orthopedic Hospital of Henan Province (Orthopedic Hospital of Henan Province), Luoyang, China, [2]Henan Provincial Rehabilitation Hospital, Luoyang, China, [3]Henan University of Chinese Medicine, Zhengzhou, China

Background: The aim of this study was to identify a panel of candidate autoantibodies against tumor-associated antigens in the detection of osteosarcoma (OS) so as to provide a theoretical basis for constructing a non-invasive serological diagnosis method in early immunodiagnosis of OS.

Methods: The serological proteome analysis (SERPA) approach was used to select candidate anti-TAA autoantibodies. Then, indirect enzyme-linked immunosorbent assay (ELISA) was used to verify the expression levels of eight candidate autoantibodies in the serum of 51 OS cases, 28 osteochondroma (OC), and 51 normal human sera (NHS). The rank-sum test was used to compare the content of eight autoantibodies in the sera of three groups. The diagnostic value of each indicator for OS was analyzed by an ROC curve. Differential autoantibodies between OS and NHS were screened. Then, a binary logistic regression model was used to establish a prediction logistical regression model.

Results: Through ELISA, the expression levels of seven autoantibodies (ENO1, GAPDH, HSP27, HSP60, PDLIM1, STMN1, and TPI1) in OS patients were identified higher than those in healthy patients ($p < 0.05$). By establishing a binary logistic regression predictive model, the optimal panel including three anti-TAAs (ENO1, GAPDH, and TPI1) autoantibodies was screened out. The sensitivity, specificity, Youden index, accuracy, and AUC of diagnosis of OS were 70.59%, 86.27%, 0.5686, 78.43%, and 0.798, respectively.

Conclusion: The results proved that through establishing a predictive model, an optimal panel of autoantibodies could help detect OS from OC or NHS at an early stage, which could be used as a promising and powerful tool in clinical practice.

Keywords: osteosarcoma, tumor-associated antigen, autoantibody, detection, early diagnosis, panel

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Xiangqian Guo,
Henan University, China
Jiantao Li,
Chinese PLA General Hospital, China

***Correspondence:**
Haiting Li
1372130667@qq.com
Linbo Xing
450078536@qq.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *09 February 2022*
Accepted: *23 March 2022*
Published: *25 April 2022*

Citation:
Luo M, Wu S, Ma Y, Liang H, Luo Y, Gu W, Fan L, Hao Y, Li H and Xing L (2022) Evaluating a Panel of Autoantibodies Against Tumor-Associated Antigens in Human Osteosarcoma. Front. Genet. 13:872253. doi: 10.3389/fgene.2022.872253

INTRODUCTION

Osteosarcoma (OS) is one of the most common primary malignancies of bone sarcomas, whose incidence rate is less than 1% of all tumors in the United States (Howlader et al., 2020; Siegel et al., 2020). Worldwide, OS occurs in about 1–3 cases per million people annually (Kansara et al., 2014). OS has a bimodal age distribution and occurs mostly in children and adolescents; another type occurs in people over 60 years old (Corre et al., 2020). In addition, once clinically diagnosed with OS, there are about 15–20% patients who

have detectable metastases, with a 5-year survival rate of ~20% (Wang et al., 2020). Surgical resection combination with chemotherapy is one of the common measures to treat the tumor (Ta et al., 2009). As OS is highly malignant and has low incidence and non-specific initial symptoms, it is often misdiagnosed or neglected. However, until now, there are no efficient biomarkers for detection in clinics. Thus, in order to treat the patients timely, it is essential to diagnose OS at an early stage with a non-invasive method, which has high specificity/sensitivity.

Gene mutations can cause the changes in the gene and abnormal expression of gene products, resulting in the occurrence and development of tumors (Sparano et al., 2013). These abnormal proteins often appear along with the development of the tumor, some of which can appear in the blood circulation system of the body which are called tumor-associated antigens (TAAs) (Tan and Zhang, 2008; Restifo et al., 2012). These TAAs can be recognized by the immune system of tumor patients, and then, anti-TAA autoantibodies were produced. An important fact is that autoantibodies can exist stably in the host for even months or years before clinical diagnosis (Lu et al., 2008). Meanwhile, the content of autoantibodies in the sera of normal humans is low, while in tumor patients it is much higher, so these autoantibodies have great potential as tumor biomarkers in the early immunodiagnosis of OS (Tan and Zhang, 2008; Bracci et al., 2012; Anderson et al., 2015). It was known to all that the anti-p53 autoantibody was used for diagnosing different tumors (Soussi, 2000). Various studies have illustrated the role of autoantibodies in diagnosing tumors, such as in hepatocellular carcinoma (Zhang and Tan, 2010), lung cancer (Huo et al., 2020), breast cancer (Qiu et al., 2018), and esophageal cancer (Pan et al., 2019). However, the function of a single autoantibody was limited; many studies have proven that a custom-made panel of autoantibodies could enhance the diagnostic performance of tumors (Chen et al., 2020; Huang et al., 2020; Okada et al., 2020).

Serological proteome analysis (SERPA) is one of the methods to screen new TAAs. In our study, new autoantibodies were discovered by SERPA, the method which was mature in our laboratory (Dai et al., 2017). Its main advantage is that it can directly screen the proteins extracted from tumor tissues or cells as antigen sources. As enzyme-linked immunosorbent assay (ELISA) is rapid, simple, and inexpensive, it becomes the most commonly used method to detect TAAs or autoantibodies as serum tumor biomarkers.

Our previous study (Li et al., 2021) has found a panel of eight candidate serum autoantibodies with SERPA. In this study, these eight candidate autoantibodies were detected by indirect ELISA. Differentially expressed autoantibodies were screened out among eight candidate autoantibodies. Then, an optimal panel of autoantibodies was established by the logistic regression statistical method. Eventually, the diagnostic value of the panel was evaluated in the detection of OS in subgroups.

METHODS AND MATERIALS

Study Population

There were 51 clinically confirmed OS cases, 28 osteochondroma (OC), and 51 NHS included in this study, which were matched by gender and age. All healthy samples were confirmed with no malignant diseases. The sera of 51 OS cases and 28 benign controls (OC) were obtained from the Luoyang Orthopedic Hospital of Henan Province (Orthopedic Hospital of Henan Province) from November 2013 to June 2016, and the healthy control of 51 cases was obtained from healthy people who were examined in the physical examination department at the same hospital in the same period. All blood samples were centrifuged at 3,000 rpm for 10 min after collection and then stored in a refrigerator at −80°C before the experiment. All subjects participating in the study have signed the informed consent form. The study had been approved by the Ethics Committee of Luoyang Orthopedic Hospital of Henan Province.

Serological Proteome Analysis

Two osteosarcoma cell lines U2-OS and Saos-2 were cultured. Specific experimental methods and results were based on our published literature (Li et al., 2021).

Enzyme-Linked Immunosorbent Assay

Eight recombinant proteins including ENO1, GAPDH, HSP27, HSP60, NPM1, PDLIM1, STMN1, and TPI1 were used as antigens to detect the corresponding autoantibodies. Purified ENO1 protein was bought from Sigma. GAPDH (ab77109), TPI1 (ab100826), HSP60 (ab78430), PDLIM1 (ab177676), and STMN1 (ab87492) were purchased from Abcam Inc (Cambridge, MA, United States). NPM1 was provided by the Henan Key Laboratory of Tumor Epidemiology.

Autoantibodies were detected by indirect ELISA. Each purified protein was diluted to an appropriate concentration with a coating solution (ENO1, GAPDH, HSP27, HSP60, NPM1, PDLIM1, STMN1, and TPI1 to final concentrations of 1.0 μg/ml). For the assay, 100 μl of the coating solution was added to each well of a 96-well ELISA plate. Then, the plate was sealed with a plastic wrap and placed in the refrigerator at 4°C for 16 h. After discarding the coating solution, 200 μl of 2% BSA was added into each well and was then placed in the refrigerator at 4°C overnight for 16 h. The plate was washed with 1 × PBST solution three times on the 96-well automatic washer. Then, 100 μl of pre-diluted serum samples diluted at 1:100 in 1% BSA in the deep well plate was added for incubating at 37°C for 1 h. After being washed with the 1 × PBST solution five times, the plates were incubated with secondary antibody, horseradish peroxidase (HRP)-conjugated goat anti-human IgG (H+L), diluted by 1:4,000 in 1% BSA at 37°C for 1 h. After also being washed with 1 × PBST solution five times, the substrate (1 mg/ml 2,2-azino-bis [3-ethylbenzthiazoline-6-sulfonic acid] with 0.005% hydrogen peroxide in citrate buffer, pH 4.6) was used as detecting reagents. The optical density (OD) was measured at 405 nm using an automated plate reader. All serum samples were assayed in duplicate.

Statistical Analysis

All statistical analyses were performed by SPSS 21.0 and GraphPad Prism 6.0 software. The Mann–Whitney U test was used to compare the expression level of autoantibodies if there were any abnormal distribution (Kolmogorov–Smirnov test). Receiver operating characteristic (ROC) curves were conducted

TABLE 1 | Basic clinical characteristics of OS, OC, and NHS.

	OS	OC	NHS
	n = 51	*n* = 28	*n* = 51
Age			
Mean, SD	24.88, 14.29	17.21, 12.02	26.35, 13.40
Range	4–64	4–52	5–59
Gender (%)			
Male	33 (64.71)	19 (67.86)	33 (64.71)
Female	18 (35.29)	9 (32.14)	18 (35.29)
TNM (%)			
I	6 (11.76)		
IIA	23 (45.10)		
IIB	9 (17.65)		
IV	9 (17.65)		
Unknown	4 (7.84)		
Degree of differentiation (%)			
1–2	4 (7.84)		
3–4	41 (80.39)		
Unknown	6 (11.76)		
Tumor size			
<5 cm	13 (25.49)		
≥5 cm	38 (74.51)		

to evaluate the diagnostic value of each autoantibody and the prediction model. The binary backward stepwise logistic regression prediction model (condition) was used to select the optimal panel of autoantibodies. The comparison between two AUCs was conducted by Medcalc 11 software. The predicting probability $p = 0.5$ was set as the cutoff point. p was two-tailed and less than 0.05 was considered significant.

RESULTS

Basic Clinical Characteristics

There were 51 new cases of OS, 28 OC, and 51 healthy people included in this study. The basic clinical characteristics are shown in **Table 1**. There was no statistically significant difference in gender and age between OS and NHS ($p > 0.05$).

Expression of Candidate Autoantibodies in Three Groups

Based on our previous study, using SERPA, eight TAAs were screened out. Here, by ELISA, seven autoantibodies whose OD values in OS were higher than those in NHS and p-values less than 0.05 were screened out (ENO1, GAPDH, HSP27, HSP60, PDLIM1, STMN1, and TPI1). More details are shown in **Figure 1** and **Supplementary Table S1**.

Expression of Eight Autoantibodies in Osteosarcoma and Control Group

ROC curves of eight autoantibodies in the sera of OS and NHS were measured, as shown in **Figure 2**. The contents of ENO1, GAPDH, HSP27, HSP60, PDLIM1, STMN1, and TPI1 autoantibodies in the sera of OS were higher than those of NHS, and the AUC was statistically significant ($p < 0.05$). The range of the AUC of seven autoantibodies was 0.617–0.716. Then, these seven TAAs were further included in the next stage for verification.

The Optimal Panel of Autoantibodies in Diagnosing Osteosarcoma Was Established

A backward stepwise (condition) logistic regression predictive model was conducted to select the optimal model among seven biomarkers in diagnosing OS. The predictive probability in diagnosing OS is as follows: PRE (P= OS, 3 TAAs) = 1/(1 + EXP (−(−6.679 + 9.686 × ENO1 + 17.286 × GAPDH +11.178 × TPI1))). Predicting probability $p = 0.5$ was set as the cutoff point. The sensitivity, specificity, Youden index, accuracy, and AUC of this model were 70.59%, 86.27%, 0.5686, 78.43%, and 0.798, respectively. The ROC curves among three groups are shown in **Figure 3**. The AUC was statistically significant ($p < 0.05$).

The Diagnostic Value of the Optimal Panel in the Early Stage and Ages Up to 19 Years Old

The subjects were divided into two subgroups as follows: early stage (I–II) and late stage (III–IV). In both stages, the AUC of OS was higher than that of OC or NHS ($p < 0.05$), except for the OS (III–IV) vs. OC ($p > 0.05$). When OS was divided by 19 years old, in these two age subgroups, the AUC of OS was higher than that of OC or NHS ($p < 0.05$). More details are shown in **Supplementary Figures S1, S2** and **Supplementary Table S2**. In addition, either in stage groups or age groups, the comparison between two subgroups (neither vs. OC or NHS) had statistical significance.

DISCUSSION

As OS has a high rate of amputation, disability, and death, it is important to detect and treat OS at an early stage (Wafa and Grimer, 2006). The non-invasive serological screening method is optimal. Autoantibodies as an autoimmune phenomenon, which have been reported in many studies, can be served as an effective means (Qin et al., 2019; Huo et al., 2020; Okada et al., 2020; Qiu et al., 2020). SERPA is a promising approach used in identifying autoantibodies (Dai et al., 2016). In the previous study, eight candidate TAAs which had statistical significance between OS and NHS were selected for further validation based on SERPA. In this study, we measured the contents of ENO1, GAPDH, HSP27, HSP60, NPM1, PDLIM1, STMN1, and TPI1 autoantibodies by ELISA. The expression levels of seven autoantibodies were higher in OS than those in NHS. However, a single biomarker was limited in diagnosing OS. Based on these data, we established a predictive model containing three biomarkers including ENO1, GAPDH, and TPI1 to help improve the diagnostic performance. Afterward, the diagnostic value was evaluated among OS, OC, and NHS.

These three TAAs have been studied in many tumors. The expression of ENO1 in the invasion and metastasis of tumor cells was concerned by many studies with high expressions in OS

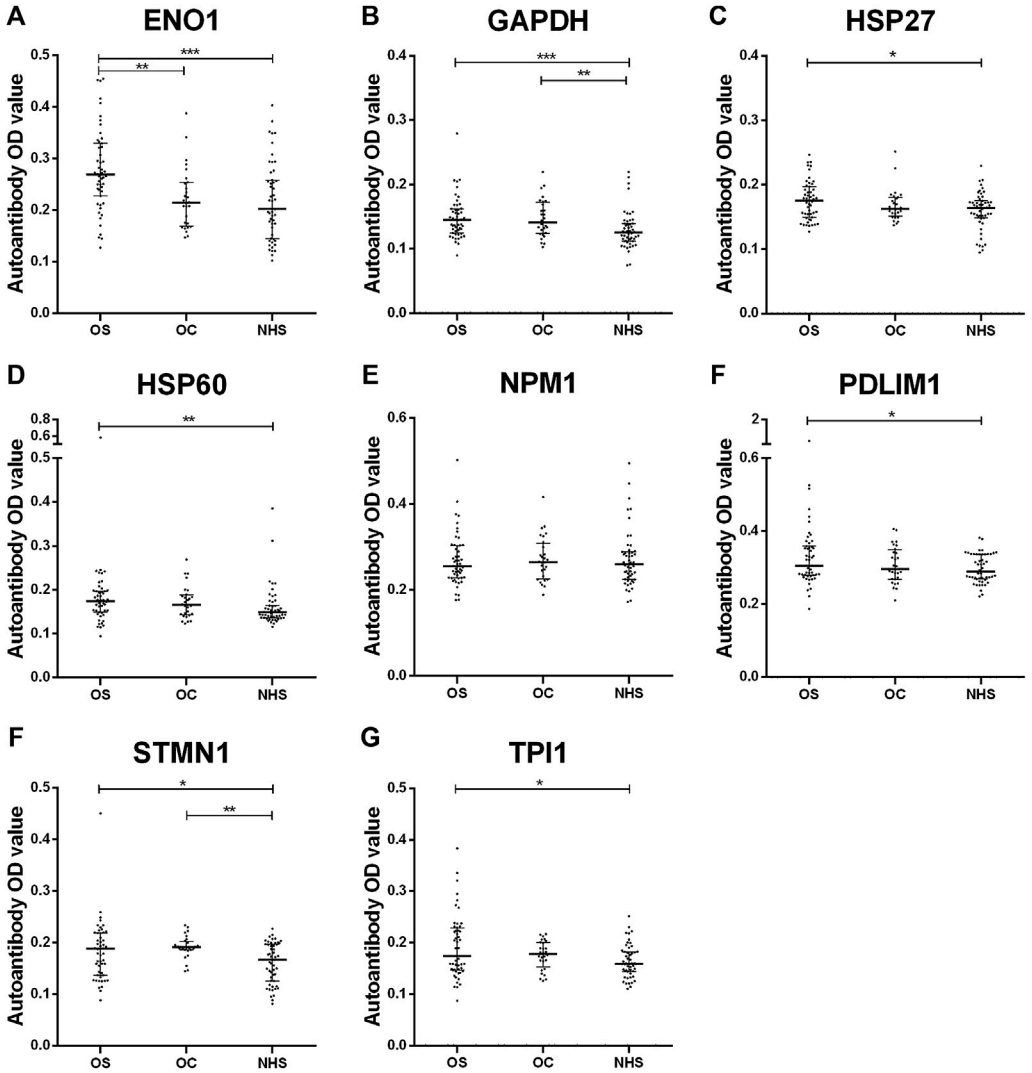


FIGURE 1 | Scatter plots of serum levels of autoantibodies against eight TAAs in OS, OC and NHS (median with interquartile). $p < 0.001$; $p < 0.01$; and $p < 0.05$. OS, osteosarcoma; OC, osteochondroma; NHS, normal human sera.

(Chen et al., 2014), breast cancer (Gao et al., 2013), lung cancer (Zhang et al., 2018), and hepatocellular carcinoma (Zhu et al., 2018; Deng et al., 2020). During cancer development and progression, glycolysis and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) played an important role, such as in colorectal adenocarcinoma (Brzozowa-Zasada et al., 2018) and breast cancer (Goes et al., 2010). Ogino et al. (2007) had proven that GAPDH mRNA was associated with lung metastasis in osteosarcoma. The poor prognosis of peripheral T-cell lymphoma (Ludvigsen et al., 2018) or hepatocellular carcinoma (Jiang et al., 2017) was associated with triosephosphate isomerase (TPI1). Research studies using TPI1 in early diagnosis of OS was hardly performed.

Many studies have shown that the diagnostic ability of a single biomarker was limited, and the combination of several ones could increase the diagnostic performance in different cancers (Zhang et al., 2016; Li et al., 2017; Wang S et al., 2018). For OS, Gao et al. (2020) reported a meta-analysis that the combination of several circulating miRNAs had a high accuracy whose sensitivity, specificity, and AUC were 0.79, 0.89, and 0.90. Three lncRNAs (RP1-261G23.7, RP11-69E11.4, and SATB2-AS1) were selected as independent prognostic factors for OS patients (Ying et al., 2020). Nevertheless, the use of a panel of autoantibodies was proven in other kinds of tumors but not in OS. Wang et al. (2019) found an optimal panel of nine autoantibodies (RalA, p62, p53, koc, p90, p16, c-myc, AHSG, and 14-3-3zeta), whose sensitivity and specificity can reach 61.4% and 85.0%, respectively. Furthermore, combined with CA125 used in clinical practice, the sensitivity, specificity, and AUC could reach 94.7, 78.2, and 0.914%. Zhang et al. (2018) selected a panel of four autoantibodies, which provided a higher diagnostic performance

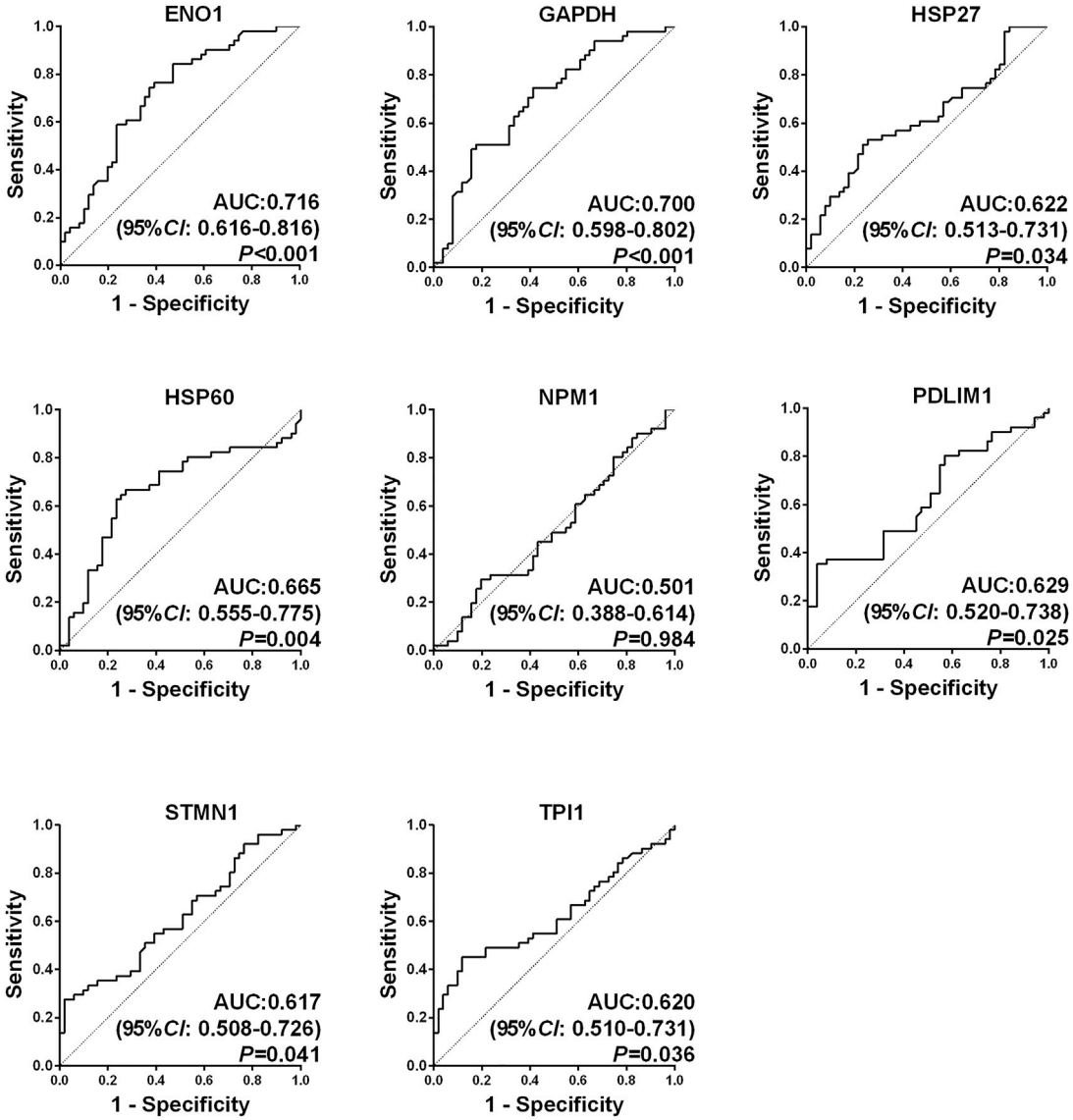


FIGURE 2 | Receiver operating characteristic (ROC) curves of eight candidate tumor-associated autoantibodies (TAAbs).

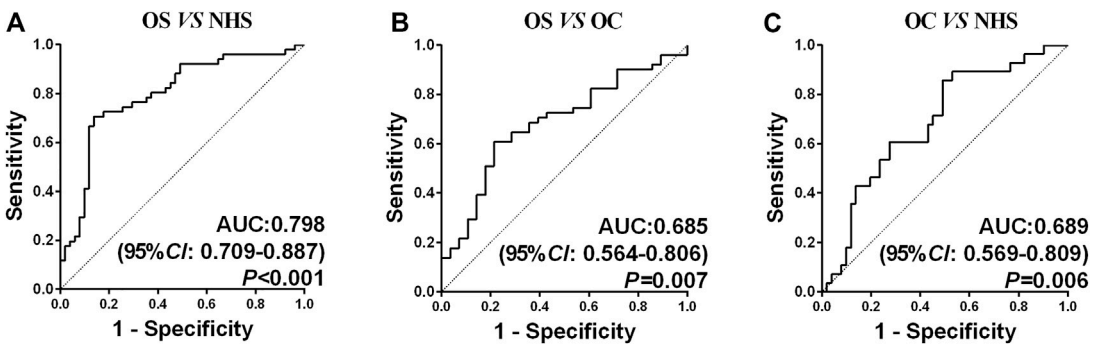


FIGURE 3 | ROC curve analysis of the prediction model with the TAA panel of OS detection. **(A)** Prediction model with three TAAs for OS detection in healthy controls. **(B)** Prediction model with three TAAs for OS detection in OC. **(C)** Prediction model with three TAAs for OC detection in healthy controls.

with AUC 0.838 in the training cohort and 0.872 in the validation cohort for ESCC detection. In hepatocellular carcinoma, six autoantibodies (Sui1, p62, RalA, p53, NY-ESO-1, and c-myc) may have both diagnostic and prognostic values in HCC (Okada et al., 2020).

Until now, studies focusing on using autoantibodies as biomarkers in early diagnosis of OS are almost rare. Our research is leading in this area. In this study, the range of the AUC of a single autoantibody was 0.617–0.716, which was difficult to meet the requirements for early diagnosis in clinical practice. By establishing the logistic regression predictive model, which was widely used to classify diseases, especially in cancers (Wang S et al., 2018; Cui et al., 2021), it proved to be a conventional analytical method. This model contained three biomarkers finally, and the sensitivity, specificity, Youden index, accuracy, and AUC of this model were 70.59%, 86.27%, 0.5686, 78.43%, and 0.798, respectively, between OS and NHS. By comparing the OS and OC, the sensitivity, specificity, Youden index, accuracy, and AUC of this model were 70.59%, 57.14%, 0.2773, 65.82%, and 0.685, respectively, and the comparison between these two AUCs was not statistically significant. The panel could help identify OS patients from neither benign tumor nor normal people, which means this panel could serve as good diagnostic biomarkers. Based on these aforementioned studies, we were confident in using the panel of autoantibodies to detect OS at an early stage; thus, the disease could be treated earlier. In addition, if the panel could be combined with common clinical indicators in the future, the results are hopeful.

In this study, OS and NHS patients were matched by age and gender. After the predictive model was established, the predictive value was calculated in benign tumors as well. It was reported that serum soluble B7-H3 concentrations were significantly higher than those in benign diseases and healthy people ($p < 0.05$) (Wang L et al., 2018). Zhao et al. (2017) presented the result that the expression of HBME-1 was higher in OS tissues than that in OC and normal bone tissues. The AUC of the HBME-1 expression was 0.864, with a sensitivity of 80.92%, specificity of 91.89%, and an accuracy of 84.51%. Xu et al. (2010) reported that the basic fibroblast growth factor (b-FGF) and endostatin were expressed higher in OS than in OC tissue. However, they were rarely focused on evaluating serological diagnostic biomarkers of OS in this field, which was one of the advantages of this study. In addition, this study did further research works on the diagnostic value of the model on the early (I–II) and late stages (III–IV), and the subgroup was divided as the subjects of age 19. In the early stage (I–II), the sensitivity, specificity, Youden index, accuracy, and AUC of diagnosis of OS were 68.42%, 86.27%, 0.5470, 78.65%, and 0.796, between OS and NHS. This indicated that this panel could be a promising and powerful tool in clinical practice, especially for people at the early stage of OS, which could help treat these patients early. While in the subgroup of age, the comparison of the AUC between these two groups did not have statistical significance.

At present, autoantibodies against tumor-associated antigens were studied by many researchers in the field of cancer for early diagnosis or prognosis. In our study, eight candidate tumor-associated antigens' autoantibodies were screened out through the SERPA technology, a sufficient screening method. On the other hand, this study included benign controls, which confirmed the results. However, there still existed some limitations. At first, the number of the subjects was limited as OS was difficult to collect, as well as the inequality of each stage. It could be better if a larger number of OS and OC samples could be collected in the future. Also, if it could be combined with other types of biomarkers, it would be better for clinical practice.

CONCLUSION

In conclusion, our results have proven that using the binary logistic regression predictive model, an optimal panel of autoantibodies could help detect OS from OC or NHS at an early stage as a promising and powerful tool in clinical practice.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Luoyang Orthopedic Hospital of Henan Province. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

LX and HL conceived the study. ML and SW conducted the study and drafted the application sections, and they contributed equally to this work. YM and HL analyzed the data. YL, WG, LF, and YH collected the serum and partly participated in the experiment. All authors read, revised, and approved the final manuscript.

FUNDING

This work was supported by the Innovation Fund of National Clinical Research Center for Orthopedics, Sports Medicine & Rehabilitation (2021-NCRC-CXJJ-PY-13), the Major Project of TCM research in Henan Province (2018ZYZD01 and 2019ZY2156), the Major Project of Science and Technology in Henan Province (192102310437), and the project of health guiding plan in Luoyang (1930005A, 1930013A, and 2101043A).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872253/full#supplementary-material

REFERENCES

Anderson, K. S., Cramer, D. W., Sibani, S., Wallstrom, G., Wong, J., Park, J., et al. (2015). Autoantibody Signature for the Serologic Detection of Ovarian Cancer. *J. Proteome Res.* 14, 578–586. doi:10.1021/pr500908n

Bracci, P. M., Zhou, M., Young, S., and Wiemels, J. (2012). Serum Autoantibodies to Pancreatic Cancer Antigens as Biomarkers of Pancreatic Cancer in a San Francisco Bay Area Case-Control Study. *Cancer* 118, 5384–5394. doi:10.1002/cncr.27538

Brzozowa-Zasada, M., Kurek, J., Piecuch, A., and Stęplewska, K. (2018). Correlation Study of GAPDH, Bcl-2, and Bax Protein Immunoexpression in Patients with Colorectal Adenocarcinoma. *pg* 13, 322–331. doi:10.5114/pg.2018.79813

Chen, K., Zhu, C., Cai, M., Fu, D., Cheng, B., Cai, Z., et al. (2014). Integrative Metabolome and Transcriptome Profiling Reveals Discordant Glycolysis Process between Osteosarcoma and normal Osteoblastic Cells. *J. Cancer Res. Clin. Oncol.* 140, 1715–1721. doi:10.1007/s00432-014-1719-y

Chen, S.-S., Li, K., Wu, J., Peng, Z.-Y., Wang, Z.-D., Wang, J.-C., et al. (2020). Stem Signatures Associated Antibodies Yield Early Diagnosis and Precise Prognosis Predication of Patients with Non-small Cell Lung Cancer. *J. Cancer Res. Clin. Oncol.* 147, 223–233. doi:10.1007/s00432-020-03325-4

Corre, I., Verrecchia, F., Crenn, V., Redini, F., and Trichet, V. (2020). The Osteosarcoma Microenvironment: A Complex but Targetable Ecosystem. *Cells* 9, 976. doi:10.3390/cells9040976

Cui, C., Duan, Y., Qiu, C., Wang, P., Sun, G., Ye, H., et al. (2021). Identification of Novel Autoantibodies Based on the Human Proteomic Chips and Evaluation of Their Performance in the Detection of Gastric Cancer. *Front. Oncol.* 11, 637871. doi:10.3389/fonc.2021.637871

Dai, L., Li, J., Xing, M., Sanchez, T. W., Casiano, C. A., and Zhang, J.-Y. (2016). Using Serological Proteome Analysis to Identify Serum Anti-nucleophosmin 1 Autoantibody as a Potential Biomarker in European-American and African-American Patients with Prostate Cancer. *Prostate* 76, 1375–1386. doi:10.1002/pros.23217

Dai, L., Qu, Y., Li, J., Wang, X., Wang, K., Wang, P., et al. (2017). Serological Proteome Analysis Approach-Based Identification of ENO1 as a Tumor-Associated Antigen and its Autoantibody Could Enhance the Sensitivity of CEA and CYFRA 21-1 in the Detection of Non-small Cell Lung Cancer. *Oncotarget* 8, 36664–36673. doi:10.18632/oncotarget.17067

Deng, F., Chen, D., Wei, X., Lu, S., Luo, X., He, J., et al. (2020). Development and Validation of a Prognostic Classifier Based on HIF-1 Signaling for Hepatocellular Carcinoma. *Aging* 12, 3431–3450. doi:10.18632/aging.102820

Gao, J., Zhao, R., Xue, Y., Niu, Z., Cui, K., Yu, F., et al. (2013). Role of Enolase-1 in Response to Hypoxia in Breast Cancer: Exploring the Mechanisms of Action. *Oncol. Rep.* 29, 1322–1332. doi:10.3892/or.2013.2269

Gao, S.-S., Wang, Y.-J., Zhang, G.-X., and Zhang, W.-T. (2020). Potential Diagnostic Value of miRNAs in Peripheral Blood for Osteosarcoma: A Meta-Analysis. *J. Bone Oncol.* 23, 100307. doi:10.1016/j.jbo.2020.100307

Goes, C. R., Bertollo, C. M., Zouain, C. S., and Goes, A. M. (2010). Glyceraldehyde-3-phosphate Dehydrogenase as a Surface Associated Antigen on Human Breast Cancer Cell Lines MACL-1 and MGSO-3. *Oncol. Rep.* 24, 677–685. doi:10.3892/or_00000907

Howlader, N., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., et al. (2020). SEER Cancer Statistics Review, 1975-2017. based on November 2019 SEER data submission, posted to the SEER web site, April 2020. National Cancer Institute. Available at: https://seer.cancer.gov/csr/1975_2017/ (Accessed April 28, 2020).

Huang, H., Luo, W., Ni, Y., Sun, S., Wang, C., and Zhang, L. (2020). The Diagnostic Efficiency of Seven Autoantibodies in Lung Cancer. *Eur. J. Cancer Prev.* 29, 315–320. doi:10.1097/CEJ.0000000000000559

Huo, Y., Guo, Z., Gao, X., Liu, Z., Zhang, R., and Qin, X. (2020). Case Study of an Autoantibody Panel for Early Detection of Lung Cancer and Ground-Glass Nodules. *J. Cancer Res. Clin. Oncol.* 146, 3349–3357. doi:10.1007/s00432-020-03309-4

Jiang, H., Ma, N., Shang, Y., Zhou, W., Chen, T., Guan, D., et al. (2017). Triosephosphate Isomerase 1 Suppresses Growth, Migration and Invasion of Hepatocellular Carcinoma Cells. *Biochem. Biophysical Res. Commun.* 482, 1048–1053. doi:10.1016/j.bbrc.2016.11.156

Kansara, M., Teng, M. W., Smyth, M. J., and Thomas, D. M. (2014). Translational Biology of Osteosarcoma. *Nat. Rev. Cancer* 14, 722–735. doi:10.1038/nrc3838

Li, J., Dai, L., Huang, M., Ma, Y., Guo, Z., Wang, X., et al. (2021). Immunoseroproteomic Profiling in Autoantibody to ENO1 as Potential Biomarker in Immunodiagnosis of Osteosarcoma by Serological Proteome Analysis (SERPA) Approach. *Oncoimmunology* 10, 1966969. doi:10.1080/2162402X.2021.1966969

Li, P., Shi, J.-X., Xing, M.-T., Dai, L.-P., Li, J.-T., and Zhang, J.-Y. (2017). Evaluation of Serum Autoantibodies against Tumor-Associated Antigens as Biomarkers in Lung Cancer. *Tumour Biol.* 39, 101042831771166. doi:10.1177/1010428317711662

Lu, H., Goodell, V., and Disis, M. L. (2008). Humoral Immunity Directed against Tumor-Associated Antigens as Potential Biomarkers for the Early Diagnosis of Cancer. *J. Proteome Res.* 7, 1388–1394. doi:10.1021/pr700818f

Ludvigsen, M., Bjerregård Pedersen, M., Lystlund Lauridsen, K., Svenstrup Poulsen, T., Hamilton-Dutoit, S. J., Besenbacher, S., et al. (2018). Proteomic Profiling Identifies Outcome-Predictive Markers in Patients with Peripheral T-Cell Lymphoma, Not Otherwise Specified. *Blood Adv.* 2, 2533–2542. doi:10.1182/bloodadvances.2018019893

Ogino, W., Takeshima, Y., Mori, T., Yanai, T., Hayakawa, A., Akisue, T., et al. (2007). High Level of Ezrin mRNA Expression in an Osteosarcoma Biopsy Sample with Lung Metastasis. *J. Pediatr. Hematol. Oncol.* 29, 435–439. doi:10.1097/MPH.0b013e3180640d18

Okada, R., Otsuka, Y., Wakabayashi, T., Shinoda, M., Aoki, T., Murakami, M., et al. (2020). Six Autoantibodies as Potential Serum Biomarkers of Hepatocellular Carcinoma: a Prospective Multicenter Study. *Int. J. Cancer* 147, 2578–2586. doi:10.1002/ijc.33165

Pan, J., Zheng, Q.-Z., Li, Y., Yu, L.-L., Wu, Q.-W., Zheng, J.-Y., et al. (2019). Discovery and Validation of a Serologic Autoantibody Panel for Early Diagnosis of Esophageal Squamous Cell Carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 28, 1454–1460. doi:10.1158/1055-9965.EPI-18-1269

Qin, J., Wang, S., Shi, J., Ma, Y., Wang, K., Ye, H., et al. (2019). Using Recursive Partitioning Approach to Select Tumor-associated Antigens in Immunodiagnosis of Gastric Adenocarcinoma. *Cancer Sci.* 110, 1829–1841. doi:10.1111/cas.14013

Qiu, C., Wang, P., Wang, B., Shi, J., Wang, X., Li, T., et al. (2020). Establishment and Validation of an Immunodiagnostic Model for Prediction of Breast Cancer. *Oncoimmunology* 9, 1682382. doi:10.1080/2162402X.2019.1682382

Qiu, J., Keyser, B., Lin, Z.-T., and Wu, T. (2018). Autoantibodies as Potential Biomarkers in Breast Cancer. *Biosensors* 8, 67. doi:10.3390/bios8030067

Restifo, N. P., Dudley, M. E., and Rosenberg, S. A. (2012). Adoptive Immunotherapy for Cancer: Harnessing the T Cell Response. *Nat. Rev. Immunol.* 12, 269–281. doi:10.1038/nri3191

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Soussi, T. (2000). p53 Antibodies in the Sera of Patients with Various Types of Cancer: a Review. *Cancer Res.* 60, 1777–1788.

Sparano, J. A., Ostrer, H., and Kenny, P. A. (2013). Translating Genomic Research into Clinical Practice: Promise and Pitfalls. *Am. Soc. Clin. Oncol. Educ. Book*, 15–23. doi:10.14694/edbook_am.2013.33.15

Ta, H. T., Dass, C. R., Choong, P. F. M., and Dunstan, D. E. (2009). Osteosarcoma Treatment: State of the Art. *Cancer Metastasis Rev.* 28, 247–263. doi:10.1007/s10555-009-9186-7

Tan, E. M., and Zhang, J. (2008). Autoantibodies to Tumor-Associated Antigens: Reporters from the Immune System. *Immunol. Rev.* 222, 328–340. doi:10.1111/j.1600-065X.2008.00611.x

Wafa, H., and Grimer, R. J. (2006). Surgical Options and Outcomes in Bone Sarcoma. *Expert Rev. Anticancer Ther.* 6, 239–248. doi:10.1586/14737140.6.2.239

Wang, L., Kang, F.-b., Zhang, G.-c., Wang, J., Xie, M.-f., and Zhang, Y.-z. (2018). Clinical Significance of Serum Soluble B7-H3 in Patients with Osteosarcoma. *Cancer Cel Int* 18, 115. doi:10.1186/s12935-018-0614-z

Wang, P., Qin, J., Ye, H., Li, L., Wang, X., and Zhang, J. (2019). Using a Panel of Multiple Tumor-associated Antigens to Enhance the Autoantibody Detection in the Immunodiagnosis of Ovarian Cancer. *J. Cel. Biochem.* 120, 3091–3100. doi:10.1002/jcb.27497

Wang, S., Qin, J., Ye, H., Wang, K., Shi, J., Ma, Y., et al. (2018). Using a Panel of Multiple Tumor-Associated Antigens to Enhance Autoantibody Detection for

Immunodiagnosis of Gastric Cancer. *Oncoimmunology* 7, e1452582. doi:10.1080/2162402X.2018.1452582

Wang, X., Qin, G., Liang, X., Wang, W., Wang, Z., Liao, D., et al. (2020). Targeting the CK1α/CBX4 axis for Metastasis in Osteosarcoma. *Nat. Commun.* 11, 1141. doi:10.1038/s41467-020-14870-4

Xu, C.-j., Song, J.-f., Su, Y.-x., and Liu, X.-l. (2010). Expression of B-FGF and Endostatin and Their Clinical Significance in Human Osteosarcoma. *Orthop. Surg.* 2, 291–298. doi:10.1111/j.1757-7861.2010.00102.x

Ying, T., Dong, J.-l., Yuan, C., Li, P., and Guo, Q. (2020). The lncRNAs RP1-261G23.7, RP11-69E11.4 and SATB2-AS1 Are a Novel Clinical Signature for Predicting Recurrent Osteosarcoma. *Biosci. Rep.* 40, BSR20191251. doi:10.1042/BSR20191251

Zhang, H.-F., Qin, J.-J., Ren, P.-F., Shi, J.-X., Xia, J.-F., Ye, H., et al. (2016). A Panel of Autoantibodies against Multiple Tumor-Associated Antigens in the Immunodiagnosis of Esophageal Squamous Cell Cancer. *Cancer Immunol. Immunother.* 65, 1233–1242. doi:10.1007/s00262-016-1886-6

Zhang, J.-Y., and Tan, E. M. (2010). Autoantibodies to Tumor-Associated Antigens as Diagnostic Biomarkers in Hepatocellular Carcinoma and Other Solid Tumors. *Expert Rev. Mol. Diagn.* 10, 321–328. doi:10.1586/erm.10.12

Zhang, L., Wang, H., and Dong, X. (2018). Diagnostic Value of α-enolase Expression and Serum α-enolase Autoantibody Levels in Lung Cancer. *J. Bras. Pneumol.* 44, 18–23. doi:10.1590/S1806-37562016000000241

Zhao, Y.-X., Gao, S.-T., Wang, J.-Q., Yao, W.-T., Wang, Y.-S., and Guo, C.-L. (2017). Correlations between Hector Battifora Mesothelial-1 (HBME-1) Expression and Clinical Pathological Characteristics and Prognosis of Osteosarcoma Patients. *Med. Sci. Monit.* 23, 665–672. doi:10.12659/msm.898820

Zhu, W., Li, H., Yu, Y., Chen, J., Chen, X., Ren, F., et al. (2018). Enolase-1 Serves as a Biomarker of Diagnosis and Prognosis in Hepatocellular Carcinoma Patients. *Cmar* 10, 5735–5745. doi:10.2147/CMAR.S182183

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Luo, Wu, Ma, Liang, Luo, Gu, Fan, Hao, Li and Xing. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fgene.2022.864499

Molecular Portrait of GISTs Associated With Clinicopathological Features: A Retrospective Study With Molecular Analysis by a Custom 9-Gene Targeted Next-Generation Sequencing Panel

Haoran Qian[1†], Na Yan[2,3†], Xiaotong Hu[4], Junchang Jiang[4], Zhengzheng Cao[2,3] and Dan Shen[2,3]*

[1]Department of General Surgery, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2]Dian Diagnostics Group Co., Ltd., Hangzhou, China, [3]Key Laboratory of Clinical in Vitro Diagnostic Techniques of Zhejiang Province, Hangzhou, China, [4]Department of Pathology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

Reviewed by:
Kausar Jabbar,
Beaumont Health, United States
Laura La Paglia,
Institute for High Performance Computing and Networking (ICAR), Italy

***Correspondence:**
Dan Shen
shendan@dazd.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 28 January 2022
Accepted: 03 March 2022
Published: 25 April 2022

Citation:
Qian H, Yan N, Hu X, Jiang J, Cao Z and Shen D (2022) Molecular Portrait of GISTs Associated With Clinicopathological Features: A Retrospective Study With Molecular Analysis by a Custom 9-Gene Targeted Next-Generation Sequencing Panel. Front. Genet. 13:864499. doi: 10.3389/fgene.2022.864499

Objectives: The study aims to investigate genetic characterization of molecular targets and clinicopathological features with gastrointestinal stromal tumors based on targeted next-generation sequencing.

Materials and Methods: We selected 106 patients with GISTs from Sir Run Run Shaw Hospital between July 2019 and March 2021. FFPE samples and paired blood samples were obtained from these patients who underwent excision of the tumor. A customized targeted-NGS panel of nine GIST-associated genes was designed to detect variants in the coding regions and the splicing sites of these genes.

Results: In total, 106 patients with a GIST were included in the study which presented with various molecular driver alterations in this study. *KIT* mutations occurred most often in GISTs (94/106, 95.92%), followed by point mutations in *PDGFRA*. *KIT* or *PDGFRA* mutations were detected to be mutually exclusive in the GIST. A total of eight patients with wide-type *KIT/PDGFRA* were characterized as WT-GISTs, according to clinical diagnosis which included six quadruple-WT GISTs, 1 *BRAF*-mutant, and 1 *NF1*-mutant GIST. In *KIT* exon 11, the most common mutation type was the codon Mutation (in-frame deletion or indels), whereas the missense mutation was the dominant type in *KIT* exon 13 and *KIT* exon 17. All variations in *KIT* exon 11 observed in this study were concentrated at a certain position of codon 550 to codon 576. Mutation in *KIT* exon 9 was mostly located at codon 502–503. Two germline pathogenic mutations were detected: NF1-R681* and KRAS-T58I. NF1-L591P was a germline mutation to be

Abbreviations: GIST, gastrointestinal stromal tumor; NCCN, National Comprehensive Cancer Network; EUS-FNAB, endoscopic ultrasound with fine-needle aspiration and biopsy; WT, wild-type (no KIT or PDGFRA somatic mutation detected); KIT, mast/stem cell growth factor receptor kit gene; PDGFRA, platelet-derived growth factor receptor A; TKI, tyrosine kinase inhibitor; NGS, next-generation Sequencing; FFPE, formalin-fixed paraffin-embedded.

identified for the first time and is not recorded in the database. The frequency of driving mutations differed between the primary anatomical site in the GIST ($p = 0.0206$). *KIT* exon 11 mutants had a lower proliferation index of Ki67 (68.66%,≤5%), while 50.00% of *KIT* exon 9 mutants had the Ki67 status greater than 10%.

Conclusion: The occurrence and development of a GIST is driven by different molecular variations. Resistance to TKIs arises mainly with resistance mutations in *KIT* or *PDGFRA* when they are the primary drivers. Targeted NGS can simultaneously and efficiently detect nine GIST-related gene mutations and provide reference for clinicians' individualized diagnosis and treatment. Our results have important implications for clinical management.

Keywords: gastrointestinal stromal tumors, molecular subtypes, clinicopathological features, next-generation sequencing, target therapy

INTRODUCTION

A gastrointestinal stromal tumor (GIST) is the most common mesenchymal malignancy of the gastrointestinal tract which originates from Cajal cells of the digestive tract and account for 3% of gastrointestinal malignant tumors. The most common clinical manifestation of a GIST is a gastric tumor or a bowel tumor. The rectum, colon, esophagus, and other sites are rare (Ducimetière et al., 2011; Boonstra et al., 2018; Florindez and Trent, 2020; Virgilio et al., 2021). Pathological examination is the most reliable method for the diagnosis of a GIST (Blay et al., 2021a). National Comprehensive Cancer Network (NCCN) guidelines recommend endoscopic ultrasound with fine-needle aspiration and biopsy (EUS-FNAB) as the first choice (Zhao et al., 2020). The expression of tumor markers in tumor tissues was detected by immunohistochemistry (Fernández et al., 2018). c-Kit (stem cell growth factor receptor, CD117) is a protein encoded by the *KIT* gene in humans. CD117 is the most important IHC marker which is expressed in 85–95% of GISTs (Ramdani et al., 2020). DOG1 is a useful marker for these tumors that the GIST does not express *KIT* on IHC. Other tumor markers include CD34, smooth muscle action (SMA), etc., (Luo et al., 2004; Malik et al., 2019; Bradea et al., 2021).

In total, 60–70% GISTs can acquire mutation of *KIT*, and 10–15% GISTs acquire mutation of platelet-derived growth factor receptor A (PDGFRA) that both promote to the occurrence and development of GISTs (Joensuu et al., 2013; Joensuu et al., 2015; Nishida et al., 2016). *KIT* and *PDGFRA* mutations play a crucial role in the pathogenesis of GISTs (Demetri et al., 2006). Chinese consensus guidelines for diagnosis and management of gastrointestinal stromal tumor recommend *KIT/PDGFRA* gene testing for CD117/DOG1-negative GIST patients, which acts as a supplement for immunohistochemical diagnosis (Li et al., 2017).

GIST is divided into three types at the molecular level based on the mutations of *KIT* and *PDGFRA*: GIST with *KIT* mutations, GIST with *PDGFRA* mutations, and non-*KIT or PDGFRA* somatic mutation (WT-GIST) (Daniels et al., 2011). WT-GIST is complex due to the existence of different subgroups with distinct molecular hallmarks. About 30% of WT-GISTs show deletion mutations of succinate dehydrogenase subunit A (SDHA) (Boikos et al., 2016). Other molecular hallmarks include mutations of neurofibromatosis type 1 (NF1), *BRAF*, or *RAS* (Corless et al., 2011a; Blay et al., 2021a). It follows that GIST is a cancer with comparatively small genetic heterogeneity. The cancer-driven pathway of a GIST is a downstream signaling pathway mediated by KIT/PDGFRA receptors (Blay et al., 2021b). The precise treatment of the cancer gene map for GISTs has become increasingly mature.

The tyrosine kinase inhibitor (TKI) imatinib is a model of targeted therapy for GISTs which can be used to treat GISTs with *KIT/PDGFRA* mutation (Park et al., 2014; Gang and Wang, 2018). However, the therapeutic response and dosage of GIST to tyrosine kinase inhibitors are closely bound up with molecular subtypes (Blay et al., 2021b). Specifically, the *KIT* exon 11 mutation is more responsive to imatinib treatment than the *KIT* exon 9 mutation or WT-GISTs. The *KIT* exon 9 mutation requires double dose of imatinib (800 mg/d) (Nishida et al., 2015; Reichardt, 2016). KIT-V654A and KIT-T670I mutations are resistant to imatinib (Guo et al., 2007). PDGFRA-D842V mutations are also characterized broadly as imatinib resistance mutations which can adjust the drug treatment strategy to 300 mg/d avapritinib (Jun et al., 2018; Heinrich et al., 2020; Smrke et al., 2020; Jones et al., 2021). Therefore, it is necessary to understand the molecular characteristics before tyrosine kinase inhibitor treatment to ensure the optimal treatment strategy. The purpose of this study was to investigate the relationship between the molecular variation and clinicopathological features in GIST patients by targeted next-generation sequencing (NGS), in order to deepen the understanding of GIST-individualized treatment.

MATERIALS AND METHODS

Patients and Tumor Samples

A total of 106 solid tumor samples from gastrointestinal stroma and paired blood samples were analyzed by a custom 9-gene targeted next-generation sequencing panel, which was obtained from Run Run Shaw Hospital between July 2019 and May 2021. All specimens were pathologically and immunohistochemically confirmed as GISTs. Tumor cells accounted for more than 20% of the tumor population. Clinical data of all patients were collected

and sorted out on the basis of age, gender, tumor location, tumor size, mitotic count, immunohistochemical detection index (CD117, CD34, DOG1, S-100, Ki-67, SDHB, SMA, and desmin), etc., (**Supplementary Table S1**). This study was approved by the internal review board of the Run Run Shaw Hospital.

Sample Preparation and DNA Extraction

The pathologist performed a histological assessment with hematoxylin and eosin-stained sections to confirm the tumor purity. Then, the tumor areas of the FFPE sections were macrodissected. Tumor cells accounted for more than 20% of the tumor population. Genomic DNA from FFPE samples of the GIST were extracted by using a QIAamp DNA FFPE Tissue Kit (QIAGEN, Dusseldorf, Germany), following the manufacturer's instructions. Paired blood samples of GIST were extracted using a QIAamp DNA Blood Midi Kit (QIAGEN, Dusseldorf, Germany), according to manufacturer's instructions. The DNA concentration was measured using a Qubit 3.0 (Thermo Fisher Scientific, Waltham, United States) fluorometer. The size distribution of DNA was analyzed using a Qsep100 (Bioptic, Taiwan, China) system.

Next-Generation Sequencing Library Preparation

This study used the CleanPlex™ (Paragon Genomics, Silicon Valley, United States) panel with an optimized manufacturer's protocol to prepare sequencing libraries. 40 ng of genomic DNA was enriched in the target region of nine GIST-related genes by multiplex PCR and then ligated with indexed sequencing adapters sequentially. Purification of DNA libraries used Agencourt AMPure XP beads (Beckman Coulter, United States). The purified NGS library was quantified using the Qubit 1×dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, United States), and its fragment size distribution was analyzed using a Qsep100 (Bioptic, Taiwan, China) system. A nine GIST-related gene panel was used to identify one or more single-nucleotide variants, insertions, deletions, duplications, and delin mutations. The nine genes include *KIT*, *PDFRA*, *KRAS*, *BRAF*, *NF1*, *SDHA*, *SDHB*, *SDHC*, and *SDHD*.

Sequencing and Bioinformatics Analysis

In order to ensure the reliability and validity of the experimental results, all library construction and sequencing were completed in a CAP-certified laboratory. Sequencing was performed on the Illumina NextSeq 500 platform (Illumina, San Diego, United States). The mean coverage depth is approximately >1000X for the tumor samples and >30X for the paired blood samples. A minimal variant frequency of 5% was designated as a mutation. The paired-end sequencing data of the libraries in the FASTQ format were mapped to the human genome (hg19) by the Burrows–Wheeler Aligner (BWA-MEM). MuTect19 with default parameters was applied to paired blood and tumor BAM files for identification of somatic single-nucleotide variants (SNV). Small insertions and deletions (indels) were detected by SCALPEL. SNV and indel annotation was performed by ANNOVAR21 using the hg19 reference genome and 2014 versions of standard databases and functional prediction programs.

Statistical Analysis

The statistical package stats of R version 4.1.1 software were used for statistical analysis. Continuous variables were reported as mean and standard deviation or median and interquartile range and compared by using the Student t test or Mann–Whitney U test. Chi-square analysis was performed toward analyses of subgroups, and Fisher's exact test was used in cases of small numbers. All tests were 2-sided, with $p \leq 0.05$ as the criterion standard for determining significance. Structural changes induced by amino acid substitution were predicted by Missense3D software. A clustering correlation heatmap with signs was performed using the Kendall correlation analysis. GO enrichment analysis and KEGG pathway analysis were established with OmicShare tools.

RESULTS

Clinicopathological Characteristics of Patients

The total of 106 patients were included in this study, of which 73 were primary tumors (68.87%) and 33 were recurrent disease (31.13%) (**Table 1**). Gender disparity in GIST incidence was not observed as the male-to-female ratio is 1.3:1 (60 male and 46 female patients). The age of the first operation composition ranged from 35 to 89 with a median age of 58. GIST arose in the non-stomach (64, 60.38%) sites more than the stomach sites (42, 39.62%). Location of the disease disparity in GIST incidence was not observed as the stomach-to-non-stomach ratio is 1:1.03. The level of the mitotic phase in the stomach is higher (Mitotic count [x/HPF] > 5, 32.43%), which is different from that of non-gastric primary GIST (Mitotic count [x/HPF] > 5, 8.33% ; p-value = 0.0389). There was no significant difference between the location of the primary focus and tumor size (**Supplementary Figure S1**, **Supplementary Table S1**).

Gene Mutation Distributions and Frequencies

Mutations in 106 patients with primary and recurrent GISTs have a preference (**Figure 1**, **Supplementary Figure S2**, **Supplementary Table S1**). *KIT* mutations occurred most often in GISTs (94 of 106 tumors, 95.92%), followed by the point mutation in *PDGFRA*. This predominant preference presented both in primary and recurrent GISTs (**Supplementary Table S1**, **Figure 1**, **Supplementary Figure S2**). *KIT* mutations or *PDGFRA* mutations were detected to be mutually exclusive here, which is in consistency with previous reports (von Mehren and Joensuu, 2018a; Blay et al., 2021a). A total of eight patients with wide-type *KIT/PDGFRA* were characterized as WT-GISTs, according to clinical diagnosis. These include six quadruple-WT (*KIT/PDGFRA/SDH/RAS*-WT) GISTs, 1 *BRAF*-mutant, and 1 *NF1*-mutant GIST.

TABLE 1 | Demographic and clinical characteristics of 106 patients with GIST.

Characteristics	No. of patients	%	Characteristic	No. of patients	%
Gender			**CD34**		
Male	60	56.60	positive	84	79.25
Female	46	43.40	negative	12	11.32
			unknown	10	9.43
Mean age of first operation [years] (range)		58 (35–89)	**SMA**		
≤55	37	34.91	positive	9	8.49
>55	69	65.09	negative	86	81.13
			unknown	11	10.38
Location of primary GIST			**S-100**		
Stomach	42	39.62	positive	1	0.94
Others	64	60.38	negative	95	89.62
			unknown	10	9.43
Tumor size (cm)			**Desmin**		
≤2	6	5.66	positive	2	1.89
>2,<5	41	38.68	negative	93	87.74
≥5	58	54.72	unknown	11	10.38
unknown	1	0.94			
Mitotic count [x/HPF]			**SDHB**		
≤5/HPF	57	53.77	positive	90	84.91
>5/HPF	29	27.36	negative	4	3.77
unknown	20	18.87	unknown	12	11.32
CD117			**Proliferation index of Ki67 (%)**		
positive	94	88.68	≤5	60	56.60
negative	3	2.83	>5,< 10	11	10.38
unknown	9	8.49	≥10	25	23.58
DOG1					
positive	85	80.19	**Disease status**		
negative	12	11.32	primary tumors	73	68.87
unknown	9	8.49	recurrent disease	33	31.13

It is worth noting that none of the first-episode patients developed resistance mutations, except for one patient who had been treated with IM 400 mg neoadjuvant therapy for 25 months preoperatively. This patient only detected *KIT* exon 11 p.W557_K558 deletion before IM therapy (**Supplementary Table S1**). However, we found secondary resistance mutations (*KIT* exon 13 p.V654A, exon 17 p.D820V, p.N822K/Y, and p.Y823D) in 13 recurrent patients (**Figure 2**, **Supplementary Table S1**, **Supplementary Figure S2**). A total of 14 patients with resistant mutations were detected in this study, and 92.86% of them relapsed.

KIT-Mutant GISTs

KIT mutations including deletion, deletion-insertion (indels), duplication, and missense mutations occurred mostly in exon 11 (74 of 94 *KIT*-mutant GISTs), followed by exon 9, exon 13, exon 17, and exon 4 (**Supplementary Figure S2, S3A**). In exon 11, the most common mutation type was the codon mutation (in-frame deletion or indels), whereas the missense mutation was the dominant type in exon 13 and exon 17 (**Supplementary Figure S2**). This preference of variant type in different exons of *KIT* was statistically significant (p = 6.011e-07) (**Table 2**).

The deletions of codon 557–558 (c.1669-1674del) in exon 11 of *KIT* were 63.04% in 46 patients, which is associated with the malignant behavior as mentioned in reports (Martin-Broto et al., 2010; Joensuu et al., 2015). The next most frequent mutation in exon 11 of *KIT* was the missense mutation of codon 559 (c.1676T > A/C/G). All variations in *KIT* exon 11 observed in this study were concentrated at a certain position of codon 550 to codon 576 (**Supplementary Table S1**), which may be defined as the hotspot region of GISTs.

Mutation in *KIT* exon 9 mostly located at codon 502–503. The duplication insertion of A502-Y503 codons account for 93.33% (14 of 15 GIST) GISTs, which was identified a variation in *KIT* exon 9. Besides, a variant at codon 476 (c.1427G > T, p.Ser476Ile) of unknown significance in *KIT* exon 9 was found in this study. Structural changes induced by amino acid substitution were predicted by Missense3D software. The results showed that the variation of Ser to Ile at position 476 resulted in hydrogen bond damage and atomic collision with surrounding amino acid residues. It was inferred that KIT-S476I may be a pathogenic mutation (**Figure 3**). Among 15 GIST patients who harbored the *KIT* exon 9 mutation, nine (60%) were recurrence patients, indicating a higher risk of relapse after surgical excision of exon 9-mutated GIST patients.

The missense mutation affecting codon V654 (c.1961T > C) in exon 13 of *KIT* was identified only in recurrence GISTs in this study. This finding is consistent with the previous understanding that V654A in exon 13 is the secondary resistance mutation acquired under the therapeutic pressure of a TKI (Nishida et al., 2008; Serrano et al., 2019). Codon L642 (c.1924A > G) in exon 13 presented in two primary GIST patients in this study.

In exon 17 of *KIT*, D820A, N822K/Y, and Y823D were observed in seven recurrent patients and one primary patient (**Figure 2**). As the secondary mutation, the activation loop mutation in exon 17 (e.g., D820A) stabilizes the active

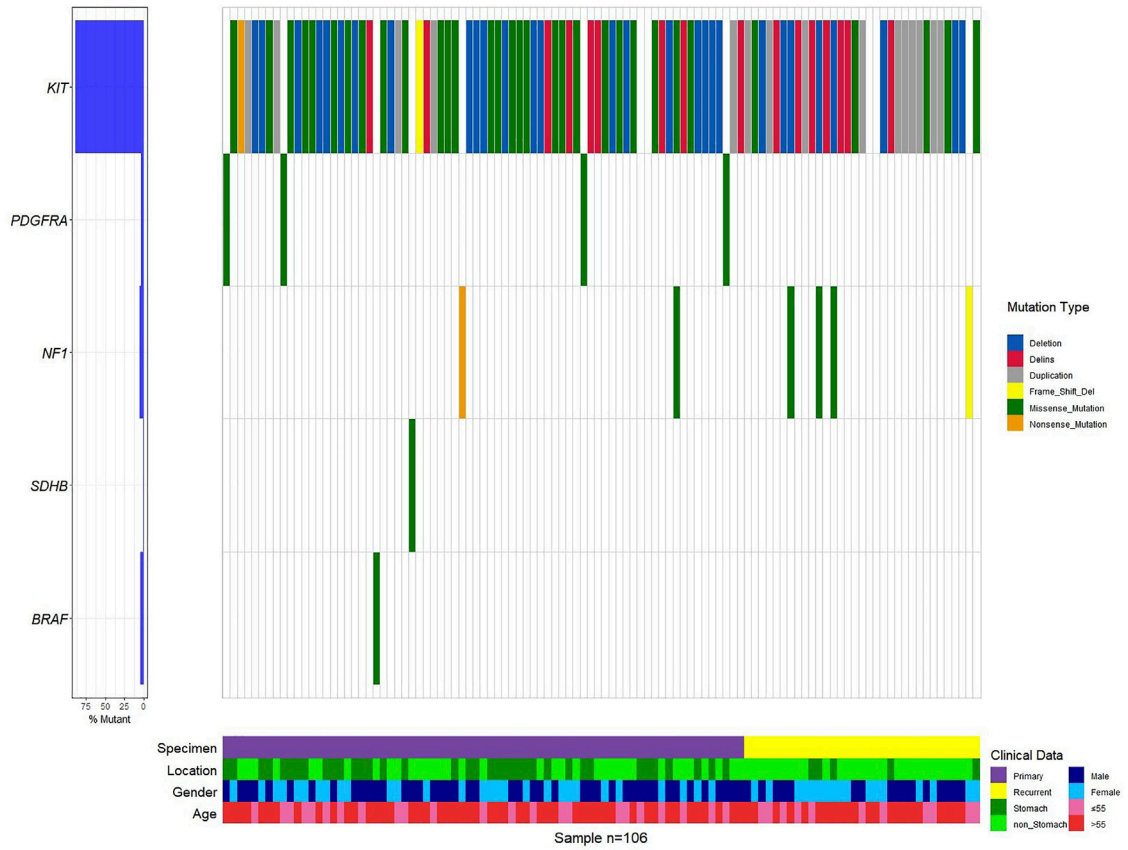


FIGURE 1 | Genetic profile of the GIST.

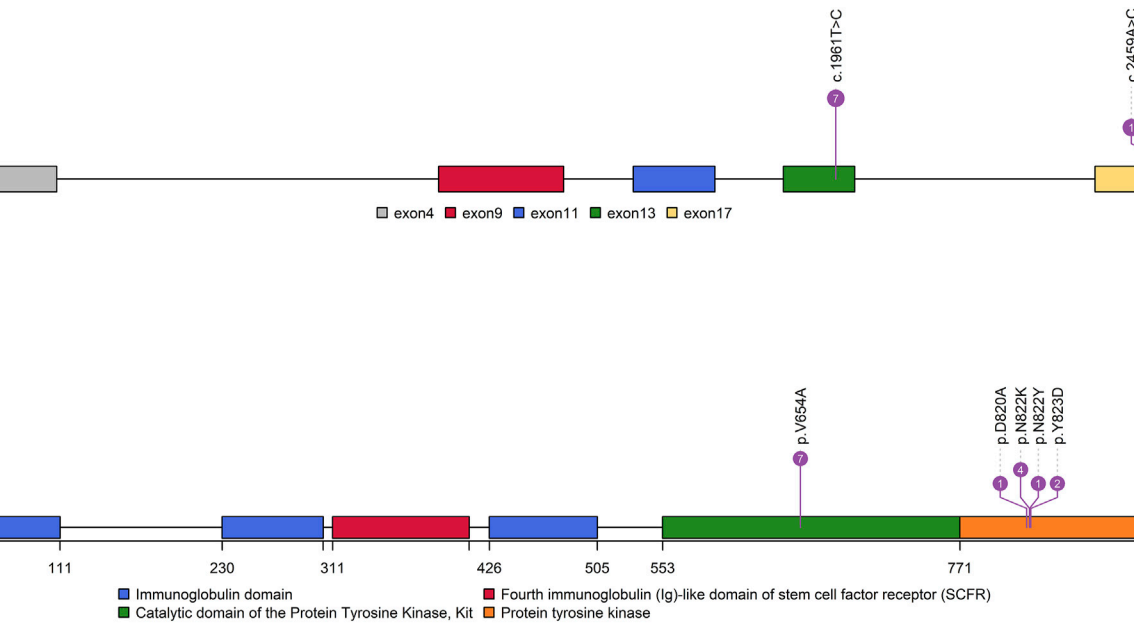


FIGURE 2 | Distribution characteristics of *KIT*-resistant mutations. **(A)**: Distribution of resistant mutations in the *KIT* gene. **(B)**: Distribution of resistant mutations in KIT functional domains.

TABLE 2 | The genomic features of 106 GIST patients.

Features		Variation type					
			Point mutation		Codon mutation		
		Number of patients %	N	%	N	%	*p* Value
Wild-type	8	7.55%	3	37.50%	1	12.50%	0.4900
Mutation	98	92.45%	49	50.00%	61	62.24%	
KIT	94	95.92%	44	46.81%	61	64.89%	6.011e-07
exon 4	1	1.06%	0	/	1	1.06%	
exon 9	15	15.96%	1	6.67%	14	14.89%	
exon 11	74	78.72%	28	37.84%	48	51.06%	
exon 13	9	9.57%	9	100.00%	0	/	
exon 17	7	7.45%	7	100.00%	0	/	
PDGFRA	4	4.08%	4	100.00%	0	/	
exon 18	4	100.00%	4	100.00%	0	/	

formation of KIT kinase, thus contributing resistance of TKI. The mutation observed in exon 17 in our study was also considered as the cause for relapse.

PDGFRA-Mutant GISTs

PDGFRA-D842V was observed in all (4/4) of GISTs with *PDGFRA* alterations in this study (**Supplementary Table S1**). It means that these patients cannot benefit from imatinib treatment.

Germline Mutation in GISTs

Two germline pathogenic mutations were detected, NF1-R681* and KRAS-T58I. We also obtained eight germline mutations of unknown clinical significance, SDHA-Y211H, SDHC-V9I, SDHB-A6S, BRAF-G69A, PDGFRA-S716R/D756N, and NF1-R2713Q/Y1292C. NF1-L591P was a germline mutation to be identified for the first time and is not recorded in the database (**Supplementary Table S2**).

Associations Between the Molecular Subgroup and Clinicopathological Characteristics of GISTs

The frequency of driving mutations differed between primary anatomical sites in GISTs ($p = 0.0321$) (**Table 3**). GISTs bearing *KIT* exon11 mutations were located most often in the gastrointestinal tract (53.6%, 30/56) and small intestine (35.7%, 20/56). Mutations in exon 9 of *KIT*-mutant GISTs all happened in the small intestine (100%, 6/6) in our cohort. In the primary rectal GIST, the most frequent driver mutations occurred in *KIT* exon 11 (100%, 5/5). All GISTs harboring the PDGFRA-D842V mutation had a unique gastric primary localization. The wide-type *KIT*/*PDGFRA* GISTs arose in the small intestine and stomach.

There was a significant difference in the *KIT* mutation rate among GIST patients in the Ki67 status group of ≤5, group of >5 and <10, and group of ≥10 ($p = 0.001333$) (**Table 4**). *KIT* exon11 mutants had a lower proliferation index of Ki67 (68.66%,≤5%), while 50.00% of *KIT* exon 9 mutants had a Ki67 status greater than 10%. Mutations in KIT 9 exon were more likely to occur in non-gastric areas (**Table 4**).

In practice, the diagnosis of GISTs is mainly based on IHC markers, including CD117, CD34, DOG1, and SDHB, and with the help of genetic analysis. Almost all GISTs overexpress CD117 or DOG1 with one exception in our cohort (**Supplementary Figure S4**). However, the CD117 expression and KIT mutation was not entirely concordant. There was one *KIT*-mutant patient who was CD117-negative but DOG1-positive. DOG1 was universally expressed in three GISTs with *PDGFRA* mutations, while one *PDGFRA*-mutant patient failed to get IHC information. The significance test of the Kendall correlation coefficient showed that the *KIT* mutation was positively correlated with CD117 ($p < 0.05$), while the *PDGFR* mutation was negatively correlated with CD34 ($p < 0.01$) (**Figure 4**).

The relationship between clinical features and the immunophenotype was depicted by **Figure 4**. CD117 ($p < 0.05$) and Ki67 ($p < 0.001$) were positively correlated with patients' mitotic count. Meanwhile, the age ($p < 0.05$) and tumor size ($p < 0.01$) were positively correlated with Ki67. In addition, the relevance between the tumor location and DOG1 was also positive. GO enrichment analysis was performed on nine GIST-related genes. Biological processes have been enriched by GO analysis. It includes cell proliferation, reproduction, developmental process, and reproductive process, which are related to the growth of cells. This may explain why there are significant correlations among CD117, Ki67, and mitotic count (**Figure 5**). Meanwhile, KEGG pathway analysis showed that nine GIST-related genes were enriched in cell growth and death, development, aging, environmental adaptation, etc. Perhaps this is why the tumor size is significantly correlated with the Ki67 index (**Figure 6**).

DISCUSSION

In the past 10 years, the remedy of GISTs has gradually been evolving from a one-size-fits-all scheme to targeted oncogene treatments for specific molecular GIST subtypes (Blay et al., 2021a). However, the effectiveness of targeted therapy varies among patients because its effectiveness depends on the genetic mutation profile of GIST tumor tissues (Debiec-Rychter et al., 2006). A variety of molecular driven mutations are present in the GIST and are directly related to the curative effect of targeted treatment (von Mehren and Joensuu, 2018b). The most frequent driver mutations occur in *KIT* and *PDGFRA*. In total, 60–70% GISTs can acquire the mutation of *KIT*, and 10–15% GISTs acquire mutation of *PDGFRA* that both promote

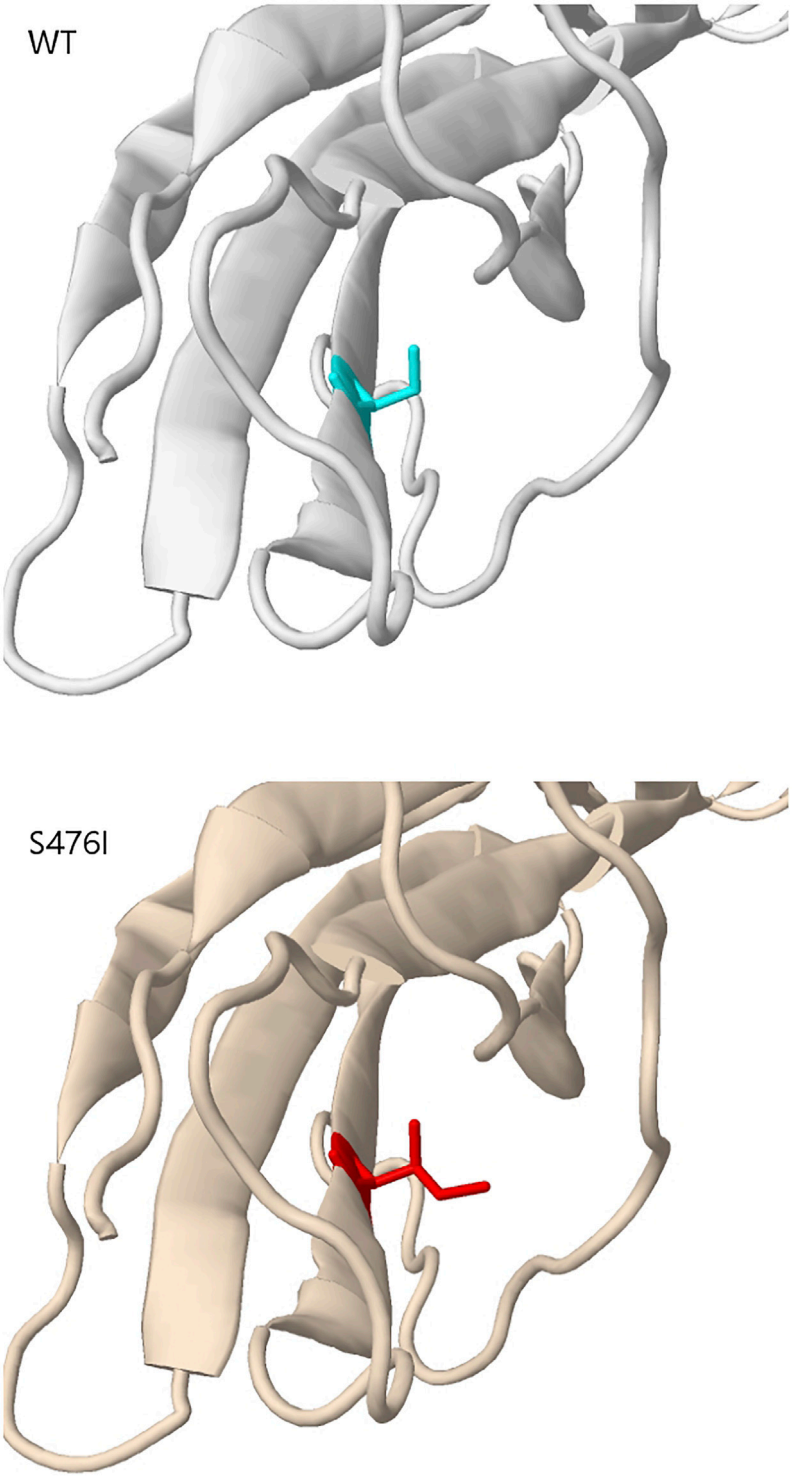


FIGURE 3 | Missense3D predicts the tertiary structure changes introduced by KIT p.Ser476Ile.

to the occurrence and development of GISTs. Around 15% of GISTs have other genetic alterations, for example, in SDH family genes, RAS family genes, *BRAF*, *NF1*, or other very rare driver gene mutations (von Mehren and Joensuu, 2018b; Blay et al., 2021b). Therefore, it is necessary to analyze the gene mutation profile of tumor tissues. Although the cost of molecular detection needs to be paid for additionally, the cost of molecular detection is lower than unnecessary adjuvant or neoadjuvant treatment.

TABLE 3 | Clinicopathological feature correlation with molecular classification.

Characteristics	KIT	PDGFRA	Wild-type	*p* Value
Gender				
Male	53	3	4	0.7065
Female	41	1	4	
Age				
≤55	33	2	2	0.6879
>55	61	2	6	
Location of GIST				
Stomach	36	4	2	0.0321
Others	58	0	6	
Tumor size (cm)				
≤2	4	1	1	0.2187
>2,<5	36	1	4	
≥5	53	2	3	
Mitotic count [x/HPF]				
≤5/HPF	50	3	1	0.0563
>5/HPF	28	0	4	
CD117				
positive	86	3	5	0.0274
negative	1	0	2	
DOG1				
positive	77	3	5	0.4762
negative	10	0	2	
CD34				
positive	77	1	6	0.03637
negative	9	2	1	
SMA				
positive	9	0	0	1
negative	76	3	7	
S-100				
positive	1	0	0	1
negative	85	3	7	
Desmin				
positive	2	0	0	1
negative	83	3	7	
SDHB				
positive	81	3	6	0.3673
negative	3	0	1	
Proliferation index of Ki67 (%)				
≤5	52	3	5	0.7805
>5,< 10	10	0	1	
≥10	24	0	1	

In this study, the molecular detection method we chose was targeted next-generation sequencing for nine GIST-related genes. Different from the traditional Sanger sequencing, NGS sequencing can detect molecular variations in multiple genes at the same time. Its outstanding advantage is higher detection sensitivity. NGS sequencing can capture the secondary drug resistance mutation earlier than Sanger sequencing that provides more valuable reference information for the treatment strategies of advanced patients. We found that four patients carried PDGFRA-D842V in 73 primary patients. The aforementioned information suggests that patients should be cautious when choosing IM for treatment. Briefly, in the current study, a gene panel consisting of 158 CDS regions in nine genes, which have clinical interest for the GIST, was tested for targeted sequencing. The genetic profiles of 106 GIST patients were comprehensively analyzed by a 9-gene targeted next-generation sequencing panel. It shows that *KIT* was the most frequent driver gene. Of all patients, 13.21% (14/106) patients had two or more mutations. *KIT* mutations or *PDGFRA* mutations were detected to be mutually exclusive in the GIST. In addition, the significance test of the Kendall correlation coefficient showed that the *KIT* mutation was negatively correlated with the *BRAF* mutation and *SDHB* mutation ($p < 0.01$). Although the analysis results showed a significant correlation, they could not represent the actual clinical situation because there was only one case of *BRAF* and *SDHB* mutation respectively. However, we found similar results in other studies (Corless et al., 2011b; Kondo et al., 2020). Different types of molecular tumor drivers are related to the primary anatomical site. GIST-bearing *KIT* mutations were located in the small intestine (51/94,54.26%) or stomach (36/94,38.30%). All of *KIT* exon 9 mutations were not observed in the stomach. Mutations of *PDGFRA* were observed in the primary gastric GIST.

Our study identified that 88.68% (94/106) of the cases carried *KIT* mutations, 3.77% (4/106) of cases harbored *PDGFRA* mutations, and 7.56% (8/106) of cases were characterized as WT-GISTs. The *KIT* mutation incidence was higher than that reported in a previously published literature study, and the *PDGFRA* mutation incidence was lower than others' published literature which could be attributed to different sample sizes, populations, and detection sensitivity. The *KIT* mutation was distributed in exons 4, 9, 11, 13, and 17 that contain different types of variation, such as point mutation, duplication, and small fragment deletions. Exon 11 (74/94, 78.72%) accounted for the highest proportion of mutations. The codon mutation (61/94, 64.89%) was the most common variant carried in *KIT* exon 11. In total, 47.54% (29/61) of patients were involved in 557/558 deletion. All patients with the *KIT* exon 9 mutation carried p.Ala502_Tyr503 duplicate mutation except for one missense mutation (p.Ser476Ile) case.

TABLE 4 | Features of KIT exons.

Features	Number of patients	%	Location of GIST		Mitotic count [x/HPF]		Tumor size (cm)			Proliferation index of Ki67 (%)		
			Stomach	Non-stomach	≤5/ HPF	>5/ HPF	≤2	>2,<5	≥5	≤5	>5,< 10	≥10
KIT	94								p-value = 0.000209	p-value = 0.1069	p-value = 0.3118	p-value = 0.001333
exon 4	1	1.06	0	1	1	0	1	0	0	1	0	0
exon 9	15	15.96	0	15	6	7	0	5	10	4	3	7
exon 11	74	78.72	35	39	42	18	3	29	41	46	7	14
exon 13	9	9.57	3	6	4	5	0	2	7	5	1	1
exon 17	7	7.45	0	7	2	4	0	3	4	0	3	4

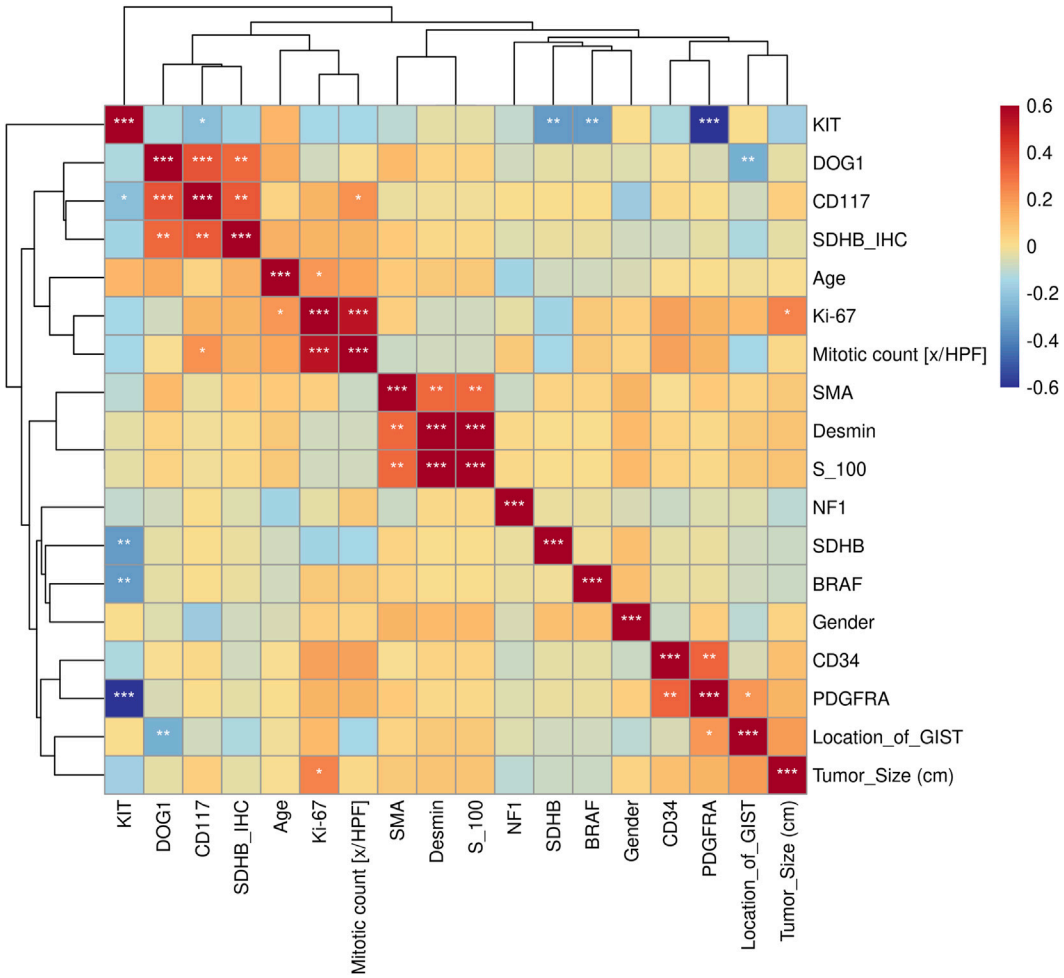


FIGURE 4 | Kendall correlation analysis between gene mutations and clinical features. Test of significance of the Kendall correlation coefficient: "*", "**", "***" represent $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

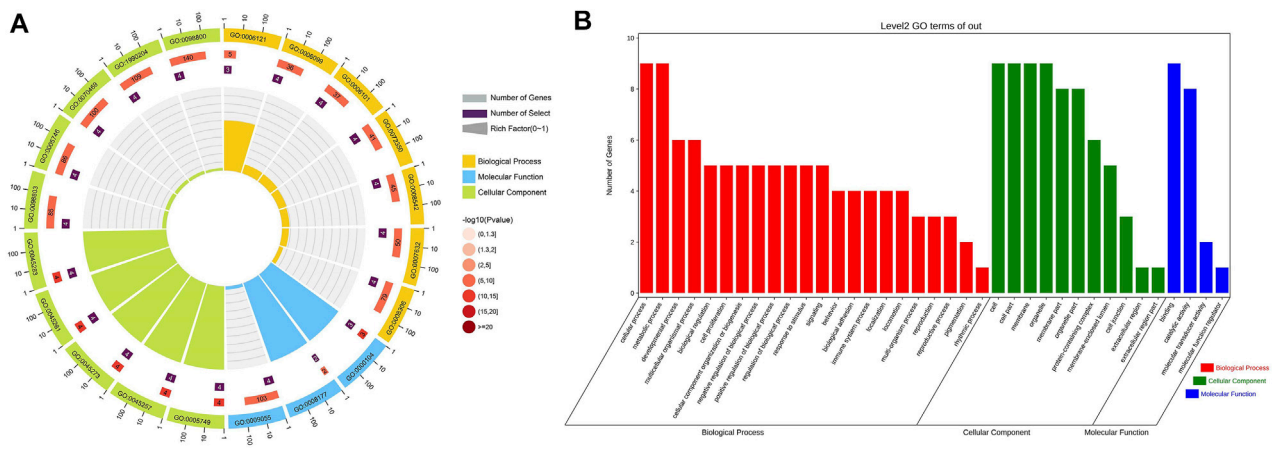


FIGURE 5 | GO functional classification results of nine genes related to GIST **(A)** Circle Diagram of GO Molecular Function Enrichment **(B)** Bar Chart of GO Molecular Function Enrichment.

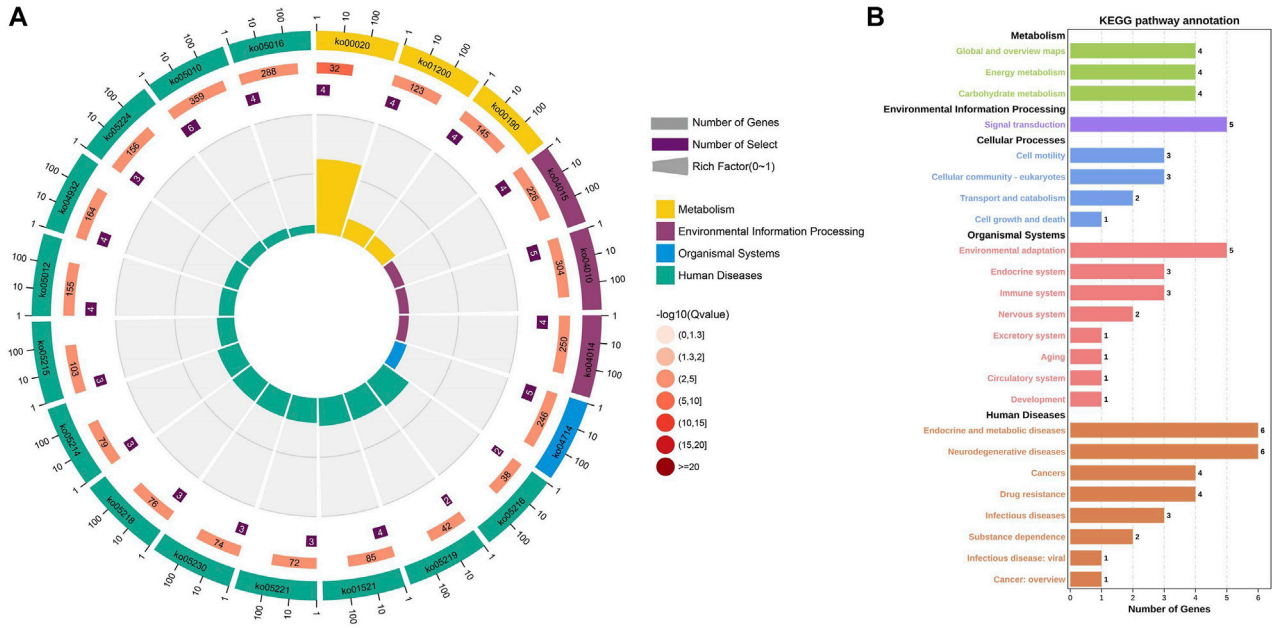


FIGURE 6 | KEGG pathway enrichment analysis of nine genes related to GIST **(A)** Circle Diagram of KEGG pathway enrichment analysis **(B)** Statistics of the enrichment number of the KEGG pathway.

Lincoln SE et al. discovered that the pathogenic germline mutations of tumor susceptibility genes had potential clinical effects, such as clinical trial qualification and earlier detection or prevention of the second primary cancer (Lincoln et al., 2020). We set up a synchronous detection of germline variation to this study and identified different pathogenic germline mutations in two patients from 106 Chinese patients with the GIST. One patient was a wild-type patient without somatic mutation, but one pathogenic germline mutation of *NF1* exon 18 p.R681 * was tested. The other patient detected somatic mutations of *KIT* exon 9 p.A502_Y503dup and exon 17 p.N822Y and developed liver and abdominal metastasis 2 years after initial surgery and IM treatment. The pathogenic mutation KRAS-T58I was detected in the germline molecular assay of this patient. In addition, we also obtained two germline mutations of unknown clinical significance, PDGFRA-S716R and NF1-R2713Q. NF1-L591P was a germline mutation to be identified for the first time and is not recorded in the database.

It is a great pity that there were several limitations to our study. The 9-gene targeted next-generation sequencing panel used in this study was too small for excavation of the complicated genetic alterations in the GIST. Therefore, according to the specific application scenario, there is space for upgrading and optimizing the analysis of the GIST disease occurrence and development and drug efficacy evaluation. Second, in our study, the relapse-free or disease-free survival analysis was not applicable due to the short median follow-up duration. Finally, the functional evaluation of the germline gene variation has not taken a closer study.

In conclusion, our study confirms the utility of the 9-gene targeted next-generation sequencing panel to efficiently identify mutations associated with GISTs. NGS can effectively expand our understanding about the specific mutations of sensitivity in individualized treatment. The occurrence and development of GIST is driven by different molecular variations. Resistance to TKIs arises mainly with resistance mutations in *KIT* or *PDGFRA* which may provide a genetic basis for developing new GIST therapeutic drugs. Our results have important implications for clinical management that supplies reference for clinicians' individualized diagnosis and treatment.

DATA AVAILABILITY STATEMENT

The raw data reported in this study has been deposited in the CNGB Nucleotide Sequence Archive (CNSA) (CNSA: https://db.cngb.org/cnsa/) with accession No. CNP0002760.

AUTHOR CONTRIBUTIONS

Conception and design: DS, HQ, and NY. Study development and methods: HQ and NY. Medical and technical support: HQ, NY, and XH. Collection and assembly of data: ZC and JJ. Data analysis and interpretation: NY, DS, and HQ. Manuscript writing: NY, DS, and HQ. Final approval of the manuscript: all authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.864499/full#supplementary-material

REFERENCES

Blay, J-Y., Kang, Y-K., Nishida, T., and von Mehren, M. (2021). Gastrointestinal Stromal Tumours. *Nat. Rev. Dis. Primers* 7 (1), 22. doi:10.1038/s41572-021-00254-5

Blay, J. Y., Kang, Y. K., Nishida, T., and Mehren, M. V. (2021). Gastrointestinal Stromal Tumours. *Nat. Rev. Dis. Primers* 7 (1), 22. doi:10.1038/s41572-021-00254-5

Boikos, S. A., Pappo, A. S., Killian, J. K., Laquaglia, M. P., Weldon, C. B., George, S., et al. (2016). Molecular Subtypes of KIT/PDGFRA Wild-type Gastrointestinal Stromal Tumors: A Report from the National Institutes of Health Gastrointestinal Stromal Tumor Clinic. *Jama Oncol.* 2 (7), 922–928. doi:10.1001/jamaoncol.2016.0256

Boonstra, P. A., Steeghs, N., Farag, S., Coevorden, F. V., Gelderblom, H., Grunhagen, D. J., et al. (2018). Surgical and Medical Management of Small Bowel Gastrointestinal Stromal Tumors: A Report of the Dutch GIST Registry. *Eur. J. Surg. Oncol.* 45 (3), 410–415. doi:10.1016/j.ejso.2018.09.013

Bradea, C., Lupascu, C., Munteanu, V., Ciobanu, D., and Vasilescu, A. (2021). Gastric GIST with Progressive Mitotic index. *Jurnalul de Chirurgie* 17 (1), 46–51. doi:10.7438/jsurg.2021.01.06

Corless, C., Barnett, C., and Heinrich, M. (2011). Gastrointestinal Stromal Tumours: Origin and Molecular Oncology. *Nat. Rev. Cancer* 11 (12), 865–878. doi:10.1038/nrc3143

Corless, C. L., Barnett, C. M., Heinrich, M. C., Corless, C. L., and Barnett, C. M. (2011). Heinrich MCGastrointestinal Stromal Tumours: Origin and Molecular Oncology. Nat Rev Cancer 11: 865-878. *Nat. Rev. Cancer* 11 (12), 865–878. doi:10.1038/nrc3143

Daniels, M., Lurkin, I., Pauli, R., Erbstösser, E., Hildebrandt, U., Hellwig, K., et al. (2011). Spectrum of KIT/PDGFRA/BRAF Mutations and Phosphatidylinositol-3-Kinase Pathway Gene Alterations in Gastrointestinal Stromal Tumors (GIST) - ScienceDirect. *Cancer Lett.* 312 (1), 43–54. doi:10.1016/j.canlet.2011.07.029

Debiec-Rychter, M., Sciot, R., Le Cesne, A., Schlemmer, M., Hohenberger, P., van Oosterom, A., et al. (2006). KIT Mutations and Dose Selection for Imatinib in Patients with Advanced Gastrointestinal Stromal Tumours. *Eur. J. Cancer* 42 (8), 1093–1103. doi:10.1016/j.ejca.2006.01.030

Demetri, G. D., van Oosterom, A. T., Garrett, C. R., Blackstein, M. E., Shah, M. H., Verweij, J., et al. (2006). Efficacy and Safety of Sunitinib in Patients with Advanced Gastrointestinal Stromal Tumour after Failure of Imatinib: a Randomised Controlled Trial. *Lancet (London, England)* 368 (9544), 1329–1338. doi:10.1016/s0140-6736(06)69446-4

Ducimetière, F., Lurkin, A., Ranchère-Vince, D., Decouvelaere, A. V., Péoc'h, M., Istier, L., et al. (2011). Incidence of Sarcoma Histotypes and Molecular Subtypes in a Prospective Epidemiological Study with Central Pathology Review and Molecular Testing. *Plos One* 6 (8), e20294. doi:10.1371/journal.pone.0020294

Fernández, J., Gómez-Ruiz Á, G. J., Olivares, V., Ferri, B., Frutos, M., Soria, T., et al. (2018). Clinical and Pathological Features of "small" GIST (≤2 Cm). What Is Their Prognostic Value? *Eur. J. Surg. Oncol. : J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol.* 44 (5), 580–586. doi:10.1016/j.ejso.2018.01.087

Florindez, J., and Trent, J. (2020). Low Frequency of Mutation Testing in the United States: An Analysis of 3866 GIST Patients. *Am. J. Clin. Oncol.* 43 (4), 270–278. doi:10.1097/COC.0000000000000659

Gang, Z., and Wang, M. (2018). Updates and Interpretations of the NCCN Clinical Practice Guidelines(2018 First Version) on Gastrointestinal Stromal Tumor. *Zhonghua Wei Chang Wai Ke Za Zhi* 23 (9), 866–871. doi:10.3760/cma.j.cn.441530-20200731-00454

Guo, T., Agaram, N. P., Wong, G. C., Hom, G., and Antonescu, C. R. (2007). Sorafenib Inhibits the Imatinib-Resistant KITT670I Gatekeeper Mutation in Gastrointestinal Stromal Tumor. *Clin. Cancer Res. Official J. Am. Assoc. Cancer Res.* 13 (16), 4874. doi:10.1158/1078-0432.ccr-07-0484

Heinrich, M. C., Jones, R. L., Mehren, M. V., Bauer, S., and George, S. (2020). Clinical Activity of Avapritinib in ≥ Fourth-Line (4L+) and PDGFRA Exon 18 Gastrointestinal Stromal Tumors (GIST). *J. Clin. Oncol.* 38 (4_Suppl. l), 826–26. doi:10.1200/jco.2020.38.4_suppl.826

Joensuu, H., Hohenberger, P., and Corless, C. L. (2013). Gastrointestinal Stromal Tumour. *Lancet (London, England)* 382 (9896), 973–983. doi:10.1016/s0140-6736(13)60106-3

Joensuu, H., Rutkowski, P., Nishida, T., Steigen, S., Brabec, P., Plank, L., et al. (2015). KIT and PDGFRA Mutations and the Risk of GI Stromal Tumor Recurrence. *J. Clin. Oncol. : official J. Am. Soc. Clin. Oncol.* 33 (6), 634–642. doi:10.1200/jco.2014.57.4970

Jones, R. L., Serrano, C., Mehren, M. V., George, S., and Bauer, S. (2021). Avapritinib in Unresectable or Metastatic PDGFRA D842V-Mutant Gastrointestinal Stromal Tumours: Long-Term Efficacy and Safety Data from the NAVIGATOR Phase I Trial. *Eur. J. Cancer* 145, 132–142. doi:10.1016/j.ejca.2020.12.008

Jun, L. I., Cai, H. H., Wang, Y., Pathology, D. O., and Hospital, T. (2018). c-Kit and PDGFRA Gene Mutation in Gastrointestinal Stromal Tumor and Their Associations with Clinicopathological Features,immunohistochemical Expression and Prognosis. *Chin. J. Clin. Exp. Pathol.* 1, 1.

Kondo, J., Huh, W., Franklin, J., Heinrich, M., Rubin, B., and Coffey, R. (2020). A Smooth Muscle-Derived, Braf-Driven Mouse Model of Gastrointestinal Stromal Tumor (GIST): Evidence for an Alternative GIST Cell-Of-Origin. *J. Pathol.* 252 (4), 441–450. doi:10.1002/path.5552

Li, J., Ye, Y., Wang, J., Zhang, B., Qin, S., and Shi, Y. (2017). Chinese Consensus Guidelines for Diagnosis and Management of Gastrointestinal Stromal Tumor. *Chin. J. Cancer Res.* 29 (004), 281–293. doi:10.21147/j.issn.1000-9604.2017.04.01

Lincoln, S., Nussbaum, R., Kurian, A., Nielsen, S., Das, K., Michalski, S., et al. (2020). Yield and Utility of Germline Testing Following Tumor Sequencing in Patients with Cancer. *JAMA Netw. open* 3 (10), e2019452. doi:10.1001/jamanetworkopen.2020.19452

Luo, J. M., Cao, F. L., Meng, C., Lin, L. J., Ma, S. Q., and Peng, S. H. (2004). Clinical Pathology and Immunohistochemistry of Gist in Xining. *J. Qinghai Med. Coll.* 7 (1), 12890. doi:10.1038/s41598-017-12622-x

Malik, F., Santiago, T., Bahrami, A., Davis, E., and Clay, M. R. (2019). Dedifferentiation in SDH-Deficient Gastrointestinal Stromal Tumor: A Report with Histologic, Immunophenotypic, and Molecular Characterization. *Pediatr. Developmental Pathol.* 1 (4), 109352661984622. doi:10.1177/1093526619846222

Martin-Broto, J., Gutierrez, A., Garcia-Del-Muro, X., Lopez-Guerrero, J. A., and Poveda, A. (2010). Prognostic Time Dependence of Deletions Affecting Codons 557 And/or 558 of KIT Gene for Relapse-free Survival (RFS) in Localized GIST: a Spanish Group for Sarcoma Research (GEIS) Study. *Ann. Oncol.* 21 (7), 1552–1557. doi:10.1093/annonc/mdq047

Nishida, T., Blay, J. Y., Hirota, S., Kitagawa, Y., and Kang, Y. K. (2015). The Standard Diagnosis, Treatment, and Follow-Up of Gastrointestinal Stromal Tumors Based on Guidelines. *Gastric Cancer* 19 (1), 3–14. doi:10.1007/s10120-015-0526-8

Nishida, T., Kanda, T., Nishitani, A., Takahashi, T., Nakajima, K., Ishikawa, T., et al. (2008). Secondary Mutations in the Kinase Domain of the KIT Gene Are Predominant in Imatinib-Resistant Gastrointestinal Stromal Tumor. *Cancer Sci.* 99 (4), 799–804. doi:10.1111/j.1349-7006.2008.00727.x

Nishida, T., Tsujimoto, M., Takahashi, T., Hirota, S., Blay, J., and Wataya-Kaneda, M. (2016). Gastrointestinal Stromal Tumors in Japanese Patients with Neurofibromatosis Type I. *J. Gastroenterol.* 51 (6), 571–578. doi:10.1007/s00535-015-1132-6

Park, S. J., Ryu, M. H., Ryoo, B. Y., Park, Y. S., Sohn, B. S., Kim, H. J., et al. (2014). The Role of Surgical Resection Following Imatinib Treatment in Patients with Recurrent or Metastatic Gastrointestinal Stromal Tumors: Results of Propensity Score Analyses. *Ann. Surg. Oncol.* 21 (13), 4211–4217. doi:10.1245/s10434-014-3866-4

Ramdani, A., Bouhout, T., Serji, B., Khannoussi, W., and Harroudi, T. E. (2020). The Outcome of Neoadjuvant Imatinib Therapy Combined with Surgery for Rectal Gastrointestinal Stromal Tumors: A Report of Three Cases and a Review of the Literature. *Cureus* 12 (12), e12100. doi:10.7759/cureus.12100

Reichardt, P. (2016). Soft Tissue Sarcomas and Gastrointestinal Stromal Tumors. *Internist* 57 (3), 245–256. doi:10.1007/s00108-016-0021-2

Serrano, C., Mariño-Enríquez, A., Tao, D., Ketzer, J., Eilers, G., Zhu, M., et al. (2019). Complementary Activity of Tyrosine Kinase Inhibitors against Secondary Kit Mutations in Imatinib-Resistant Gastrointestinal Stromal Tumours. *Br. J. Cancer* 120 (6), 612–620. doi:10.1038/s41416-019-0389-6

Smrke, A., Gennatas, S., Huang, P., and Jones, R. L. (2020). Avapritinib in the Treatment of PDGFRA Exon 18 Mutated Gastrointestinal Stromal Tumors. *Future Oncol.* 16 (22), 1639–1646. doi:10.2217/fon-2020-0348

Virgilio, E., Annicchiarico, A., Pagliai, L., Morini, A., Romboli, A., and Montali, F. (2021). Inguinal GIST: A Systematic Literature Review of Primary and Metastatic Cases. *Anticancer Res.* 41 (1), 21–25. doi:10.21873/anticanres.14748

von Mehren, M., and Joensuu, H. (2018). Gastrointestinal Stromal Tumors. *J. Clin. Oncol.* 36 (2), 136–143. doi:10.1200/jco.2017.74.9705

von Mehren, M., and Joensuu, H. (2018). Gastrointestinal Stromal Tumors. *J. Clin. Oncol. : official J. Am. Soc. Clin. Oncol.* 36 (2), 136–143. doi:10.1200/jco.2017.74.9705

Zhao, W., Zhao, G., and Wang, M. (2020). Updates and Interpretations of the NCCN Clinical Practice Guidelines (2019 6th Version) on Gastrointestinal Stromal Tumor. *Zhonghua wei chang wai ke Za Zhi = Chin. J. Gastrointest. Surg.* 23 (9), 866–871. doi:10.3760/cma.j.cn.441530-20200731-00454

Conflict of Interest: Authors DS, NY, and ZC were employed by the company Dian Diagnostics Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Qian, Yan, Hu, Jiang, Cao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fgene.2022.860611

Anti-POSTN and Anti-TIMP1 Autoantibodies as Diagnostic Markers in Esophageal Squamous Cell Carcinoma

Weihong Xie[1†], *Guiying Sun*[2,3†], *Jicun Zhu*[2,3], *Huimin Wang*[3,4], *Zhuo Han*[2,3] and *Peng Wang*[1,2,3*]

[1]Department of Oral and Maxillofacial Surgery, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, China, [2]Department of Epidemiology and Statistics and State Key Laboratory of Esophageal Cancer Prevention & Treatment, College of Public Health, Zhengzhou University, Zhengzhou, China, [3]Henan Key Laboratory of Tumor Epidemiology and State Key Laboratory of Esophageal Cancer Prevention & Treatment, Zhengzhou University, Zhengzhou, China, [4]Academy of Medical Science, Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Xue Li,
Zhejiang University, China
Jintang He,
Genentech, Inc., United States
Yashu Liu,
Iaronde Inc., United States

***Correspondence:**
Peng Wang
wangpeng1658@hotmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 23 January 2022
Accepted: 29 March 2022
Published: 26 April 2022

Citation:
Xie W, Sun G, Zhu J, Wang H, Han Z and Wang P (2022) Anti-POSTN and Anti-TIMP1 Autoantibodies as Diagnostic Markers in Esophageal Squamous Cell Carcinoma. Front. Genet. 13:860611. doi: 10.3389/fgene.2022.860611

Esophageal cancer is one of the most commonly diagnosed malignant gastrointestinal tumors. The aim of the study was to explore the diagnostic values of anti-POSTN and anti-TIMP1 autoantibodies in esophageal squamous cell carcinoma (ESCC). Differentially expressed genes (DEGs) associated with esophageal cancer were screened out by the LIMMA method in the Gene Expression Profiling Interactive Analysis (GEPIA) platform. Search Tool for the Retrieval of Interacting Genes (STRING) was used to construct the protein–protein interaction (PPI) based on highly DEGs. The candidate hub genes were the intersection genes calculated based on degree and Maximal Clique Centrality (MCC) algorithms *via* Cytoscape. A total of 370 participants including 185 ESCC patients and 185 matched normal controls were enrolled in enzyme-linked immunosorbent assay (ELISA) to detect the expression levels of autoantibodies corresponding to POSTN and TIMP1 proteins. A total of 375 DEGs with high expression were obtained in esophageal cancer. A total of 20 hub genes were acquired using the cytoHubba plugin by degree and MCC algorithms. The expression levels of anti-POSTN and anti-TIMP1 autoantibodies were higher in the sera of ESCC patients ($p < 0.05$). Anti-POSTN autoantibody can diagnose ESCC patients with an AUC of 0.638 at the specificity of 90.27% and sensitivity of 27.57%, and anti-TIMP1 autoantibody can diagnose ESCC patients with an AUC of 0.585 at the specificity of 90.27% and sensitivity of 20.54% ($p < 0.05$). In addition, anti-POSTN and anti-TIMP1 autoantibodies can distinguish ESCC patients from normal controls in most clinical subgroups ($p < 0.05$). In conclusion, anti-POSTN and anti-TIMP1 autoantibodies may be considered the potential biomarkers in the clinical diagnosis of ESCC.

Keywords: autoantibody, POSTN, TIMP1, immunodiagnosis, esophageal squamous cell carcinoma, biomarker

INTRODUCTION

Esophageal cancer (EC) is one of the most common malignant tumors of the digestive tract and ranks sixth among malignant tumors in mortality worldwide (Sung et al., 2021). GLOBOCAN 2020 shows that there are 604,100 newly diagnosed EC cases worldwide and 544,076 people died in the same period (Sung et al., 2021). Many patients are in the advanced stage at the initial diagnosis and have poor prognosis because of the occult onset and no obvious early symptoms (Li et al., 2021). The 5-year survival rate can reach as high as 60%, if patients with EC could be diagnosed early and treated with surgery (He et al., 2021). At present, the commonly used clinical methods for diagnosing EC include endoscopy, CT scan, barium meal examination, and pathological biopsy, but they are expensive and invasive and cannot be used for screening (Guo et al., 2018). Traditional tumor serological markers such as cancer antigen 12-5 (CA12-5), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCCA), and carbohydrate antigen199 (CA19-9) are used as auxiliary markers of clinical diagnosis for EC, but the sensitivity and specificity of these markers are poor (Kosugi et al., 2004; Mroczko et al., 2008). Therefore, it is of great clinical value to explore better diagnostic markers of EC.

Bioinformatics can conduct in-depth analysis of open biological databases such as tissues, cell genes, and proteins, and provide a potential theoretical basis for cancer early diagnosis and treatment (Wang Y. et al., 2021). Many scholars have used the database to analyze differentially expressed genes (DEGs) in the progression of lung cancer, gastric cancer, breast cancer, and other tumors (Ren et al., 2020; Wang H. et al., 2021; Shan et al., 2021). The proteins encoded by these DEGs could be used as molecular markers for tumor diagnosis and prognosis (Zhao et al., 2021). Moreover, autoantibodies against tumor-associated antigens (TAAbs) can exist stably in the serum of cancer patients and can be detected months or even years before the onset of clinical symptoms (Tan and Zhang, 2008). Therefore, they have the potential to be biomarkers for early immunodiagnosis of cancers. Many studies have reported higher levels of TAAbs in the serum of patients with cancer, such as hepatocellular carcinoma, ovarian cancer, breast cancer, and esophageal cancer (Katchman et al., 2017; Xu and Liu, 2017; Loke and Lee, 2018; Zheng et al., 2018). At present, there is no recognized marker for the detection of esophageal cancer, and the aim of the present study was to identify novel TAAbs to improve the sensitivity and specificity.

In China, esophageal squamous cell carcinoma (ESCC) patients account for more than 90% of esophageal cancer patients (Cao and Sun, 2016). In this study, we analyzed the EC-related data from TCGA (The Cancer Genome Atlas) and GTEx (the Genotype-Tissue Expression) databases to identify the

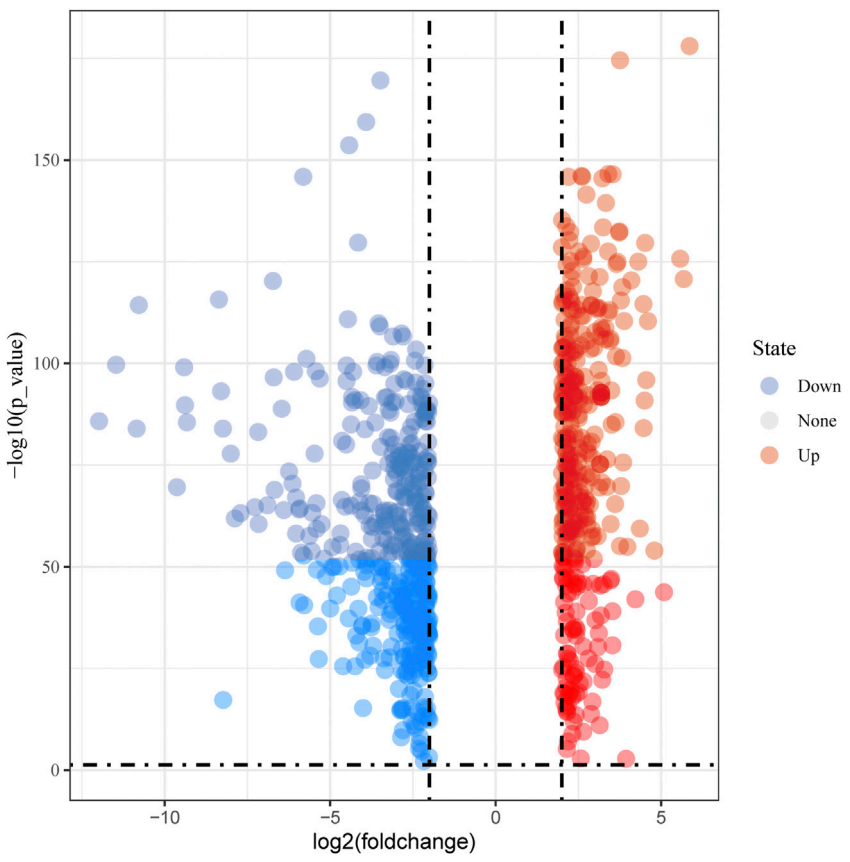


FIGURE 1 | Volcano plot of differentially expressed genes.

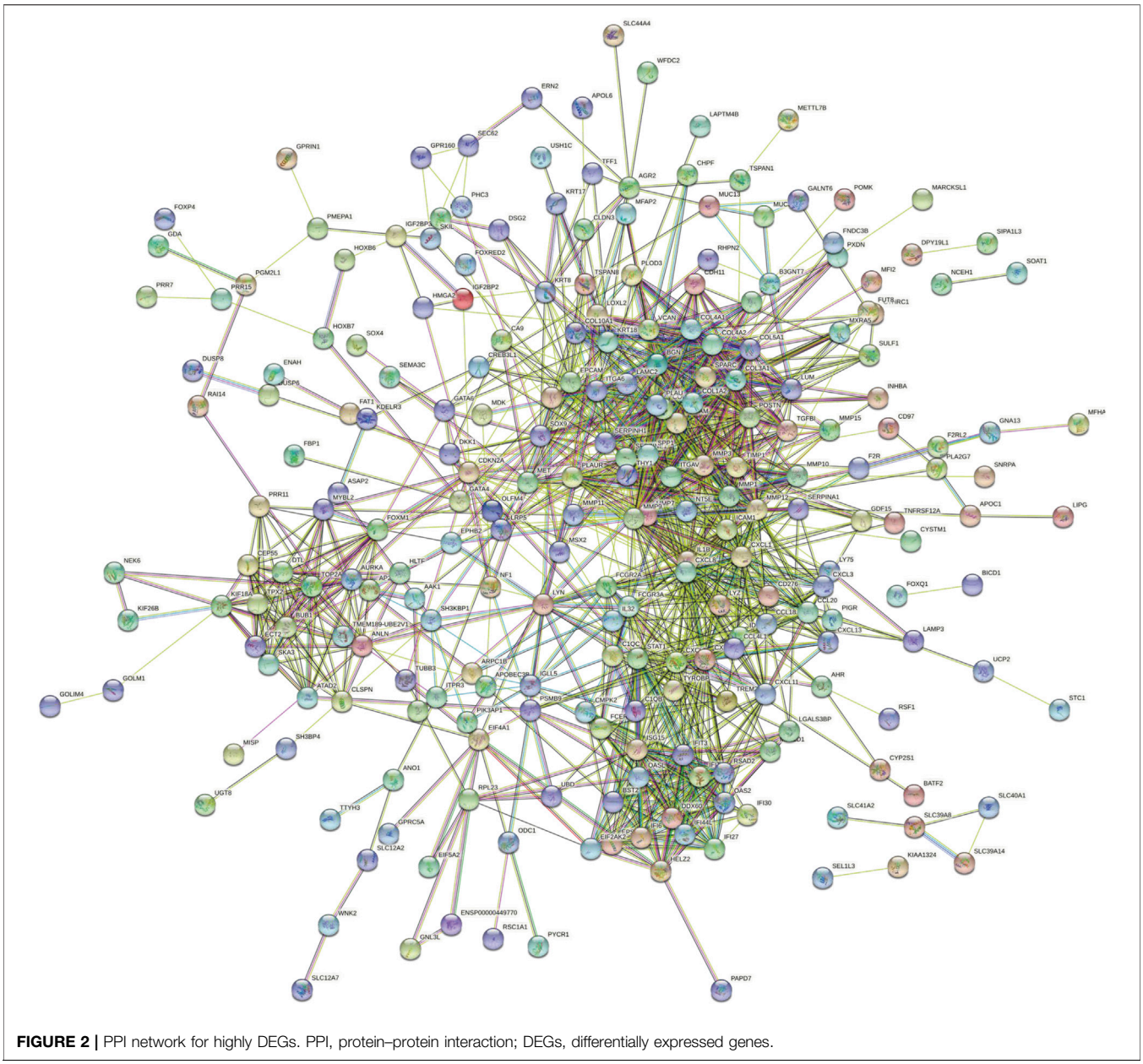


FIGURE 2 | PPI network for highly DEGs. PPI, protein–protein interaction; DEGs, differentially expressed genes.

differentially expressed genes in esophageal cancer. Then, we further attained the hub genes in the highly expressed differential genes *via* degree and Maximal Clique Centrality (MCC) algorithms, and the proteins encoded by them were regarded as candidate TAAs. Finally, enzyme-linked immunosorbent assay (ELISA) was used to evaluate the diagnostic value of the corresponding autoantibodies of TAAs for ESCC.

MATERIALS AND METHODS

Screening of Candidate Hub Genes

Based on TCGA (The Cancer Genome Atlas) and GTEx (the Genotype-Tissue Expression) databases, we used the GEPIA (http://gepia2.cancer-pku.cn/#degenes) web server to screen DEGs associated with esophageal cancer by the LIMMA method (Tang et al., 2019). DEGs satisfying the criteria with adjusted $p < 0.01$ and |log2 fold change (FC)|>2 were designated as statistically significant. The volcano map of DEGs was drawn using the SangerBox tools, a free online platform for data analysis (http://www.sangerbox.com/tool).

STRING (Search Tool for the Retrieval of Interacting Genes, https://cn.string-db.org/) was adopted to construct the protein–protein interaction (PPI) based on highly DEGs (Franceschini et al., 2013). The cytoHubba plug-in Cytoscape (version 3.8.2) was used to calculate hub nodes to select the top 20 DEGs (Shannon et al., 2003). The intersection genes calculated

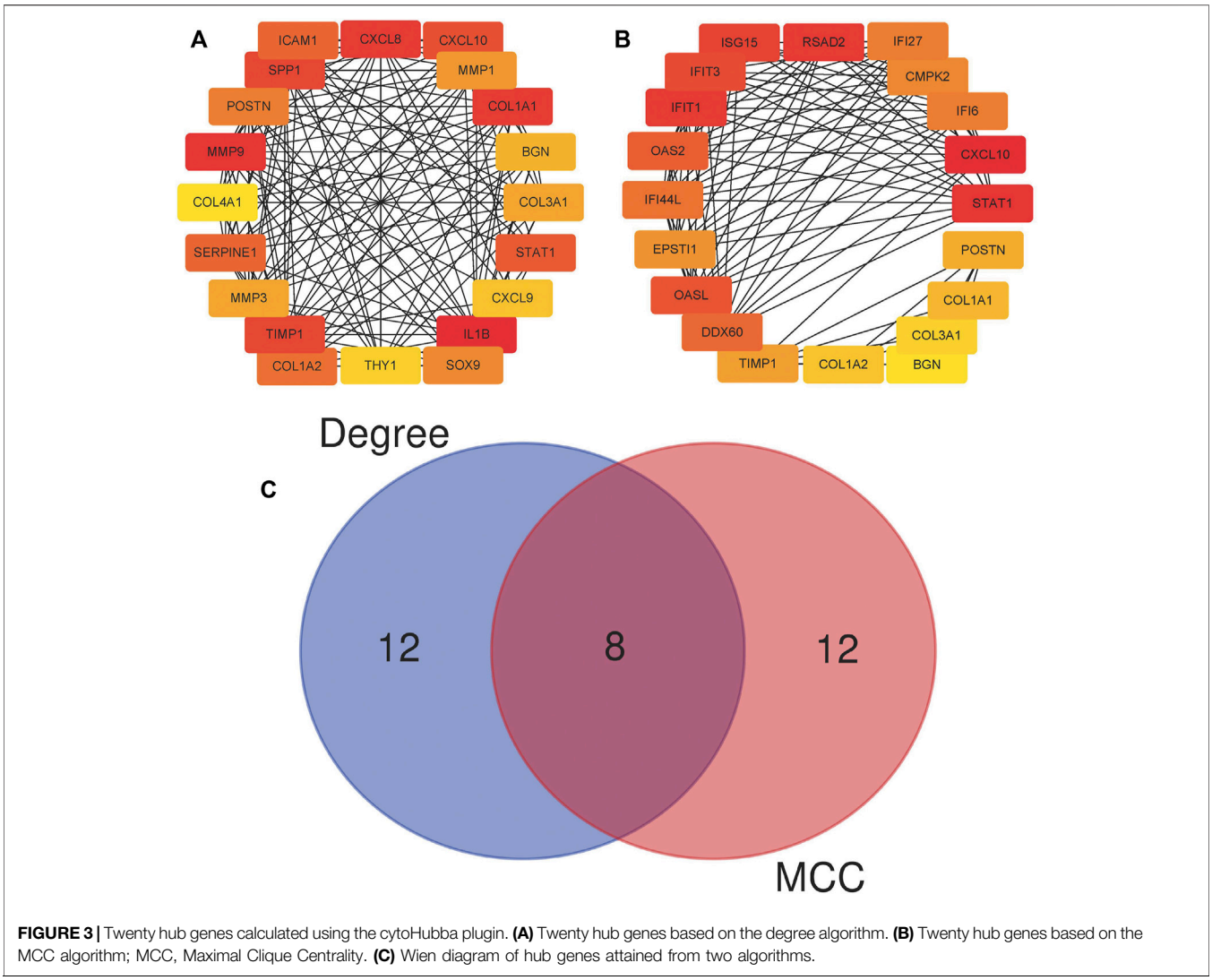


FIGURE 3 | Twenty hub genes calculated using the cytoHubba plugin. **(A)** Twenty hub genes based on the degree algorithm. **(B)** Twenty hub genes based on the MCC algorithm; MCC, Maximal Clique Centrality. **(C)** Wien diagram of hub genes attained from two algorithms.

based on degree and MCC algorithms were used as the candidate hub genes.

Enzyme-Linked Immunosorbent Assay

Through bioinformatics methods, we know that POSTN and TIMP1 were highly expressed in patients with esophageal cancer. Therefore, we further detect the expression levels of autoantibodies corresponding to these two proteins in the serum of patients with ESCC by ELISA. Purified recombinant proteins POSTN and TIMP1 were purchased from the CLOUD-CLONE CORP (Wuhan, China). Horseradish peroxidase (HRP)-conjugated mouse anti-human IgG (Wuhan Aoko Biotechnology Co. Ltd.) was used as the secondary antibody. The coated concentrations of POSTN and TIMP1 proteins were both 0.5 ng/ml. Each ELISA plate included six repeated serum samples and two blank controls. The repeated serum samples were used to normalize the difference between plates, and the blank controls were consulted for quality control. The detailed steps of ELISA were described in our previous study (Wang et al., 2018). The microplate reader was performed to measure the optical density (OD) of wells at 450 and 620 nm. The absorbance difference between 450 and 620 nm was used for the subsequent analysis.

Study Participants

The serum samples of 185 patients with ESCC included in this study came from a third class hospital in Henan Province. They were all patients with ESCC diagnosed by histopathology and without any treatment and did not suffer from other cancer diseases. The serum samples of 185 normal controls were received from the specimen bank of Henan Key Laboratory of Tumor Epidemiology. The normal controls excluded autoimmune diseases, esophageal cancer, and related diseases.

Statistical Analysis

IBM SPSS 22.0 and GraphPad Prism 9.1.1 were used in the study. All statistical analyses were based on the two-tailed test, and $p < 0.05$ was considered to be statistically significant. The non-parametric test was adopted to compare the expression levels

TABLE 1 | Characteristics of study participants.

Group	ESCC patient	Normal control
Number	185	185
Gender		
Male, *n* (%)	126 (68.11)	122 (65.95)
Female, *n* (%)	59 (31.89)	63 (34.05)
Age		
Mean age ± SD (years)	64.2 ± 8.3	64.0 ± 7.2
Age range (years)	42–88	53–93
Tumor site		
Upper thorax, *n* (%)	22 (11.89)	—
Middle thorax, *n* (%)	66 (35.68)	—
Lower thorax, *n* (%)	39 (21.08)	—
Unknown, *n* (%)	58 (31.35)	—
Family tumor history		
Yes, *n* (%)	47 (25.41)	—
No, *n* (%)	136 (73.51)	—
Unknown, *n* (%)	2 (1.08)	—
Histological grade		
High, *n* (%)	3 (1.62)	—
Medium, *n* (%)	46 (24.86)	—
Low, *n* (%)	40 (21.62)	—
Unknown, *n* (%)	96 (51.89)	—
TNM stage		
I, *n* (%)	65 (35.14)	—
II, *n* (%)	28 (15.14)	—
III, *n* (%)	43 (23.24)	—
IV, *n* (%)	25 (13.51)	—
Unknown	24 (12.97)	—
Lymph node metastasis		
Positive, *n* (%)	73 (39.46)	—
Negative, *n* (%)	101 (54.59)	—
Unknown, *n* (%)	11 (5.95)	—
Distant metastasis		
Yes, *n* (%)	19 (10.27)	—
No, *n* (%)	118 (63.78)	—
Unknown, *n* (%)	48 (25.95)	—

ESCC, esophageal squamous cell carcinoma; SD, standard deviation.

of autoantibodies between ESCC patients and normal controls. The OD value corresponding to the maximum Youden's index when the specificity is greater than 90% was determined as the cutoff value. The Chi-square test was employed to analyze the positive rates of autoantibodies in different clinical subgroups in all ESCC patients. The receiver operating characteristic curve (ROC) was used to evaluate the diagnostic value of the autoantibodies in different groups. The positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV), accuracy, and Youden's index were calculated to estimate the diagnostic value of the two autoantibodies.

RESULTS

Identified Potential Hub Genes

By analyzing the gene expression profiles of esophageal cancer tissues and adjacent tissues, the differentially expressed genes in esophageal cancer tissues were significantly higher or lower than those in normal esophageal tissues. After differential analysis, a total of 375 DEGs with high expression and 449 DEGs with low expression were obtained in esophageal cancer, and the results of the visualization are in a volcano plot (**Figure 1**). The PPI network generated by STRING is presented in **Figure 2** for highly DEGs. The 20 hub genes calculated using the cytoHubba plugin were IL1B, MMP9, CXCL8, COL1A1, SPP1, TIMP1, CXCL10, STAT1, ICAM1, SERPINE1, COL1A2, POSTN, SOX9, MMP1, MMP3, COL3A1, BGN, CXCL9, THY1, and COL4A1 by degree algorithm (**Figure 3A**), and the 20 hub genes calculated using the cytoHubba plugin were CXCL10, STAT1, IFIT1, RSAD2, ISG15, IFIT3, OASL, OAS2, DDX60, IFI44L, IFI27, IFI6, CMPK2, EPSTI1, TIMP1, POSTN, COL1A1, COL1A2, COL3A1, and BGN by MCC algorithm (**Figure 3B**).

A total of eight genes were selected by both algorithms, including CXCL10, STAT1, POSTN, TIMP1, COL1A1, COL1A2, COL3A1, and BGN (**Figure 3C**). According to the importance ranking of these eight genes in the two algorithms and the query of the relevant literature, we finally determined POSTN and TIMP1 as the target genes of this study. In terms of importance ranking, according to the comprehensive importance ranking calculated by the two algorithms, the order of the eight genes was CXCL10, STAT1, TIMP1, COL1A1, POSTN, COL1A2, BGN, and COL3A1. In addition, POSTN functioned as a cell adhesion molecule and participated in many biological processes, including cell adhesion, invasion, metastasis, and tumor angiogenesis (Bao et al., 2004; Kudo et al., 2006; Siriwardena et al., 2006). TIMP1 promotes the growth of human keratinocytes and several other cell types, inhibits apoptosis, and promotes growth (Gasson et al., 1985; Bertaux et al., 1991; Guedez et al., 1998). A high expression of TIMP1 has a significant correlation with a poor prognosis of cancer (Jackson et al., 2017). Therefore, the proteins encoded by POSTN and TIMP1 have the possibility of being potential tumor-associated antigens, and we further detected the level of anti-TAA autoantibodies in the subjects' serum by ELISA experimental.

Characteristics of Study Participants

The expression levels of anti-POSTN and anti-TIMP1 autoantibodies obtained by bioinformatics methods in ESCC patients and normal controls were verified by ELISA. Sera of 185 ESCC patients and 185 normal controls were used in ELISA. The study was approved by the Medical Ethics Committee of Zhengzhou University, and informed consent was obtained from all participants. The detailed clinical information of 370 participants is described in **Table 1**. There was no significant difference in gender (p = 0.658) and age (p = 0.223) between ESCC patients and normal controls.

The Diagnostic Values of Anti-POSTN and Anti-TIMP1 Autoantibodies in Esophageal Squamous Cell Carcinoma

The expression levels of anti-POSTN and anti-TIMP1 autoantibodies in ESCC patients and normal controls were detected by ELISA. The expression level of anti-POSTN autoantibody in ESCC patients was distinctly higher than that in the normal controls (mean ± SD: 0.292 ± 0.149 vs. 0.224 ± 0.087) (**Figure 4A**). Anti-POSTN autoantibody can diagnose ESCC patients with an AUC of 0.638 at the specificity

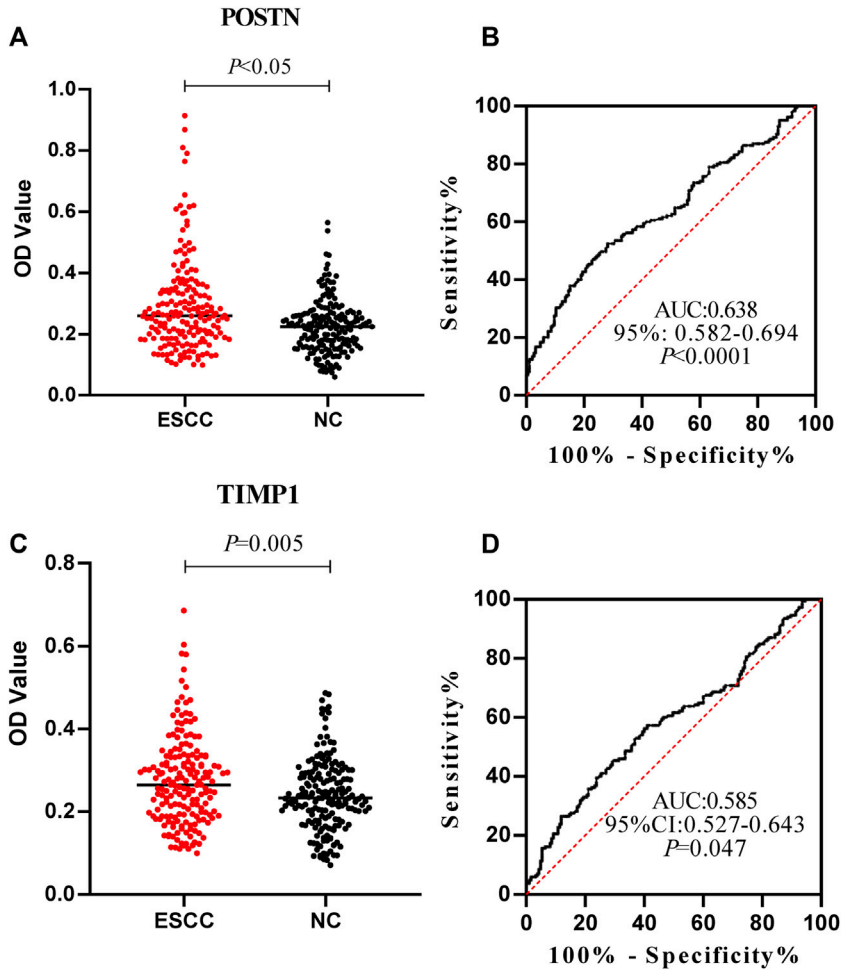


FIGURE 4 | Expression level and diagnostic value of anti-POSTN and anti-TIMP1 autoantibodies in ESCC patients and normal controls. **(A,C)** Scatter plot described the expression level of anti-POSTN and anti-TIMP1 autoantibodies in the ESCC group and normal control group. **(B,D)** Receiver operating characteristic curve of anti-POSTN and anti-TIMP1 autoantibodies in diagnosing ESCC patients and normal controls. ESCC, esophageal squamous cell carcinoma; NC, normal controls.

of 90.27% and sensitivity of 27.57% (**Figure 4B**). The expression level of anti-TIMP1 autoantibody in ESCC patients was elevated compared with that of normal controls (mean ± SD: 0.274 ± 0.106 vs. 0.240 ± 0.085) (**Figure 4C**). Anti-TIMP1 autoantibody can diagnose ESCC patients with an AUC of 0.585 at the specificity of 90.27% and sensitivity of 20.54% (**Figure 4D**). In order to further evaluate the diagnostic values of anti-POSTN and anti-TIMP1 autoantibodies in ESCC, other diagnostic parameters were also calculated, as shown in **Table 2**. Although the specificities of the two biomarkers were same, the sensitivity and accuracy of anti-POSTN were higher than those of anti-TIMP1. In terms of the diagnostic test, when the positive likelihood ratio was larger and the negative likelihood ratio was smaller, the diagnostic effect of biomarkers was better. **Table 2** indicates that the diagnostic value of anti-POSTN autoantibody was more excellent than that of anti-TIMP1 autoantibody. All in all, the diagnosis effect of anti-POSTN for ESCC was better.

The Values of Anti-POSTN and Anti-TIMP1 Autoantibodies in Diagnosing Esophageal Squamous Cell Carcinoma Patients of Different Clinical Features

The positive rate of autoantibodies in patients was calculated by taking the mean value of autoantibodies in normal controls plus the expression level of standard deviation as the cutoff value. The diagnostic values of anti-POSTN and anti-TIMP1 autoantibodies in ESCC patients of different clinicopathological characteristics, including lymphatic metastasis, distance metastasis, differentiation, TNM stage, family tumor history, gender, and age, were further explored.

TABLE 2 | Diagnostic value of anti-POSTN and anti-TIMP1 autoantibodies in diagnosing ESCC patients.

	AUC	95% CI	Se (%)	Sp (%)	YI	+LR	−LR	PPV (%)	NPV (%)	Accuracy (%)
POSTN	0.638	0.582–0.694	27.57	90.27	0.18	2.83	0.80	73.91	44.52	59.00
TIMP1	0.585	0.527–0.643	20.54	90.27	0.11	2.11	0.88	67.86	46.82	55.00

Se, sensitivity; Sp, specificity; AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; +LR, positive likelihood ratio; −LR, negative likelihood ratio.

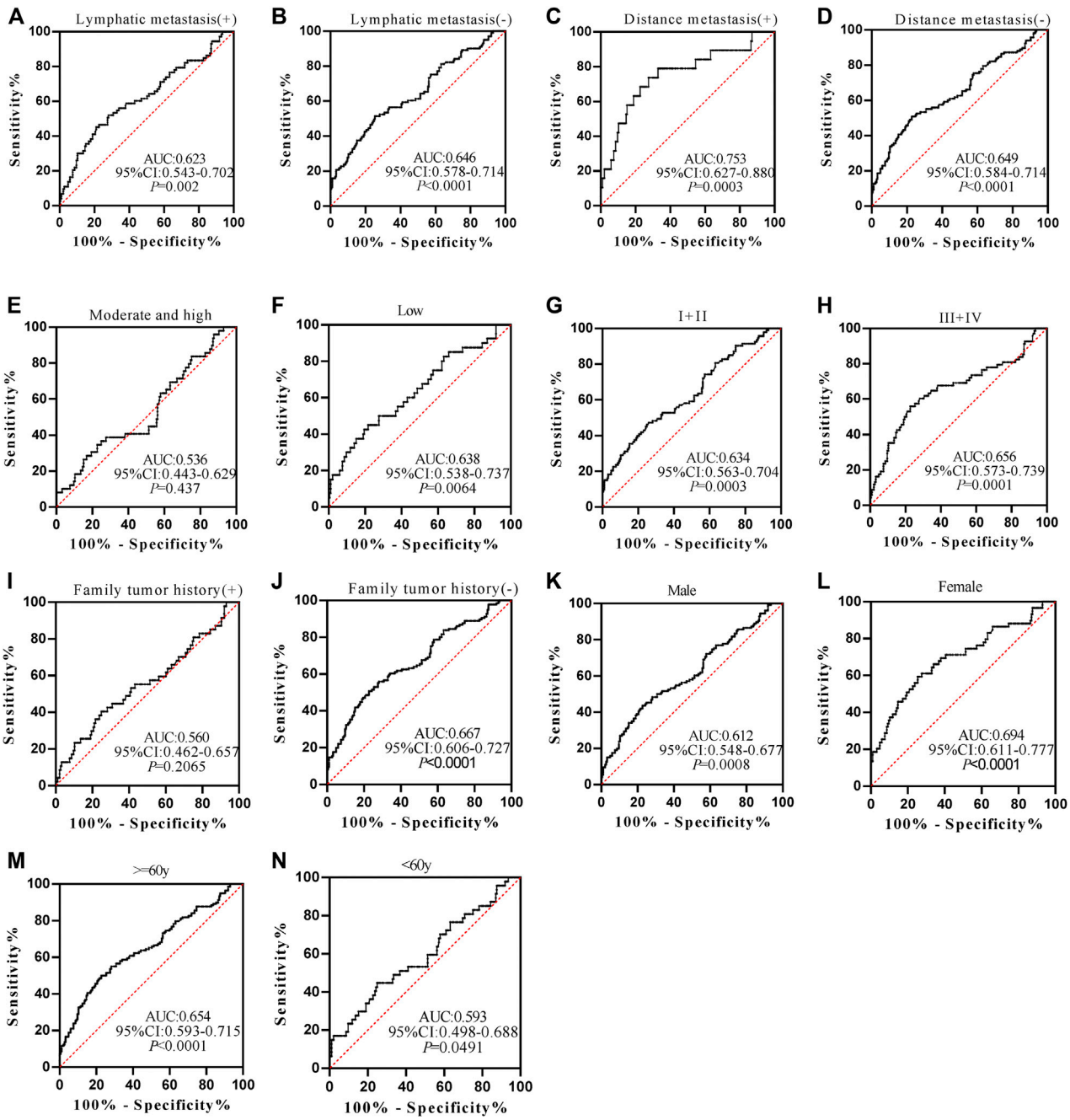


FIGURE 5 | The performance of anti-POSTN autoantibody in ESCC patients with different clinical characteristics.

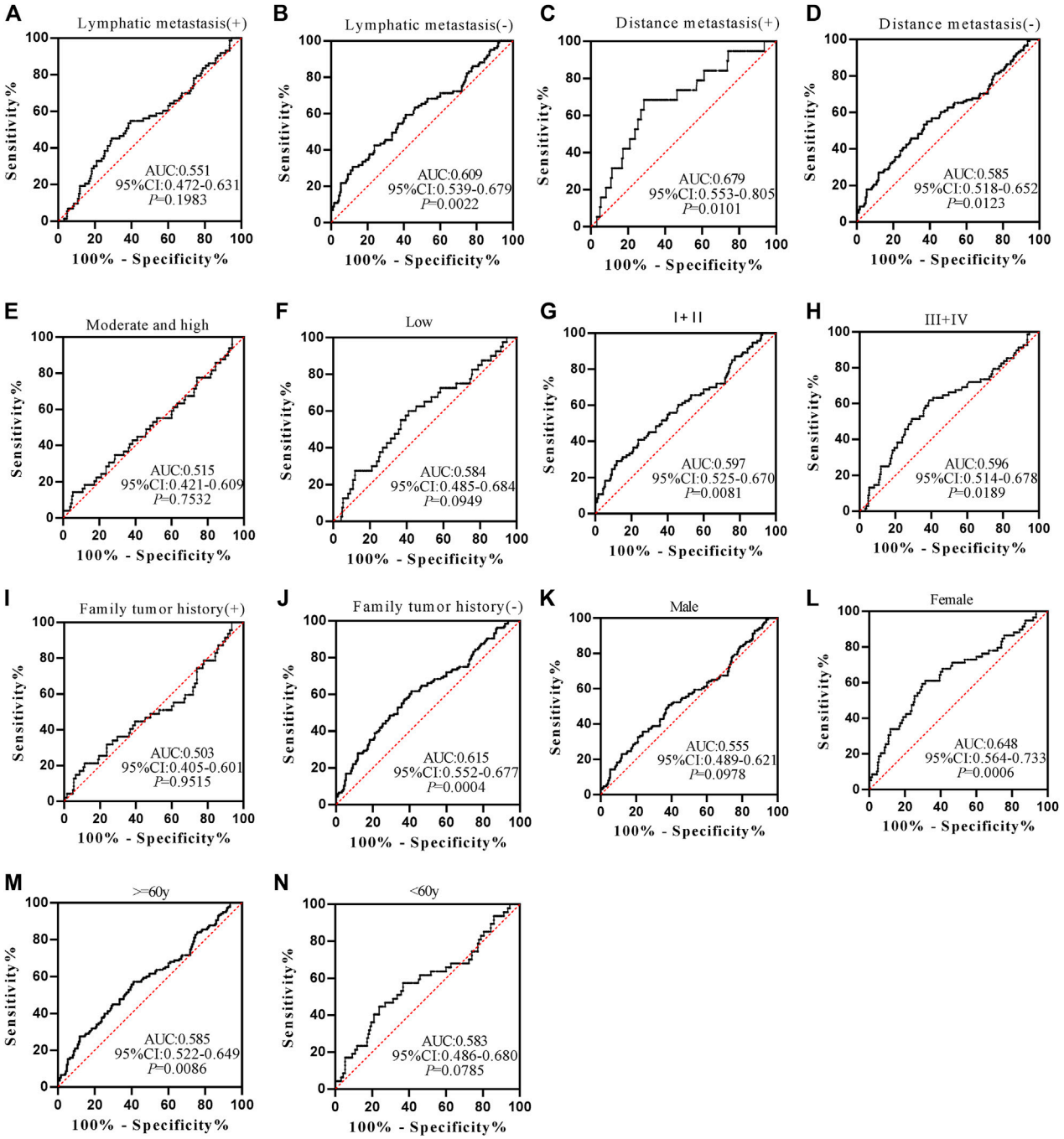


FIGURE 6 | The value of anti-TIMP1 autoantibody in diagnosing ESCC patients of different clinical subgroups.

Anti-POSTN autoantibody can diagnose ESCC patients from normal controls in most subgroups except for patients with family tumor history and moderate and high differentiation, and it showed marginal difference in diagnosing ESCC patients younger than 60 years old (p = 0.0491) (**Figures 5A–N**). The AUCs of anti-POSTN autoantibody in diagnosing ESCC patients with different clinical characteristics ranged from 0.612 to 0.753 (p<0.05). The minimum AUC of 0.612 was observed in male ESCC patients (**Figure 5K**), and the maximum AUC of 0.753 was presented in patients with distance metastasis (**Figure 5C**). Anti-TIMP1

TABLE 3 | Positive frequencies of autoantibodies in subgroups.

Subgroup	Anti-POSTN		Anti-TIMP1	
	Positive rate (%)	*p*	Positive rate (%)	*p*
Lymphatic metastasis				
Positive (73)	30.14	0.723	19.18	0.087
Negative (101)	32.67		30.69	
Differentiation				
Moderate and high (49)	18.37	0.124	18.37	0.305
Low (40)	32.50		27.50	
Distance metastasis				
Positive (19)	47.37	0.324	31.58	0.575
Negative (118)	35.59		27.12	
TNM stage				
I + II (93)	31.18	0.583	29.03	0.943
III + IV(68)	35.29		25.00	
Family tumor history				
Positive (47)	23.40	0.133	21.28	0.571
Negative (136)	35.29		27.94	
Age				
≥60 years (138)	38.10	0.242	27.54	0.579
<60 years (47)	25.53		23.40	
Gender				
Male (126)	29.37	0.193	23.02	0.118
Female (59)	38.98		33.90	

autoantibody can diagnose ESCC patients from normal controls in most subgroups ($p < 0.05$), but it failed to distinguish ESCC patients from normal controls in patients with lymphatic metastasis, a history of family tumor, different degrees of differentiation, and patients younger than 60 years old as well as male patients ($p > 0.05$) (**Figures 6A–N**). The AUC values of anti-TIMP1 autoantibody in different clinical subgroups ranged from 0.585 to 0.679($p<0.05$). The minimum AUC of 0.585 was observed in both patients older than 60 and patients with no distance metastasis (**Figures 6D,M**), and the maximum AUC of 0.679 was detected in patients with distance metastasis (**Figure 6C**).

The positive rates of anti-POSTN and anti-TIMP1 autoantibodies were of no difference in clinical subgroups ($p > 0.05$) (**Table 3**). Furthermore, the AUC values of the two autoantibodies were not statistically different in all clinical subgroups ($p > 0.05$).

DISCUSSIONS

EC is a common malignant tumor worldwide. There are no obvious clinical symptoms in the early stage of EC, and there are no serum test markers that can be used for minimally invasive detection. Recent studies have identified many types of biomarkers based on GEO and TCGA datasets. For example, Zhao et al. (2021) reported that GXYLT2 might be a potential diagnostic and prognostic marker in gastric cancer based on a comprehensive analysis. Liang et al. (2019) indicated that FKBP10 may be a potential therapeutic target for the treatment of gastric cancer *via* bioinformatics analysis and immunohistochemical verification. Yang et al. (2021) confirmed that TRIB3 was a potential prognostic marker and therapeutic target for bladder cancer through bioinformatics analysis and cell function experiment. In this study, we performed differential analysis, PPI analysis, and hub gene calculation on TCGA data and GTEx data related to EC. It was gratifying that we identified two TAAbs (POSTN and TIMP1) with a potential diagnostic value for ESCC through experiment verification. To the best of our knowledge, the association of anti-POSTN and anti-TIMP1 autoantibodies with ESCC has not been reported.

Periostin (POSTN), playing a crucial role in some biological processes, is considered to be associated with tumor progression (Kudo, 2017). In several malignant tumors, such as colorectal cancer, ovarian cancer, and hepatocellular carcinoma, high POSTN expression was confirmed to be associated with poor prognosis (Lv et al., 2013; Sung et al., 2016; Deng et al., 2019). Moreover, Jin and Yang (2019) showed that POSTN could be a promising potential diagnostic biomarker for head and neck squamous cell carcinoma. Moreover, limited studies have indicated a high POSTN expression was due to poor prognostic factors for ESCC based on immunohistochemistry (Wang et al., 2014; Lv et al., 2017; Ishibashi and Tsujimoto, 2021). Based on the aforementioned research, we proposed to assume that the autoantibody produced by the protein encoded by POSTN can be used as a marker for ESCC diagnosis. In this study, we confirmed the value of anti-POSTN autoantibody in the diagnosis of ESCC, with an AUC of 0.638.

Tissue inhibitors of metalloproteinases (TIMPs) are proverbial inhibitors of metalloproteinases and composed of four structurally related members (TIMP1, TIMP2, TIMP3, and TIMP4), associated with tumor invasion and angiogenesis

(Baker et al., 2002; Jiang et al., 2002). Among four TIMPs, TIMP1 overexpression or TIMP3 silencing is considered to be associated with tumor progression (Jackson et al., 2017). In gastric cancer, Mroczko et al. (2009) demonstrated that TIMP1 expression of pre-treated serum and plasma correlated significantly with the presence of distant metastases. In addition, Kozłowski et al. (2013) reported that high levels of TIMP1 in serum were related to progression and worse prognosis of patients with EC. In the current study, we found there was no significant difference of positive rates of anti-TIMP1 autoantibody between subgroups.

However, the study still has some limitations. This study is a retrospective study, so prospective studies should be carried out to verify the results. Based on the bioinformatics analyses and experimental verification presented in this study, we concluded from these results that anti-POSTN and anti-TIMP1 autoantibodies could be considered potential diagnostic markers for ESCC. We hope that our results will benefit future studies and improve the diagnosis of ESCC patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

WX and PW contributed to study concepts. WX and GS analyzed and interpreted the patient data and were major contributors in writing the manuscript. WX, GS, and JZ contributed to experimental studies. GS, JZ, and HW contributed to data analysis. WX, GS, JZ, HW, ZH, and PW contributed to manuscript revision. All authors participated in the critical revision of the manuscript for important intellectual content.

FUNDING

This work was supported by the Funded Project of International Training of High-level Talents in Henan Province (No number).

REFERENCES

Baker, A. H., Edwards, D. R., and Murphy, G. (2002). Metalloproteinase Inhibitors: Biological Actions and Therapeutic Opportunities. *J. Cel Sci* 115 (Pt 19), 3719–3727. doi:10.1242/jcs.00063

Bao, S., Ouyang, G., Bai, X., Huang, Z., Ma, C., Liu, M., et al. (2004). Periostin Potently Promotes Metastatic Growth of colon Cancer by Augmenting Cell Survival *via* the Akt/PKB Pathway. *Cancer Cell* 5 (4), 329–339. doi:10.1016/s1535-6108(04)00081-9

Bertaux, B., Hornebeck, W., Eisen, A. Z., and Dubertret, L. (1991). Growth Stimulation of Human Keratinocytes by Tissue Inhibitor of Metalloproteinases. *J. Invest. Dermatol.* 97 (4), 679–685. doi:10.1111/1523-1747.ep12483956

Cao, X., and Sun, X. (2016). Incidence and Trend of Esophageal Cancer. *Chin. J. Clin. Oncol.* 43 (21), 932–936. doi:10.3969/j.issn.1000-8179.2016.21.984

Deng, X., Ao, S., Ao, S., Hou, J., Li, Z., Lei, Y., et al. (2019). Prognostic Significance of Periostin in Colorectal Cancer. *Chin. J. Cancer Res.* 31 (3), 547–556. doi:10.21147/j.issn.1000-9604.2019.03.16

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration. *Nucleic Acids Res.* 41 (Database issue), D808–D815. doi:10.1093/nar/gks1094

Gasson, J. C., Golde, D. W., Kaufman, S. E., Westbrook, C. A., Hewick, R. M., Kaufman, R. J., et al. (1985). Molecular Characterization and Expression of the Gene Encoding Human Erythroid-Potentiating Activity. *Nature* 315 (6022), 768–771. doi:10.1038/315768a0

Guedez, L., Stetler-Stevenson, W. G., Wolff, L., Wang, J., Fukushima, P., Mansoor, A., et al. (1998). *In Vitro* suppression of Programmed Cell Death of B Cells by Tissue Inhibitor of Metalloproteinases-1. *J. Clin. Invest.* 102 (11), 2002–2010. doi:10.1172/jci2881

Guo, L., Zhang, L., and Zhao, J. (2018). CT Scan and Magnetic Resonance Diffusion-weighted I-maging in the D-iagnosis and T-reatment of E-sophageal C-ancer. *Oncol. Lett.* 16 (6), 7117–7122. doi:10.3892/ol.2018.9532

He, S., Xu, J., Liu, X., and Zhen, Y. (2021). Advances and Challenges in the Treatment of Esophageal Cancer. *Acta Pharmaceutica Sinica B* 11 (11), 3379–3392. doi:10.1016/j.apsb.2021.03.008

Ishibashi, Y., Tsujimoto, H., Einama, T., Mochizuki, S., Kouzu, K., Nomura, S., et al. (2021). Correlation between Immunoinflammatory Measures and Periostin Expression in Esophageal Squamous Cell Carcinoma: A Single-Center, Retrospective Cohort Study. *Ann. Surg. Oncol.* 28 (2), 1228–1237. doi:10.1245/s10434-020-08765-3

Jackson, H. W., Defamie, V., Waterhouse, P., and Khokha, R. (2017). TIMPs: Versatile Extracellular Regulators in Cancer. *Nat. Rev. Cancer* 17 (1), 38–53. doi:10.1038/nrc.2016.115

Jiang, Y., Goldberg, I. D., and Shi, Y. E. (2002). Complex Roles of Tissue Inhibitors of Metalloproteinases in Cancer. *Oncogene* 21 (14), 2245–2252. doi:10.1038/sj.onc.1205291

Jin, Y., and Yang, Y. (2019). Identification and Analysis of Genes Associated with Head and Neck Squamous Cell Carcinoma by Integrated Bioinformatics Methods. *Mol. Genet. Genomic Med.* 7 (8), e857. doi:10.1002/mgg3.857

Katchman, B. A., Chowell, D., Wallstrom, G., Vitonis, A. F., LaBaer, J., Cramer, D. W., et al. (2017). Autoantibody Biomarkers for the Detection of Serous Ovarian Cancer. *Gynecol. Oncol.* 146 (1), 129–136. doi:10.1016/j.ygyno.2017.04.005

Kosugi, S.-i., Nishimaki, T., Kanda, T., Nakagawa, S., Ohashi, M., and Hatakeyama, K. (2004). Clinical Significance of Serum Carcinoembryonic Antigen, Carbohydrate Antigen 19-9, and Squamous Cell Carcinoma Antigen Levels in Esophageal Cancer Patients. *World J. Surg.* 28 (7), 680–685. doi:10.1007/s00268-004-6865-y

Kozłowski, M., Laudański, W., Mroczko, B., Szmitkowski, M., Milewski, R., and Łapuć, G. (2013). Serum Tissue Inhibitor of Metalloproteinase 1 (TIMP-1) and Vascular Endothelial Growth Factor A (VEGF-A) Are Associated with Prognosis in Esophageal Cancer Patients. *Adv. Med. Sci.* 58 (2), 227–234. doi:10.2478/ams-2013-0017

Kudo, A. (2017). Introductory Review: Periostin-Gene and Protein Structure. *Cell. Mol. Life Sci.* 74 (23), 4259–4268. doi:10.1007/s00018-017-2643-5

Kudo, Y., Ogawa, I., Kitajima, S., Kitagawa, M., Kawai, H., Gaffney, P. M., et al. (2006). Periostin Promotes Invasion and anchorage-independent Growth in the Metastatic Process of Head and Neck Cancer. *Cancer Res.* 66 (14), 6928–6935. doi:10.1158/0008-5472.can-05-4540

Li, J., Xu, J., Xu, J., Zheng, Y., Gao, Y., He, S., et al. (2021). Esophageal Cancer: Epidemiology, Risk Factors and Screening. *Chin. J. Cancer Res.* 33 (5), 535–547. doi:10.21147/j.issn.1000-9604.2021.05.01

Liang, L., Zhao, K., Zhu, J. H., Chen, G., Qin, X. G., and Chen, J. Q. (2019). Comprehensive Evaluation of FKBP10 Expression and its Prognostic Potential in Gastric Cancer. *Oncol. Rep.* 42 (2), 615–628. doi:10.3892/or.2019.7195

Loke, S. Y., and Lee, A. S. G. (2018). The Future of Blood-Based Biomarkers for the Early Detection of Breast Cancer. *Eur. J. Cancer* 92, 54–68. doi:10.1016/j.ejca.2017.12.025

Lv, Y.-J., Wang, W., Ji, C.-S., JiaXie, M. R., Xie, M.-R., and Hu, B. (2017). Association between Periostin and Epithelial-Mesenchymal Transition in Esophageal Squamous Cell Carcinoma and its Clinical Significance. *Oncol. Lett.* 14 (1), 376–382. doi:10.3892/ol.2017.6124

Lv, Y., Wang, W., Jia, W.-D., Sun, Q.-K., Li, J.-S., Ma, J.-L., et al. (2013). High-level Expression of Periostin Is Closely Related to Metastatic Potential and Poor Prognosis of Hepatocellular Carcinoma. *Med. Oncol.* 30 (1), 385. doi:10.1007/s12032-012-0385-7

Mroczko, B., Groblewska, M., Łukaszewicz-Zając, M., Bandurski, R., Kędra, B., and Szmitkowski, M. (2009). Pre-treatment Serum and Plasma Levels of Matrix Metalloproteinase 9 (MMP-9) and Tissue Inhibitor of Matrix Metalloproteinases 1 (TIMP-1) in Gastric Cancer Patients. *Clin. Chem. Lab. Med.* 47 (9), 1133–1139. doi:10.1515/cclm.2009.253

Mroczko, B., Kozłowski, M., Groblewska, M., Łukaszewicz, M., Nikliński, J., Jelski, W., et al. (2008). The Diagnostic Value of the Measurement of Matrix Metalloproteinase 9 (MMP-9), Squamous Cell Cancer Antigen (SCC) and Carcinoembryonic Antigen (CEA) in the Sera of Esophageal Cancer Patients. *Clinica Chim. Acta* 389 (1-2), 61–66. doi:10.1016/j.cca.2007.11.023

Ren, Y., Deng, R., Zhang, Q., Li, J., Han, B., and Ye, P. (2020). Bioinformatics Analysis of Key Genes in Triple Negative Breast Cancer and Validation of Oncogene PLK1. *Ann. Transl Med.* 8 (24), 1637. doi:10.21037/atm-20-6873

Shan, M.-j., Meng, L.-b., Guo, P., Zhang, Y.-m., Kong, D., and Liu, Y.-b. (2021). Screening and Identification of Key Biomarkers of Gastric Cancer: Three Genes Jointly Predict Gastric Cancer. *Front. Oncol.* 11, 591893. doi:10.3389/fonc.2021.591893

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Siriwardena, B. S. M. S., Kudo, Y., Ogawa, I., Kitagawa, M., Kitajima, S., Hatano, H., et al. (2006). Periostin Is Frequently Overexpressed and Enhances Invasion and Angiogenesis in Oral Cancer. *Br. J. Cancer* 95 (10), 1396–1403. doi:10.1038/sj.bjc.6603431

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Sung, P.-L., Jan, Y.-H., Lin, S.-C., Huang, C.-C., Lin, H., Wen, K.-C., et al. (2016). Periostin in Tumor Microenvironment Is Associated with Poor Prognosis and Platinum Resistance in Epithelial Ovarian Carcinoma. *Oncotarget* 7 (4), 4036–4047. doi:10.18632/oncotarget.6700

Tan, E. M., and Zhang, J. (2008). Autoantibodies to Tumor-Associated Antigens: Reporters from the Immune System. *Immunol. Rev.* 222, 328–340. doi:10.1111/j.1600-065X.2008.00611.x

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: an Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* 47 (W1), W556–w560. doi:10.1093/nar/gkz430

Wang, H., Huang, L., Chen, L., Ji, J., Zheng, Y., and Wang, Z. (2021a). Identification of Novel Biomarkers Related to Lung Squamous Cell Carcinoma Using Integrated Bioinformatics Analysis. *Comput. Math. Methods Med.* 2021, 1–18. doi:10.1155/2021/9059116

Wang, S., Qin, J., Ye, H., Wang, K., Shi, J., Ma, Y., et al. (2018). Using a Panel of Multiple Tumor-Associated Antigens to Enhance Autoantibody Detection for Immunodiagnosis of Gastric Cancer. *Oncoimmunology* 7 (8), e1452582. doi:10.1080/2162402x.2018.1452582

Wang, W., Sun, Q. K., He, Y. F., Ma, D. C., Xie, M. R., Ji, C. S., et al. (2014). Overexpression of Periostin Is Significantly Correlated to the Tumor Angiogenesis and Poor Prognosis in Patients with Esophageal Squamous Cell Carcinoma. *Int. J. Clin. Exp. Pathol.* 7 (2), 593–601.

Wang, Y., Wang, P., Liu, M., Zhang, X., Si, Q., Yang, T., et al. (2021b). Identification of Tumor-Associated Antigens of Lung Cancer: SEREX Combined with Bioinformatics Analysis. *J. Immunological Methods* 492, 112991. doi:10.1016/j.jim.2021.112991

Xu, Y.-W., Liu, C.-T., Huang, X.-Y., Huang, L.-S., Luo, Y.-H., Hong, C.-Q., et al. (2017). Serum Autoantibodies against STIP1 as a Potential Biomarker in the Diagnosis of Esophageal Squamous Cell Carcinoma. *Dis. Markers*, 2017, 1–9. doi:10.1155/2017/5384091

Yang, J., Lin, J., An, J., Zhao, Y., Jing, S., Yu, M., et al. (2021). TRIB3 Promotes the Malignant Progression of Bladder Cancer: An Integrated Analysis of Bioinformatics and *In Vitro* Experiments. *Front. Genet.* 12, 649208. doi:10.3389/fgene.2021.649208

Zhao, Y., Hu, S., Zhang, J., Cai, Z., Wang, S., Liu, M., et al. (2021). Glucoside Xylosyltransferase 2 as a Diagnostic and Prognostic Marker in Gastric Cancer *via* Comprehensive Analysis. *Bioengineered* 12 (1), 5641–5654. doi:10.1080/21655979.2021.1967067

Zheng, S.-X., Xiang, B.-D., Long, J.-M., Qu, C., Mo, Z.-J., Li, K., et al. (2018). Diagnostic Value of Serum SMP30 and Anti-SMP30 Antibody in Hepatocellular Carcinoma. *Lab. Med.* 49 (3), 203–210. doi:10.1093/labmed/lmy004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xie, Sun, Zhu, Wang, Han and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fgene.2022.870191

Altered Expressions of NF1 and NF1-Related microRNAs as Biomarkers in the Diagnosis of Undifferentiated Pleomorphic Sarcoma

Peng Zhang[1]*, *Lingling Huang*[1], *Pengwei Ma*[2] and *Xiaoying Niu*[1]

[1]Department of Bone and Soft Tissue Cancer, The Affiliated Cancer Hospital of Zhengzhou University (Henan Cancer Hospital), Zhengzhou, China, [2]Department of Toxicology, College of Public Health, Zhengzhou University, Zhengzhou, China

Objective: Undifferentiated pleomorphic sarcoma (UPS) is a highly malignant, aggressive, and pleomorphic subtype of soft tissue sarcoma in adults. However, UPS is difficult to be diagnosed due to the lack of specific morphological and immunophenotypic features. Here, we aimed to identify new biomarkers for the diagnosis of UPS.

Methods: The mRNA and protein expression of neurofibromin 1 (NF1) in 68 pairs of UPS and adjacent normal tissues were detected by qRT-PCR and immunohistochemistry, and the correlation between the NF1 protein expression and clinicopathological characteristics was analyzed. Then, differentially expressed microRNAs (DE miRNAs) were identified between the UPS tumor tissue and matched adjacent normal tissue using Hisep sequencing, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG). The DE miRNAs of the regulating NF1 gene were also identified using the TargetScan and miRanda databases and validated by qRT-PCR.

Results: Compared with the adjacent normal tissue, both mRNA and protein expressions of NF1 in the UPS tumor tissue were significantly decreased, and the positive rate of NF1 protein was associated with the tumor size, metastasis, and recurrence. A total of 125 known DE miRNAs were identified from the screened miRNAs based on $|\log_2(\text{Fold Change}) \geq 5$ and $p\text{-value} < 0.05$ (A total of 82 upregulated and 43 downregulated DE miRNAs in the UPS tissue). Target genes regulated by the DE miRNAs were enriched in pathways of metabolisms, RNA degradation, PI3K-Akt, and Rap1 pathway. In total, 11 miRNAs which were predicted to regulate the *NF1* gene were screened. After verification, the relative expressions of hsa-miR-199a-3p and hsa-miR-34a-5p were increased and decreased in the UPS tumor tissue compared with those in the adjacent normal tissue, respectively.

Conclusion: NF1 and NF1-related microRNAs including hsa-miR-199a-3p and hsa-miR-34a-5p may be novel biomarkers in the diagnosis of undifferentiated pleomorphic sarcoma (UPS).

Keywords: soft tissue sarcomas, undifferentiated pleomorphic sarcoma, NF1, microRNA, biomarkers

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Weifeng Liu,
Peking University, China
Yi Luo,
Central South University, China

***Correspondence:**
Peng Zhang
heartnesszp@126.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *06 February 2022*
Accepted: *11 March 2022*
Published: *26 April 2022*

Citation:
Zhang P, Huang L, Ma P and Niu X (2022) Altered Expressions of NF1 and NF1-Related microRNAs as Biomarkers in the Diagnosis of Undifferentiated Pleomorphic Sarcoma. Front. Genet. 13:870191. doi: 10.3389/fgene.2022.870191

INTRODUCTION

Soft tissue sarcomas (STSs) are rare malignant tumors of mesenchymal origin with high aggressiveness and heterogeneity, including more than 50 histological subtypes. Incidence of STSs accounts for 1% in adult malignant tumors (Gamboa et al., 2020). The incidence rate of STSs is low, but they are malignant and easy for metastasis and reoccurrence. Undifferentiated pleomorphic sarcoma (UPS) is a highly malignant and pleomorphic subtype of STSs in adults (Winchester et al., 2018), accounting for more than 20% of STSs, which is the highest than the other subtypes of STSs (Ray-Coquard et al., 2018). However, UPS is considered an exclusionary diagnosis due to the lack of specific morphological and immunophenotypic features (Roland et al., 2016), and its molecular changes have not yet been clarified.

Genetic alterations are always associated with oncogenesis. Neurofibromin 1 (NF1) is derived from neurofibromatosis type I, which is an autosomal dominant genetic disease with the mutation of the *NF1* gene. In addition, *NF1* is a tumor suppressor gene; NF1 mutation was first demonstrated in the malignant peripheral nerve sheath tumor (MPNST), which is a subtype of STSs (Amirnasr et al., 2020). In addition to that, its mutations were found in other subtypes of STSs. For example, there was the deletion of the *NF1* gene in pediatric rhabdomyosarcoma (Walther et al., 2016); 20% cases with mutation in the *NF1* gene were observed in 86 liposarcomas using the whole-exome sequencing (Kanojia et al., 2015). In a mouse study, the mice with NF1 deletion developed either high-grade myogenic sarcomas or MPNSTs (Dodd et al., 2013), and in our previous study, we found that individuals carrying the TC/CC genotype for NF1 rs2905789 may be susceptible to STSs (Zhang et al., 2021). Since UPS is a main kind of myogenic sarcomas and the constituent ratio of UPS is the highest among the subtypes of STSs, NF1 may be involved in the development of UPS. NF1 protein is a tumor suppressor protein; impaired NF1 may increase the Ras activity and then activate the PI3K-Akt signal pathway, which regulates tumor cell growth, survival, and angiogenesis. However, there is no report on the expression of NF1 in UPS.

MicroRNAs (miR or miRNAs) are a class of single-stranded 19–25 ribonucleotide noncoding RNAs (Saliminejad et al., 2019). They are involved in regulating the expression of about 60% of the coding genes and are widely involved in almost all the physiological processes, such as cell differentiation, proliferation, and apoptosis (Wang J. et al., 2021). In recent years, many studies have found that miRNAs are abnormally expressed in the tumor tissue and have confirmed that miRNAs are closely related to tumorigenesis (Li et al., 2020, Tang et al., 2021). Previous studies have demonstrated that antioncogene *NF1* may be regulated by some miRNAs during the occurrence of lung squamous cell carcinoma (Guo et al., 2020), ovarian cancer (Su et al., 2019), and melanoma (Stark et al., 2015), but the miRNA profile in the UPS sarcoma tissue and miRNAs, which regulate NF1 in the initiation of UPS, are still unknown.

In this study, we first investigated the expression level, clinical significance, and signal pathway of NF1 in the UPS sarcoma tissue; then, we identified the miRNA profile between the UPS tumor tissue and adjacent normal tissue and screened the miRNAs that may regulate the *NF1* gene that are expressed differentially. More understanding of molecular changes involved in UPS may provide more clues and ideas of identifying biomarkers for the diagnosis and treatment of UPS patients.

MATERIALS AND METHODS

Subjects and Sample Collection

A total of 68 patients with primary UPS without radiotherapy and chemotherapy before surgery at the Department of Bone and Soft Tissue Sarcoma, Henan Cancer Hospital in China from 2006 to December 2016 were chosen. Among the patients, there were 39 males and 29 females, aged 55.58 ± 3.94 years.

The paired UPS tissue and normal muscle tissue which are more than 5 cm away from the sarcoma were surgically obtained, which were stored for immunohistochemistry and RNA isolation. All the patients were surveyed by a questionnaire about the basic information and signed the informed consent, and the study was approved by the Medical Ethics Committee of Henan Cancer Hospital.

Immunohistochemistry

The protein expression of NF1, as well as the phosphorylated proteins of p-Akt, p-mTOR, and p-S6 in the adjacent normal tissue and UPS sarcoma tissue of 68 patients, was assessed by immunohistochemistry (IHC). Briefly, the specimens were deparaffinized, blocked with goat serum for 30 min, and incubated with the rabbit antihuman NF1 antibody (1: 100, Beijing Boosen Biotechnology Co., Ltd., China) at 4°C overnight; then, they were incubated with biotinylated goat antirabbit immunoglobulin at a concentration of 1:100 at 37°C for 30 min. All the immunohistochemical-stained tissue sections were assessed independently by two pathologists in a blinded manner, and a consensus was reached for each score. If there was a disagreement between these two pathologists, a third pathologist would be invited, and the three pathologists would come to the final conclusion. Scoring was based on the percentage of the positively stained cells (0=<5%; 1 = 6–25%; 1 = 6–25%; 2 = 26–50%; 3 = 51–75%; and 4 = 76–100%) under five high-power vision fields, and the intensity of staining was graded as negative (score 0), weak (score 1), moderate (score 2), or strong (score 3). The final staining scores were calculated as percentage × staining intensity. Therefore, the final scores were score 0 = value 0–1 (**negative expression**), score 1 = value 2–4 (**low expression**), and score 2 = value 5–8 and score 3 = value 9–12 (**high expression**). The ratio of the positive area to the total area for the protein expression was analyzed by ImageJ software. NF1 protein was analyzed by two ranks of expression, and p-Akt, p-mTOR, and p-S6 were analyzed using the ratio of the positive area.

RNA Extraction

The total RNA including miRNAs was isolated from approximately 100 mg of three pairs of the sarcoma tissue and adjacent normal tissue using TRIzol reagent (Invitrogen). The concentration and quality of the RNA were determined by using a NanoDrop Spectrophotometer (NanoDrop Technologies,

Thermo Fisher, United States). The individual aliquots of RNA from the UPS tissue or adjacent normal tissue were pooled for Hisep deep sequencing. The extracted RNA solution of 68 pairs of tissues is stored at −80°C for the quantitative real-time PCR (qRT-PCR).

Hisep Sequencing

The RNA pools consisted of the same amounts of total RNA from the UPS tissue and matched adjacent normal tissue, and the total RNAs were used for Hisep deep sequencing. First, the total RNA was separated by polyacrylamide gel electrophoresis, and the small RNA regions of 18–30 nucleotides were excised. Then, 5′-adapter and 3′-adapter were ligated to the small RNAs, and the small RNA-adapter molecules were reverse transcribed and amplified. Finally, the two miRNA libraries were constructed and Hisep sequenced with Illumina Hiseq 2500 (Illumina, Inc., United States). Compared with the adjacent normal tissue, a limma test was used to identify the differentially expressed miRNAs in the UPS tissue using the DEGseq R language package. Finally, the dysregulated miRNAs were chosen based on $|\log_2(\text{Fold Change})| \geq 2$ and p-value < 0.05.

GO and KEGG Analysis

Using the databases of TargetScan and miRanda and taking the intersection, we predicted the target genes of the differentially expressed miRNAs. The target genes were annotated from four aspects of functional items, such as the biological process, molecular function, cellular component, and signaling pathway using the databases of the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). The differentially expressed miRNAs of the regulating *NF1* gene were identified using the TargetScan and miRanda databases.

qRT-PCR

The miDETECT A Track™ miRNA qRT-PCR Starter Kit (Ribobio Biotechnology Co., Guangzhou, China) was used for miRNA detection. First, poly (A) tails were added to RNAs, and the poly (A) tailing reaction system was made as follows: 1.0 μL total RNA, 2.0 μL 5×poly (A) polymerase buffer, 1.0 μL poly (A) polymerase, and RNase-free water were added to 10.0 μL. The mixture was kept at 37°C for 1 h. Second, cDNA was synthesized with specific primers. We made the reverse transcription reaction system as follows: 4.0 μL RTase mix, 2.0 μL mi*DETECT* A Track™ Uni-Reverse Primer, and 10.0 μL tail product. The reaction mixtures were incubated at 42°C for 1 h, at 72°C for 10 min, and then stored at 4°C. The amplification reaction was performed in a 20 μL volume containing 2.0 μL cDNA, 0.5 μL mi*DETECT* A Track™ miRNA-Forward Primer and Reverse Primer, respectively (Ribobio Biotechnology Co., Guangzhou, China), 10.0 μL of 2×SYBR Green Mix and 7.0 μL RNase-free water. qPCR was performed as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 2 s, 60°C for 20 s, and 70°C for 10 s. U6 snRNA was used as a control for the normalization of tissue miRNA levels. The primer sequences are as follows: for human NF1: forward 5′-TGG GACATTCGCCTCTTAAC-3′ and reverse 5′-ACACATGCA AAATGGGAACA-3'; human GAPDH: forward 5′-GGAAGC TTGTCATCAATGGAAATC-3′ and reverse 5′-TGATGA CCCTTTTGGCTCCC-3'.

The relative expression levels of miRNAs and NF1 mRNA were calculated using a comparative Ct ($2^{-\Delta\Delta Ct}$) method.

Statistical Analysis

The SPSS21.0 software package was used for the data statistical analysis. The histogram was drawn with GraphPad Prism 5. The continuity variable was expressed as the mean ± SD, and the quantitative data were analyzed using two independent samples. Student's t-test and the Pearson's Chi-squared (x^2) test was used to evaluate the positive rates of NF1 between the UPS sarcoma tissue and adjacent normal tissue and the difference between the protein levels of NF1 staining in the UPS sarcoma tissue and clinicopathological characteristics, respectively. The p value less than 0.05 was considered statistically significant.

RESULTS

The Expression of NF1 in the UPS Sarcoma Tissue and Adjacent Normal Tissue

Figure 1A demonstrated that the NF1 mRNA level in the UPS tissue was decreased compared to that in the adjacent normal tissue, and the difference was statistically significant (p < 0.05). NF1 protein was stained mainly in the nucleus, and a high expression of protein was considered as a positive expression. In the 68 cases of the adjacent normal tissue, there were 48 cases with NF1 high protein expression (**Figure 1B** a) and 20 cases with NF1 low/ no protein expression (**Figure 1B** b); In the 68 cases of the UPS sarcoma tissue, there were 27 cases with NF1 high protein expression (**Figure 1B** c) and 41 cases with NF1 low/no protein expression (**Figure 1B** d). In the adjacent normal tissue, the positive rate of NF1 protein was 70.59%, and in the UPS tissue, the positive rate of NF1 protein was 39.71%. The positive rate of NF1 protein in the UPS sarcoma tissue was decreased compared to that in the adjacent tissue, and the difference was statistically significant (p < 0.05) (**Table 1**).

Correlation Between the NF1 Protein Expression and Clinicopathological Characteristics of UPS Patients

We divided the high expression and low/negative expression groups based on the protein expression of NF1 and assessed the correlation between the NF1 protein expression and clinicopathological characteristics of the UPS patients. The results (**Table 2**) showed that the NF1 protein expression was associated with the tumor size (x^2 = 7.372 and p = 0.007), distant metastasis (x^2 = 9.378 and p = 0.002), and recurrence (x^2 = 4.300 and p = 0.047), but NF1 was not associated with gender, age, lymph node metastasis, location, or 5-year survival.

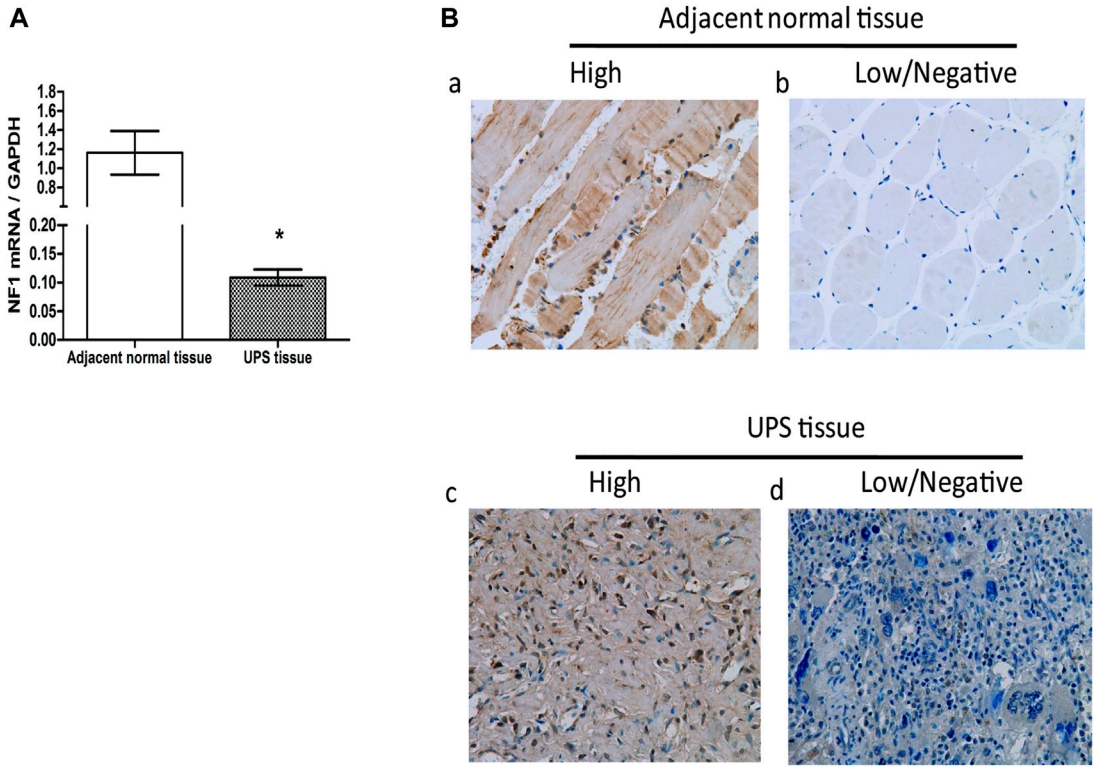


FIGURE 1 | Expression of NF1 in the UPS sarcoma tissue and adjacent normal tissue. **(A)**: NF1 mRNA expression in the UPS sarcoma tissue and adjacent normal tissue by qRT-PCR, *: vs. adjacent normal tissue, $p < 0.05$; **(B)** Representative images of the protein expression of NF1 in the UPS sarcoma tissue and adjacent normal tissue by immunohistochemistry (200×) (a): High protein expression of NF1 in the adjacent normal tissue; (b): Low or negative protein expression of NF1 in the adjacent normal tissue; (c): High protein expression of NF1 in the UPS sarcoma tissue; (d): Low or negative protein expression of NF1 in the UPS sarcoma tissue.

The Protein Levels of p-Akt, p-mTOR, and p-S6 Were Increased in the UPS Sarcoma Tissue

The phosphorylated protein of p-Akt, p-mTOR, and p-S6 in the adjacent normal tissue and UPS sarcoma tissue from the 68 patients were detected by immunohistochemistry (**Figure 2A**). p-Akt and p-mTOR were stained brown color in both the cytoplasm and nucleus, and p-S6 was stained in the cytoplasm. The ratios of the positive area to the total area of p-Akt, p-mTOR, and p-S6 determined by ImageJ software were (17.44 ± 2.73)%, (18.32 ± 1.90)%, and (22.52 ± 2.39)%, which were significantly increased compared to those in the adjacent tissue ((6.41 ± 1.60)%, (7.25 ± 1.74)%, and (4.29 ± 1.50)%, respectively) ($p < 0.05$) (**Figure 2B**).

Hisep Sequencing of Noncoding RNA in the Tissue

Hisep sequencing yielded a total of 22,273,789 and 22,506,782 raw sequence reads in the adjacent normal tissue and UPS tissue, respectively. The types and numbers of the small RNAs (sRNAs) mapped to the genome by SOAP and bowtie are shown in **Supplementary Table S1** and **Supplementary Figures S1A, B**. **Supplementary Figure S1C** demonstrated that the size distribution of the reads indicated that the majority of sRNAs were 22 nt in the tissue. The percentages of miRNAs in total sRNAs in the adjacent normal tissue and UPS tissue were 67.76 and 77.97%, respectively (**Supplementary Table S1**), which indicated that the sRNAs were highly enriched in the miRNA sequences.

Analysis of Differentially Expressed miRNAs

A total of 496 known differentially expressed miRNAs were identified from the screened miRNAs with Hisep sequencing based on $|\log_2(\text{Fold Change})| \geq 2$ and p-value < 0.05. The scatter plot demonstrated in **Figure 3** showed that there were many dysregulated miRNAs between the adjacent normal tissue and UPS tissue. In order to pick up more important miRNAs with a differential expression, more restrictive criteria were used as follows: more than 50 copies, $|\log_2(\text{Fold Change})| \geq 5$ and p-value < 0.05. According to these principles, 125 differentially expressed known miRNAs were identified, of which 82 were upregulated (**Supplementary Table S2**) and 43 downregulated (**Supplementary Table S3**) in the UPS tissue. Top 10 of the upregulated and downregulated differentially expressed known miRNAs are shown in **Table 3**.

TABLE 1 | Protein expression of NF1 in the UPS tissue and adjacent normal tissue.

Group	N	NF1 (n, %)	
		High	Low/Negative
Adjacent normal tissue	68	48(70.59)	20 (29.41)
UPS tissue	68	27(39.71)	41 (60.29)
χ^2		13.110	
P		**0.001**	

Bold values represents p < 0.05.

The Analysis of Gene Ontology Enrichment and Kyoto Encyclopedia of Genes and Genomes Pathways With Target Genes Regulated by DE miRNAs

A total of 51,449 target genes which may be regulated by 496 DE miRNAs were identified from the intersection of TargetScan and miRanda databases. In order to investigate the possible mechanism for the initiation of UPS, we conducted the analysis of GO enrichment including the biological process (BP), cellular component (CC), molecular function (MF), and KEGG signal pathway. The BP demonstrated the physical activities regulated by the target genes associated with DE miRNAs, which were presented with the cellular process, single-organism process, metabolic process, and biological regulation (**Figure 4A**). The CC showed the location of transcription and protein translation of the target genes. The cell, cell part, organelle, and membrane were in the CC category (**Figure 4B**). The MF defined the role of the target genes regulated by DE miRNAs. The target genes were predicted to regulate binding, catalytic activity, molecular transducer activity, and receptor activity (**Figure 4C**). These target genes were also enriched in some important signal pathways, such as the metabolic pathway, RNA degradation, PI3K-Akt pathway, and Rap1 pathway (**Figure 4D**).

DE miRNAs Regulating NF1 Were Predicted and Verified

To screen the miRNAs which were differentially expressed and predicted to regulate the gene of NF1 from 125 DE known miRNAs ($\log_2$ (Fold Change) | ≥5 and *p*-value < 0.05), first, the miRNAs that may regulate NF1 were chosen from the databases of TargetScan and miRanda and then were intersected with the screened 125 DE known miRNAs; finally, 11 miRNAs were screened, which are demonstrated in **Table 4**.

Based on the literature studies and fold change values of the differential expression, hsa-miR-127-3p and hsa-miR-199a-3p, which were the first and the third of upregulation, and hsa-miR-34a-5p, which was the first of the downregulation in the UPS tissue, were selected to be verified in the 68 pairs of the UPS tissue and adjacent normal tissue. **Figure 5** illustrated that the relative expression of the human miR-199a-3p and miR-34a-5p were increased and decreased in the UPS tumor tissue compared with those in the adjacent normal tissue, respectively, and the difference was statistically significant ($p < 0.05$). In addition, there was an upward trend of the level of human miR-127-3p in the UPS tissue compared to that in the adjacent normal tissue, but there was no significant difference ($p > 0.05$).

TABLE 2 | Correlation between the NF1 protein expression and clinicopathological characteristics of UPS patients

Characteristic	N	NF1 protein expression		x^2	*p*
		High (*n* = 27)	Low/Negative (*n* = 41)		
Gender					
Male	39	12	27	3.051	0.081
Female	29	15	14		
Age (years)					
<55	24	11	13	0.582	0.446
≥55	44	16	28		
Tumor size (cm)					
<5	16	11	5	7.372	**0.007**
≥5	52	16	36		
Lymph node metastasis					
Yes	23	7	16	1.248	0.264
No	45	20	25		
Distant metastasis					
Yes	19	2	17	9.378	**0.002**
No	49	27	41		
Location					
Trunk	32	12	20	0.123	0.726
Extremity	36	15	21		
Recurrence					
Yes	28	7	21	4.300	**0.047**
No	40	20	20		
5-year survival					
Survival	45	20	25	1.248	0.305
Dead	23	7	16		

Bold values represents p < 0.05.

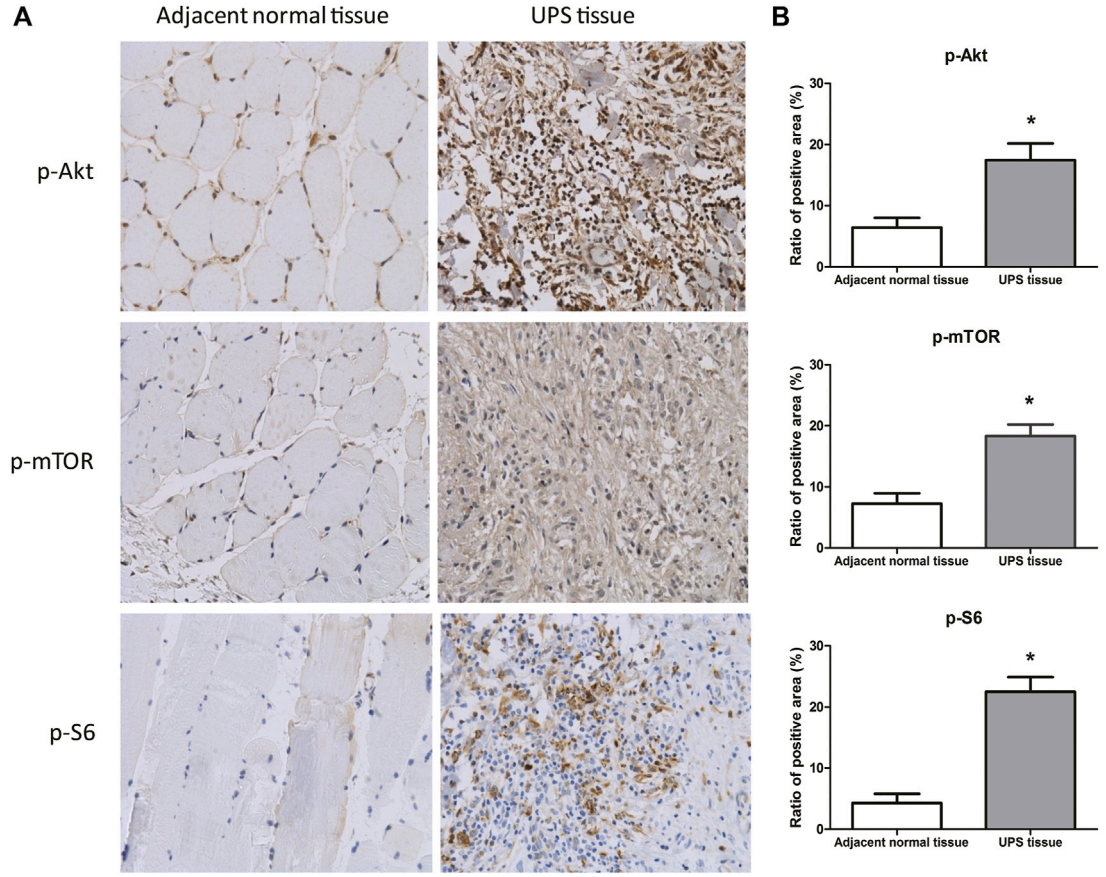


FIGURE 2 | Protein expression of p-Akt, p-mTOR, and p-S6 in the UPS sarcoma tissue and adjacent normal tissue. **(A)**: Representative images of the protein expression of p-Akt, p-mTOR, and p-S6 in the UPS sarcoma tissue and adjacent normal tissue by immunohistochemistry (200×); **(B)**: The comparison of the ratios of the positive area to the total area of p-Akt, p-mTOR, and p-S6 between the UPS sarcoma tissue and adjacent normal tissue. *: vs. adjacent normal tissue, $p < 0.05$.

DISCUSSION

The *NF1* gene is mapped to human chromosome 17 and encodes neurofibromin. Point mutations and genomic deletions of the *NF1* gene were found in 10.5% of the myxofibrosarcomas and 8% of the pleomorphic liposarcomas in 207 samples of STSs (Barretina et al., 2010). However, the NF1 expression in UPS has not been explored. In this study, we detected the mRNA and protein expression of NF1 in the UPS sarcoma tissue and adjacent normal tissue and analyzed the clinicopathological characteristics. Our data demonstrated that not only NF1 mRNA was decreased but also the positive rate of NF1 protein in the UPS tissue (39.71%) was significantly decreased compared with that in the adjacent normal tissue (70.59%). A research reported that 58.1% (93/160) of the gastric cancer samples were NF1-positive as compared to 94.4% (151/160) of the matched normal tissue samples ($p < 0.001$) (Liu et al., 2017), which was consistent with our results. In addition, the NF1 protein expression was associated with the tumor size and distant metastasis of UPS in the present study. In the gastric cancer study, the NF1 protein was associated with the T stage and TNM stage (18). Actually, the T stage means the size or direct extent of the primary tumor, equal to the tumor size. Therefore, our data were consistent with the results of gastric cancer. But NF1 was not associated with gender, age, or lymph node metastasis in UPS. Elzagheid et al. demonstrated that the NF expression showed a borderline correlation with gender; tumors of women showed a higher NF expression than those of males, but there was no significant difference ($p = 0.068$) (Elzagheid et al., 2016). Further study on the expression of NF1 in UPS is of significance to understand the pathogenesis of UPS.

miRNA expression deregulation has been found in many types of cancers (Xu et al., 2014; Komoll et al., 2021), and some reports have indicated that the differentially expressed miRNAs may be employed for the diagnosis of sarcomas (Fricke et al., 2015; Zhou et al., 2016; Smolle et al., 2017). Pazzaglia et al. (2017) performed 12 miRNA expression analyses between 59 primary STS samples (27 leiomyosarcomas (LMS) and 32 UPS) and 10 normal control tissue using the TaqMan microRNA array and found that compared with the normal control tissue, there was a statistical difference only on the increasing expression of miR-152 in STS samples, and the researchers from Finland conducted miRNA profiling on 10 LMS and 10 UPS samples; two cultured human mesenchymal stem cell samples were used as controls.

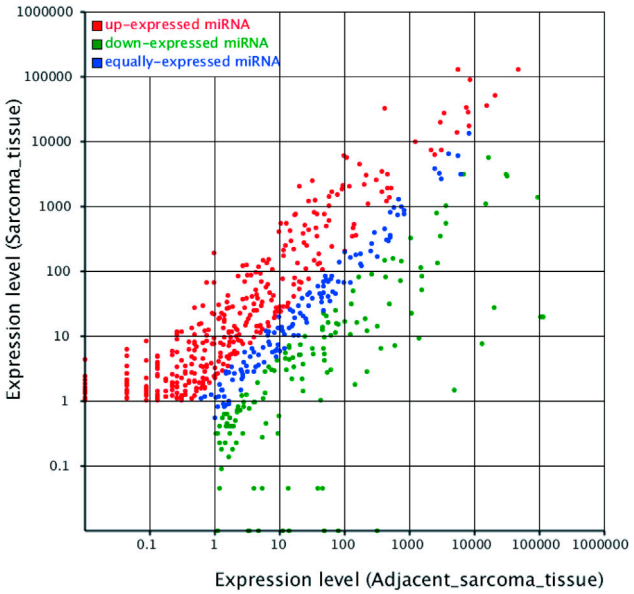


FIGURE 3 | Scatter plot indicating the differentially expressed miRNAs between the adjacent normal tissue and UPS tissue. The principles of identifying the differentially expressed miRNAs were based on $|\log_2(\text{Fold Change})| \geq 2$ and p-value <0.05. Each dot represented one miRNA. The red dots represented upregulated miRNAs, the green dots represented downregulated miRNAs, and the blue dots represented equally-expressed miRNAs.

The results of the miRNA microarray demonstrated that miR-320a and miR-199b-5p were differentially expressed between LMS and UPS, and more samples were detected to confirm the microarray data using qRT-PCR (Guled et al., 2014). This study was carried out to look for the differentially expressed miRNAs for the diagnosis between LMS and UPS. While Pazzaglia et al. took LMS and UPS together as soft tissue sarcoma, the differentially expressed miRNAs in STSs may be less compared to normal control because the miRNA profiles are different in each subtype of STSs. However, in this study, we detected and identified the miRNA profile using miRNA sequencing between the UPS tumor tissue and adjacent normal tissue for the first time. Our data showed that 125 differentially expressed miRNAs were identified in the UPS tissue, according to $|\log_2(\text{Fold Change})| \geq 5$ and p-value < 0.05.

GO analysis was performed to analyze the function of these target genes regulated by the DE miRNAs, which included transcriptional regulation, protein binding, metabolic regulation, and signal transduction. Furthermore, KEGG analysis found that these target genes were enriched in some important signal pathways, such as the metabolic pathway, RNA degradation, PI3K-Akt pathway, and Rap1 pathway. The PI3K/Akt signaling pathway participates in the phosphorylation of numerous molecules that control the cell growth, proliferation, metabolism, and survival. This pathway is commonly activated in cancers (Ghafouri-Fard et al., 2021), including soft tissue sarcoma (Serrano et al., 2016). The PI3K/Akt signaling pathway is related

TABLE 3 | Top 10 upregulated and downregulated differentially expressed known miRNAs in the UPS tumor tissue compared with the adjacent normal tissue.

miRNA ID	Copy number in the adjacent normal tissue	Normalized miRNA level in the adjacent normal tissue	Copy number in the UPS tissue	Normalized miRNA level in the UPS tissue	Fold change	*p* value
hsa-miR-138-5p	460	20.44	44,805	2,011.56	98.42	0
hsa-miR-376a-3p	725	32.21	55,637	2,497.87	77.54	0
hsa-miR-155-5p	2,233	99.21	135,439	6,080.64	61.29	0
hsa-miR-146a-5p	2,421	107.57	125,240	5,622.75	52.27	0
hsa-miR-409-3p	240	10.66	12,023	539.78	50.62	0
hsa-miR-299-3p	61	2.71	2,713	121.80	44.94	0
hsa-miR-31-5p	630	27.99	26,972	1,210.93	43.26	0
hsa-miR-146b-5p	397	17.64	16,985	762.56	43.23	0
hsa-miR-21-3p	392	17.42	16,359	734.45	42.17	0
hsa-miR-214-5p	289	12.84	12,059	541.40	42.16	0
hsa-miR-133a-3p	2,579,106	114,592	438	20	−5,827.40	0
hsa-miR-1-3p	2,398,804	106,581	440	20	−5,395.38	0
hsa-miR-206	295,147	13,114	171	8	−1,708.13	0
hsa-miR-499a-5p	459,811	20,430	612	27	−743.55	0
hsa-miR-95-3p	31,853	1,415	203	9	−155.29	0
hsa-miR-378a-5p	13,374	594	159	7	−83.24	0
hsa-miR-378d	4,958	220	62	3	−79.14	0
hsa-miR-378a-3p	2,171,985	96,504	30,941	1,389	−69.47	0
hsa-miR-486-5p	8,200	364	144	6	−56.35	0
hsa-miR-34a-5p	24,240	1,077	492	22	−48.76	0

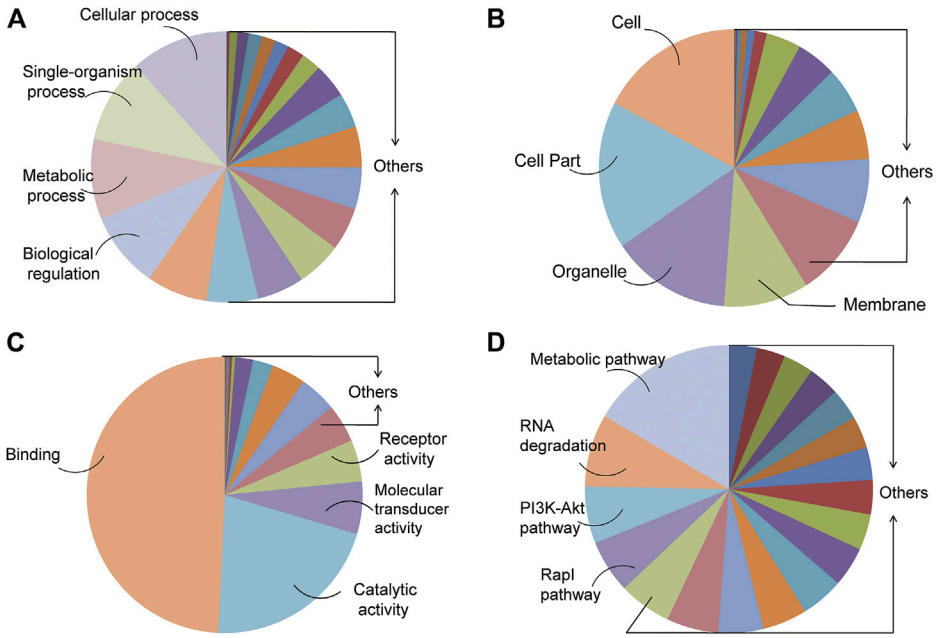


FIGURE 4 | GO and KEGG analysis of genes targeted by the DE miRNAs. **(A–C)** Top four GO biological processes (BPs), cellular components (CCs), and molecular functions (MFs) enriched in target genes regulated by all the DE miRNAs. **(D)** KEGG enrichment analysis of signal pathways.

TABLE 4 | Eleven DE miRNAs regulated the *NF1* gene were predicted using bioinformatics.

Name of the regulated gene	miRNA name	Fold change	*p*	Expression level in the tumor tissue
NF1	**hsa-miR-127-3p**	26.03	0	up
NF1	hsa-miR-654-5p	23.96	0	up
NF1	**hsa-miR-199a-3p**	22.33	0	up
NF1	hsa-miR-199b-3p	22.33	0	up
NF1	hsa-miR-7-5p	20.15	0	up
NF1	hsa-miR-942-5p	8.21	6.67E-147	up
NF1	**hsa-miR-34a-5p**	−48.76	0	down
NF1	hsa-miR-196-5p	−30.30	0	down
NF1	hsa-miR-144-3p	−20.18	3.93E-285	down
NF1	hsa-miR-193-3p	−20.21	0	down
NF1	hsa-miR-24-3p	−13.61	0	down

Bold values represents that the three NF1- related microRNAs were chosen and verified.

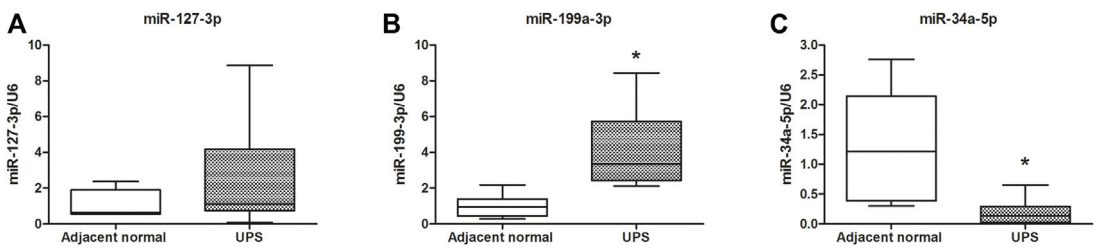


FIGURE 5 | Validation by qRT-PCR of the three miRNAs regulating the *NF1* gene. **(A–C)** were the expression and comparison of miR-127-3p, miR-199a-3p, and miR-34a-5p in 68 pairs of the UPS tumor tissue and adjacent normal tissue. U6 was used as the reference gene. *: vs. adjacent normal tissue, $p < 0.05$.

closely to the tumor suppressor gene *NF1* that could play the role by inhibiting this signaling pathway (Gutmann et al., 2017). Our data showed that the NF1 expression was decreased, but the protein expressions of p-Akt, p-mTOR, and p-S6 were significantly higher in the UPS sarcoma tissue than those in the adjacent normal tissue, which confirmed and verified that the NF1-inhibited PI3K-Akt-mTOR-S6 signal pathway was associated with the occurrence and growth of UPS. Ras-associated protein 1 (Rap1) is a member of the Ras family of small G proteins that regulate a variety of signaling pathways involved in proliferation, differentiation, polarity, and apoptosis (Zhang et al., 2019). Studies revealed that Rap1 is not only a tumor suppressor gene but also a conditioned oncoprotein (Hattori and Minato, 2003). An increased and abnormal activation of Rap1 can lead to tumor formation and malignant development (Shah et al., 2019). The pathways need to be explored further.

miRNAs usually play their function by negatively regulating the gene expression (Lee et al., 2016). When the target gene of miRNA is an oncogene, the decreased expression of miRNA could cause the upregulation of oncogene, which will eventually lead to excessive cell proliferation, reduced apoptosis, and tumor formation. When the target genes of miRNAs are tumor suppressor genes, the miRNA function of oncogenes and the increased expression of these miRNAs can cause downregulation of their target genes (tumor suppressor genes) and promote tumorigenesis. In this study, 11 miRNAs were screened, which were predicted to regulate the gene of NF1 from 125 DE known miRNAs. hsa-miR-127-3p, hsa-miR-199a-3p, and hsa-miR-34a-5p were validated by qRT-PCR in 68 pairs of the UPS tissue and adjacent normal tissue.

The previous literature studies have reported that the differential expression of hsa-miR-127-3p in different tumors is not consistent. For example, hsa-miR-127-3p is downregulated in oral squamous cell carcinoma and inhibits tumor proliferation and metastasis as a tumor suppressor (Ji et al., 2019). Jiang et al. found that hsa-miR-127-3p is downregulated in the glioblastoma tissue compared with the normal brain tissue through the next-generation sequencing analysis of miRNA (Jiang et al., 2014). However, the upregulation of hsa-miR-127-3p in the colorectal cancer tissue was associated with KRAS mutation (Mosakhani et al., 2012). Also, in this study, the results of miRNA sequencing demonstrated that the expression of hsa-miR-127-3p was increased in the UPS tumor tissue, but there was no statistical significance of hsa-miR-127-3p in the later verification, which needs to be validated with more samples.

Many reports have demonstrated that hsa-miR-199a-3p is a kind of oncogenic miRNAs. Zhenqiang Wang (Wang et al., 2014) found that hsa-miR-199a-3p was significantly upregulated in the gastric cancer cell lines and tissue and that hsa-miR-199a-3p dramatically increased the cell proliferation and suppressed cell apoptosis both *in vitro* and *in vivo*, which is a tumor promoter. Another study (Wan et al., 2013) reported that the hsa-miR-199a-3p expression was significantly upregulated in the colorectal cancers tissue than the normal adjacent tissue, and a high hsa-miR-199a-3p expression contributed to more advanced lymphatic invasion, lymph node metastasis, liver metastases, late TNM stage of colorectal cancer, and shorter overall survival rate. Furthermore, the hsa-miR-199a-3p inhibitor could markedly inhibit the colon cancer cell proliferation and induce more cell apoptosis. But the downregulation of hsa-miR-199a-3p was observed in the tissue of esophageal squamous cell carcinoma (Hou et al., 2021) and hepatocellular carcinoma (Giovannini et al., 2018). From the researches abovementioned, we found that the expression of hsa-miR-199a-3p is not consistent. In the present study, the hsa-miR-199a-3p expression is increased in the UPS tissue.

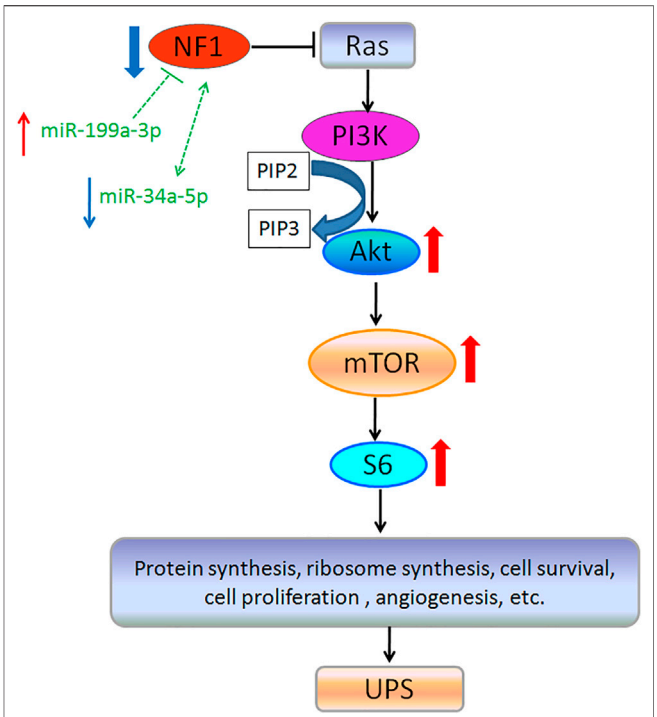


FIGURE 6 | Conclusion image of NF1 and NF1-related miRNAs involved in signal pathways and occurrence of UPS. Note: Blue arrows: downregulated expression; Red arrows: upregulated expression; and Dotted line: Need to be validated.

The relative expression and function of hsa-miR-34a-5p is similar. The expression level of hsa-miR-34a-5p is decreased in neuroblastoma, and hsa-miR-34a-5p may act as a tumor suppressor (Wang Y. et al., 2021). Ding et al. (2017) reported that the level of hsa-miR-34a-5p was downregulated in the ovarian cancer cells, and the hsa-miR-34a-5p overexpression suppressed the ovarian cancer cell proliferation and triggered apoptosis. In addition, hsa-miR-34a-5p was downregulated in the pancreatic cancer tissue, which was associated with proliferation, metastasis, and invasion of the pancreatic cancer cells (Sun et al., 2018). In this study, hsa-miR-34a-5p is downregulated in the UPS tumor tissue, which was consistent with the previous studies performed in other cancers. It inferred that hsa-miR-34a-5p might be a tumor suppressor.

In a word, in this study, first the mRNA and protein expression of NF1 was decreased in the UPS tumor tissue,

and NF1 protein was associated with some clinical characteristic, such as tumor size, distant metastasis, and recurrence *via* the PI3K-Akt-mTOR-S6 signal pathway. Second, we identified the miRNA profile and screened 125 known differently expressed miRNAs of the UPS tumor tissue compared with the matched adjacent normal tissue using microRNA sequencing. In total, 11 miRNAs that may regulate the *NF1* gene was screened among the differently expressed miRNAs, and three miRNAs were verified; we found that hsa-miR-199a-3p was upregulated, and hsa-miR-34a-5p was downregulated (**Figure 6**). In this study, the altered expression of NF1 and NF1-related microRNAs, such as hsa-miR-199a-3p and hsa-miR-34a-5p may be biomarkers, which provide a basis for the novel strategy of UPS diagnosis in future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA806207, PRJNA806207.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Henan Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PZ designed the study and wrote the manuscript; LH and PM performed the experiments; XN collected samples.

FUNDING

This study was financially supported by the grant of the Education Department of Henan Province (No. 20B320042) and the grant of Scientific and technological innovation outstanding young talent training project from Health Commission of Henan Province (No. YXKC2021031).

ACKNOWLEDGMENTS

The authors appreciate everyone who contributed to finish this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.870191/full#supplementary-material

Supplementary Figure S1 | Size distribution and annotation of small RNAs from the adjacent normal tissue and UPS tissue. The proportions of various kinds of small RNAs detected in the adjacent normal tissue **(A)** and UPS tissue **(B)**. **(C)** Frequency percent of length distribution (nt) of the annotated reads of the miRNAs from the adjacent normal tissue and UPS tissue.

REFERENCES

Amirnasr, A., Verdijk, R. M., van Kuijk, P. F., Kartal, P., Vriends, A. L. M., French, P. J., et al. (2020). Deregulated microRNAs in Neurofibromatosis Type 1 Derived Malignant Peripheral Nerve Sheath Tumors. *Sci. Rep.* 10 (1), 2927. doi:10.1038/s41598-020-59789-4

Barretina, J., Taylor, B. S., Banerji, S., Ramos, A. H., Lagos-Quintana, M., Decarolis, P. L., et al. (2010). Subtype-specific Genomic Alterations Define New Targets for Soft-Tissue Sarcoma Therapy. *Nat. Genet.* 42 (8), 715–721. doi:10.1038/ng.619

Ding, N., Wu, H., Tao, T., and Peng, E. (2017). NEAT1 Regulates Cell Proliferation and Apoptosis of Ovarian Cancer by miR-34a-5p/BCL2. *Ott* Vol. 10, 4905–4915. doi:10.2147/ott.s142446

Dodd, R. D., Mito, J. K., Eward, W. C., Chitalia, R., Sachdeva, M., Ma, Y., et al. (2013). NF1 Deletion Generates Multiple Subtypes of Soft-Tissue Sarcoma that Respond to MEK Inhibition. *Mol. Cancer Ther.* 12 (9), 1906–1917. doi:10.1158/1535-7163.mct-13-0189

Elzagheid, A., Emaetig, F., Elsaghayer, W., Torjman, F., Latto, M., Syrjänen, K., et al. (2016). Neurofibromin Expression Is Associated with Aggressive Disease and Poor Outcome in Colorectal Carcinoma. *Ar* 36 (10), 5301–5306. doi:10.21873/anticanres.11102

Fricke, A., Ullrich, P. V., Heinz, J., Pfeifer, D., Scholber, J., Herget, G. W., et al. (2015). Identification of a Blood-Borne miRNA Signature of Synovial Sarcoma. *Mol. Cancer* 14, 151. doi:10.1186/s12943-015-0424-z

Gamboa, A. C., Gronchi, A., and Cardona, K. (2020). Soft-tissue Sarcoma in Adults: An Update on the Current State of Histiotype-specific Management in an Era of Personalized Medicine. *CA A. Cancer J. Clin.* 70 (3), 200–229. doi:10.3322/caac.21605

Ghafouri-Fard, S., Abak, A., Tondro Anamag, F., Shoorei, H., Majidpoor, J., and Taheri, M. (2021). The Emerging Role of Non-coding RNAs in the Regulation of PI3K/AKT Pathway in the Carcinogenesis Process. *Biomed. Pharmacother.* 137, 111279. doi:10.1016/j.biopha.2021.111279

Giovannini, C., Fornari, F., Dallo, R., Gagliardi, M., Nipoti, E., Vasuri, F., et al. (2018). MiR-199-3p Replacement Affects E-Cadherin Expression through Notch1 Targeting in Hepatocellular Carcinoma. *Acta Histochem.* 120 (2), 95–102. doi:10.1016/j.acthis.2017.12.004

Guled, M., Pazzaglia, L., Borze, I., Mosakhani, N., Novello, C., Benassi, M. S., et al. (2014). Differentiating Soft Tissue Leiomyosarcoma and Undifferentiated Pleomorphic Sarcoma: A miRNA Analysis. *Genes Chromosomes Cancer* 53 (8), 693–702. doi:10.1002/gcc.22179

Guo, L., Li, B., Yang, J., Shen, J., Ji, J., and Miao, M. (2020). Fibroblast-derived E-xosomal microRNA-369 P-otentiates M-igration and I-nvasion of L-ung S-quamous C-ell C-arcinoma C-ells via NF1-mediated MAPK S-ignaling P-athway. *Int. J. Mol. Med.* 46 (2), 595–608. doi:10.3892/ijmm.2020.4614

Gutmann, D. H., Ferner, R. E., Listernick, R. H., Korf, B. R., Wolters, P. L., and Johnson, K. J. (2017). Neurofibromatosis Type 1. *Nat. Rev. Dis. Primers* 3, 17004. doi:10.1038/nrdp.2017.4

Hattori, M., and Minato, N. (2003). Rap1 GTPase: Functions, Regulation, and Malignancy. *J. Biochem.* 134 (4), 479–484. doi:10.1093/jb/mvg180

Hou, G., Wang, Y., Zhang, M., Hu, Y., Zhao, Y., Jia, A., et al. (2021). miR-199a-3p Suppresses Progression of Esophageal Squamous Cell Carcinoma through Inhibiting mTOR/p70S6K Pathway. *Anti-cancer drugs* 32 (2), 157–167. doi:10.1097/cad.0000000000000983

Ji, L., Zhu, Z. N., He, C. J., and Shen, X. (2019). MiR-127-3p Targets KIF3B to Inhibit the Development of Oral Squamous Cell Carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 23 (2), 630–640. doi:10.26355/eurrev_201901_16877

Jiang, H., Jin, C., Liu, J., Hua, D., Zhou, F., Lou, X., et al. (2014). Next Generation Sequencing Analysis of miRNAs: MiR-127-3p Inhibits Glioblastoma Proliferation and Activates TGF-β Signaling by Targeting SKI. *OMICS: A J. Integr. Biol.* 18 (3), 196–206. doi:10.1089/omi.2013.0122

Kanojia, D., Nagata, Y., Garg, M., Lee, D. H., Sato, A., Yoshida, K., et al. (2015). Genomic Landscape of Liposarcoma. *Oncotarget* 6 (40), 42429–42444. doi:10.18632/oncotarget.6464

Komoll, R.-M., Hu, Q., Olarewaju, O., von Döhlen, L., Yuan, Q., Xie, Y., et al. (2021). MicroRNA-342-3p Is a Potent Tumour Suppressor in Hepatocellular Carcinoma. *J. Hepatol.* 74 (1), 122–134. doi:10.1016/j.jhep.2020.07.039

Lee, E., Ito, K., Zhao, Y., Schadt, E. E., Irie, H. Y., and Zhu, J. (2016). Inferred miRNA Activity Identifies miRNA-mediated Regulatory Networks Underlying Multiple Cancers. *Bioinformatics* 32 (1), 96–105. doi:10.1093/bioinformatics/btv531

Li, X., Zeng, Z., Wang, J., Wu, Y., Chen, W., Zheng, L., et al. (2020). MicroRNA-9 and Breast Cancer. *Biomed. Pharmacother.* 122, 109687. doi:10.1016/j.biopha.2019.109687

Liu, D., Zhang, Y., Li, Y., and Fan, K. (2017). Neurofibromatosis Type-1 Is a Prognostic Indicator in Human Gastric Carcinoma. *Oncotarget* 8 (47), 82910–82919. doi:10.18632/oncotarget.19876

Mosakhani, N., Sarhadi, V. K., Borze, I., Karjalainen-Lindsberg, M.-L., Sundström, J., Ristamäki, R., et al. (2012). MicroRNA Profiling Differentiates Colorectal Cancer According to KRAS Status. *Genes Chromosom. Cancer* 51 (1), 1–9. doi:10.1002/gcc.20925

Pazzaglia, L., Novello, C., Conti, A., Pollino, S., Picci, P., and Benassi, M. S. (2017). miR-152 Down-Regulation Is Associated with MET Up-Regulation in Leiomyosarcoma and Undifferentiated Pleomorphic Sarcoma. *Cell Oncol.* 40 (1), 77–88. doi:10.1007/s13402-016-0306-4

Ray-Coquard, I., Serre, D., Reichardt, P., Martín-Broto, J., and Bauer, S. (2018). Options for Treating Different Soft Tissue Sarcoma Subtypes. *Future Oncol.* 14 (10s), 25–49. doi:10.2217/fon-2018-0076

Roland, C. L., May, C. D., Watson, K. L., Al Sannaa, G. A., Dineen, S. P., Feig, R., et al. (2016). Analysis of Clinical and Molecular Factors Impacting Oncologic Outcomes in Undifferentiated Pleomorphic Sarcoma. *Ann. Surg. Oncol.* 23 (7), 2220–2228. doi:10.1245/s10434-016-5115-5

Saliminejad, K., Khorram Khorshid, H. R., Soleymani Fard, S., and Ghaffari, S. H. (2019). An Overview of microRNAs: Biology, Functions, Therapeutics, and Analysis Methods. *J. Cel Physiol* 234 (5), 5451–5465. doi:10.1002/jcp.27486

Serrano, C., Romagosa, C., Hernández-Losa, J., Simonetti, S., Valverde, C., Moliné, T., et al. (2016). RAS/MAPK Pathway Hyperactivation Determines Poor Prognosis in Undifferentiated Pleomorphic Sarcomas. *Cancer* 122 (1), 99–107. doi:10.1002/cncr.29733

Shah, S., Brock, E. J., Ji, K., and Mattingly, R. R. (2019). Ras and Rap1: A Tale of Two GTPases. *Semin. Cancer Biol.* 54, 29–39. doi:10.1016/j.semcancer.2018.03.005

Smolle, M. A., Leithner, A., Posch, F., Szkandera, J., Liegl-Atzwanger, B., and Pichler, M. (2017). MicroRNAs in Different Histologies of Soft Tissue Sarcoma: A Comprehensive Review. *Int. J. Mol. Sci.* 18 (9), 1960. doi:10.3390/ijms18091960

Stark, M. S., Bonazzi, V. F., Boyle, G. M., Palmer, J. M., Symmons, J., Lanagan, C. M., et al. (2015). miR-514a Regulates the Tumour Suppressor NF1 and Modulates BRAFi Sensitivity in Melanoma. *Oncotarget* 6 (19), 17753–17763. doi:10.18632/oncotarget.3924

Su, J., Ruan, S., Dai, S., Mi, J., Chen, W., and Jiang, S. (2019). NF1 Regulates Apoptosis in Ovarian Cancer Cells by Targeting MCL1 via miR-142-5p. *Pharmacogenomics* 20 (3), 155–165. doi:10.2217/pgs-2018-0161

Sun, Z., Zhang, B., and Cui, T. (2018). Long Non-coding RNA XIST Exerts Oncogenic Functions in Pancreatic Cancer via miR-34a-5p. *Oncol. Rep.* 39 (4), 1591–1600. doi:10.3892/or.2018.6245

Tang, S., Li, S., Liu, T., He, Y., Hu, H., Zhu, Y., et al. (2021). MicroRNAs: Emerging Oncogenic and Tumor-Suppressive Regulators, Biomarkers and Therapeutic Targets in Lung Cancer. *Cancer Lett.* 502, 71–83. doi:10.1016/j.canlet.2020.12.040

Walther, C., Mayrhofer, M., Nilsson, J., Hofvander, J., Jonson, T., Mandahl, N., et al. (2016). Genetic Heterogeneity in Rhabdomyosarcoma Revealed by SNP Array Analysis. *Genes Chromosomes Cancer* 55 (1), 3–15. doi:10.1002/gcc.22285

Wan, D., He, S., Xie, B., Xu, G., Gu, W., Shen, C., et al. (2013). Aberrant Expression of miR-199a-3p and its Clinical Significance in Colorectal Cancers. *Med. Oncol.* 30 (1), 378. doi:10.1007/s12032-012-0378-6

Wang, J., Lv, N., Lu, X., Yuan, R., Chen, Z., and Yu, J. (2021). Diagnostic and Therapeutic Role of microRNAs in Oral Cancer (Review). *Oncol. Rep.* 45 (1), 58–64. doi:10.3892/or.2020.7854

Wang, Y., Guan, E., Li, D., and Sun, L. (2021). miRNA-34a-5p Regulates Progression of Neuroblastoma via Modulating the Wnt/β-Catenin Signaling Pathway by Targeting SOX4. *Medicine (Baltimore)* 100 (20), e25827. doi:10.1097/md.0000000000025827

Wang, Z., Ma, X., Cai, Q., Wang, X., Yu, B., Cai, Q., et al. (2014). MiR-199a-3p Promotes Gastric Cancer Progression by Targeting ZHX1. *FEBS Lett.* 588 (23), 4504–4512. doi:10.1016/j.febslet.2014.09.047

Winchester, D., Lehman, J., Tello, T., Chimato, N., Hocker, T., Kim, S., et al. (2018). Undifferentiated Pleomorphic Sarcoma: Factors Predictive of Adverse Outcomes. *J. Am. Acad. Dermatol.* 79 (5), 853–859. doi:10.1016/j.jaad.2018.05.022

Xu, L., Li, M., Wang, M., Yan, D., Feng, G., and An, G. (2014). The Expression of microRNA-375 in Plasma and Tissue Is Matched in Human Colorectal Cancer. *BMC cancer* 14, 714. doi:10.1186/1471-2407-14-714

Zhang, L., Cui, M., Song, L., Zhang, M., and Zhang, J. (2019). Function, Significance, and Regulation of Rap1b in Malignancy. *Crit. Rev. Eukaryot. Gene Expr.* 29 (2), 151–160. doi:10.1615/CritRevEukaryotGeneExpr.2019025997

Zhang, P., Liu, J., Li, X., Gao, M., Feng, F., Wang, W., et al. (2021). NF1 and PTEN Gene Polymorphisms and the Susceptibility to Soft Tissue Sarcomas in a Chinese Population: A Case-Control Study. *Exp. Mol. Pathol.* 118, 104603. doi:10.1016/j.yexmp.2021.104603

Zhou, C., Tan, W., Lv, H., Gao, F., and Sun, J. (2016). Hypoxia-inducible microRNA-488 Regulates Apoptosis by Targeting Bim in Osteosarcoma. *Cel Oncol.* 39 (5), 463–471. doi:10.1007/s13402-016-0288-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Huang, Ma and Niu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 02 May 2022
doi: 10.3389/fgene.2022.864655

Pan-Cancer Analyses Reveal Oncogenic and Immunological Role of PLOD2

*Qiqi Xu, Na Kong, Yiguo Zhao, Quan Wu, Xin Wang, Xiaodong Xun and Pengji Gao**

General Surgery of Beijing Jishuitan Hospital, The Fourth Clinical Medical College of Peking University, Beijing, China

Some previous studies have shown that *PLOD2* has some value in tumorigenesis. However, the broad significance of *PLOD2* has not been discussed in depth. This study was aimed at elaborated and summarized the value of PLOD2 in various tumors. First, we integrated GTEx, The Cancer Genome Atlas and Cancer Cell Line Encyclopedia databases to analyze the expression of *PLOD2*, and found that it was expressed differently in normal tissues and significantly highly expressed in most tumors compared with normal tissues. Second, our analysis revealed that *PLOD2* expression was negatively correlated with the prognosis of several tumors. For gastric cancer, the median overall survival time was significantly higher in the *PLOD2* low expression group [HR 0.616 (95%CI 0.442–0.858), $p = 0.004$]. Third, for tumor immunity, *PLOD2* was significantly associated with tumor infiltration, including immune infiltrating cells; immune checkpoint expression; immune microenvironment scores (immune score, stromal score and estimate scores); immunotherapy-related scores (tumor mutational burden, microsatellite instability, tumor neoantigen burden); expression of DNA repair genes Mismatch Repairs and methyltransferase; and enrichment analyses identified *PLOD2*-associated terms and pathways. Lastly, twenty pairs of gastric cancer and adjacent immunohistochemistry showed that *PLOD2* was significantly overexpressed in gastric cancer ($p < 0.001$). Collectively, *PLOD2* played a significant role in tumorigenesis and maybe serve as a potential biomarker for diagnosis and prognosis in cancers.

Keywords: PLOD2, prognostic biomarker, pan-cancer analysis, immune infiltration, cancer immunity

OPEN ACCESS

Edited by:
Apeng Chen,
Chinese Academy of Agricultural Sciences, China

Reviewed by:
Wallax Ferreira,
Evandro Chagas Institute, Brazil
Xiaoying Hou,
Jianghan University, China

****Correspondence:***
Pengji Gao
1011210370@bjmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *28 January 2022*
Accepted: *07 April 2022*
Published: *02 May 2022*

Citation:
Xu Q, Kong N, Zhao Y, Wu Q, Wang X, Xun X and Gao P (2022) Pan-Cancer Analyses Reveal Oncogenic and Immunological Role of PLOD2. Front. Genet. 13:864655. doi: 10.3389/fgene.2022.864655

INTRODUCTION

The incidence of malignant neoplasms has increased at an alarming rate over the past few decades (Bray et al., 2021; Sung et al., 2021). Pan-cancer analysis aims to examine the similarities and differences between genomic and cellular changes found in different tumor types (Chang et al., 2013; Gentles et al., 2015). Pan-cancer analysis projects, such as the Cancer Cell Line Encyclopedia (CCLE) and The Cancer Genome Atlas (TCGA), were created based on the evaluation of different human cancer cell lines and tissues at the epigenomic, genomic, proteomic, and transcriptomic levels. TCGA provides medical researchers with irreplaceable genomic, epigenomic, transcriptomic, and clinical

Abbreviations: CCLE, cancer cell line encyclopedia; TCGA, the cancer genome atlas; PLOD2, procollagen-lysine,2-oxoglutarate 5-dioxygenase 2; ECM, extracellular matrix; TMB, tumor mutational burden; MSI, microsatellite instability; TNB, tumor neoantigen burden; GSEA, gene set enrichment analysis; GTEx, genotype-tissue expression; OS, overall survival; DFI, disease free interval; DSS, disease specific survival; PFI, progression free interval; KM, Kaplan-Meier.

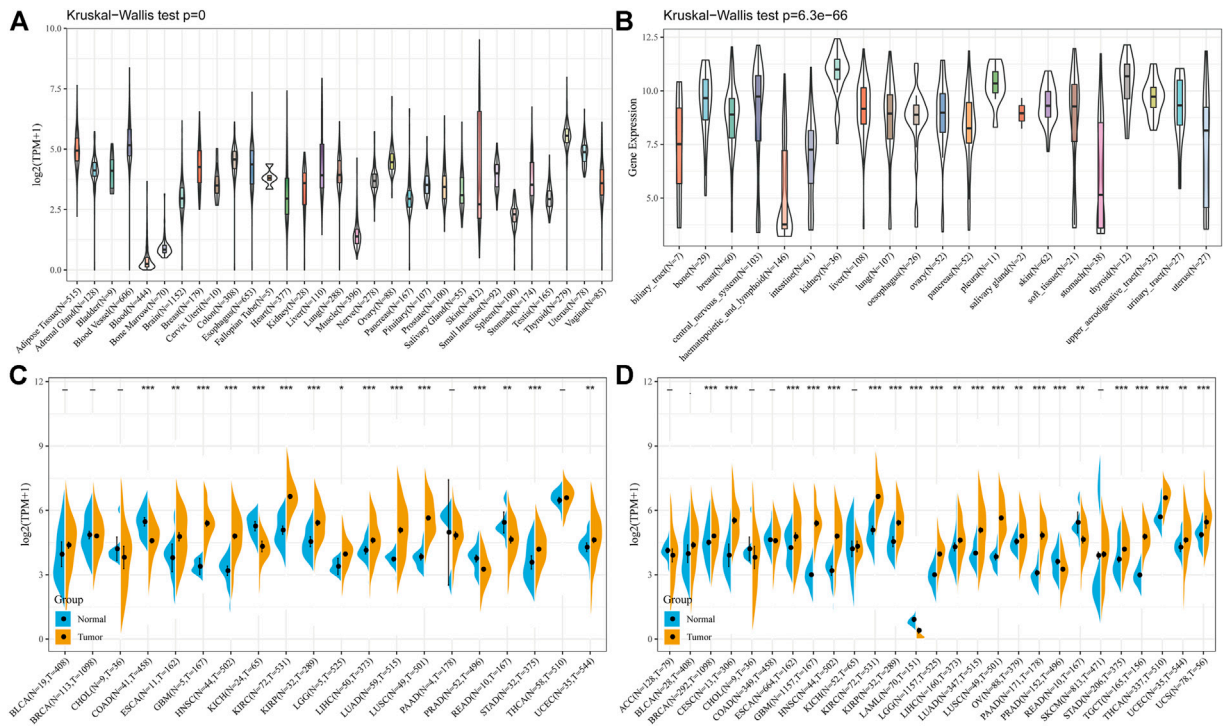


FIGURE 1 | The *PLOD2* expression level in human pan-cancer analyses. **(A)** Expression of *PLOD2* in normal tissues in GTEx. **(B)** Expression of *PLOD2* in CCLE. **(C)** The level of *PLOD2* in TCGA. **(D)** The expression level in TCGA combined with GTEx. The blue and yellow bar graphs indicate normal and tumor tissues, respectively. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. The significance of the two groups of samples passed the Wilcox test.

data (Hutter and Zenklusen, 2018). What's more, it has boosted the study of tumor immunology and immunotherapy (Thorsson et al., 2018). Pan-cancer analysis has made an important contribution to the development of life science and medicine. For example, on 4 February 2020, Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and TCGA published six articles in Nature (Author Anonyms, 2020), proposing the most comprehensive cancer genome analysis so far. Different from the previous focus on protein coding regions, this time is to analyze the whole genome of cancer. This program covers six aspects: pan-cancer analysis of whole genomes (ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020); analyses of non-coding somatic drivers in 2,658 cancer whole genomes (Rheinbay et al., 2020); the repertoire of mutational signatures in human cancer (Alexandrov et al., 2020); patterns of somatic structural variation in human cancer genomes (Li Y et al., 2020); the evolutionary history of 2,658 cancers (Gerstung et al., 2020); genomic basis for RNA alterations in cancer (Calabrese et al., 2020).

PLOD2 (Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2) was a member of PLOD family (*PLOD1*, *PLOD2*, *PLOD3*), which encodes a special protein (also known as LH2, TLH2 and BRKS2) mediating the formation of stabilized collagen cross-links (Genecards, 2021). Collagen crosslinking played a key role in extracellular matrix (Du et al., 2017). Various studies have shown the extracellular matrix (ECM) to be closely to tumor cell growth and metastasis (Gilkes et al., 2014; Tadeo et al., 2016). Upregulation of *PLOD2* has been observed in several malignancies such as bladder cancer (Miyamoto et al., 2016), lung cancer (Kocher et al., 2021), gastric cancer (Song et al., 2021), head and neck squamous cell cancer (Xin et al., 2021), breast cancer (Gilkes et al., 2013), etc. Kiyozumi et al. showed that PLOD2 was significantly associated with peritoneal dissemination in gastric cancer (Kiyozumi et al., 2018). In the metastatic group, PLOD2 was significantly highly expressed, both at the mRNA and protein level. Silencing PLOD2 significantly reduced cell invasiveness and migration *in vitro*. Further experiments showed that this was mainly regulated by HIF-1a in hypoxia condition. Another study on PLOD2 and 5-FU resistance in gastric cancer showed that PLOD2 could enhance 5-FU resistance by regulating BCRP and inhibit cell apoptosis by affecting the expression of Bax and Bcl2 (Wang et al., 2020). Downregulation of PLOD2 facilitated the sensitivity of gastric cancer to 5-FU *in vivo*. Generally, *PLOD2* plays an important role in both tumor growth and closely related to the prognosis of patients. Although a number of studies have been carried out on *PLOD2*, no single study exists which could overall evaluate its effects on considerable types of cancers. To understand the functions (especially cancer immunity) of *PLOD2* in different tumors, a comprehensive pan-cancer analysis was necessary.

To that end, we will elucidate the expression of *PLOD2* in 33 different malignant tumors in the following aspects and focus on gastric cancer. All in all, the results of our study provide

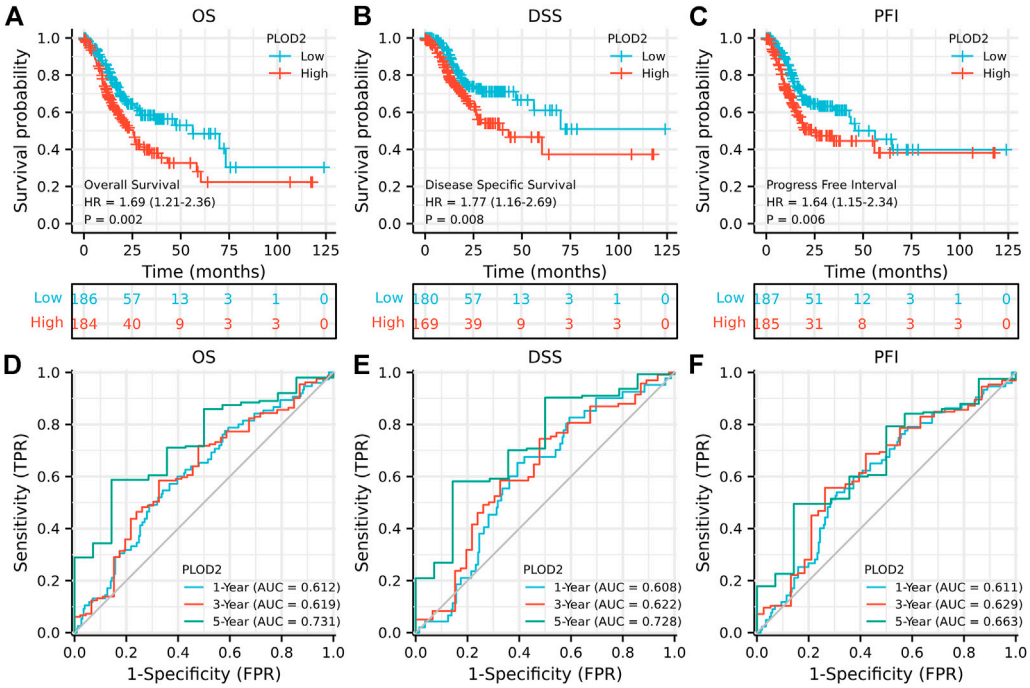


FIGURE 2 | Prognostic analysis of *PLOD2* gene signature in STAD in the TCGA set. **(A)** overall survival **(B)** disease-specific survival **(C)** progression-free interval **(D)** ROC of overall survival **(E)** ROC of disease-specific survival **(F)** ROC of progression-free interval.

information regarding the role of *PLOD2* in tumors, reveal the relationship between *PLOD2* and tumor-immune interactions, and clarify the potential underlying mechanisms.

MATERIALS AND METHODS

Patient Datasets and Processes

This study has been approved by the Ethics Committee of Beijing Jishuitan Hospital (202004-58). First, we analyzed the *PLOD2* gene expression levels in each normal tissue using the GTEx (Genotype-Tissue Expression) database (https://xena.ucsc.edu/). Second, the data of each tumor cell line were downloaded from the CCLE database (https://sites.broadinstitute.org/ccle), and the expression levels of 21 tissues were analyzed according to the tissue source. Third, we obtained gene expression differences between cancer and para-cancer tissues in individual tumor samples from the TCGA database (https://portal.gdc.cancer.gov/). Fourth, considering the small number of normal samples in TCGA database, we integrated data from GTEx and TCGA database to analyze the differences expression in multiple tumors.

Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2 Expression and its Survival-Associated Cancers

The differences of *PLOD2* gene expression were compared according to TNM stages of different tumors (data from TCGA database). Next, univariable and multivariable Cox regression analysis was used to compare the relationship between different *PLOD2* expression (divided into high and low expression groups with the median cutoff value) and prognosis. Prognosis includes OS (overall survival; period from the start of treatment to death from any cause), DSS (disease specific survival; cancer survival in the absence of other causes of death), and PFI (progression free interval; period from the start of treatment to disease progression or death from any cause). Subsequently, our findings were verified in the GSE84433 cohort.

Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2 and Tumor Immunity

We used CIBERSORT(Newman et al., 2019) to explore the association of *PLOD2* gene expression with the level of immune infiltration in different types of cancer. CIBERSORT (https://CIBERSORT.stanford.edu/) is a tool for deconvolution of expression matrices of immune cell subtypes based on the principle of linear support vector regression, and can be used to estimate immune cell infiltration with data from RNA-Seq. Then we used xCell algorithm (Aran et al., 2017) and MCP-Counter algorithm (Becht et al., 2016) to verify the result of CIBERSORT.

In the tumor microenvironment, immune cells and stromal cells are two major non-tumor components (Bejarano et al., 2021; Dey et al., 2021). The immune score and stromal score calculated based on ESTIMATE algorithm (Yoshihara et al., 2013). The ESTIMATE algorithm produces three scores on the basis of single

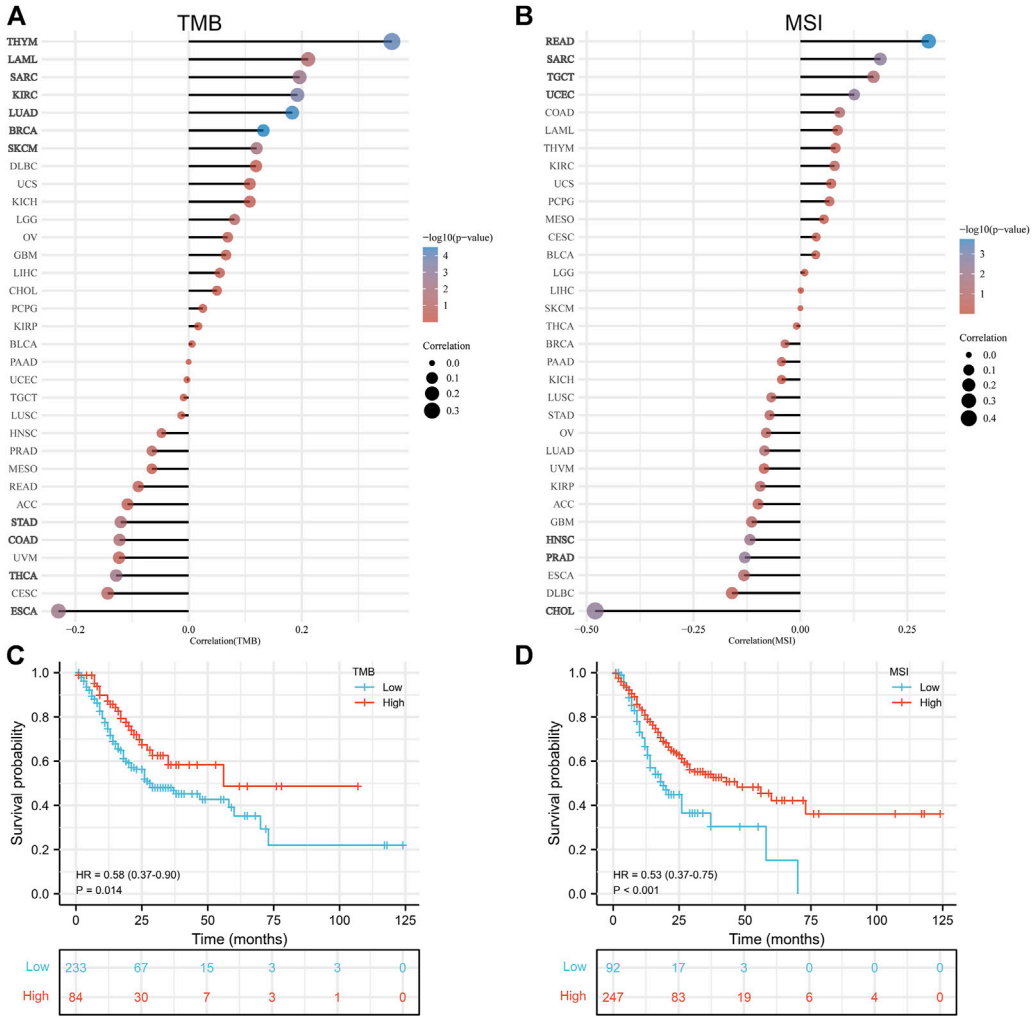


FIGURE 3 | Correlation analysis of *PLOD2* expression with tumor mutational burden and microsatellite instability. **(A)** with TMB **(B)** with MSI **(C)** relationship between TMB and overall survival **(D)** relationship between MSI and overall survival.

sample Gene Set Enrichment Analysis (ssGSEA): stromal score, immune score, and estimate score. In this study, we estimated these 3 scores and then calculated the relationship between these scores and *PLOD2* expression.

Furthermore, we examined the correlation between *PLOD2* expression and immune checkpoint-related genes (*BTLA*, *CD200*, *TNFRSF14*, *NRP1*, *LAIR1*, *TNFSF4*, *CD244*, *LAG3*, *ICOS*, *CD40LG*, *CTLA4*, *CD48*, *CD28*, *CD200R1*, *HAVCR2*, *ADORA2A*, *CD276*, *KIR3DL1*, *CD80*, *PDCD1*, *LGALS9*, *CD160*, *TNFSF14*, *IDO2*, *ICOSLG*, *TMIGD2*, *VTCN1*, *IDO1*, *PDCD1LG2*, *HHLA2*, *TNFSF18*, *BTNL2*, *CD70*, *TNFSF9*, *TNFRSF8*, *CD27*, *TNFRSF25*, *VSIR*, *TNFRSF4*, *CD40*, *TNFRSF18*, *TNFSF15*, *TIGIT*, *CD274*, *CD86*, *CD44*, *TNFRSF9*, shown in **Supplementary Figure S4**) using R software.

Next, we analyzed the relationship between *PLOD2* expression and TMB, MSI, and TNB. TMB is usually defined as the number of somatic nonsynonymous mutations or all mutations occurring per MB in the gene region detected by whole-exome sequencing or targeted sequencing in one tumor sample (Passaro et al., 2020). Somatic mutations calculated by TMB include point mutations and insertion/deletion mutations (Valero et al., 2021). TNB is an indicator of the total number of neoantigens in tumor cells, usually expressed as the number of tumor neoantigens per million bases of tumor genomic region (Wang et al., 2021). The combination of TMB and TNB can better predict the efficacy of immunotherapy. MSI, the insertion or loss of base pairs in microsatellite regions due to replication errors, was first identified in colorectal cancer and is thought to be a feature of hereditary non-polyposis colorectal cancer (Lynch syndrome) (Vilar and Gruber, 2010) and has since been found in a variety of sporadic tumors.

We downloaded the *PLOD2* genetic mutation data, transcriptome data, and clinical data from the TCGA database. To identify the somatic mutations of the patients with *PLOD2* in the TCGA database, mutation data were downloaded and visualized using the "maftools" package in R software.

TABLE 1 | Univariate and multivariate Cox regression analysis of overall survival.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p* value	Hazard ratio (95% CI)	*p* value
Age	367				
≤65	163	References			
>65	204	1.620 (1.154–2.276)	**0.005**	1.731 (1.194–2.508)	**0.004**
Gender	370				
Male	237	References			
Female	133	0.789 (0.554–1.123)	0.188		
T stage	362				
T1&T2	96	References			
T3&T4	266	1.719 (1.131–2.612)	**0.011**	1.252 (0.741–2.118)	0.401
N stage	352				
N0	107	References			
N1	97	1.629 (1.001–2.649)	**0.049**	1.300 (0.712–2.374)	0.394
N2	74	1.655 (0.979–2.797)	0.060	1.359 (0.647–2.853)	0.418
N3	74	2.709 (1.669–4.396)	**<0.001**	2.032 (0.952–4.339)	0.067
M stage	352				
M0	327	References			
M1	25	2.254 (1.295–3.924)	**0.004**	2.038 (1.094–3.799)	**0.025**
Pathologic stage	347				
Stage I and II	160	References			
Stage III and IV	187	1.947 (1.358–2.793)	**<0.001**	1.144 (0.600–2.181)	0.683
PLOD2	370				
Low	186	References			
High	184	1.693 (1.215–2.360)	**0.002**	1.484 (1.034–2.131)	**0.032**

The bold values means P<0.05.

Horizontal histogram showed the genes have the higher mutation frequency in patients with *PLOD2*.

Finally, we evaluated the relationship between the expression of *PLOD2* and 5 DNA repair genes (MMRs: *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*) and 4 methyltransferases (*DNMT1*, *DNMT2*, *DNMT3A*, *DNMT3B*) genes.

Gene Set Enrichment Analysis

Using JAVA (http://software.broadinstitute.org/gsea/index.jsp), we conducted GSEA to assess for possible underlying mechanisms based on the 'Molecular Signatures Database' of c5.all.v7.1.symbols and c2.cp.kegg.v7.1.symbols. When the number of random sample arrangements was 100 and the significance threshold was $p < 0.05$, R software and Bioconductor (http://bioconductor.org/) were applied to visualize our results.

Immunohistochemical Staining

Tissue sections were prepared from the paraffin-embedded tissue samples. Then *PLOD2* immunostaining was performed according to the instructions (proteintech 21214-1-AP, China). Immunohistochemical scoring was performed by semi-quantitative analysis (20 pairs of gastric cancer and adjacent tissues). Two pathologists analyzed and scored the immunohistochemistry of gastric tissue. Each slice was randomly observed for 5 high-power visual fields, and scored according to the percentage of positive cells (0–5%, 6–25%, 26–50%, 51–75%, 76–100% were recorded as 0, 1, 2, 3, and 4 points respectively) and the intensity of staining (0, 1, 2, and 3 points respectively for non-staining, light, medium, and deep). The total score was the sum of staining intensity and percentage of positive cells. Next, we validated the expression of PLOD2 in STAD and normal tissues in the Human Protein Atlas (HPA) database (Uhlén et al., 2015).

Statistical Methods

The Wilcoxon log-rank test was used to determine the presence or absence of a markedly increased sum of gene expression z-scores for tumor tissues, as compared to adjacent normal tissues. The difference in *PLOD2* expression between different tumor stages was compared using the Kruskal–Wallis H test. Survival was analyzed using the K-M curves, log-rank test, and Cox proportional hazards regression model. Spearman's test was used for correlation analysis. R language (version 3.6.0; R Foundation) was used for all analyses. A two-sided $p < 0.05$ indicated a statistically significant difference.

RESULTS

Pan-Cancer Expression Landscape of Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2

Firstly, we analyzed the expression levels of *PLOD2* in 7858 normal samples using the GTEx dataset. As shown in **Figure 1A**, the differences in *PLOD2* gene expression were significant ($p < 0.001$) in 31 tissues. Subsequently, we analyzed

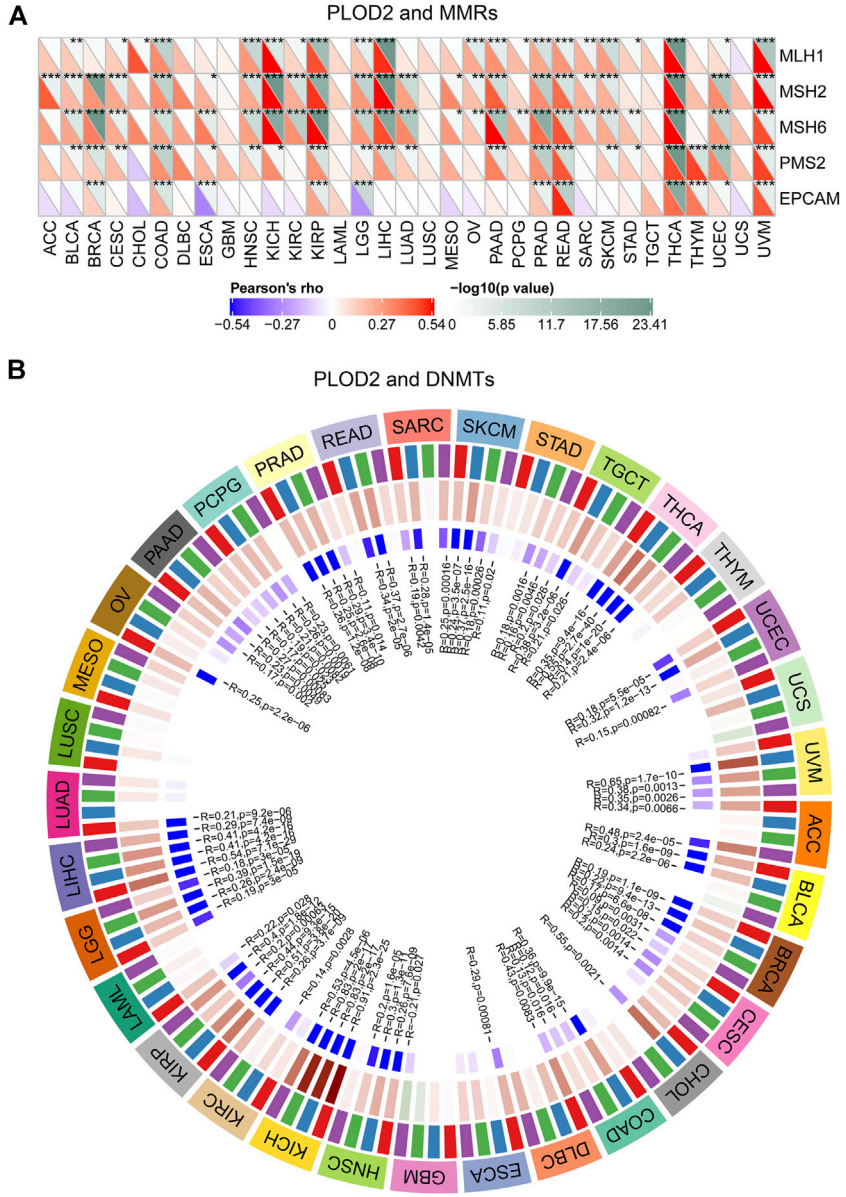


FIGURE 4 | Relationship between *PLOD2* expression and MMRS and methyltransferase in pan-cancer. **(A)** Relationship between *PLOD2* expression and mutation of 5 MMRs genes. **(B)** Relationship between 4 methyltransferases and *PLOD2* expression. Red, blue, green, and purple colors are indicated *DNMT1*, *DNMT2, DNMT3A*, and *DNMT3B*, respectively.

data downloaded from the CCLE database for each tumor cell line. There were significant differences in expression among the 21 tumor cell lines (**Figure 1B**), with the highest in renal tumors. Further, we obtained the differences in *PLOD2* from the TCGA database between cancer and para-cancer in individual tumor samples; and as shown in **Figure 1C**, *PLOD2* was highly expressed in 11 (ESCA, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, STAD, UCEC) of 20 different tumors, lowly expressed in 4 tumors (COAD, KICH, PRAD, READ). Finally, considering the small number of normal samples in TCGA database, we integrated data from GTEx and TCGA database to analyze the differences expression in multiple tumors (**Supplementary Table S1**). As shown in **Figure 1D**, *PLOD2* was highly expressed in 18 tumors (BRCA, CESC, ESCA, GBM, HNSC, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, OV, PAAD, STAD, TGCT, THCA, UCEC, UCS) and lowly expressed in 3 tumors (LAML, PRAD, READ).

To assess the levels of gene expression for all tumor stages, we compared *PLOD2* expression in patients with different stages. As demonstrated in **Supplementary Figure S1**, *PLOD2* expression was upregulated at the advanced stages in BLCA, COAD, HNSC, KIRC, KIRP, LIHC, LUAD, and READ.

TABLE 2 | Correlation analysis of *PLOD2* expression with immune scores, stromal scores, and estimate scores.

Cancer type	Immune scores		Stromal scores		Estimate scores	
	R	P	R	P	R	P
ACC	0.056	0.624	0.103	0.366	0.080	0.484
BLCA	**0.221**	<0.001	**0.266**	<0.001	**0.260**	<0.001
BRCA	0.038	0.203	**0.192**	<0.001	**0.108**	<0.001
CESC	**−0.168**	0.003	−0.056	0.332	**−0.126**	0.028
CHOL	0.042	0.806	0.255	0.133	0.089	0.605
COAD	**0.117**	0.012	**0.327**	<0.001	**0.243**	<0.001
DLBC	−0.145	0.326	**0.606**	<0.001	**0.328**	0.024
ESCA	0.044	0.579	**0.281**	<0.001	**0.180**	0.022
GBM	**0.192**	0.013	**0.310**	<0.001	**0.263**	<0.001
HNSC	0.014	0.751	**0.227**	<0.001	**0.132**	0.003
KICH	−0.103	0.415	0.095	0.449	−0.015	0.908
KIRC	**0.111**	0.010	**0.283**	<0.001	**0.204**	<0.001
KIRP	−0.089	0.131	0.058	0.329	−0.018	0.763
LAML	0.035	0.672	**0.309**	<0.001	**0.179**	0.028
LGG	**0.476**	<0.001	**0.512**	<0.001	**0.505**	<0.001
LIHC	0.023	0.663	−0.055	0.293	−0.017	0.739
LUAD	0.030	0.500	**0.142**	0.001	**0.095**	0.031
LUSC	−0.033	0.462	**0.100**	0.025	0.030	0.498
MESO	0.016	0.883	0.209	0.054	0.095	0.382
OV	**0.173**	<0.001	**0.307**	<0.001	**0.257**	<0.001
PAAD	**0.228**	0.002	**0.501**	<0.001	**0.369**	<0.001
PCPG	**0.159**	0.032	**0.331**	<0.001	**0.252**	<0.001
PRAD	**0.089**	0.049	**0.227**	<0.001	**0.163**	<0.001
READ	0.068	0.384	**0.371**	<0.001	**0.245**	0.001
SARC	**−0.198**	0.001	**−0.181**	0.004	**−0.202**	0.001
SKCM	0.006	0.904	**0.159**	<0.001	0.076	0.097
STAD	0.053	0.307	**0.416**	<0.001	**0.252**	<0.001
TGCT	**−0.326**	<0.001	**0.189**	0.018	−0.141	0.079
THCA	**−0.262**	<0.001	**−0.152**	<0.001	**−0.233**	<0.001
THYM	**−0.316**	<0.001	**0.192**	0.037	−0.053	0.565
UCEC	**0.234**	<0.001	**−0.145**	<0.001	**−0.211**	<0.001
UCS	−0.154	0.255	−0.006	0.963	−0.107	0.431
UVM	0.140	0.216	−0.166	0.141	−0.151	0.180

The correlation coefficient with p *value less than 0.05 is expressed in bold.*

Screening of Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2 Survival Associated Cancers

In the OS analysis, Cox regression identified that high *PLOD2* expression was a risk factor for CESC ($p < 0.001$), CHOL ($p = 0.032$), HNSC ($p = 0.005$), KICH ($p < 0.001$), KIRC ($p = 0.001$), KIRP ($p = 0.034$), LGG ($p < 0.001$), LIHC ($p < 0.001$), LUAD ($p = 0.001$), MESO ($p = 0.020$), PAAD ($p = 0.006$), SARC ($p = 0.005$), and STAD ($p = 0.001$); however, it appeared to be a protective factor in LAML ($p = 0.024$) and SKCM ($p = 0.049$), as shown in **Supplementary Figure S2A**. The Cox regression analysis of DSS indicated that high *PLOD2* expression is a risk factor in CESC ($p < 0.001$), CHOL ($p = 0.038$), ESCA ($p = 0.045$), HNSC ($p = 0.020$), KICH ($p < 0.001$), KIRC ($p < 0.001$), LGG ($p < 0.001$), LIHC ($p < 0.001$), LUAD ($p = 0.001$), MESO ($p = 0.007$), PAAD ($p = 0.004$), SARC ($p = 0.005$) and STAD ($p = 0.006$) as illustrated in **Supplementary Figure S2B**. The Cox regression analysis of PFI revealed that higher *PLOD2* expression is a risk factor in CESC($p < 0.001$), ESCA ($p = 0.049$), KICH ($p < 0.001$), KIRC ($p < 0.001$), LGG ($p < 0.001$), LIHC ($p = 0.001$), LUAD ($p = 0.004$), MESO ($p = 0.028$), PAAD ($p = 0.015$), SARC ($p < 0.001$) and STAD ($p = 0.047$). In the DFI analysis, Cox regression identified that high *PLOD2* expression was a risk factor for CESC ($p = 0.010$), CHOL ($p = 0.038$), KIRC ($p = 0.023$), LIHC ($p = 0.019$), LUAD ($p = 0.022$), PAAD ($p < 0.001$) and SARC ($p = 0.007$).

We further analyzed the relationship between *PLOD2* gene expression and prognosis in gastric cancer patients in detail (**Figure 2**). **Figures 2A–C** showed the relationship between *PLOD2* gene expression and OS, DSS and PFI, respectively. In the OS analysis, as illustrated in **Figure 2A**, we split cases into high-risk and low-risk groups according to the median expression. The median OS time was significantly higher in

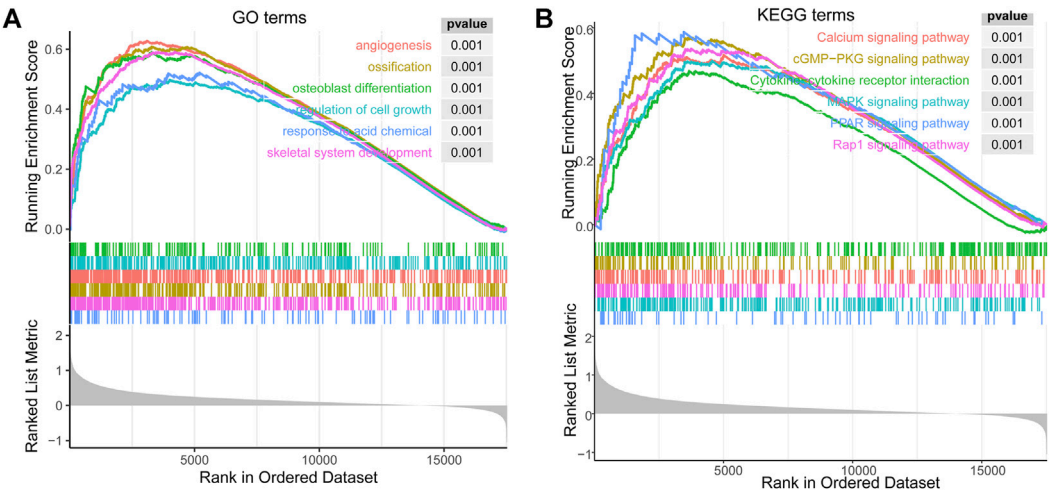


FIGURE 5 | The enrichment results of GSEA correlated with *PLOD2* expression in STAD. **(A)** top 6 significant enrichment GO terms. **(B)** top 6 KEGG terms.

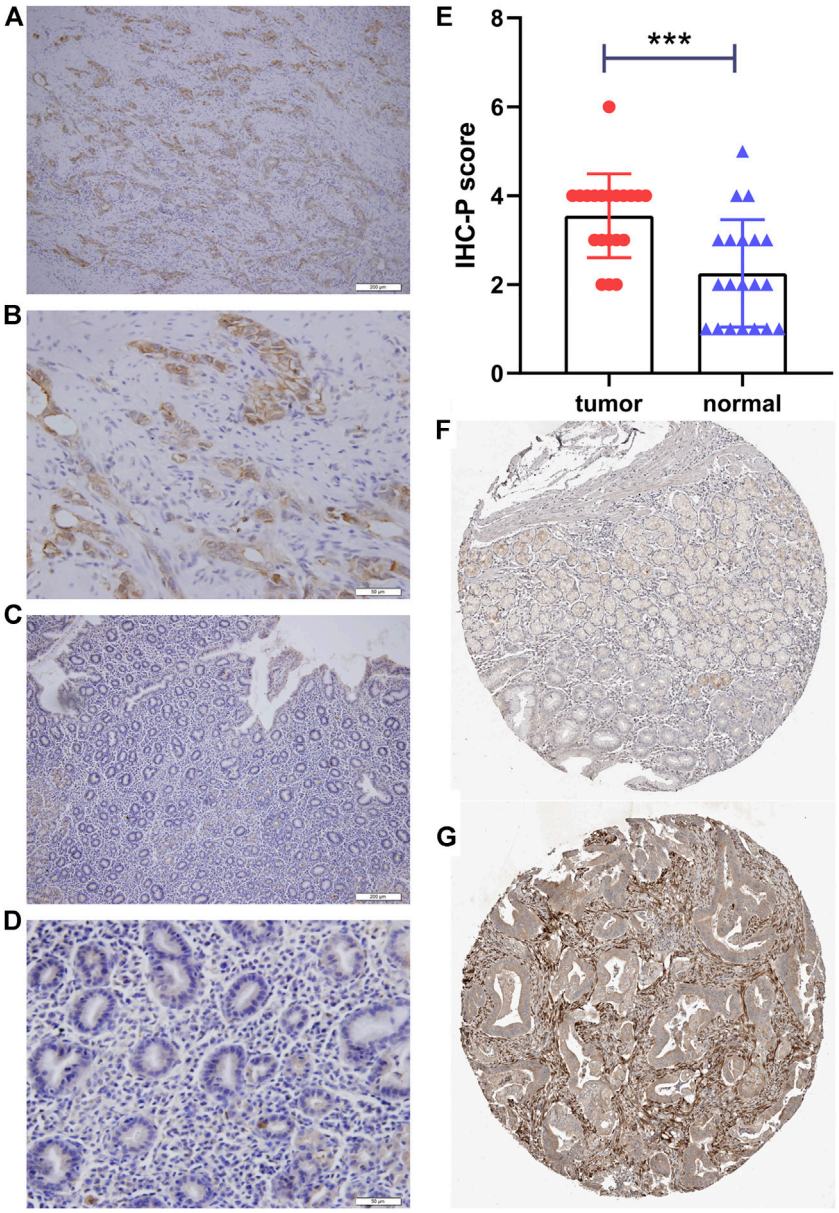
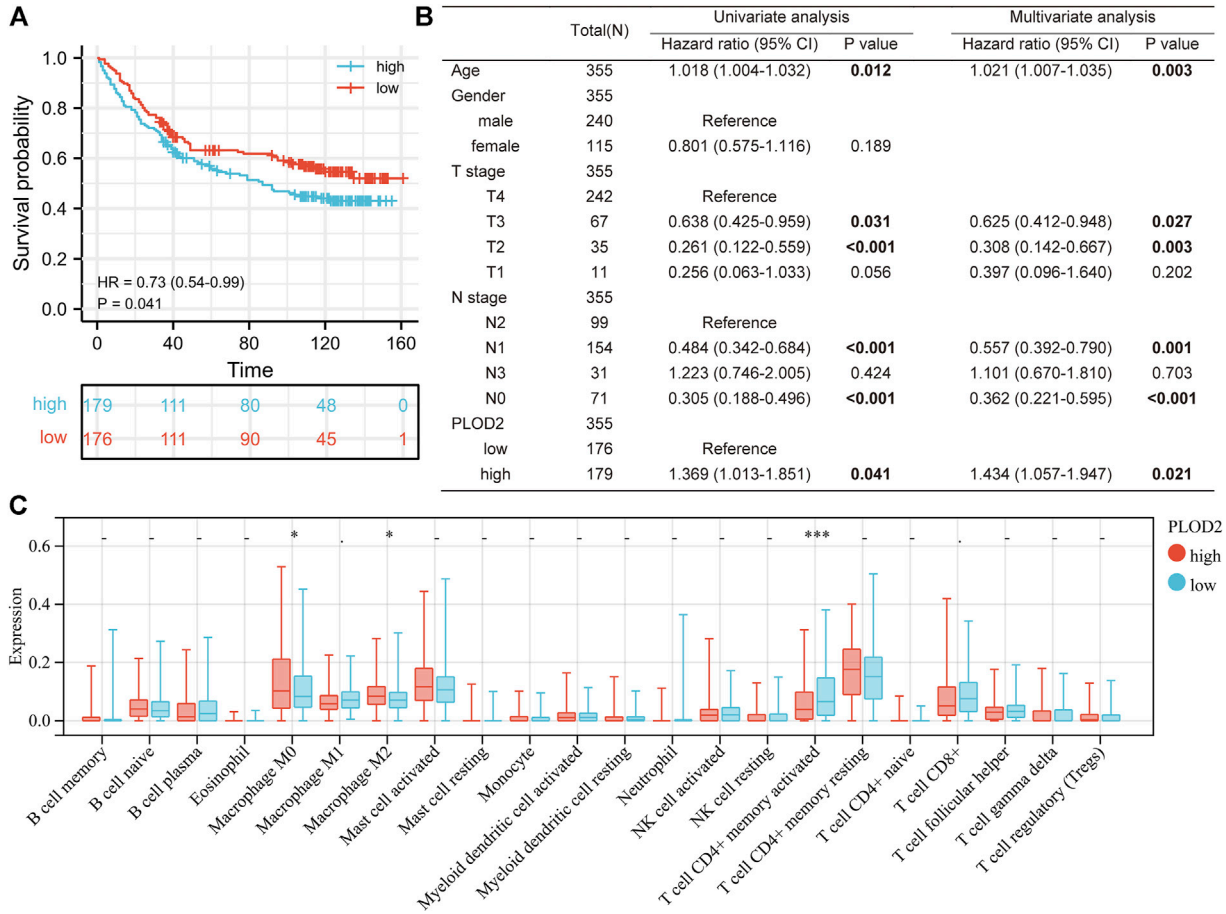


FIGURE 6 | Immunohistochemical staining of *PLOD2*. **(A)** gastric cancer, ×10. **(B)** gastric cancer, ×40. **(C)** adjacent tissue, ×10. **(D)** adjacent tissue, ×40. **(E)** Statistical analysis of cancer and adjacent tissue. **(F)** validation the expression of PLOD2 in normal tissues (Patient ID 1650, staining low) in the Human Protein Atlas (HPA) database **(G)** validation the expression of PLOD2 in STAD (Patient ID 2557, staining medium) in the HPA database ***$p < 0.001$.

the *PLOD2* low expression group than in the high expression group [HR 1.69 (95% CI 1.21–2.36), $p = 0.002$]. Time-dependent receiver operating characteristic (ROC) curves were plotted and the area under curve (AUC) values of the classifier to predict 1-, 3-, and 5-year OS were 0.612, 0.619, and 0.731, respectively (**Figure 2D**). In DSS analysis (**Figure 2B**) and PFI analysis (**Figure 2C**), hazard ratio was 1.77 (95% CI 1.16–2.69, $p = 0.008$) and 1.64 (95% CI 1.15–2.34, $p = 0.006$), respectively. AUC values of the classifier to predict 1-, 3-, and 5-year DSS were 0.608, 0.622, and 0.728, respectively (**Figure 2E**). AUC values of the classifier to predict 1-, 3-, and 5-year PFI were 0.611, 0.629, and 0.663, respectively (**Figure 2F**). As shown in **Figure 3A**, in the GSE84433 validation cohort (355 patients remained after deleting 2 patients who survived less than 1 month), the overall survival time of the PLOD2 low expression group was also significantly longer than that of the high expression group [HR 0.73, 95% CI (0.54–0.99), $p = 0.041$].

As shown in **Table 1**, univariate analysis showed that age ($p = 0.005$), T stage, N stage, M stage, pathologic stage and PLOD2 expression were significantly correlated with OS (all $p < 0.05$). However, in multivariate analysis, only age [HR: 1.731 (95% CI



	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Age	355	1.018 (1.004-1.032)	**0.012**	1.021 (1.007-1.035)	**0.003**
Gender	355				
male	240	Reference			
female	115	0.801 (0.575-1.116)	0.189		
T stage	355				
T4	242	Reference			
T3	67	0.638 (0.425-0.959)	**0.031**	0.625 (0.412-0.948)	**0.027**
T2	35	0.261 (0.122-0.559)	**<0.001**	0.308 (0.142-0.667)	**0.003**
T1	11	0.256 (0.063-1.033)	0.056	0.397 (0.096-1.640)	0.202
N stage	355				
N2	99	Reference			
N1	154	0.484 (0.342-0.684)	**<0.001**	0.557 (0.392-0.790)	**0.001**
N3	31	1.223 (0.746-2.005)	0.424	1.101 (0.670-1.810)	0.703
N0	71	0.305 (0.188-0.496)	**<0.001**	0.362 (0.221-0.595)	**<0.001**
PLOD2	355				
low	176	Reference			
high	179	1.369 (1.013-1.851)	**0.041**	1.434 (1.057-1.947)	**0.021**

FIGURE 7 | Validation PLOD2 expression in GSE84433 cohort. **(A)** survival curve of high and low PLOD2 expression. **(B)** univariate and multivariate analysis of overall survival in GSE84433 cohort. **(C)** immune infiltration difference between high and low PLOD2 expression in GSE84433 cohort, based on CIBERSORT deconvolution method.

1.194–2.508), p = 0.004], M stage [HR: 2.038 (95% CI 1.094–3.799), p = 0.025] and PLOD2 expression [HR: 1.484 (95% CI 1.034–2.131), p = 0.032] were significantly correlated with prognosis. In the GSE84433 validation cohort, both univariate [HR 1.369 95% CI (1.013–1.851), p = 0.041] and multivariate [HR 1.434 95% CI (1.057–1.947), p = 0.021] analysis, the overall survival of PLOD2 high expression group was significantly worse than that of low expression group (**Figure 3B**).

Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2 Level and Immune Infiltration

Tumor-infiltrating lymphocytes are independent predictors of lymph node status and survival in cancer precursors. Immune infiltrating cells were analyzed using CIBERSORT. Although the tumor types differed, the relationship between *PLOD2* expression and immune cells was similar. For example, *PLOD2* expression was significantly negatively correlated with memory B cell, activated NK cell, Plasma cell, CD8 T cell, follicular helper T cell, regulatory T cell in most tumors; and significantly positively correlated with Macrophages M0, Macrophages M1, Macrophages M2, activated Mast cell, resting NK cell, CD4 memory activated T cell, CD4 memory resting T cell. As shown in **Figure 4C**, PLOD2 was significantly correlated with 19 immune checkpoints in patients with STAD, of which 80% (14/19, CD200, CD276, CD28, CD44, CD80, CD86, HAVCR2, LAIR1, NRP1, PDCD1LG2, TNFRSF25, TNFRSF9, TNFSF14, TNFSF18, TNFSF4) was positively correlated.

Next, we analyzed the relationship between *PLOD2* expression and infiltrating immune cells in gastric cancer based on the xCELL algorithm. As shown in **Figure 4B**, the proportion of T cell CD4$^+$ Th1, Macrophage, Macrophage M1, Plasmacytoid dendritic cell, B cell, Monocyte, Neutrophil, and Endothelial cell were significantly higher in the *PLOD2* high expression group than low expression group. Contrarily, the proportion of B cell plasma, microenvironment score, T cell CD8$^+$ effector memory, T cell CD8$^+$ central memory, Class-switched memory B cell, B cell memory, Granulocyte-monocyte progenitor, Hematopoietic stem cell and stroma score were higher in the *PLOD2* low expression group. we also used MCP-Counter deconvolution

TABLE 3 | Relationship between PLOD2 expression and clinicopathology.

Characteristic	Total	PLOD2-high	PLOD2-low	*p*
n	20	12	8	
age, mean ± SD	62 ± 13	61 ± 11	64 ± 16	0.594
gender, *n* (%)				>0.999
female	7 (35%)	4 (33.3%)	3 (37.5%)	
male	13 (65%)	8 (66.7%)	5 (62.5%)	
stageT, *n* (%)				0.170
1 or 2	10 (50.0%)	4 (33.3%)	6 (75%)	
3 or 4	10 (50.0%)	8 (66.7%)	2 (25%)	
stageN, *n* (%)				0.289
0	3 (15%)	3 (25%)	0 (0%)	
1	1 (5%)	0 (0%)	1 (12.5%)	
2	8 (40%)	4 (33.3%)	4 (50%)	
3	8 (40%)	5 (41.7%)	3 (37.5%)	
stageM, *n* (%)				>0.999
0	18 (90%)	11 (91.7%)	7 (87.5%)	
1	2 (10%)	1 (8.3%)	1 (12.5%)	
clinical stage, *n* (%)				0.065
1 or 2	9 (45%)	3 (25%)	6 (75%)	
3 or 4	11 (55%)	9 (75%)	2 (25%)	
differentiation grade, *n* (%)				0.465
poorly	11 (55%)	5 (41.7%)	6 (75%)	
moderately	8 (40%)	6 (50%)	2 (25%)	
well	1 (5%)	1 (8.3%)	0 (0%)	
signet-ring, *n* (%)				0.603
non–signet-ring	15 (75%)	8 (66.7%)	7 (87.5%)	
signet-ring	5 (25%)	4 (33.3%)	1 (12.5%)	
Lauren type, *n* (%)				0.052
diffuse	11 (55%)	5 (41.7%)	6 (75%)	
intestinal	3 (15%)	1 (8.3%)	2 (25%)	
mixed	6 (30%)	6 (50%)	0 (0%)	
HER-2, *n* (%)				>0.999
0 or 1+	12 (60%)	7 (58.3%)	5 (62.5%)	
2+ or 3+	8 (40%)	5 (41.7%)	3 (37.5%)	

methods to verify our results (**Supplementary Figures S7A,B**). In the GSE84433 validation cohort, macrophage M0 ($p < 0.05$) and macrophage M2 ($p < 0.05$) increased significantly in the PLOD2 high expression group, while T cell CD4 + memory activated ($p < 0.001$) decreased significantly (**Figure 3C**).

Numerous studies indicated that the tumor immune microenvironment has an important role in tumor development. As shown in **Table 2**, in 9 kinds of tumors (BLCA, COAD, GBM, KIRC, LGG, OV, PAAD, PCPG, and PRAD), immune scores were positively correlated with the expression of *PLOD2* and negatively correlated with the expression of CESC, SARC, TGCT, THCA, THYM, and UCEC. For stromal scores, 22 kinds of tumors (BLCA, BRCA, COAD, DLBC, ESCA, GBM, HNSC, KIRC, LAML, LGG, LUAD, LUSC, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT and THYM) were positively correlated with the expression of *PLOD2*. Combining immune scores and stromal scores gives estimate scores. *PLOD2* gene expression was positively correlated with estimate scores in 17 tumors (BLCA, BRCA, COAD, DLBC, ESCA, GBM, HNSC, KIRC, LAML, LGG, LUAD, OV, PAAD, PCPG, PRAD, READ, and STAD, in **Supplementary Figure S3**).

Finally, we collected more than forty common immune checkpoint genes to analyze the relationship between *PLOD2* gene expression and immune checkpoint gene expression (**Supplementary Figure S4**). Among them, *TNFRSF14*, *NRP1*, *LAIR1*, *TNFSF4*, *CD276*, *CD80*, *PDCD1LG2*, *CD274*, *CD86*, and *CD44* were significantly positively correlated with *PLOD2* expression.

Correlation Analysis With TMB and MSI

In general, high TMB is associated with better OS; higher TMB is associated with better response to immune checkpoint inhibition. The association between TMB and *PLOD2* expression was evaluated, as seen in **Figure 5A**. *PLOD2* expression was positively correlated with TMB in BRCA ($p < 0.001$), LUAD ($p < 0.001$), THYM ($p < 0.001$), KIRC ($p < 0.001$), SARC ($p = 0.003$), SKCM ($p = 0.009$), and LAML ($p = 0.041$); but negatively correlated with ESCA ($p = 0.004$), THCA ($p = 0.005$), STAD ($p = 0.021$) and COAD ($p = 0.022$). In addition, in the low expression group of PLOD2, there was a significant negative correlation between PLOD2 and TMB (r = -0.160, $p = 0.030$), but there was no significant correlation in the high expression group (r = 0.013, $p = 0.862$). As shown in **Figure 5C**, TMB was significantly correlated with OS. OS in the TMB high group was significantly longer than that in the low group [HR 0.58 95% CI (0.37–0.90), $p = 0.014$].

PLOD2 was positively correlated with MSI in READ ($p < 0.001$), SARC ($p = 0.003$), UCEC ($p = 0.003$), and TGCT ($p =$

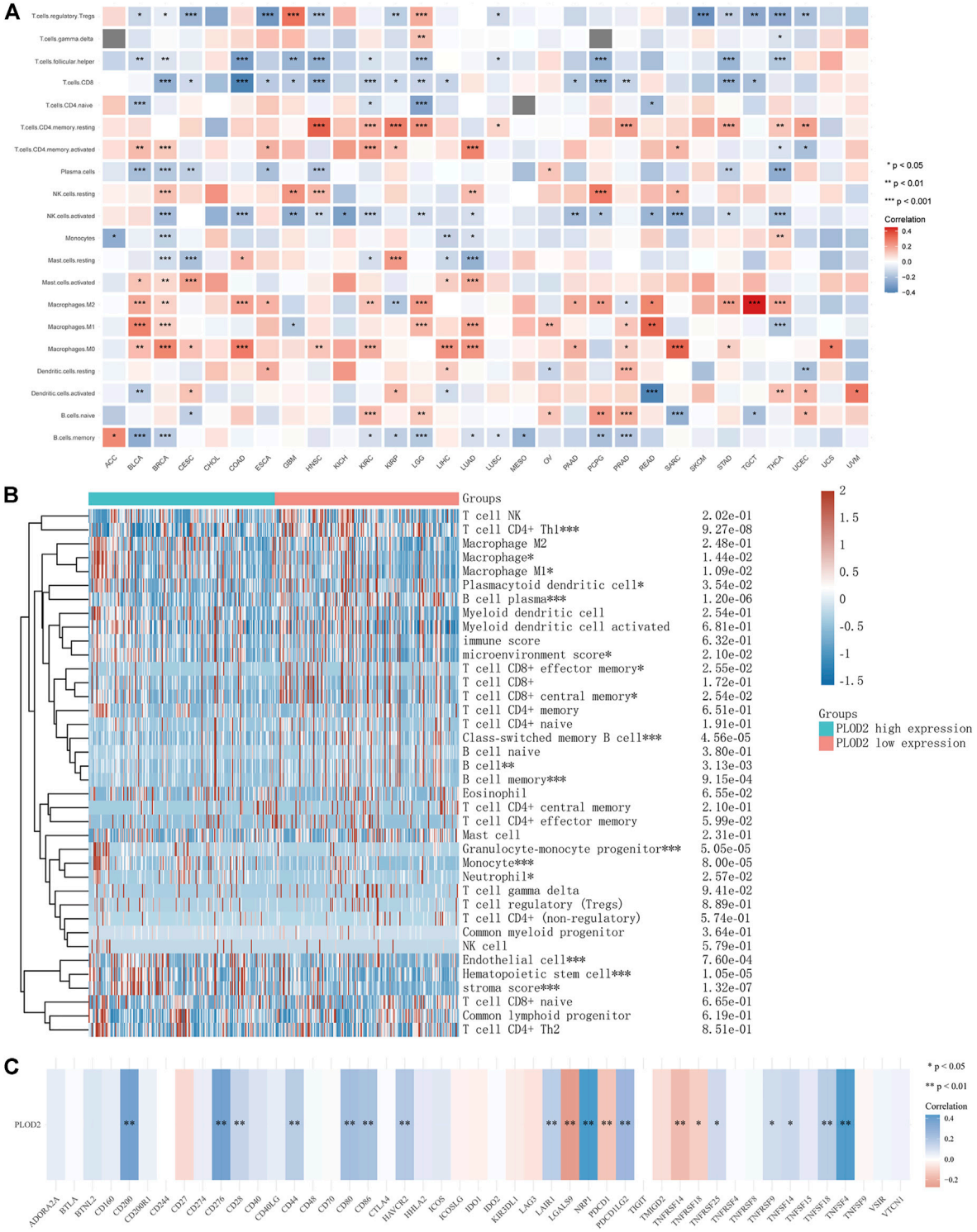


FIGURE 8 | Relationship between *PLOD2* expression and immune cells in pan-cancer. **(A)** CIBERSORT predicts that *PLOD2* expression is correlated with immunocytes. **(B)** Heatmap of *PLOD2* expression and infiltrating immune cells in STAD based on the xCELL algorithm. **(C)** relationship between PLOD2 and immune checkpoint in STAD.

0.048); but negatively correlated with CHOL (p = 0.003), PRAD (p = 0.004) and HNSC (p = 0.008), as shown in **Figure 5B**. With similarly, the results of TMB, PLOD2 was negatively correlated with MSI in the low expression group (r = -0.148, p = 0.043), but not in the high expression group (r = -0.049, p = 0.505). The OS of high MSI group was significantly

longer than that of low group, in **Figure 5D** [HR 0.53 95% CI (0.37–0.75), $p < 0.001$].

Relationship Between Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 2 Somatic Mutation, Mismatch Repairs and DNA Methyltransferase

We downloaded mutect-processed mutation data from TCGA to analyze the mutation of *PLOD2* gene in these tumors. As **Supplementary Figure S6A** demonstrated, the proportion of *PLOD2* mutations in each tumor, which ranged from 5.09% (UCEC) to 0.23% (OV). **Supplementary Figure S6B** shown the distribution of mutations in the top 3 tumors, UCEC (5.09%), COAD (2.76%) and STAD (2.75%). In STAD, the most common of the was Missense Mutation, followed by Frame Shift Del and Nonsense Mutation.

MMRs (Mismatch Repairs) was the repair of nucleotide sequences to normal in DNA molecules containing mismatched bases. Thus, the MMR system was a safety and security system *in vivo* that maintains the integrity and stability of genetic material. As shown in **Figure 6A**, the expression of *PLOD2* was positively correlated with *MLH1*, *MSH2*, and *MSH6* in a variety of tumors. In STAD, *PLOD2* expression was positively correlated with *MLH1*, *MLH6*, *and PSM2*.

In addition, a close relationship was observed between *PLOD2* expression and mutations in 4 methyltransferases (*DNMT1*, *DNMT2*, *DNMT3A*, *DNMT3B*) in several cancer types (**Figure 6B**). For example, in STAD, POLD2 expression was positively correlated with *DNMT3A* (marked green, R = 0.18, $p = 0.002$) and *DNMT3B* (marked purple, R = 0.16, $p = 0.005$) expression.

Functional Analysis

The biological effect of *PLOD2* expression was assessed using GSEA. In STAD, **Figure 7A** showed top 6 significant enrichment GO terms: angiogenesis (GO:0001525); ossification (GO: 0001503); osteoblast differentiation (GO:0001649); regulation of cell growth (GO:0001558); response to acid chemical (GO: 0001101); skeletal system development (GO:0001501). The top 6 KEGG terms also showed significant enrichment (**Figure 7B**): Calcium signaling pathway (hsa04020); cGMP−PKG signaling pathway (hsa04022); Cytokine−cytokine receptor interaction (hsa04060); MAPK signaling pathway (hsa04010); PPAR signaling pathway (hsa03320); Rap1 signaling pathway (hsa04015).

Immunohistochemical Staining

As shown in **Table 3**, the average age of the 20 gastric cancer patients was 62 years old, and women accounted for 35%. Among the clinical stages, stage III or IV accounted for 55%, signet ring cell carcinoma accounted for 25%, and diffuse, intestinal and mixed accounted for 55%, 15%, and 30% respectively. We note that clinical stage III or IV in PLOD2-high group accounted for 75%, and Lauren's type was mainly diffuse and mixed; while clinical stage III or IV in PLOD2-high group accounted for 25%. Lauren's classification was mainly diffuse and intestinal. However, due to the small sample size, there was no significant difference. **Figure 8** showed the immunohistochemical staining results of 20 pairs of gastric cancer and corresponding adjacent tissues. In the gastric cancer group, *PLOD2* was significantly overexpressed, while the expression in adjacent tissues was low. The median values of the two groups were 4 and 2 respectively. There was significant difference in staining score ($p < 0.001$, **Figure 8E**). **Figures 8A,B** were cancer tissues, and **Figures 8C,D** were adjacent tissues. The results of HPA database also showed that the expression of PLOD2 in STAD was higher than that normal gastric tissue (**Figures 8F,G**).

DISCUSSION

The present study aimed to demonstrate a comprehensive workflow for pan-cancer analysis and to extensively investigate the role of *PLOD2* as it related to various cancers. Based on our results, we found that *PLOD2* overexpression was associated with prognosis in a variety of tumors (CHOL, HNSC, KIRC, KIRP, LAML, LUAD, MESO, PAAD, SARC, SKCM, and STAD) based on Cox proportional risk models and KM survival analysis. We focused on the relationship between PLOD2 expression and STAD. Univariate and multivariate analysis showed that the overall survival time of PLOD2 high expression group was significantly less than low group. Immunohistochemical results also showed that the PLOD2 expression in gastric cancer was significantly higher than normal tissues.

In order to investigate the research status of PLOD2 and gastric cancer, we conducted a search on the PubMed using the following search strategy: ["stomach neoplasms" (Title/Abstract) OR "stomach neoplasms" (MeSH Terms) OR "gastric adenocarcinoma" (Title/Abstract)] AND ["PLOD2" (Title/Abstract)]. Finally, seven studies were found (Kiyozumi et al., 2018; Li S. S et al., 2020; Luo et al., 2020; Wang et al., 2020; Dai et al., 2021; Li et al., 2021; Song et al., 2021). Dai et al. constructed a prognostic model for five genes including *PLOD2*, which was subsequently validated by RT-PCR in normal tissue and gastric cancer cell lines (Dai et al., 2021). However, their study did not perform analyses related to tumor immunity (including infiltrating immune cells, TMB, etc.). Similarly, Li J et al.(Li et al., 2021), Li SS et al. (Li S. S et al., 2020), Luo et al. (Luo et al., 2020), and Song et al.(Song et al., 2021) were also constructed multiple genes (including *PLOD2*) prognostic model, but all lacked tumor immune-related analysis or only had comparisons of different immune cell classifications. Kiyozumi et al. study showed that *PLOD2* promotes cell invasion and migration in gastric cancer under hypoxic conditions and leads to dissemination to the peritoneum, *in vitro* (Kiyozumi et al., 2018). This might be even better when coupled with a *PLOD2* knockout or overexpression mouse model. Wang et al. conducted a study on the relationship between *PLOD2* gene expression and gastric

cancer chemotherapy (Wang et al., 2020). Their results showed that *PLOD2* knockdown in BGC823 cells significantly reduced the IC_{50} value of 5-FU, which contributed to the reduction of migration and invasion and promoted apoptosis of gastric cancer cells. The opposite result appeared in *PLOD2* overexpressing MGC803 cells. *In vivo* experiments showed that knockdown of *PLOD2* gene enhanced the inhibitory effect of 5-FU on the growth of transplanted tumors in nude mice. It is particularly unfortunate that the study was only cellular and animal-based, and extrapolation to human gastric cancer requires further validation. In brief, all of the above studies have their own merits and there were many areas for further improvement also.

Our study showed that both OS, DSS, and PFI suggested that the prognosis of PLOD2 high group of was significantly worse than that low group. This may be related to the following reasons. Firstly, the immune cell infiltration in the low group was more abundant (DC, M1 macrophages, CD4 + T cells, CD8 + T cells higher than PLOD2 low group (**Figure 4B**). Secondly, immune checkpoint gene was also significantly overexpressed in the high expression group. Thirdly, the tumor stroma score in the high group was significantly higher than low group (shown in **Figure 4B**). This indicates that the proportion of non-immune cells [e.g., cancer associated fibroblasts (Chen and Song, 2019)] was aplenty in the high group. Derks et al. showed that the infiltration of non-immune cells (such as fibroblasts and stromal cells) was associated with poor prognosis in gastroesophageal adenocarcinomas (Derks et al., 2020). In addition, immunohistochemical results showed that in the high PLOD2 group, the clinical stages were mainly stage III and stage IV, and the proportion of signet ring cell carcinoma was also higher. Signet ring cell carcinomas was usually "cold tumor" (i.e., lack of immune infiltration) (Garcia-Pelaez et al., 2021; Monster et al., 2022). Therefore, we speculated that the high expression of PLOD2 and poor prognosis may be related to immune infiltration and pathological types. However, further animal experiments were needed to prove it.

In conclusion, we have found that *PLOD2* can serve as a valuable prognostic biomarker for some tumors, especially gastric cancer. We believe that these findings may lay the groundwork for prospective functional experiments and eventually have an impact in clinical work.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Beijing Jishuitan Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

QX and PG designed the manuscript. QX wrote and completed the manuscript. QX, NK, YZ, QW, XW, and XX completed the data download and analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by Beijing JST Research Funding (ZR-201919).

ACKNOWLEDGMENTS

We thanks to Sangerbox platform (http://sangerbox.com/) and Assistant for Clinical Bioinformatics platform (https://www.aclbi.com/) for providing support in data analysis and plotting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.864655/full#supplementary-material

REFERENCES

Alexandrov, L. B., Kim, J., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., et al. (2020). The Repertoire of Mutational Signatures in Human Cancer. *Nature* 578 (7793), 94–101. doi:10.1038/s41586-020-1943-3

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1

Author Anonyms (2020). The Era of Massive Cancer Sequencing Projects Has Reached a Turning point. *Nature* 578 (7793), 7–8. doi:10.1038/d41586-020-00308-w

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Bejarano, L., Jordão, M. J. C., and Joyce, J. A. (2021). Therapeutic Targeting of the Tumor Microenvironment. *Cancer Discov.* 11 (4), 933–959. doi:10.1158/2159-8290.cd-20-1808

Bray, F., Laversanne, M., Weiderpass, E., and Soerjomataram, I. (2021). The Ever-increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide. *Cancer* 127 (16), 3029–3030. doi:10.1002/cncr.33587

Calabrese, C., Davidson, N. R., Davidson, N. R., Demircioğlu, D., Fonseca, N. A., He, Y., et al. (2020). Genomic Basis for RNA Alterations in Cancer. *Nature* 578 (7793), 129–136. doi:10.1038/s41586-020-1970-0

Chang, K., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764

Chen, X., and Song, E. (2019). Turning Foes to Friends: Targeting Cancer-Associated Fibroblasts. *Nat. Rev. Drug Discov.* 18 (2), 99–115. doi:10.1038/s41573-018-0004-1

Dai, W., Xiao, Y., Tang, W., Li, J., Hong, L., Zhang, J., et al. (2021). Identification of an EMT-Related Gene Signature for Predicting Overall Survival in Gastric Cancer. *Front. Genet.* 12, 661306. doi:10.3389/fgene.2021.661306

Derks, S., de Klerk, L. K., Xu, X., Fleitas, T., Liu, K. X., Liu, Y., et al. (2020). Characterizing Diversity in the Tumor-Immune Microenvironment of Distinct Subclasses of Gastroesophageal Adenocarcinomas. *Ann. Oncol.* 31 (8), 1011–1020. doi:10.1016/j.annonc.2020.04.011

Dey, P., Kimmelman, A. C., and DePinho, R. A. (2021). Metabolic Codependencies in the Tumor Microenvironment. *Cancer Discov.* 11 (5), 1067–1081. doi:10.1158/2159-8290.cd-20-1211

Du, H., Pang, M., Hou, X., Yuan, S., and Sun, L. (2017). PLOD2 in Cancer Research. *Biomed. Pharmacother.* 90, 670–676. doi:10.1016/j.biopha.2017.04.023

Garcia-Pelaez, J., Barbosa-Matos, R., Gullo, I., Carneiro, F., and Oliveira, C. (2021). Histological and Mutational Profile of Diffuse Gastric Cancer: Current Knowledge and Future Challenges. *Mol. Oncol.* 15 (11), 2841–2867. doi:10.1002/1878-0261.12948

Genecards (2021). GeneCards Symbol: PLOD2. Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLOD2 (Accessed August 2nd, 2021).

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The Prognostic Landscape of Genes and Infiltrating Immune Cells across Human Cancers. *Nat. Med.* 21 (8), 938–945. doi:10.1038/nm.3909

Gerstung, M., Jolly, C., Jolly, C., Leshchiner, I., Dentro, S. C., Gonzalez, S., et al. (2020). The Evolutionary History of 2,658 Cancers. *Nature* 578 (7793), 122–128. doi:10.1038/s41586-019-1907-7

Gilkes, D. M., Bajpai, S., Wong, C. C., Chaturvedi, P., Hubbi, M. E., Wirtz, D., et al. (2013). Procollagen Lysyl Hydroxylase 2 Is Essential for Hypoxia-Induced Breast Cancer Metastasis. *Mol. Cancer Res.* 11 (5), 456–466. doi:10.1158/1541-7786.mcr-12-0629

Gilkes, D. M., Semenza, G. L., and Wirtz, D. (2014). Hypoxia and the Extracellular Matrix: Drivers of Tumour Metastasis. *Nat. Rev. Cancer* 14 (6), 430–439. doi:10.1038/nrc3726

Hutter, C., and Zenklusen, J. C. (2018). The Cancer Genome Atlas: Creating Lasting Value beyond its Data. *Cell* 173 (2), 283–285. doi:10.1016/j.cell.2018.03.042

Icgc/Tcga Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer Analysis of Whole Genomes. *Nature* 578 (7793), 82–93. doi:10.1038/s41586-020-1969-6

Kiyozumi, Y., Iwatsuki, M., Kurashige, J., Ogata, Y., Yamashita, K., Koga, Y., et al. (2018). PLOD2 as a Potential Regulator of Peritoneal Dissemination in Gastric Cancer. *Int. J. Cancer* 143 (5), 1202–1211. doi:10.1002/ijc.31410

Kocher, F., Tymoszuk, P., Amann, A., Sprung, S., Salcher, S., Daum, S., et al. (2021). Deregulated Glutamate to Pro-collagen Conversion Is Associated with Adverse Outcome in Lung Cancer and May Be Targeted by Renin-Angiotensin-Aldosterone System (RAS) Inhibition. *Lung Cancer* 159, 84–95. doi:10.1016/j.lungcan.2021.06.020

Li, J., Pu, K., Li, C., Wang, Y., and Zhou, Y. (2021). A Novel Six-Gene-Based Prognostic Model Predicts Survival and Clinical Risk Score for Gastric Cancer. *Front. Genet.* 12, 615834. doi:10.3389/fgene.2021.615834

Li, S.-S., Lian, Y.-F., Huang, Y.-L., Huang, Y.-H., and Xiao, J. (2020). Overexpressing PLOD Family Genes Predict Poor Prognosis in Gastric Cancer. *J. Cancer* 11 (1), 121–131. doi:10.7150/jca.35763

Li, Y., Roberts, N. D., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., et al. (2020). Patterns of Somatic Structural Variation in Human Cancer Genomes. *Nature* 578 (7793), 112–121. doi:10.1038/s41586-019-1913-9

Luo, T., Du, Y., Duan, J., Liang, C., Chen, G., Jiang, K., et al. (2020). Development and Validation of a Scoring System Based on 9 Glycolysis-Related Genes for Prognosis Prediction in Gastric Cancer. *Technol. Cancer Res. Treat.* 19, 1533033820971670. doi:10.1177/1533033820971670

Miyamoto, K., Seki, N., Matsushita, R., Yonemori, M., Yoshino, H., Nakagawa, M., et al. (2016). Tumour-suppressive miRNA-26a-5p and miR-26b-5p Inhibit Cell Aggressiveness by Regulating PLOD2 in Bladder Cancer. *Br. J. Cancer* 115 (3), 354–363. doi:10.1038/bjc.2016.179

Monster, J. L. J., Kemp, M., Gloerich, R. S., and van der Post, R. (2022). Diffuse Gastric Cancer: Emerging Mechanisms of Tumor Initiation and Progression. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1877, 188719. doi:10.1016/j.bbcan.2022.188719

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining Cell Type Abundance and Expression from Bulk Tissues with Digital Cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2

Passaro, A., Stenzinger, A., and Peters, S. (2020). Tumor Mutational Burden as a Pan-Cancer Biomarker for Immunotherapy: The Limits and Potential for Convergence. *Cancer Cell* 38 (5), 624–625. doi:10.1016/j.ccell.2020.10.019

Rheinbay, E., Nielsen, M. M., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., et al. (2020). Analyses of Non-coding Somatic Drivers in 2,658 Cancer Whole Genomes. *Nature* 578 (7793), 102–111. doi:10.1038/s41586-020-1965-x

Song, W., Bai, Y., Zhu, J., Zeng, F., Yang, C., Hu, B., et al. (2021). A Novel Prognostic Model Based on Epithelial-Mesenchymal Transition-Related Genes Predicts Patient Survival in Gastric Cancer. *World J. Surg. Onc* 19 (1), 216. doi:10.1186/s12957-021-02329-9

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tadeo, I., Berbegall, A. P., Castel, V., García-Miguel, P., Callaghan, R., Påhlman, S., et al. (2016). Extracellular Matrix Composition Defines an Ultra-high-risk Group of Neuroblastoma within the High-Risk Patient Cohort. *Br. J. Cancer* 115 (4), 480–489. doi:10.1038/bjc.2016.210

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T.-H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48 (4), 812–830. doi:10.1016/j.immuni.2018.03.023

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-Based Map of the Human Proteome. *Science* 347 (6220), 1260419.

Valero, C., Lee, M., Hoen, D., Wang, J., Nadeem, Z., Patel, N., et al. (2021). The Association between Tumor Mutational burden and Prognosis Is Dependent on Treatment Context. *Nat. Genet.* 53 (1), 11–15. doi:10.1038/s41588-020-00752-4

Vilar, E., and Gruber, S. B. (2010). Microsatellite Instability in Colorectal Cancer-The Stable Evidence. *Nat. Rev. Clin. Oncol.* 7 (3), 153–162. doi:10.1038/nrclinonc.2009.237

Wang, X., Guo, J., Dai, M., Wang, T., Yang, T., Xiao, X., et al. (2020). PLOD2 Increases Resistance of Gastric Cancer Cells to 5-fluorouracil by Upregulating BCRP and Inhibiting Apoptosis. *J. Cancer* 11 (12), 3467–3475. doi:10.7150/jca.41828

Wang, P., Chen, Y., and Wang, C. (2021). Beyond Tumor Mutation Burden: Tumor Neoantigen Burden as a Biomarker for Immunotherapy and Other Types of Therapy. *Front. Oncol.* 11, 672677. doi:10.3389/fonc.2021.672677

Xin, W., Zhao, C., Jiang, L., Pei, D., Zhao, L., and Zhang, C. (2021). Identification of a Novel Epithelial-Mesenchymal Transition Gene Signature Predicting Survival in Patients with HNSCC. *Pathol. Oncol. Res.* 27, 585192. doi:10.3389/pore.2021.585192

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xu, Kong, Zhao, Wu, Wang, Xun and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fgene.2022.836199

DEAD-Box Helicase 27 Triggers Epithelial to Mesenchymal Transition by Regulating Alternative Splicing of Lipoma-Preferred Partner in Gastric Cancer Metastasis

Yirong Jin[1†], Suzhen Yang[2†], Xiaoliang Gao[1], Di Chen[1], Tingting Luo[3], Song Su[1], Yanting Shi[1], Gang Yang[1], Lei Dong[2] and Jie Liang[1]**

[1]State Key Laboratory of Cancer Biology, National Clinical Research Center for Digestive Diseases, Air Force Military Medical University, Xi'an, China, [2]Department of Digestive Disease and Gastrointestinal Motility Research Room, Xi'an Jiaotong University, Xi'an, China, [3]Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, School of Medicine, Northwest University, Xi'an, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Massimo Zollo,
University of Naples Federico II, Italy
Yaseswini Neelamraju,
University of Virginia, United States

***Correspondence:**
Lei Dong
dong556@126.com
Jie Liang
liangjie@fmmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 15 December 2021
Accepted: 06 April 2022
Published: 04 May 2022

Citation:
Jin Y, Yang S, Gao X, Chen D, Luo T, Su S, Shi Y, Yang G, Dong L and Liang J (2022) DEAD-Box Helicase 27 Triggers Epithelial to Mesenchymal Transition by Regulating Alternative Splicing of Lipoma-Preferred Partner in Gastric Cancer Metastasis. Front. Genet. 13:836199. doi: 10.3389/fgene.2022.836199

DEAD-box helicase 27 (DDX27) was previously identified as an important mediator during carcinogenesis, while its role in gastric cancer (GC) is not yet fully elucidated. Here, we aimed to investigate the mechanism and clinical significance of DDX27 in GC. Public datasets were analyzed to determine DDX27 expression profiling. The qRT-PCR, Western blot, and immunohistochemistry analyses were employed to investigate the DDX27 expression in GC cell lines and clinical samples. The role of DDX27 in GC metastasis was explored *in vitro* and *in vivo*. Mass spectrometry, RNA-seq, and alternative splicing analysis were conducted to demonstrate the DDX27-mediated molecular mechanisms in GC. We discovered that DDX27 was highly expressed in GCs, and a high level of DDX27 indicated poor prognosis. An increased DDX27 expression could promote GC metastasis, while DDX27 knockdown impaired GC aggressiveness. Mechanically, the LLP expression was significantly altered after DDX27 downregulation, and further results indicated that LPP may be regulated by DDX27 *via* alternative splicing. In summary, our study indicated that DDX27 contributed to GC malignant progression *via* a prometastatic DDX27/LPP/EMT regulatory axis.

Keywords: DDX27, gastric cancer metastasis, LPP, alternative splicing, EMT

INTRODUCTION

Gastric cancer was responsible for over 1 million new cases and 0.7 million deaths in 2020, which is still the main cause of cancer morbidity and mortality worldwide (Sung et al., 2021). Despite advances in early diagnosis and treatment, the incidence of recurrent and metastatic gastric cancer remains stubbornly high (Uemura et al., 2001). Metastasis is one of the most significant features of advanced GC, and a large proportion of GC patients in advanced stages benefit less than expected mainly due to metastasis (Soerjomataram and Bray 2021). GC is a prevalent but heterogeneous disease, arising from genetic accumulation and epigenetic alterations (Tan and Yeoh 2015). This gradual accumulation of gene alterations leads to continuous growth and metastatic advantage of neoplastic cells and substantially promotes GC malignant progression (Hayakawa et al., 2021). Thus, it is an urgent necessity to identify specific targets contributing to GC metastasis.

DDX27 belongs to the DEAD-box RNA helicases family, which has been verified to participate in carcinogenesis (Mazurek et al., 2012; Zhang et al., 2019; Mo et al., 2021). Accumulative evidence suggests that DDX27 serves as an oncogene in various cancers, such as colorectal cancer (CRC), hepatocellular carcinoma (HCC), and GC. For example, DDX27 accelerates CRC progression by forming a DDX27-NPM1-NFκB functional axis (Tang et al., 2018). Also, Tsukamoto et al. reported that DDX27 accelerated GC proliferation by inducing TP53-dependent cell cycle arrest (Tsukamoto et al., 2015). Meanwhile, DDX27 causes chemotherapeutic resistance toward epirubicin and cisplatin in the GC cells (Zhou et al., 2015). Additionally, DDX27 clues the unfavorable outcomes of HCC patients (Wang et al., 2015). Nevertheless, the role of DDX27 in malignant tumors still remains largely unclear, especially in the GC metastatic process.

Alternative splicing (AS) plays an indispensable role in enlarging gene expression patterns and enriching protein diversity. Dysregulation of splicing variants has always been strongly associated with tumor malignancy, including poor differentiation, metastasis, and poor prognosis of patients (Hönig et al., 2002; Dardenne et al., 2012; Germann et al., 2012). Abnormal expression levels and activities of splicing factors linked with abnormal AS patterns were extensively detected in different kinds of tumors (Rahman et al., 2020). A study using 32 different types of cancer tissues revealed that abnormal AS events always happen during cancer progression (Kahles et al., 2018). Meanwhile, the DDX family is generally involved in spliceosome assemblies and functioned as one of the canonical regulators of splicing in RNA metabolism (Ameur et al., 2020), which is gaining growing attention in cancer research for its vital roles in AS of many tumor-associated genes, such as macroH2A1 histones (NFAT5) (Germann et al., 2012) and CD44 (Hönig et al., 2002).

Lipoma-preferred partner (LPP) is located at chromosome 3q27-q28, owned by the zyxin family of proteins (Ngan et al., 2018). It is primarily expressed in the cell periphery of focal adhesion, thus participating in cytoskeletal organization, cell motion, and mechanosensing (Grunewald et al., 2009). The malignant role of LPP within carcinogenesis has been previously revealed. For example, LPP mediates breast cancer cells metastasize to lungs by participating in the formation of invadopodia and regulating cell motility. LPP promotes tumor angiogenesis and confers chemoresistance to ovarian cancer (Leung et al., 2018). In addition, LPP has been noticed to be regulated in an alternative splicing manner (Petit et al., 1996), which contains a number of transcription variants (Yamanaka et al., 2008).

Here, we elucidated the critical role of DDX27 in gastric cancer progression and found that DDX27 promoted the GC EMT process by regulating the alternative splicing of LPP.

MATERIALS AND METHODS

Public Datasets

Public datasets were obtained from the Gene Expression Omnibus (GEO) database, the Cancer Genome Atlas (TCGA) program, and Gene Expression Profiling Interactive Analysis 2 (GEPIA2). Each dataset was applied for separate usage, according to the data characteristics respectively. In details, gastric cancer transcriptome data are fully or partly acquired from the data generated by TCGA Research Network: https://www.cancer.gov/tcga. GEPIA2 (http://gepia2.cancer-pku.cn/) was performed to analyze the difference of DDX27 between gastrointestinal tumors and healthy specimens at translation levels. The cBioPortal (https://www.cbioportal.org/) database and the gastric cancer dataset are employed to analyze the mutation information of DDX27 and clinical characteristics in gastric cancer. The effects of DDX27 and LPP expression on the survival time among the GC patients were analyzed by the Kaplan–Meier plotter (http://kmplot.com/analysis/).

Patients and Specimens

A total of 104 GC specimens (52 paired primary tumor and normal tissues) were obtained from Xijing hospital, Air Force Military University, between the year of 2018 and 2020, in accordance with ethical approvals. Fresh tissue specimens were divided into three parts: the first part is formalin-fixed, the second is paraffin-embedded, and the last was kept at liquid nitrogen. The study was approved by the Clinical Research Ethics Committee of the Air Force Military Medical University and Xijing Hospital. Informed consent was obtained from all the subjects involved in the study.

Cell Culture

All the cell lines (MKN-28, AGS, BGC-823, SNU-1, MKN-45, HGC-27, and GES-1) were restored in our laboratory for research. DMEM (Gibco, Grand Island, NY, United States) was used to culture HGC-27 only, while the others were cultured in RPMI1640 (Gibco). For the process, 10% fetal bovine serum (FBS, Gibco, Grand Island, NY, United States), 100 μg/ml streptomycin, and 100 IU/ml penicillin were added into a medium to culture cells at 37°C in a humidified 5% CO_2 atmosphere.

Quantitative Reverse Transcription–Polymerase Chain Reaction

Following the instructions of the manufacturer, the total RNA in cell lines and tissues was extracted using the RNAiso reagent (TaKaRa, Japan). PrimeScript RT Master Mix (TaKaRa) was selected to perform reverse transcription. Synthesized cDNA was used to conduct real-time PCR by the SYBR Premix Ex Taq II (TaKaRa). The CFX96 RealTime PCR detection system (Bio-Rad, CA, United States) was adopted for detection. The determined relative expression levels of target genes were obtained using the $2^{-\Delta\Delta Ct}$ method, and β-actin was chosen as the internal control for mRNAs. The primers used are listed in **Supplementary Table S1**.

Western Blot

The total proteins from cell or tissue lysates were transferred onto the nitrocellulose membranes after being resolved by SDS-PAGE. The membranes were blocked in 5% fat-free milk and then

incubated with primary specific antibodies at 4°C overnight (DDX27: ab177938, Abcam; LPP: #3389, CST, MA, United States; E-cadherin: #14472, CST; Vimentin: #5741, CST; N-cadherin: #13116, CST; ZO-1: #13663, CST; β-actin: #3700, CST). Appropriate HRP-conjugated secondary antibody was used to further incubate the protein band for 1 h. The Molecular Imager ChemiDoc XRS + Imaging System was used for membrane visualization, and protein quantification was obtained by Image Lab software.

Immunohistochemistry and Evaluation

IHC staining was performed on paraffin sections or microarrays. In brief, all the slides were dewaxed at 65°C for 2 h, and each slide got antigen repairment with ethylenediaminetetraacetic acid (EDTA) or citric acid. Hydrogen peroxide was used to eliminate endogenous peroxidase (EGPO). After blocking by using serum for 1 h, the slides were incubated with specific primary antibodies. At last, the sections were incubated with horseradish peroxidase (HRP)-labeled secondary antibody, followed by coloring with diaminobenzidine (Dako). All the slides were counterstained with hematoxylin.

The IHC protein expression level was evaluated by two independent pathologists using semi-quantitative scoring of the intensity methods. The scoring system contains two parameters, including the percentage of positive cells and the intensity of staining. The percentage of cells was divided into four degrees: < 1% (1), 1–25% (2), 26–75% (3), and >75% (4). Negative (0), weak (1), moderate (2), and strong (3) were conducted to evaluate the intensity of staining. The product of the two parameters was equaled to get a final score. The final score 0–6 was graded as low, and 8–12 was defined as high, which represented "low expression" and "high expression" of the target molecules.

Stable Cell Lines Construction

To silence the target genes, the siRNA hairpin sequence was cloned into the GV248-Puro lentivirus vector, and a fragment of cDNA was cloned into the GV492-Puro lentivirus vector to overexpress the target gene. Corresponding negative controls (LV-Control and shControl) were also constructed by Shanghai GeneChem Company. Hitrans G (GeneChem, China) was applied for cell infection, following the instruction. In brief, the cells were cultured in OPTI-MEM (Gibco, United States) and transfected with lentivirus for 24 h; then, the medium was replaced with fresh complete media containing puromycin (Merck Millipore, Germany) for stable cell selection for at least 1 week. The sequences used are exhibited in **Supplementary Table S1**. The efficiency of lentivirus was detected using the real-time PCR and Western blot.

Trans-Well Assay

Corning chambers containing the 8-μm pore polycarbonate membrane filter were selected to perform the trans-well migration assays. The 8×10^4 cells were added to the upper chamber after being suspended with a serum-free medium. A measure of 600 μl of culture medium containing 20% FBS was added into the bottom chamber. The cells on top of the membrane were wiped off after 24 h of incubation. The cells migrated to the bottom surface of the filter, were fixed, and then stained with 1% crystal violet. Each trans-well insert was microscopically imaged in five randomly selected regions at x20 in triplicates. The cells that migrated to the bottom surface were shot for statistical analysis. Similarly, invasion assays were performed, following the same protocol as migration assays, except that the 8-μm microporous filter was coated with the configured Matrigel (Corning) to examine the cell invasive abilities.

Wound Healing

The cells were seeded in 6-well plates. When grew to 90% confluence, the attached cell layers were carefully scratched by a 200-μl pipette tip to generate wounds. The remaining adherent cells were incubated with the serum-free medium and then photographed at different time points (0, 24 and 48 h) using the inverted microscope (Olympus, Tokyo, Japan).

Metastasis Assays *In Vivo*

BALB/c nude mice were raised in specific comfortable and pathogen-free conditions. The mice (ten, 4–6-week mice per group, 18–22 g) were injected with cells transferred with recombinant lentiviruses *via* the tail vein to construct the vivo metastatic model. Separated metastatic nodes in lung specimens were separated from the sacrificed mice and fixed with paraformaldehyde (1%) about 6–8 weeks later. After H&E staining, the nodes were counted to evaluate cell metastasis potentials. The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of the Air Force Military Medical University.

Gene Set Enrichment Analysis

GSEA was performed by GSEA 4.1.0 software (GSEA, United States), and the gene set database was set as ftp.broadinstitute.org://pub/gsea/gene_sets/h.all.v7.4.symbols. The number of permutations was set at 1000 and the permutation type as phenotype. For single-gene GSEA, TCGA rpkm data were divided into two groups. The top 25% was identified as high expression, and the bottom 75% was perceived as low expression, followed by GSEA analysis.

Transcriptome Sequencing and Alternative Splicing Analysis

RNAseq was conducted between the DDX27 knockdown and control cell lines. The transcriptome sequencing was performed on the Illumina sequencing platform. Illumina PE libraries (~300bp) were constructed for sequencing, and the sequencing data obtained were quality controlled, after which the transcriptome data were analyzed using bioinformatics tools. The clean data are compared with the reference genome to obtain the whole transcriptomic information. RPKM (Reads per Kilobase per Million Reads) values as a measure of the gene expression. For the AS analysis, we used rMATS (replicate Multivariate Analysis of Transcript Splicing, University of California) software for the variable splicing analysis of the differences. I is the exon inclusion reads, and S is the exon skipping reads, LI is the effective length of the

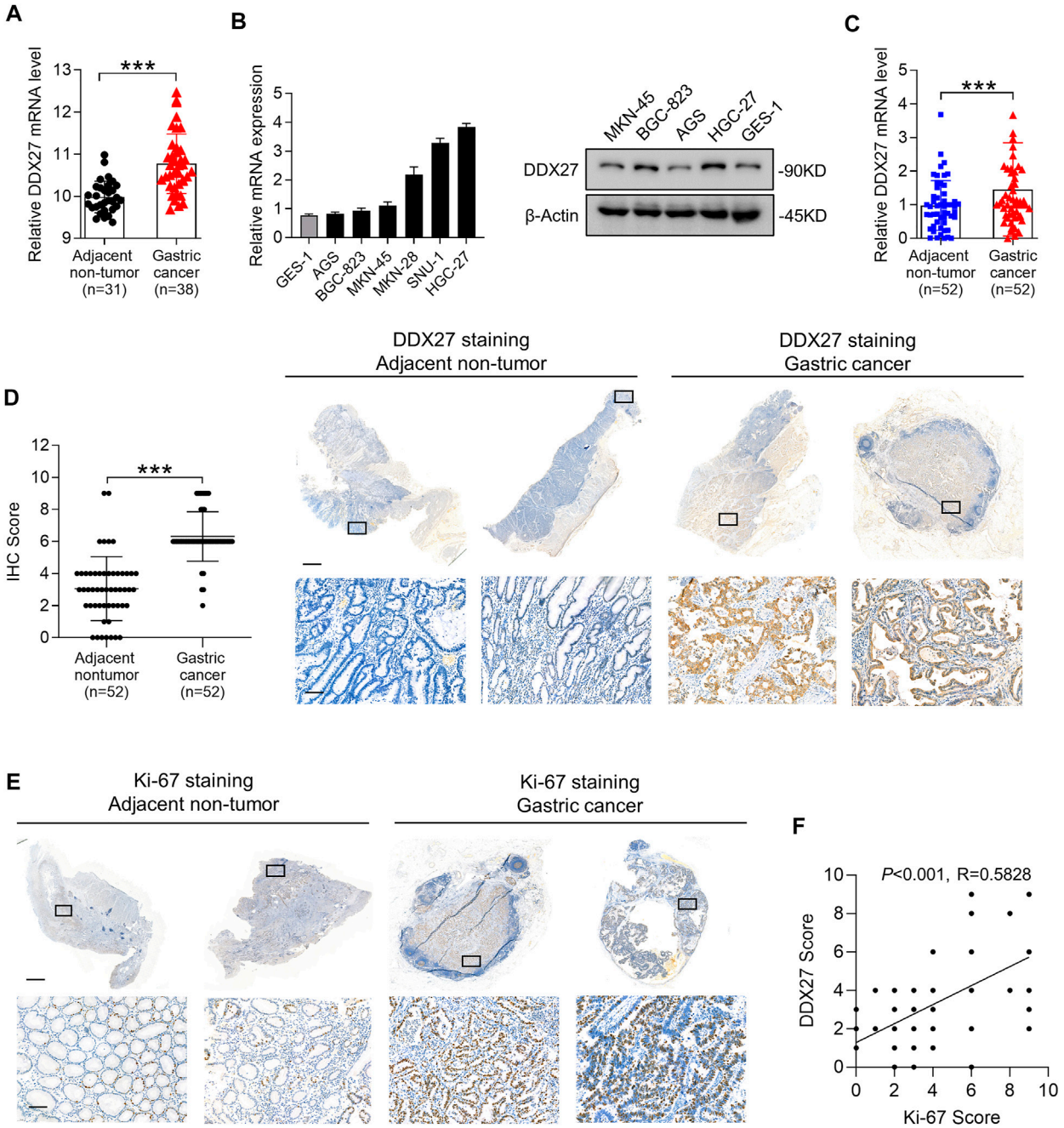


FIGURE 1 | DDX27 exhibits an elevated expression profile in GC. **(A)** DDX27 transcript levels between the adjacent nontumor and GC tissues in the GEO database (GSE13911, ***$p < 0.001$, by Student's t test). **(B)** Quantitative reverse transcription PCR (qRT-PCR) (left) and Western blot (right) analyses of the DDX27 expression in the normal gastric epithelial cell line (GES-1) and gastric cancer cell lines (ASG, BGC-823, MKN-45, MKN-28, SNU-1, and HGC-27). **(C)** DDX27 expression was analyzed by qRT-PCR in paired surgically resected GC ($n = 52$) and adjacent nontumor tissues ($n = 52$) (***$p < 0.001$, by Student's t test). **(D)** IHC score of DDX27 in clinically paired collected adjacent nontumor ($n = 52$) and GC tissues ($n = 52$) (left, ***$p < 0.001$, by Student's t test)). Representative images of staining were exhibited in the right. Scale bars: 1000 μm (up) or 100 μm (below). **(E)** IHC staining images of Ki-67 in the adjacent nontumor ($n = 52$) and GC tissues ($n = 52$). Scale bars:1000 μm (up) or 100 μm (below). **(F)** Correlation analysis between DDX27 and the Ki-67 IHC staining score. ($p < 0.001$, $R = 0.5828$, by Pearson correlation method). $n \geq 3$, the data are presented as the mean ± standard deviation (SD). A high DDX27 expression indicates poor clinical prognosis.

inclusion isoform (length of the inclusion form), and LS is the effective length of the skipping isoform (length of the skipping form); then, the estimated value of ψ is

$$\hat{\psi} = \frac{I/L_I}{I/L_I + S/Ls}.$$

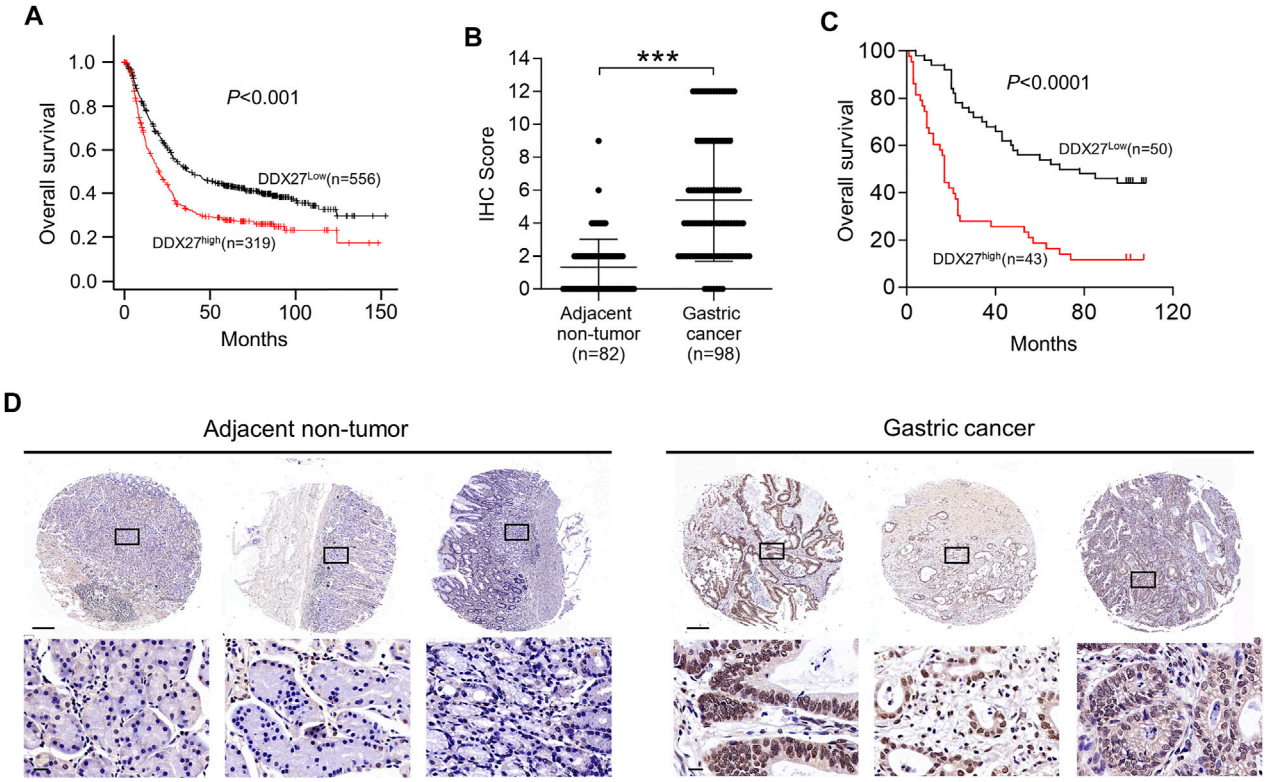


FIGURE 2 | High DDX27 expression indicated a poor clinical prognosis. **(A)** Kaplan-Meier analysis of overall survival (OS) between GC patients with high ($n = 319$) or low ($n = 556$) DDX27 expression from the dataset. ($p < 0.001$, by log-rank test). **(B,D)** IHC scoring of paired adjacent nontumor ($n = 82$) and GC tissues ($n = 98$) [**(B)**, ***$p < 0.001$, by Student's t-test] with the representative images of IHC staining for DDX27 protein levels **(D)**. Scale bars: 250 μm (up) or 20 μm (below). **(C)** Kaplan–Meier analysis was used to evaluate the DDX27 expression in patients with GC form tissue microarray ($p < 0.001$, by log-rank test). $n \geq 3$, the data are presented as the mean ± SD. DDX27 is a critical mediator of GC metastasis *via* EMT.

The alternative splicing threshold is set as p value < 0.05, difference $|\Delta\psi|>0.1$. GO term annotation and KEGG pathway associated with alternative splicing are analyzed with $p < 0.05$.

Statistical Analysis

All data were expressed as the mean ± standard errors. GraphPad Prism 8.4.0 software (GraphPad Software, United States) and SPSS software (IBM SPSS, Armonk, NY, United States) were employed for the analysis. Considering the data characteristics, differential analysis was performed by the Wilcoxon rank-sum test, two-group t-test, ANOVA, and Student's t test. The cut-off value for the high or low expression was defined with ROC analysis. The Cox regression model, Kaplan–Meier curve, and log-rank test were applied for the survival analysis. Spearman correlation analysis was utilized for correlation assessment. $p < 0.05$ was considered statistically significant.

RESULTS

DEAD-Box Helicase 27 Exhibits an Elevated Expression Profile in Gastric Cancer

Large amounts of studies have regarded DDX27 as vital cancer-promoting genes in distinct tumors (Tsukamoto et al., 2015; Tang et al., 2018). However, the definite role of DDX27 in GC remains largely uncharacterized. Therefore, we first assessed the DDX27 expression among various gastrointestinal tumors. TCGA gene expression profiling showed that the high expression of DDX27 was exhibited in the gastrointestinal tumor tissues, especially in stomach adenocarcinoma (STAD) (**Supplementary Figure S1A**). From the Gene Expression Omnibus (GEO) database, we again found that GC tissues contained a higher level of DDX27 than the adjacent nontumor tissues (**Figure 1A**).

Subsequently, we detected the DDX27 mRNA and protein levels in different GC cell lines and noticed that compared to the normal gastric epithelial cell line (GES-1), DDX27 was significantly upregulated in the GC cells (AGS, BGC-823, MKN-45, MKN-28, SNU-1, and HGC-27) (**Figure 1B**). By analyzing the paired clinical collected GC and adjacent nontumor tissues with qRT-PCR, an elevated mRNA level of DDX27 was found in the GC tissues rather than in nontumor tissues (**Figure 1C**, $n = 52$).

Ki-67 is a representative marker of cancer malignancy, and the paired clinical specimens were utilized to observe the expression of DDX27 and Ki67 with IHC analysis. IHC staining indicated that DDX27 and Ki-67 were upregulated in the GC tissues and mainly distributed in the nucleus (**Figures 1D,E**). In addition, the DDX27 IHC score exhibited a positive correlation with the Ki67

TABLE 1 | Association between DDX27 and clinicopathological parameters of GC patients.

Characteristic	Total	DDX27		*p*
		Low	High	
Gender				
Female	36	20	16	0.823447
Male	62	33	29	
Age (years)				
≤60	35	21	14	0.380858
>60	63	32	31	
Location				
Cardia	11	4	7	0.648887
Fundus/body of stomach	29	17	12	
Antrum	50	27	23	
Diffuse	7	4	3	
Tumor size				
≤6 cm	61	36	25	0.126739
>6 cm	30	15	20	
Depth of invasion				
T 1–3	77	49	28	0.000479
T 4	20	4	16	
Metastases to other organs				
M 0	89	51	38	0.044156
M 1	9	2	7	
Number of positive lymph nodes				
≤5	57	36	21	0.043427
>5	38	16	22	

PS: bold fonts represent p < *0.05.*

IHC score in the GC tissues (**Figure 1F**). Overall, we can draw the conclusion that DDX27 was significantly upregulated in GC and may participate in facilitating the GC malignancy development.

High DEAD-Box Helicase 27 Expression Indicates Poor Clinical Prognosis

To explore the clinical signature of DDX27 in GC, we conducted Kaplan–Meier (KM) analysis and discovered GC patients with a higher DDX27 expression always had worse overall survival (OS) ($p < 0.001$; **Figure 2A**). IHC staining of tissue microarray containing GC ($n = 98$) exhibited a relatively high expression level of DDX27 in GC compared with the nontumor tissues ($n = 82$) ($p < 0.001$; **Figures 2B,D**).

In line with the results of **Figure 2A**, IHC staining of the tissue microarray showed that GC patients with a high DDX27 expression tend to have a worse prognosis ($p < 0.0001$; **Figure 2C**). Moreover, DDX27 had a positive correlation with the depth of invasion, lymph node metastasis, and distant organ metastasis ($p < 0.05$, **Table 1**). Univariate analysis revealed that tumor size, invasion depth, lymph node metastasis, distant organic invasion, AJCC, and DDX27 expression all were associated with OS ($p < 0.05$, **Table 2**), and multivariate Cox regression analysis validated that a higher DDX27 expression was an independent risk factor for the GC patient prognosis and could predict patient's shorter OS ($p < 0.05$, **Table 3**).

TABLE 2 | Prognostic factors in patients with gastric cancer by univariate analysis.

Characteristic	*n*	HR	95% CI	*p*
Gender				
Female	35	1.223855	0.748886–2.000066	0.420199
Male	58			
Age (years)				
≤60	32	1.05675	0.6329–1.7643997	0.832845
>60	61			
Location				
Cardia	10	0.905378	0.647746–1.265480	0.560687
Fundus/body of stomach	26			
Antrum	49			
Diffuse	7			
Tumor size				
≤6 cm	61	2.09	1.2619–3.46166	0.004190
>6 cm	30			
Depth of invasion				
T 1–3	73	1.9813	1.13713–3.4523	0.015797
T 4	19			
Metastases to other organs				
M 0	85	3.948667	1.8432–8.58917	0.000436
M 1	8			
Number of positive lymph nodes				
≤5	54	2.7105	1.6359–4.49099	0.000109
>5	36			
AJCC				
1/2	39	2.069	1.24078–3.450429	0.005323
3/4	53			
DDX27				
Low	50	3.085071	1.877780–5.068573	0.000009
High	43			

PS: bold fonts represent p < *0.05.*

In addition, we investigated the DDX27 genetic alterations among a GC cohort (OncoSG, 2018) with whole-genome sequencing and found that DDX27 amplification account for about 21%, indicating a high frequency of amplification of DDX27 among the clinical GC patients. Specific clinical features of DDX27 amplification are shown in **Supplementary Figure S1B**. All these results suggested that DDX27 have the potential of being a potential indicator in GC prognosis.

DEAD-Box Helicase 27 is a Critical Mediator of Gastric Cancer Metastasis *via* Epithelial Mesenchymal Transition

The findings we have gained suggested us that the biological function of DDX27 may promote GC progression. To unveil the role of DDX27 in tumor malignancy, we successfully constructed a stable DDX27 overexpression (LV-DDX27) and knockdown (shDDX27) cell models in three GC cell lines (AGS, HGC-27, and BGC823) using recombinant lentivirus infection. The transfection efficiency is shown in **Figures 3A,B**.

Metastasis has been verified as a prototypical feature in tumor progression and a core hallmark of malignant behavior (Lambert et al., 2017). Therefore, we performed functional assays to fully characterize the effects of DDX27 during GC metastasis. Just as Tsukamoto et al. mentioned previously, DDX27 may contribute to GC invasiveness, yet they did not explore further (Tsukamoto et al., 2015). Our *in vitro* assay results confirmed their assumptions that overexpressed DDX27 could promote GC metastasis (**Figures 3C,D** left), while DDX27 inhibition substantially impaired GC cell motility (**Figures 3C,D** right and **Supplementary Figure S1C**). In *in vivo* lung metastatic models, when DDX27 was downregulated, the incidence of *in vivo* lung metastasis was significantly reduced (**Figures 3E**, **Supplementary Figure S1D**).

EMT is considered to be one of the most obvious traits in tumor metastasis, while N-cadherin upregulation and E-cadherin downregulation are typical indicators of EMT progress (Dongre and Weinberg 2019; Lambert and Weinberg 2021). Thus, we detected some typical EMT markers in DDX27 stable up- or downregulated cell lines and found that DDX27 could increase the protein levels of vimentin and N-cadherin and reduce ZO-1 and E-cadherin expressions in the GC cells (**Figure 2F**). To further illustrate the function of DDX27 in the EMT process, the Gene Set Enrichment Analysis (GSEA) was proceeded, and the results suggested that EMT was one of the most impaired pathways in the DDX27-deficient GC cells compared with the controls (**Supplementary Figure S1E**).

In this part, we determined that DDX27 could strengthen the metastatic capacity of the GC cells, and DDX27 may realize its promotive effects by triggering the EMT process in GC.

Lipoma-Preferred Partner is Indispensable for DEAD-Box Helicase 27–Mediated Gastric Cancer Migration and Invasion

To illustrate the underlying mechanisms in DDX27-mediated GC metastasis, we first performed mass spectrometry (MS) between DDX27 knockdown and control cells to seek the DDX27-related potential functional proteins. MS showed that LPP was remarkably changed when DDX27 was downregulated (**Figure 4A**; $p < 0.05$, $\log_2$ fold change <-1). Considering that the DDX family has regulative potentials in alternative splicing, we performed whole-transcriptome sequencing (RNA-seq) and alternative splicing analysis. A total of 18 genes, including LPP, were identified as candidate genes by the intersection analysis of data from MS and alternative splicing analysis (**Supplementary Figure S2A**). Interestingly, from MS and RNA-seq, we found that the LPP protein level was significantly decreased with tiny changes of its whole transcript level after DDX27 was inhibited, which clued us that there may exist a special regulative manner between DDX27 and LPP. Based on the comprehensive results of MS, transcriptome sequencing, and alternative splicing analysis, we finally chose LPP as the DDX27 downstream target for further studies.

LPP-encoded protein localizes at the cell periphery in focal adhesions, involving in cell adhesion, cell motility, and cell-substrate cytoskeletal interactions (Jin et al., 2007; Ngan et al., 2017). Although LPP has been reported as an oncogene in various tumor malignancies (Colas et al., 2012; Leung et al., 2018), its role in GC metastasis is not fully confirmed. To better understand the role of LPP in GC, we studied its expression in the gastric cell lines and TCGA dataset. We found not only the protein level of LPP in the GC cell lines (**Supplementary Figure S2B**) but also the mRNA level of LPP in the GC tissues, which showed a high tendency (**Supplementary Figure S2C**). Kaplan–Meier analysis showed that GC patients with a high LPP expression represented a poor OS by conducting the TCGA STAD cohort analysis (**Supplementary Figure S2D**).

Based on the previous results, we detected the LPP expression in overexpressed or silenced DDX27 cell models. We found that LPP was obviously increased in the DDX27-overexpressed cells and was significantly inhibited when DDX27 was downregulated

TABLE 3 | Multivariate analysis using the Cox proportional hazards model.

Characteristic	*n*	HR	95% CI	*p*
Tumor size				
≤6 cm	61	1.982423	1.143308–3.437395	0.014815
>6 cm	30			
Depth of invasion				
T 1–3	73	1.169362	0.619943–2.205699	0.628930
T 4	19			
Metastases to other organs				
M 0	85	1.648962	0.682601–3.983401	0.266385
M 1	8			
Number of positive lymph nodes				
≤5	54	2.456772	1.223911–4.931509	0.011462
>5	36			
AJCC				
1/2	39	0.796997	0.396643–1.601450	0.523923
3/4	53			
DDX27				
Low	50	2.561683	1.437016–4.566559	0.001427
High	43			

PS: bold fonts represent p < *0.05.*

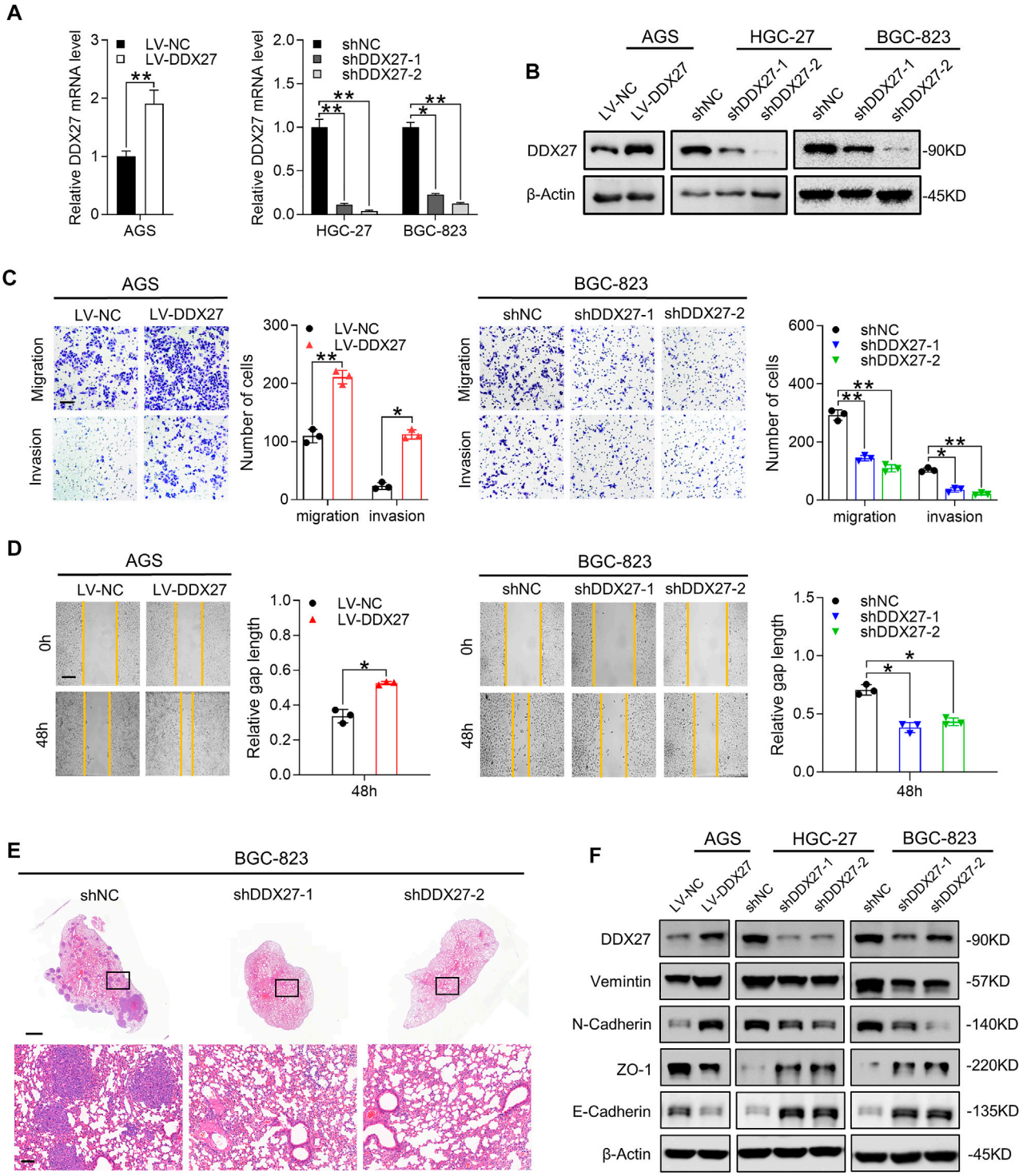


FIGURE 3 | DDX27 is a critical mediator of GC metastasis *via* EMT. **(A,B)** Transcript level **(A)** and protein expression **(B)** of DDX27 in DDX27-knockdown and overexpressed cell models (**$p < 0.01$, *$p < 0.05$, by ANOVA). **(C)** Metastatic ability assessed by trans-well assay for DDX27-overexpressed and -knockdown cell models (left, overexpression vs. control; right, knockdown vs. control). Scale bars: 12.5 μm (**$p < 0.01$, *$p < 0.05$, by ANOVA). **(D)** Wound healing and statistical analysis for indicated cells (left, overexpression vs. control; right, knockdown vs. control) Scale bars: 25 μm. (*$p < 0.05$, by ANOVA). **(E)** Hematoxylin-eosin **(H,E)** staining for metastatic nodules in the dissected lung specimen from nude mice (10 mice in each group). Scale bars:1,000 μm (up) or 100 μm (below). **(F)** Protein level of EMT markers (E-cadherin, N-cadherin, vimentin, and ZO-1) among the indicated GC cells. $n \geq 3$, the data are presented as mean ± SD. LPP is indispensable for DDX27-mediated GC migration and invasion.

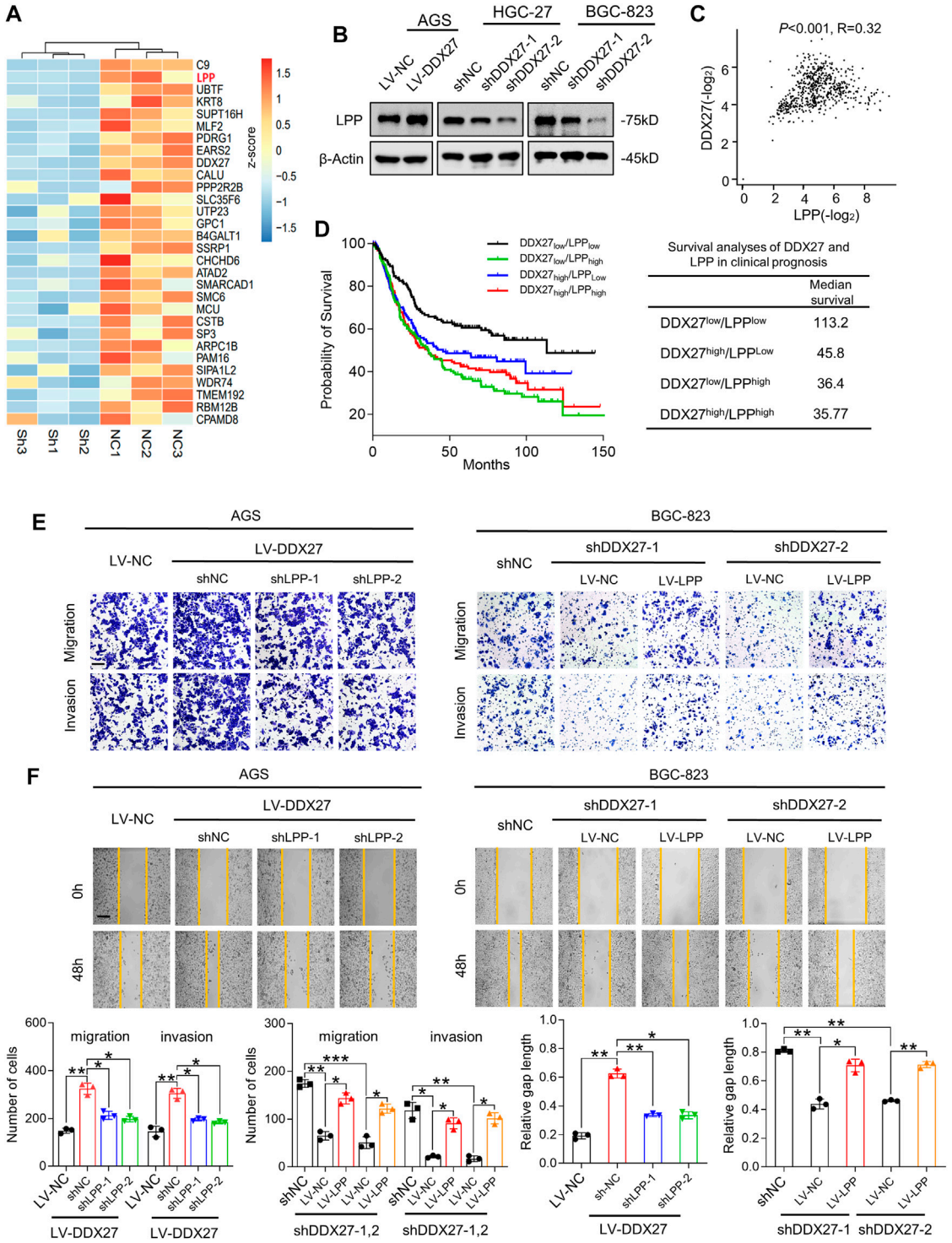


FIGURE 4 | LPP is indispensable for DDX27-mediated GC migration and invasion. **(A)** Heatmap generated from mass spectrometry analyses of the protein samples isolated from HGC-27 knockdown and control cells (log_2 fold change <−1; $p < 0.05$). LPP was identified as one of the significantly downregulated genes. **(B)** Western blot analysis of the LPP expression in DDX27 overexpressing and silencing GC cells. **(C)** Positive correlation between DDX27 and LPP was analyzed *via* TCGA and GTEx datasets of the stomach. ($p < 0.001$, $R = 0.32$, by Spearman correlation method). **(D)** Survival analysis of DDX27 and LPP in clinical prognosis ($n = 630$, left) and the median survival in different groups (right, by log-rank test). **(E)** Trans-well assays showing the migratory and invasive abilities between LPP knockdown (or overexpression) and corresponding control cells with the stable DDX27 overexpression (or knockdown). Scale bars: 12.5 μm, statistical analyses are exhibited below. (***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$, by ANOVA). **(F)** Wound healing was performed to detect cell motility between LPP knockdown (or overexpression) and corresponding control cells with a stable DDX27 overexpression (or knockdown). Scale bars: 25 μm, statistical analyses are exhibited as follows. (**$p < 0.01$ and *$p < 0.05$, by ANOVA), $n \geq 3$, the data are presented as the mean ± SD. LPP can be regulated by DDX27 through alternative splicing.

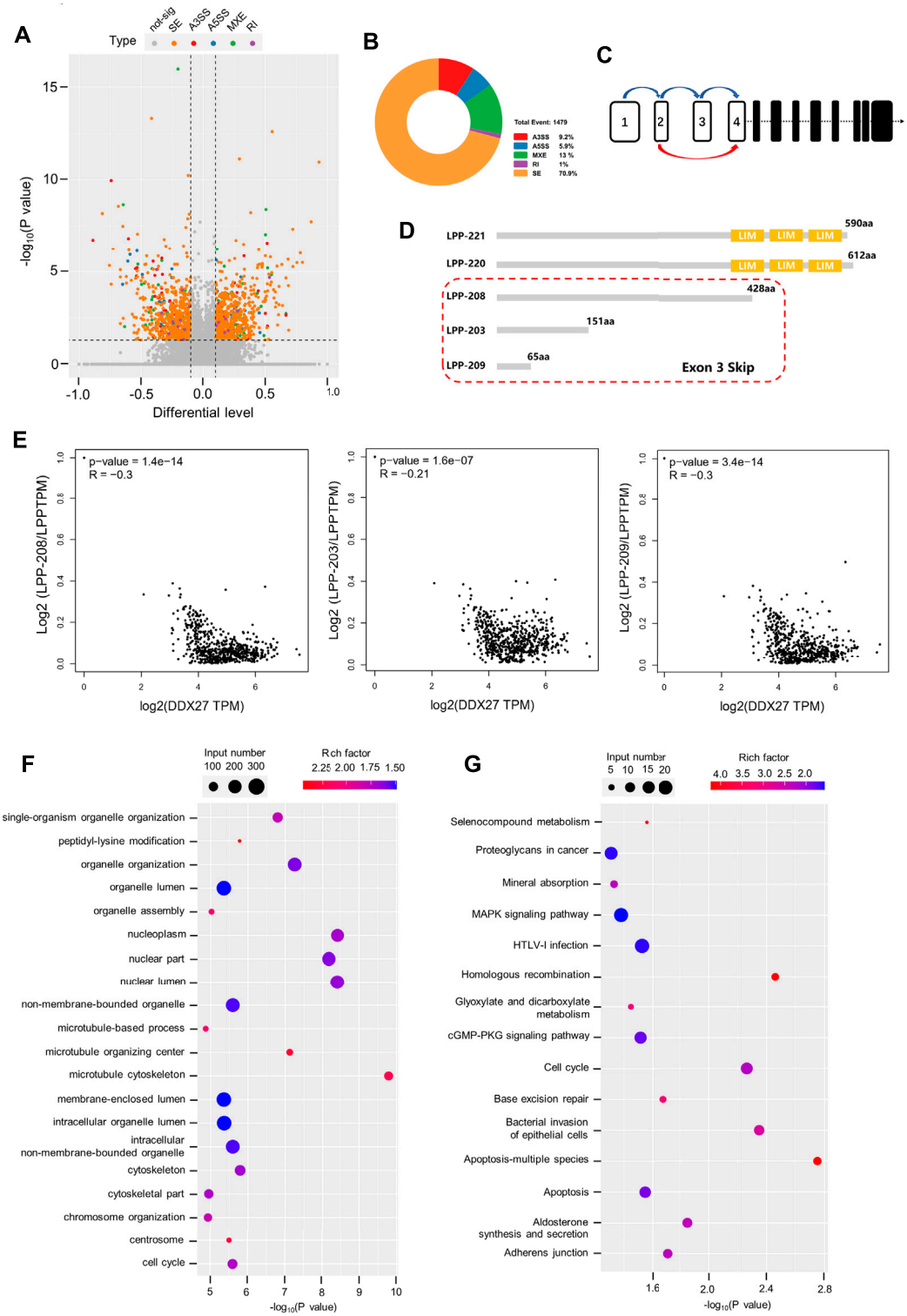


FIGURE 5 | LPP can be regulated by DDX27 through alternative splicing. **(A)** Volcano plot reflects the distribution of 1,169 different alternative splicing genes. (p < 0.05, difference of alternative splicing events $|\Delta\psi|$>0.1). **(B)** A total of 1,479 alternative splicing events were found in the alternative splicing analysis (p value < 0.05, $|\Delta\psi|$ >0.1), of which exon skipping accounted for a large proportion. **(C)** Increased exon skipping events in exon 3 of LPP in the gastric cancer cells with DDX27 knockdown. **(D)** Schematic diagram of the protein domains of several LPP transcript variants. LPP-221 and LPP-220 are the long transcript variants of LPP whose translated proteins have complete functional domains, while LPP-203, LPP-208, and LPP-209 are transcript variants of exon 3 skipped, which lacks the characteristic LIM domain. **(E)** Spearman correlation analysis results of the relative expression of LPP transcript variants and DDX27 expression. (Relative expression of LPP transcript variants was normalized with the total LPP). **(F,G)** GO/KEGG enrichment for different alternative splicing genes. The horizontal axis indicates the significance of the enrichment (expressed as –log10 (p value); the vertical axis indicates an enriched GO Terms/KEGG pathway (p value < 0.05).

(**Figure 4B**). A positive correlation between DDX27 and LPP was observed in the TCGA gastric cancer cohort (**Figure 4C**). Noticeably, after dividing the GC patients into four groups according to the median expression of DDX27 and LPP, we analyzed the effects of two markers on clinical survival and found that the low expressions of DDX27 and LPP could indicate a better prognosis, while either a high DDX27 or LPP could lead to a short overall survival (**Figure 4D**). In the rescue assays, through silencing LPP in the DDX27 overexpressed GC cells, we found that improved metastatic ability of the GC cells caused by the DDX27 overexpression was impaired after LPP inhibition. Meanwhile, when LPP was overexpressed in the DDX27-knockdown cells, the metastatic ability of cells inhibited by DDX27 impairment was recovered (**Figures 4E,F** and **Supplementary Figure S2F**). The lentivirus efficiency of LPP is shown in **Supplementary Figure S2E**. We found out that LPP may be a downstream functional target of DDX27.

In addition, the single-gene GSEA using the TCGA GC cohort identified epithelial mesenchymal transition (EMT) as the main hallmarks of the LPP high expression group (**Supplementary Figure S3**), and we discovered the DDX27 overexpression caused EMT activation which could be alleviated by LPP inhibition, while the LPP overexpression could partly recover DDX27-induced EMT impairment (**Supplementary Figure S2G**). The aforementioned evidence indicated that LPP took part in DDX27-promoted metastasis by modulating the EMT process in GC cells.

Lipoma Preferred Partner can be Regulated by DEAD-Box Helicase 27 Through Alternative Splicing

Accumulative evidence revealed that the DEAD box family can participate in the cancer malignant process by regulating the alternative splicing of downstream RNAs. For example, DDX17 is reported to promote HCC metastasis by regulating the alternative splicing of PXN-AS1 (Zhou et al., 2021). However, in CRC, DDX56 is identified as a novel oncogene and prognostic biomarker that promotes alternative splicing of WEE1 (Kouyama et al., 2019). In this study, the Volcano plot data from alternative splicing analysis exhibited distribution of different alternative splicing genes in the DDX27-knockdown GC cells (**Figure 5A**) and showed that skipping exon (SE) was the predominant alternative splicing event after DDX27 knockdown (**Figure 5B**). Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments were performed, and the results exhibited that different splicing genes were significantly enriched in the cytoskeleton and adhesive junctions which hinted us that DDX27 may regulate cell mobility *via* alternative splicing (**Figures 5F,G**).

Notably, we noticed that SE events of the third exon (chr3: 188225404–188225527) of the LPP full-length transcript were increased in the DDX27-knockdown cell lines (**Figure 5C**). LPP belongs to the LIM domain protein subfamily, characterized by a N-terminal proline-rich region and three C-terminal LIM domains (Ngan et al., 2018). After searching the ensemble database, we found that LPP has a variety of transcript variants, and none of these transcript variants with the deletion of the third exon can be translated into a relatively complete LPP protein structure containing three LIM domains, which is indispensable for the LPP protein binding to the cytoskeleton and adhesion (**Figure 5D**). As exhibited in **Figure 5E**, we found that a high expression of DDX27 was negatively correlated with the relative expression of LPP transcripts with the third exon depletion (LPP-203, LPP-208, and LPP-209) from the TCGA gastric cancer cohorts.

Taking together, we noticed that DDX27 may regulate the protein expression of LPP by reducing the skipping exon (SE) events on the third exon of LPP full-length transcripts, thereby enhancing the translation of LPP proteins with functional domains and promoting the cell motility and metastatic ability of the gastric cancer cells.

DISCUSSION

The DEAD box family has been suggested to act as multifunctional roles in carcinogenesis processes (Heerma van Voss et al., 2015; Heerma van Voss et al., 2017; Hashemi et al., 2019). DDX27 has also been found aberrantly overexpressed in various tumors, while its function in GC metastasis is still unknown. Here, we investigated the effects and clinical significance of DDX27 in gastric cancer and proved that DDX27 could promote the migratory and invasive ability of GC through the LPP-mediated EMT process.

For the role of DDX27 in GC, a recent study has suggested that DDX27 could regulate cell proliferation in the GC cells through the inhibition of cell cycle progression independent of apoptosis (Tsukamoto et al., 2015). Zhou et al. found that DDX27 knockdown sensitized the gastric cancer cells to epirubicin or cisplatin, and conversely its overexpression reduced the eirenicon-induced DNA damage and apoptosis (Zhou et al., 2015). They mainly focused on studying the effects of DDX27 in GC proliferation and chemotherapy. This is the first time we verified that DDX27 could also regulate GC migration and invasion. We found that in the human GC samples, DDX27 maintained a significantly upregulated expression at the mRNA and protein levels, which could also reinforce GC metastasis *in vitro* and *in vivo*. The highly expressed DDX27 indicated GC patients' poor prognosis.

Mass spectrometry and RNA-seq analysis were conducted to screen out the DDX27 downstream candidates and found that LPP, which belongs to the zyxin family of LIM proteins, might interact with DDX27 in GC. LPP has been confirmed as a regulator of mesenchymal/fibroblast cell motility (Ngan et al., 2018). Pathologically, LPP is commonly emerged as a critical tumor inducer, accounting for tumor initiation, metastasis, and drug resistance (Ngan et al., 2013; Kuriyama et al., 2016). LPP acted as a tumor metastatic accelerator by virtue of being located inside the adhesion and promoting invadopodia formation (Kiepas et al., 2020). However, the role of LPP in GC is still not clear , so we studied the role of LPP in this study and determined that LPP was a putative target of DDX27 in GC cells.

Members of the DEAD box family of RNA helicases have been found to be involved in affecting RNA maturation, RNP

assembly, and RNAs' ultimate destiny (Linder and Jankowsky 2011). We supposed that DDX27 can also influence the alternative splicing process of LPP by influencing the RNA structure. We demonstrated our assumption with RNA-seq and alternative splicing analysis and discovered that exon 3 (chr3:188225404–188225527) skipped events of LPP transcript that happened in DDX27-knockdown cell lines. In functional rescue experiments, DDX27-conferred metastasis advantage to GC cells was abolished by LPP inhibition.

Current studies demonstrate that both DDX27 and LPP are involved in the tumor EMT process. For example, DDX27 contributes to CRC development by driving the EMT process (Tang et al., 2018). The formation of the FOXD1/CYTOR/LPP axis has been proven to be indispensable to induce EMT in oral squamous cell cancer (OSCC) (Chen et al., 2021). Here, we similarly noticed that DDX27 can regulate the EMT process in GC cells, and its effects were abrogated if LPP was knocked down, indicating that LPP is necessary in the DDX27-triggered EMT process. In addition, activation of an epithelial-mesenchymal transition (EMT) has been linked to the formation of neoplastic stem cells (Lambert and Weinberg 2021), and DDX27 has been noticed to also regulate the cancer stem cell (CSC) activity in CRC (Yang et al., 2019). There may exist a complicated interaction among DDX27, EMT, and CSCs waiting to be illuminated.

The high level of DDX27 in GC, together with its remarkable prometastasis effects, prompted that DDX27 performed a vital role in accelerating GC metastasis and might be a novel prognostic risk factor for GC patients. Our study has limitations that the detailed regulatory mechanisms of how DDX27 acts to influence the alternative splicing of LPP need more evidence to prove, and it is also the focus of our later research. In summary, we unveiled that the protumorigenic function of DDX27 was mediated through a cancer-regulating functional axis DDX27/LPP/EMT in gastric cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Clinical Research Ethics Committee of the Air Force Military Medical University and Xijing Hospital. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the Institutional Animal Care and Use Committee of the Air Force Military Medical University.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This research was funded by the National Natural Science Foundation of China, grant numbers: 81772650, 81421003, and 81572302; the Independent Funds of the Key Laboratory, grant number: CBSKL 2015Z12; the Airforce Aoxiang Foundation and Young Changjiang Scholars of the Ministry of Education (JL); the Central Government Guide the Development of Local Science and Technology Special Funds, grant number: 2016ZY-HM-01; and the Shaanxi Provincial Science and Technology Plan Project, grant number: 2016KTZDSF02-02.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.836199/full#supplementary-material

REFERENCES

Ameur, L. B., Marie, P., Thenoz, M., Giraud, G., Combe, E., Claude, J.-B., et al. (2020). Intragenic Recruitment of NF-Kb Drives Splicing Modifications upon Activation by the Oncogene Tax of HTLV-1. *Nat. Commun.* 11, 3045. doi:10.1038/s41467-020-16853-x

Chen, S., Yang, M., Wang, C., Ouyang, Y., Chen, X., Bai, J., et al. (2021). Forkhead Box D1 Promotes EMT and Chemoresistance by Upregulating lncRNA CYTOR in Oral Squamous Cell Carcinoma. *Cancer Lett.* 503, 43–53. doi:10.1016/j.canlet.2020.11.046

Colas, E., Muinelo-Romay, L., Alonso-Alconada, L., Llaurado, M., Monge, M., Barbazan, J., et al. (2012). ETV5 Cooperates with LPP as a Sensor of Extracellular Signals and Promotes EMT in Endometrial Carcinomas. *Oncogene* 31, 4778–4788. doi:10.1038/onc.2011.632

Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Espinoza, M. P., et al. (2012). Splicing Switch of an Epigenetic Regulator by RNA Helicases Promotes Tumor-Cell Invasiveness. *Nat. Struct. Mol. Biol.* 19, 1139–1146. doi:10.1038/nsmb.2390

Dongre, A., and Weinberg, R. A. (2019). New Insights into the Mechanisms of Epithelial-Mesenchymal Transition and Implications for Cancer. *Nat. Rev. Mol. Cel Biol* 20, 69–84. doi:10.1038/s41580-018-0080-4

Germann, S., Gratadou, L., Zonta, E., Dardenne, E., Gaudineau, B., Fougère, M., et al. (2012). Dual Role of the Ddx5/ddx17 RNA Helicases in the Control of the Pro-migratory NFAT5 Transcription Factor. *Oncogene* 31, 4536–4549. doi:10.1038/onc.2011.618

Grunewald, T. G. P., Pasedag, S. M., and Butt, E. (2009). Cell Adhesion and Transcriptional Activity - Defining the Role of the Novel Protooncogene LPP. *Translational Oncol.* 2, 107–116. doi:10.1593/tlo.09112

Hashemi, V., Masjedi, A., Hazhir-karzar, B., Tanomand, A., Shotorbani, S. S., Hojjat-Farsangi, M., et al. (2019). The Role of DEAD-box RNA Helicase P68 (DDX5) in the Development and Treatment of Breast Cancer. *J. Cel Physiol* 234, 5478–5487. doi:10.1002/jcp.26912

Hayakawa, Y., Nakagawa, H., Rustgi, A. K., Que, J., and Wang, T. C. (2021). Stem Cells and Origins of Cancer in the Upper Gastrointestinal Tract. *Cell Stem Cell* 28, 1343–1361. doi:10.1016/j.stem.2021.05.012

Heerma van Voss, M. R., van Diest, P. J., and Raman, V. (2017). Targeting RNA Helicases in Cancer: The Translation Trap. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1868, 510–520. doi:10.1016/j.bbcan.2017.09.006

Heerma van Voss, M. R. H., Vesuna, F., Trumpi, K., Brilliant, J., Berlinicke, C., de Leng, W., et al. (2015). Identification of the DEAD Box RNA Helicase DDX3 as a Therapeutic Target in Colorectal Cancer. *Oncotarget* 6, 28312–28326. doi:10.18632/oncotarget.4873

Hönig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W., and Berget, S. M. (2002). Regulation of Alternative Splicing by the ATP-dependent DEAD-Box RNA Helicase P72. *Mol. Cel Biol* 22, 5698–5707. doi:10.1128/mcb.22.16.5698-5707.2002

Jin, L., Kern, M. J., Otey, C. A., Wamhoff, B. R., and Somlyo, A. V. (2007). Angiotensin II, Focal Adhesion Kinase, and PRX1 Enhance Smooth Muscle Expression of Lipoma Preferred Partner and its Newly Identified Binding Partner Palladin to Promote Cell Migration. *Circ. Res.* 100, 817–825. doi:10.1161/01.RES.0000261351.54147.de

Kahles, A., Lehmann, K. V., Toussaint, N. C., Hüser, M., Stark, S. G., Sachsenberg, T., et al. (2018). Comprehensive Analysis of Alternative Splicing across Tumors from 8,705 Patients. *Cancer Cell* 34, 211–e6. e216. doi:10.1016/j.ccell.2018.07.001

Kiepas, A., Voorand, E., Senecal, J., Ahn, R., Annis, M. G., Jacquet, K., et al. (2020). The SHCA Adapter Protein Cooperates with Lipoma-Preferred Partner in the Regulation of Adhesion Dynamics and Invadopodia Formation. *J. Biol. Chem.* 295, 10535–10559. doi:10.1074/jbc.RA119.011903

Kouyama, Y., Masuda, T., Fujii, A., Ogawa, Y., Sato, K., Tobo, T., et al. (2019). Oncogenic Splicing Abnormalities Induced by DEAD -Box Helicase 56 Amplification in Colorectal Cancer. *Cancer Sci.* 110, 3132–3144. doi:10.1111/cas.14163

Kuriyama, S., Yoshida, M., Yano, S., Aiba, N., Kohno, T., Minamiya, Y., et al. (2016). LPP Inhibits Collective Cell Migration during Lung Cancer Dissemination. *Oncogene* 35, 952–964. doi:10.1038/onc.2015.155

Lambert, A. W., Pattabiraman, D. R., and Weinberg, R. A. (2017). Emerging Biological Principles of Metastasis. *Cell* 168, 670–691. doi:10.1016/j.cell.2016.11.037

Lambert, A. W., and Weinberg, R. A. (2021). Linking EMT Programmes to normal and Neoplastic Epithelial Stem Cells. *Nat. Rev. Cancer* 21, 325–338. doi:10.1038/s41568-021-00332-6

Leung, C. S., Yeung, T.-L., Yip, K.-P., Wong, K.-K., Ho, S. Y., Mangala, L. S., et al. (2017). Cancer-associated Fibroblasts Regulate Endothelial Adhesion Protein LPP to Promote Ovarian Cancer Chemoresistance. *J. Clin. Invest.* 128, 589–606. doi:10.1172/jci95200

Linder, P., and Jankowsky, E. (2011). From Unwinding to Clamping - the DEAD Box RNA Helicase Family. *Nat. Rev. Mol. Cel Biol* 12, 505–516. doi:10.1038/nrm3154

Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., and Stillman, B. (2012). DDX5 Regulates DNA Replication and Is Required for Cell Proliferation in a Subset of Breast Cancer Cells. *Cancer Discov.* 2, 812–825. doi:10.1158/2159-8290.Cd-12-0116

Mo, J., Liang, H., Su, C., Li, P., Chen, J., and Zhang, B. (2021). DDX3X: Structure, Physiologic Functions and Cancer. *Mol. Cancer* 20, 38. doi:10.1186/s12943-021-01325-7

Ngan, E., Kiepas, A., Brown, C. M., and Siegel, P. M. (2018). Emerging Roles for LPP in Metastatic Cancer Progression. *J. Cel Commun. Signal.* 12, 143–156. doi:10.1007/s12079-017-0415-5

Ngan, E., Northey, J. J., Brown, C. M., Ursini-Siegel, J., and Siegel, P. M. (2013). A Complex Containing LPP and α-Actinin Mediates TGFβ-Induced Migration and Invasion of ErbB2-Expressing Breast Cancer Cells. *J. Cel Sci* 126, 1981–1991. doi:10.1242/jcs.118315

Ngan, E., Stoletov, K., Smith, H. W., Common, J., Muller, W. J., Lewis, J. D., et al. (2017). LPP Is a Src Substrate Required for Invadopodia Formation and Efficient Breast Cancer Lung Metastasis. *Nat. Commun.* 8, 15059. doi:10.1038/ncomms15059

Petit, M. M. R., Mols, R., Schoenmakers, E. F. P. M., Mandahl, N., and Van de Ven, W. J. M. (1996). LPP,the Preferred Fusion Partner Gene ofHMGICin Lipomas, Is a Novel Member of the LIM Protein Gene Family. *Genomics* 36, 118–129. doi:10.1006/geno.1996.0432

Rahman, M. A., Krainer, A. R., and Abdel-Wahab, O. (2020). SnapShot: Splicing Alterations in Cancer. *Cell* 180, 208. e201. doi:10.1016/j.cell.2019.12.011

Soerjomataram, I., and Bray, F. (2021). Planning for Tomorrow: Global Cancer Incidence and the Role of Prevention 2020-2070. *Nat. Rev. Clin. Oncol.* 18, 663–672. doi:10.1038/s41571-021-00514-z

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tan, P., and Yeoh, K.-G. (2015). Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* 149, 1153–1162. e1153. doi:10.1053/j.gastro.2015.05.059

Tang, J., Chen, H., Wong, C.-C., Liu, D., Li, T., Wang, X., et al. (2018). DEAD-box Helicase 27 Promotes Colorectal Cancer Growth and Metastasis and Predicts Poor Survival in CRC Patients. *Oncogene* 37, 3006–3021. doi:10.1038/s41388-018-0196-1

Tsukamoto, Y., Fumoto, S., Noguchi, T., Yanagihara, K., Hirashita, Y., Nakada, C., et al. (2015). Expression of DDX27 Contributes to colony-forming Ability of Gastric Cancer Cells and Correlates with Poor Prognosis in Gastric Cancer. *Am. J. Cancer Res.* 5, 2998–3014.

Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., et al. (2001). Helicobacter pyloriInfection and the Development of Gastric Cancer. *N. Engl. J. Med.* 345, 784–789. doi:10.1056/NEJMoa001999

Wang, D., Zhu, Z.-Z., Jiang, H., Zhu, J., Cong, W.-M., Wen, B.-J., et al. (2015). Multiple Genes Identified as Targets for 20q13.12-13.33 Gain Contributing to Unfavorable Clinical Outcomes in Patients with Hepatocellular Carcinoma. *Hepatol. Int.* 9, 438–446. doi:10.1007/s12072-015-9642-0

Yamanaka, M., Anada, Y., Igarashi, Y., and Kihara, A. (2008). A Splicing Isoform of LPP1, LPP1a, Exhibits High Phosphatase Activity toward FTY720 Phosphate. *Biochem. Biophysical Res. Commun.* 375, 675–679. doi:10.1016/j.bbrc.2008.07.165

Yang, C., Li, D., Bai, Y., Song, S., Yan, P., Wu, R., et al. (2019). DEAD-box Helicase 27 Plays a Tumor-Promoter Role by Regulating the Stem Cell-like Activity of Human Colorectal Cancer Cells. *Ott* Vol. 12, 233–241. doi:10.2147/ott.S190814

Zhang, H., Zhang, Y., Zhu, X., Chen, C., Zhang, C., Xia, Y., et al. (2019). DEAD Box Protein 5 Inhibits Liver Tumorigenesis by Stimulating Autophagy *via* Interaction with p62/SQSTM1. *Hepatology* 69, 1046–1063. doi:10.1002/hep.30300

Zhou, H. Z., Li, F., Cheng, S. T., Xu, Y., Deng, H. J., Gu, D. Y., et al. (2021). DDX17-regulated Alternative Splicing that Produced an Oncogenic Isoform of PXN-AS1 to Promote HCC Metastasis. *Hepatology* 75, 847–865. doi:10.1002/hep.32195

Zhou, J., Yong, W.-P., Yap, C. S., Vijayaraghavan, A., Sinha, R. A., Singh, B. K., et al. (2015). An Integrative Approach Identified Genes Associated with Drug Response in Gastric Cancer. *Carcinogenesis* 36, 441–451. doi:10.1093/carcin/bgv014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Jin, Yang, Gao, Chen, Luo, Su, Shi, Yang, Dong and Liang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 05 May 2022
doi: 10.3389/fgene.2022.848391

Transmembrane Protein 170B is a Prognostic Biomarker and Associated With Immune Infiltrates in Pancreatic Adenocarcinoma

Zilong Zhang[1,2,3†], *Jin Shang*[1,2,3†], *Zonglin Dai*[1,2,3], *Yutong Yao*[1,2,3], *Ying Shi*[3], *Deyuan Zhong*[1,2,3], *Yuxin Liang*[1,2,3], *Chunyou Lai*[1,2], *Qinyan Yang*[1,2], *Tianhang Feng*[1,2,3] *and Xiaolun Huang*[1,2,3]*

[1]*Department of Hepatobiliary-Pancreatic Surgery, Cell Transplantation Center, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China,* [2]*Chinese Academy of Sciences Sichuan Translational Medicine Research Hospital, Chengdu, China,* [3]*School of Medicine, University of Electronic Science and Technology of China, Chengdu, China*

OPEN ACCESS

Edited by:
Jian-Guo Zhou, University of Erlangen Nuremberg, Germany

Reviewed by:
Changjing Cai, Central South University, China
Robert Torphy, University of Colorado Denver, United States

***Correspondence:**
Xiaolun Huang
huangxiaolun@med.uestc.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *04 January 2022*
Accepted: *07 April 2022*
Published: *05 May 2022*

Citation:
Zhang Z, Shang J, Dai Z, Yao Y, Shi Y, Zhong D, Liang Y, Lai C, Yang Q, Feng T and Huang X (2022) Transmembrane Protein 170B is a Prognostic Biomarker and Associated With Immune Infiltrates in Pancreatic Adenocarcinoma. Front. Genet. 13:848391. doi: 10.3389/fgene.2022.848391

Background: Pancreatic adenocarcinoma (PAAD) is among the most common types of cancer with a poor prognosis. Transmembrane protein 170B (TMEM170B) has been reported to suppress breast cancer proliferation, metastasis, and tumorigenesis and is related to prognosis. However, its role in PAAD and the underlying molecular mechanisms are yet to be investigated.

Patients and methods: We performed a comprehensive analysis of RNA sequencing data obtained from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases to determine TMEM170B expression. Immunostaining and real-time polymerase chain reaction (RT-PCR) were done to determine TMEM170B expression in human pancreatic cancer cell lines and tissue specimens. Furthermore, the correlation of TMEM170B with clinicopathological features and PAAD prognosis was investigated, and the mechanisms were explored through enrichment analysis and immune cell infiltration analysis.

Results: TCGA and GEO dataset analysis revealed that TMEM170B expression in PAAD tissue samples was significantly lower than that in non-tumorous tissues, which was further confirmed by immunohistochemistry and RT-PCR. Low TMEM170B expression was associated with poor differentiation ($p = 0.014$). Multivariate analysis identified that TMEM170B is an independent indicator for overall survival [hazard ratio (HR) = 0.116, 95% confidence interval (CI) = 0.014–0.995; $p = 0.049$] and disease-free survival (HR = 0.19, 95% CI = 0.04–0.910; $p = 0.038$) in patients with PAAD. Additionally, TMEM170B was involved in immune-related gene sets, including those related to chemokine signaling pathways and innate and adaptive immunity. High TMEM170B expression was linked to antitumor immune microenvironment with a high infiltration of B cells, T cells, dendritic cells, monocytes, M1 macrophages, neutrophil, and natural killer cells and a low infiltration of Tregs and myeloid-derived suppressor cells (all $p < 0.05$).

Plain Language Summary: There is an urgent need to identify clinical prognostic biomarkers and targeted drugs for pancreatic cancer treatment. In this study, the expression status and

prognostic value of transmembrane protein 170B (TMEM170B) in pancreatic adenocarcinoma were elucidated. Furthermore, TMEM170B, as a tumor suppressor gene, induced antitumor immune effects, including increased tumor infiltration of immune effector cells and reduced levels of inhibitory immune molecules and regulatory cells. Therefore, TMEM170B could be regarded as a novel target in preventing the progression of pancreatic cancer.

Conclusion: The findings suggest that low TMEM170B expression is remarkably correlated with poor PAAD prognosis, which might provide a therapeutic target for PAAD.

Keywords: TMEM170B, prognosis, pancreatic cancer, biomarker, immune infiltration

1 INTRODUCTION

Pancreatic cancer is a common malignant tumor of the digestive tract and the seventh leading cause of cancer-related death worldwide in 2020, with approximately 4,96,000 newly diagnosed cases and 4,66,000 deaths Sung et al. (2021). Owing to its inapparent symptoms and rapid progression, 80% of patients with pancreatic cancer fail to meet the criteria for radical resection Connor and Gallinger, (2021). Furthermore, the high rates of late recurrence and metastasis and numerous postoperative complications also result in unsatisfying surgical outcomes and poor prognosis. For the reasons mentioned above, pancreatic adenocarcinoma (PAAD) has one of the highest death rates of any solid organ malignancy, with overall 5-years survival of less than 8% Connor and Gallinger, (2021); Di Marco et al. (2016); Siegel et al. (2018). In recent years, significant improvements in molecular diagnosis and targeted biological therapy have facilitated the discovery of novel genetic biomarkers and drug targets. Such advancements have helped in elucidating the molecular mechanisms of pancreatic cancer and have gradually becoming important means for improving the overall prognosis of the patients Cassetta and Kitamura, (2018); Mace et al. (2018); O'Reilly et al. (2019); Waddell et al. (2015). Additionally, tumor-related genes have continuously been found to be associated with the incidence and development of pancreatic cancer Duan et al. (2013); Shi and Cao, (2014); Wang et al. (2018); Yang et al. (2016); Yi et al. (2015).

TMEM170B, a transmembrane protein and a member of the TMEM170 family, is composed of 132 amino acids, with sequences that are highly conserved from invertebrates to mammals. Its important paralog is TMEM170A, which has been reported to be a new regulator of endoplasmic reticulum (ER) and nuclear envelope morphogenesis Christodoulou et al. (2016). A transcriptome-wide association study revealed that TMEM170A expression is significantly associated with the risk of pancreatic cancer Zhong et al. (2020). Moreover, a recent study suggested that low TMEM170B expression is an independent predictor of poor overall survival (OS) in patients with breast carcinoma (BRCA) ($p < 0.01$) and that TMEM170B could act as a tumor suppressor to antagonize the protumorigenic Wnt/β-catenin signaling pathway Li et al. (2018). Therefore, TMEM170B is a potentially attractive therapeutic target for future anticancer drugs although its role in pancreatic cancer remains unknown.

Hence, the present study aimed to investigate the role of TMEM170B as a prognostic biomarker and therapeutic target in pancreatic cancer and explore the possible underlying mechanisms of TMEM170B in the development of pancreatic cancer.

2 MATERIAL AND METHODS

2.1 Patients and Histological Specimens

After surgical resection at the Sichuan Provincial People's Hospital between January 2018 and August 2020, 38 fresh paired tissue and adjacent nontumor tissue samples were obtained and divided into two groups for preservation in 10% formalin for histological examination and immediate freezing in liquid nitrogen and storage at −80°C until further use for qRT-PCR. All patients with pancreatic adenocarcinoma (PAAD) were histologically and clinically diagnosed and assessed according to the TNM staging system (7th version) of the International Union against Cancer. Follow-up data for the patients were collected until 31 October 2021 or their date of death. Demographic, clinical, and blood parameters of all the patients were collected on the day of admission. Clinicopathological data for the patients with pancreatic ductal adenocarcinoma are presented in **Table 1**. OS was defined as the interval from the postoperative period until death or the last follow-up date. Disease-free survival (DFS) was calculated from post-operation until disease progression, death, or last follow-up.

2.2 Data Source and Processing

To assess TMEM170B mRNA levels in PAAD tissues and normal tissues, the online tools employed were provided by the Sangerbox (http://www.sangerbox.com/tool), based on The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/) database, and by the Gene Expression database of Normal and Tumor tissues 2 (GENT2, http://gent2.appex.kr), based on the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database. Gene presentation datasets (GSE32676, GSE16515, and GSE71729) were retrieved from the GEO database. Differentially expressed genes (DEGs) were screened with the adjusted p-value of<0.01 and |log fold change| of ≥2.

2.3 Cell Culture

Human pancreatic cancer cell lines PANC-1, AsPC-1, and BxPC-3 and the human pancreatic duct cell line HPDE6-C7 were obtained from the American Tissue Culture Collection

(Manassas, VA, United States). All cells were cultured in Roswell Park Memorial Institute (RPMI)-1,640 medium containing 10% fetal bovine serum, followed by incubation at 37°C in 5% CO_2. After a few passages (2–4), the cells in the logarithmic growth phase were used for further experiments.

2.4 RNA Extraction and Quantitative Real-Time Polymerase Chain Reaction (RT-PCR)

Total RNA from fresh frozen tissues and cells was extracted using TRIzol reagent (Invitrogen, NY, United States) according to the manufacturer's instructions. RNA was quantified using a NanoDrop (Thermo Fisher Scientific, MA, United States). Subsequently, 1 μg of total RNA was used to produce cDNA using PrimeScript RT reagent (Takara, Kusatsu, and Japan), and qRT-PCR was performed using TB Green™ Premix Ex Taq™ II (Takara, Japan) in the CFX96 Real-time System (Bio-Rad). β-Actin was used for normalization purposes. Primers were designed based on the TMEM170B and β-Actin mRNA sequences in GenBank. The primers used were: TMEM170B, forward primer: 5′-TTCCTCTGGGCTCTCTTC TCT-3′, reverse primer: 5′-CTGCTGCACTGGTAATCATCG-3′, β-Actin, forward primer: 5′-CCTGAAGTACCCCATCGAGC-3′, reverse primer: 5′-AGGGATAGCACAGCCTGGAT-3′. The relative expression levels of TMEM170B and β-actin were calculated using the comparative CT ($2^{-\Delta\Delta CT}$) method. Each sample was analyzed in triplicate.

2.5 Immunohistochemistry

Tissue samples were fixed in 10% formalin for at least 24 h, embedded in paraffin, and sections of 3-μm thickness were obtained. The tissue sections were then placed in Tris-ethylenediaminetetraacetic acid (pH 9.0) antigen repair solution and repaired in a pressure cooker for 5 min. Endogenous peroxidase activity was blocked with 3% H_2O_2 solution for 25 min at room temperature in a dark condition. Thereafter, the sections were incubated with 10% goat serum (HyClone, United States) in phosphate buffered saline (PBS) at room temperature for 1 h. The sections were then incubated overnight at 4°C with anti-TMEM170B antibody (1:100 dilution, PA5-63072, Thermo Fisher Scientific). After washing off the antibodies, the sections were incubated with rabbit secondary antibodies at 37°C.

The histology of the different tissues was analyzed microscopically (BX51, Olympus, Tokyo, Japan). Image capturing was performed using DP2-BSW software (Olympus). The immunohistochemistry (IHC) scores were obtained using a semi-quantitative method by comparing the staining intensity with the proportion of the stained cells. The H score system was employed for scoring the positive cell distribution as follows: the recording of the distribution (positive cell percentage) was determined using A = 0–100; the staining intensity was evaluated as B and scored as follows: strong (dark brown): 3; mild (brown): 2; light (light brown): 1; and none (not stained): 0. The following formula was used to determine the H scores in terms of intensity and distribution: H score = 1 × A1 +2 × A2 +3 × A3, $0 \leq H \leq 300$.

2.6 Immunofluorescence

The tissue sections and cells on glass coverslips were fixed with 4% paraformaldehyde for 20 min and permeabilized with 0.1% Triton X-100 for 15 min at room temperature. After blocking with 5% BSA for 60 min at room temperature, the sections and cells were incubated overnight at 4°C with anti-TMEM170B (1: 200 dilution), anti-$CD4^+$(1:200 dilution, Proteintech)/anti-$CD8^+$(1:200 dilution, Proteintech)/anti-CD11b (1:100 dilution, Proteintech)/anti-CD33 (1:25 dilution, Proteintech). After three washes with PBS, the sections and cells were stained with fluorescent labeled secondary antibodies for 1 h at 37°C. The coverslips were stained with 4′,6-diamidino-2-phenylindole (Sigma) and examined under a confocal microscope (LSM 800, Zeiss) with a 63 × /1.40 oil-immersion objective lens.

2.7 Correlation and Gene Enrichment Analysis

Correlation analysis between TMEM170B and other mRNAs in PAAD was performed using TCGA data, and the Pearson correlation coefficient was calculated. The top 300 genes that were most positively associated with TMEM170B were selected for enrichment analysis to reflect the role of TMEM170B. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of genes were performed using the functional annotation tool in the Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov) and Gene Set Enrichment Analysis (GSEA) Reimand et al. (2019). The Reactome Knowledgebase provides both as an archive of biological processes and as a tool for discovering functional relationships in data Jassal et al. (2020). GSEA was performed using the gseGO, gseKEGG, and gsePathway functions of the R package "clusterProfiler."

2.8 Immune Cells Infiltration Analysis

The level of immune and stromal fraction was scored by Estimation of Stromal Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) based on log2-transformed TPM data. Tumor IMmune Estimation Resource 2.0 (TIMER 2.0, http://timer.cistrome.org/) algorithm database provided immune cell infiltration expression profiles by TIMER, CIBERSORT Chen et al. (2018), Microenvironment Cell Populations-counter (MCP-counter) Becht et al. (2016), EPIC Racle et al. (2017), these algorithms were used to determine the Spearman correlation between TMEM170B expression levels in PAAD and the infiltration of different immune cells. The relationship between immune cell infiltration and prognosis of patients was further analyzed. p values <0.05 were considered to denote statistical significance.

2.9 Statistical Analysis

The clinical characteristics of the patients were compared using the Fisher's exact test for categorical variables and Wilcoxon rank-sum test for continuous variables. The continuous and categorical variables were expressed as mean (standard deviation) and proportions (percentages), respectively. Multivariate time-dependent Cox proportional DFS and OS hazard ratios (HRs) were fitted based on significant univariate factors. Survival analyses were performed using the Kaplan-Meier method and log rank test. The IHC results were quantitatively analyzed using ImageJ software (MEDIA CYBERNETICS, United States). All p values were two-sided tests, with $p < 0.05$ indicating statistical significance. All data analyses were performed

using SPSS version 27.0 (IBM Corp., NY, United States) and R 3.6.0 (The R Foundation for Statistical Computing, Vienna, Austria).

3 RESULTS

3.1 Downregulation of TMEM170B Expression in Human Pancreatic Cancer

GENT2 tool was utilized to examine the mRNA levels of TMEM170B in pan-cancer tissues. We first determined that TMEM170B was significantly downregulated in conditions such as PAAD, breast cancer, oral cancer, ovary cancer, and thyrioid cancer (**Figure 1A**). In pancreatic cancer data set GSE32676 from the GEO, the top 24 ranked DEGs were selected to form a heat map (**Figure 1B**). All DEGs constituted the volcano plot of the differential genes (**Figure 1C**). Compared with normal tissues, TMEM170B expression was frequently downregulated in the PAAD tissues from GSE32676 ($p < 0.001$) (**Figure 1D**). In addition, consistent with the above results, TMEM170B was obviously low expressed in PAAD tissues from TCGA ($p = 0.023$), GSE16515 ($p = 0.005$), and GSE32676 ($p < 0.001$) datasets (**Figure 1E**).

To determine the low expression of TMEM170B in human pancreatic cancer cell lines and tissues, the mRNA and protein levels of TMEM170B were assayed using RT-PCR and IHC, respectively. Our results revealed that TMEM170B expression was considerably lower in pancreatic cancer cell lines and tissues than in normal pancreatic duct epithelial cells and normal pancreatic tissues, respectively (**Figures 2A–E**). Moreover, IHC and IF demonstrated that TMEM170B was localized in the plasma membrane and cytoplasmic regions of pancreatic cancer cells (**Figures 2C,E**) and that 63.2% of the pancreatic cancer tissue samples were TMEM170B positive (**Figure 2D**).

The above results confirmed that TMEM170B is expressed at lower levels in PAAD when compared with its levels in adjacent normal tissues.

3.2 Significance of TMEM170B Expression and Prognosis of PAAD in the TCGA Cohort

The TCGA cohort was used to further investigate TMEM170B expression and its correlation with clinical prognosis among patients with PAAD. Patients with low TMEM170B expression had worse OS ($p < 0.001$) and disease-free survival (DFS) ($p = 0.0028$) than those with high TMEM170B expression (**Figure 3A**). The receiver operating characteristic (ROC) analysis showed that TMEM170B expression had robust prognostic predictive performance (areas under the ROC curve (AUC) = 0.708, 95% CI = 0.652–0.763, **Figure 3B**). Subsequently, the clinical significance of TMEM170B expression in patients with PAAD was systematically assessed using this cohort. The downregulation of TMEM170B expression was common in human PAAD. Low TMEM170B expression group usually have a history of alcohol abuse ($p < 0.05$), higher histologic grade ($p < 0.05$, G3 vs. G1) and more dead event ($p < 0.001$), but TMEM170B expression is not associated with sex, age, primary tumor (pT) stage, lymph node invasion, metastasis, and residual tumor (**Figure 3C**).

3.3 Association Between TMEM170B Expression and Clinicopathological Features of PAAD

Tumor tissues were divided into low and high TMEM170B expression based on the median values of H-score of each core. Since the median of H-score was 61, the patients were

TABLE 1 | Association between TMEM170B expression and clinicopathological characteristics in 38 pancreatic adenocarcinoma patients.

Variable	All Patients (*n* = 38) (%)	TMEM170B expression		*p* Value (χ2-test)
		Positive (+) (*n* = 24)	Negative (–) (*n* = 14)	
Gender (male/female)	25 (65.8)/13 (34.2)	17 (70.8)/7 (29.2)	8 (57.1)/6 (42.9)	0.61
Age (years) (< 60/≥60)	18 (47.4)/20 (52.6)	11 (45.8)/13 (54.2)	7 (50)/7 (50)	0.8
Tumor size (cm) (< 5/≥5)	23 (60.5)/15 (39.5)	14 (58.3)/10 (41.7)	9 (64.3)/5 (35.7)	0.72
Tumor number (Single/Multiple)	27 (71.1)/11 (28.9)	18 (75)/5 (25)	9 (64.3)/6 (35.7)	0.40
Tumor cellularity (< 61%/≥61%)	23 (60.5)/15 (39.5)	17 (70.8)/7 (29.2)	6 (42.9)/8 (57.1)	0.09
Differentiation				
Well	7 (18.4)	5 (20.8)	2 (14.3)	
Moderate	22 (62.9)	17 (70.8)	5 (35.7)	0.014*
Poor	9 (23.7)	2 (8.4)	7 (50)	
pT classification				
T1–2	5 (13.2)	4 (16.7)	1 (7.1)	
T3	17 (44.7)	12 (50)	5 (35.7)	0.33
T4	16 (42.1)	8 (33.3)	8 (57.2)	0.94
Lymph node metastasis (Absent/Present)	22 (62.9)/16 (37.1)	14 (58.3)/10 (41.7)	8 (57.2)/6 (42.8)	
pTNM stage I-II/III-IV	11 (28.9)/27 (71.1)	8 (33.3)/16 (66.7)	3 (21.4)/11 (78.6)	0.68
CEA (< 5/≥5)	21 (55.3)/17 (44.7)	15 (62.5)/9 (37.5)	6 (42.8)/8 (57.2)	0.24
CA19-9 (< 35/≥35)	17 (44.7)/21 (55.3)	10 (41.7)/14 (58.3)	7 (50)/7 (50)	0.62
Cancer progression, n (%) (no/yes)	26 (68.4)/12 (31.6)	22 (91.7)/2 (8.3)	4 (28.6)/10 (71.4)	<0.001
Death, n (%) (no/yes)	29 (76.3)/9 (23.7)	23 (95.8)/1 (4.2)	6 (42.9)/8 (57.1)	<0.001

*NOTE: p: primary; TNM: tumor-node-metastasis; CEA: carcinoembryonic antigen; CA19-9: carbohydrate antigen 19–9; *: p < 0.05.*

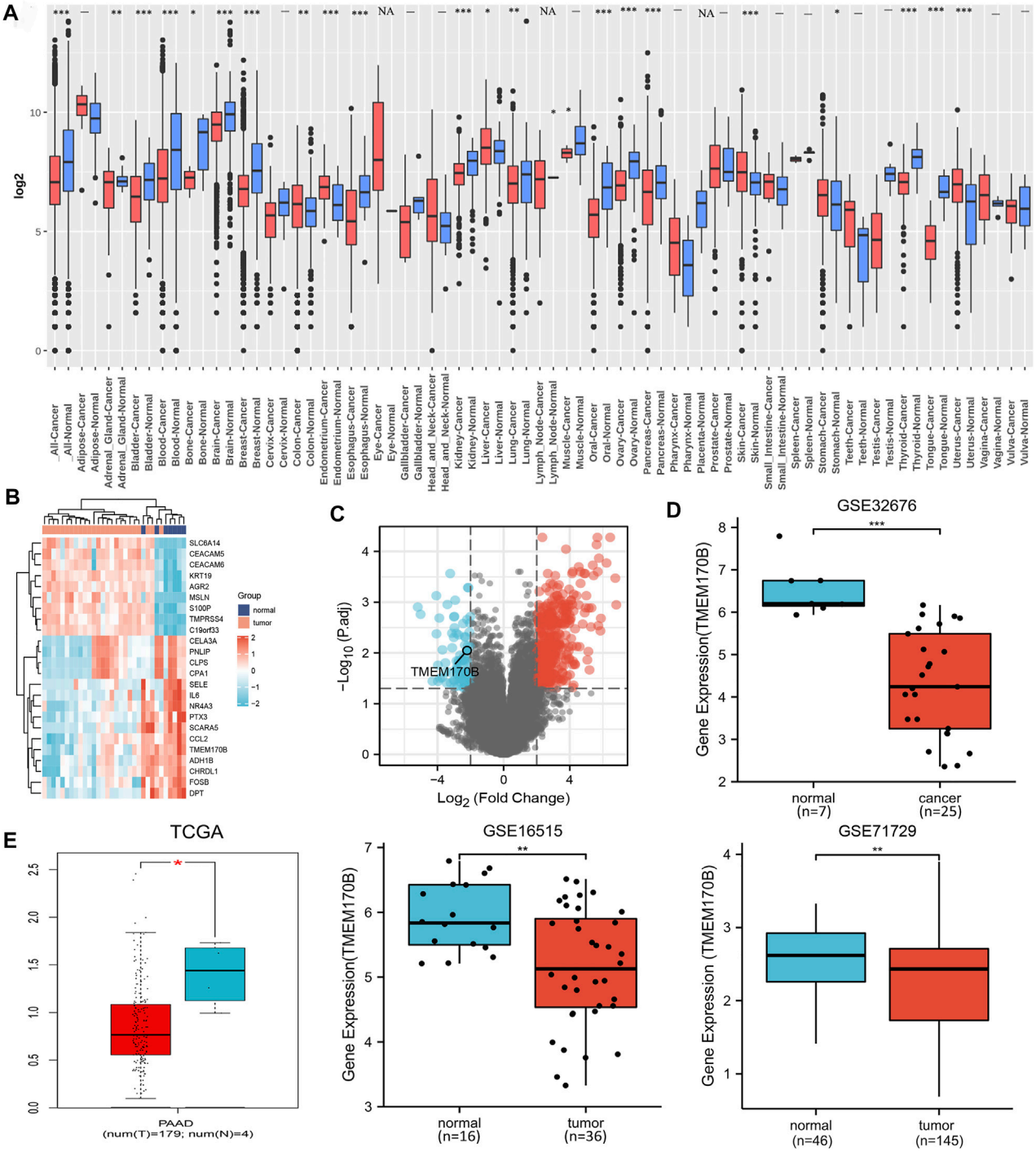


FIGURE 1 | TMEM170B mRNA expression levels in different tumors and corresponding adjacent normal tissues. **(A)** The pan-cancer TMEM170B expression status was analyzed using GENT2 based on the GEO database. **(B)** The heat map of the top 24 ranked DEGs. **(C)** The volcano plot of all DEGs, red represents up-regulated, blue represents down-regulated genes. **(D,E)** The expressions of TMEM170B in pancreatic adenocarcinoma and the corresponding nontumor tissues were inferred by analyzing the TCGA and GEO databases. ns: not significant. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

divided into two groups: H scores <61 was defined as low TMEM170B group, H scores ≥61 was high TMEM170B group, respectively. A total of 38 patients with complete clinical data were included to analyze the relationship between TMEM170B expression and various clinicopathological factors using the chi-squared test. The results illustrated that low TMEM170B expression had a significant correlation with poor differentiation (p = 0.014). Nevertheless, TMEM170B

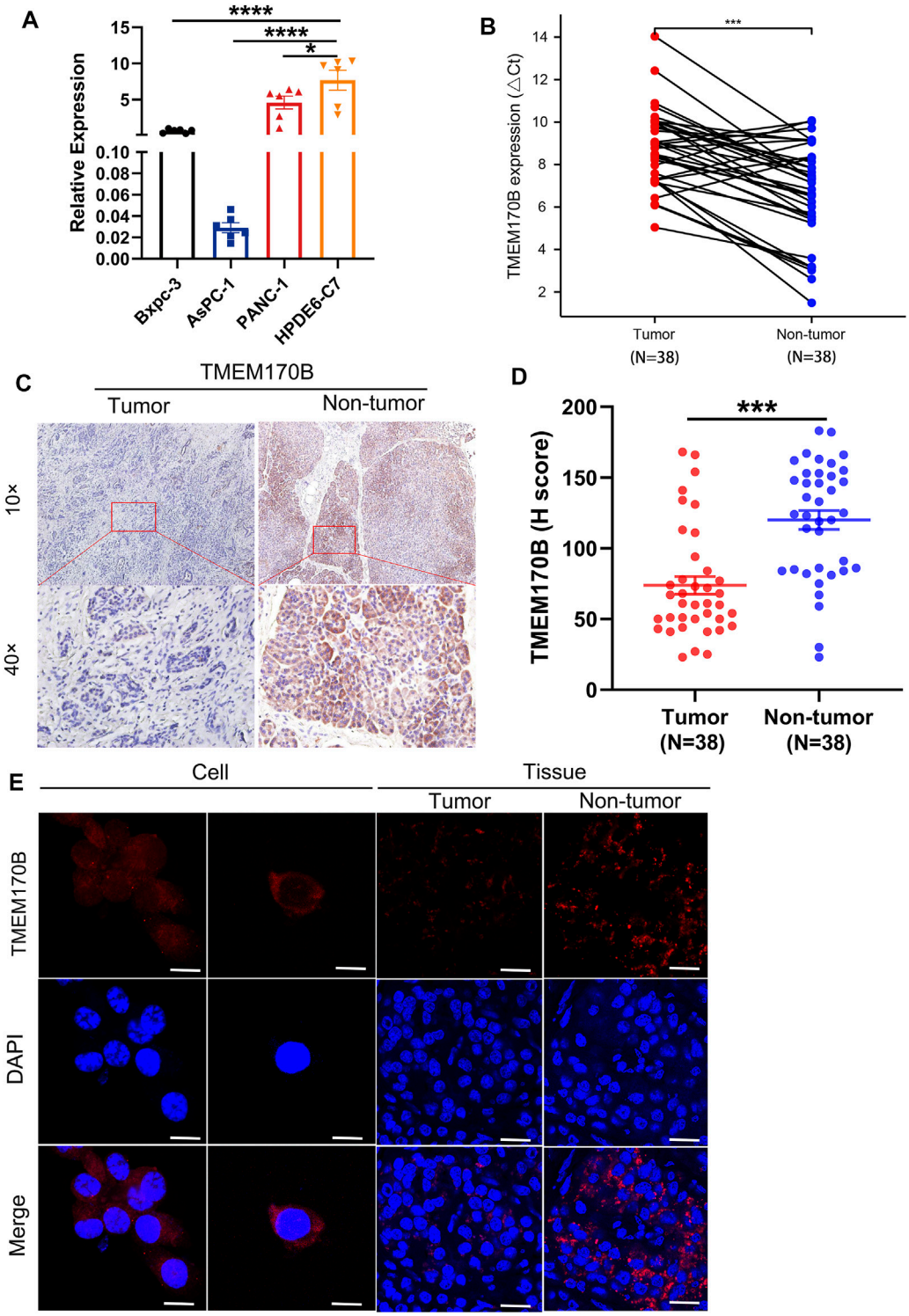


FIGURE 2 | TMEM170B expression was downregulated in human pancreatic cancer cell lines and tissues. **(A)** TMEM170B mRNA levels were low in pancreatic cancer cell lines. **(B)** Real-time PCR analysis of TMEM170B mRNA expression in 38 pairs of pancreatic adenocarcinoma (PAAD) tissue and paired nontumor tissue. **(C)** Representative immunohistochemistry (IHC) images of TMEM170B in 38 pairs of matched tumor tissues and adjacent nontumor tissues. **(D)** IHC quantification was performed using ImageJ software, and TMEM170B protein expression in PAAD tissues was lower than that in nontumor tissues. **(E)** The subcellular localization of TMEM170B was analyzed with immunostaining (cell: scale bar, 20 μm; tissue: scale bar, 30 μm); ns: not significant. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

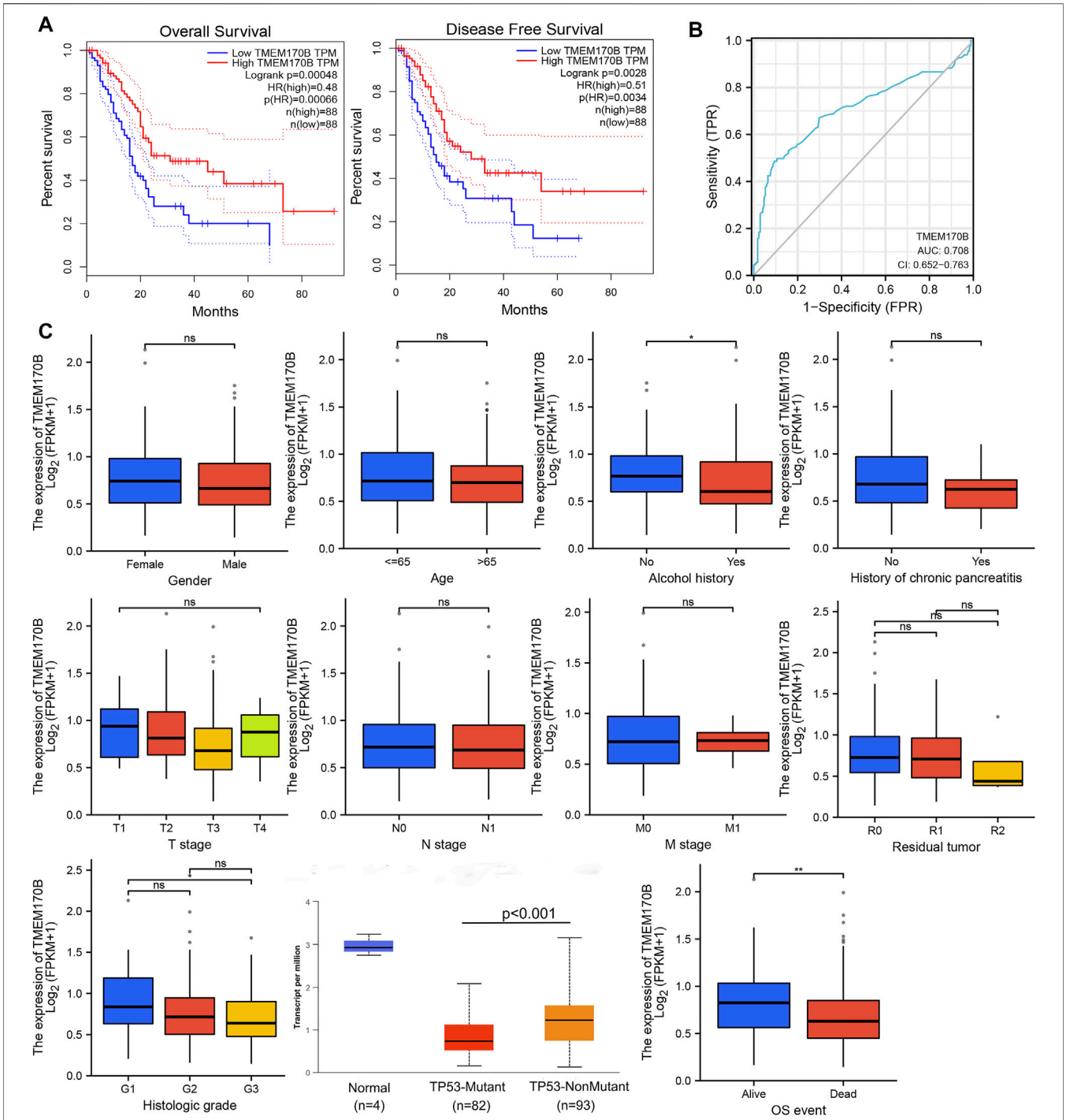


FIGURE 3 | Significance of TMEM170B expression and prognosis of pancreatic adenocarcinoma (PAAD) in the TCGA cohort. **(A)** The Kaplan–Meier survival curve analysis showed that low TMEM170B expression was related to poor overall survival and relapse-free survival. **(B)** Receiver operator characteristic curve analysis of TMEM170B expression in PAAD. **(C)** Relationship between TMEM170B expression and clinicopathological factors. ns: not significant; *$p < 0.05$; **$p < 0.01$.

expression was not significantly associated with sex, age, tumor size or number, pathological tumor-node-metastasis (pTNM) stage, pT classification, serum carcinoembryonic antigen, and carbohydrate antigen 19–9 (CA19-9) levels (**Table 1**). The Cox regression analysis of the forecast factors of OS by univariate analysis was showed in **Figure 4A**. A histopathological comparison of the degrees of differentiation and TMEM170B expression of tumor tissues also indicates that low expression of

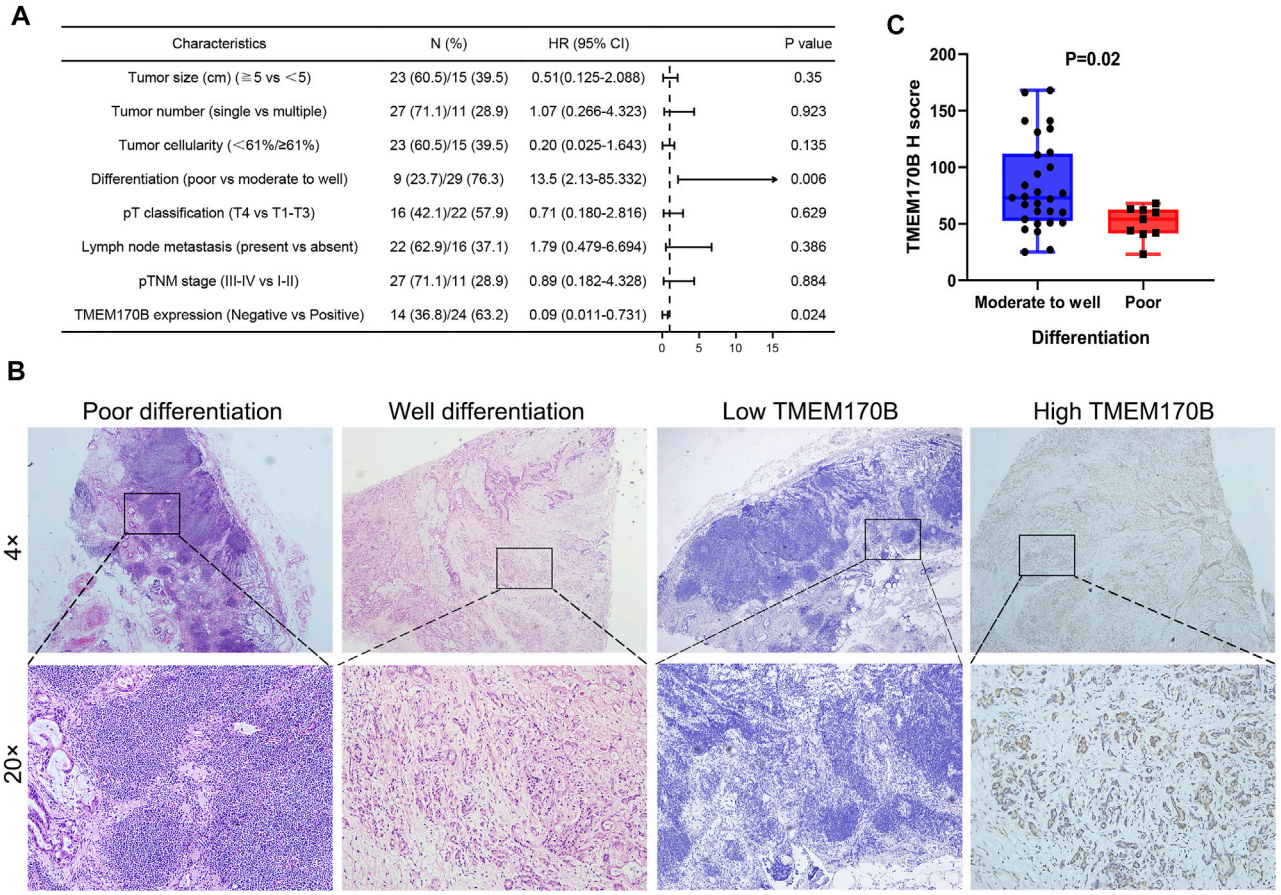


FIGURE 4 | Correlation of TMEM170B expression with the degree of differentiation in PAAD. **(A)** The Cox regression analysis of the forecast factors of OS by univariate analysis. **(B)** The histopathological comparison of the degrees of differentiation and TMEM170B expression of the same specimen. **(C)** TMEM170B Expression in Moderate to well differentiation group and poor group.

TMEM170B in PAAD was correlated with less tumor differentiation (**Figure 4B**). Moderate to well differentiation groups had generally higher TMEM170B expression relative to poor groups (p = 0.02, **Figure 4C**).

3.4 Correlation of TMEM170B Expression With Clinical Prognosis in PAAD

To determine whether TMEM170B expression levels were associated with clinical PAAD progression, we checked the association between TMEM170B expression and the clinical outcomes in our patients. Kaplan-Meier analysis revealed that patients in the TMEM170B-negative group had significantly worse OS (p = 0.005) and DFS (p = 0.0052) than those in the TMEM170B-positive group (**Figure 5A**). Univariate analysis identified poor differentiation (p = 0.006) and low TMEM170B expression (p = 0.024) as factors significantly associated with OS. Multivariate analysis indicated that differentiation [HR = 7.26, 95% CI = 1.141–46.147; p = 0.036] and TMEM170B expression (HR = 0.12, 95% CI = 0.014–0.995, and p = 0.049) were independent prognostic factors for OS. Similarly, univariate analysis demonstrated that poor differentiation (p = 0.013) and low TMEM170B expression (p = 0.016) were pivotal factors for DFS. However, multivariate analysis signified that only TMEM170B expression was an independent predictor (HR 0.19, 95% CI 0.04–0.91; p = 0.038) and not differentiation (HR 3.29, 95% CI 0.848–12.734; p = 0.085) (**Table 2**). Kaplan-Meier analysis also showed that moderate and poor differentiation had worse prognostic performance when compared with well differentiation (**Figure 5B**). In subgroup analysis, the moderate to well differentiation group with high TMEM170B expression achieved an average OS of 9.5 months, which was longer than that of the other groups (**Figure 5C**).

3.5 Correlation and Enrichment Analyses of TMEM170B in PAAD

For functional and pathway enrichment analyses of TMEM170B-related molecules, GO terms and KEGG enrichment pathway were visualized using the cluster Profiler R software package. According to functional enrichment and GO analyses, the genes were mainly enriched during biological activities such as leukocyte migration, cell–cell adhesion mediator activity,

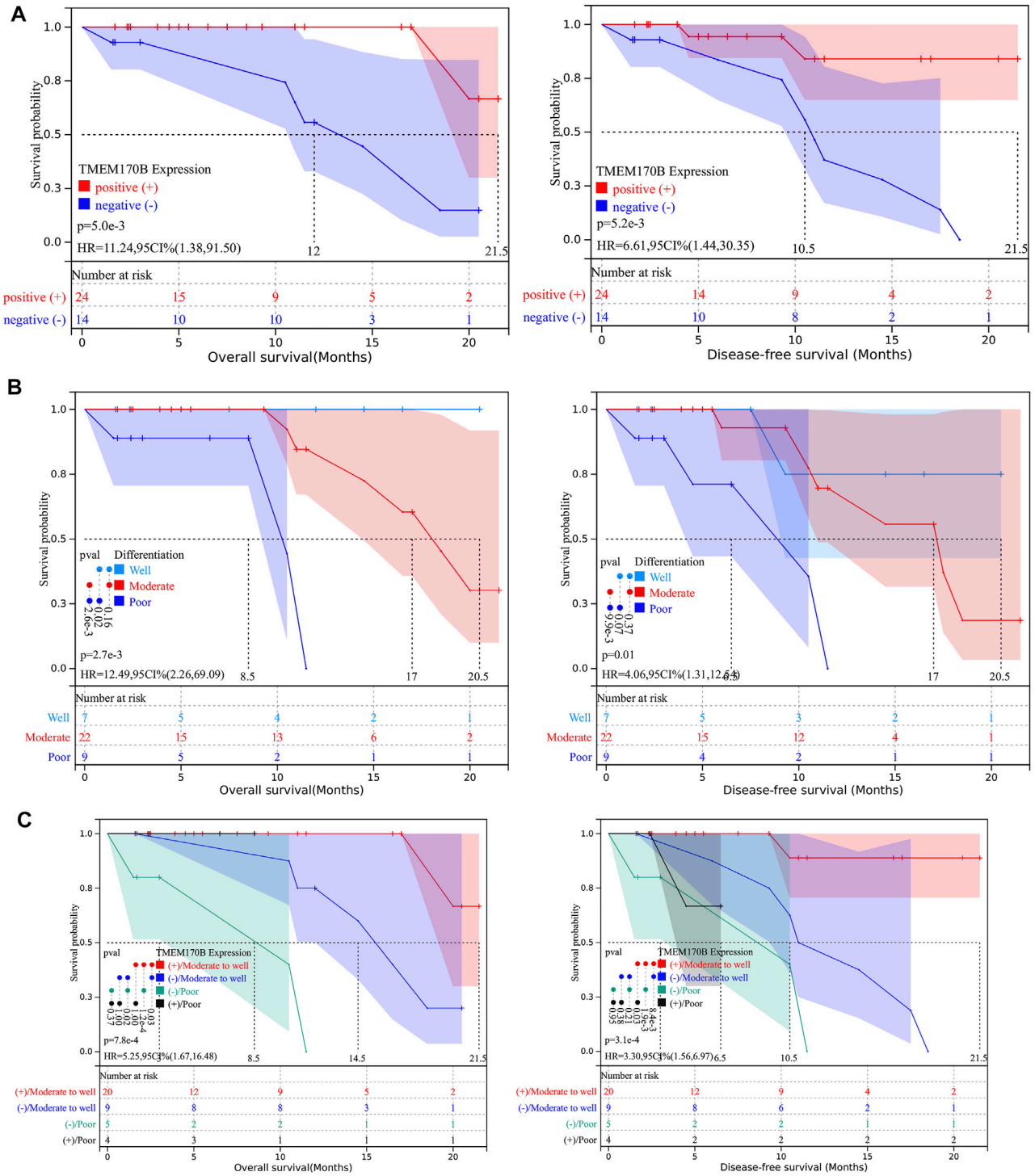


FIGURE 5 | TMEM170B expression and the degree of differentiation correlated with survival outcomes in our cohort. **(A)** Low TMEM170B expression was significantly associated with poor clinical outcomes. **(B)** The outcomes in the poor differentiation group were worse than those of the moderate and high differentiation groups. **(C)** Association of TMEM170B expression with the outcomes in poor differentiation or moderate to high differentiation: Subgroup analysis.

ameboidal-type cell migration and epithelial cell migration, all of which were correlation with the activities of immune cells (**Figures 6A–C**). Moreover, the KEGG pathway mainly involved cytokine–cytokine receptor interaction, malaria, the IL-17 signaling pathway, and the p53 signaling pathway (**Figure 6D**). Most importantly, GSEA was used to further

TABLE 2 | Univariate and multivariate Cox analyses of clinicopathological factors and clinical outcomes.

OS characteristics	Univariate analysis			Multivariate analysis		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Gender (male vs. female)	1.35	0.319–5.728	0.682			
Age (years) (≧60 vs .< 60)	0.65	0.173–2.427	0.337			
Tumor size (cm) (≧5 vs. < 5)	0.51	0.125–2.088	0.35			
Tumor number (single vs. multiple)	1.07	0.266–4.323	0.923			
Tumor cellularity (< 61% vs .≥ 61%)	0.20	0.025–1.643	0.135			
Differentiation (poor vs. moderate to well)	13.5	2.13–85.332	0.006**	7.26	1.141–46.147	0.036*
pT classification (T4 vs. T1-T3)	0.71	0.180–2.816	0.629			
Lymph node metastasis (present vs. absent)	1.79	0.479–6.694	0.386			
pTNM stage (III-IV vs. I-II)	0.89	0.182–4.328	0.884			
CEA (≧5 vs. < 5)	0.45	0.114–1.735	0.244			
CA19-9 (≧35 vs. < 35)	1.11	0.295–4.17	0.877			
TMEM170B expression (Negative vs. Positive)	0.09	0.011–0.731	0.024*	0.12	0.014–0.995	0.049*
DFS	**Univariate Analysis**			**Multivariate Analysis**		
	HR	**95% CI**	***p*-value**	**HR**	**95% CI**	***p*-value**
Gender (male vs. female)	0.6	0.16–2.217	0.44			
Age (years) (≧60 vs .< 60)	0.51	0.160–1.640	0.26			
Tumor size (cm) (≧5 vs. < 5)	0.78	0.237–2.546	0.677			
Tumor number (single vs. multiple)	1.14	0.358–3.605	0.828			
Tumor cellularity (< 61% vs .≥ 61%)	0.137	0.018–1.064	0.057			
Differentiation (poor vs. moderate to well)	5.43	1.421–20.776	0.013*	3.29	0.848–12.734	0.085
pT classification (T4 vs. T1-T3	0.78	0.216–2.812	0.704			
Lymph node metastasis (present vs. absent)	1.66	0.524–5.243	0.39			
pTNM stage (III-IV vs. I-II)	0.81	0.169–3.868	0.791			
CEA (≧5 vs. < 5)	0.47	0.124–1.742	0.256			
CA19-9 (≧35 vs. < 35)	2.45	0.716–8.398	0.153			
TMEM170B expression (Negative vs. Positive)	0.15	0.033–0.702	0.016*	0.19	0.04–0.910	0.038*

*Notes: OS, overall survival; DFS, Disease-free survival; HR, hazard ratio; CI, confidence interval, *p < 0.05, **p < 0.01.*

search for KEGG and Reactome pathways. GSEA-KEGG revealed that cytokine–cytokine receptor interaction, the chemokine signaling pathway, natural killer cell mediated cytotoxicity and the T cell receptor signaling pathway were significantly enriched (**Figure 6E**). Furthermore, G protein-coupled receptor (GPCR) ligand binding, interleukin signaling, and platelet activation signaling and aggregation were found to be significantly enriched in Reactome pathway analysis (**Figure 6F**). According the above results, we thus reasoned that TMEM170B may be involved in immune cell infiltration.

3.6 Correlation Between Immune Cell Infiltration, Immunotherapeutic Benefits and TMEM170B in PAAD

Immune cells in the tumor microenvironment play a central role in tumor-associated immune responses, which may regulate tumor progression and determine the prognosis. We further explore the correlation between TMEM170B expression and the infiltration level of different immune cells in PAAD. First and foremost, immuneScore, StromalScore and ESTIMATEScore were in the TMEM170B high-expression cluster than in the TMEM170B low-expression cluster (**Figure 7A**). ssGSEA with Spearman's rank correlation was applied to the analysis of correlation between TMEM170B expression and infiltration levels of 24 immune cell types, the result showed that TMEM170B expression was negatively correlated with the infiltration of Th2 cells and NK CD56bright cells (**Figure 7B**). By TIMER, CIBERSORT, MCP-counter and EPIC algorithms, the results indicated that TMEM170B expression was positively correlated with the infiltration of antitumor immune cells, including B cells, $CD8^+$T cells, $CD4^+$T cells, dendritic cells (DCs), natural killer cells (NKs), neutrophils, monocyte, and M1 macrophage (**Figures 7C,D**). Moreover, a negative correlation between TMEM170B expression and immune infiltration of myeloid-derived suppressor cells (MDSCs) and T cell regulatory (Tregs) was noted (**Figure 7E**). Elevated monocyte and NKs and reduced MDSCs infiltration were associated with good prognosis in patients with PAAD (**Figure 7F**). The expression of immune checkpoint (ICP) genes, such as PDCD1 (PD1), CD274 (PDL1), CTLA4, LAG3, and HAVCR2 (TIM3) has been utilized in predicting the response of patients to immune checkpoints therapy in a variety of cancers including PAAD. In gene-by-gene correlation analysis, these common ICP genes were positively correlated with TMEM170B expression (**Figures 7G,H**). This illustrated that TMEM170B might have value in predicting immunotherapeutic benefits.

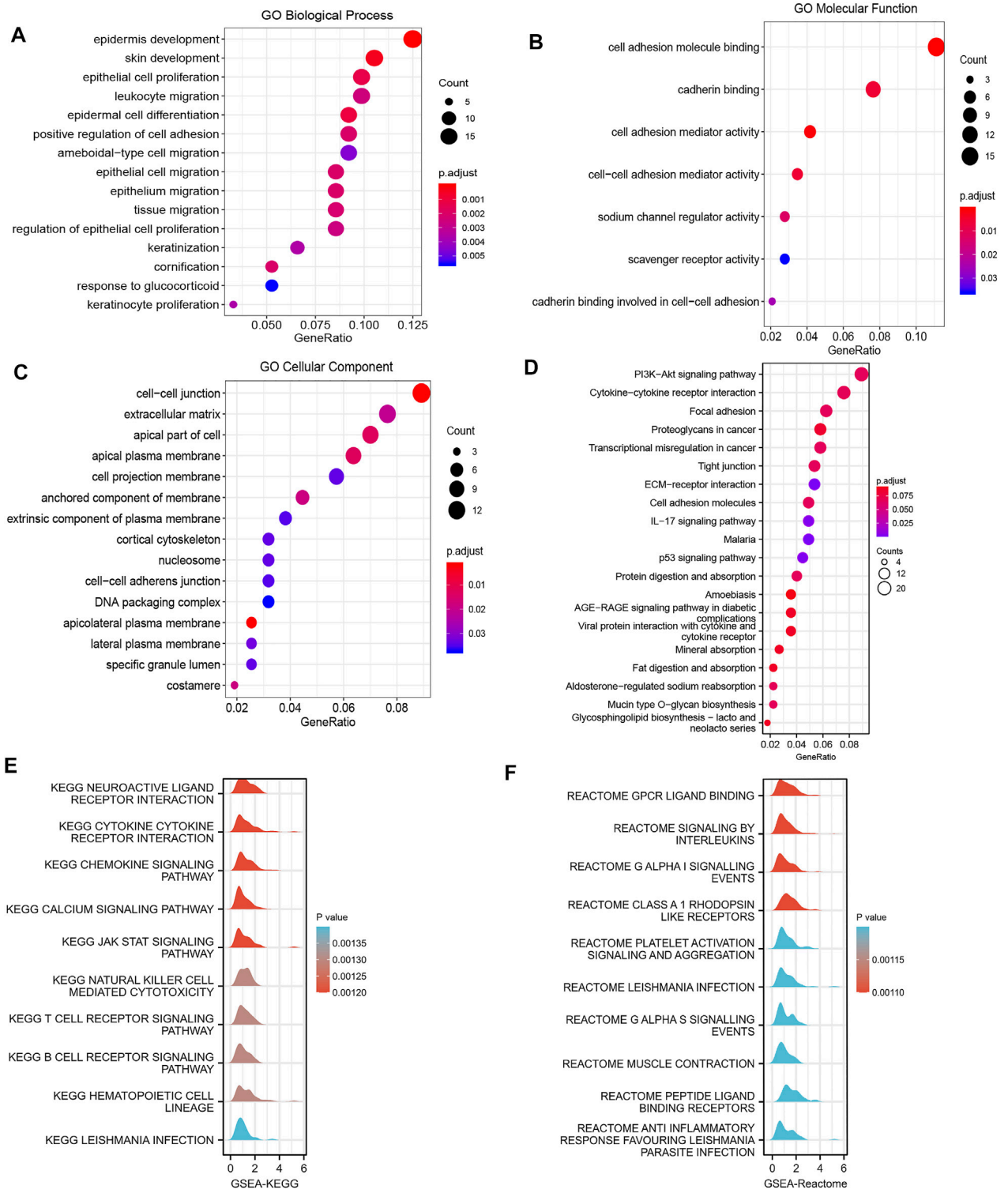


FIGURE 6 | Function and pathway enrichment analyses of TMEM170B in pancreatic adenocarcinoma. **(A–C)** Significant GO terms of the top 300 genes most positively associated with TMEM170B, including biological processes, molecular function, and cell component. **(D)** Significant KEGG pathways of TMEM170B-related genes. **(E,F)** Significant GSEA results of the top 300 genes most positively associated with TMEM170B, including KEGG pathways and Reactome pathways.

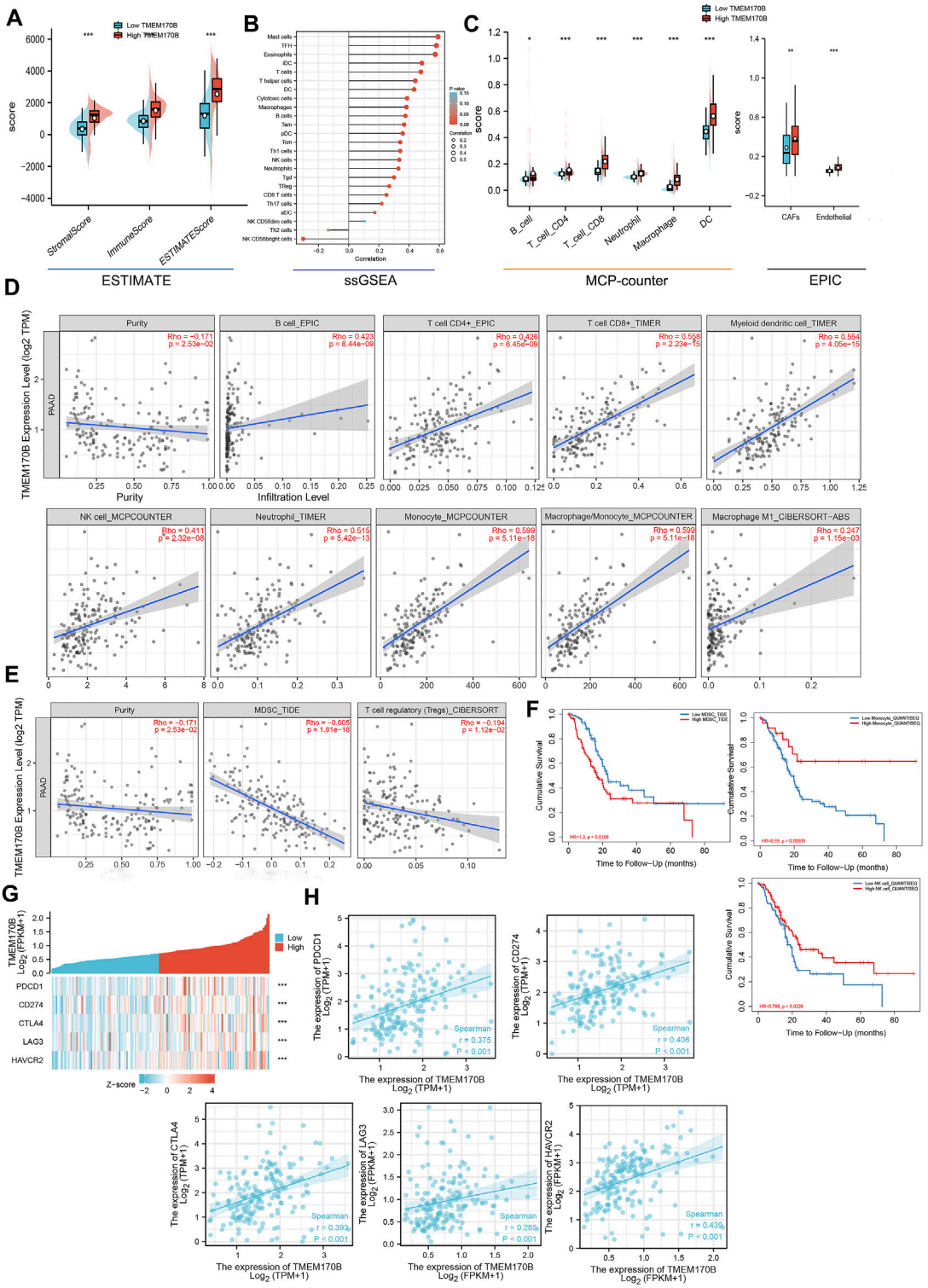


FIGURE 7 | Correlation analysis between TMEM170B expression and immune infiltration of multiple immune cells. **(A)** Violin plot showing the differential StromalScore, ImmuneScore, and ESTIMATEScore of high and low TMEM170B expression. **(B)** The correlation between TMEM170B expression level and 24 immune cell types. **(C–D)** The correlation between TMEM170B expression and antitumor immune cell infiltration, including that of B cells, $CD8^+$T cells, $CD4^+$T cells, myeloid dendritic cells, natural killer cells, neutrophils, monocyte and M1 macrophages. **(E)** The correlation between TMEM170B expression and the infiltration of immunosuppressive cells, including myeloid-derived suppressor cells and regulatory T cells. **(F)** Kaplan-Meier analysis results based on the extent of monocyte, natural killer cells and yeloid-derived suppressor cells infiltration in PAAD. Red represents high immune cells infiltration, and blue represents low immune cells infiltration. **(G,H)** The gene-by-gene correlation analysis of common ICP genes and TMEM170B expression. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

3.7 TMEM170B Expression Is Correlated With the Infiltration Levels of CD4$^+$ T Cells, CD8$^+$T Cells, and MDSCs

Next, the immunofluorescence method was performed to validate whether immune cell infiltration was correlated with TMEM170B expression in our cohort. Since T cells and MDSCs were the most prevalent cells among the infiltrated immune cells correlated with TMEM170B expression, and previous studies have shown that CD4$^+$T Cells, CD8$^+$T cells and MDSCs are the major factors responsible for impacting pancreatic cancer progression, invasiveness, metastasis, and patients' survival Foley et al. (2016); Thyagarajan et al. (2019), we further explored the correlation between CD4$^+$T Cells, CD8$^+$T cells, and MDSCs infiltration and TMEM170B expression. It was observed in tumor center that the tumor tissue presented a high TMEM170B expression level, the infiltration Levels of CD4$^+$T Cells, CD8$^+$T cells was relatively active. While infiltration of CD4$^+$T Cells, CD8$^+$T cells was significantly low in areas with a corresponding low TMEM170B expression status (**Figures 8A,B**). Conversely, we observed that the MDSCs (labeled by CD11b$^+$CD33$^+$) was obviously recruited in tumor regions with low TMEM170B expression level (**Figure 8C**). Therefore, these results preliminarily verified our inferences in the above analysis that TMEM170B was indeed related to CD4$^+$, CD8$^+$T cells recruitment, and MDSCs blocking.

4 DISCUSSION

Owing to the surge in the prevalence of obesity, diabetes, alcohol consumption, smoking, and pancreatitis, both the incidence and mortality rates of pancreatic cancer have increased steadily in Europe and the United States over the past decades Connor and Gallinger, (2015). Despite the recent therapeutic advances, the prognosis of the patients is yet to improve significantly. PAAD is a highly invasive and metastatic type of pancreatic cancer that results in poor clinical outcomes in the patients Hackeng et al. (2016). However, its pathogenesis and underlying mechanisms are yet to be elucidated Fang et al. (2017); Hoskins et al. (2016); Klein et al. (2018; Zheng et al. (2016). Thus, there is an urgent need to identify novel markers and therapeutic targets for the early diagnosis of pancreatic cancer and prevention of metastasis. Furthermore, research in this direction is likely to aid in developing effective targeted molecular therapy for pancreatic cancer.

TMEM170B was mapped to human chromosome 6p24.2, which belongs to the TMEM family. TMEM170B and its paralog, TMEM170A, have shown diverse but specific expression profiles in different tumor cells and tissues and are involved in several biological processes, such as gene expression, invasion, proliferation, tumor apoptosis, normal cellular function, and disease pathogenesis Zhong et al. (2020). Current research has shown that TMEM170B negatively regulates canonical Wnt signaling in breast cancer cells and exerts an inhibitory effect on breast cancer growth by inhibiting CTNNB1 stabilization and nucleus translocation, which reduces the activity of Wnt targets Li et al. (2018).

The present study is the first to show that TMEM170B expression is significantly decreased in human pancreatic cancer cell lines and PAAD tissues. More importantly, univariate and multivariate analyses identified TMEM170B expression level and the degree of differentiation as the independent prognostic factors for OS and DFS in patients with PAAD. This finding suggests that low TMEM170B expression and poor differentiation are significantly related to poor prognosis. Consistently, low expression levels of TMEM170B are closely associated with poor differentiation and *TP53* mutation in patients with PAAD.

As a tumor suppressor gene, the tumor suppressor effect of *TP53* gene has been demonstrated in multiple tumors Olivier et al. (2010). Mutations in *TP53* indicate a poor prognosis in pancreatic cancer Guerra et al. (2011; Lu et al. (2020). Our study showed that low expression of TMEM170B is significantly related to *TP53* mutations. KEGG analysis also revealed that TMEM170B is involved in the p53 signaling pathway, which indicates that the presence of TMEM170B may inhibit the occurrence of *TP53* mutations and then exert a tumor suppressor effect in PAAD.

The tumor microenvironment (TME) and its individual immune cells play important roles in tumor initiation, progression, and metastasis Pattabiraman and Weinberg, (2014). A study showed that immune suppressive cells in pancreatic cancer maintain a tumor-friendly environment, with the majority of immune cells being macrophages and exhausted lymphocytes, based on single-cell RNA sequencing Lin et al. (2020). As such, we explored the biological functions and complex mechanisms of TMEM170B, and the correlation between immune cell infiltration and TMEM170B. GO results showed that TMEM170B is involved in leukocyte migration and cell–cell junction, whereas KEGG and GSEA indicated that TMEM170B is closely associated with immune-related pathways, cytokine–cytokine receptor interaction, chemokine signaling pathway, and T and B cell receptor signaling pathways in PAAD.

By analyzing the relationship between TMEM170B and immune cell infiltration, we found that low TMEM170B expression is often associated with MDSCs and Tregs infiltration into tumors, which aids in immune evasion. Moreover, low TMEM170B expression is linked to decreased levels of antitumor immune cells, such as B cells, CD8$^+$T cells, CD4$^+$T cells, NKs, DCs, and M1 monocytes. Therefore, activating TMEM170B may increase lymphocyte infiltration and accordingly decrease the immune suppressive cells within the TME. Tissue staining showed that TMEM170B had strong positive correlations with tumour-infiltrating CD4$^+$ and CD8$^+$ cells, and negative correlations with MDSCs.

Immunotherapy, which is used in cancer treatment to block the immune checkpoints such as programmed cell death-1 (PD-1) and programmed cell death-ligand 1(PD-L1), has shown persisting clinical responses and prolonging survival Feng et al. (2017). Hence, TMEM170B can be a potential prognostic biomarker and immunotherapy agent in combination therapy regimens to improve pancreatic cancer treatment.

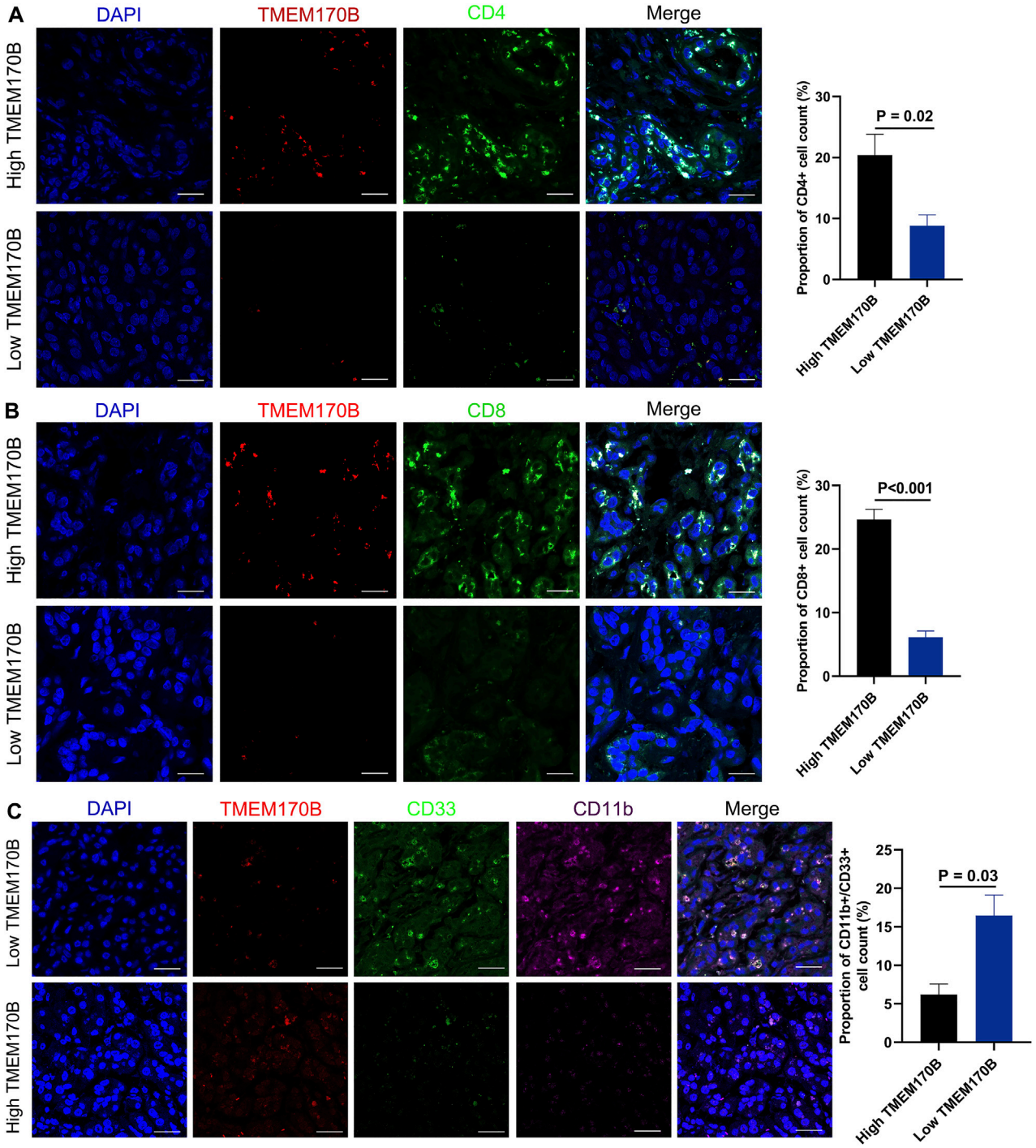


FIGURE 8 | Correlation Between TMEM170B Expression and the Infiltration Levels of $CD4^+$ T Cells, CD8+T cells, and MDSCs in PAAD tissues. **(A)** TMEM170B expression was positively correlated with the infiltration of $CD4^+$T cells. **(B)** TMEM170B expression was positively correlated with the infiltration of $CD8^+$T cells. **(C)** TMEM170B expression was negatively correlated with the infiltration of MDSCs. Scale bar, 30 μm.

In conclusion, this study is the first to confirm that human pancreatic cancer cells and tissues have decreased expressed of TMEM170B, which suggests that TMEM170B can serve as an independent prognostic predictor following surgery in patients with PAAD. Moreover, our results showed that the antitumor effects might be explained by the reduced number of immunosuppressive cells (MDSCs and Tregs) and the infiltration of antitumor immune cells ($CD8^+$T cells, $CD4^+$T cells, and M1 macrophage) into the TME. Thus, TMEM170B could be regarded as a novel target to address cancer progression. Future research

endeavors to design novel drugs that can activate TMEM170B are expected to counteract the immunosuppressive microenvironment and improve the response to immunotherapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital ethics committee. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZZ and XH contributed to the study conception and design. Material preparation, data collection and analysis were performed by ZZ and JS, and DZ. YY and YS contributed to the literature search. The first draft of the manuscript was written by ZZ and all authors commented on previous versions of the manuscript. CL, QY and TF reviewed the paper and gave suggestions on the revision of the article. All authors read and approved the final manuscript.

FUNDING

This work was supported by the following project funds: National Basic Research Program of China (973 program) (2015CB964703), Sichuan Province Science and Technology Support Program (2016JY0020), Sichuan Province Science and Technology Support Program (2018HH0062), and Sichuan Province Science and Technology Support Program (2021YFH0187).

REFERENCES

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Cassetta, L., and Kitamura, T. (2018). Targeting Tumor-Associated Macrophages as a Potential Strategy to Enhance the Response to Immune Checkpoint Inhibitors. *Front. Cel Dev. Biol.* 6, 38. doi:10.3389/fcell.2018.00038

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Christodoulou, A., Santarella-Mellwig, R., Santama, N., and Mattaj, I. W. (2016). Transmembrane Protein TMEM170A Is a Novel Regulator of ER and NE Morphogenesis in Human Cells. *J. Cel Sci* 129, 1552–1565. doi:10.1242/jcs.175273

Connor, A. A., and Gallinger, S. (2015). Hereditary Pancreatic Cancer Syndromes. *Surg. Oncol. Clin. North America* 24, 733–764. doi:10.1016/j.soc.2015.06.007

Connor, A. A., and Gallinger, S. (2021). Pancreatic Cancer Evolution and Heterogeneity: Integrating Omics and Clinical Data. *Nat. Rev. Cancer* 22, 131–142. doi:10.1038/s41568-021-00418-1

Duan, L., Hu, X.-q., Feng, D.-y., Lei, S.-y., and Hu, G.-h. (2013). GPC-1 May Serve as a Predictor of Perineural Invasion and a Prognosticator of Survival in Pancreatic Cancer. *Asian J. Surg.* 36, 7–12. doi:10.1016/j.asjsur.2012.08.001

Fang, J., Jia, J., Jia, J., Makowski, M., Xu, M., Wang, Z., et al. (2017). Functional Characterization of a Multi-Cancer Risk Locus on chr5p15.33 Reveals Regulation of TERT by ZNF148. *Nat. Commun.* 8. doi:10.1038/ncomms15034

Feng, M., Xiong, G., Cao, Z., Yang, G., Zheng, S., Song, X., et al. (2017). PD-1/PD-L1 and Immunotherapy for Pancreatic Cancer. *Cancer Lett.* 407, 57–65. doi:10.1016/j.canlet.2017.08.006

Foley, K., Kim, V., Jaffee, E., and Zheng, L. (2016). Current Progress in Immunotherapy for Pancreatic Cancer. *Cancer Lett.* 381, 244–251. doi:10.1016/j.canlet.2015.12.020

Guerra, C., Collado, M., Navas, C., Schuhmacher, A. J., Hernández-Porras, I., Cañamero, M., et al. (2011). Pancreatitis-induced Inflammation Contributes to Pancreatic Cancer by Inhibiting Oncogene-Induced Senescence. *Cancer Cell* 19, 728–739. doi:10.1016/j.ccr.2011.05.011

Hackeng, W. M., Hruban, R. H., Offerhaus, G. J. A., and Brosens, L. A. A. (2016). Surgical and Molecular Pathology of Pancreatic Neoplasms. *Diagn. Pathol.* 11, 47. doi:10.1186/s13000-016-0497-z

Hoskins, J. W., Ibrahim, A., Emmanuel, M. A., Manmiller, S. M., Wu, Y., O'Neill, M., et al. (2016). Functional Characterization of a chr13q22.1 Pancreatic Cancer Risk Locus Reveals Long-Range Interaction and Allele-specific Effects onDIS3expression. *Hum. Mol. Genet.* 25, ddw300–4738. doi:10.1093/hmg/ddw300

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 48, D498–D503. doi:10.1093/nar/gkz1031

Klein, A. P., Wolpin, B. M., Risch, H. A., Stolzenberg-Solomon, R. Z., Mocci, E., Zhang, M., et al. (2018). Genome-wide Meta-Analysis Identifies Five New Susceptibility Loci for Pancreatic Cancer. *Nat. Commun.* 9, 556. doi:10.1038/s41467-018-02942-5

Li, M., Han, Y., Zhou, H., Li, X., Lin, C., Zhang, E., et al. (2018). Transmembrane Protein 170B Is a Novel Breast Tumorigenesis Suppressor Gene that Inhibits the Wnt/β-Catenin Pathway. *Cell Death Dis* 9, 91. doi:10.1038/s41419-017-0128-y

Lin, W., Noel, P., Borazanci, E. H., Lee, J., Amini, A., Han, I. W., et al. (2020). Single-cell Transcriptome Analysis of Tumor and Stromal Compartments of Pancreatic Ductal Adenocarcinoma Primary Tumors and Metastatic Lesions. *Genome Med.* 12, 80. doi:10.1186/s13073-020-00776-9

Lu, S.-W., Pan, H.-C., Hsu, Y.-H., Chang, K.-C., Wu, L.-W., Chen, W.-Y., et al. (2020). IL-20 Antagonist Suppresses PD-L1 Expression and Prolongs Survival in Pancreatic Cancer Models. *Nat. Commun.* 11, 4611. doi:10.1038/s41467-020-18244-8

Mace, T. A., Shakya, R., Pitarresi, J. R., Swanson, B., McQuinn, C. W., Loftus, S., et al. (2018). IL-6 and PD-L1 Antibody Blockade Combination Therapy Reduces Tumour Progression in Murine Models of Pancreatic Cancer. *Gut* 67, 320–332. doi:10.1136/gutjnl-2016-311585

Marco, M. D., Grassi, E., Durante, S., Vecchiarelli, S., Palloni, A., Macchini, M., et al. (2016). State of the Art Biological Therapies in Pancreatic Cancer. *Wjgo* 8, 55–66. doi:10.4251/wjgo.v8.i1.55

Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspect. Biol.* 2, a001008. doi:10.1101/cshperspect.a001008

O'Reilly, E. M., Oh, D.-Y., Dhani, N., Renouf, D. J., Lee, M. A., Sun, W., et al. (2019). Durvalumab with or without Tremelimumab for Patients with Metastatic Pancreatic Ductal Adenocarcinoma. *JAMA Oncol.* 5, 1431–1438. doi:10.1001/jamaoncol.2019.1588

Pattabiraman, D. R., and Weinberg, R. A. (2014). Tackling the Cancer Stem Cells - what Challenges Do They Pose? *Nat. Rev. Drug Discov.* 13, 497–512. doi:10.1038/nrd4253

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *eLife* 6. doi:10.7554/eLife.26476

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. doi:10.1038/s41596-018-0103-9

Shi, S., and Cao, H. (2014). Shikonin Promotes Autophagy in BXPC-3 Human Pancreatic Cancer Cells through the PI3K/Akt Signaling Pathway. *Oncol. Lett.* 8, 1087–1089. doi:10.3892/ol.2014.2293

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics, 2018. *CA: A Cancer J. Clinicians* 68, 7–30. doi:10.3322/caac.21442

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Thyagarajan, A., Alshehri, M. S. A., Miller, K. L. R., Sherwin, C. M., Travers, J. B., and Sahu, R. P. (2019). Myeloid-Derived Suppressor Cells and Pancreatic Cancer: Implications in Novel Therapeutic Approaches. *Cancers* 11, 1627. doi:10.3390/cancers11111627

Waddell, N., Pajic, M., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., et al. (2015). Whole Genomes Redefine the Mutational Landscape of Pancreatic Cancer. *Nature* 518, 495–501. doi:10.1038/nature14169

Wang, L., Xiong, L., Wu, Z., Miao, X., Liu, Z., Li, D., et al. (2018). Expression of UGP2 and CFL1 Expression Levels in Benign and Malignant Pancreatic Lesions and Their Clinicopathological Significance. *World J. Surg. Onc* 16, 11. doi:10.1186/s12957-018-1316-7

Yang, Z., Li, D., Liu, Z., Miao, X., Yang, L., Zou, Q., et al. (2017). BIRC7 and KLF4 Expression in Benign and Malignant Lesions of Pancreas and Their Clinicopathological Significance. *Cbm* 17, 437–444. doi:10.3233/CBM-160660

Yi, X.-P., Han, T., Li, Y.-X., Long, X.-Y., and Li, W.-Z. (2015). Simultaneous Silencing of XIAP and Survivin Causes Partial Mesenchymal-Epithelial Transition of Human Pancreatic Cancer Cells via the PTEN/PI3K/Akt Pathway. *Mol. Med. Rep.* 12, 601–608. doi:10.3892/mmr.2015.3380

Zheng, J., Huang, X., Tan, W., Yu, D., Du, Z., Chang, J., et al. (2016). Pancreatic Cancer Risk Variant in LINC00673 Creates a miR-1231 Binding Site and Interferes with PTPN11 Degradation. *Nat. Genet.* 48, 747–757. doi:10.1038/ng.3568

Zhong, J., Jermusyk, A., Wu, L., Hoskins, J. W., Collins, I., Mocci, E., et al. (2020). A Transcriptome-wide Association Study Identifies Novel Candidate Susceptibility Genes for Pancreatic Cancer. *J. Natl. Cancer Inst.* 112, 1003–1012. doi:10.1093/jnci/djz246

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Shang, Dai, Yao, Shi, Zhong, Liang, Lai, Yang, Feng and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 May 2022
doi: 10.3389/fgene.2022.872026

Prognostic Value of *SPOCD1* in Esophageal Squamous Cell Carcinoma: A Comprehensive Study Based on Bioinformatics and Validation

Zhizhong Lin[1], Lin Chen[1], Tingting Wu[2,3], Yiping Zhang[1], Xinyi Huang[1], Yuanmei Chen[4], Junqiang Chen[1] and Yuanji Xu[1]**

[1]Department of Radiation Oncology, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China, [2]The School of Nusing, Fujian Medical University, Fuzhou, China, [3]Shengli Clinical Medical College of Fujian Medical University, Fuzhou, China, [4]Department of Thoracic Surgery, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China

OPEN ACCESS

Edited by:
Ruowen Zhang, Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Cen Wu, Kansas State University, United States
Yanmei Zou, Huazhong University of Science and Technology, China
Jing Wang, Sun Yat-sen University Cancer Center (SYSUCC), China

***Correspondence:**
Yuanji Xu
xuyuanji@fjmu.edu.cn
Tingting Wu
itt2016@126.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *09 February 2022*
Accepted: *20 April 2022*
Published: *11 May 2022*

Citation:
Lin Z, Chen L, Wu T, Zhang Y, Huang X, Chen Y, Chen J and Xu Y (2022) Prognostic Value of SPOCD1 in Esophageal Squamous Cell Carcinoma: A Comprehensive Study Based on Bioinformatics and Validation. Front. Genet. 13:872026. doi: 10.3389/fgene.2022.872026

In the study, we aimed to explore and analyze the potential function of SPOC Domain Containing 1 (SPOCD1) in esophageal squamous cell carcinoma (ESCC). We performed a comprehensive analysis of gene expression of SPOCD1 and its corresponding clinicopathological features in ESCC. In particular, the correlation between SPOCD1 and ESCC was evaluated using a wide range of analysis tools and databases, including TCGA, GTEx, GenePattern, CellMiner, GDSC, and STRING datasets. Different bioinformatics analyses, including differential expression analysis, mutation analysis, drug sensitivity analysis, function analysis, pathway analysis, co-expression network analysis, immune cell infiltration analysis, and survival analysis, were carried out to comprehensively explore the potential molecular mechanisms and functional effects of SPOCD1 on the initiation and progression of ESCC. The expression of SPOCD1 was upregulated in ESCC tissues compared to those in normal tissues. In the high SPOCD1 expression group, we found apparent mutations in TP53, TTN, and MUC16 genes, which were 92, 36, and 18%, respectively. GO and KEGG enrichment analysis of SPOCD1 and its co-expressed genes demonstrated that it may serve as an ESCC oncogene by regulating the genes expression in the essential functions and pathways of tumorigenesis, such as glycosaminoglycan binding, Cytokine-cytokine receptor interaction, and Ras signaling pathway. Besides, the immune cell infiltration results revealed that SPOCD1 expression was positively correlated with Macrophages M0 and Mast cells activated cells, and negatively correlated with plasma cells and T cells follicular helper cell infiltration. Finally, ESCC patients with high expression of SPOCD1 indicated poor overall survival. qRT-PCR demonstrated that the SPOCD1 expression in ESCC tissues was significantly higher than adjacent tissues ($p < 0.001$). Our study indicated that SPOCD1 was increased in ESCC tissues. The current data support the oncogenic role of SPOCD1 in the occurrence and development of ESCC. Most importantly, SPOCD1 might be an independent prognostic factor for ESCC patients.

Keywords: SPOCD1, esophageal squamous cell carcinoma, novel biomarkers, prognosis, immune infiltration

INTRODUCTION

Esophageal cancer (EC) is the sixth most common cause of cancer death and the seventh most common malignant tumor globally, with a survival rate of less than 20% (Torre et al., 2015; Bray et al., 2018). EC is mainly divided into two pathological types: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EA). ESCC is the predominant type in Asian countries (Abnet et al., 2018). Despite recent progress in surgery, radiotherapy, chemotherapy, and immunotherapy, the 5-years survival rate of ESCC patients remains dismal (Pennathur et al., 2013; Wang et al., 2018; Xu et al., 2019). The outcome of ESCC remains extremely poor, which is mainly due to its insensitive to chemotherapy and rapid progress. Due to the lack of reliable biomarkers, most patients have experienced the advanced stage at the time of diagnosis. Distant metastasis usually occurs at this stage, resulting in a poor outcome. In addition, patients with advanced ESCC often suffer from severe pain that makes it difficult to be treated. Therefore, it is in dire need to find promising diagnostic biomarkers and novel therapeutic targets of ESCC, which will help reveal the molecular mechanisms of ESCC tumorigenesis and metastasis.

SPOC Domain Containing 1 (SPOCD1), also known as PPP1R146, is a protein coding gene. An important paralog of this gene is Death Inducer-Obliterator 1 (DIDO1). It encodes a protein that belongs to the transcription factor S-II (TFIIS) family of transcription factors. Alternate splicing results in multiple transcript variants. It is widely expressed in many tissues and has been reported in various tumors. Most notably, some studies have explored the biological function of SPOCD1 in human cancer, such as gastric cancer, clear cell renal cell carcinoma, ovarian cancer, osteosarcoma, and glioma (Zhu et al., 2017; Liang et al., 2018; Liu et al., 2018; Sakaguchi et al., 2018; Liu et al., 2020). In addition, some studies have demonstrated that SPOCD1 is an essential executor of piRNA-directed *de novo* DNA methylation in recent years (Zoch et al., 2020). These studies revealed the underlying value of SPOCD1 in tumors and may provide a new direction for further understanding the specific role and molecular mechanism of SPOCD1. Nonetheless, there is no study of SPOCD1 in ESCC at present. In addition, SPOCD1 is considered to be an attractive therapeutic and prognostic target for cancer, and inhibition of SPOCD1 may be a feasible cancer treatment strategy. Consequently, we selected SPOCD1 as the research object and explored the potential function and mechanism of SPOCD1 through various kinds of analyses in ESCC.

So far, the analysis on the SPOCD1 role in ESCC remains largely unknown. In the present study, we performed a comprehensive analysis of the gene expression of SPOCD1 and its corresponding clinicopathological features in ESCC. In particular, the correlation between SPOCD1 and ESCC was evaluated using a wide range of analysis tools and databases, including TCGA, GTEx, GenePattern, CellMiner, GDSC, and STRING datasets. Different bioinformatics analyses, including differential expression analysis, single nucleotide polymorphism (SNP)/copy number variation (CNV) analysis, drug sensitivity analysis, function analysis, pathway analysis, co-expression network analysis, immune cell infiltration analysis, and survival analysis, were carried out to comprehensively explored the potential molecular mechanisms and functional effects of SPOCD1 on the initiation and development of ESCC. Furthermore, we explored the expression of SPOCD1 in ESCC tissues, and the expression was validated by quantitative real-time polymerase chain reaction (qRT-PCR) in 21 ESCC patients. Together, the study provides evidence for the clinical and functional significance of SPOCD1, as well as its potential as a novel strategy for cancer treatment, providing momentous evidence for exploring the role of SPOCD1 in ESCC.

METHODS

Sample Collection and Pretreatment

Esophageal squamous cell carcinoma (ESCC) tissues (n = 6) and adjacent normal esophageal tissues (n = 6) were retrieved from surgical patients in Fujian Cancer Hospital from July 1st to 17 July 2020. All patients were not treated before surgery. The clinical data for RNA sequencing of patients with esophageal squamous cell carcinoma in our institution was shown in **Supplementary Table S1**. The specimens after surgical resection were immediately placed in liquid nitrogen and then transferred to the refrigerator at −80°C for preservation. All patients involved in the study received written informed consent for biological research. The research scheme (including specimen collection) was reviewed and approved by the Biomedical Ethics Committee of the Fujian Cancer Hospital (batch number: K2021-027-01). All procedures were in line with the guidelines of the Helsinki Declaration of the World Medical Association. The clinicopathological staging and classification of the patients were in accordance with the criteria of the American Joint Committee on Cancer (AJCC).

In addition, the count data, SNP data, CNV, and matched clinical data (n = 78) of ESCC RNA-seq were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/projects/) by GDC software. Due to the deficiency of normal control sample (n = 1) in ESCC in TCGA database, the RNA-seq count data of esophageal normal control samples (n = 663) combined with TCGA combined with GTEx and ESCC samples (n = 78) were obtained by UCSC Xena (https://xenabrowser.net/datapages/) (The data was corrected in batches by combat package). The analysis process was shown below (**Figure 1**).

RNA Sequencing

The above six pairs of esophageal tissues were used for the sequencing of RNA. The total RNA was extracted by RNA extraction kit (Qiagen). The samples were sequenced on Illumina HiSeq and produced an average of 42.38 million reads. We next used the TopHat2 (v2.1.1) to map them to the human reference genome (GRCh38/hg38). In addition, the splicing sequence of RNA were identified by the CIRC explorer program (v2.2.3) together with the fusion connection gained from the TopHat2. RNA quantification, quality evaluation, library construction, and sequencing were carried

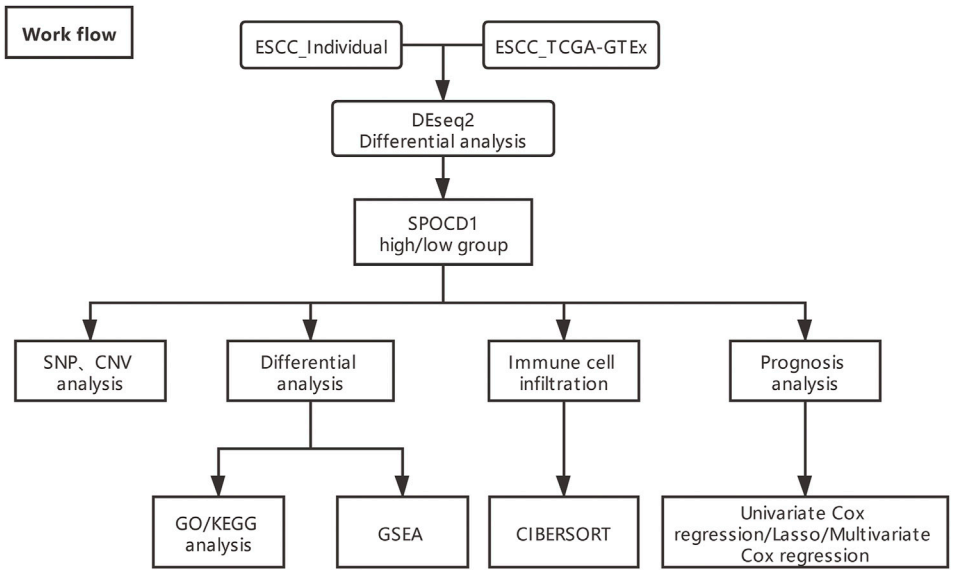


FIGURE 1 | The overall analysis flow chart (ESCC_Individual is self-test data). ESCC, esophageal squamous cell carcinoma; TCGA, The Cancer Genome Atlas; SPOCD1, SPOC Domain Containing 1; SNP, single nucleotide polymorphism; CNV, copy number variation; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene set enrichment analysis.

out by Shanghai Life genes Biotechnology Co., Ltd. Raw data in fastq format was processed through perl scripts. We also use HTSeq v0.6.1 (https://htseq.readthedocs.io/en/master/) calculation to get the read count and TPM of each gene.

Screening of Differential Genes Between Normal and Tumor Samples

The COUNT value of difference analysis of RNA sequencing data of TCGA, GTEx and our medical center were standardized before difference analysis. First of all, the principal component analysis (PCA) analysis of the sequencing data of our center and the TCGA combined with the GTEx data set was carried out through the FactoMine R package (Lê et al., 2008). Then Deseq2 R package (Love et al., 2014) was used to screen the differentially expressed genes between tumor and normal samples in our center and TCGA combined with GTEx data set. The differential genes met the requirements of p-value < 0.05 and | log2 Fold Change | > 1. The differential genes obtained from the two data sets were intersected, and the interested gene of SPOCD1 was selected for follow-up analysis. Finally, we employed ggpubr R package (ggpubr, 2017) to draw the expression box diagram of SPOCD1 in the two data sets.

Analysis of SPOCD1 Expression and Mutation Data

According to the expression of SPOCD1, the tumor samples of our center and TCGA-ESCC were divided into high SPOCD1 expression group and low SPOCD1 expression group, and the differential genes in the high expression group and low expression group were screened by the Deseq2 package. The differential genes satisfied the requirements of p-value < 0.05 and | log2 Fold Change | > 1. Besides, the heat map and volcano map were drawn by heatmap R package (Kolde, 2015) and ggplot2 R package (Ginestet, 2011), respectively, to show the overall expression and differential expression of SPOCD1-related genes.

The expression of SPOCD1 in various tissues of the human body was sorted out by GTEx database (https://gtexportal.org/home/), and visualized by ggplot2 R package. The somatic mutation data of TCGA-ESCC patients were extracted by maftools R package (Mayakonda and Koeffler, 2018). The somatic mutation data of patients with high expression of SPOCD1 were collected and analyzed to excavate the high frequency mutation genes of patients with high expression of SPOCD1. Subsequently, we identified statistically significant amplification and deletion in the high SPOCD1 expression group and the low SPOCD1 expression group by using genomic identification of essential targets in cancer (GISTIC) 2.0 (https://cloud.genepattern.org/gp/pages/index.jsf).

Related Drug Analysis of SPOCD1

The data of mRNA expression profile and drug activity of SPOCD1 gene were downloaded from CellMiner database (https://discover.nci.nih.gov/cellminer/). CellMiner is a web-based tool that contains genomic and pharmacological information for researchers to use transcripts and drug response data from NCI-60 cell lines compiled by the US. National Cancer Institute (Reinhold et al., 2012). The transcriptional expression levels of 22,379 genes, 360 microRNA, and 20,503 compounds in drug responses are available on the CellMiner website. We calculated the

correlation between SPOCD1 gene expression and compound sensitivity by Pearson correlation analysis.

In addition, we used the genomics of drug sensitivity in cancer (GDSC) database (www.cancerrxgene.org/) (Yang et al., 2013) and pRRophetic algorithm (Paul et al., 2014) to construct a ridge regression model based on GDSC cell line expression profile and TCGA gene expression profile, which contributes to predict the IC50 values of high and low SPOCD1 expression groups for common anticancer drugs.

Functional Enrichment Analysis

ClusterProfiler R package (Yu et al., 2012) was used to analyze the enrichment of differential genes by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), respectively. Adj. p-value < 0.05 was considered to be statistically significant. Pathview R package (Luo and Cory, 2013) was used to visualize the pathways of KEGG enrichment. The gene expression matrix was analyzed by the gene set enrichment analysis (GSEA) using clusterProfiler R package, and "c2. cp.kegg.v7.0. symbols.gmt" was selected as the reference gene set. False discovery rate (FDR) < 0.25.

Analysis of PPI Interaction Network and miRNA-Hub Gene Regulatory Network

The STRING protein-protein interaction database (https://www.string-db.org/) was performed to analyze the interaction between genes. After the results were derived, the hub genes were further screened by the CytoHubba plug-in in Cytoscape (Chin et al., 2014). In addition, Networkanalyst was used to regulate and analyze the Hub gene-miRNA, and the miRNA-target gene was predicted according to the "minimum number of network connections". Finally, the results were derived from Networkanalyst, and the miRNA-hub gene regulatory network was drawn by Cytoscape software.

Analysis of Immune Cell Infiltration and Correlation Analysis

CIBERSORT (http://CIBERSORT.stanford.edu/) and LM22 characteristic gene matrices were used to predict the proportion of 22 immune cells in all samples of the data set. We used CIBERSORT R package (Newman et al., 2015) to evaluate the abundance of 22 kinds of immune cells in the TCGA-ESCC data set and calculate the correlation between 22 kinds of immune cells. Then, by integrating the information of SPOCD1 expression, the Pearson correlation between the expression of SPOCD1 and immune cell infiltration was further calculated.

Prognostic Analysis and Model Construction of SPOCD1

Firstly, we analyze the survival of SPOCD1 in the TCGA-ESCC dataset through the survival R package (Therneau, 2012). In addition, to obtain SPOCD1-related prognostic genes, we conducted a preliminary screening by univariate Cox regression with a p-value < 0.1. Then glmnet R package (Friedman et al., 2010) was used to carry out lasso regression to further remove overfitted variables. Finally, independent prognostic factors were determined by multivariate cox regression analysis, and the results were visualized by a risk heat map. Among them, the formula for calculating the prognostic risk score was as follows: risk score = expression value of gene 1 *β1 + expression value of gene 2 *β2 +... + expression value of gene n * βn, where β is the regression coefficient obtained in the process of calculation. According to the results of multivariate cox analysis, independent prognostic factors were screened according to $p < 0.05$, and risk heat map and scatter map were drawn.

qRT-PCR Validation for the Expression of SPOCD1

Twenty-one ESCC tissues and adjacent tissues from ESCC patients were obtained from July 2020 to June 2021 in Fujian Cancer Hospital. The clinical data for qRT-PCR of patients with esophageal squamous cell carcinoma in our institution was shown in **Supplementary Table S2**. qRT-PCR was applied to verify the expression of the target SPOCD1 using twenty-one pairs of ECA and adjacent normoal esophageal tissues. The primers of SPOCD1 were listed in **Table 1**, purchased from BioSune (Shanghai, China). An RTIII All-in-One Mix with dsDNase (Monad Biotech Co., Ltd, Shanghai, China) was used to synthesize cDNA from 1 μg of total RNA. The qRT-PCR analyses were conducted on the StepOnePlus Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific Co., Ltd, US) by the Hieff ®qPCR SYBR® Green Master Mix, High Rox (Yeasen, Biotechnology Co., Ltd, Shanghai, China). The reaction was: 95°C for 10 min, then 41 cycles of 95°C for 15 s and 60°C for 1 min, last 95°C for 15 s. The reference gene was GAPDH, and the relative levels of gene expression were calculated by the $2^{-\Delta\Delta Ct}$ method.

Statistical Analysis

All tests were two-sided, and a p-value less than 0.05 was considered to indicate a statistically significant difference. The statistical analysis was performed with R software (version 4.0.2).

RESULTS

Screening of Differential Genes Between Normal and Tumor Samples

We sorted out the sequencing data of our center and the TCGA combined with the GTEx data set. The results of PCA analysis showed that the normal samples and tumor samples were completely separated in our center, and most of the normal samples and tumor samples were separated in the data set of TCGA and GTEx. The two data sets indicated a good state of sample separation (**Figures 2A,B**). We then used the Deseq2 R package to analyze the differences in count data. Finally, 4,435 differential genes were found in the sequencing data of our center, and 4,126 differential genes were found in the TCGA combined

TABLE 1 | The primers of SPOCD1 and GAPDH.

Primers	Forward primers(5′ to 3′)	Reverse primers(5′ to 3′)
qPCR-SPOCD1	CATGGACCAGACACTGACCC	CTCCCAGTCCTTGCAGATGT
qPCR-GAPDH	GGAGCGAGATCCCTCCAAAAT	GGCTGTTGTCATACTTCTCATGG

with the GTEx data set. After screening, the SPOCD1 gene was discovered to be differentially expressed in both data sets.

Consequently, we compared the difference of SPOCD1 between tumor samples and normal samples in the two data sets after selecting SPOCD1 as the target gene. It was revealed that the expression of SPOCD1 in tumors was significantly higher than that in normal tissues (**Figures 2C,D**). The clinical data of patients with esophageal squamous cell carcinoma in TCGA was listed in **Table 2**.

Related Gene Analysis of SPOCD1

Based on the expression value of SPOCD1, we divided the tumor samples from our center data and TCGA data set into high SPOCD1 expression group and low SPOCD1 expression group according to the median value. It was shown that a total of 1,335 differential genes were found in our center tumor data set and 1,295 differential genes were found in the TCGA tumor data set. The volcanic map depicted the expression level of differential genes, and the overall difference was displayed by the heat map (**Figures 3A–D**).

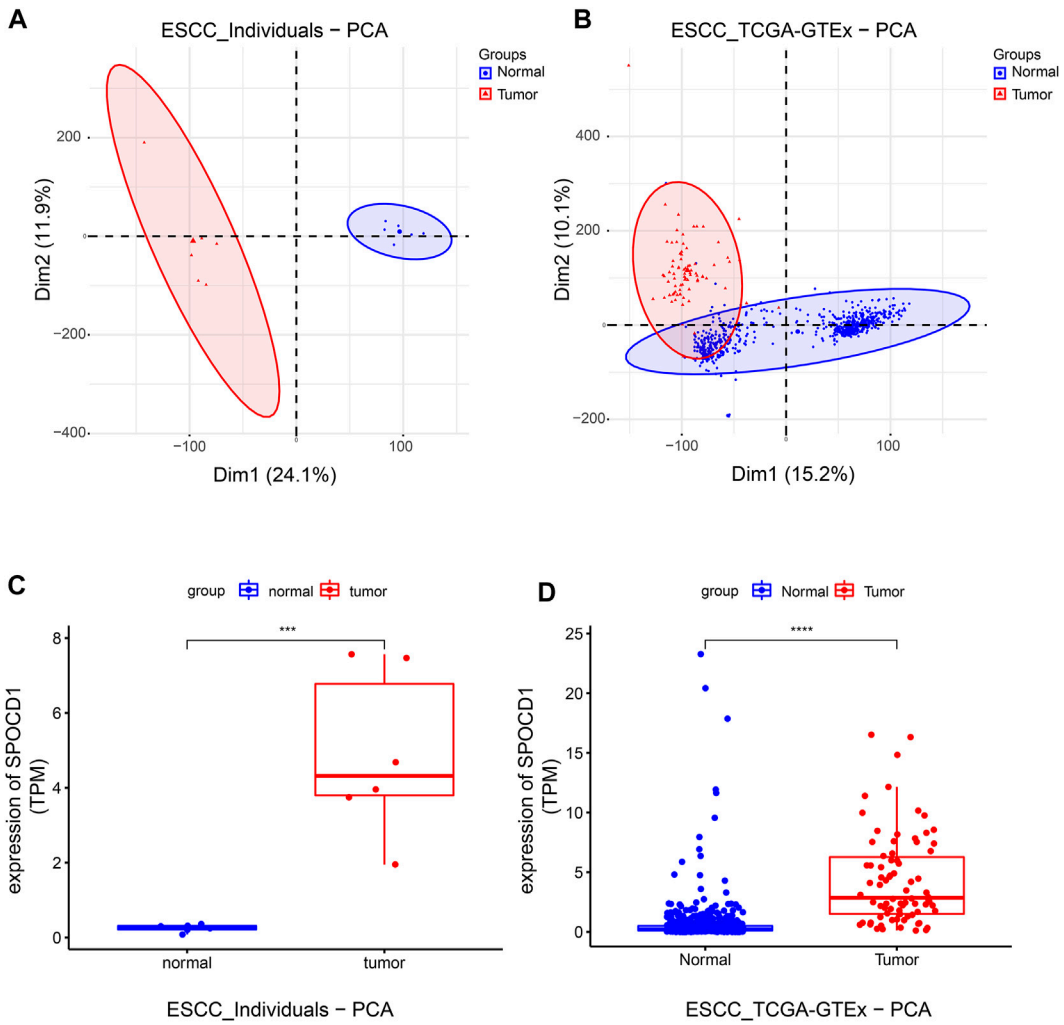


FIGURE 2 | The grouping of tumor and normal samples and the detection of SPOCD1 expression. **(A)** The results of PCA analysis of the sequencing data of our center samples; **(B)** The results of PCA analysis of the sequencing data of TCGA combined with GTEx dataset samples; **(C)** The sequencing data of our center showed that the expression of SPOCD1 in esophageal squamous cell carcinoma was significantly higher than that in the normal group; **(D)** The sequencing data of TCGA combined with GTEx data set showed that the expression of SPOCD1 in esophageal squamous cell carcinoma was significantly higher than that in the normal group. ***$p < 0.001$, ****$p < 0.0001$. SPOCD1, SPOC Domain Containing 1; ESCC, esophageal squamous cell carcinoma; TCGA, The Cancer Genome Atlas; PCA, principal component analysis.

TABLE 2 | Clinical data of patients with esophageal squamous cell carcinoma in TCGA.

Characteristic	Levels	Overall
N		95
Gender, n (%)	Female	15(15.79%)
	Male	80(84.21%)
Age, n (%)	≤60	56(58.95%)
	>60	39(41.05%)
Pathologic stage, n (%)	Stage I	14(14.74%)
	Stage II	59(62.11%)
	Stage III	14(14.74%)
	Stage IV	8(8.42%)
OS event, n (%)	Alive	77(81.05%)
	Dead	18(18.95%)

Correlation Analysis of SPOCD1 Expression and Mutation

To explore the expression characteristics of SPOCD1 in human tissues, we extracted the expression data of SPOCD1 from the GTEx database. The results showed that it was highly expressed in testis tissue and lung tissue (**Figure 4A**). As for ECSS, the tumor patients in TCGA-ESCC were divided into two groups: high expression of SPOCD1 group and low expression of SPOCD1 group. In the high expression group, we found apparent mutations in tumor protein P53 (TP53), Titin (TTN), and mucin 16 (MUC16) genes, which were 92, 36, and 18%, respectively (**Figure 4B**). Based on the mutation site of SPOCD1, we mapped the mutation information of SPOCD1 (**Figure 4C**). In addition, by collating the CNV information of TCGA-ESCC, we calculated the changes of CNV in the groups with high expression of SPOCD1 and low expression of SPOCD1 by the GISTIC 2.0 algorithm. As shown in **Supplementary Figure S1**, there was no significant change in CNV in the groups between high and low expression of SPOCD1.

Drug Sensitivity Analysis and Drug Prediction

In order to evaluate the potential effect of SPOCD1 on drug response, we analyzed the Pearson correlation between the expression of the SPOCD1 gene in the NCI-60 cell line and the activity of antineoplastic drugs retrieved from CellMiner database. According to the order of correlation, we selected the first eight drugs most related to AURKA (Aurora Kinase A). AURKA is a protein coding gene. The protein encoded by this gene is a cell cycle-regulated kinase that appears to be involved in microtubule formation and/or stabilization at the spindle pole during chromosome segregation. The encoded protein is found at the centrosome in interphase cells and at the spindle poles in mitosis. This gene may play a role in tumor development and progression. A processed pseudogene of this gene has been found on chromosome 1, and an unprocessed pseudogene has been found on chromosome 10. As shown in **Figure 5**, there was a positive correlation between SPOCD1 and Bieomycin, Cabozantinib, Rapamycin, Everolimus, Abiraterone, Zoledronoate, Temsirolimus, and Staurosporine (**Figures 5A–H**). In addition, we analyzed the IC50 value of drugs in high and low expression groups of SPOCD1 through the GDSC database. No significant difference was presented in common antineoplastic drugs such as cisplatin ($p > 0.05$) (**Supplementary Figure S2**), but there were significant differences in Bortezomib and Doxorubicin ($p < 0.05$) (**Figures 5I–P**).

Functional Enrichment Analysis

To further explore the function of SPOCD1, we intersected the differential genes obtained from the high and low SPOCD1 expression groups in the two data sets. Finally, 129 common differential genes were identified (**Figure 6A**). GO analysis showed that these differential genes were mainly associated with sulfur compound binding, heparin binding, and glycosaminoglycan binding (**Figure 6B**). The results of the KEGG analysis were shown in **Figure 6C**. The pathways of these differential genes were enriched mainly including the Estrogen signaling pathway, Cytokine-cytokine receptor interaction, Ras signaling pathway, and *Staphylococcus aureus* infection. At the same time, we further demonstrated the pathway of hsa04060: Cytokine-cytokine receptor interaction, which enriches most genes (**Figure 6D**). Lastly, the pathway of GSEA enrichment mainly referred to Rickman head and neck cancer, Atgttaa miR302c, Bosco epithelial differentiation module, Benporath PRC2 Targets, Anastassiou Multicancer invasiveness signature, GO adaptive immune response, GO adaptive immune response Based on somatic, and GO antigen binding (**Figures 7A–H**). NES stands for normalized Enrichment score. Because the ES is calculated according to whether the gene in the analyzed dataset appears in a functional gene set, but the number of gene contained in each functional gene set is different, and the correlation between different functional gene set and data is also different. Therefore, it is necessary to standardize the ES to compare the enrichment degree of dataset in different functional gene set. That is, the NES = ES of a function gene set/the average ES of all random combinations of the dataset. NES is the main statistic. NES >0, this pathway is enriched in high expression group. NES <0, this pathway is enriched in low expression group. The GO, KEGG, and GSEA enrichment analysis results of differential genes were shown in **Tables 3–5**, respectively.

Protein-Protein Interaction Network and miRNA-Target Gene Regulatory Network

We constructed a PPI network based on the String database to reveal the potential relationship between differential genes. When the minimum interaction score was 0.4, only 47 of the 129 differential genes interacted with other gene pairs (**Figure 8A**). The PPI network consists of 47 characteristic genes and 51 edges, and the average node degree was 0.93. Then we further identified the most relevant genes in the PPI network through the Cytohubba plug-in. We also used the cytohuabba plug-in in cytoscape to screen for differential genes. After screening, Cytohubba identified 10 genes that can be regarded as hub genes: POMC, PENK, TH, NGFR, CXCL11, MMP9, KRT24, KRT40, KRT23, and KRT78 (**Figure 8B**). Then, we predicted the 10 potential miRNA regulating hub genes through the Networkanalyst database,

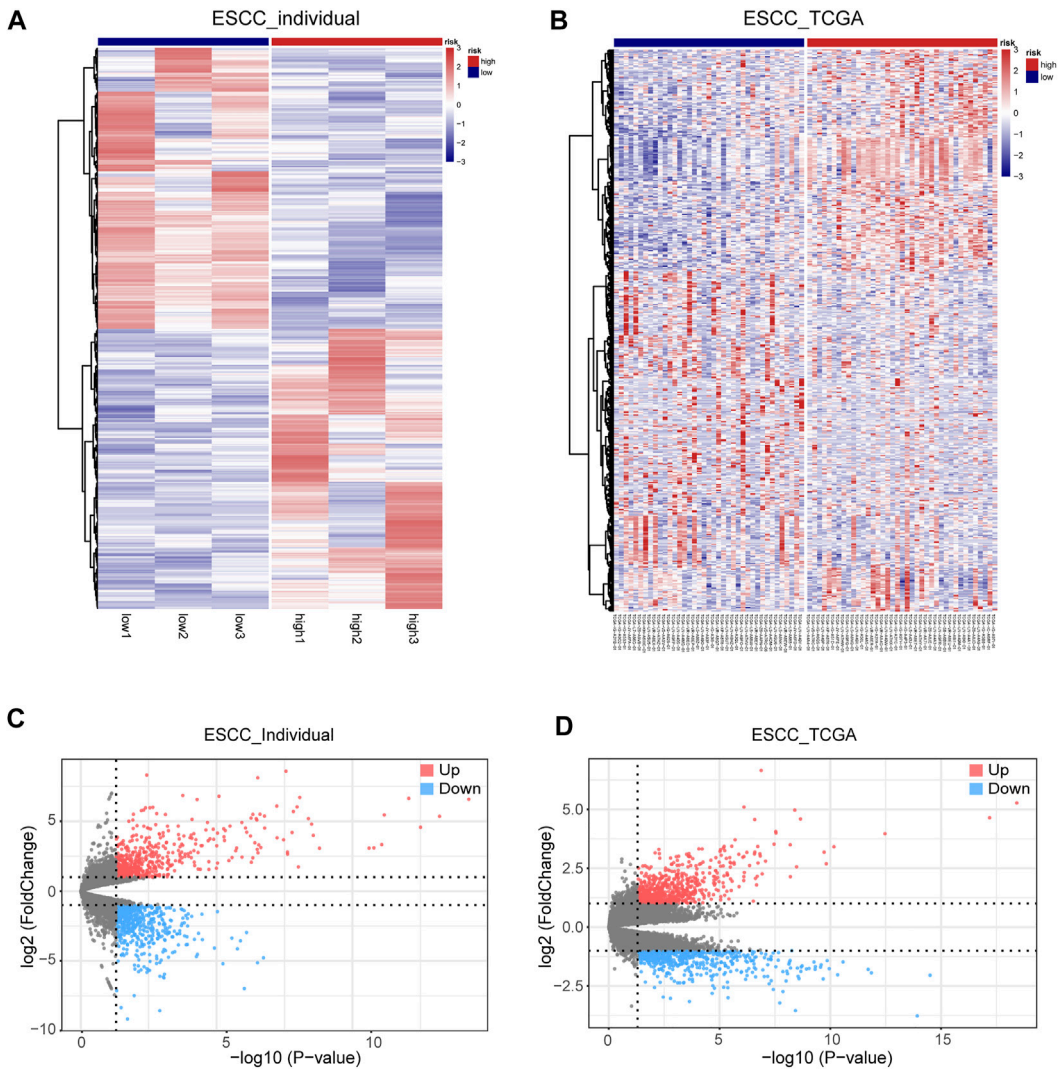


FIGURE 3 | The condition of differential expression of high SPOCD1 expression group and low SPOCD1 expression group. **(A,B)** The heat map of differentially expressed genes in high and low SPOCD1 expression was divided into two groups according to the median value of SPOCD1. Red represents upregulated differential genes, and blue represents downregulated differential genes; **(C,D)** The volcano map of differential genes showed that red represents upregulated differential genes, blue represents downregulated differential genes, and gray represents non-differential genes. SPOCD1, SPOC Domain Containing 1; ESCC, esophageal squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

and the screening condition was "Trim current network to minimum connected network" (minimum link network). Finally, four miRNA regulatory subnetworks are obtained, including 4 nodes, 89 edges, and 87 seeds (**Figure 8C**).

Immune Cell Infiltration and Correlation Analysis

To analyze the relationship between the expression of SPOCD1 and immune cell infiltration, we calculated the proportion of immune cell infiltration in the tumor microenvironment by the CIBERSORT algorithm. **Figure 9A** and **Figure 9B** demonstrated the panorama of immune cell infiltration in the tumor microenvironment of ESCC and the correlation of immune cell score, respectively. After integrating SPOCD1 gene expression and immune cell infiltration score, we comprehensively analyzed the infiltrating immune cells, which were significantly correlated with SPOCD1 expression. The results showed that SPOCD1 expression was positively correlated with Macrophages M0 and Mast cells activated cells, and negatively correlated with plasma cells and T cells follicular helper cell infiltration (**Figure 9C**).

Prognosis Analysis of SPOCD1

To assess the prognostic value of SPOCD1, the survival information of patients was obtained from the TCGA-ESCC dataset. It was found that the prognosis of the group with high expression of SPOCD1 was significantly worse than that of the group with low expression of SPOCD1 (**Figure 10A**). To

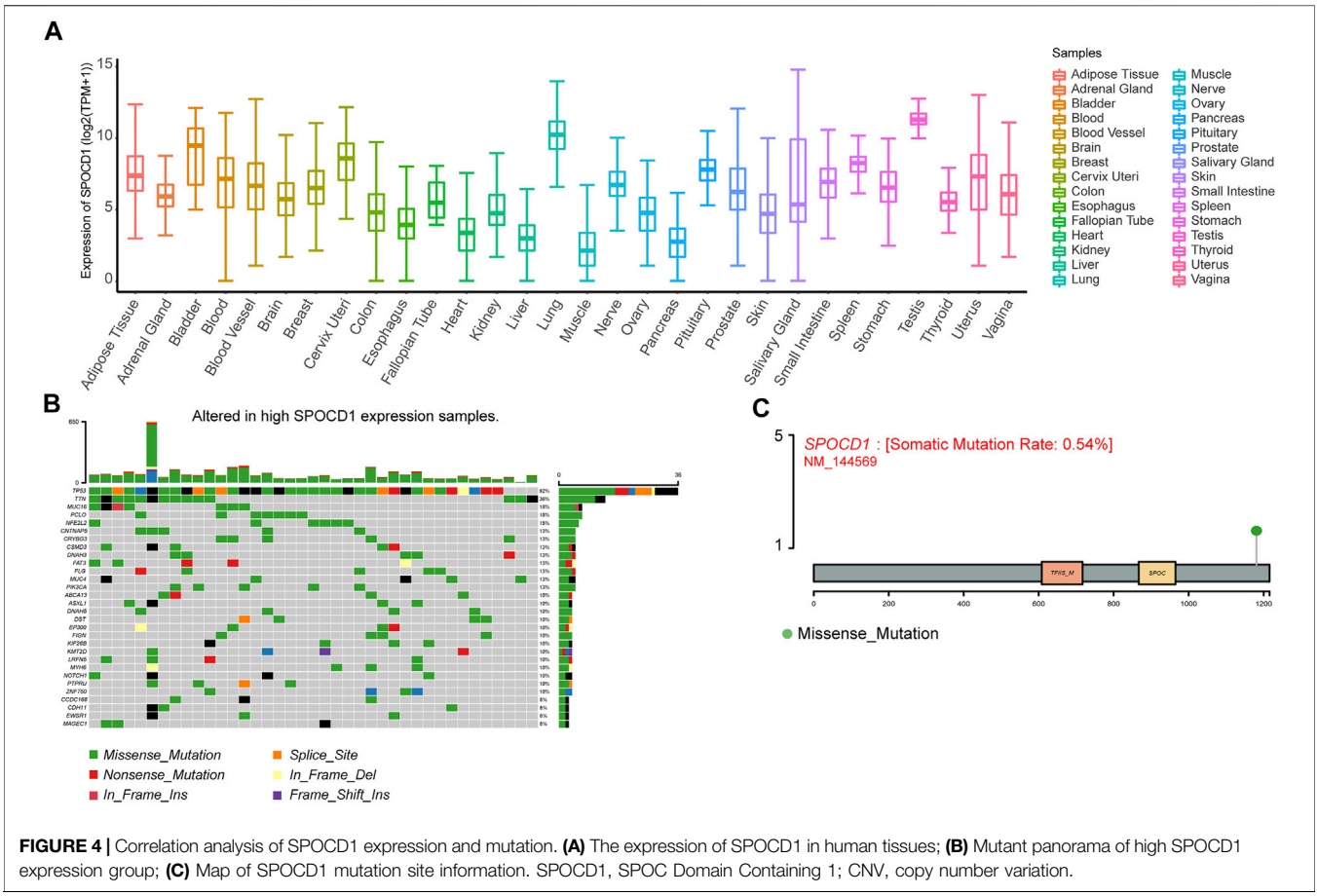


FIGURE 4 | Correlation analysis of SPOCD1 expression and mutation. **(A)** The expression of SPOCD1 in human tissues; **(B)** Mutant panorama of high SPOCD1 expression group; **(C)** Map of SPOCD1 mutation site information. SPOCD1, SPOC Domain Containing 1; CNV, copy number variation.

analyze the prognostic genes associated with SPOCD1, we divided the patients into high expression groups and low expression groups. Univariate cox analysis showed that 17 genes satisfied the condition of $p < 0.1$. Afterwards, lasso regression was used to screen again to remove the factors of over-fitting. The final eight genes were identified to meet the conditions. The eight genes were AJAP1, APOC1, CT83, GTSF1, KLRC2, KRT23, MYCL, and STRA6 (**Figures 10B,C**). Finally, multivariate cox analysis showed that only MYCL, AJAP1, APOC1, and KRT23 were independent prognostic factors. The prognosis of patients with the low-risk score was significantly better than those with the high-risk (**Figure 10D**). Although the survival group and death group were not statistically significant ($p > 0.05$), the risk score of the death group was significantly higher than that of the survival group (**Figure 10E**). As a result, we presented the expression and prognostic risk scores of the four genes in the form of a heat map and scatter map (**Figure 10F**). The result of SPOCD1 related prognostic genes was shown in **Table 6**.

Validation of the Expression of Target SPOCD1 in ESCC

Through qRT-PCR, the SPOCD1 expression in ESCC tissues was verified to be significantly higher than that of adjacent tissues using paired sample *t* test and Wilcoxon rank sum test in **Figures 11A,B** (both $p < 0.001$).

DISCUSSION

Although the treatment of ESCC is developing rapidly, the prognosis of ESCC patients remains unsatisfied in recent years (Pasquali et al., 2017; Baba et al., 2018). ESCC remains a severe burden on the health system in the world. This is mainly due to the high heterogeneity of ESCC, which leads to significant differences in patients' responses to treatment (Marshall and Djamgoz, 2018). Even if the patients receive similar treatment and are at the same stage, the clinical results and prognosis are diverse (Navin et al., 2011; Gerlinger et al., 2012). With the continuous development of bioinformatics, a wide range of disease prediction and molecular mechanisms have been gradually acknowledged. Although several studies have been concentrated on this field in the past few years, finding efficient ESCC biomarkers is still an important issue. Therefore, it is critical to identify novel, reliable diagnostic, and targeted therapeutic molecular biomarkers for ESCC, which will improve the effectiveness of diagnosis and treatment, and even increase our understanding of the pathogenesis mechanisms.

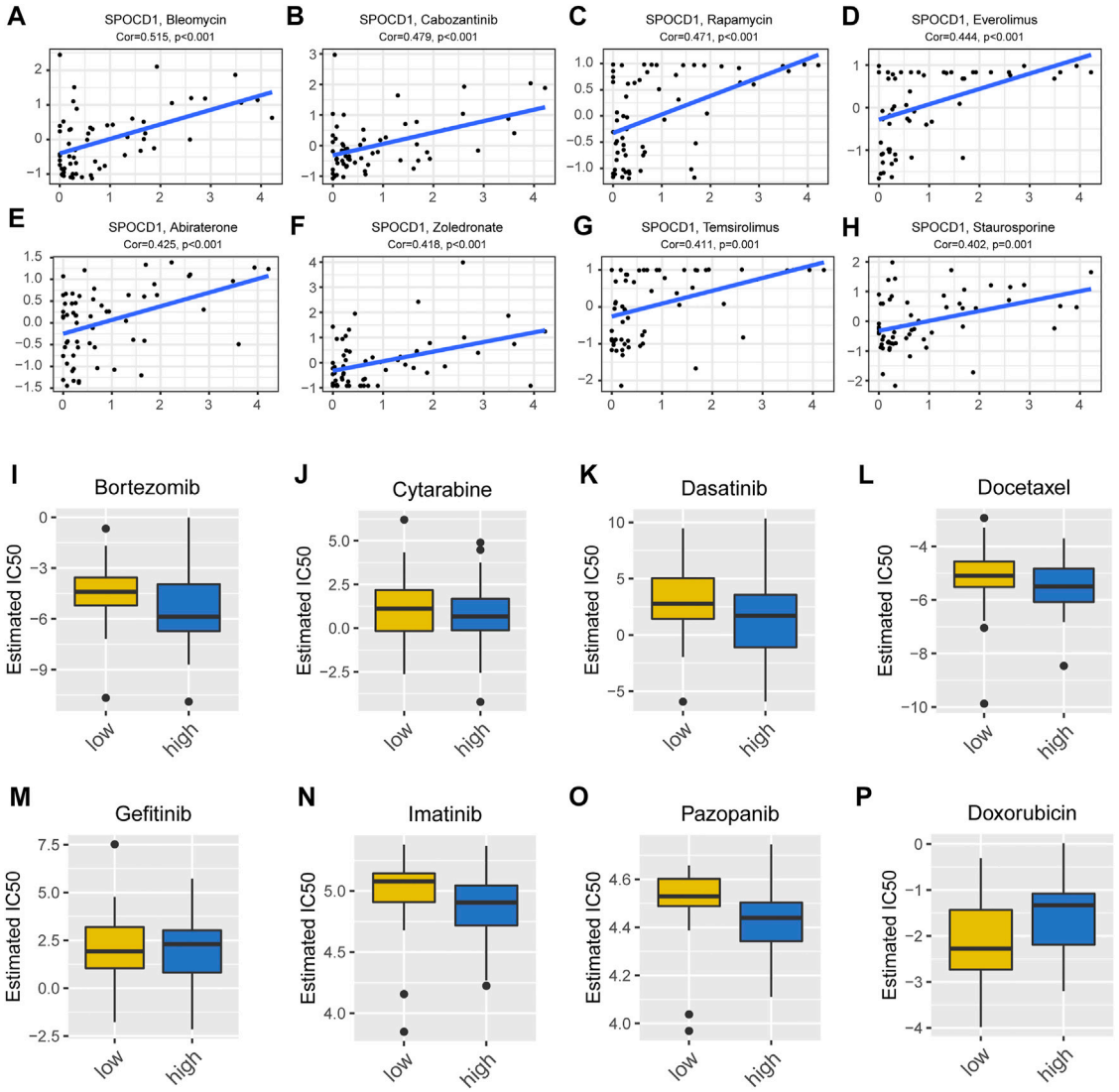


FIGURE 5 | Drug sensitivity analysis and drug prediction. **(A–H)** The top eight drugs with the most significant correlation with SPOCD1 in CellMiner database; **(I–P)** The differences in IC50 values for drugs prediction with high and low SPOCD1 expression from the GDSC database. SPOCD1, SPOC Domain Containing 1; GDSC, genomics of drug sensitivity in cancer.

In recent years, many studies have revealed that the expression of SPOCD1 was significantly related to the development and occurrence of the tumor, and they may be used as ideal biomarkers of the tumor. SPOCD1 belongs to the TFIIS family that participates in the regulation of development (Kimura et al., 2006). Emerging evidence has indicated that the expression of SPOCD1 is a prognostic factor in several tumors, including gastric cancer, clear cell renal cell carcinoma, ovarian cancer, osteosarcoma, and glioma (Zhu et al., 2017; Liang et al., 2018; Liu et al., 2018; Sakaguchi et al., 2018; Liu et al., 2020). It has been revealed that SPOCD1 can be used to distinguish patients with progressive and non-progressive bladder cancer (van der Heijden et al., 2016). Besides, SPOCD1 promotes the metastasis and proliferation of glioma cells *via* PTX3 (Liu et al., 2018). It also inhibits cell apoptosis and promotes cell proliferation in osteosarcoma through VEGF-A (Liang et al., 2018). Knockout of it decreased gastric cancer invasive activity, cell proliferation, and migration (Zhu et al., 2017). Nevertheless, whether SPOCD1 plays a role in ESCC remains unclear. Herein, we provided a comprehensive analysis of SPOCD1 to illustrate the potential clinical significance and roles in ESCC.

The purpose of this study was to explore the potential role and mechanism of SPOCD1 in the pathogenesis of ESCC through our own sequencing data, ESCC tissue and normal esophageal tissue of TCGA and GTEx dataset. In the present study, 4,435 differential genes were found in the sequencing data of our center, and 4,126 differential genes were found in the TCGA combined with the GTEx data set. After screening, we found that the SPOCD1 gene was differentially expressed in both data sets.

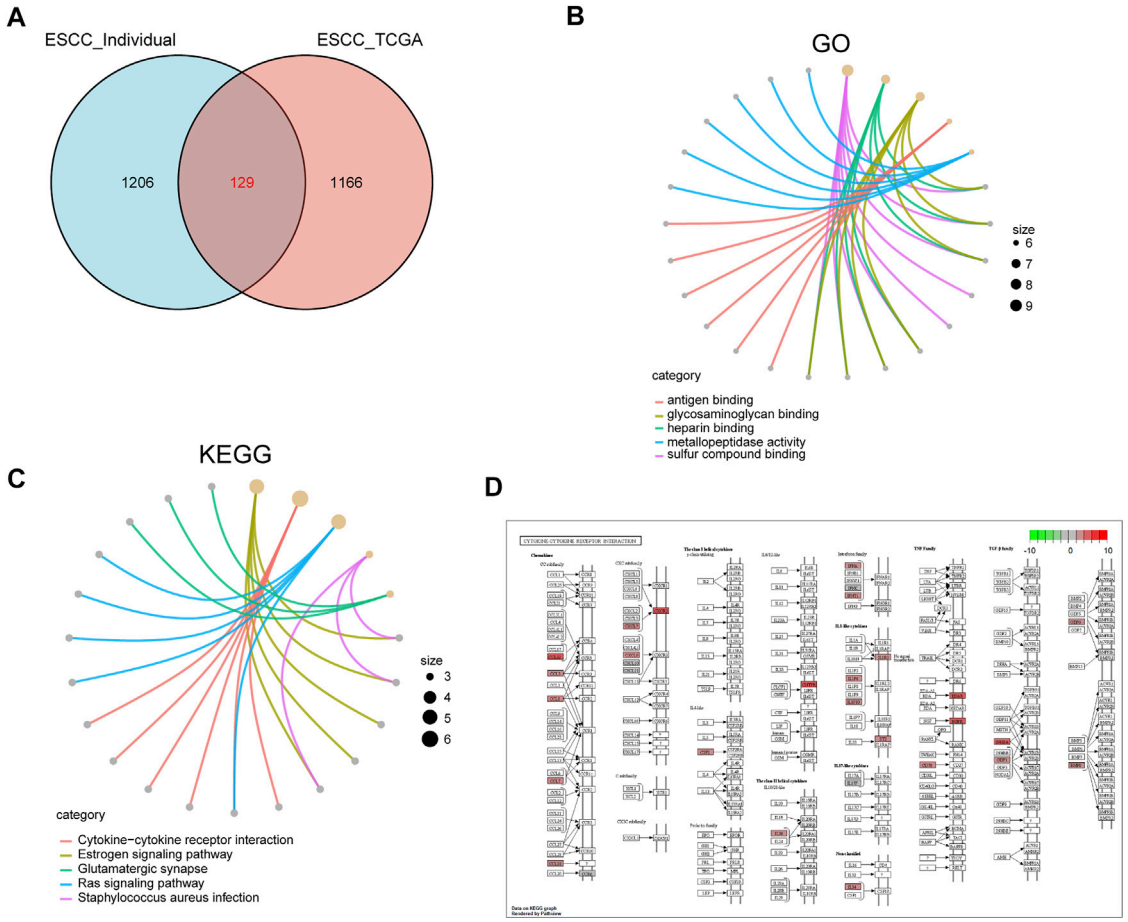


FIGURE 6 | GO and KEGG analysis of differentially expressed genes in high and low SPOCD1 expression groups. **(A)** The differentially expressed genes of high and low SPOCD1 expression group in the sequencing data of our center and TCGA data set were intersected; **(B)** GO biological function enrichment analysis, the size of the dot represents the proportion of gene enrichment in this function; **(C)** KEGG pathway enrichment analysis, the annotation is the same as before; **(D)** Cytokine-cytokine receptor interaction, the pathway of significant enrichment of KEGG; the red represents the degree of gene enrichment, the redder the color, the more obvious the upregulation of the gene in the pathway; the greener the color, the more obvious the downregulation of the gene in the pathway. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; SPOCD1, SPOC Domain Containing 1; TCGA, The Cancer Genome Atlas.

Therefore, we selected SPOCD1 as the target gene and used bioinformatics methods to perform the mutation analysis, drug sensitivity analysis, function analysis, pathway analysis, co-expression network analysis, and immune cell infiltration analysis between patients with low and high expression of SPDCD1. Finally, we explored the prognostic significance of SPOCD1.

Our results indicated that the expression of SPOCD1 was significantly increased in ESCC tissues in our independent verification tests and TCGA and GTEx datasets. Interestingly, the same results can be confirmed in other studies. Several studies revealed that SPOCD1 was highly expressed in glioma and gastric cancer (Zhu et al., 2017; Liang et al., 2018). Notably, the analysis of the results showed that the overexpression of SPOCD1 was associated with advanced clinicopathological features and poor outcomes. On the contrary, the expression of SPOCD1 was downregulated in bladder cancer (van der Heijden et al., 2016). The inconsistent expression of SPOCD1 in different tissues may mean that the transcriptional spectrum of different tissue types is different. Given its biological importance in other areas, previous studies and our current study might indicate that SPOCD1 plays a broad regulatory role in human carcinogenesis. These findings supported the suggestion that SPOCD1 might be a promising molecular target for the diagnosis and prognosis of ESCC patients.

It is known to all that SPOCD1 is an oncogene in several tumor types, and plays an essential role in the occurrence and development of tumors. However, its functional way and molecular mechanism in ESCC remain to be elucidated. Therefore, we analyze the function and pathway analysis of SPOCD1. GO and KEGG enrichment analysis of SPOCD1 and its co-expressed genes demonstrated that it may act as an ESCC oncogene by regulating the genes expression in the essential functions and pathways of tumorigenesis, such as glycosaminoglycan binding, Cytokine-cytokine receptor interaction, and Ras signaling pathway. Cytokines can limit

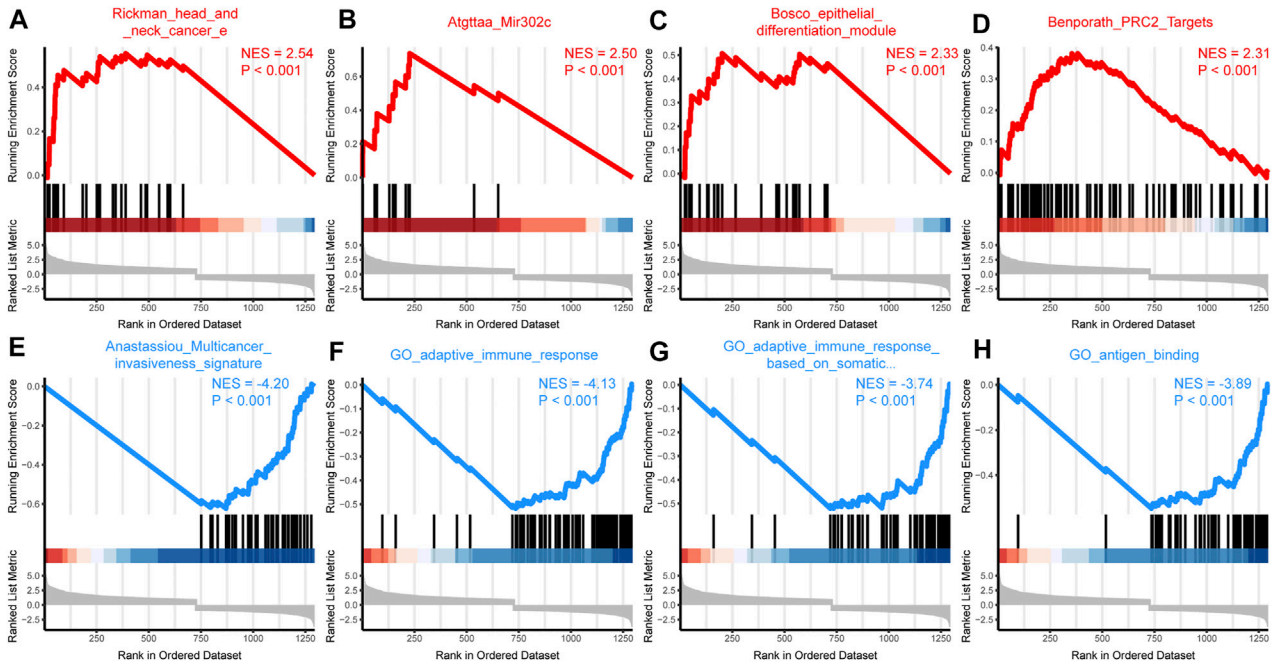


FIGURE 7 | GSEA enrichment analysis. **(A–D)** According to the NES value, the differentially expressed genes were enriched in the first four pathways in the high SPOCD1 expression group. The NES value represents the normalized enrichment fraction, and the higher the NES value is, the more genes are enriched in this pathway. The *p*-value reflects the credibility of the enrichment results; **(E–H)** According to the NES value, the differentially expressed genes were enriched in the first four pathways in the low SPOCD1 expression group. The NES value represents the normalized enrichment fraction, and the higher the NES value is, the more genes are enriched in this pathway. The *p*-value reflects the credibility of the enrichment results. GSEA, gene set enrichment analysis; NES, normalized enrichment score; SPOCD1, SPOC Domain Containing 1.

TABLE 3 | Results of GO enrichment analysis of differential genes.

ID	Description	p.adjust	geneID	Count
GO:1,901,681	sulfur compound binding	0.01	COL5A3/FGFR1/SMOC1/GSTM3/GSTM2/PLA2G2D/FGF9/CCL8/CXCL11	9
GO:0008201	heparin binding	0.01	COL5A3/FGFR1/SMOC1/PLA2G2D/FGF9/CCL8/CXCL11	7
GO:0005539	glycosaminoglycan binding	0.03	COL5A3/FGFR1/SMOC1/PLA2G2D/FGF9/CCL8/CXCL11	7
GO:0003823	antigen binding	0.03	IGHV3-72/IGLV3-21/IGHV3-66/LILRB4/KLRC2/IGHV4-34	6
GO:0008237	metallopeptidase activity	0.04	MMP20/CPA6/VASH2/MMP12/MMP9/MMP8	6

TABLE 4 | Results of KEGG enrichment analysis of differential genes.

ID	Description	*p*-value	Count
hsa04915	Estrogen signaling pathway	0.00139	5
hsa04060	Cytokine-cytokine receptor interaction	0.00827	6
hsa04014	Ras signaling pathway	0.01258	5
hsa05150	Staphylococcus aureus infection	0.02001	3
hsa04724	Glutamatergic synapse	0.03124	3
hsa04930	Type II diabetes mellitus	0.03128	2
hsa04010	MAPK signaling pathway	0.03143	5
hsa04973	Carbohydrate digestion and absorption	0.03254	2
hsa00520	Amino sugar and nucleotide sugar metabolism	0.03383	2
hsa04015	Rap1 signaling pathway	0.03741	4
hsa05207	Chemical carcinogenesis - receptor activation	0.03853	4
hsa00480	Glutathione metabolism	0.04626	2

TABLE 5 | Results of GSEA enrichment analysis.

Description	setSize	NES	p.adjust	Rank
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	69	−4.221612585	**2.86E-08**	525
ANASTASSIOU_MULTICANCER_INVASIVENESS_SIGNATURE	42	−4.197893077	2.86E-08	426
GO_ADAPTIVE_IMMUNE_RESPONSE	77	−4.126203037	2.86E-08	561
POOLA_INVASIVE_BREAST_CANCER_UP	58	−4.04773592	2.86E-08	540
BENPORATH_PRC2_TARGETS	76	2.30528993	0.000990457	382
BOSCO_EPITHELIAL_DIFFERENTIATION_MODULE	27	2.325602524	0.00687682	201
ATGTTAA_MIR302C	11	2.501002207	0.000703345	227
RICKMAN_HEAD_AND_NECK_CANCER_E	27	2.54007954	0.000990457	389

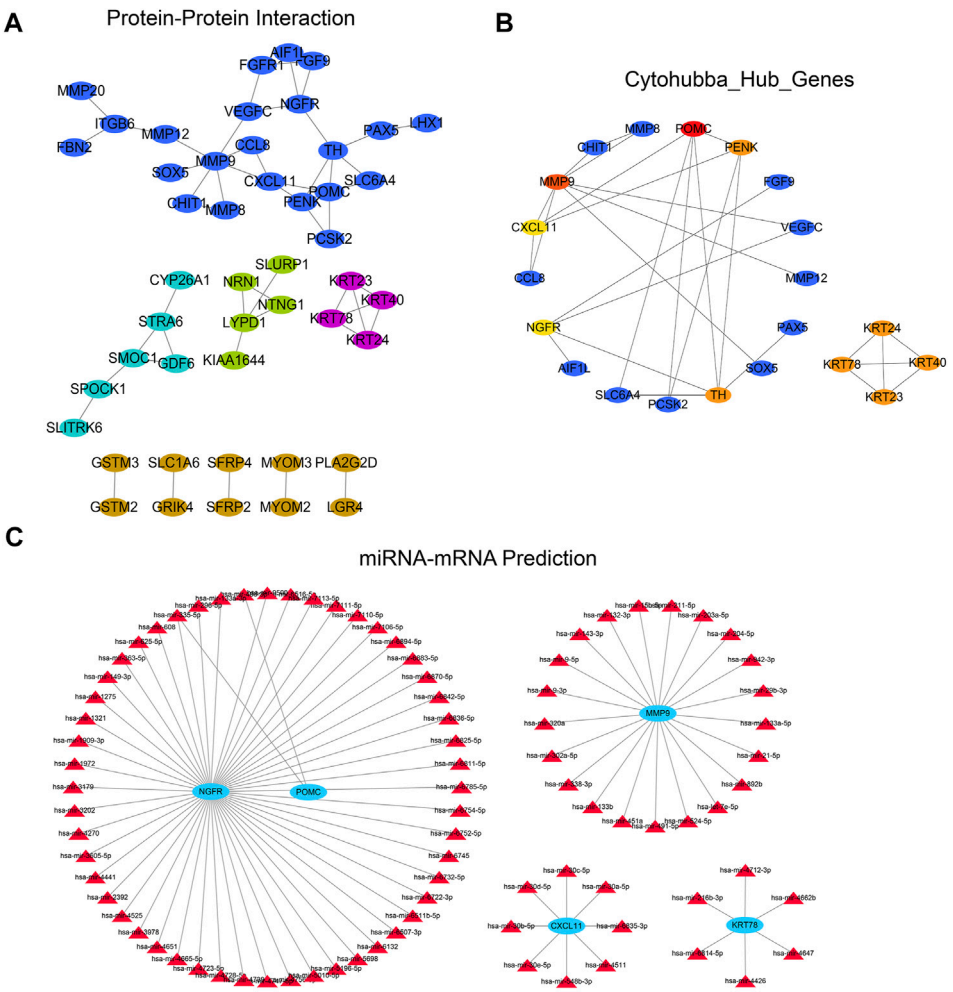


FIGURE 8 | PPI and miRNA regulatory network analysis of differential genes. **(A)** PPI regulatory network established by differential genes; **(B)** Two subnetworks and 10 hub genes were further screened by Cytohubba plug-in; **(C)** According to the miRNA-hub gene regulatory network predicted by the Networkanalyst database to regulate hub genes. The Networkanalyst database predicted the miRNA-hub gene regulatory network of regulatory hub genes.PPI, protein-protein interaction.

the growth of tumor cells through pro-apoptotic or anti-proliferation activity, or by stimulating the activity of immune cells to tumor cells (Berraondo et al., 2019). Ras is involved in the regulation of intracellular signaling pathways, which are involved in fundamental cellular processes such as cell polarity, growth, differentiation, migration, and apoptosis, and eventually lead to cancer (Shields et al., 2000; McKay and Morrison, 2007). Furthermore, we also carried out GSEA to investigate the enriched gene sets and critical pathways. Interestingly, the result showed that the

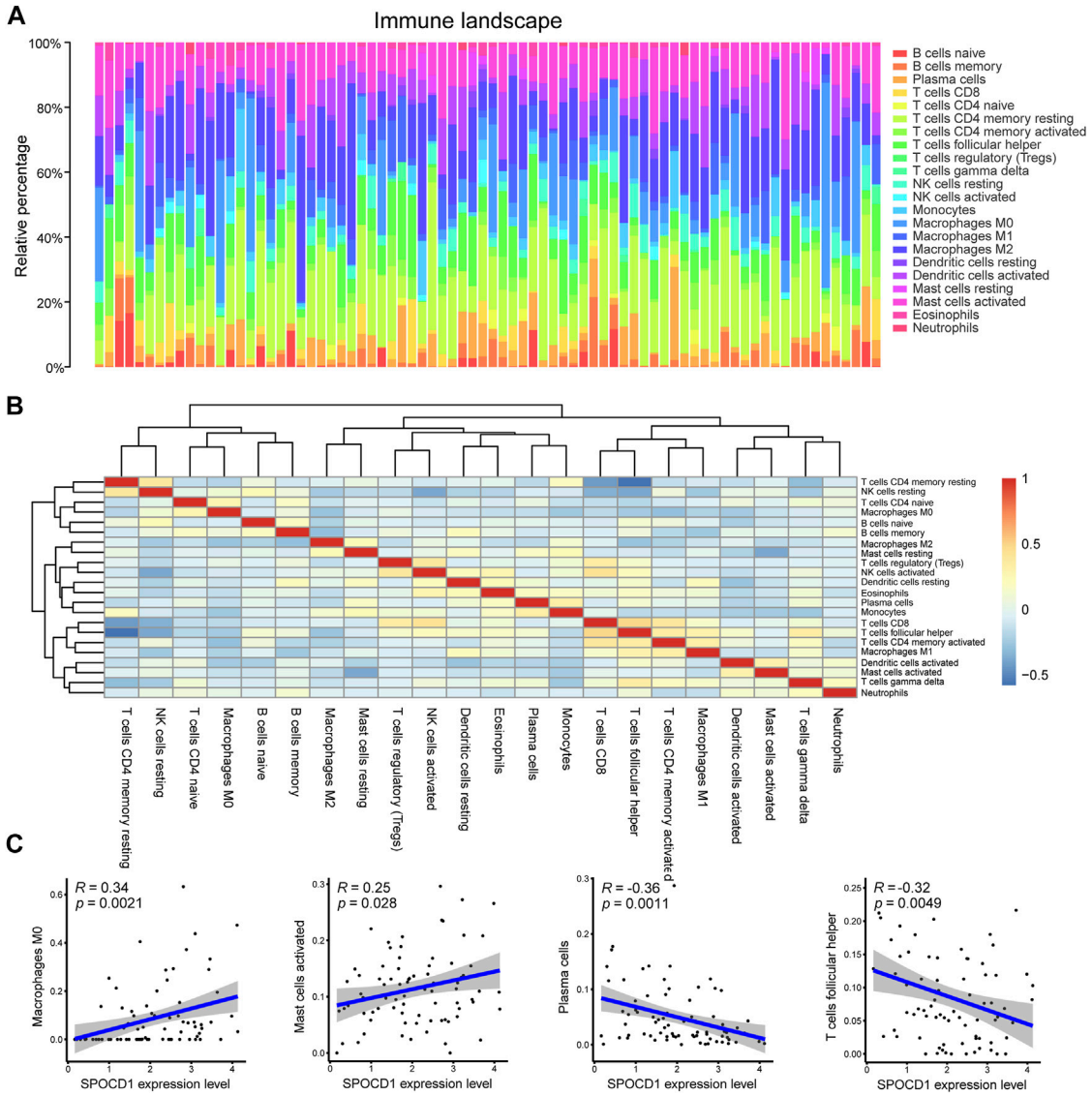


FIGURE 9 | Evaluation and correlation analysis of immune cell infiltration. **(A)** Panoramagram of 22 kinds of immune cell infiltration in TCGA-ESCC tumor samples; **(B)** Correlation heat map of 22 kinds of immune cell score; **(C)** Correlation between SPOCD1 gene expression and immune cell infiltration. TCGA, The Cancer Genome Atlas; ESCC, esophageal squamous cell carcinoma; SPOCD1, SPOC Domain Containing 1.

pathway of GSEA enrichment mainly included miR-302c, PRC2 Targets, Multicancer invasiveness signature, adaptive immune response, and adaptive immune response Based on somatic, which was similar to previous findings. Previous studies have indicated that miR-302c expression was significantly correlated with overall survival of locally advanced adenocarcinomas of the gastroesophageal junction (Odenthal et al., 2015). In addition, some literature also demonstrated that PRC2 represses tumor suppressor genes and promotes tumorigenesis (Ha and Kim, 2012; Hu et al., 2012). In the study, multicancer invasiveness signature and adaptive immune response was correlated with the expression of SPOCD1, which revealed that SPOCD1 may play a crucial role in immune response and modulating cancer invasion in ESCC.

To further explore the role of SPOCD1 in tumorigenesis, we also carried out drug sensitivity and immune cell infiltration analysis. The results revealed that SPOCD1 expression was positively correlated with Macrophages M0 and Mast cells activated cells, and negatively correlated with plasma cells and T cells follicular helper cell infiltration. Our functional analysis indicated that the pathway that enriches the most genes was Cytokine-cytokine receptor interaction. Cytokine is the primary regulator of the immune system, and it is also an effective but complex immune mediator. Cytokines can enlarge and activate immune cells and promote the invasion

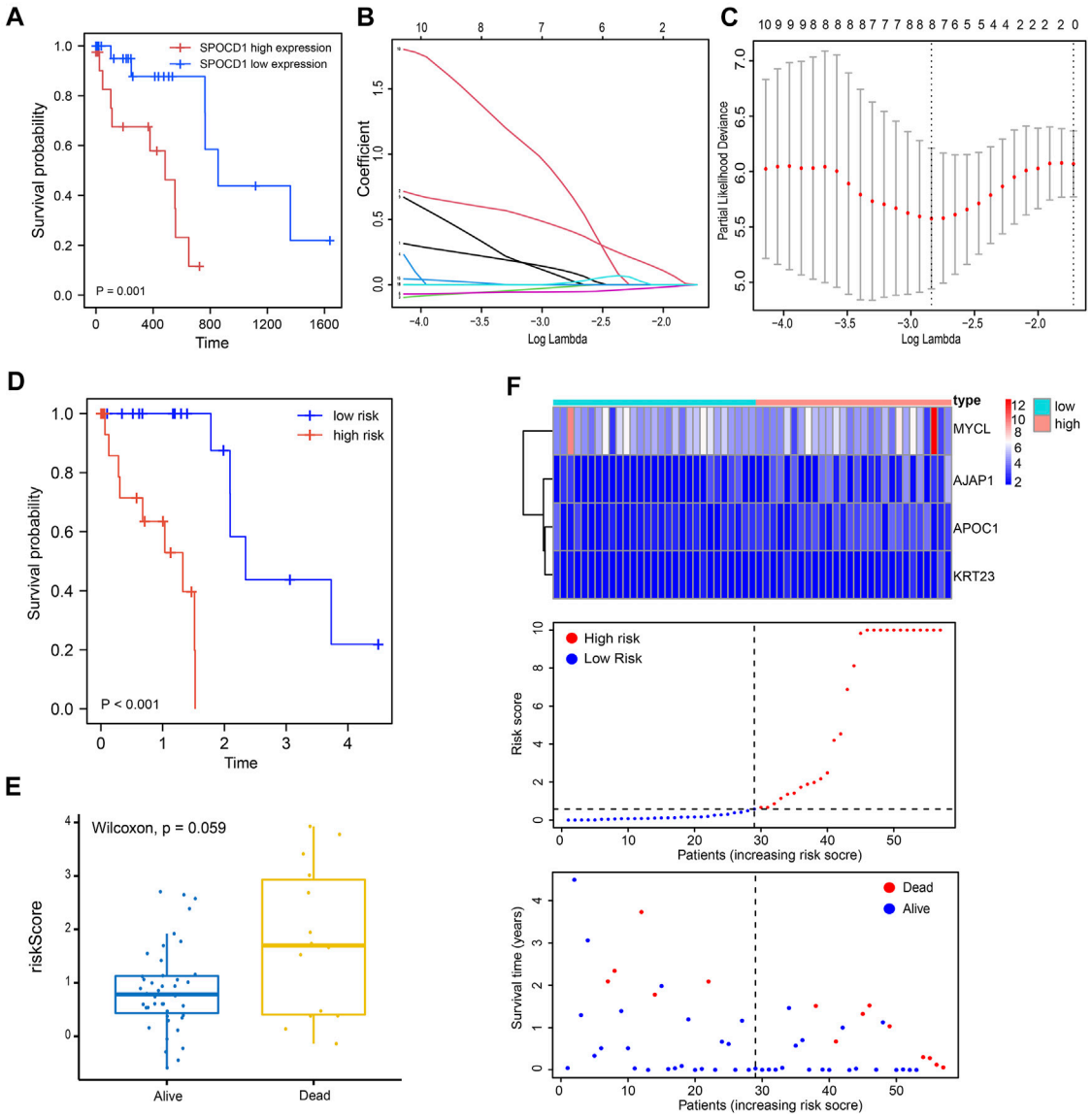


FIGURE 10 | Prognosis correlation analysis of SPOCD1. **(A)** The prognosis of the low SPOCD1 expression group was significantly better than that of the high expression group; **(B,C)** From Lasso regression to multivariate Cox regression: λ and cv maps showed that eight factors were screened; **(D)** The prognosis of patients with the high-risk score was worse than that of patients with low risk; **(E)** The risk score of the death group was significantly higher than that of the survival group; **(F)** Risk heat map and scatter map of independent prognostic genes in multivariate Cox regression. SPOCD1, SPOC Domain Containing 1.

of immune cells to tumors. The manufacture of cytokine-based drugs is a daunting challenge, and an in-depth understanding of the biological role of cytokines is needed to take advantage of their anti-tumor activity while minimizing toxicity.

In the high SPOCD1 expression group, we found certain mutations in TP53, TTN, and MUC16 genes, which were 92, 36, and 18%, respectively. TP53 is a well-known tumor-associated gene for its ability to regulate the malignancy of ESCC cells. Previous studies have shown that TTN mutations are associated with a better response of solid tumors to immune checkpoint inhibitors (Jia et al., 2019). Besides, MUC16 mutation is associated with prognosis and maybe a site affecting tumor prognosis and progression (Yang et al., 2020). In our study, ESCC patients with high SPOCD1 expression indicated poor overall survival, suggesting the expression of SPOCD1 as an underlying factor for the outcome of patients with ESCC. In addition, the mutations that mediate the expression of SPOCD1 possibly have an influence on the development of ESCC. Hence, inhibiting the expression of SPOCD1 might improve the prognosis and therapeutic efficiency of patients with ESCC.

Our multivariate cox analysis showed that only MYCL, AJAP1, APOC1, and KRT23 were independent prognostic

TABLE 6 | Univariate Cox regression, lasso regression, and multivariate Cox regression were used to identify SPOCD1 related prognostic genes.

Gene	Univariate cox		Lasso	Multivariate cox	
	HR	*p*-Value	Coef	Coef	*p*-Value
MYO7B	1.22	0.00	—	—	—
MYCL	1.00	0.01	0.00	0.00	0.01
ITGB6	1.00	0.01	—	—	—
APOC1	1.21	0.02	0.44	1.25	0.00
TRPM1	22.86	0.02	—	—	—
SPOCK1	1.07	0.02	—	—	—
GALNT6	1.01	0.03	—	—	—
HK3	1.19	0.03	—	—	—
SLITRK6	1.08	0.03	—	—	—
FGF9	1.18	0.03	—	—	—
STRA6	1.17	0.03	0.01	—	—
KLRC2	1.72	0.04	0.06	1.60	0.11
AJAP1	1.13	0.05	0.11	0.49	0.00
MMP12	1.10	0.06	—	—	—
GTSF1	0.87	0.07	−0.05	—	—
CT83	0.88	0.09	−0.02	−0.33	0.07
KRT23	1.80	0.10	0.78	2.96	0.00

factors. MYCL is an oncogene deregulated in human cancers, which supports tumorigenic progression and processes. As therapeutic target, it has been found to be amplified and overexpressed in some malignancies, including gastric cancer and lung cancer (Albihn et al., 2010; Chen et al., 2015; Masso-Valles et al., 2020). Besides, the transmembrane adherens junctions-associated protein-1 (AJAP1) targets the membrane of epithelial cells. Previous research showed that AJAP1 is an independent prognostic factor of squamous cell carcinoma of the esophagus. In ESCC, AJAP1 might serve as a tumor suppressor and that AJAP1 transcription is modulated by hypermethylation (Tanaka et al., 2015). Several researches revealing APOC1 to be a diagnostic and prognostic marker for gastric cancer and colorectal cancer (Yi et al., 2019; Shen et al., 2021). Finally, keratin 23 (KRT23) belongs to the acidic type I keratins (Moll, 1998). KRT23 knockdown decreases proliferation and affects the DNA damage response of colon cancer cells (Birkenkamp-Demtroder et al., 2013). Research have revealed that KRT23 is a subtype-specific prognostic factor for gastric cancer (Min et al., 2017).

Some limitations of our study had to be noted. At present, with the development of high-throughput technology, gene expression profile has become a critical molecular biomarker to identify the phenotype or outcome of ESCC (Zhan et al., 2016). Our study indicated that SPOCD1 might be a momentous biomarker for predicting prognosis in ESCC. Nonetheless, the correlation between the expression of SPOCD1 and the biological mechanisms in ESCC has not been fully clarified. Therefore, further *in vitro* and *in vivo* experiments are needed to validate the biological mechanism of SPOCD1. In addition, more clinical studies are also necessary to identify whether it is an independent prognostic biomarker. Finally, some statistical problems are worth mentioning. 1) A problem of using univariate cox analysis is that this approach ignores the correlation among genes, resulting in inaccurate subset for downstream analysis, especially given that the subset is so small. The network based variable selection methods perform regularized variable selection while incorporating correlations as networks (Wu et al., 2018; Ren et al., 2019), bypassing those disadvantages. 2) Due to the heterogeneity of disease, the heavy-tailed distributions and outliers in the clinical outcomes are widely observed. 3) Our future study should perform variable selection on a much larger scale. The additional gain of using robust network based variable selection is that the identified model is usually more stable and can be easily reproduced. 4) For reproducible research, the gene signature should be reported. The regnet R package can be one of the potential tools for reliable analysis (Wu et al., 2018; Ren et al., 2019). All in all, better statistical methods and more samples are needed to verify our findings in the future.

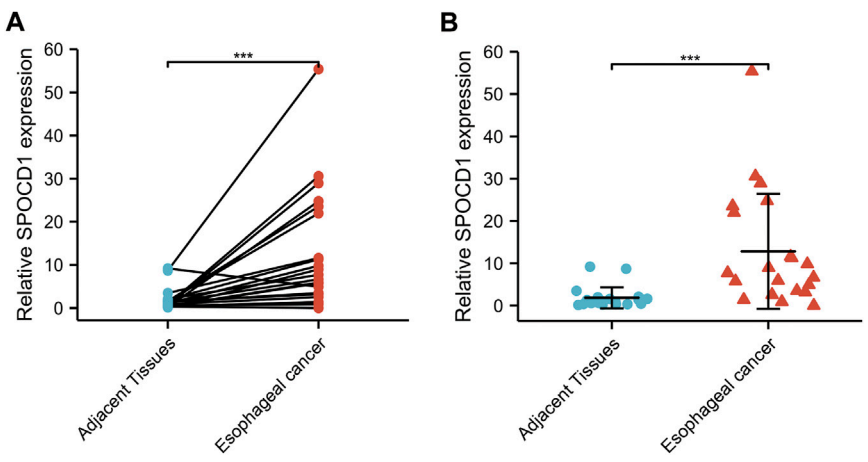


FIGURE 11 | The SPOCD1 expression in ESCC (n = 21), and adjacent normal tissues (n = 21) was evaluated by qRT-PCR. **(A)** The results were analyzed using paired sample *t* test; **(B)** The results were analyzed using Wilcoxon rank sum test. Results expressed as mean ± standard deviation (SD). ***$p < 0.001$, ***$p < 0.001$.

CONCLUSION

In summary, our study indicated that the expression of SPOCD1 was increased in ESCC tissues. The current data support the oncogenic role of SPOCD1 in the occurrence and development of ESCC. Most importantly, SPOCD1 might be an independent prognostic factor for ESCC patients. Our study provided novel evidence into the role of SPOCD1 in the tumorigenesis of ESCC and may promote the development of specific treatments or diagnostics. Further deep investigation and well-designed studies about the exact mechanism of SPOCD1 in ESCC are needed.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194116.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Biomedical Ethics Committee of the Fujian Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YX conceived the project and contributed equally to the whole study. LC and TW participated in data acquisition and interpretation. YZ and XH performed the experiments. ZL wrote the manuscript. YX, YC, and JC revised the article. All authors read and approved the final manuscript.

FUNDING

The project was funded by the grants of Fujian Provincial Clinical Research Center for Cancer Radiotherapy and Immunotherapy (Grant number: 2020Y2012), and the National Clinical Key Specialty Construction Program.

ACKNOWLEDGMENTS

We sincerely thank the public databases, including TCGA, GTEx, GenePattern, CellMiner, GDSC, and STRING datasets for providing open access.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872026/full#supplementary-material

Supplementary Figure S1 | Correlation analysis of SPOCD1 expression and mutation. **(A)**. The CNV of the group with high expression of SPOCD1; **(B)**. The CNV of the group with low expression of SPOCD1. SPOCD1, SPOC Domain Containing 1; CNV, copy number variation.

Supplementary Figure S2 | Drug sensitivity analysis and drug prediction. **(A–D)**. The differences in IC50 values for drugs prediction with high and low SPOCD1 expression from the GDSC database. SPOCD1, SPOC Domain Containing 1; GDSC, genomics of drug sensitivity in cancer.

REFERENCES

Abnet, C. C., Arnold, M., and Wei, W.-Q. (2018). Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 154 (2), 360–373. doi:10.1053/j.gastro.2017.08.023

Albihn, A., Johnsen, J. I., and Henriksson, M. A. (2010). MYC in Oncogenesis and as a Target for Cancer Therapies. *Adv. Cancer Res.* 107, 163–224. doi:10.1016/s0065-230x(10)07006-5

Baba, Y., Yoshida, N., Kinoshita, K., Iwatsuki, M., Yamashita, Y. I., Chikamoto, A., et al. (2018). Clinical and Prognostic Features of Patients with Esophageal Cancer and Multiple Primary Cancers: A Retrospective Single-Institution Study. *Ann. Surg.* 267 (3), 478–483. doi:10.1097/sla.0000000000002118

Berraondo, P., Sanmamed, M. F., Ochoa, M. C., Etxeberria, I., Aznar, M. A., Perez-Gracia, J. L., et al. (2019). Cytokines in Clinical Cancer Immunotherapy. *Br. J. Cancer* 120 (1), 6–15. doi:10.1038/s41416-018-0328-y

Birkenkamp-Demtroder, K., Hahn, S. A., Mansilla, F., Thorsen, K., Maghnouj, A., Christensen, R., et al. (2013). Keratin23 (KRT23) Knockdown Decreases Proliferation and Affects the DNA Damage Response of Colon Cancer Cells. *PLoS One* 8 (9), e73593. doi:10.1371/journal.pone.0073593

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Chen, S., Tang, J., Huang, L., and Lin, J. (2015). Expression and Prognostic Value of Mycl1 in Gastric Cancer. *Biochem. Biophys. Res. Commun.* 456 (4), 879–883. doi:10.1016/j.bbrc.2014.12.060

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C.-Y. (2014). CytoHubba: Identifying Hub Objects and Sub-networks from Complex Interactome. *BMC Syst. Biol.* 8 (S4), S11. doi:10.1186/1752-0509-8-s4-s11

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33 (1), 1–22. doi:10.18637/jss.v033.i01

Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 366 (10), 883–892. doi:10.1056/nejmoa1113205

ggpubr (2017). Ggpubr: 'ggplot2' Based Publication Ready Plots. Available at: https://cran.rproject.org/web/packages/ggpubr/index.html

Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc. Ser. A (Statistics Soc.* 174 (1), 245–246. doi:10.1111/j.1467-985x.2010.00676_9.x

Ha, S. Y., and Kim, S. H. (2012). Co-expression of Bmi1 and EZH2 as an Independent Poor Prognostic Factor in Esophageal Squamous Cell Carcinoma. *Pathol. Res. Pract.* 208 (8), 462–469. doi:10.1016/j.prp.2012.05.012

Hu, H., Yang, Y., Ji, Q., Zhao, W., Jiang, B., Liu, R., et al. (2012). CRL4B Catalyzes H2AK119 Monoubiquitination and Coordinates with PRC2 to Promote Tumorigenesis. *Cancer Cell* 22 (6), 781–795. doi:10.1016/j.ccr.2012.10.024

Jia, Q., Wang, J., He, N., He, J., and Zhu, B. (2019). Titin Mutation Associated with Responsiveness to Checkpoint Blockades in Solid Tumors. *JCI Insight* 4 (10). doi:10.1172/jci.insight.127901

Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., et al. (2006). Diversification of Transcriptional Modulation: Large-Scale Identification and Characterization of Putative Alternative Promoters of Human Genes. *Genome Res.* 16 (1), 55–65. doi:10.1101/gr.4039406

Kolde, M. R. (2015). Package 'pheatmap'. *R package* 1 (7), 790

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* 25. doi:10.18637/jss.v025.i01

Liang, J., Zhao, H., Hu, J., Liu, Y., and Li, Z. (2018). SPOCD1 Promotes Cell Proliferation and Inhibits Cell Apoptosis in Human Osteosarcoma. *Mol. Med. Rep.* 17 (2), 3218–3225. doi:10.3892/mmr.2017.8263

Liu, D., Yang, Y., Yan, A., and Yang, Y. (2020). SPOCD1 Accelerates Ovarian Cancer Progression and Inhibits Cell Apoptosis via the PI3K/AKT Pathway. *Onco Targets Ther.* 13, 351–359. doi:10.2147/ott.s200317

Liu, Q., Wang, X. Y., Qin, Y. Y., Yan, X. L., Chen, H. M., Huang, Q. D., et al. (2018). SPOCD1 Promotes the Proliferation and Metastasis of Glioma Cells by Up-Regulating PTX3. *Am. J. Cancer Res.* 8 (4), 624–635.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Luo, W., and Cory, B. (2013). Pathview: an R/Bioconductor Package for Pathway-Based Data Integration and Visualization. *Bioinformatics* 29 (14), 1830–1831. doi:10.1093/bioinformatics/btt285

Marshall, H. T., and Djamgoz, M. B. A. (2018). Immuno-Oncology: Emerging Targets and Combination Therapies. *Front. Oncol.* 8, 315. doi:10.3389/fonc.2018.00315

Masso-Valles, D., Beaulieu, M. E., and Soucek, L. (2020). MYC, MYCL, and MYCN as Therapeutic Targets in Lung Cancer. *Expert Opin. Ther. Targets* 24 (2), 101–114. doi:10.1080/14728222.2020.1723548

Mayakonda, A., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Research* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

McKay, M. M., and Morrison, D. K. (2007). Integrating Signals from RTKs to ERK/MAPK. *Oncogene* 26 (22), 3113–3121. doi:10.1038/sj.onc.1210394

Min, L., Zhao, Y., Zhu, S., Qiu, X., Cheng, R., Xing, J., et al. (2017). Integrated Analysis Identifies Molecular Signatures and Specific Prognostic Factors for Different Gastric Cancer Subtypes. *Transl. Oncol.* 10 (1), 99–107. doi:10.1016/j.tranon.2016.11.003

Moll, R. (1998). Cytokeratins as Markers of Differentiation in the Diagnosis of Epithelial Tumors. *Subcell. Biochem.* 31, 205–262.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour Evolution Inferred by Single-Cell Sequencing. *Nature* 472 (7341), 90–94. doi:10.1038/nature09807

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Odenthal, M., Hee, J., Gockel, I., Sisic, L., Schmitz, J., Stoecklein, N. H., et al. (2015). Serum microRNA Profiles as Prognostic/predictive Markers in the Multimodality Therapy of Locally Advanced Adenocarcinomas of the Gastroesophageal Junction. *Int. J. Cancer* 137 (1), 230–237. doi:10.1002/ijc.29363

Pasquali, S., Yim, G., Vohra, R. S., Mocellin, S., Nyanhongo, D., Marriott, P., et al. (2017). Survival after Neoadjuvant and Adjuvant Treatments Compared to Surgery Alone for Resectable Esophageal Carcinoma: A Network Meta-Analysis. *Ann. Surg.* 265 (3), 481–491. doi:10.1097/sla.0000000000001905

Paul, G., Nancy, C., and Huang, R. S. (2014). pRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Pennathur, A., Gibson, M. K., Jobe, B. A., and Luketich, J. D. (2013). Oesophageal Carcinoma. *Lancet* 381 (9864), 400–412. doi:10.1016/s0140-6736(12)60643-6

Reinhold, W. C., Sunshine, M., Liu, H., Varma, S. H., Kohn, K. W., Morris, J., et al. (2012). CellMiner: a Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. *Cancer Res.* 72 (14), 3499–3511. doi:10.1158/0008-5472.CAN-12-1370

Ren, J., Du, Y., Li, S., Ma, S., Jiang, Y., and Wu, C. (2019). Robust Network-Based Regularization and Variable Selection for High-Dimensional Genomic Data in Cancer Prognosis. *Genet. Epidemiol.* 43 (3), 276–291. doi:10.1002/gepi.22194

Sakaguchi, T., Yoshino, H., Sugita, S., Miyamoto, K., Yonemori, M., Osako, Y., et al. (2018). Bromodomain Protein BRD4 Inhibitor JQ1 Regulates Potential Prognostic Molecules in Advanced Renal Cell Carcinoma. *Oncotarget* 9 (33), 23003–23017. doi:10.18632/oncotarget.25190

Shen, H. Y., Wei, F. Z., and Liu, Q. (2021). Differential Analysis Revealing APOC1 to Be a Diagnostic and Prognostic Marker for Liver Metastases of Colorectal Cancer. *World J. Clin. Cases* 9 (16), 3880–3894. doi:10.12998/wjcc.v9.i16.3880

Shields, J. M., Pruitt, K., McFall, A., Shaub, A., and Der, C. J. (2000). Understanding Ras: 'it Ain't over 'til It's over'. *Trends Cell Biol.* 10 (4), 147–154. doi:10.1016/s0962-8924(00)01740-2

Tanaka, H., Kanda, M., Koike, M., Iwata, N., Shimizu, D., Ezaka, K., et al. (2015). Adherens Junctions Associated Protein 1 Serves as a Predictor of Recurrence of Squamous Cell Carcinoma of the Esophagus. *Int. J. Oncol.* 47 (5), 1811–1818. doi:10.3892/ijo.2015.3167

Therneau, T. (2012). A Package for Survival Analysis. *R. package* 2, 37–42.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global Cancer Statistics, 2012. *CA A Cancer J. Clin.* 65 (2), 87–108. doi:10.3322/caac.21262

van der Heijden, A. G., Mengual, L., Lozano, J. J., Ingelmo-Torres, M., Ribal, M. J., Fernandez, P. L., et al. (2016). A Five-Gene Expression Signature to Predict Progression in T1G3 Bladder Cancer. *Eur. J. Cancer* 64, 127–136. doi:10.1016/j.ejca.2016.06.003

Wang, M., Smith, J. S., and Wei, W. Q. (2018). Tissue Protein Biomarker Candidates to Predict Progression of Esophageal Squamous Cell Carcinoma and Precancerous Lesions. *Ann. N.Y. Acad. Sci.* 1434 (1), 59–69. doi:10.1111/nyas.13863

Wu, C., Zhang, Q., Jiang, Y., and Ma, S. (2018). Robust Network-Based Analysis of the Associations between (Epi)genetic Measurements. *J. Multivar. Anal.* 168, 119–130. doi:10.1016/j.jmva.2018.06.009

Xu, Y.-W., Peng, Y.-H., Xu, L.-Y., Xie, J.-J., and Li, E.-M. (2019). Autoantibodies: Potential Clinical Applications in Early Detection of Esophageal Squamous Cell Carcinoma and Esophagogastric Junction Adenocarcinoma. *Wjg* 25 (34), 5049–5068. doi:10.3748/wjg.v25.i34.5049

Yang, W., Jorge, S., Patricia, G., Edelman, E. J., Howard, L., Simon, F., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yang, Y., Zhang, J., Chen, Y., Xu, R., Zhao, Q., Guo, W., et al. (2020). MUC16, and TTN Genes Mutation Correlated with Prognosis, and Predicted Tumor Mutation Burden and Immunotherapy Efficacy in Gastric Cancer and Pan-Cancer. *Clin. Transl. Med.* 10 (4), e155. doi:10.1002/ctm2.155

Yi, J., Ren, L., Wu, J., Li, W., Zheng, X., Du, G., et al. (2019). Apolipoprotein C1 (APOC1) as a Novel Diagnostic and Prognostic Biomarker for Gastric Cancer. *Ann. Transl. Med.* 7 (16), 380. doi:10.21037/atm.2019.07.59

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhan, C., Yan, L., Wang, L., Jiang, W., Zhang, Y., Xi, J., et al. (2016). Landscape of Expression Profiles in Esophageal Carcinoma by the Cancer Genome Atlas Data. *Dis. Esophagus* 29 (8), 920–928. doi:10.1111/dote.12416

Zhu, M., Yan, C., Ren, C., Huang, X., Zhu, X., Gu, H., et al. (2017). Exome Array Analysis Identifies Variants in SPOCD1 and BTN3A2 that Affect Risk for Gastric Cancer. *Gastroenterology* 152 (8), 2011–2021. doi:10.1053/j.gastro.2017.02.017

Zoch, A., Auchynnikava, T., Berrens, R. V., Kabayama, Y., Schopp, T., Heep, M., et al. (2020). SPOCD1 Is an Essential Executor of piRNA-Directed De Novo DNA Methylation. *Nature* 584 (7822), 635–639. doi:10.1038/s41586-020-2557-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lin, Chen, Wu, Zhang, Huang, Chen, Chen and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fgene.2022.891270

The Landscape of the Tumor-Infiltrating Immune Cell and Prognostic Nomogram in Colorectal Cancer

Jiateng Zhong[1,2,3,4†], Yu Qin[1,2†], Pei Yu[2], Weiyue Xia[2], Baoru Gu[2], Xinlai Qian[2], Yuhan Hu[2], Wei Su[1] and Zheying Zhang[1,2]**

[1]Department of Pathology, The First Affiliated Hospital of Xinxiang Medical University, Xinxiang, China, [2]Department of Pathology, Xinxiang Medical University, Xinxiang, China, [3]Department of Gynecology, The Fourth Clinical College of Xinxiang Medical University, Xinxiang, China, [4]Xinxiang Key Laboratory of Tumor Vaccine and Immunotherapy, Institute of Precision Medicine, Xinxiang Medical University, Xinxiang, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Liang Wang,
China Medical University, China
Qiang Zhang,
University of Michigan, United States
Gang Wang,
Xuzhou Medical University, China

***Correspondence:**
Wei Su
hnswei@163.com
Zheying Zhang
081023@xxmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 07 March 2022
Accepted: 11 April 2022
Published: 12 May 2022

Citation:
Zhong J, Qin Y, Yu P, Xia W, Gu B, Qian X, Hu Y, Su W and Zhang Z (2022) The Landscape of the Tumor-Infiltrating Immune Cell and Prognostic Nomogram in Colorectal Cancer. Front. Genet. 13:891270. doi: 10.3389/fgene.2022.891270

Tumor-infiltrating immune cells are associated with prognosis and immunotherapy targets in colorectal cancer (CRC). The recently developed CIBERSORT method allows immune cell analysis by deconvolution of high-throughput data onto gene expression. In this study, we analyzed the relative proportions of immune cells in GEO (94 samples) and TCGA (522 samples) CRC data based on the CIBERSORT method. A total of 22 types of tumor-infiltrating immune cells were evaluated. Combined with GEO and TCGA data, it was found that naive B cells, M2 macrophages, and resting mast cells were highly expressed in normal tissues, while M0 macrophages, M1 macrophages, activated mast cells, and neutrophils were highly expressed in tumors. Moreover, we constructed a prognostic model by infiltrating immune cells that showed high specificity and sensitivity in both the training (AUC of 5-year survival = 0.699) and validation (AUC of 5-year survival = 0.844) sets. This provides another basis for clinical prognosis. The results of multiple immunofluorescence detection showed that there were differences in the results of bioinformatics analysis. Neutrophils were highly expressed in normal tissues, and M2 macrophages were highly expressed in tumor tissues. Collectively, our data suggested that infiltrating immune cells in CRC may be an important determinant of prognosis and immunotherapy.

Keywords: CIBERSORT, immunotherapy, colorectal cancer, prognosis, nomogram

INTRODUCTION

Colorectal cancer (CRC) is a common malignant tumor of the digestive tract, and its incidence ranks third among the top ten malignant tumors (Bray et al., 2018). Most of the CRC has been found in the late stages, often accompanied by metastasis, which is an important reason for the high mortality rate of CRC patients (Xi and Xu, 2021). The treatment is an important factor affecting the survival rate of advanced CRC. Clinical treatment of advanced CRC mainly includes chemotherapy, radiotherapy, and traditional Chinese medicine treatment. Chemotherapy is the main treatment for advanced CRC, which can reduce cancer to a certain extent, but it also increases the toxicity and side effects (Sun et al., 2021). There is still a lack of targets that can help to select personalized treatment options.

The progress of tumors is related to not only the growth and dissemination of tumor cells but also the infiltrating immune cells. It is the interaction between these different types of cells that promotes the growth of tumors (Binnewies et al., 2018). In the past 20 years, there have been a lot of reports on the correlation between the infiltration of intratumoral immune cells and prognosis in solid tumors (Nirmal et al., 2018; Edlund et al., 2019; Kim et al., 2019). Immunotherapy is a kind of therapy that regulates T-cell activity by co-suppression or co-stimulation. The therapeutic methods of high anti-tumor immune response have shown remarkable clinical effects (Waldman et al., 2020; Zhang and Zhang, 2020). Immunocytes infiltrated in tumors are most likely to be used as drug targets to improve patient survival.

With the development of immunological checkpoint therapy, the distribution of infiltrating immune cells in tumors has become a research topic. The previous studies have mostly used flow cytometry or immunohistochemistry to assess the composition of infiltrating immune cells in tumors (Finotello and Trajanoski, 2018), but these methods are difficult to detect in large quantities. Researchers recently developed the CIBERSORT analysis tool, a new bioinformatics tool, which is a deconvolution algorithm developed by Bindea G (Chen et al., 2018). It can estimate the cell composition of complex tissues based on standardized gene expression data.

Here, we use CIBERSORT to quantify the expression of 22 kinds of immune cells in CRC and normal colon tissues. For the first time, we analyze the relationship between immune cells and survival, tumor size, lymphatic metastasis, and blood vessel metastasis of CRC patients. Also, we establish a prognostic nomogram for predicting survival by immune cells. It is hoped that this study will be helpful for prognosis judgment and immunotherapy in CRC patients.

MATERIALS AND METHODS

Data Collection

The study utilized data from the public databases. Gene expression profile datasets for common CRC patients were obtained from the GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/gds), and TCGA databases (The Cancer Genome Atlas, https://portal.gdc.cancer.gov/). The GEO data selected three datasets for the patient to be paired with their normal tissues, namely, GSE32323 (platform GPL750), GSE110223 (platform GPL96), and GSE110224 (platform GPL750), for a total of 94 samples, including 47 normal control samples and 47 tumor tissue samples. The samples of inappropriate values were removed, leaving a total of 67 samples, including 30 normal samples and 37 tumor tissue samples. The TCGA dataset and corresponding clinical information were downloaded from TCGA official website, a total of 522 samples, containing 42 normal control samples and 480 tumor tissue samples. After removing unreasonable data, 245 samples remained which included 13 normal control samples and 232 tumor tissue samples.

After downloading the GEO data, the probe name was annotated, the three-chip data were combined into one, and then the experimental batch correction was performed. Similarly, after downloading TCGA data, the data were sorted into a matrix. The data were then normalized. All the aforementioned steps were completed by R software.

Assessing the Composition of Immune Cells Infiltrated in Tumors

This study uses CIBERSORT (http://cibersort.stanford.edu/) to assess the relative proportion of 22 invasive immune cell types in each tumor tissue. CIBERSORT is an assessment tool that analyzes the abundance of specific cell types in mixed cells. The analysis can be performed using gene expression profile data. The immune cells infiltrated in tumor tissue can be analyzed for a value for subsequent analysis. These tumor-infiltrating immune cell types include: M1 macrophages, M2 macrophages, M0 macrophages, T follicular helper cells, resting memory CD4 T cells, activated memory CD4 T cells, γδ T cells, CD8 T cells, regulatory T cells, naive CD4 T cells, resting NK cells, activated NK cells, resting mast cells, activated mast cells, memory B cells, resting dendritic cells, activated dendritic cells, naive B cells, monocytes, neutrophils, eosinophils, and plasma cells. For each tumor sample, the sum of the evaluated immune cell type scores is equal to 1. The results with a value of $p < 0.05$ were considered eligible for further analysis. The process is completed by R software.

Nomograms

A nomogram model was established and validated to predict the prognosis of CRC. Then, the corresponding clinical information from TCGA data was downloaded, and the samples were divided randomly into training and validation (3:1) sets for evaluation and verification. The following clinical information was collected from the database: patient, gender, age, TNM stage, and survival data.

Univariate analysis of TCGA immune cell infiltration results and clinical variables, using $p < 0.05$ variables, naive B cells, activated T cells CD4 memory, T cells follicular helper, M0 macrophages, M1 macrophages, and M2 macrophages established a nomogram of 3-year and 5-year survival rates for patients with CRC.

The method of using the nomogram is as follows: the points on the score of each variable of the patient are matched to the scores on the topmost horizontal line, and then the scores of each variable are added to obtain the total score of each patient, and finally the total score corresponds to the bottom-end 3- and 5-year survival rates.

Multiplex Immunofluorescence Staining (mIF)

A total of three pairs of colorectal samples and adjacent normal tissues were collected from The First Affiliated Hospital of Xinxiang Medical College. The experiments were performed consistently with the guidelines and regulations of the Committee of Xinxiang Medical University. Informed consent was obtained from all patients. Tissues were fixed

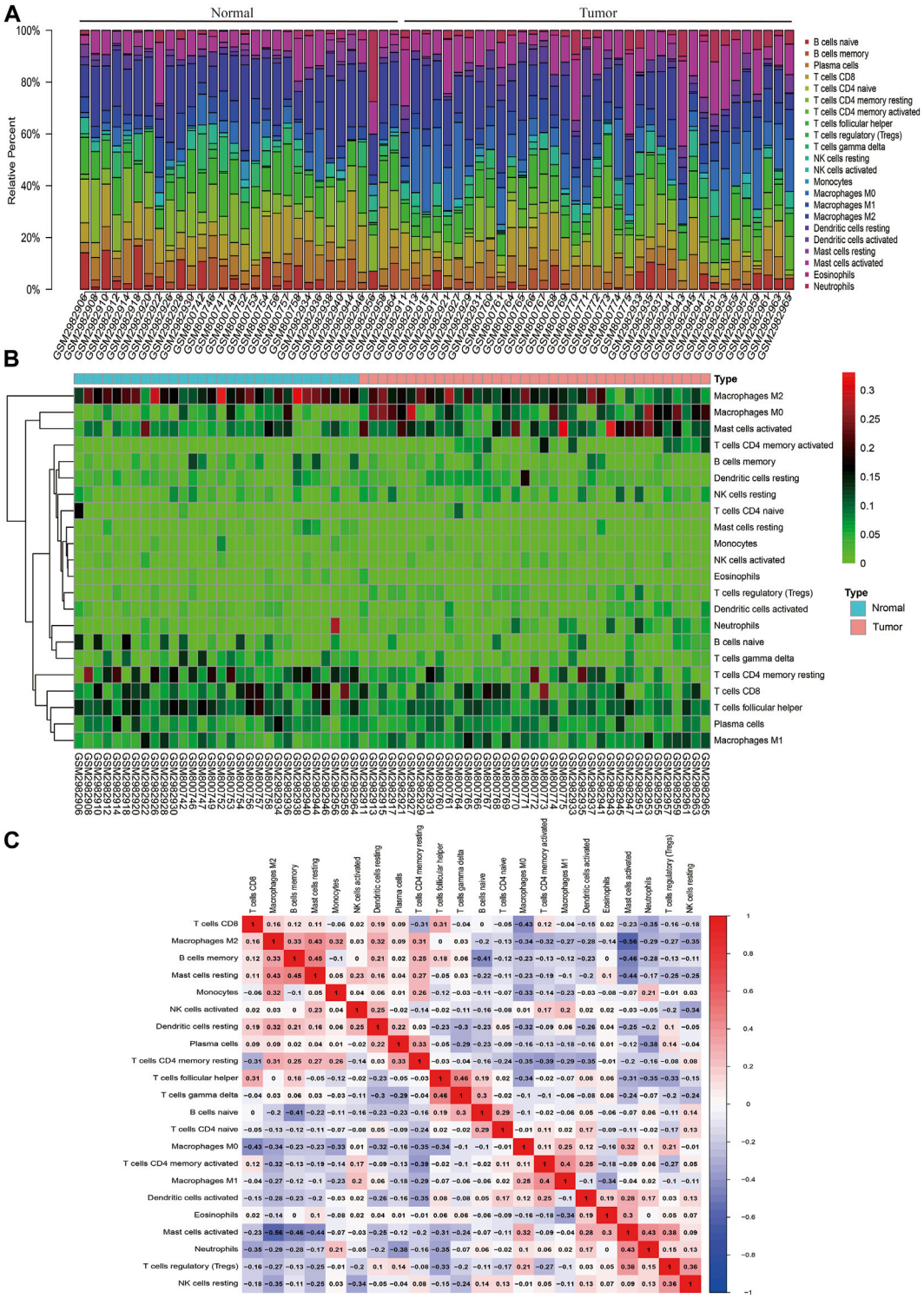


FIGURE 1 | Landscape of immune infiltration in normal and CRC tissues based on GEO data. **(A)** Composition of immune infiltration between paired cancer and normal tissues. The height of the colored column represents the proportion of immune cells. **(B)** Heat map of the 22 immune cell proportions. The genes shown in red are upregulated, and the genes in green are downregulated. The horizontal axis shows the clustering information of samples which were divided into two major clusters. **(C)** Correlation matrix of 22 immune cell expressions. The redder the color, the higher the positive correlation. The bluer the color, the higher the negative correlation.

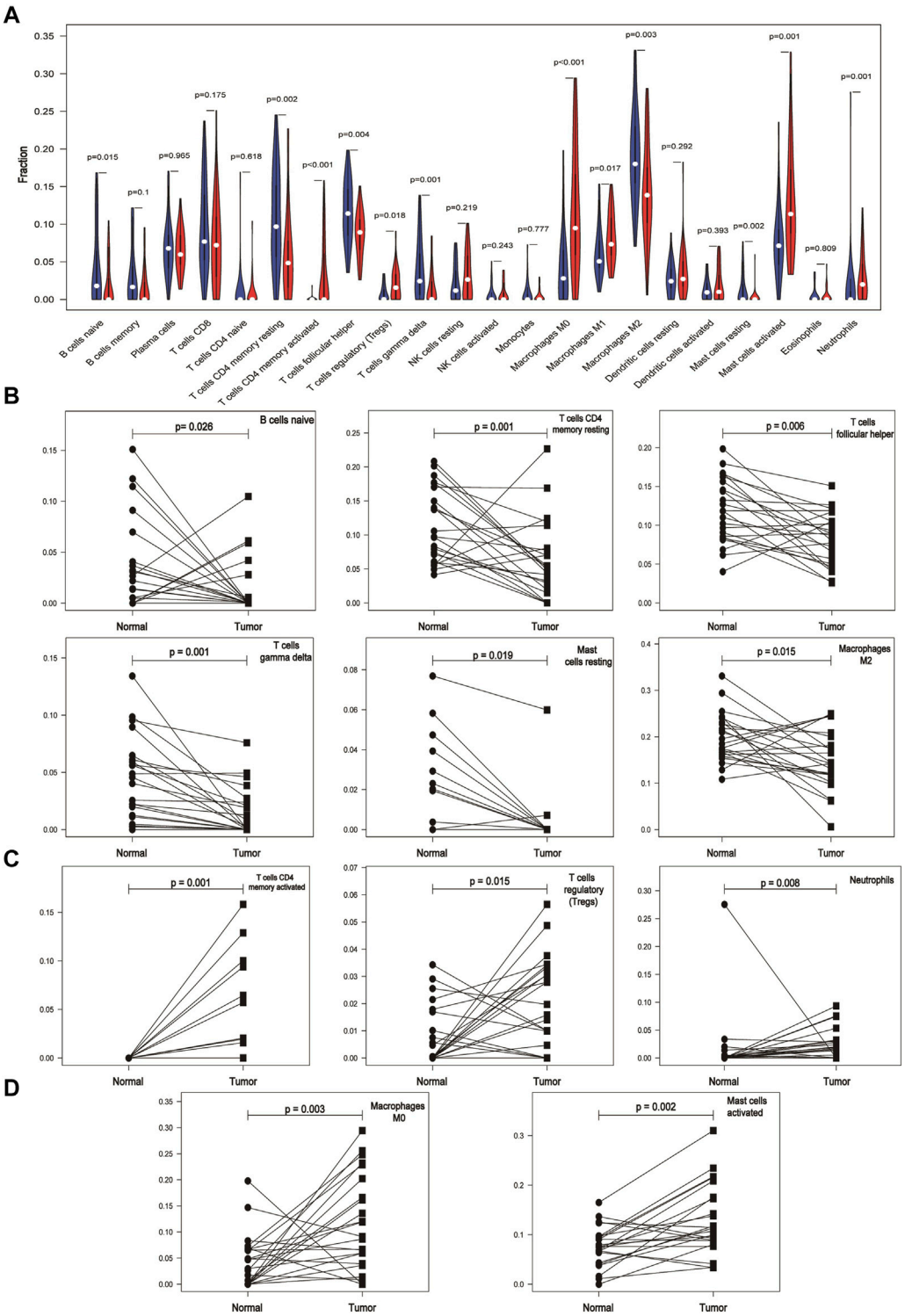


FIGURE 2 | Differential expression of immune cells in cancer and normal tissues based on the GEO database. **(A)** Expression of different immune cells in normal and cancer tissues. The white dots inside the violin indicate the median value. **(B)** Expression of immune cells is lower in cancer tissues than in paired normal tissues. **(C)** Immune cells are mainly expressed in cancer tissues. **(D)** Expression of immune cells is higher in cancer tissues than in paired normal tissues.

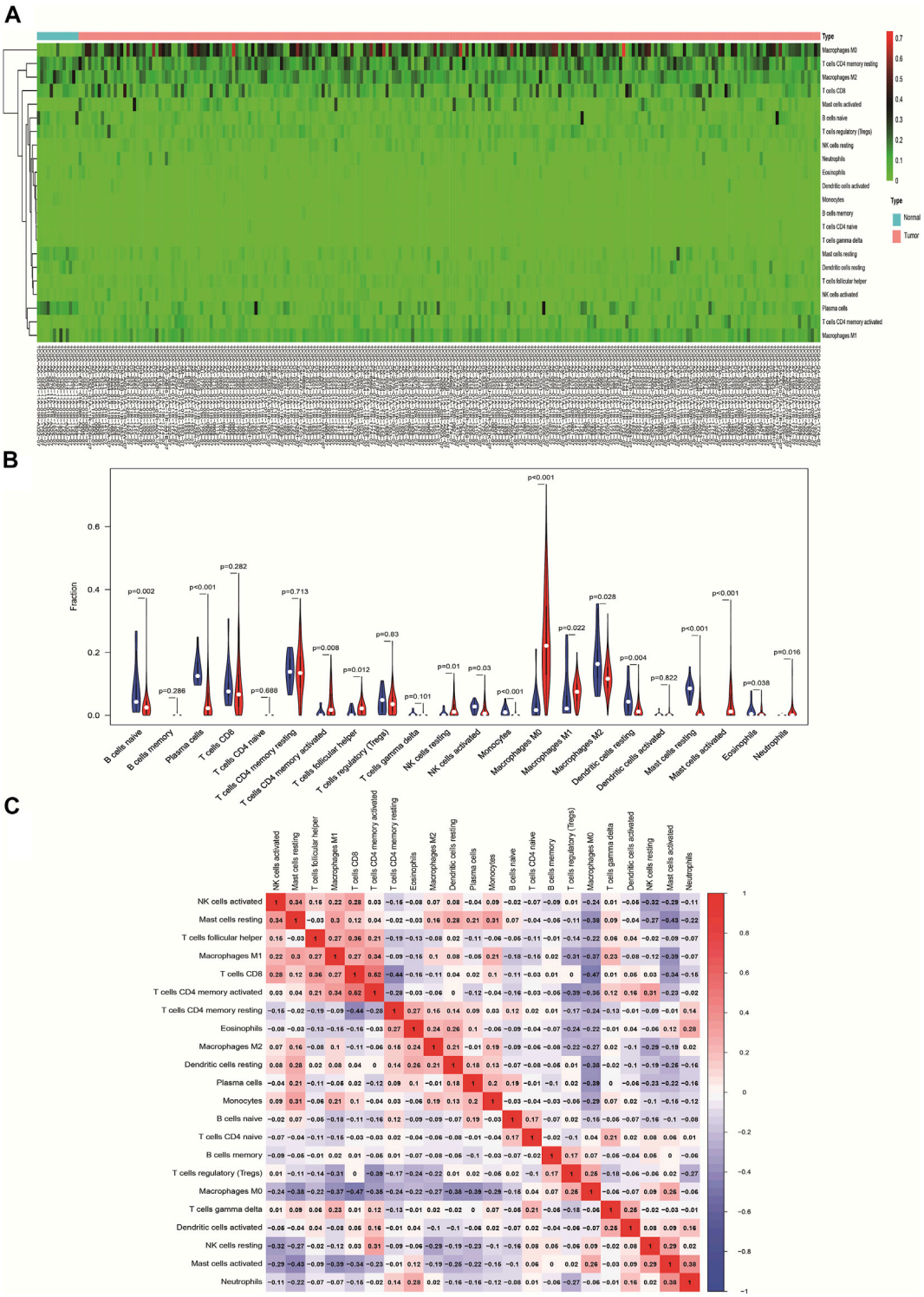


FIGURE 3 | Landscape of immune infiltration in normal and CRC tissues based on TCGA data. **(A)** Heat map of the 22 immune cell expressions. The genes shown in red are upregulated, and the genes in green are downregulated. **(B)** Violin map of the 22 immune cell expressions in normal and CRC tissues. **(C)** Correlation matrix of 22 immune cell expressions based on TCGA data. The redder the color, the higher the positive correlation. The bluer the color, the higher the negative correlation.

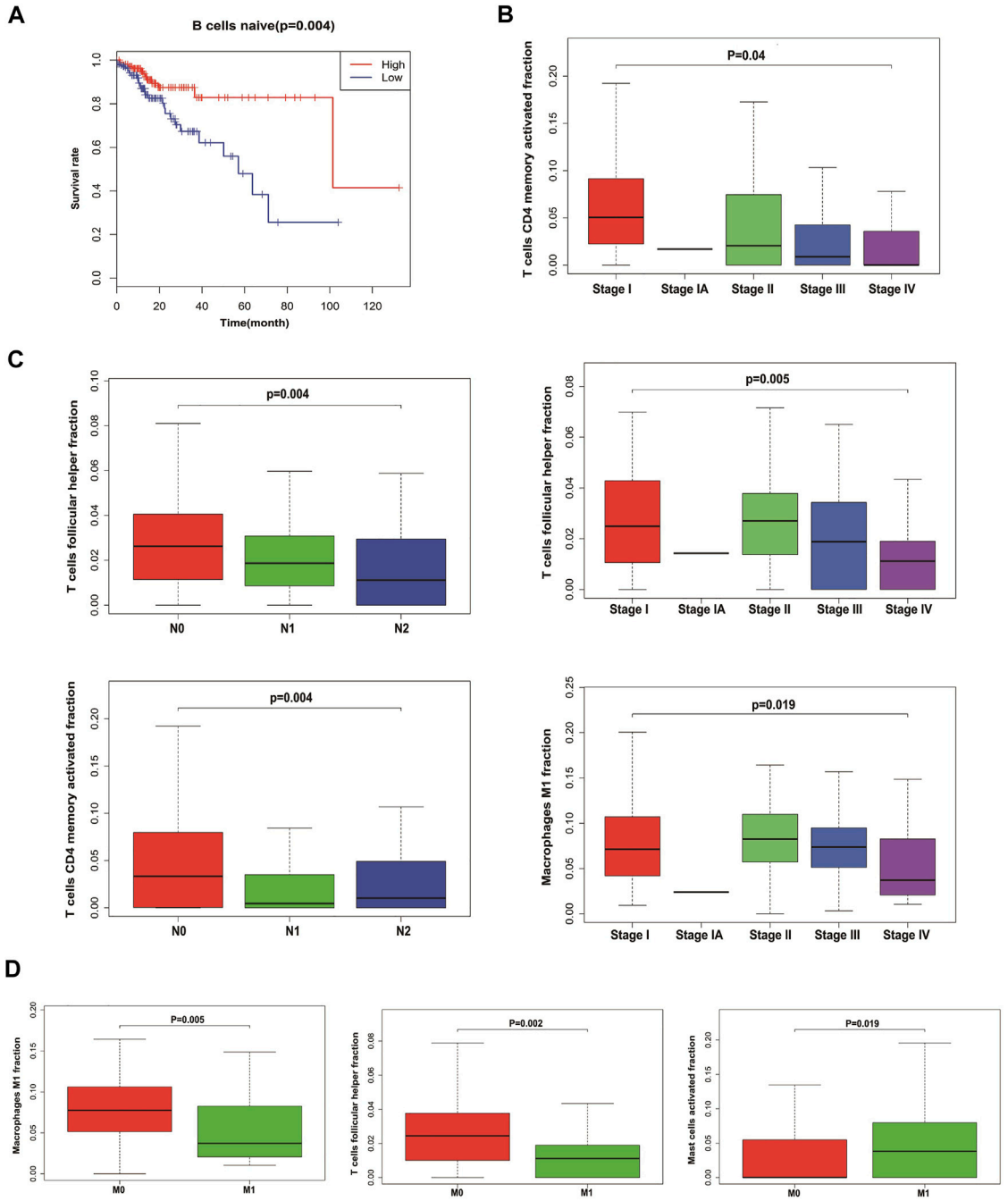


FIGURE 4 | Relationship between tumor-infiltrating immune cells and clinical characters of CRC. **(A)** Survival plots of naive B cells. Data were analyzed using the Kaplan–Meier plotter. Patients with naive B cell expression above the median are indicated in the red line and below the median in the green line. **(B)** Immune cells with the highest correlation with CRC stages. **(C)** Immune cells with the highest correlation with lymph node metastasis in CRC. **(D)** Immune cells with the highest correlation with distant metastasis in CRC.

and embedded for paraffin section, and the paraffin section was dewaxed and antigen repair and serum sealing were performed; the first primary antibody was added, the corresponding HRP-labeled secondary antibody was added, and CY3 fluorescence enhancer was added. Microwave treatment was followed by the addition of the second primary antibody, the corresponding HRP-labeled secondary antibody, and the FITC fluorescence enhancer. A third primary antibody was added, and the corresponding HRP-labeled secondary antibody and a 647 fluorescence enhancement agent were added. After

microwave treatment, the fourth primary antibody was added. DAPI was used to re-stain the nucleus, quenching by spontaneous fluorescence, sealing the slices, and placing the slices under a scanner to collect images. During the final imaging, the nuclei stained with DAPI are blue under the excitation of ultraviolet, and the positive expression is the corresponding fluorescein-labeled red light, pink light, green light, and purple light. All the slides were scanned and observed using a Pannoramic MIDI (3DHISTECH) and CaseViewer C.V.2.4. The antibodies CD19, CD163, C-kit, CD68, NOS2, mast cell tryptase, and CD16 were purchased from Servicebio Technology Co. Ltd. DAPI is used to label the nucleus, CD19 is used to label naive B cells, CD163 is used to label M2 macrophages, C-kit labeled the resting mast cells, CD68 labeled cell M0, NOS2 labeled M1 macrophages, mast cell tryptase labeled activated mast cells, and CD16 labeled neutrophils (Jensen et al., 2014; Yamaguchi et al., 2016; Lv et al., 2017; Sanz et al., 2019; Valent et al., 2019; Lu et al., 2020; Lai et al., 2021).

Statistical Analysis

The relationship between immune cell type and clinic variables was analyzed by univariate Cox regression. The overall survival was analyzed by Kaplan–Meier curves and assessed using a log-rank test. Nomogram construction was according to multivariate Cox regression and Wilcox test results. The C-index was calculated, and the receiver operating characteristic (ROC) and calibration curve was drawn to evaluate the performance of the nomogram model. The differences in the expression of immune cells between tumor and non-tumor tissue were assessed using an unpaired *t*-test, with $p < 0.05$ being considered statistically significant. All analyses were performed using R version 3.5.1.

RESULTS

Analysis of the Immune Cell Infiltration Composition in CRC Tissues Based on the GEO Database

CIBERSORT algorithm was used to evaluate the distribution of immune cells in tumors and normal tissues of 47 paired CRC patients in the GEO database (**Figure 1A**). The detailed results are provided in **Supplementary Table S1**. We can see that there are great differences in the composition of immune cells in different samples, which indicates that the composition of different types of immune cells may be used to diagnose and judge the prognosis of tumors. Hierarchical cluster analysis was used to analyze the difference of immune cell distribution between tumor and normal tissues. We found that there were significant differences in M0 macrophages, M1, M2, and activated mast cells (**Figure 1B**). It can be seen from the correlation thermogram that there is a significant negative correlation between M2 macrophages and activated mast cells and a strong positive correlation between M0 macrophages (**Figure 1C**). From **Figure 1C**, we can find out the correlation between different immune cells, which will play an important role in the analysis of the mechanism in the future.

Expression of Immune Cells in Cancer and Normal Tissues Based on the GEO Database

By comparing the expression of different types of immune cells in normal and cancer tissues, we found that the expression of naive B cells, resting T cells CD4 memory, activated T cells CD4 memory, T cells follicular helper, regulatory T cells (Tregs), γδ T cells, M0 macrophages, M2 macrophages, resting mast cells, activated mast cells, and neutrophils in cancer and normal tissues was significantly different ($p < 0.05$) (**Figure 2A**). M1 macrophages were statistically significant in the comparison between the normal and tumor groups of the violin map, but the difference was not statistically significant in the comparison of the paired samples. The fraction of naive B cells, resting T cells CD4 memory, T cells follicular helper, γδ T cells, resting mast cells, and M2 macrophages was higher in normal adjacent tissue than in cancer tissue (**Figure 2B**). Activated T cells CD4 memory, Tregs, and neutrophils were mainly present in cancer (**Figure 2C**). M0 macrophages and activated mast cells were increased in cancer compared to normal (**Figure 2D**). Other immune cells were not significantly altered between tissues.

The Landscape of Immune Infiltration in CRC Based on TCGA Database

We used data of TCGA database to validate the results of immunocyte infiltration of GEO data. The distribution of immune cells in each sample is plotted as a histogram, with which it can be visually found that the distribution of immune cells in different samples is significantly different and consistent with the GEO results (**Supplementary Figure S1**). Compared with normal tissues, CRC tissues contained a high proportion of M0 macrophages and activated T cells CD4 memory, whereas the M2 macrophages and resting T cells CD4 memory fraction was relatively low in cancer (**Figure 3A**). From the violin map, we can also see that naive B cells, plasma cells, activated NK cells, monocytes, M2 macrophages, resting dendritic cells, resting mast cells, and eosinophils are expressed higher in normal tissues than in cancer tissues ($p < 0.05$). The activated T cells CD4 memory, T cells follicular helper, resting NK cells, M0 macrophages, M1 macrophages, activated mast cells, and neutrophils are mainly expressed in cancer tissues ($p < 0.05$) (**Figure 3B**). The correlation heat map results are basically consistent with GEO results (**Figure 3C**).

The Prognostic and Clinical Values of Tumor-Infiltrating Immune Cells in CRC

We analyzed the relationship between infiltrating immune cells and the prognosis of patients with CRC. Through Kaplan–Meier analysis, we found that only naive B cells had statistical significance in the prognosis of patients (**Figure 4A**).

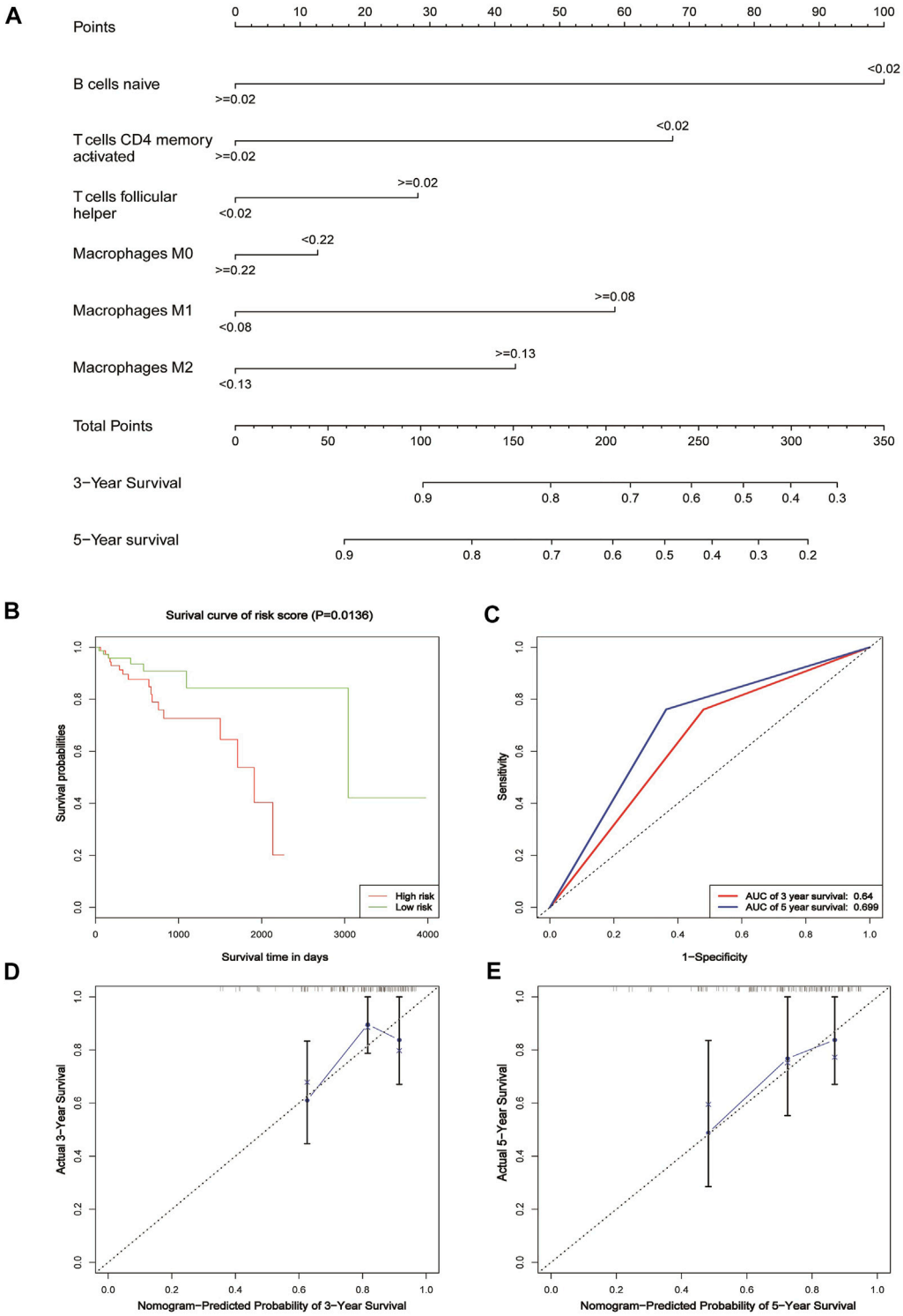


FIGURE 5 | Nomogram for patients with CRC. **(A)** Nomogram for predicting 3- and 5-year survival for CRC patients based on tumor-infiltrating immune cells. **(B)** Kaplan–Meier estimates of patients' survival status and time using the median risk score cut-off, which divided patients into low-risk and high-risk groups based on the training cohort. **(C)** ROC analysis of the sensitivity and specificity of the survival time by the tumor-infiltrating immune cells based on the risk score for the training cohort. **(D)** Calibration curve for the prediction of 3-year overall survival based on the training cohort. **(E)** Calibration curve for the prediction of 5-year overall survival based on the training cohort.

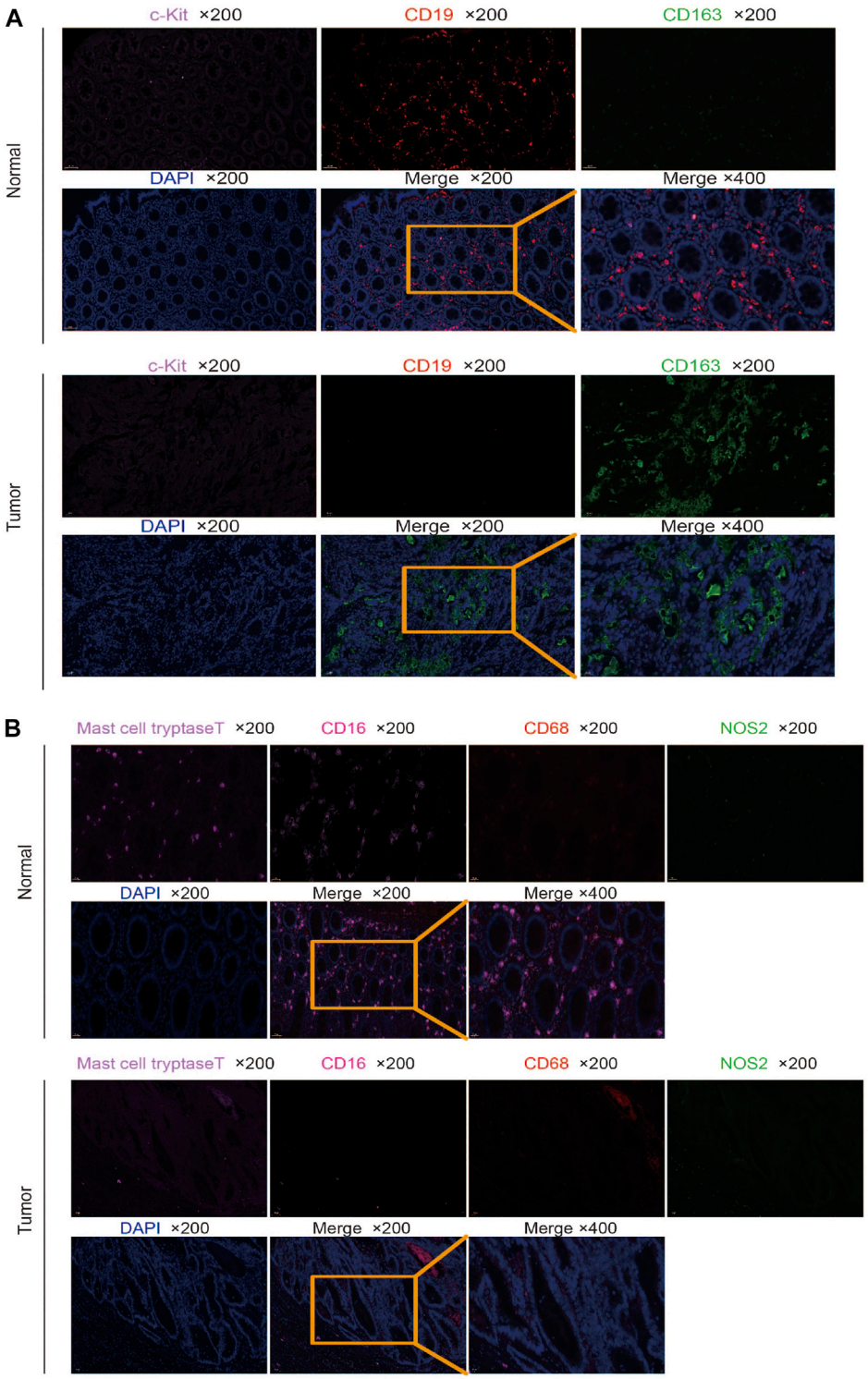


FIGURE 6 | Multiplex immunofluorescence detection of different marker expressions. **(A)** Detection of CD19, CD163, and C-kit expression in tumor and normal tissues by multiplex immunofluorescence. **(B)** Detection of CD68, NOS2, mast cell tryptase, and CD16 expressions in tumor and normal tissues by multiplex immunofluorescence.

Associated with the tumor stage are activated T cells CD4 memory, T cells follicular helper, and M1 macrophages (**Figure 4B**). We used Wilcoxon's test to find that activated T cells CD4 memory and T cells follicular helper showed significant differences in lymph node metastasis and non-metastasis groups (**Figure 4C**). Also, the cells associated with tumor hematogenous metastasis are T cells follicular helper, M1 macrophages, and activated mast cells (**Figure 4D**).

Establishing and Validating a Nomogram Model to Predict CRC Prognosis

In order to predict the overall survival rate of CRC patients individually, we selected the independent prognostic factor naive B cells as one variable. The other variables differentially expressed in different clinical traits were screened out by the Wilcox test, as shown in **Figure 4**. The immune cells with too low expression levels were removed. Finally, we used naive B cells, activated T cells CD4 memory, T cells follicular helper, M0 macrophages, M1 macrophages, and M2 macrophages to establish a nomogram model to predict the overall survival rate of CRC patients for 3–5 years (**Figure 5A**). In the training group, Kaplan–Meier analysis showed that stratified risk factors can distinguish the overall prognostic survival rate (p = 0.0136) (**Figure 5B**). The AUC values predicting 3-year and 5-year survival rates were 0.64 and 0.699, respectively (**Figure 5C**). The calibration curve verifies model credibility (**Figures 5D,E**). In the validation group, Kaplan–Meier analysis showed that stratified risk factors could better distinguish the overall prognostic survival rate (p = 0.0026) (**Supplementary Figure 2A**), and the AUC values predicting 3-year and 5-year survival rates were 0.836 and 0.844, respectively (**Supplementary Figure 2B**).

Different Markers Expressed Differently in CRC and Normal Tissues

We used multiple immunofluorescence techniques to detect the expression of CD19, CD163, C-kit, CD68, NOS2, mast cell tryptase, and CD16 markers in colorectal cancer tissues and surrounding normal tissues on the same slide. C-kit and CD19 were mainly expressed in normal tissues, and their weak expression in tumors was consistent with the results of bioinformatics analysis, while CD163 was more strongly expressed in tumors than in normal tissues (**Figure 6A**). Mast cell tryptase was widely expressed in tumor parenchyma but weakly and regionally strongly expressed in normal tissues. CD16 expression was stronger in normal tissue than in tumor tissue. The expression of CD68 and NOS2 was stronger in tumor stroma than in normal tissue but weaker in tumor parenchyma than in normal tissue (**Figure 6B**).

DISCUSSION

Tumor immunotherapy is an important therapy after surgery, radiotherapy, chemotherapy, and targeted therapy. Immune drug therapy has great challenges and great potential (Ventola, 2017). It is important to evaluate the infiltration of immune cells in different tumors. Based on the gene expression data in TCGA and GEO databases, this study analyzed the distribution of tumor-infiltrating immune cells in CRC. The relationship between infiltrating immune cells and clinical characteristics was also studied to reveal its prognostic value in CRC. Finally, we established a nomogram using infiltrating immune cells to assess the prognosis of patients.

By comparing the GEO datasets and TCGA data results in immune cell infiltration, we found that the highly expressed immune cells in CRC were M0 macrophages, M1 macrophages, activated mast cells, and neutrophils. Naive B cells, M2 macrophages, and resting mast cells are highly expressed in normal tissues. The macrophages can be classified into classical M1 and M2 according to their functions (Orihuela et al., 2016). M1 macrophages participate in the inflammatory response and anti-tumor immunity (Ubil et al., 2018), while M2 macrophages can promote tumor development (Genard et al., 2017). M0 macrophages are formed by mononuclear cells that have not yet to the M1 or M2 macrophage polarization (Zajac et al., 2013). The macrophages in the tumor stroma are called tumor-associated macrophages (TAM) (Petty et al., 2021). Some studies have shown that TAM may have an M2 phenotype, which is closely related to tumor angiogenesis and lymphangiogenesis, and is involved in the process of tumor occurrence, growth, invasion, and metastasis (Riabov et al., 2014). Other studies have shown that TAM has an M1 phenotype and can secrete immunomodulatory factors and enzymes to play an anti-tumor immune role (Hobson-Gutierrez and Carmona-Fontaine, 2018). In breast cancer, TAM plays a dominant role in tumor promotion, and the expression of TAM is negatively correlated with the prognosis of patients (Zhao et al., 2017). However, in lung cancer, the results were just the opposite, i.e., TAM infiltration was positively correlated with the prognosis of patients (Conway et al., 2016). The role of TAM in tumors is not clear. The results of this study suggested that the expression of M2 macrophages in CRC tissues was lower than that in normal tissues, but it was not related to TNM stage, tumor size, and presence or absence of hematological and lymphatic metastasis. The expression of M1 macrophages in the tumor was higher than that in normal tissue and correlated with lymphatic metastasis. This result is consistent with another report using the CIBERSORT method to analyze immune cell infiltration in CRC tissues (Xiong et al., 2018). These results suggest that M1 macrophages and M2 may play different roles in different tumors or at different stages of tumors. No matter which research results show the important role of macrophages in tumors, further experiments are needed to verify their function.

As seen from the correlation heat map, M2 macrophages are significantly negatively correlated with activated mast cells and positively correlated with resting mast cells, and resting mast cells and activated mast cells are negatively correlated. Many literature studies have reported that mast cells have the effect of promoting cancer (Molderings et al., 2017; Aponte-Lopez et al., 2018; Komi and Redegeld, 2020), and we speculate that there may be a regulatory relationship among them, which plays a role in the tumor together.

Kaplan–Meier analysis showed that only naive B cells had statistical significance in the prognosis of patients. A naive B cell is a B cell that has not been exposed to an antigen. Once exposed

to an antigen, the naive B cell either becomes a memory B cell or a plasma cell that secretes antibodies specific to the antigen that was originally bound. Little has been reported about the relationship between the initial B cells and the tumor, and there may be unknown functions awaiting further exploration.

The occurrence and development of CRC have great complexity and heterogeneity in cell origin, histological grade, clinical stage, recurrence, and metastasis. There is still a lack of practical, inexpensive prognostic assessment tools in clinical work. The nomogram is a mapping method based on the results of multivariate analysis. It can integrate multiple clinical–pathological indicators to determine the probability of a certain clinical event in a particular individual (Balachandran et al., 2015). It has become a relatively simple and accurate prediction method and has been applied to the prognosis of tumors. We used these six risk factors, naive B cells, activated T cells CD4 memory, T cells follicular helper, M0 macrophages, M1 macrophages, and M2 macrophages, to construct a nomogram to assess the patient's 3-year and 5-year survival rate. The nomogram model established in this study divided the patients with CRC into a high-risk group and low-risk group, which can better distinguish the prognosis survival rate of patients. In the verification group, the model AUC value reached 0.844, and the correction curve also showed that the model had a good consistency with the actual incidence. The nomogram model constructed by this study can predict the prognosis of individual patients and individualized treatment for patients with CRC.

We verified marker expression with multiplex immunofluorescence. It was found that naive B cells and resting mast cells were mainly found in normal tissues, but few in tumor tissues, which was consistent with the results of bioinformatics analysis. However, some results are inconsistent with bioinformatics analysis. The M2 macrophage cells were mainly found in tumor tissues. The activated mast cells were widely infiltrated, but weakly in tumor parenchyma, and have a strong regional expression in normal tissues. The neutrophils were mainly found in normal tissues compared to tumor tissues. The M0 macrophage and M1 macrophage cells in tumor stroma were stronger than that in normal tissue, and the infiltration in tumor parenchyma was weaker than that in normal tissue. The inconsistency between the multiple immunofluorescence verification results and the bioinformatics analysis results may be because the data in TCGA and GEO databases are RNA expression, and RNA expression may be inconsistent with protein expression. In addition, the expression of different markers in tumor parenchyma and stroma is different. Bioinformatics analysis is for the whole tissue, while multiple immunofluorescence can see the location and number of different immune cells in different regions. Moreover, different patients have different conditions, such as the impact of radiotherapy and chemotherapy, which may also lead to the existence of inconsistency.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Xinxiang Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZZ and WS designed the experiments. JZ, YQ, PY, WX, BG, XQ, and YH performed the experiments and analyzed the data. ZZ wrote the manuscript. All authors contributed to the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. U1804173), Henan Province Young and Middle-Aged Health Science and Technology Innovation Talent Project (No. YXKC2021044), Innovation and Entrepreneurship Training Program for College Students of Xinxiang Medical College (No. 202010472045), Xinxiang Medical University Research Funding (No. XYBSKYZZ201632), Joint construction Project of Henan Medical Science and Technology Research Plan (No. LHGJ20190452), Natural Science Foundation of Henan Province (No. 202300410326), Zhongyuan Qianren Jihua of Henan Province (No. ZYQR201810153), and Teaching and Research Cultivation Project of School of Basic Medical Sciences, Xinxiang Medical College (No. JCYXYKY202021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.891270/full#supplementary-material

REFERENCES

Aponte-López, A., Fuentes-Pananá, E. M., Cortes-Muñoz, D., and Muñoz-Cruz, S. (2018). Mast Cell, the Neglected Member of the Tumor Microenvironment: Role in Breast Cancer. *J. Immunol. Res.* 2018, 1–11. doi:10.1155/2018/2584243

Balachandran, V. P., Gonen, M., Smith, J. J., and DeMatteo, R. P. (2015). Nomograms in Oncology: More Than Meets the Eye. *Lancet Oncol.* 16, e173–e180. doi:10.1016/S1470-2045(14)71116-7

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat. Med.* 24, 541–550. doi:10.1038/s41591-018-0014-x

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68, 394–424. doi:10.3322/caac.21492

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol. (Clifton, N.J.)* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Conway, E. M., Pikor, L. A., Kung, S. H. Y., Hamilton, M. J., Lam, S., Lam, W. L., et al. (2016). Macrophages, Inflammation, and Lung Cancer. *Am. J. Respir. Crit. Care Med.* 193, 116–130. doi:10.1164/rccm.201508-1545CI

Edlund, K., Madjar, K., Mattsson, J. S. M., Djureinovic, D., Lindskog, C., Brunnström, H., et al. (2019). Prognostic Impact of Tumor Cell Programmed Death Ligand 1 Expression and Immune Cell Infiltration in NSCLC. *J. Thorac. Oncol.* 14, 628–640. doi:10.1016/j.jtho.2018.12.022

Finotello, F., and Trajanoski, Z. (2018). Quantifying Tumor-Infiltrating Immune Cells from Transcriptomics Data. *Cancer Immunol. Immunothercii* 67, 1031–1040. doi:10.1007/s00262-018-2150-z

Genard, G., Lucas, S., and Michiels, C. (2017). Reprogramming of Tumor-Associated Macrophages with Anticancer Therapies: Radiotherapy versus Chemo- and Immunotherapies. *Front. Immunol.* 8, 828. doi:10.3389/fimmu.2017.00828

Hobson-Gutierrez, S. A., and Carmona-Fontaine, C. (2018). The Metabolic axis of Macrophage and Immune Cell Polarization. *Dis. Model. Mech.* 11, 034462. doi:10.1242/dmm.034462

Jensen, B. M., Frandsen, P. M., Raaby, E. M., Schiøtz, P. O., Skov, P. S., and Poulsen, L. K. (2014). Molecular and Stimulus-Response Profiles Illustrate Heterogeneity between Peripheral and Cord Blood-Derived Human Mast Cells. *J. Leukoc. Biol.* 95, 893–901. doi:10.1189/jlb.0712354

Kim, Y., Shin, Y., and Kang, G. H. (2019). Prognostic Significance of CD103+ Immune Cells in Solid Tumor: a Systemic Review and Meta-Analysis. *Sci. Rep.* 9, 3808. doi:10.1038/s41598-019-40527-4

Komi, D. E. A., and Redegeld, F. A. (2020). Role of Mast Cells in Shaping the Tumor Microenvironment. *Clinic Rev. Allerg Immunol.* 58, 313–325. doi:10.1007/s12016-019-08753-w

Lai, C.-Y., Tseng, P.-C., Chen, C.-L., Satria, R. D., Wang, Y.-T., and Lin, C.-F. (2021). Different Induction of PD-L1 (CD274) and PD-1 (CD279) Expression in THP-1-Differentiated Types 1 and 2 Macrophages. *Jir* Vol. 14, 5241–5249. doi:10.2147/JIR.S329921

Lu, Y., Huang, Y., Huang, L., Xu, Y., Wang, Z., Li, H., et al. (2020). CD16 Expression on Neutrophils Predicts Treatment Efficacy of Capecitabine in Colorectal Cancer Patients. *BMC Immunol.* 21, 46. doi:10.1186/s12865-020-00375-8

Lv, R., Bao, Q., and Li, Y. (2017). Regulation of M1-type and M2-type Macrophage Polarization in RAW264.7 Cells by Galectin-9. *Mol. Med. Rep.* 16, 9111–9119. doi:10.3892/mmr.2017.7719

Molderings, G. J., Zienkiewicz, T., Homann, J., Menzen, M., and Afrin, L. B. (2017). Risk of Solid Cancer in Patients with Mast Cell Activation Syndrome: Results from Germany and USA. *Results from Germany and USAF1000Res* 6, 1889. doi:10.12688/f1000research.12730.1

Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., and Freeman, T. C. (2018). Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunol. Res.* 6, 1388–1400. doi:10.1158/2326-6066.CIR-18-0342

Orihuela, R., McPherson, C. A., and Harry, G. J. (2016). Microglial M1/M2 Polarization and Metabolic States. *Br. J. Pharmacol.* 173, 649–665. doi:10.1111/bph.13139

Petty, A. J., Owen, D. H., Yang, Y., and Huang, X. (2021). Targeting Tumor-Associated Macrophages in Cancer Immunotherapy. *Cancers* 13, 5318. doi:10.3390/cancers13215318

Riabov, V., Gudima, A., Wang, N., Mickley, A., Orekhov, A., and Kzhyshkowska, J. (2014). Role of Tumor Associated Macrophages in Tumor Angiogenesis and Lymphangiogenesis. *Front. Physiol.* 5, 75. doi:10.3389/fphys.2014.00075

Sanz, I., Wei, C., Jenks, S. A., Cashman, K. S., Tipton, C., Woodruff, M. C., et al. (2019). Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. *Front. Immunol.* 10, 2458. doi:10.3389/fimmu.2019.02458

Sun, Q., He, M., Zhang, M., Zeng, S., Chen, L., Zhao, H., et al. (2021). Traditional Chinese Medicine and Colorectal Cancer: Implications for Drug Discovery. *Front. Pharmacol.* 12, 685002. doi:10.3389/fphar.2021.685002

Ubil, E., Caskey, L., Holtzhausen, A., Hunter, D., Story, C., and Earp, H. S. (2018). Tumor-secreted Pros1 Inhibits Macrophage M1 Polarization to Reduce Antitumor Immune Response. *J. Clin. Invest.* 128, 2356–2369. doi:10.1172/JCI97354

Valent, P., Bonadonna, P., Hartmann, K., Broesby-Olsen, S., Brockow, K., Butterfield, J. H., et al. (2019). Why the 20% + 2 Tryptase Formula Is a Diagnostic Gold Standard for Severe Systemic Mast Cell Activation and Mast Cell Activation Syndrome. *Int. Arch. Allergy Immunol.* 180, 44–51. doi:10.1159/000501079

Ventola, C. L. (2017). Cancer Immunotherapy, Part 3: Challenges and Future Trends. *P T* 42, 514–521.

Waldman, A. D., Fritz, J. M., and Lenardo, M. J. (2020). A Guide to Cancer Immunotherapy: from T Cell Basic Science to Clinical Practice. *Nat. Rev. Immunol.* 20, 651–668. doi:10.1038/s41577-020-0306-5

Xi, Y., and Xu, P. (2021). Global Colorectal Cancer burden in 2020 and Projections to 2040. *Translational Oncol.* 14, 101174. doi:10.1016/j.tranon.2021.101174

Xiong, Y., Wang, K., Zhou, H., Peng, L., You, W., and Fu, Z. (2018). Profiles of Immune Infiltration in Colorectal Cancer and Their Clinical Significant: A Gene Expression-based Study. *Cancer Med.* 7, 4496–4508. doi:10.1002/cam4.1745

Yamaguchi, T., Fushida, S., Yamamoto, Y., Tsukada, T., Kinoshita, J., Oyama, K., et al. (2016). Tumor-associated Macrophages of the M2 Phenotype Contribute to Progression in Gastric Cancer with Peritoneal Dissemination. *Gastric Cancer* 19, 1052–1065. doi:10.1007/s10120-015-0579-8

Zajac, E., Schweighofer, B., Kupriyanova, T. A., Juncker-Jensen, A., Minder, P., Quigley, J. P., et al. (2013). Angiogenic Capacity of M1- and M2-Polarized Macrophages Is Determined by the Levels of TIMP-1 Complexed with Their Secreted proMMP-9. *Blood* 122, 4054–4067. doi:10.1182/blood-2013-05-501494

Zhang, Y., and Zhang, Z. (2020). The History and Advances in Cancer Immunotherapy: Understanding the Characteristics of Tumor-Infiltrating Immune Cells and Their Therapeutic Implications. *Cell Mol Immunol* 17, 807–821. doi:10.1038/s41423-020-0488-6

Zhao, X., Qu, J., Sun, Y., Wang, J., Liu, X., Wang, F., et al. (2017). Prognostic Significance of Tumor-Associated Macrophages in Breast Cancer: a Meta-Analysis of the Literature. *Oncotarget* 8, 30576–30586. doi:10.18632/oncotarget.15736

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhong, Qin, Yu, Xia, Gu, Qian, Hu, Su and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 18 May 2022
doi: 10.3389/fgene.2022.834445

Systems-Level Mapping of Cancer Testis Antigen 1b/a to Sarcoma Pathways Identifies Activated Ran Binding-2 E3 SUMO-Protein Ligase and Transducin-Like Enhancer Protein 1

Nikolaos A. Papanikolaou[1]*, *Prodromos Hytiroglou*[2], *Pavlina Pantelidou*[2], *Athanasios G. Papavassiliou*[3] and *Lloyd L. Old*[4†]

[1]Laboratory of Biological Chemistry, Department of Medicine, Section of Biological Sciences and Preventive Medicine, School of Medicine, Aristotle University of Thessaloniki, Macedonia, Greece, [2]Department of Medicine, Laboratory of Pathology, School of Medicine, Aristotle University of Thessaloniki, Macedonia, Greece, [3]Department of Biological Chemistry, Medical School, National and Kapodistrian University of Athens, Athens, Greece, [4]Ludwig Institute for Cancer Research, Memorial Sloan Kettering Institute for Cancer Research, New York City, NY, United States

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Pabitra Kumar Parua,
Albert Einstein College of Medicine, United States
Tupa Basuroy,
Massachusetts General Hospital and Harvard Medical School, United States

***Correspondence:**
Nikolaos A. Papanikolaou
papnik@med.auth.gr

†Deceased

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 12 January 2022
Accepted: 14 March 2022
Published: 18 May 2022

Citation:
Papanikolaou NA, Hytiroglou P, Pantelidou P, Papavassiliou AG and Old LL (2022) Systems-Level Mapping of Cancer Testis Antigen 1b/a to Sarcoma Pathways Identifies Activated Ran Binding-2 E3 SUMO-Protein Ligase and Transducin-Like Enhancer Protein 1. Front. Genet. 13:834445. doi: 10.3389/fgene.2022.834445

Here we describe the identification of genes and their encoded proteins that are expressed in advanced grade tumors by reconstruction of a sarcoma cancer testis gene 1b/a (*catg1b/a*) network. CTAG1B/A is an ortholog of the yeast/Drosophila transcription factor Pcc1p, and a member of the KEOPS transcription complex. It has been implicated in telomere maintenance and transcriptional regulation through association with chromatin remodeling factors and is only expressed during adult testis germ cell differentiation. C*tag1b/a* is re-activated in synovial sarcomas and myxoid liposarcomas but not in differentiated liposarcomas. We mapped CTAG1B/A protein to sarcoma transcription pathways with gene set expression analysis (GSEA) and using independent samples, we immunohistochemically identified expression of at least two network neighbors, RANBP2, and TLE1, thus validating our approach. This work demonstrates that mapping unknown genes to functional pathways by network reconstruction is a powerful tool that can be used to identify candidate oncoproteins.

Keywords: CtaG, GSEA, network, sarcoma, RanBP2, TLE1, transcription

INTRODUCTION

The cancer testis genes (*ctags*) are a heterogeneous group of genes, that is, exclusively expressed in gametogenic germ cells (Ait-Tahar et al., 2009), and trophoblasts (Chang and Nevins, 2006), but is aberrantly re-expressed in different tumors (Ait-Tahar et al., 2009; Cho et al., 2006; Cooke and Nanjappa, 2015). The absence of expression in normal adult tissues makes these genes and their protein products unique as therapy targets and as potential diagnostic tools or markers (DeRisi and Iyer, 1999; DeRisi et al., 1997; Epping et al., 2005). With the exception of the SCP-1 (Fordham et al., 2020) and OY-TES-1 CT proteins (Gioutlakis et al., 2017), virtually nothing is known about their biological functions either in gametogenesis, trophoblast biology or tumor development. Identifying the biological functions of *ctag* gene products by mapping them to active pathways is therefore

critical for defining their role in normal development and in cancer growth. In addition, analysis of their networks, will facilitate the identification of novel gene products with hitherto unknown functions (Gnjatic et al., 2006; Gjerstorff et al., 2007; Gjerstorff et al., 2007). A serious obstacle to this endeavor, however, is the random patterns of expression in tumors and the relative lack of human germ cell development models (Hecker et al., 2008; Grigoriadis et al., 2009).

Ctag1b/a was the first cancer testis antigen gene to be discovered in cancer patients using an *in vivo* antibody response and successfully targeted by a vaccine (Hei et al., 2019). It is localized on the Xq28 region of the X chromosome encoding a 180 long amino acid protein. *Ctag1b/a* is found in approximately one third of all melanoma, lung, and esophageal tumors, from which it was first isolated, as well as in liver, gastric, prostate, ovarian and bladder cancers (Henderson et al., 2005). It occurs in 80% of synovial sarcomas, a rare but highly aggressive tumor type (Hua et al., 2014), and also in myxoid liposarcomas (Iura et al., 2015; Irie et al., 2018), a type of aggressive mesenchymal tumor but not in well differentiated liposarcomas (Jiang et al., 2017). Single and double immunohistochemical studies in human germ cells have revealed that CTAG1B/A protein is associated with highly proliferating germ cells (Jungbluth et al., 2001). Its expression during male germ development is limited to a narrow window spanning mitosis and meiosis (I) events at the types A and B spermatogonium stages (Kakimoto et al., 2019). In melanoma cells the CTAG1B/A protein is found in complexes with MAGEC1, another cancer testis antigen protein and a putative transcription factor, which is consistent with the narrow window of expression of both proteins in differentiating germ cells. The CTAG1B/A protein is a homolog of the yeast Pcc1p transcription factor (Kisseleva-Romanova et al., 2006)] and a member of EKC complexes, which control the yeast cell cycle, raising the possibility that CTAG1B/A is a germ cell-specific transcription factor. The above data suggest that the two proteins may have common functions mediated by a CTAG1B/A/MAGEC1/KEOPS complex (Kalejs and Erenpreisa, 2005).

We reasoned that by extracting gene and transcriptional modules that are co-expressed or co-regulated with *ctag1b/a*, we could map it to tumor pathways and use its network properties to identify network neighborhood genes that are also expressed in sarcoma samples (Klapa et al., 2013; Kobayashi and Surani, 2018). Functional modules of co-expressed genes reflect coherent changes in biological properties beyond individual genes and characterize the behavior of cells at higher levels of organization (Liggins et al., 2010). Using transcriptomics samples, we have extracted *ctag1b/a* co-expression modules from different tumor types in which *ctag1b/a* is re-expressed, and reconstructed a sarcoma *ctag1b/a* network. Next, we identified several network neighbors that are candidates involved in sarcoma biology. Using different sarcoma samples, we immunohistochemically confirmed expression of some of the *ctag1b/a* network neighbors. *Ctag1b/a* is enriched in transcription factor pathways, suggesting that it is likely a germ cell-specific transcription factor, that is, re-expressed, albeit randomly, in adult sarcomas and in various other tumors. Thus, by using reverse network engineering, we demonstrate that it is possible to discover and map co-expressed genes in sarcoma biology pathways.

MATERIALS AND METHODS

Microarray Data Collection and Analysis of *Ctag1b/a* Differential Expression

Gene expression data were downloaded as text (txt.) files from the Gene Expression Omnibus (GEO) at NCBI. 32 microarrays deposited at NCBI's GEO or EBI's Array Express databases were retrieved and analyzed for *ctag1b/a* differential expression between normal vs. primary or normal vs. metastatic tissue, and isolated glioblastoma stem cells vs. normal adult human tissues (**Supplementary File S1**).

Mining Interactions and Interrelationships

First and second level interactors of the CTAG1B/A protein were extracted from the PICKLE database (Klapa et al., 2013; Gioutlakis et al., 2017) which contains comprehensive data from all known interaction databases and integrates knowledge with genetic data. Data were filtered for confirmed physical/functional interactions and, in order to minimize false positives, only those with an E value > 0.7 were retained.

Extraction of Co-regulated *Ctag1b/a* Gene Modules From Sarcoma Samples

Ctag1b/a co-expression gene modules were extracted from synovial and myxoid (grade 3) sarcomas as well as various benign fibrosarcomas (grade1-2) using the three step strategy shown in **Figure 2A** (**Supplementary Appendix SB** for the computational methods). Our working hypothesis was that genes that are co-expressed in sarcomas with *ctag1b/a* are also likely to be co-regulated if they are in the network neighborhood of *ctag1b/a*. We implemented our strategy as follows: First, we established whether *ctag1b/a* is differentially expressed between two different classes of samples, such as, for example, between normal vs. metastatic tissue, benign vs. primary tumor, etc, (**Supplementary File S1**). Next, we subjected the samples to pair-wise analysis using either Student's t test (null hypothesis: The medians/means of expression between two classes of samples are not significantly different at $p < 0.01$) or Mann-Whitney's non-parametric statistic for rank distribution in order to find if the expression levels of *ctag1b/a* between two sample classes differed significantly between classes. We have found that *ctag1b/a* was induced at significant levels in eight out of 32 downloaded microarrays ($p < 0.01$, **Supplementary File S1**, datasets where *ctag1b/a* is induced significantly are in red). Microarray data with $p > 0.01$ were rejected. The choice of the two tests depended on the number of replicates in each sample class. Only three sets of data passed the strict cutoff point $p = < 0.01$, microarrays GDS1209a and b and GDS2736. A second reason for rejecting the other data was the random distribution of *ctg1b/a* re-expression in both advanced (metastatic) and non-metastatic tumor samples. We then derived gene modules that are co-

expressed either positively or negatively with *ctag1b/a* using gene set enrichment analysis (GSEA) as described below.

Gene Set Enrichment Analysis for Gene Sets Showing Co-Regulation With Ctag1b/a

We extracted ctag1b/a gene modules in two steps, as described in the previous section. Next, we used gene set enrichment analysis (GSEA, **Supplementary Appendix SB**) to obtain gene sets enriched either for CTAG1B/A_POS or negative expression, (CTAG1B/A_NEG) (Parameters and results in **Supplementary File S2**).

Ctag1b/a Synovial Network Reconstruction

The retrieved binary interaction data from the PICKLE database were collected in an Excel file and used a txt file in Cytoscape. Only data with an E value equal or greater than 0.7 were considered. Using as queries the sarcoma genes in the GSEA-extracted modules, the genes found in the indicated gene collections (Papanikolaou gene sets compiled by the Papanikolaou lab, **Figure 2**), mitosis-meiosis gene sets, and the genes in the Coxpred database that are known to be co-regulated with *ctag1b/a* as well as their first and second-level physical/genetic interactors, we reconstructed a *ctag1b/a* synovial network with Cytoscape and analyzed its neighborhood network properties using the MCODE algorithm for clustering highly interconnected genes and for finding multiprotein complexes (modules) and also cytoHubba for identifying hub genes and sub-networks in the neighborhood of *ctag1b/a*.

Gene Ontology Analysis

We subjected the top 20 genes from each of the CTAG1B/A_POS and CTAG1B/A_NEG modules to systems-level gene ontology (GO) analysis with the GATHER algorithm (Chang and Nevins, 2006). Enrichment for functional GO pathways was performed based on statistical significance using a GATHER algorithm-based Bayesian approach (**Supplementary Files S4, S5**, respectively).

Immunohistochemistry

Sections of formalin-fixed, paraffin-embedded tissues derived from synovial sarcomas were obtained from the archives of the Pathology Laboratory, Department of Medicine, Aristotle University. Ganlioneuroma IHCs were a gift from Dr. Achim Jungbluth, Laboratory of Pathology, Ludwig Institute for Cancer Research in New York, United States. In all cases, slides were reviewed and representative blocks were chosen for immunohistochemical analysis. The following primary monoclonal (MAb) or polyclonal (PAb) antibodies were used at concentrations ranging from 1:200 to 1:1,000 (primary antibody): CTAG1B/ (sc-53869), CT7-33 mouse MAb against MAGEC1 (sc-53868), Ranbp2 (ABCAM, PAB 2938TRS), TLE1 (ABCAM, PAb, 15587syn), MDM2 (Santa-Cruz Biotechnology, D-7, sc-13161). Tissue sections were deparaffinized in xylene and graded alcohols. The stains were performed in an automated staining system (Bond-max, Vision Biosystems, Germany). For antigen retrieval the slides were heated at 99°C in a 10 mM sodium citrate solution (pH 6.0) for 10 min using a vegetable steamer (Oster 5712 food steamer, Maitland, Florida, United States). Labeling was visualized using 3, 3′-diaminobenzidine (DAB, Vector Laboratories, Burlingame, California, United States) or NovaRED (Vector Laboratories). Alternatively, the Thermo Fisher Scientific MultiVision Polymer Detection System (Thermo Fisher Scientific, Lab Vision Corporation, Fremont, California, United States) was used for detection, and development was achieved using 3, 3′-diaminobenzidine (Vector Laboratories) followed by application of NovaRED (Vector Laboratories). Slides were counterstained with Mayer's haematoxylin. For negative controls non-immune serum of the same species was applied as primary antibody.

RESULTS

Re-Activation of Ctag1b/a Expression in Different Tumors

In order to map *ctag1b*/a to cancer pathways and to discover co-expressed genes in its network neighborhood, we exploited its re-activation in synovial and myxoid liposarcomas but not in dedifferentiated liposarcomas or myofibrosarcomas and leiomysosarcomas (**Figure 1**) and mined its physical and genetic interactions. We then re-constructed a *ctag1b/a* sarcoma network and used its properties to identify genes in the neighborhood of *ctag1b/a*. It has been reported that expression of *ctag1b/a* (aliases include: *ny-eso-1*, *ct6.1*, *ctag2*, *lage-2*, *lage-2a* and *lage2b*) is re-activated, albeit randomly, in several different tumor types but without the distinct expression profile seen in sarcomas (**Supplementary File S1**, **Figure 1**). We therefore focused on sarcoma samples and using the three-step strategy shown in **Figure 2**, we extracted *ctag1b/a* co-expression modules (step 2 in **Figure 2A**), as described in Section 2. We then subjected the gene lists to network analysis and identified *ctag1b/a* network neighbors and key hubs linked with shortest paths (with the cytoHubba algorithm in Cytoscape) in the *ctag1/a* neighborhood. Finally, using immunohistochemical staining of separate, independent samples, we examined the protein expression status in independent sarcoma or control ganglioneuroma samples and confirmed their expression. Two of these genes, *tle1* and *ranbp2*, have previously been implicated in sarcoma biology, thus validating our approach. Our findings have potential applications for the identification of novel synovial sarcoma sub-networks and genes with hitherto unknown functions.

The *ctag1b/a* query in the Gene Expression Omnibus (GEO) database returned 4033 *ctag1b/a* expression profiles for different tumor types. Of these, only thirty two (Scanlan et al., 2002b) merited further examination (**Supplementary File S1**). Ten out of thirty two (31%) microarray samples exhibit significant ($p < 0.05$) differences in *ctag1b/a* expression between samples (**Supplementary File S1**). Microarray data with $p \geq 0.01$ were not considered further. In contrast, in two studies (GDS2736 and GDS1209) almost all samples of stage 3 synovial sarcomas and myxoid liposarcomas uniformly express *ctag1b/a* whereas various

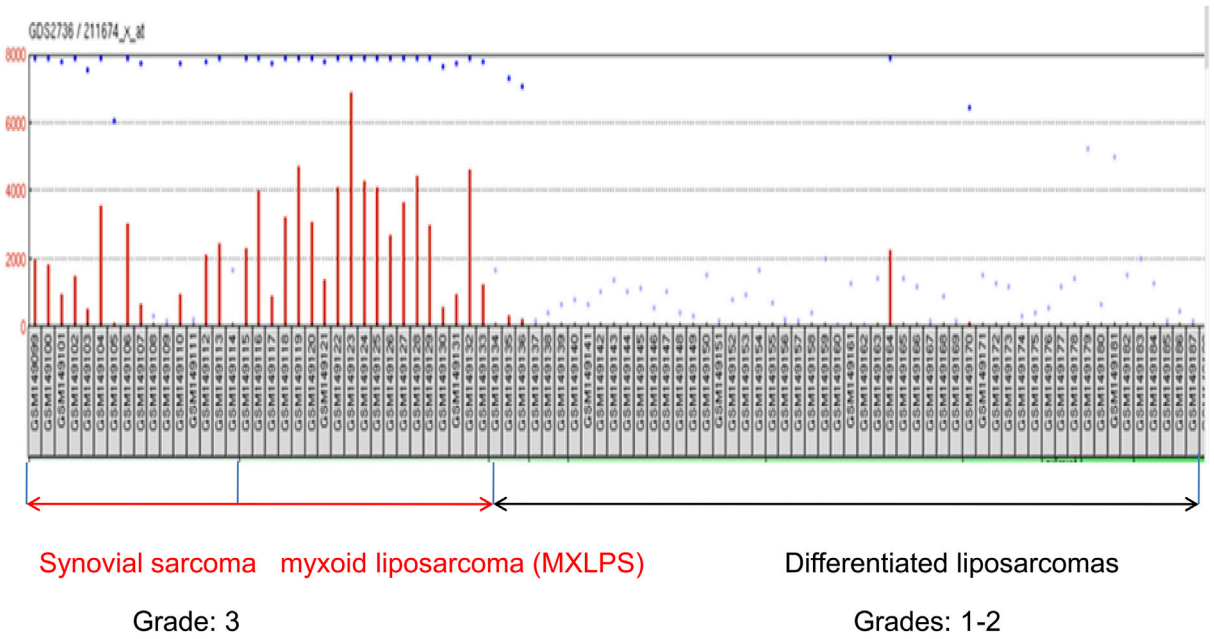


FIGURE 1 | Expression of *ctag1b/a* mRNA in different sarcoma samples retrieved from the GDS2736 archive. Notice that with the exception of two samples, various differentiated sarcomas of grades 1–2 uniformly lack expression of *ctag1b/a*. In contrast, *ctg1b/a* is re-expressed in most (grade 3) synovial and myxoid liposarcomas samples as well as dedifferenatiated liposarcomas. Expression unitson the Y axis are arbitrary transcriptomics units. Retrieved samples are designated by GSM, followed by a number.

differentiated liposarcomas do not show re-activation (**Figure 1** and **Table 1**). With the exception of sarcoma, melanoma and CD133 positive glioblastoma samples, *ctag1b/a*, is randomly re-activated in most samples examined and does not follow any pattern, that is, characteristic of stage or tumor type. Thus, in three melanoma studies (GDS 1375, GDS 1965, and GDS1078) *ctag1b/a* is re-activated in 33% of samples. Of the fourteen breast cancer samples only one group (GSE3156) exhibits re-activation. All others display a non-uniform, dispersed re-expression pattern (Cooke and Nanjappa, 2015). On the other hand, 33% of CD133 positive glioblastoma samples exhibit re-activation whereas CD133 negative ones do not re-express it at all.

We focused on sarcoma samples because they exhibited a distinct *ctag1b/a* expression pattern. We subjected the different groups of samples to pair-wise analysis using either the Mann-Whitney non-parametric statistic for rank distribution of *ctag1b/a* expression levels, or Student's paired t-test (**Figure 2B**). The choice of the two tests depended on the number of replicates in each sample class. *Ctag1b/a* was found to be significantly overexpressed ($p < 10^{-7}$) in myxoid liposarcomas and synovial sarcomas but not in lipomas or leiomyosarcomas, a significant finding.

Identification of Ctag1b/a Synovial Re-Expression Gene Module Signatures

Using GSEA we extracted two phenotypic classes from a compendium of ten sarcoma types in two studies (GDS1209), representing gene modules whose co-expression pattern either correlates (CTAG1B/A-POS) or anti-correlates (CTAG1B/A_NEG) with *ctag1b/a* re-expression (data in **Supplementary Files S2, S3**). The GSEA-generated phenotypic classes (No expression vs. expression, **Supplementary Figures S1A,B**) were separated according to their GSEA-computed enrichment score (ES) and plotted against their *p* value. *Ctag1b/a* was used as an index gene (details in Computational Methods section; **Supplementary Files S2, S3** contain all genes identified above a cutoff threshold normalized enrichment score (NES) (NES = enrichment score normalized to mean enrichment of random samples of the same size) as having either a positive or negative correlation with *ctag1b/a*, the index gene in GSEA. The extracted gene set-containing phenotypic classes were assigned a positive or negative enrichment score (ES, **Supplementary Figure S1A**, upper left panel, A) depending on their rank position in the Kolomogorov-Smirnov list (**Supplementary Figure S1A**, lower middle panel).

Genes that are enriched in the CTAG1B/A-POS phenotype are up-regulated (positive ES), whereas those in the CTAG1B/A_NEG phenotype are down-regulated, hence the negative ES. The gene sets within each phenotype represent actual biological pathways involved in processes enriched with respect to *ctag1b/a* re-activation and are assigned by GSEA a NES and a *p* value which determine their position in the ranked list (**Supplementary Files S2, S3**). The lower the NES value the more significant the gene set. For example, the top gene set enriched in this phenotype is the GPCRD_rhodopsin_like set with 169 gene members and a NES of −3.8 (see *p*-value plots sheet in **Supplementary File S2**). *Ctag1b/a* was assigned a score equal to zero and was used as an index gene in GSEA, followed by

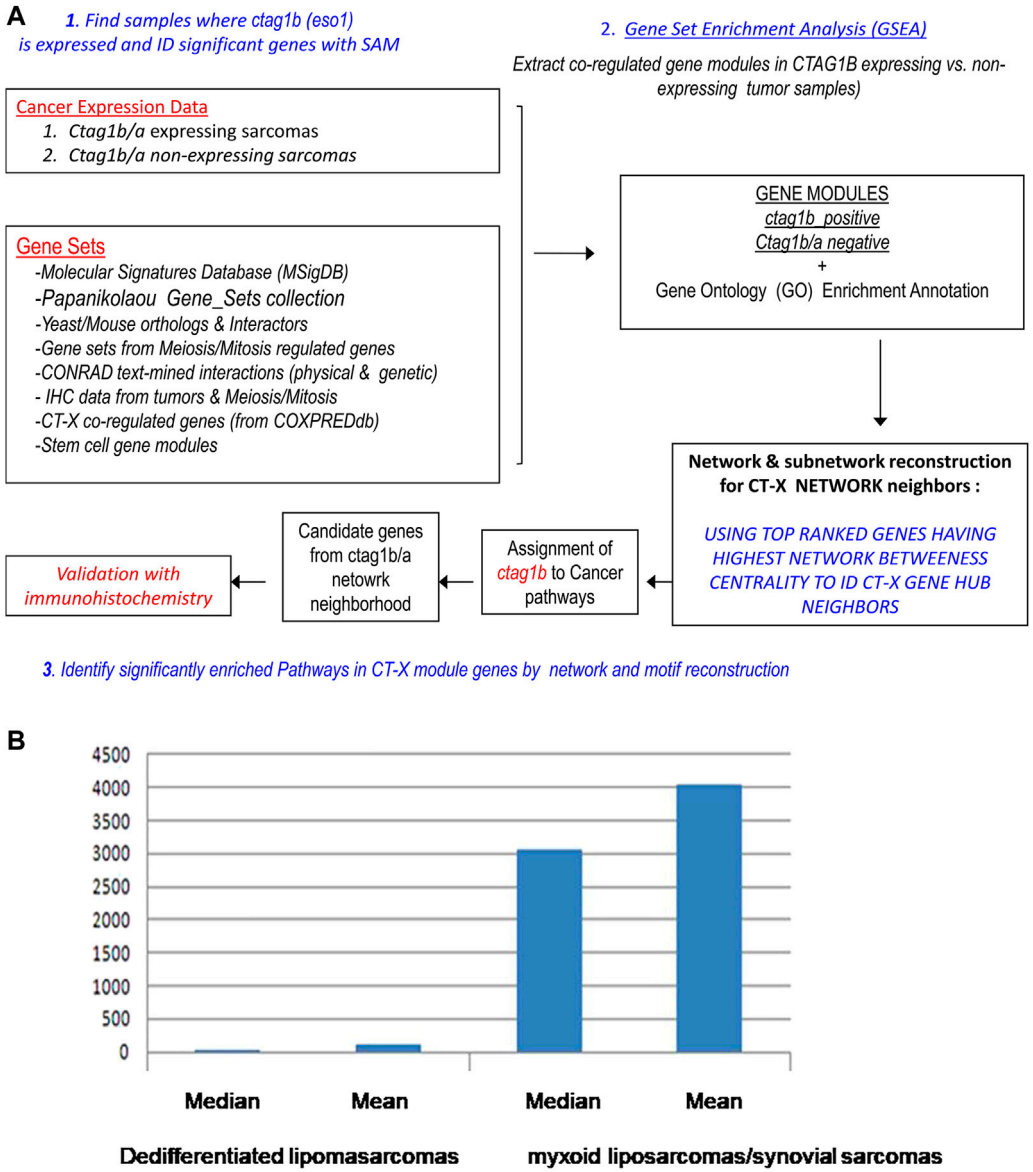


FIGURE 2 | (A) Strategy for the extraction and characterization of synovial *ctag1b/a* co-expression modules, reconstruction of a synovial *ctag1b/a* network and identification of its neighbor genes in the network. **(B)** Median and Mean values of *ctag1b/a* expression in differentiated liposarcomas or myxoid liposarcomas/synovial sarcomas. Pairwise comparison was with the T statistic (*p* value was equal to 3.3×10^{-7}). The null hypothesis is that there is no difference in *ctag1b/a* expression between the two groups of tumors. Expression units on Y axis are arbitrary transcriptomics units.

computation of the KS sum for all genes in the array. The smaller the KS sum for a particular gene the closer that gene is to *ctag1b/a* in terms of expression profile enrichment. On the other hand, the larger the KS score the more distant in terms of expression profile is the gene. Thus, the gene closest to *ctag1b/a* in this list is LOC349160 (**Supplementary File S3**), an RNA gene, and belongs to an lncRNA class in the human genome followed by the rest in the list going downwards. Of the 1,245 gene sets in the MSig database, 1,171 are positively upregulated with *ctag1b/*, whereas only 74 of 1,245 are downregulated in the negative phenotype (**Supplementary File S2**), suggesting that *ctag1b/a* re-activation is linked to many activated sarcoma pathways. Notable among the twenty five top genes found to be co-expressed with *ctag1b/a* in sarcomas are protocadherin β3, tumor necrosis factor alpha-induced protein 8-like 3, demethylase jumonji domain containing 2C, cyclin-dependent kinase 6, and glycine dehydrogenase (decarboxylating) (**Supplementary File S3**).

Gene ontology enrichment (GO) analysis showed that the CTAG1B/A-POS gene module is almost exclusively enriched in genes for transcription factor activity (**Supplementary Figure S1A**, lower right panel). In contrast, the CTAG1B/A_NEG module (**Supplementary Figure S1A**, lower right panel) is enriched in structural and macromolecular synthesis genes

TABLE 1 | Different tumor types used in extracting *ctag1b/a* modules from trascriptomics data. Shown in red are samples of advanced sarcoma types as well as melanomas that express *ctag1b/a*. Samples were grouped by tumor type and ctag1b/a expression status.

Sample grouping of tumors by *ctag1b/a* expression status		
Tumor type	**Number of samples**	**ctag1b/a expression status**
Primary/normal melanoma	3	No
nNSCLC tissue (lung cancer cell lines/non-metastatic	3	No
Lipomas	16	No
Fibromyosarcoma	8	No
NSCLC	3	Yes
Myxoid liposarcoma	15	Yes
Synovial sarcoma	10	Yes
Metastatic melanoma	5	Yes
Total	63	—

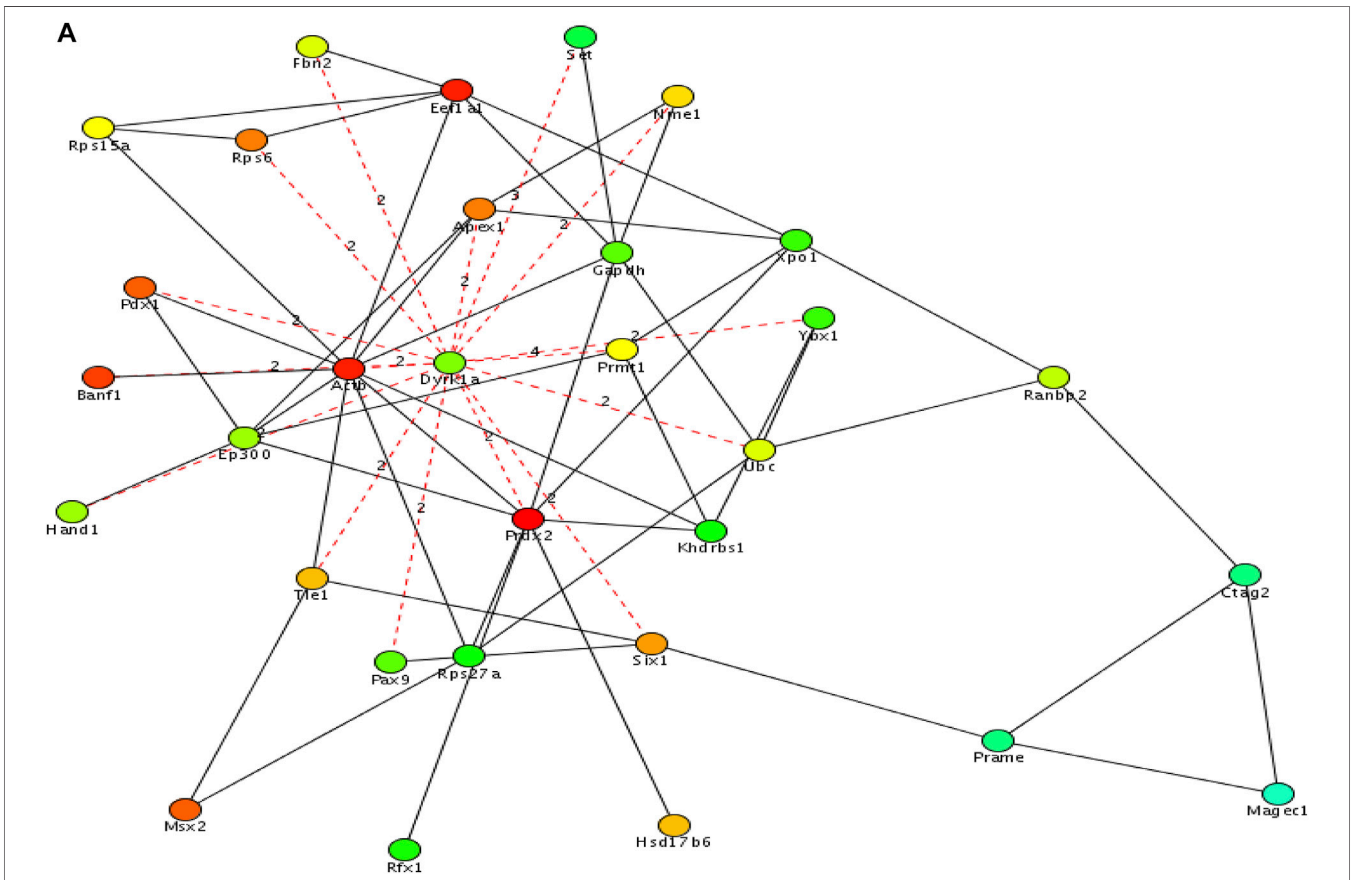


FIGURE 3 | (A) *Ctag1b/a* STRING network. *ctag1b, ctag1a* are aliases of *ctag1b/a*. The network neighborhood was reconstructed from first and second level interactors of *ctag1b/a*, *pasd1* and ZIC3 are transcription factors. *Pasd1* links the *ctag1b/a* group (lower group containing the *mages, cage1* and others) with chromatin accessory proteins (upper group containing *hdac1, chd4* and others). Network was generated with Cytoscape. **(B)** The ctag1b/a network neighborhood. Shortest paths, shown in broken lines, between *ctag1b/a, magec1, prame* and top important genes in the network **(Table 2)** were calculated with the Dijkstra algorithm within Cytoscape. Top ranked nodes/bottlenecks are shown in different colors with each rank in a different color with highest ranked nodes in red, followed by pink and green for lower ranked nodes. Dashed red lines are calculated shortest paths between nodes in the *ctag1b/a* network neighborhood in sarcoma samples. Solid lines: Functional or physical edges (links) between nodes. Broken lines: Shortest paths (computed with the Dijkstra algorithm within Cytoscape) from the network in Panel 3A. *Ctag2* is an alias of *ctag1b/a* (also *ctag1b* or *ctag1a*). Expression of *tle1*and *ranbp2* was tested in independent archived synovial sarcomas (Laboratory of Pathology, Aristotle University School of Medicine, Macedonia, Greece) and ganglioneuromas (Ludwig Institute for Cancer Research, New York, NY United States). The networks were generated with Cytoscape. **(C)** Top GO categories for genes in *ctag1a/magec1* network neighborhood (Panel 3A and **Table 2**). Note: Regulation of transcription and Transcription regulation GO categories contain complementary though not identical genes. *p* values were calculated with the hypergeometric method using random sets of genes from the GO database as reference sets shown in Panel 3B.

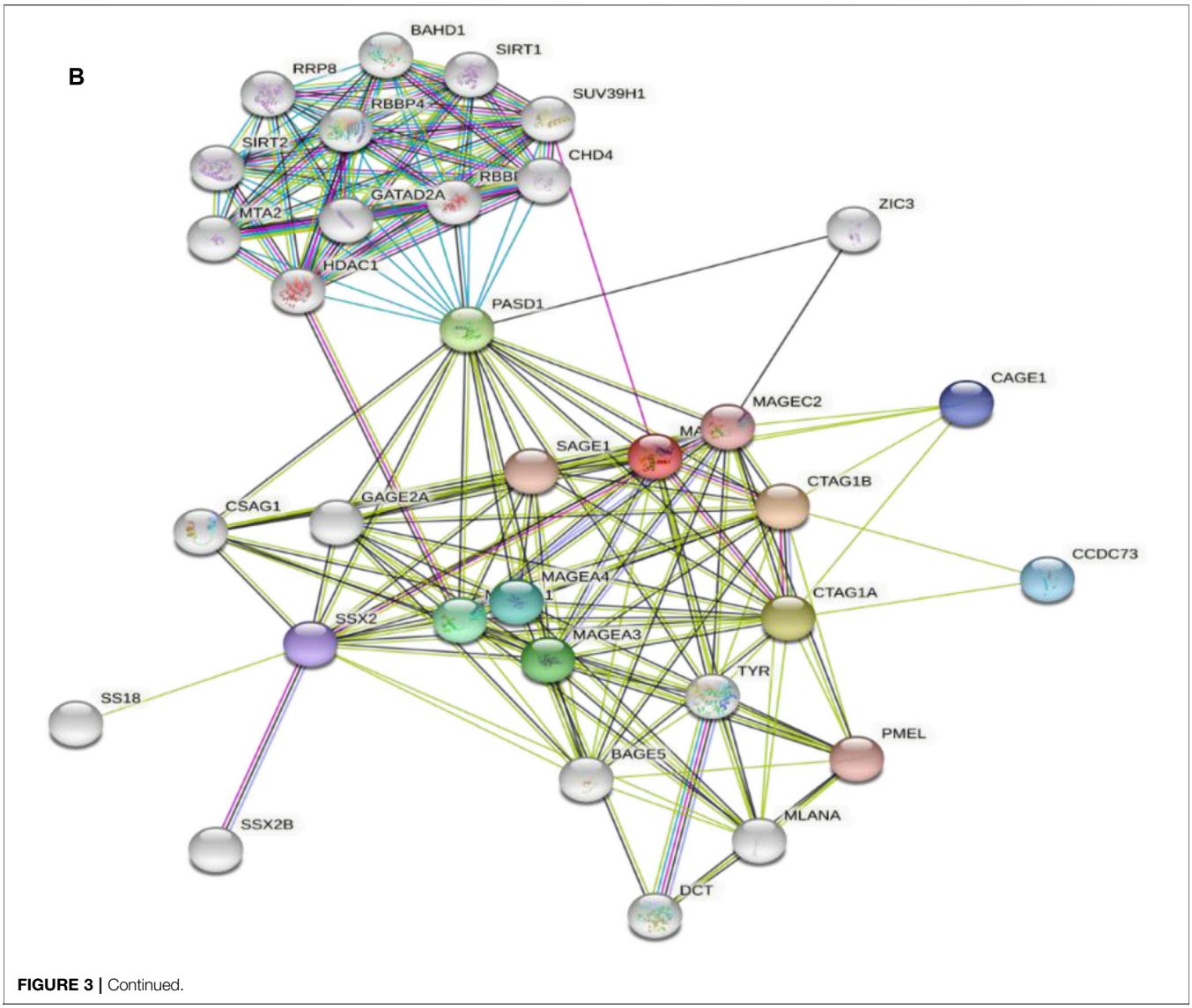


FIGURE 3 | Continued.

such as ribosomal genes. Several genes in the CTAG1B/A-POS module, such as *mdm2*, and *cdk6*, have established oncogenic roles, whereas others remain uncharacterized. Examples include the zinc finger gene 588 (*znf588*), protocadherin β3, laminin α1 and cyclin b3 (**Supplementary File S3**, sheet 1: 50 co-expressed genes/Heat Map).

Although the genes in each CTAG1B/A-POS module in the microarray datasets were different, their enrichment with KEGG pathways, GO functions and transcription factor binding sites (**Supplementary Figure S1A**, upper right panel) suggested that they belonged to gene groups involved in morphogenesis (GO: 0009653), cell motility, organogenesis, cell migration etc, as indicated by their decreasing ln (Bayes factor) [**Supplementary File S4** and 5 under ln (Bayes factor)]. In contrast, the CTAG1B/A_NEG module consistently contained ribosomal genes in all ten microarrays (**Supplementary File S4**). GO annotation enrichment analysis of these genes returned protein biosynthesis (GO: 0006412), macromolecular biosynthesis and other enriched annotations as top functional categories.

Re-Construction of a Ctag1b/a Synovial Network

Using the genes and their first and second-level physical and genetic interactors (**Figure 2**, step 2), we constructed a *ctag1b/a*-centered sarcoma network with Cytoscape. The sarcoma network consists of 1,140 nodes (genes), and 7,972 edges (interactions/interrelationships). It has a clustering coefficient equal to 0.163 and an average path length equal to 3.443 (**Figure 3A** shows part of the *ctag1b/a* network) values that are typical for biological networks. Notable among members of this network is *pasd1* which encodes a transcription factor linking a group of genes encoding a cluster of chromatin proteins such as *hdac1*, *chd4*, *mta1* (**Figure 3A**, upper group). *Pasd1* and

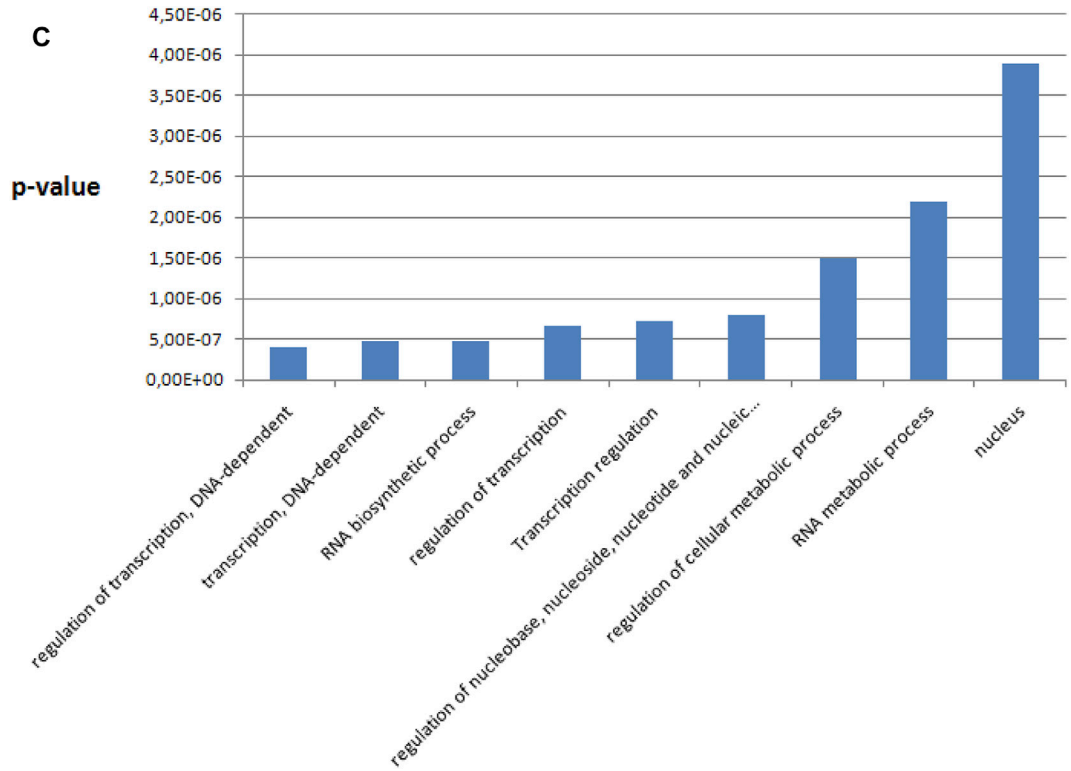


FIGURE 3 | Continued.

zic3 link the chromatin gene cluster with the *ctag1b/a* neighborhood and *with magec2*.

Several features are apparent in the sub-network neighborhood in **Figure 3B**: First, the network neighborhood suggests that *ctag1b/a* (aliased as *ctag2* in the sub-network), *prame* and *magec1*, form a clique. Also, *prame* is linked to *six1* and *ctag2* to *ranbp2*. Top bottlenecks were computed using Dijkstra's algorithm in Cytoscape and the top physical complexes were extracted with the MCODE algorithm. Lastly, we calculated the shortest paths between the top 20 hubs (**Table 2** and **Figure 3B**), including *ctag1b/a* in the *ctag1b/a* network neighborhood (**Figure 4B**). The rationale behind this approach was based on the fact that top nodes (hubs) that are also bottlenecks organize biological networks into broader functional groups and contribute to network stability (robustness). On the other hand, shortest paths (**Figure 3B**, broken lines) among top network bottlenecks indicate how close nodes (genes/proteins) are in a network and therefore can help map unknown genes to pathways or suggest biological roles as well as indicating that signal transduction is efficient. Notably, *prame, six1* (a known oncogene), *magec1* and *ranbp2* (a large GTP binding protein) are close neighbors of the *ctag1b/a* sub-network (**Figure 3B**). Interestingly, though shortest paths to *ctag1b/a* go through *dyrk1a*, suggesting a central role for this gene in this sub-network (**Figure 3B**), we were not able to procure antibodies to test its presence. GO analysis shows that the *ctag1b/a* neighborhood is enriched in genes involved exclusively in transcription (**Figure 3C**). Lastly, the *ssx2* gene, that is, found in the *ctag1b/a* subnetwork (**Figure 3A**), is involved in 80% of all synovial sarcomas as fusion protein products SSXT-SSX1 or SSXT-SSX2, derived from translocation t (X; 18) (p11.2; q11.2).

Validation of Expression of Ranbp2 and tle1 in Sarcoma Samples by Immunohistochemistry

We confirmed expression of some of the *ctag1b/a* network neighbors in advanced grade sarcomas (**Figure 3B** and **Table 2**) by using independent samples from the Department of Pathology at Aristotle University School of Medicine. Immunohistochemical staining for MAGEC1, a protein, that is, found in intracellular complexes with CTAG1B/A, showed extensive (80%) positivity in synovial sarcoma cells in one of seven cases evaluated (**Figure 4A**), whereas the remaining six cases were negative. On the other hand, staining for CTAG1B/A showed variable positivity of synovial sarcoma cells in six of seven cases. The CTAG1B/A positivity was extensive in three cases, involving 90–100% of the neoplastic cells (**Figure 4A**), and focal in the other three cases, involving 7–25% of the cells (data not shown). Interestingly, staining for MDM2 control revealed extensive positivity (data not shown).

RANBP2 and TLE1 exhibited extensive positivity in several different, independent ganglioneuromas (samples DSCN1432-1,436, **Figure 4B**), albeit with a different pattern. In the case of RANBP2, there are small yet distinct individual cells scattered

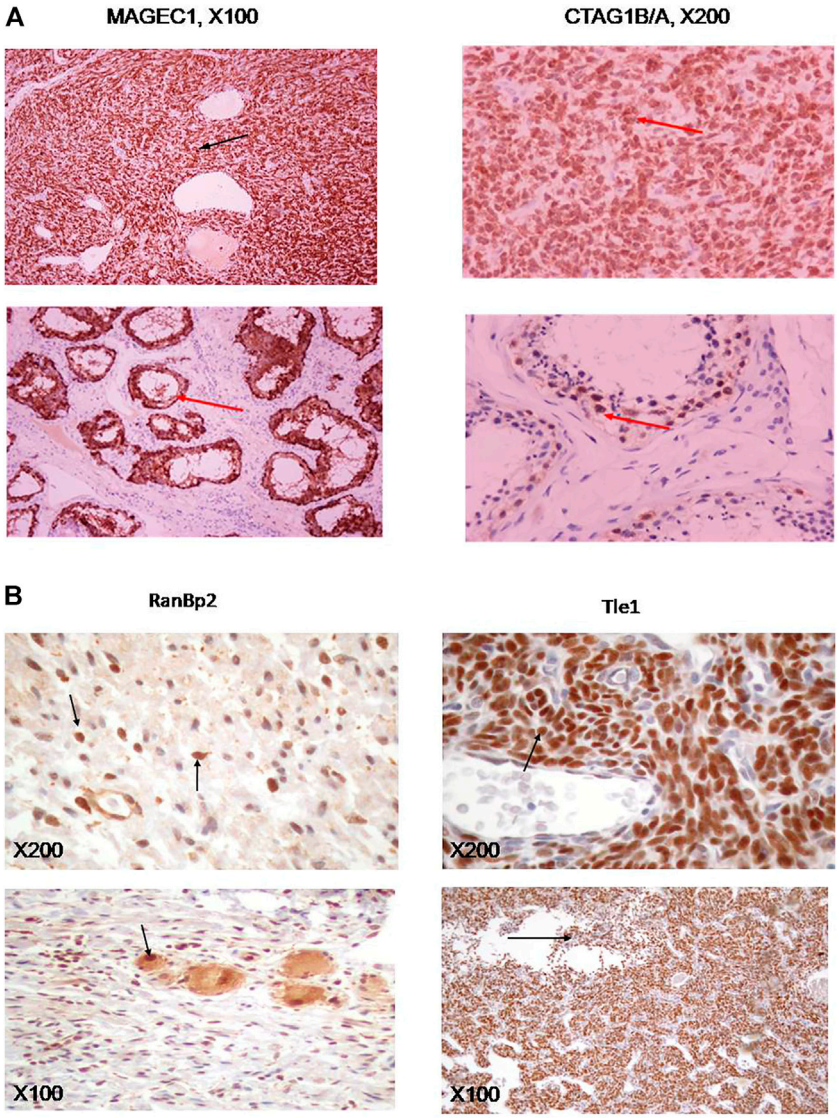


FIGURE 4 | (A) Immunohistochemical stains of synovial sarcoma specimens for MAGEC1 (left panel), or CTAG1B/A (right panel), as well as appropriate controls (testicular tissue, lower panels). **(B)** Immunohistichemical stains of ganglioneuroma samples for RANBP2 (left panels) and synovial sarcomas for TLE1 (right panels). Antibodies used are described in Materials and Methods. Magnification was at 100× (upper panels) or 40× (lower panels) on a NIKON Diaphot 200/300 inverted microscope.

throughout the field (**Figure 4B**, left upper and lower panels) whereas in the case of TLE1, the entire field of vision is covered with positive cells (results not shown), similar to synovial sarcomas (**Figure 4B**, upper right and lower panels). The lack of resources prevented us from testing expression of some of the other highly promising candidates such as PASD1 and DYRK1A.

DISCUSSION

Mesenchymal tumors that arise from soft tissues are rare and belong to a varieity of subtypes, hindering our understanding of the underlying molecular pathology. In this work we have identified *ranbp2* and *tle1* as co-expressed network neighbors of *ctag1b/a*, *magec1* and *prame* in synovial sarcomas (*tle1*) as well as in ganglioneuromas (*ranbp2*), and confirmed the expression of the encoded proteins in independent samples. The network proximity of these genes in the *ctag1b/a* network, and by extension to the *ctag1b/a*, *prame* and *magec1* network clique (**Figure 3B**), suggests that although *ctag1b/a* expression is not uniform within the malignant tissue, it is likely involved in synovial sarcoma biology. Alternatively, re-expression could be the result of epigenetic factors. Rearrangement of their sub-networks reveals that the *ctag1b/a, magec1 and prame* clique (**Figures 3B**, **5**) is maximally two degrees apart from *ranbp2, tle1, dyrk1a and six1* in the network, thus linking *cta1b/a* to the oncogenic pathways of *six1*

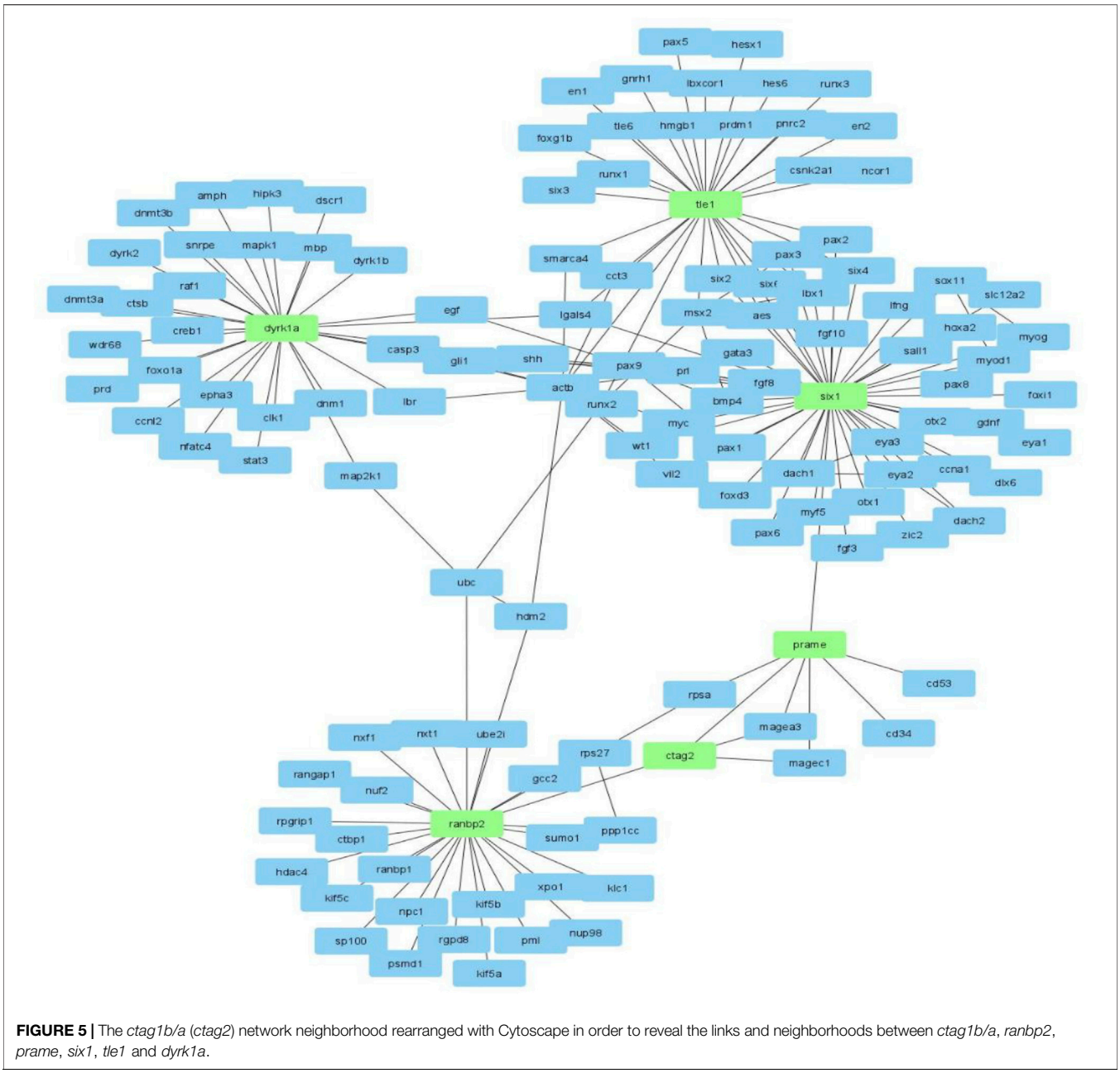


FIGURE 5 | The *ctag1b/a* (*ctag2*) network neighborhood rearranged with Cytoscape in order to reveal the links and neighborhoods between *ctag1b/a*, *ranbp2*, *prame*, *six1*, *tle1* and *dyrk1a*.

(cyclinA1 or transcriptional misregulation in cancer via the Wnt/Hedgehog/Notch-mediated oncogenesis), dyrk1a (cell cycle control at the G1-G1/S phases), *tle1* (NFκB-mediated control of gene transcription), and *ranbp2* (mitotic cell cycle control and Notch pathways).

It is noteworthy that *ctag1b/a* is re-activated in CD133 positive but not in CD133 negative stem cells, suggesting that it could have an active role in cancer stem cell biology (Rojas-Benítez et al., 2013). Consistent with CTAG1B/A being a homolog of the yeast and Drosophila transcription factor PCC1P, the synovial *ctag1b/a* neighborhood is overwhelmingly enriched for transcription processes. Of the genes comprising the *ctag1b/a* network neighborhood, *ranbp2* and *six1* are directly linked in the *ctag1b/a*, *prame* and *magec1* clique (**Figure 4B**). Moreover, *ctag1b/a* is linked to a subnetwork of chromatin protein-encoding genes via *pasd1* which also encodes a putative transcription factor (**Figure 3A**). PASD1 is a cancer-associated antigen that can stimulate autologous T-cell responses, and it is therefore considered to be a potential immunotherapeutic target for the treatment of various hematopoietic malignancies (Scanlan et al., 2000; Scanlan et al., 2002a). Another gene of interest is *zic3* (**Figure 3A**) which encodes a member of the ZIC family of C2H2-type zinc finger proteins. ZIC3 probably functions as a transcription factor in early stages of left-right body axis formation. Mutations in this gene cause X-linked visceral heterotaxy, which includes congenital heart disease and left-right axis defects in organs.

TABLE 2 | Important synovial sarcoma nodes identified with Cystoscape. Nodes found to be hubs (important nodes with numerous edges) and close expression neighbors of *ctag1b/a* with the gene set enrichment (GSEA) module analysis method are shown in red. Prdx2 is highlighted yellow as the top-ranked node.

Name	Score	Rank
Prdx2	46.490.800.000	1
Actb	28.369.400.000	2
Eef1a1	10.148.500.000	3
Banf1	9.006.900.000	4
Msx2	7.618.400.000	5
Pdx1	7.540.200.000	6
Apex1	7.087.200.000	7
Rps6	6.248.100.000	8
Six1	5.960.900.000	9
Hsd17b6	5.881.600.000	10
Tle1	5.716.000.000	11
Nme1	5.370.300.000	12
Prmt1	5.295.800.000	13
Rps15a	5.137.900.000	14
Ubc	5.057.900.000	15
Fbn2	4.915.400.000	16
Ranbp2	4.616.200.000	17
Hand1	4.609.600.000	18
Ep300	4.241.300.000	19
Dyrk1a	4.204.200.000	20
Gapdh	3.295.900.000	21
Pax9	3.096.700.000	22
Xpo1	3.056.900.000	23
Ybx1	2.976.900.000	24
Rfx1	2.928.000.000	25
Khdrbs1	2.921.900.000	26
Rps27a	2.640.800.000	27
Set	2.640.300.000	28
Hdm2	2.560.800.000	29
Anp32a	2.498.400.000	30

The *ctag1b/*a sub-network neighborhood includes, among others, the *six1* gene (**Figure 3A**), a known oncogene that enhances tumorigenesis by activation of *cyclin a1* expression and *ranbp2*, a gene of the RAS super-family, that is, associated with the nuclear membrane and which immunolocalizes to the nuclear pore complex. It encodes a small GTP binding protein belonging to the RAS superfamily, that is, essential for the translocation of molecules through the nuclear pore complex. *Ranbp2* is implicated in different cellular functions through interactions with other proteins and has been implicated in the Ran-GTPase cycle. It directly interacts with the E2 enzyme UBC9 enhancing SUMO1 transfer from UBC9 to the SUMO1 target SP100. These findings place sumoylation at the cytoplasmic filaments of the nuclear pore complex suggesting that, for some substrates, modification and nuclear import are linked events. Notably, synovial sarcomas express the fusion oncoprotein SYT-SSX1 which enhances symoylation of NCOA3 through interaction with the SUMO E3 ligase, PIASy(Sun et al., 2011). Whether RANBP2 and SYT-SSX1 act in common sumoylation pathways remains to be found. Lastly, a RANBP2-ALK fusion oncoprotein is expressed in epitheliod inflammatory myofibroblastic sarcomas (eIMS) and combined targeting of CD30 and RANBP2-ALK shows therapeutic promise (Segal et al., 2004; Silva et al., 2007).

The vertebrate *six* genes are homologs of the Drosophila "sine oculis" (so) gene, which are expressed primarily in the developing visual system of the fly. Members of the *six* gene family encode proteins that are characterized by a divergent DNA-binding homeodomain and an upstream SIX domain, which may be involved both in determining DNA-binding specificity and in mediating protein-protein interactions. Genes of the *six* family are involved in vertebrate and insect development and in maintaining the differentiated state of tissues. Our results as well as those of others implicate Six1 in sarcoma growth (Stockert et al., 1998; Simpson et al., 2005; Sun et al., 2011) however its functional links to CTAG1B/A or to RANBP2 remain to be established and underscore the complexity of sarcoma biology.

Tle1 is notable not only for being an expression neighbor of *ctag1b/a* in the GSEA-identified module but also a network neighbor that has been proposed as a diagnostic immunohistochemical marker for synovial sarcoma (Terry et al., 2007). Our network data suggest that ctag1a, magec1 and prame form a clique, where all three nodes are connected by edges to each other. Usually cliques are formed by functionally or physically and genetically linked genes. They could, therefore, affect each other's biology in synovial sarcoma. PRAME is a known transcriptional repressor protein (Epping et al., 2005), that is, expressed in human melanomas and recognized by cytolytic T lymphocytes. Like CTAG1B/A (and MAGEC1), PRAME is not expressed in normal tissues, except in testis. The protein is a repressor of retinoic acid receptor, and confers a growth advantage to cancer cells *via* this function. CTAG1B/A and MAGEC1 proteins physically interact with each other in intracellular protein complexes and with a few RNA polymerase II subunits, thus raising the issue of whether they function as specialized transcription complexes in developing germ cells (spermatophytes and spermatids). Interestingly, expression of CTAG1B/A and PRAME has recently been correlated with tumor grade and poor prognosis in myxoid sarcomas, supporting our hypothesis that CTAG1B/A is active in sarcoma biology (Iura et al., 2015).

Our findings that the CTAG1B/A network is enriched with transcription factors (**Figures 3C**, **5**) and its ability to compensate for the absence of the yeast transcription factor, PCC1P, support our hypothesis that CTAG1A complexes (likely with MAGEC1) are part of gametogenic germ cell transcription programs that could also be recruited to support tumor growth. Specifically, the PCC1P protein is part of an EKC/KEOPS multiprotein complex found in diverse organisms such as the Archaea and Drosophila (Hecker et al., 2008). Moreover, in yeast, this complex has been implicated in telomere maintenance, transcriptional regulation, bud site selection and chemical modification of tRNAs. In Drosophila, it has been linked to the TOR pathway via the protein kinase Bud32/PRPK, which regulates growth signals for nutrition status (Yu et al., 2018). Thus, the presence of *tle1* and *ranbp2* in the *ctag1b/a* network neighborhood tentatively link *ctag1b/a* to sarcoma biology.

CONCLUSION

Using sarcoma transcriptomics data deposited in GEO, we have re-constructed a sarcoma *ctag1b/*a network, identified sub-network neighbors, and determined that it is enriched with

transcription factor-encoding and chromatin-related genes. The combined data suggest that *ctag1b/*a is part of transcription complexes specific for germ cell development and differentiation. It is re-activated in advanced myxoid liposarcomas and synovial sarcomas but not in various differentiated liposarcomas or ganglioneuromas, thus indirectly implicating it in sarcoma biology. We have identified several genes in sarcoma samples that are network neighbors of *ctag1b/a* including *ranbp2*, *tle1, six1* and *prame* and using independent sarcoma samples, we immunohistochemically confirmed expression of the encoded proteins for *tle1* and *ranbp2*. Finally, the confirmation of expression of *ranbp2* and *tle1* as well as the presence of *six1* in its network neighborhood suggests that *ctag1b/a* may be part of transcriptional proteins that functionally interact with proteins involved in sumoylation (*via ranbp2*) and the cell cycle (*via six1, pasd1* and *zic3*) in sarcoma biology (Liu et al., 2018; Martínez-Arroyo et al., 2015; Ono et al., 2001; Türeci et al., 1998; Xu et al., 2015).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

NP conceived, performed the computations, wrote the manuscript, and interpreted the data. PH and PP performed the immunohistochemistry and interpreted the data. Before his passing, LO contributed to the initial design of the study. AP assisted in data interpretation and in editing the manuscript. NP had the overall supervision of the study. All authors, except LO, have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by Hilton-Ludwig Metastasis Grant CRA-2431.9-0-CB-3838 to NP.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.834445/full#supplementary-material

Supplementary Figure S1 | (A) Upper left panel. Top fifty genes identified with GSEA that show positive (CTAG1_POS) or negative (CTAG1_NEG) patterns of co-expression with ctag1b/a. Ctag1b/a (middle of Figure, left panel) was used as an index gene to compute the statistical distance of the other genes with GSEA as described in Computational section. Genes at the top and bottom are the most distant whereas those nearest ctag1b/a are those with the highest index of correlation (details in **Supplementary Material**). Upper right panels. Top enriched pathways of Ctag1_Positive or Negative modules computed as described in the Methods section. **(B)** Lower left panel: Heat map of gene expression status in phenotypic class lists of ctag1b/a co-expressed genes in microarray datasets retrieved from GEO **(Supplementary File S1)**. Lower middle panel: Gene setsranked by Kolomogorov-Smirnov (KS) statistics resulting from comparison with the MSigDb (GSEA site, MIT) and with the Papanikolaou gene sets collection. Lower right panel: Computed ES score for genes lists and rank in ordered sets.

REFERENCES

Ait-Tahar, K., Liggins, A. P., Collins, G. P., Campbell, A., Barnardo, M., Lawrie, C., et al. (2009). Cytolytic T-Cell Response to the PASD1 Cancer Testis Antigen in Patients with Diffuse Large B-Cell Lymphoma. *Br. J. Haematol.* 146 (4), 396–407. doi:10.1111/j.1365-2141.2009.07761.x

Chang, J. T., and Nevins, J. R. (2006). GATHER: A Systems Approach to Interpreting Genomic Signatures. *Bioinformatics* 22 (23), 2926–2933. doi:10.1093/bioinformatics/btl483

Cho, H. J., Caballero, O. L., Gnjatic, S., Andrade, V. C., Colleoni, G. W., Vettore, A. L., et al. (2006). Physical Interaction of Two Cancer-Testis Antigens, MAGE-C1 (CT7) and NY-ESO-1 (CT6). *Cancer Immun.* 6 (December), 12.

Cooke, P., and Nanjappa, M. (2015). How to Make a Human Germ Cell. *Asian J. Androl.* 17 (3), 441–442. doi:10.4103/1008-682X.151401

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* 278 (5338), 680–686. doi:10.1126/science.278.5338.680

DeRisi, J. L., and Iyer, V. R. (1999). Genomics and Array Technology. *Curr. Opin. Oncol.* 11 (1), 76–79. doi:10.1097/00001622-199901000-00015

Epping, M. T., Wang, L., Edel, M. J., Carlée, L., Hernandez, M., and Bernards, R. (2005). The Human Tumor Antigen PRAME Is a Dominant Repressor of Retinoic Acid Receptor Signaling. *Cell* 122 (6), 835–847. doi:10.1016/j.cell.2005.07.003

Fordham, A. M., Xie, J., Gifford, A. J., Wadham, C., Morgan, L. T., Mould, E. V. A., et al. (2020). CD30 and ALK Combination Therapy Has High Therapeutic Potency in RANBP2-ALK-Rearranged Epithelioid Inflammatory Myofibroblastic Sarcoma. *Br. J. Cancer* 123 (7), 1101–1113. doi:10.1038/s41416-020-0996-2

Gioutlakis, A., Klapa, M. I., and Moschonas, N. K. (2017). PICKLE 2.0: A Human Protein-Protein Interaction Meta-Database Employing Data Integration via Genetic Information Ontology. *PLOS ONE* 12 (10), e0186039. doi:10.1371/journal.pone.0186039

Gjerstorff, M. F., Kock, K., Nielsen, O., and Ditzel, H. J. (2007). MAGE-A1, GAGE and NY-ESO-1 Cancer/Testis Antigen Expression during Human Gonadal Development. *Hum. Reprod. (Oxford, England)* 22 (4), 953–960. doi:10.1093/humrep/del494

Gnjatic, S., Nishikawa, H., Jungbluth, A. A., Güre, A. O., Ritter, G., Jäger, E., et al. (2006). NY-ESO-1: Review of an Immunogenic Tumor Antigen. *Adv. Cancer Res.* 95, 1–30. doi:10.1016/S0065-230X(06)95001-5

Grigoriadis, A., Caballero, O. L., Hoek, K. S., da Silva, L., Chen, Y.-T., Shin, S. J., et al. (2009). CT-X Antigen Expression in Human Breast Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 106 (32), 13493–13498. doi:10.1073/pnas.0906840106

Hecker, A., Lopreiato, R., Graille, M., Collinet, B., Forterre, P., Libri, D., et al. (2008). Structure of the Archaeal Kae1/Bud32 Fusion Protein MJ1130: A Model for the Eukaryotic EKC/KEOPS Subcomplex. *Embo J.* 27 (17), 2340–2351. doi:10.1038/emboj.2008.157

Hei, S. M., Wei, H. J., Chen, H., and Wang, J. G. (2019). Pathological Significance of NY-ESO-1 Expression in the Diagnosis of Myxoid Liposarcoma. *Zhonghua Bing Li Xue Za Zhi* 48 (3), 225–230. doi:10.3760/cma.j.issn.0529-5807.2019.03.011

Henderson, S. R., Guiliano, D., Presneau, N., McLean, S., Frow, R., Vujovic, S., et al. (2005). A Molecular Map of Mesenchymal Tumors. *Genome Biol.* 6 (9), R76. doi:10.1186/gb-2005-6-9-r76

Hua, L., Fan, L., Aichun, W., Yongjin, Z., Qingqing, C., and Xiaojian, W. (2014). Inhibition of Six1 Promotes Apoptosis, Suppresses Proliferation, and Migration of Osteosarcoma Cells. *Tumor Biol.* 35 (3), 1925–1931. doi:10.1007/s13277-013-1258-1

Irie, N., Sybirna, A., and Surani, M. A. (2018). What Can Stem Cell Models Tell Us about Human Germ Cell Biology? *Curr. Top. Dev. Biol.* 129, 25–65. doi:10.1016/bs.ctdb.2018.02.010

Iura, K., Kohashi, K., Hotokebuchi, Y., Ishii, T., Maekawa, A., Yamada, Y., et al. (2015). Cancer-testis Antigens PRAME and NY-ESO-1 Correlate with Tumour Grade and Poor Prognosis in Myxoid Liposarcoma. *J. Path: Clin. Resclinical Res.* 1 (3), 144–159. doi:10.1002/cjp2.16

Jiang, Q., Tong, H.-X., Hou, Y.-Y., Zhang, Y., Li, J.-L., Zhou, Y.-H., et al. (2017). Identification of EML4-ALK as an Alternative Fusion Gene in Epithelioid Inflammatory Myofibroblastic Sarcoma. *Orphanet J. Rare Dis.* 12 (1), 97. doi:10.1186/s13023-017-0647-8

Jungbluth, A. A., Antonescu, C. R., Busam, K. J., Iversen, K., Kolb, D., Coplan, K., et al. (2001). Monophasic and Biphasic Synovial Sarcomas Abundantly Express Cancer/Testis Antigen NY-ESO-1 but Not MAGE-A1 or CT7. *Int. J. Cancer* 94 (2), 252–256. doi:10.1002/ijc.1451

Kakimoto, T., Matsumine, A., Kageyama, S., Asanuma, K., Matsubara, T., Nakamura, T., et al. (2019). Immunohistochemical Expression and Clinicopathological Assessment of the Cancer Testis Antigens NY-ESO-1 and MAGE-A4 in H-igh-grade S-oft-tissue S-arcoma. *Oncol. Lett.* 17 (4), 3937–3943. doi:10.3892/ol.2019.10044

Kalejs, M., and Erenpreisa, J. (2005). Cancer/Testis Antigens and Gametogenesis: A Review and 'Brain-Storming' Session. *Cancer Cel Int* 5 (1), 4. doi:10.1186/1475-2867-5-4

Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J.-C., Ilan, L., Hofmann, K., et al. (2006). Yeast Homolog of a Cancer-Testis Antigen Defines a New Transcription Complex. *Embo J.* 25 (15), 3576–3585. doi:10.1038/sj.emboj.7601235

Klapa, M. I., Tsafou, K., Theodoridis, E., Tsakalidis, A., and Moschonas, N. K. (2013). Reconstruction of the Experimentally Supported Human Protein Interactome: What Can We Learn? *BMC Syst. Biol.* 7 (1), 96. doi:10.1186/1752-0509-7-96

Kobayashi, T., and Surani, M. A. (2018). On the Origin of the Human Germline. *Dev.* 145 dev150433. doi:10.1242/dev.150433

Liggins, A. P., Lim, S. H., Soilleux, E. J., Pulford, K., and Banham, A. H. (2010). A Panel of Cancer-Testis Genes Exhibiting Broad-Spectrum Expression in Haematological Malignancies. *Cancer Immun.* 10 (August), 8.

Liu, H., Wei, W., Wang, X., Guan, X., Chen, Q., Pu, Z., et al. (2018). miR-23b-3p Promotes the Apoptosis and Inhibits the Proliferation and Invasion of Osteosarcoma Cells by Targeting SIX1. *Mol. Med. Rep.* 18 (6), 5683–5692. doi:10.3892/mmr.2018.9611

Martínez-Arroyo, A. M., Míguez-Forján, J. M., Remohí, J., Pellicer, A., and Medrano, J. V. (2015). Understanding Mammalian Germ Line Development with *In Vitro* Models. *Stem Cell Dev.* 24 (18), 2101–2113. doi:10.1089/scd.2015.0060

Ono, T., Kurashige, T., Harada, N., Noguchi, Y., Saika, T., Niikawa, N., et al. (2001). Identification of Proacrosin Binding Protein Sp32 Precursor as a Human Cancer/Testis Antigen. *Proc. Natl. Acad. Sci. U.S.A.* 98 (6), 3282–3287. doi:10.1073/pnas.041625098

Rojas-Benítez, D., Ibar, C., and Glavic, Á. (2013). TheDrosophilaEKC/KEOPS Complex. *Fly* 7 (3), 168–172. doi:10.4161/fly.25227

Scanlan, M. J., Altorki, N. K., Gure, A. O., Williamson, B., Jungbluth, A., Chen, Y.-T., et al. (2000). Expression of Cancer-Testis Antigens in Lung Cancer: Definition of Bromodomain Testis-specific Gene (BRDT) as a New CT Gene, CT9. *Cancer Lett.* 150 (2), 155–164. doi:10.1016/s0304-3835(99)00385-7

Scanlan, M. J., Welt, S., Gordon, C. M., Chen, Y-T., Gure, A. O., Stockert, E., et al. (2002a). Cancer-Related Serological Recognition of Human Colon Cancer: Identification of Potential Diagnostic and Immunotherapeutic Targets. *Cancer Res.* 62 (14), 4041–4047.

Scanlan, M. J., Gure., A. O., Jungbluth., A. A., Old., L. J., and Chen., Y.-T. (2002b). Cancer/Testis Antigens: An Expanding Family of Targets for Cancer Immunotherapy. *Immunological Rev.* 188 (October), 22–32. doi:10.1034/j.1600-065x.2002.18803.x

Segal, E., Friedman, N., Koller, D., and Regev, A. (2004). A Module Map Showing Conditional Activity of Expression Modules in Cancer. *Nat. Genet.* 36 (10), 1090–1098. doi:10.1038/ng1434

Silva, W. A., Gnjatic, S., Ritter, E., Chua, R., Cohen, T., Hsu, M., et al. (2007). PLAC1, a Trophoblast-specific Cell Surface Protein, Is Expressed in a Range of Human Tumors and Elicits Spontaneous Antibody Responses. *Cancer Immun.* 7 (November), 18.

Simpson, A. J. G., Caballero, O. L., Jungbluth, A., Chen, Y.-T., and Old, L. J. (2005). Cancer/Testis Antigens, Gametogenesis and Cancer. *Nat. Rev. Cancer* 5 (8), 615–625. doi:10.1038/nrc1669

Stockert, E., Jäger, E., Chen, Y.-T., Scanlan, M. J., Gout, I., Karbach, J., et al. (1998). A Survey of the Humoral Immune Response of Cancer Patients to a Panel of Human Tumor Antigens. *J. Exp. Med.* 187 (8), 1349–1354. doi:10.1084/jem.187.8.1349

Sun, Y., Perera, J., Rubin, B. P., and Huang, J. (2011). SYT-SSX1 (Synovial Sarcoma Translocated) Regulates PIASy Ligase Activity to Cause Overexpression of NCOA3 Protein. *J. Biol. Chem.* 286 (21), 18623–18632. doi:10.1074/jbc.M110.176693

Terry, J., Saito, T., Subramanian, S., Ruttan, C., Antonescu, C. R., Goldblum, J. R., et al. (2007). TLE1 as a Diagnostic Immunohistochemical Marker for Synovial Sarcoma Emerging from Gene Expression Profiling Studies. *Am. J. Surg. Pathol.* 31 (2), 240–246. doi:10.1097/01.pas.0000213330.71745.39

Türeci, Ö., Chen, Y.-T., Sahin, U., Güre, A. O., Zwick, C., Villena, C., et al. (1998). Expression ofSSX Genes in Human Tumors. *Int. J. Cancer* 77 (1), 19–23. doi:10.1002/(sici)1097-0215(19980703)77:1<19::aid-ijc4>3.0.co;2-2

Xu, Y., Li, X., Ren, S., Zhao, C., Hou, L., Ye, Z., et al. (2015). Predictive and Prognostic Effect of CD133 and Cancer-Testis Antigens in Stage Ib-IIIA Non-small Cell Lung Cancer. *Int. J. Clin. Exp. Pathol.* 8 (5), 5509–5518.

Yu, C., Zhang, B., Li, Y.-L., and Yu, X.-R. (2018). SIX1 Reduces the Expression of PTEN via Activating PI3K/AKT Signal to Promote Cell Proliferation and Tumorigenesis in Osteosarcoma. *Biomed. Pharmacother.* 105 (September), 10–17. doi:10.1016/j.biopha.2018.04.028

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Papanikolaou, Hytiroglou, Pantelidou, Papavassiliou and Old. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fgene.2022.844542

Identification of Candidate Therapeutic Genes for More Precise Treatment of Esophageal Squamous Cell Carcinoma and Adenocarcinoma

Aneta Polewko-Klim[1†], Sibo Zhu[2,3,4†], Weicheng Wu[2,3], Yijing Xie[2,3], Ning Cai[2,3], Kexun Zhang[2,3], Zhen Zhu[2,3], Tao Qing[3], Ziyu Yuan[3], Kelin Xu[2,3], Tiejun Zhang[2,3], Ming Lu[3,5], Weimin Ye[3,6], Xingdong Chen[3,4], Chen Suo[2,3,7]* and Witold R. Rudnicki[1,8]*

[1]Institute of Computer Science, University in Bialystok, Białystok, Poland, [2]Department of Epidemiology, School of Public Health, Fudan University, Shanghai, China, [3]Fudan-Taizhou Institute of Health Sciences, Taizhou, China, [4]State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China, [5]Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, China, [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, [7]Shanghai Institute of Infectious Disease and Biosecurity, Shanghai, China, [8]Computational Centre, University of Bialystok, Białystok, Poland

The standard therapy administered to patients with advanced esophageal cancer remains uniform, despite its two main histological subtypes, namely esophageal squamous cell carcinoma (SCC) and esophageal adenocarcinoma (AC), are being increasingly considered to be different. The identification of potential drug target genes between SCC and AC is crucial for more effective treatment of these diseases, given the high toxicity of chemotherapy and resistance to administered medications. Herein we attempted to identify and rank differentially expressed genes (DEGs) in SCC *vs*. AC using ensemble feature selection methods. RNA-seq data from The Cancer Genome Atlas and the Fudan-Taizhou Institute of Health Sciences (China). Six feature filters algorithms were used to identify DEGs. We built robust predictive models for histological subtypes with the random forest (RF) classification algorithm. Pathway analysis also be performed to investigate the functional role of genes. 294 informative DEGs (87 of them are newly discovered) have been identified. The areas under receiver operator curve (AUC) were higher than 99.5% for all feature selection (FS) methods. Nine genes (i.e., ERBB3, ATP7B, ABCC3, GALNT14, CLDN18, GUCY2C, FGFR4, KCNQ5, and CACNA1B) may play a key role in the development of more directed anticancer therapy for SCC and AC patients. The first four of them are drug targets for chemotherapy and immunotherapy of esophageal cancer and involved in pharmacokinetics and pharmacodynamics pathways. Research identified novel DEGs in SCC and AC, and detected four potential drug targeted genes (ERBB3, ATP7B, ABCC3, and GALNT14) and five drug-related genes.

Keywords: esophageal cancer (ESCA), drug target genes, Feature Selection (FS), random forest (RF), ensemble learning (EL)

OPEN ACCESS

Edited by:
Ruowen Zhang, Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Manoj Kumar Kashyap, Amity University Gurgaon, India
Mamoon Rashid, King Abdullah International Medical Research Center (KAIMRC), Saudi Arabia

***Correspondence:**
Xingdong Chen
xingdongchen@fudan.edu.cn
Chen Suo
suochen@fudan.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *28 December 2021*
Accepted: *20 April 2022*
Published: *19 May 2022*

Citation:
Polewko-Klim A, Zhu S, Wu W, Xie Y, Cai N, Zhang K, Zhu Z, Qing T, Yuan Z, Xu K, Zhang T, Lu M, Ye W, Chen X, Suo C and Rudnicki WR (2022) Identification of Candidate Therapeutic Genes for More Precise Treatment of Esophageal Squamous Cell Carcinoma and Adenocarcinoma. Front. Genet. 13:844542. doi: 10.3389/fgene.2022.844542

INTRODUCTION

Esophageal cancer (SCA) is a very aggressive condition. In 2018, there were an estimated 17,290 new cases of esophageal cancer and 15,850 deaths in the United States alone (Noone et al., 2018). Although its prognosis has gradually improved due to advances in treatment and surgical techniques, the overall survival remains poor, with only 10–22% patients showing survival of >5 years after diagnosis in Europe, the United States, and China (Dubecz et al., 2012). Such a low survival outcome is mainly attributable to late diagnosis and lack of effective treatment methods.

Esophageal cancer represents a heterogeneous group of cancers and consists of two main histological subtypes: squamous cell carcinoma and esophageal adenocarcinoma. Generally, SCC is associated with worse prognosis than AC (Enzinger and Mayer, 2003), but it is dependent on cancer progression (Shimada et al., 2013). SCC and AC are increasingly being considered as separate conditions with different etiologies, epidemiology, histopathology, and other biological behavior (Tustumi et al., 2016; Lagergren et al., 2017). Furthermore, recent studies have reported distinct differences in their genomic profiles (Wang et al., 2015; Salem et al., 2016; Network, 2017), and the number of different biomarkers between SCC and AC is in the order of thousands (Greenawalt et al., 2007; Lin et al., 2017). Analyses involving The Cancer Genome Atlas (TCGA) Research Network have shown that with respect to the overall genomic landscape, SCC and AC are more similar to non-esophageal cancers than to each other (Network, 2017; Salem et al., 2018).

Despite profound biological and clinical differences between SCC and AC, the standard therapy and drugs used in chemotherapy remain largely similar (Lordick et al., 2016). A combination of platinums, taxanes, anthracyclines, or pyrimidine analogs is usually prescribed to patients with esophageal cancer, regardless of the pathological subtypes (Abdo et al., 2017; Davidson et al., 2017). Bang et al. (2010) reported that therapies targeting *HER2* (trastuzumab) and vascular endothelial growth factor receptor 2 (ramucirumab) are highly effective for gastroesophageal junction cancer. Davidson et al. (Davidson et al., 2017)found that patients with AC showed a significantly higher response rate to first-line fluoropyrimidine-based chemotherapy than those with SCC. Earlier identification of drug-related genes with a high difference in their expression levels between SCC and AC can be helpful for understanding the differences in the clinical response of patients with esophageal cancer to different anticancer drugs, given the high toxicity of chemotherapy and resistance to administered medications. For instance, (Abdo et al., 2017) suggested that information pertaining to the overexpression of genes encoding drug molecular targets could help oncologists in decision making; the screening of nine genes (*HER2, EGFR, PD-L1, ERCC1, TUBB3, TS, RCF, TOPOI,* and *TOPO2A*) was recommended to ensure more effective immunotherapy and chemotherapy outcomes in patients with SCA.

Herein we aimed to identify novel biomarkers with the intention of improving diagnosis, as well as potential drug target genes and molecular candidate drugs to achieve effective treatment of SCC and AC. We used the heterogeneous ensemble feature selection method to identify the most informative biomarkers for the classification of the subtypes of esophageal cancer and the random forest machine learning algorithm (Breiman, 2001) to evaluate the quality of the set of the selected features. The ensemble filter method is based on six diverse filtering FS methods for reducing the risk of omitting biological relevant biomarkers. Such advanced machine learning methods have not been previously used for the classification of SCC and AC. Furthermore, we primarily focused on specific targets of drug action, such as membrane proteins, which are pivotal for drug development, because most therapeutics target membrane proteins are responsible for altering cellular signaling. We specifically studied membrane proteins affected by differentially expressed genes (DEGs) between SCC and AC and characterized relevant genes, which should enable individualized drug development. Additionally, we analyzed gene-gene interactions using the GeneMANIA software (Warde-Farley et al., 2010).

MATERIALS AND METHODS

Preprocessing and Integration of Datasets

Gene-level RNA-seq analyses of esophageal carcinoma were performed (two experiments): RNA-sequencing data from TCGA esophageal cancer project (TCGA-SCA, https://portal.gdc.cancer.gov/projects/TCGA-ESCA) and NODE-SCC data deposited in the National Omics Data Encyclopedia of China (accession no.: OEP000138, http://www.biosino.org/node/project/detail/OEP000138). TCGA-SCA mRNA data have been previously investigated (Zeng et al., 2017), but the main analysis was focused on the identification of molecular targets for prognostic analysis and diagnosis with reference to normal esophageal tissues (Zhan et al., 2016), not for the classification of SCA subtypes. TCGA-SCA dataset contains data of White and Asian patients; patients with SCC (n = 87, 87% male) had a median age of 57 years (range, 36–90) and those with AC (n = 71, 83% male) had a median age of 71 years (range, 27–86). The second dataset was generated by the Fudan University, containing data of 43 tumor tissue samples obtained from Asian patients with SCC (75% male) with a median age of 69 years (range, 50–82).

Data pre-processing and all analyses were conducted using the open-source statistical software R v3.4.3 (R Core Team, 2017a). Data preparation includes four main subprocesses: cleaning, transformation, merging, and reduction. TCGA-SCA dataset contains gene level RNA-seq data of 158 tumor tissue samples and 20,501 biomarkers, whereas the NODE-SCC dataset contains data of 43 tumor tissue samples obtained from patients with SCC and 21,309 biomarkers.

After cleaning up and Log2 transformation of the data, both datasets were subjected to unsupervised biomarker set selection. For this purpose, the following criteria were applied using the R package genefilter (Gentleman et al., 2020): (1) robust coefficient of variation of RNA-seq expression level (GE) > 0.05 and (2) at least 10% samples having GE > 0.45 (the number of biomarkers

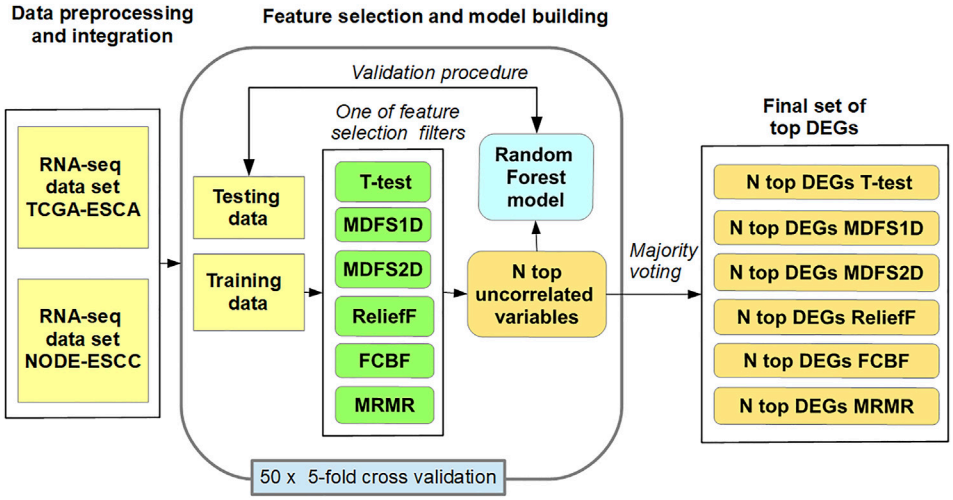


FIGURE 1 | Procedures involved in selecting the most informative biomarkers.

rapidly decreases below this GE threshold) for TCGA-SCA dataset and −4.35 for the NODE-SCC dataset. The software package BRB-ArrayTools (Simon, 2020) includes a detailed description of this reduction procedure. These standard preprocessing procedures are particularly important when using statistical methods because RNA-tags with low expression measurement range are not normally distributed. TCGA-SCA and NODE-SCC datasets were merged (COM-SCA dataset); i.e., the pairs of biomarkers belonging to the same gene were integrated. The COM-SCA dataset contains 201 samples (130 patients with SCC and 71 with AC) and 16,596 overlapping biomarkers. The ComBat function in "SVA" R package (Leek et al., 2018) was used for removing batch effects between the two experiments and races.

Statistical Analysis

To quantify feature distribution in COM-SCA dataset, the statistical analysis was performed. The 67% DEGs in SCC group and 61% DEGs in AC group have a fairly symmetrical distribution of data and the value of skewness ranges from -0.5 to 0.5. The Levene's test showed that 74% DEGs have variances equal in these groups, the Bartlett test showed 60%. The Kolmogorov-Smirnov test confirmed the normality distribution in 86% DEGs in both groups of patients. Considering the above, the normal distribution of biomarkers in both groups was assumed. The Welch *t*-test was used for the differential expression analysis of RNA-seq data, as one of six used feature selection methods (**Supplementary Table S1**).

Feature Selection and Prediction Model Building

To validate the FS process, machine learning models for discerning SCC from AC were built using selected markers as explanatory variables. To this end, we applied the random forest algorithm (Breiman, 2001) as implemented in the randomForest package (Liaw and Weiner 2002) in R (R Core Team, 2017b). Random forest is considered to be one of the best off-the-shelf classifier algorithms that can be applied to nearly all classes of problems. The conclusion of a very thorough study devoted to testing multiple algorithms on numerous publicly available datasets (Fernandez-Delgado and Cernadas., 2014) was that random forest is the best overall classification algorithm, that generally gives good results, very rarely gives bad results, and in many cases gives best ones. These conclusions were based on analysis of results of the application of 179 algorithms belonging to 17 broad families of algorithms on 121 diverse datasets.

Considering the problem of an unbalanced dataset (Luque et al., 2019), the area under the receiver operator characteristic curve and Matthews correlation coefficient (MCC) were used as measures of classification performance.

Figure 1 shows the methods used for the identification of the most informative biomarkers and building prediction RF models. Individual RF models from RNA-seq data were constructed in 50 loops with the following procedure:

(1) The dataset was randomly split into five equal partitions;
(2) Insignificant genes between SCC and AC were ranked/filtered out using Ttest, MDFS1D, MDFS2D, FCBF, MRMR, and ReliefF on four partitions (training set);
(3) Highly correlated features (Spearman's rank correlation coefficient >0.7) were rejected from the ranked list;
(4) Random forest classifier was built on the training set using the top-N features;
(5) Model quality was evaluated on the remaining partition (test set);
(6) Steps 2–5 were repeated for all k-partitions and each FS method.

The final number of top features used for model building was experimentally established. In addition, the quality of predictive RF models and stability of feature selection as a function of the

number of top features were measured for all FS methods. The stability of feature selection was gauged by the similarity of different sets of relevant variables in cross-validations with the Lustgarten stability measure (ASM) (Lustgarten et al., 2009). All FS methods used the same cross-validation splits.

Identification of Key Genes

The complete list of key genes was derived using the following procedure:

(1) Top-N DEGs were selected from each of the 250 ranked lists for each FS method independently;
(2) A set of N genes with the highest frequency of occurrence among the 250 lists was identified for each FS method independently;
(3) From the six gene sets corresponding to the FS methods, key genes were selected;
(4) Log2 fold change (FC) of normalized RNA-seq gene expression between SCC and AC was calculated using the formula Log2FC = Log2 (GE_{SCC}/GE_{AC}), wherein GE_{SCC} and GE_{AC} represent the mean value of normalized RNA-seq gene expression level for SCC and AC, respectively, for each gene. The key gene list was then sorted according to absolute Log2FC values;
(5) Membrane protein-encoding genes and their association with well-known drugs were subsequently identified.

The Human Protein Atlas database (Uhlen et al., 2015) was used for selecting membrane protein-encoding genes. Data pertaining to drugs and drug–gene interactions were collected from several databases, namely, DGIdb (Cotto et al., 2018), DrugBank (Wishart et al., 2006), and Therapeutic Target DB (Wang et al., 2020), and additional information was obtained from ApexBio, FDA Approved Drugs, ClinicalTrials.gov, PharmGKB, and GeneCards.

RESULTS

Informative Biomarkers

We investigated the molecular markers that could distinguish between the two main histological subtypes of esophageal cancer. To identify DEGs from the full combined RNA-seq datasets (COMSCA), we used six feature filters, namely Welch *t*-test (Ttest) (Welch, 1947), one- and two-dimensional FS filters based on information theory (MDFS1D and MDFS2D, respectively) (Piliszek et al., 2019), fast correlation-based filter (FCBF) (Yu and Liu, 2003), the ReliefF algorithm (Kononenko, 1994), and minimum redundancy and maximum relevance (MRMR) (Ding and Peng, 2005).

Using the methods Ttest, MDFS1D, and MDFS2D, which could identify relevant predictor features but did not remove redundant ones, 7142 unique relevant genes were identified (refer to the Venn diagram in **Supplementary Figure S1**) in the entire data set.

The other three FS methods, namely ReliefF, MRMR, and FCBF, either returned just a ranking of features (ReliefF, MRMR) or a set of top non-redundant informative ones (MRMR and FCBF). In particular, FCBF identified only 59 relevant variables, all of which were found by all other algorithms as well.

TABLE 1 | Comparison of feature selection methods.

Metric	Ttest	MDFS1D	MDFS2D	FCBF	ReliefF	MRMR
AUC	0.996	0.998	0.997	0.994	0.996	0.999
MCC	0.994	0.997	0.996	0.991	0.993	0.998
ASM	0.52	0.43	0.40	0.05	0.34	0.53

Note: The first two rows display AUC and MCC obtained for RF classifier on 100 most relevant genes selected with each feature selection method. The last row displays the stability of these, which was measured using ASM. Fifty repeats of 5-fold cross-validation were performed. Standard deviation of mean AUC and MCC was <0.001. See notation in the main text.

MDFS1D identified the highest number of relevant features (5437), and this number was used as the limit of relevant variables returned by MRMR and ReliefF. The final number of unique DEGs identified by at least one method was 8246. Although this number is bound to include false positive data, it shows how distinct SCC and AC are at the molecular level.

A much smaller number of features is sufficient to build a machine learning model that can distinguish between SCC and AC with high precision levels (**Supplementary Figures S2, S3**). In the current study, a high average predictive power of random forest model (AUC = 0.994) was already achieved for 20 features for all filters. However, small sets of most relevant features showed instability in 5-fold cross-validation repeated 50 times (**Supplementary Figure S4**). For all algorithms, except FCBF, the maximal stability value of the Lustgarten measure for sets of top N features as N approaching 100 (**Table 1**). In contrast, the FCBF method attempts to minimize redundancy in the set of features. This optimization increases instability, because it amplifies small random differences in relevance observed in the different repeats of cross-validation. ASM values, indicating stability, were <0.6 in the entire studied range of top-N uncorrelated features. These instabilities could be attributed to a high number of highly relevant variables with very similar levels of relevance. Random fluctuations due to differences in the composition of samples in cross-validation lead to large changes in the relative rankings of features in different samples. To minimize the influence of these fluctuations, top 100 features from each algorithm were used for further analysis, ensuring that most relevant genes were a part of them. The good predictive power for all methods (**Table 1**) and the small overlap between the six sets of top 100 genes from the six algorithms suggested that each algorithm identified different aspects of relevance.

The final list of relevant genes ultimately included six lists of 100 genes identified by six independent FS methods. Overall, 294 genes represented the key set of biomarkers that could be used to distinguish between AC and SCC. The complete list of genes is shown in **Supplementary Table S1**. More than 46% genes in this set showed a high difference in expression levels between AC and SCC, with abs (Log2FC) ≥ 3.0 (FC = fold change between SCC and AC). Under-expressed genes were the most prevalent (59%).

Potential Therapeutic Targets

Herein we focused on membrane protein-encoding genes that are one of the most important macromolecules for drug development. In total, 84 genes from the list of the most

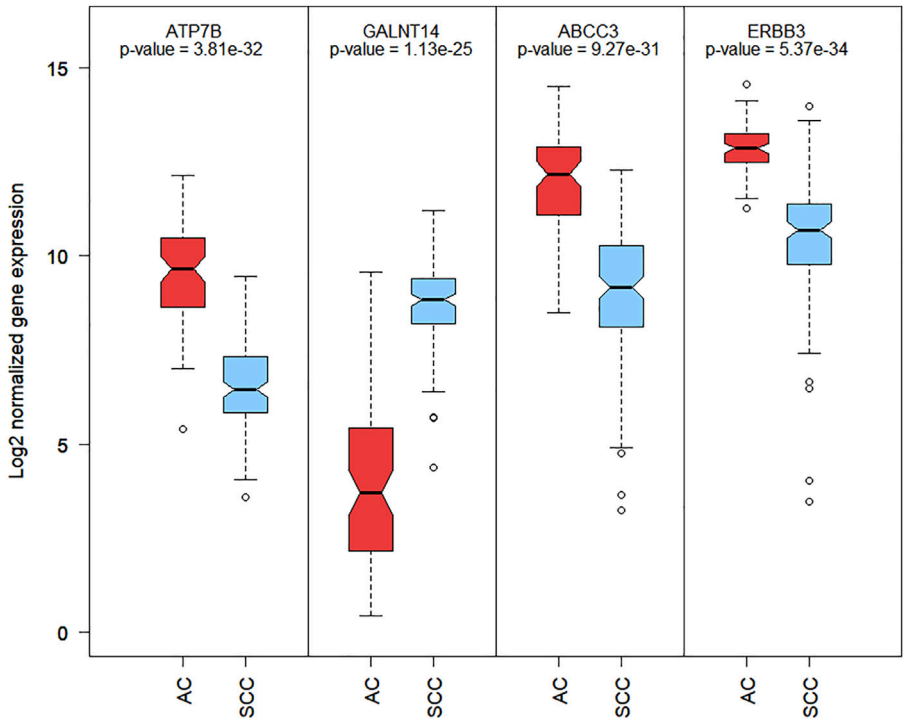


FIGURE 2 | Boxplot of Log2 normalized RNA-Seq gene expression of 4 membrane encoding genes related with SCA anti-cancer drugs. Boxplot contains the *p*-value of mean differential expression between AC and SCC patients groups with a two-sample *t*-test.

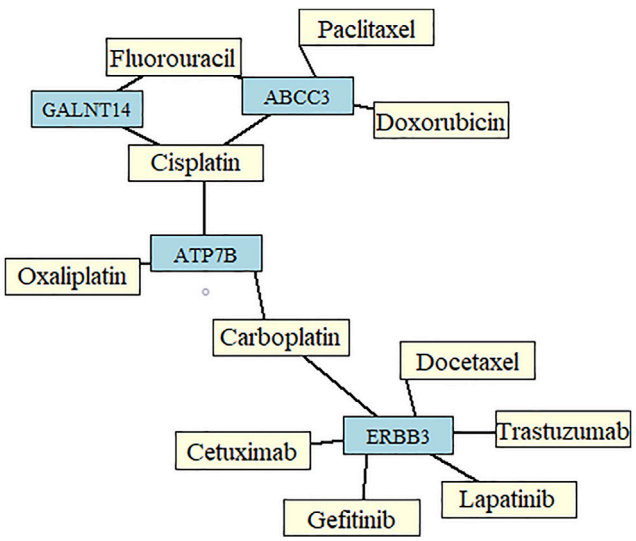


FIGURE 3 | Network of drugs important for chemotherapy in patients with esophageal cancer, and genes identified from the set of the most informative biomarkers predictive of the two main histological subtypes of esophageal cancer.

important biomarkers were labeled as "membrane proteins" or "predicted membrane proteins" in the Human Protein Atlas database. Of these 84 genes, 44 were related to drugs (**Supplementary Table S2**).

The most common drugs used for anticancer therapy in the case of patients with SCA include carboplatin, paclitaxel, platinol, epirubicin, docetaxel, fluorouracil, oxaliplatin, irinotecan, cetuximab, lapatinib, trastuzumab, doxorubicin, cisplatin, leucovorin, capecitabine, gefitinib, ramucirumab, mitomycin, bleomycin, and amethopterin (Abdo et al., 2017; Huang and Yu, 2018). We found that four of the 84 membrane protein-encoding genes were drug targets for chemotherapy and immunotherapy of esophageal cancer and involved in pharmacokinetics and pharmacodynamics pathways. Three genes were under-expressed Erb-B2 receptor tyrosine kinase 3 (*ERBB3*), ATPase copper-transporting beta (*ATP7B*), and ATP-binding cassette subfamily c member 3 (*ABCC3*) and one was overexpressed polypeptide N-acetylgalactosaminyltransferase 14 (*GALNT14*) in SCC *vs*. AC. Among these four genes, *GALNT14* showed the highest difference in expression levels between SCC and AC (Log2FC = 4.62, **Figure 2**). A network of anticancer drugs related to the four genes is shown in **Figure 3**.

To identify new potential therapeutic targets that may affect the choice of SCA therapy, we identified genes with the highest difference in expression levels between SCC and AC, and arranged them according to the absolute value of Log2FC for the 84 genes predicted to encode membrane proteins. Biological functions of each of these genes are listed in **Supplementary Table S2**. Top 10 over/under-expressed genes in SCC *vs*. AC and examples of drugs associated with them are shown in **Table 2**. Although these drugs are not necessarily chemotherapy drugs, they provide new insights into targeted therapy for SCC and AC.

TABLE 2 | Top 10 membrane protein-encoding genes that were under- or overexpressed in SCC *vs*. AC

Top down-expressed membrane protein encoding genes in SCC *vs*. AC			
No	**Gene**	**Log2FC**	**Drugs**
1	*CLDN18*	7.57	CLAUDIXIMAB
2	*TM4SF5*	7.27	
3	*SLC6A20*	6.03	
4	*TM4SF4*	5.87	
5	*SI*	5.86	ACARBOSE, SCOPOLAMINE, DEXAMETHASONE, MIFEPRISTONE, STREPTOZOTOCIN, FURAN, SODIUM BETA-NICOTINAMIDE ADENINE DINUCLEOTIDE PHOSPHATE, HEXAMETHYLENEBISACETAMIDE
6	*GPR128*	5.77	
7	*GUCY2C*	5.49	LINACLOTIDE, PLECANATIDE, PANITUMUMAB, PIRIBEDIL (CHEMBL1371770), PHOSPHORIC ACID, LINACLOTIDE ACETATE, GUANOSINE MONOPHOSPHATE, CYCLIC GMP
8	*CLDN2*	5.29	CALCIUM
9	*GPR35*	5.16	PROSCILLARIDIN, BUMETANIDE, FUROSEMIDE, TRANSTORINE, PAMOIC ACID, ZAPRINAST, PYRANTEL, KYNURENIC ACID
10	*MIA2*	5.06	
Top up-expressed membrane protein encoding genes in SCC *vs*. AC			
No	**Gene**	**Log2FC**	**Drugs**
1	*DSC1*	4.64	CALCIUM
2	*GALNT14*	4.62	FLUOROURACIL, CARBOQUONE, MITOXANTRONE, CISPLATIN, CALCIUM, MANGANESE
3	*DMRT2*	4.61	
4	*SLC35F3*	4.26	
5	*KCNQ5*	4.24	CELECOXIB, IRINOTECAN, EZOGABINE, TEDISAMIL, FLINDOKALNER, CHEMBL317935, LINOPIRDINE, CHEMBL342375, DALFAMPRIDINE, GUANIDINE HYDROCHLORIDE, NERISPIRDINE, POTASSIUM, TETRAETHYLAMMONIUM
6	*RYR1*	3.91	DANTROLENE SODIUM, SURAMIN, MAGNESIUM, cA2, ADENOSINE TRIPHOSPHATE, CAFFEINE, PROCAINE, RYANODINE, DANTROLENE, SURAMIN, RUTHENIUM RED, CAFFEINE, DANTROLENE, TETRACAINE
7	*SLC6A2 (SLC6A5)*	3.75	AMPHETAMINE, GUANADREL, GUANETHIDINE, REBOXETINE, MIRTAZAPINE, LOXAPINE, DOXEPIN, AMOXAPINE, MAZINDOL, ERGOTAMINE, COCAINE, PHENMETRAZINE, SIBUTRAMINE, NOMIFENSINE, CHLORPHENIRAMINE POLISTIREX, GINKGO, CRX-119, AMINEPTINE, DEBRISOQUIN, BICIFADINE, MMDA, MIANSERIN, TAPENTADOL, TRAMADOL, BUPROPION, LEVOMILNACIPRAN, ATOMOXETINE, CITALOPRAM, CLOMIPRAMINE, DESIPRAMINE, DESVENLAFAXINE, DEXTROAMPHETAMINE, DOTHIEPIN, DULOXETINE, IMIPRAMINE, LOFEPRAMINE, METHYLPHENIDATE, MAPROTILINE, MILNACIPRAN, DIETHYLPROPION, NEFAZODONE, NISOXETINE, AMITRIPTYLINE, NORTRIPTYLINE, VENLAFAXINE, PHENELZINE, PROTRIPTYLINE, QUETIAPINE, PSEUDOEPHEDRINE, PHENTERMINEGUANADREL SULFATE, ZOTEPINE, AMITIFADINE, EDIVOXETINE, BETHANIDINE, TRAZODONE, DEXMETHYLPHENIDATE, PETHIDINE, PAROXETINE, TEDATIOXETINE, TESOFENSINE, PHENTERMINE, PROTRIPTYLINE MIRTAZAPINE, DEXTROMETHORPHAN, METHAMPHETAMINE, NORTRIPTYLINE, AMOXAPINE, TRIMIPRAMINE, DOPAMINE, SIBUTRAMINE, CHLORPHENAMINE, ORPHENADRINE, ESCITALOPRAM, KETAMINE, EPHEDRA, EPHEDRINE, GINKGO BILOBA, XEN2174, 3,4-METHYLENEDIOXYMETHAMPHETAMINE, PHENDIMETRAZINE, LEVONORDEFRIN, REBOXETINE GLYCINE, BITOPERTIN, CHEMBL88588, CILNIDIPINE
8	*IGSF11*	3.71	
9	*CACNA1B*	3.71	VERAPAMIL, GABAPENTIN, LEVETIRACETAM, Z160, AGMATINE, CELECOXIB, CNV2197944, VERAPAMIL, GABAPENTIN ENACARBIL, AMLODIPINE, AGMATINE, SAFINAMIDE, RALFINAMIDE, ZICONOTIDE, BEPRIDIL HYDROCHLORIDE, CLEVIDIPINE, ELPETRIGINE, ATAGABALIN, ZICONOTIDE ACETATE, IMAGABALIN, PREGABALIN, AMLODIPINE, GABAPENTIN, LEVETIRACETAM, CILNIDIPINE, NITRENDIPINE
10	*DLK2*	3.49	

Note: Where applicable, each gene is accompanied by a list of drugs that were associated with it in at least one of the following databases: DrugBank, PharmGKB, ClinicalTrials.gov, DGIdb, and FDA Approved Drugs. Standard deviation of the expression values at the Log2FC level was <0.002.

Gene-Gene Interaction Network

To conduct the gene-gene (G-G) functional interaction analysis of key biomarkers, we used the GeneMANIA online software. All membrane protein-encoding genes (**Supplementary Table S2**) were used as input data for G-G network. To find the hub genes, we ranked genes by the number of edges they shared with other genes and the difference in expression levels between SCC and AC.

The functional associations between 57 of 84 genes were observed (**Figure 4**). Three genes, namely the mucin 1 (*MUC1*), the gap junction protein alpha 1 (*GJA1*), and *KCNQ5* with a high number of gene-gene interactions (more than 11 edges) and a high difference in expression levels between SCC and AC (abs (Log2FC) > 2.8) were considered as hub genes. Within the 57-gene network, we identified two sub-networks. In the first sub-network, the *MUC1* gene is a primary hub gene linked with 21 significant genes, such as *GALNT14, ABCC3,* and *ERBB3*. In the second sub-network, there are *KCNQ1* and *GJA1* hub genes linked with 12 significant genes, such as *GUCY2C*, and ERBB3. Seven genes are shared by both sub-networks (*GALNT14, TMEM144, KALRN, IGSF11, GP2*, and *ERBB3*).

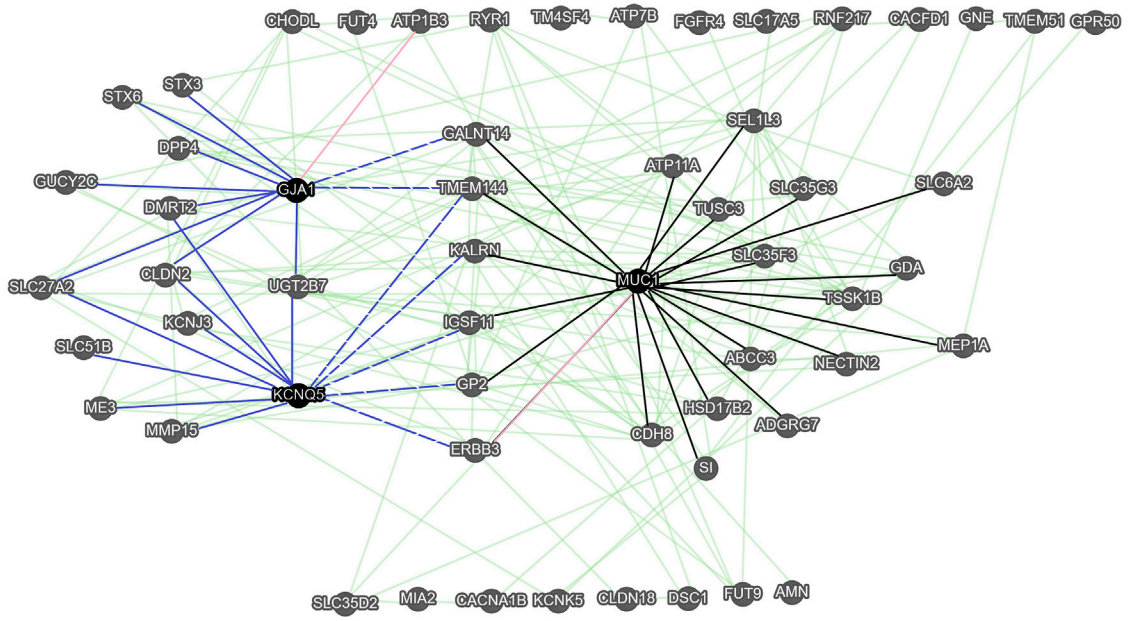


FIGURE 4 | The gene-gene interaction network of membrane protein-encoding genes obtained with GeneMANIA. Green edges correspond to the functional associations between genes (nodes), while pink edges represent the predicted gene-gene interaction. The black edges correspond to genes functionally associated with the MUC1 hub gene (sub-network 1). The solid blue edges correspond to genes functionally associated with the KCNQ5 and GJA1 hub genes (sub-network 2), while the blue dashed edges correspond to genes functionally associated with the KCNQ5, GJA1, and MUC1.

DISCUSSION

In this study, we aimed to identify the most informative molecular markers to distinguish between SCC and AC and to characterize pertinent genes in terms of their potential utility in individualized cancer treatment. We used a robust two-step protocol for identifying the most informative RNA-seq biomarkers important for cancer diagnosis and potentially druggable genes crucial for SCA treatment.

Informative Biomarkers

SCC and AC significantly differ at the molecular level (Lin et al., 2017; Network, 2017). Previous studies have confirmed these results, and owing to the use of a more sensitive approach based on ensemble FS, even stronger differences have become known. We herein identified 8246 DEGs, of which, 5434 (65.9%) have not been mentioned in previous studies (Greenawalt et al., 2007; Lin et al., 2017). Further, 81.7% of the 3443 genes identified by Lin *et al.* (Lin et al., 2017) and 64.7% of the 546 genes identified by Greenawalt *et al.* (Greenawalt et al., 2007) were identified in this study. These differences could be attributed to (1) the absence of some genes in our dataset and (2) using a more stringent method for multiple testing correction (for Ttest, MDFS-1D, and MDFS-2D).

We observed that for different FS methods, the overlap between the sets of selected biomarkers was low (**Supplementary Figures S1, S5**). This is a manifestation of the well-known problem that different selection methods tend to produce different biological signatures (Abusamra, 2013). Such differences can also be due to different approaches for FS implemented in different algorithms (Bommert et al., 2020). Furthermore, application of methods that reduce redundancy in the feature set result in decreased stability of the set of biomarkers (Polewko-Klim and Rudnicki, 2020). Nevertheless, as per ontological analyses, biological functions captured by different gene subsets are rather similar (Dessi et al., 2013). In this study, we constructed highly effective predictive models (AUC >0.994) using only top 20 features returned by any of the applied FS methods. Overlooking genes that are important for biological and functional interpretation of differences between datasets is possible when only one FS method is used for identifying relevant features. Thus, the application of diverse FS methods is highly recommended.

Among the 294 genes that facilitated differentiation between AC and SCC, 87 have not been previously reported as relevant (Greenawalt et al., 2007; Lin et al., 2017). These newly discovered relevant genes can potentially be used as drug targets.

Further, although subtypes can be easily determined by pathologists, these genes can facilitate automatic pathological identification in clinical settings with extremely high sensitivity and specificity. The set of genes presented in **Supplementary Table S1** can be used for SCC and AC diagnosis.

Specific Therapeutic Agents for SCC and AC

We focused on membrane protein-encoding genes that were drug targets for chemotherapy and immunotherapy and involved in pharmacokinetics and pharmacodynamics pathways. *ERBB3,*

ATP7B, ABCC3, and *GALNT14* were particularly interesting, as genes encoding them are already related to 11 common anti-SCA drugs (**Figure 3**). Both, overexpression (*GALNT14*) or under-expression (*ERBB3, ATP7B, ABCC3*) of these genes in SCC relative to AC may contribute to a different response to these common anticancer drugs. Interestingly, according to GeneCards, *GALNT14* is overexpressed by > 5-fold in the esophagus–mucosa in normal tissues.

Considering the upregulated expression of *GALNT14* in SCC *vs*. AC, *GALNT14* appears to be a promising potential therapeutic target for SCC. *GALNT14* is an antitumor agent and therapeutic response predictor for concurrent chemoradiotherapy wherein the platinum-based drugs fluorouracil and cisplatin are used for advanced SCC (Honing et al., 2014; Tsou et al., 2017). Gebski et al. (2007) reported that neoadjuvant chemotherapy with cisplatin and 5-fluorouracil led to relatively better survival of patients with AC. *GALNT14* genotype is also a potential predictor of the response to the first course of 5-fluorouracil, mitoxantrone, and cisplatin chemotherapy in patients with advanced hepatocellular carcinoma (Liang et al., 2011). The overexpression of *GALNT14* is a strong biomarker correlated to the sensitivity of Apo2L/TRAIL-based anticancer therapy. *GALNT14* alters cell migration and cellular morphology, and its overexpression causes malignancies, such as those of the breast, ovarium, lungs, and skin (Erdal et al., 2017), so it is a good predictor of therapeutic outcomes, particularly of chemotherapy, in multiple cancers (Lin and Yeh, 2020).

ERBB3, ATP7B, and *ABCC3* are also reportedly promising drug targets. *ERBB3 (HER3)* is a member of the epidermal growth factor receptor family of receptor tyrosine kinases. A comprehensive analysis of *EGFR, HER2,* and *HER3* coexpression and dimerization that were observed in the two histopathological subtypes of SCA has been previously performed (Fichter et al., 2014). Fichter *et al.* (Fichter et al., 2014) suggested that preclinical investigations of antibody-dependent cellular cytotoxicity elicited by trastuzumab and pertuzumab can be very important in AC, namely, these drugs indicate an effect in AC cancer cells with high *HER2* expression and *HER2* homodimers. *ATP7B* is a key mediator of cellular cisplatin, carboplatin, and oxaliplatin accumulation, these platinum-based drugs are widely used in modern cancer therapeutics (Li et al., 2016). Li *et al.* reported that *ATP7B* overexpression plays a key role in platinum resistance in SCC (Li et al., 2016). *ABCC3* is a transporter and inducer of cisplatin and an inhibitor of doxorubicin. It is involved in cellular resistance to chemotherapy with fluorouracil in patients with SCC (Zhou et al., 2008) and is also a putative biomarker of resistance to antimitotic agents, such as paclitaxel, used in breast cancer treatment (O'Brien et al., 2008).

Novel Drug-Related Genes

We also detected several new drug-related genes with a high Log2FC value, which were associated with cancer promotion, transformation, and progression (**Supplementary Table S2**) and thus relevant for targeted treatment of SCA. Three genes were under-expressed in SCC *vs*. AC, namely claudin 18 (*CLDN18*), guanylyl cyclase C (*GUCY2C*), and fibroblast growth factor receptor 4 (*FGFR4*), and two genes were overexpressed, namely potassium voltage-gated channel subfamily Q member 5 (*KCNQ5*) and calcium voltage-gated channel subunit alpha1 B (*CACNA1B*). These genes showed a high difference in their expression levels between SCC and AC and are already associated with certain drugs. For example, the tight junction molecule claudin-18 isoform 2 (*CLDN18.2*) is a target for claudiximab, which is a first-in-class chimeric monoclonal antibody used for the treatment of gastric cancer (Singh et al., 2017). Aka *et al.* (Aka et al., 2017) reported that *GUCY2C* is a potentially ideal target antigen for colorectal cancer immunotherapy and that supplementation with linaclotide (*GUCY2C* ligand) is a novel and promising strategy for tumor prevention. Zhongwei et al. (Xin et al., 2018) suggested that blocking *FGFR4* significantly suppressed the malignant behaviors of SCC, indicating that *FGFR4* is a potential target for SCC treatment. *KCNQ5* interacts with celecoxib and is a promising drug for prevention/treatment of several cancers, such as colon, breast, prostate, and head and neck cancers (Toloczko-Iwaniuk et al., 2019). *CACNA1B* is useful for evaluating the efficacy of chemoradiotherapy against SCC (Sasaki et al., 2017).

Considering the significant differences in gene expression levels of *ERBB3, ATP7B, ABCC3, GALNT14, CLDN18, GUCY2C, FGFR4, KCNQ5,* and *CACNA1B* between patients with SCC and AC, we recommend conducting further preclinical research on them. Future studies are warranted to investigate how these genes can be used to develop more effective chemotherapy and immunotherapy treatment methods for patients with SCA, as well as options for novel drug use associated with those genes with large fold change in SCC and AC.

Gene-Gene Interaction

The G-G interaction analysis via GeneMANIA indicated three pivotal hub genes, namely *MUC1, GJA1*, and *KCNQ5*. The most linked of them, the *MUC1*, is an oncogene that exhibits extensive glycosylation *in vivo*. The aberrant glycosylation and overexpression of *MUC1* gene in cancer cells may lead to cancer invasion, metastasis, angiogenesis, and apoptosis by virtue of its participation in intracellular signaling processes and the regulation of related biomolecules (Chen et al., 2021). Moreover, mucin 1 protein coded by *MUC1* is an important barrier to the penetration of drugs and takes part in the inhibition of apoptosis in tumor cells. It has been widely recognized as one of the most promising molecular targets in cancer therapy (Lee et al., 2021). For example, the overexpression of this membrane-bound glycoprotein limits the effectiveness of 5-fluorouracil treatment in patients with pancreatic cancer (Kalra and Campbell 2009), and decreases sensitivity of cisplatin in SCC (Zhao et al., 2021). Mucin 1 plays a key role in trastuzumab resistance in breast cancer (Hosseinzadeh et al., 2022). The high-expression of *MUC1* is associated with a poor prognosis for esophageal cancer patients (Song et al., 2003), contributes to SCC metastasis

(Ye et al., 2011), and plays a pivotal role in the progression to AC (Adil Butt et al., 2017).

GJA1 and *KCNQ5* are hub genes of one gene module and both are significantly up-regulated in SCC *vs.* AC. *GJA1* (encoding *Cx43*) is a member of the connexin family that possesses both tumor-suppressive, and oncogenic functions (Aasen et al., 2019). The misregulation of connexins affects a process of cell differentiation, inflammation, and cell death (Katturajan and Evan Prince, 2021). *GJA1* is a highly attractive target for delivering drugs directly into the cytoplasm of cancer cells, due to the permeability of gap junction channels to small molecules and macromolecules (Bonacquisti and Nguyen, 2019). The silencing of *GJA1* gene may cause a reduction of paclitaxel efficiency in gastric cancer (Zhao et al., 2019), and cisplatin-resistance in lung cancer (Luo et al., 2021). The high *GJA1* expression in SCC cancer cells is associated with poor survival of patients (Tanaka et al., 2016). The exact role of *KCNQ5* in SCA cancer tumor genesis and progression is not known. But recent studies have shown that this oncogene is a potential prognostic biomarker for gastrointestinal cancer (Shorthouse et al., 2020) and a promising biomarker for early colorectal cancer detection (Cao et al., 2021). Considering the over-expression of *GJA1* and *KCNQ5* in SCC *vs.* AC, they can be promising molecular targets for SCC.

DATA AVAILABILITY STATEMENT

Data from TCGA is publicly available. The datasets presented in this study can be found in an online repository, http://www.biosino.org. The data accession number is OEP000138. NODE-SCC data is available upon reasonable request from the corresponding author through the repository.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Fudan-Taizhou Institute of Health Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CS and XC contributed to conception and design of the study. XC, WY, and ML generated the data. SZ, TQ, WW, AP-K, ZY, KX, TZ, KZ, ZZ, and NC organized and preprocessed the datasets. AP-K developed methodology, performed the statistical analysis and wrote the first draft of the manuscript. WR and CS provided statistical advise. SZ interpreted the results and wrote Discussions of the manuscript. YX, WR and CS revised the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This study was funded by the National Key Research and Development Program of China (2017YFC0907000 to XC and 2019YFC1315804 to CS) and Shanghai Municipal Science and Technology Major Project (ZD2021CY001 to CS). This study was also supported by the National Science Centre, Poland (grant Miniatura 2 no. 2018/02/X/ST6/02571).

ACKNOWLEDGMENTS

We would like to thank the interviewers and technicians at the Fudan-Taizhou Institute of Health Sciences for their invaluable contribution to data collection and sample preparation, the staff at the Taixing Center for Disease Control and Prevention for helping us with field work, and the staff at the Taixing People's Hospital for their assistance with sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.844542/full#supplementary-material

REFERENCES

Aasen, T., Leithe, E., Graham, S. V., Kameritsch, P., Mayán, M. D., Mesnil, M., et al. (2019). Connexins in Cancer: Bridging the Gap to the Clinic. *Oncogene* 38, 4429–4451. doi:10.1038/s41388-019-0741-6

Abdo, J., Agrawal, D. K., and Mittal, S. K. (2017). "Targeted" Chemotherapy for Esophageal Cancer. *Front. Oncol.* 7, 63. doi:10.3389/fonc.2017.00063

Abusamra, H. (2013). A Comparative Study of Feature Selection and Classification Methods for Gene Expression Data of Glioma. *Procedia Comput. Sci.* 23, 5–14. doi:10.1016/j.procs.2013.10.003

Aka, A. A., Rappaport, J. A., Pattison, A. M., Sato, T., Snook, A. E., and Waldman, S. A. (2017). Guanylate Cyclase C as a Target for Prevention, Detection, and Therapy in Colorectal Cancer. *Expert Rev. Clin. Pharmacol.* 10 (5), 549–557. doi:10.1080/17512433.2017.1292124

Bang, Y.-J., Van Cutsem, E., Feyereislova, A., Chung, H. C., Shen, L., Sawaki, A., et al. (2010). Trastuzumab in Combination with Chemotherapy versus Chemotherapy Alone for Treatment of HER2-Positive Advanced Gastric or Gastro-Oesophageal Junction Cancer (ToGA): a Phase 3, Open-Label, Randomised Controlled Trial. *Lancet* 376 (9742), 687–697. doi:10.1016/s0140-6736(10)61121-x

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., and Lang, M. (2020). Benchmark for Filter Methods for Feature Selection in High-Dimensional Classification Data. *Comput. Statistics Data Analysis* 143, 106839. doi:10.1016/j.csda.2019.106839

Bonacquisti, E. E., and Nguyen, J. (2019). Connexin 43 (Cx43) in Cancer: Implications for Therapeutic Approaches via Gap Junctions. *Cancer Lett.* 442, 439–444. doi:10.1016/j.canlet.2018.10.043

Breiman, L. (2001). Random Forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324

Butt, M. A., Pye, H., Haidry, R. J., Oukrif, D., Khan, S.-U. -R., Puccio, I., et al. (2017). Upregulation of Mucin Glycoprotein MUC1 in the Progression to Esophageal Adenocarcinoma and Therapeutic Potential with a Targeted Photoactive Antibody-Drug Conjugate. *Oncotarget* 8 (15), 25080–25096. doi:10.18632/oncotarget.15340

Cao, Y., Zhao, G., Yuan, M., Liu, X., Ma, Y., Cao, Y., et al. (2021). KCNQ5 and C9orf50 Methylation in Stool DNA for Early Detection of Colorectal Cancer. *Front. Oncol.* 10, 621295. doi:10.3389/fonc.2020.621295

Chen, W., Zhang, Z., Zhang, S., Zhu, P., Ko, J. K.-S., and Yung, K. K.-L. (2021). MUC1: Structure, Function, and Clinic Application in Epithelial Cancers. *Int. J. Mol. Sci.* 22 (12), 6567. doi:10.3390/ijms22126567

Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., et al. (2018). DGIdb 3.0: a Redesign and Expansion of the Drug-Gene Interaction Database. *Nucleic acids Res.* 46 (D1), D1068–D1073. doi:10.1093/nar/gkx1143

Davidson, M., Chau, I., Cunningham, D., Khabra, K., Iveson, T., Hickish, T., et al. (2017). Impact of Tumour Histological Subtype on Chemotherapy Outcome in Advanced Oesophageal Cancer. *World. J. Gastrointest. Oncol.* 9 (8), 333–340. doi:10.4251/wjgo.v9.i8.333

Dessì, N., Pascariello, E., and Pes, B. (2013). A Comparative Analysis of Biomarker Selection Techniques. *Biomed Res. Int.* 2013, 1–10. doi:10.1155/2013/387673

Ding, C., and Peng, H. (2005). Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *J. Bioinform. Comput. Biol.* 03 (2), 185–205. doi:10.1142/S0219720005001004

Dubecz, A., Gall, I., Solymosi, N., Schweigert, M., Peters, J. H., Feith, M., et al. (2012). Temporal Trends in Long-Term Survival and Cure Rates in Esophageal Cancer: a SEER Database Analysis. *J. Thorac. Oncol.* 7 (2), 443–447. doi:10.1097/JTO.0b013e3182397751

Enzinger, P. C., and Mayer, R. J. (2003). Esophageal Cancer. *N. Engl. J. Med.* 349 (23), 2241–2252. doi:10.1056/NEJMra035010

Ferńandez-Delgado, M., and Cernadas, E. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *J. Mach. Learn. Res.* 15, 3133

Fichter, C. D., Timme, S., Braun, J. A., Gudernatsch, V., Schöpflin, A., Bogatyreva, L., et al. (2014). EGFR, HER2 and HER3 Dimerization Patterns Guide Targeted Inhibition in Two Histotypes of Esophageal Cancer. *Int. J. Cancer* 135 (7), 1517–1530. doi:10.1002/ijc.28771

Gebski, V., Burmeister, B., Smithers, B. M., Foo, K., Zalcberg, J., and Simes, J. (2007). Survival Benefits from Neoadjuvant Chemoradiotherapy or Chemotherapy in Oesophageal Carcinoma: a Meta-Analysis. *Lancet Oncol.* 8 (3), 226–234. doi:10.1016/S1470-2045(07)70039-6

Gentleman, R., Carey, V. J., Huber, W., and Hahne, F. (2020). Genefilter: Methods for Filtering Genes from High-Throughput Experiments. *R package version 3.11*. doi:10.18129/B9.bioc.genefilter

Greenawalt, D. M., Duong, C., Smyth, G. K., Ciavarella, M. L., Thompson, N. J., Tiang, T., et al. (2007). Gene Expression Profiling of Esophageal Cancer: Comparative Analysis of Barrett's Esophagus, Adenocarcinoma, and Squamous Cell Carcinoma. *Int. J. Cancer* 120 (9), 1914–1921. doi:10.1002/ijc.22501

Honing, J., Smit, J. K., Muijs, C. T., Burgerhof, J. G. M., de Groot, J. W., Paardekooper, G., et al. (2014). A Comparison of Carboplatin and Paclitaxel with Cisplatinum and 5-fluorouracil in Definitive Chemoradiation in Esophageal Cancer Patients. *Ann. Oncol.* 25 (3), 638–643. doi:10.1093/annonc/mdt589

Hosseinzadeh, A., Merikhian, P., Naseri, N., Eisavand, M. R., and Farahmand, L. (2022). MUC1 Is a Potential Target to Overcome Trastuzumab Resistance in Breast Cancer Therapy. *Cancer Cell Int.* 22 (110), 22. doi:10.1186/s12935-022-02523-z

Huang, F.-L., and Yu, S.-J. (2018). Esophageal Cancer: Risk Factors, Genetic Association, and Treatment. *Asian J. Surg.* 41 (3), 210–215. doi:10.1016/j.asjsur.2016.10.005

Kalra, A. V., and Campbell, R. B. (2009). Mucin Overexpression Limits the Effectiveness of 5-FU by Reducing Intracellular Drug Uptake and Antineoplastic Drug Effects in Pancreatic Tumours. *Eur. J. Cancer* 45, 164–173. doi:10.1016/j.ejca.2008.10.008

Katturajan, R., and Evan Prince, S. (2021). A Role of Connexin 43 on the Drug-Induced Liver, Kidney, and Gastrointestinal Tract Toxicity with Associated Signaling Pathways. *Life Sci.* 280, 119629. doi:10.1016/j.lfs.2021.119629

Kononenko, I. (1994). "Estimating Attributes: Analysis and Extensions of RELIEF," in *Machine Learning: ECML-94. ECML 1994. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence*. Editors F. Bergadano and L. De Raedt, 171–182. doi:10.1007/3-540-57868-4_57

Lagergren, J., Smyth, E., Cunningham, D., and Lagergren, P. (2017). Oesophageal Cancer. *Lancet* 390 (10110), 2383–2396. doi:10.1016/s0140-6736(17)31462-9

Lee, D.-H., Choi, S., Park, Y., and Jin, H.-s. (2021). Mucin1 and Mucin16: Therapeutic Targets for Cancer Therapy. *Pharmaceuticals* 14 (10), 1053. doi:10.3390/ph14101053

Leek, J. T., Johnson, W. E., Parker, H. S., Fertig, E. J., Jaffe, A. E., Zhang, Y., et al. (2018). Surrogate Variable Analysis. *R package version 3.8*. doi:10.18129/B9.bioc.sva

Li, Z.-h., Zheng, R., Chen, J.-t., Jia, J., and Qiu, M. (2016). The Role of Copper Transporter ATP7A in Platinum-Resistance of Esophageal Squamous Cell Cancer (ESCC). *J. Cancer* 7 (14), 2085–2092. doi:10.7150/jca.16117

Liang, K.-H., Lin, C.-C., and Yeh, C.-T. (2011). GALNT14 SNP as a Potential Predictor of Response to Combination Chemotherapy Using 5-FU, Mitoxantrone and Cisplatin in Advanced HCC. *Pharmacogenomics* 12 (7), 1061–1073. doi:10.2217/pgs.11.43

Liaw, A., and Wiener, M. (2007). Classification and Regression by randomForest. *R. news* 2 (3), 18

Lin, E. W., Karakasheva, T. A., Lee, D.-J., Lee, J.-S., Long, Q., Bass, A. J., et al. (2017). Comparative Transcriptomes of Adenocarcinomas and Squamous Cell Carcinomas Reveal Molecular Similarities that Span Classical Anatomic Boundaries. *PLoS Genet.* 13 (8), e1006938. doi:10.1371/journal.pgen.1006938

Lin, W.-R., and Yeh, C.-T. (2020). GALNT14: An Emerging Marker Capable of Predicting Therapeutic Outcomes in Multiple Cancers. *Int. J. Mol. Sci.* 21 (4), 1491. doi:10.3390/ijms21041491

Lordick, F., Mariette, C., Haustermans, K., Obermannová, R., and Arnold, D. (2016). Oesophageal Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* 27 (5), v50–v57. doi:10.1093/annonc/mdw329

Luo, J., Jin, Y., Li, M., and Dong, L. (2021). Tumor Suppressor miR613 Induces Cisplatin Sensitivity in Non small Cell Lung Cancer Cells by Targeting GJA1. *Mol. Med. Rep.* 23 (5), 385. doi:10.3892/mmr.2021.12024

Luque, A., Carrasco, A., Martín, A., and de las Heras, A. (2019). The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix. *Pattern Recognit.* 91, 216–231. doi:10.1016/j.patcog.2019.02.023

Lustgarten, J. L., Gopalakrishnan, V., and Visweswaran, S. (2009). Measuring Stability of Feature Selection in Biomedical Datasets. *AMIA Symp.*, 406

Network, T. C. G. A. R. (2017). Integrated Genomic Characterization of Oesophageal Carcinoma. *Nature* 541 (7636), 169–175. doi:10.1038/nature20805

Noone, A. M. H. N., Krapcho, M., Miller, D., Brest, A., Yu, M., et al. (2018). *SEER Cancer Statistics Review, 1975–2015*. Bethesda, MD: National Cancer Institute.

O'Brien, C., Cavet, G., Pandita, A., Hu, X., Haydu, L., Mohan, S., et al. (2008). Functional Genomics Identifies ABCC3 as a Mediator of Taxane Resistance in HER2-Amplified Breast Cancer. *Cancer Res.* 68 (13), 5380–5389. doi:10.1158/0008-5472.CAN-08-0234

Piliszek, R., Mnich, K., Migacz, S., Tabaszewski, P., Sułecki, A., Polewko-Klim, A., et al. (2019). MDFS: MultiDimensional Feature Selection in R. *R J.* 11 (1), 198–210. doi:10.32614/RJ-2019-019

Polewko-Klim, A., and Rudnicki, W. R. (2020). Analysis of Ensemble Feature Selection for Correlated High-Dimensional RNA-Seq Cancer Data. *Comput. Sci. ICCS* 12139, 525–538. doi:10.1007/978-3-030-50420-5_39

R Core Team (2017a). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

R CoreTeam (2017b). R: A Language and Environment for Statistical Computing. [Online]. Available at: https://www.R-project.org/

Raina, D., Uchida, Y., Kharbanda, A., Rajabi, H., Panchamoorthy, G., Jin, C., et al. (2014). Targeting the MUC1-C Oncoprotein Downregulates HER2 Activation and Abrogates Trastuzumab Resistance in Breast Cancer Cells. *Oncogene* 33 (26), 3422–3431. doi:10.1038/onc.2013.308

Salem, M. E., Puccini, A., Xiu, J., Raghavan, D., Lenz, H.-J., Korn, W. M., et al. (2018). Comparative Molecular Analyses of Esophageal Squamous Cell Carcinoma, Esophageal Adenocarcinoma, and Gastric Adenocarcinoma. *Oncologist* 23 (11), 1319–1327. doi:10.1634/theoncologist.2018-0143

Salem, M. E., Xiu, J., Khan, S. A., Wang, H., Smaglo, B. G., El-Deiry, W. S., et al. (2016). Comparative Molecular Analyses of Esophageal Cancer: Adenocarcinoma *vs.* Squamous Cell Carcinomas and Impact on Outcome. *J. Clin. oncol.* 34 (15), 4035. doi:10.1200/JCO.2016.34.15_suppl.4035

Sasaki, H., Aoyagi, K., Muto, M., and Takahashi, H. (2017). *Method for Evaluating Efficacy of Chemoradiotherapy against Squamous Cell CarcinomaUS 20170292955 A1 United States*. Available at: https://patents.google.com/patent/US10969390B2/en

Shimada, M., Nishimura, R., Nogawa, T., Hatae, M., Takehara, K., Yamada, H., et al. (2013). Comparison of the Outcome between Cervical Adenocarcinoma and Squamous Cell Carcinoma Patients with Adjuvant Radiotherapy Following Radical Surgery: SGSG/TGCU Intergroup Surveillance. *Mol. Clin. Oncol.* 1 (4), 780–784. doi:10.3892/mco.2013.112

Shorthouse, D., Rahrmann, E., Kosmidou, C., Greenwood, B., Hall, M. W. J., Devonshire, G., et al. (2020). *KCNQ Gene Family Members Act as Both Tumor Suppressors and Oncogenes in Gastrointestinal Cancers bioRxiv*. doi:10.1101/2020.03.10.984039

Simon, R. (2020). BRB-ArrayTools Development Team. [Online]. Available: https://brb.nci.nih.gov/BRB-ArrayTools/Documentation.html

Singh, P., Toom, S., and Huang, Y. (2017). Anti-claudin 18.2 Antibody as New Targeted Therapy for Advanced Gastric Cancer. *J. Hematol. Oncol.* 10, 0473. doi:10.1186/s13045-017-0473-4

Song, Z.-B., Gao, S. S., Yi, X. N., Li, Y. J., Wang, Q. M., Zhuang, Z. H., et al. (2003). Expression of MUC1 in Esophageal Squamous-Cell Carcinoma and its Relationship with Prognosis of Patients from Linzhou City, a High Incidence Area of Northern China. *World. J. Gasterentrol.* 9, 404. doi:10.3748/wjg.v9.i3.404

Tanaka, T., Kimura, M., Ishiguro, H., Mizoguchi, K., and Takeyama, H. (2016). Connexin 43 Expression Is Associated with Poor Survival in Patients with Esophageal Squamous Cell Carcinoma. *Mol. Clin. Oncol.* 4, 989. doi:10.3892/mco.2016.828

Tołoczko-Iwaniuk, N., Dziemiańczyk-Pakieła, D., Nowaszewska, B. K., Celińska-Janowicz, K., and Miltyk, W. (2019). Celecoxib in Cancer Therapy and Prevention - Review. *Curr. Drug. Target.* 20, 302–315. doi:10.2174/1389450119666180803121737

Tsou, Y.-K., Liang, K.-H., Lin, W.-R., Chang, H.-K., Tseng, C.-K., and Yeh, C.-T. (2017). GALNT14 Genotype as a Response Predictor for Concurrent Chemoradiotherapy in Advanced Esophageal Squamous Cell Carcinoma. *Oncotarget* 8, 29151. doi:10.18632/oncotarget.16253

Turgut, M., Kılıc, K. D., Basaloglu, H. K., and Turgut, M. (2017). Critical Role of a Novel Biological Marker GALNT14 Expression in Different Cancer Types. *J. Brain Tumors Neurooncol* 02 (1). e106. doi:10.4172/2475-3203.1000e106

Tustumi, F., Takeda, F. R., Kimura, C. M. S., Sallum, R. A. A., Ribeiro Junior, U., and Cecconello, I. (2016). Esophageal Carcinoma: Is Squamous Cell Carcinoma Different Disease Compared to Adenocarcinoma? A Transversal Study in a Quaternary High Volume Hospital in Brazil. *Arq. Gastroenterol.* 53(1), 44–48. doi:10.1590/s0004-28032016000100009

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-Based Map of the Human Proteome. *Science* 347, 1260419. doi:10.1126/science.1260419

Wang, K., Johnson, A., Ali, S. M., Klempner, S. J., Bekaii-Saab, T., Vacirca, J. L., et al. (2015). Comprehensive Genomic Profiling of Advanced Esophageal Squamous Cell Carcinomas and Esophageal Adenocarcinomas Reveals Similarities and Differences. *Oncologist* 20, 1132. doi:10.1634/theoncologist.2015-0156

Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., et al. (2020). Therapeutic Target Database 2020: Enriched Resource for Facilitating Research and Early Development of Targeted Therapeutics. *Nucleic Acids Res.* 48, D1031–D1041. doi:10.1093/nar/gkz981

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Welch, B. L. (1947). The Generalization of 'Student's' Problem when Several Different Population Varlances Are Involved. *Biometrika* 34(1-2), 28–35. doi:10.1093/biomet/34.1-2.28

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a Comprehensive Resource for In Silico Drug Discovery and Exploration. *Nucleic Acids Res.* 34, D668–D672. doi:10.1093/nar/gkj067

Xin, Z., Song, X., Jiang, B., Gongsun, X., Song, L., Qin, Q., et al. (2018). Blocking FGFR4 Exerts Distinct Anti-tumorigenic Effects in Esophageal Squamous Cell Carcinoma. *Thorac. Cancer* 9, 1687–1698. doi:10.1111/1759-7714.12883

Ye, Q., Yan, Z., Liao, X., Li, Y., Yang, J., Sun, J., et al. (2011). MUC1 Induces Metastasis in Esophageal Squamous Cell Carcinoma by Upregulating Matrix Metalloproteinase 13. *Lab. Invest.* 91, 778–787. doi:10.1038/labinvest.2011.12

Yu, L., and Liu, H. (2003). Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution." in *Proceedings, Twentieth International Conference on Machine Learning*. Editor T. Fawcett and N. Mishra (Washington, DC: Proceedings, Twentieth International Conference on Machine Learning) 2, 856–863.

Zeng, J.-H., Xiong, D.-D., Pang, Y.-Y., Zhang, Y., Tang, R.-X., Luo, D.-Z., et al. (2017). Identification of Molecular Targets for Esophageal Carcinoma Diagnosis Using miRNA-Seq and RNA-Seq Data from the Cancer Genome Atlas: a Study of 187 Cases. *Oncotarget* 8, 35681–35699. doi:10.18632/oncotarget.16051

Zhan, C., Yan, L., Wang, L., Jiang, W., Zhang, Y., Xi, J., et al. (2016). Landscape of Expression Profiles in Esophageal Carcinoma by the Cancer Genome Atlas Data. *Dis. Esophagus* 29, 920–928. doi:10.1111/dote.12416

Zhao, X., Yu, C., Zheng, M., and Sun, J. (2019). Prognostic Value of the mRNA Expression of Gap Junction α Members in Patients with Gastric Cancer. *Oncol. Lett.* 18 ,1669. doi:10.3892/ol.2019.10516

Zhao, Y.-Q., Wu, T., Wang, L.-F., Yin, B., Shi, M., Jiang, B., et al. (2021). Targeting MUC1-C Reverses the Cisplatin Resistance of Esophageal Squamous Cell Carcinoma *In Vitro* and *In Vivo*. *Transl. Cancer Res. TCR* 10, 645–655. doi:10.21037/tcr-20-2495

Zhou, S.-F., Wang, L.-L., Di, Y., Xue, C., Duan, W., Li, C., et al. (2008). Substrates and Inhibitors of Human Multidrug Resistance Associated Proteins and the Implications in Drug Development. *Curr. Med. Chem.* 15, 1981–2039. doi:10.2174/092986708785132870

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Polewko-Klim, Zhu, Wu, Xie, Cai, Zhang, Zhu, Qing, Yuan, Xu, Zhang, Lu, Ye, Chen, Suo and Rudnicki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fgene.2022.876253

NAP1L1 Functions as a Novel Prognostic Biomarker Associated With Macrophages and Promotes Tumor Progression by Influencing the Wnt/β-Catenin Pathway in Hepatocellular Carcinoma

Bingbing Shen[1†], *Wenjie Zhu*[1†], *Xinyuan Liu*[1,2†] *and Jianxin Jiang*[1*]

[1]*Department of Hepatobiliary Surgery, Renmin Hospital of Wuhan University, Wuhan, China,* [2]*Department of Hepatic–Biliary-Pancreatic Surgery, The Affiliated Hospital of Guizhou Medical University, Guiyang, China*

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Sze Keong Tey,
Nanyang Technological University, Singapore
Anna Kan,
Sun Yat-sen University Cancer Center, China

***Correspondence:**
Jianxin Jiang
rm002979@whu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *15 February 2022*
Accepted: *08 April 2022*
Published: *19 May 2022*

Citation:
Shen B, Zhu W, Liu X and Jiang J (2022) NAP1L1 Functions as a Novel Prognostic Biomarker Associated With Macrophages and Promotes Tumor Progression by Influencing the Wnt/β-Catenin Pathway in Hepatocellular Carcinoma. Front. Genet. 13:876253. doi: 10.3389/fgene.2022.876253

Hepatocellular carcinoma (HCC) is regarded as one of the universal cancers in the world. Therefore, our study is based on clinical, molecular mechanism and immunological perspectives to analyze how *NAP1L1* affects the progression of HCC. To begin with, the gene expression datasets and clinical data of GSE14520, GSE76427, ICGC, and TCGA are originated from GEO, ICGC, and TCGA databases. Subsequently, DEG screening was performed on data using R studio, and we finally found that 2,145 overlapping DEGs were screened from four datasets at the end. Then, we used R studio to filter the survival-related genes of the GSE76427 and ICGC datasets, and we screened out 101 survival-related genes. Finally, 33 common genes were screened out from 2,145 overlapping DEGs and 101 survival-related genes. Then, *NAP1L1* was screened from 33 common genes using the CytoHubba plug-in in Cytoscape software. Furthermore, ground on GEO, ICGC, and TCGA databases, the survival analysis, clinical feature analysis, univariate/multivariate regression analysis, and multiple GSEA were used to study NAP1L1. The Conclusion claimed that HCC patients with higher expression levels of NAP1L1 had a poorer prognosis than those with lower expression levels. Thus, we believe that *NAP1L1* is an independent prognostic factor for HCC. In order to shed light on NAP1L1's molecular mechanism promoting the progression of HCC closely, the GSEA tool was applied to complete the GSEA of the four datasets. Furthermore, the results confirmed that *NAP1L1* could promote HCC progression by regulating the G2/M transition of the cell cycle and *Wnt* signaling

Abbreviations: *NAP1L1*, nucleosome assembly protein 1-like one; HCC, hepatocellular carcinoma; GEO, Gene Expression Omnibus; ICGC, International Cancer Genome Consortium Database; TCGA, The Cancer Genome Atlas; FPKM, the fragments per kilobase million; STRING, The Search Tool for the Retrieval of Interacting Genes; TIMER, The Tumor Immune Estimation Resource; TCIA, The Cancer Immunome Atlas; TISIDB, Tumor-immune System Interactions Database; DEGs, differentially expressed genes; PPI, protein–protein Interactions; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene Set Enrichment Analysis; IPS, immuno pheno score; qRT-PCR; quantitative real-time PCR; DMEM, Dulbecco's modified eagle medium; FBS, fetal bovine serum; OS, overall survival; HPA, Human Protein Atlas; TILs, tumor-infiltrating lymphocyte; NC, negative control; *PRDM8*, PRDI-BF1 and RIZ Homology Domain Containing eight; TAMs, tumor-associated macrophages.

pathway. Western blot and flow cytometry were also performed to understand those mechanisms in this study. The result of Western blot showed that *NAP1L1* silencing led to downregulation of *CDK1* and *β-catenin* proteins; the result of flow cytometry showed that cell numbers in the G2 phase were significantly increased when *NAP1L1* was silenced. Thus, we claimed that *NAP1L1* might promote HCC progression by activating the *Wnt* signaling pathway and promoting cell cycle G2/M transition. In addition, ground on GSE14520 and GSE76427 datasets, and ICGC and TCGA databases, the correlation between *NAP1L1* and immune cells was analyzed in HCC patients. At the same time, the TISIDB online database and the TIMER online database were testified to the association between *NAP1L1* and immune cells. Hence, the summary shows that *NAP1L1* was connected with a certain amount of immune cells. We can speculate that *NAP1L1* may influence macrophages to promote HCC progression through some potential mechanisms.

Keywords: NAP1L1, Bioinformatics, cell cycle, Wnt/β-catenin, immune, hepatocellular carcinoma, clinical

1 INTRODUCTION

According to the 2020 world cancer data, the incidence of HCC ranked the seventh, and the mortality rate related to HCC ranked the fourth in the world (Sung et al., 2021). Surgical and non-surgical therapies are the primary therapy methods used for treating HCC. The surgical treatment mainly covers liver resection, liver transplantation, and radiofrequency ablation, and the non-surgical treatment is based on targeted immunotherapy such as sorafenib and monoclonal antibodies (Chen et al., 2020). Although there are tremendous treatment methods for HCC, the extreme unbalance of survival and recurrence of HCC patients is a significant difficulty to treat HCC now. Therefore, we place more value on studying the potential molecular mechanisms of HCC progression for working on the perfection of the strategies of a comprehensive therapy.

Bioinformatics has become a hot medical analysis tool to filtrate key genes. GSE14520, GSE76427, ICGC, and TCGA gene expression datasets, and clinical data were obtained from GEO, ICGC, and TCGA databases. Then, we searched for DEGs ground on the GEO microarray, ICGC, and TCGA datasets and screened survival-related genes in the GSE76427 and the ICGC datasets through R studio. Next, we obtained 33 common genes between the 2,145 DEGs and the 101 survival-related genes through the Venn online tool. Subsequently, we analyzed the 33 common genes through a PPI network. Cytoscape software, including the MCODE and the CtyoHubba plug-ins, screened out the essential modules by the PPI network and hub genes among all common genes. GEO, ICGC, and TCGA databases illustrated the correlation of hub genes with the survival and prognosis of HCC patients. The intrinsic mechanisms that hub genes use to influence HCC progression were found by analyzing multiple databases. Finally, some experiments based on its potential mechanism were carried out. Although most of our studies were based on bioinformatics analysis, this research is critical and crucial for accelerating medical progression and making a minor contribution to dig the occurrence mechanisms of HCC in the medical area. The workflow of the specific analysis is shown in **Figure 1**.

2 METHODS AND MATERIALS

2.1 The Acquisitions of Gene Expression and Clinical Data

The gene expression and clinical data covering age, gender, stage, and survival data were rooted in the GEO (https://www.ncbi.nlm.nih.gov/geo/), ICGC (https://icgc.org/), and TCGA databases (https://tcga-data.nci.nih.gov/). Gene expression and clinical data of the GSE14520 series (Roessler et al., 2010) and the GSE76427 series (Grinchuk et al., 2018) originated from the GEO database. Gene expression and clinical data of GSE14520 based on the GPL3921 platform [HT_HG-U133A] (Affymetrix HT Human Genome U133A Array), including 241 non-tumor specimens and 247 tumor specimens (including 227 paired HCC specimens). Gene expression and clinical data of GSE76427 based on the GPL10558 platform (Illumina HumanHT-12 V4.0 expression bead chip), including 52 non-tumor specimens and 115 tumor specimens (including 52 paired HCC specimens). The gene expression files (ICGC-LIRI-JP) of the ICGC database based on the Illumina HiSeq RNA Seq platform, including 202 non-tumor specimens and 243 tumor specimens (including 202 paired HCC specimens), were obtained from the ICGC database. The RNA sequencing and clinical data (TCGA-LIHC-FPKM) based on the Illumina HiSeq RNA-Seq platform, including 50 non-tumor specimens and 374 tumor specimens (including 50 paired HCC specimens), were obtained from the TCGA database.

2.2 The Identification of DEGs and Survival-Related Genes, and Screening out Common Genes

The DEGs of non-tumor specimens and tumor specimens were identified with R studio by dealing with the raw data of GSE14520, GSE76427, ICGC, and TCGA databases. In the four datasets, the RMA package was utilized to standardize the entire raw data, the Affy package was utilized to evaluate the quality of data, and the Limma package was utilized to filtrate DEGs whose standard was deemed as the $|\log FC| > 2$ as well as

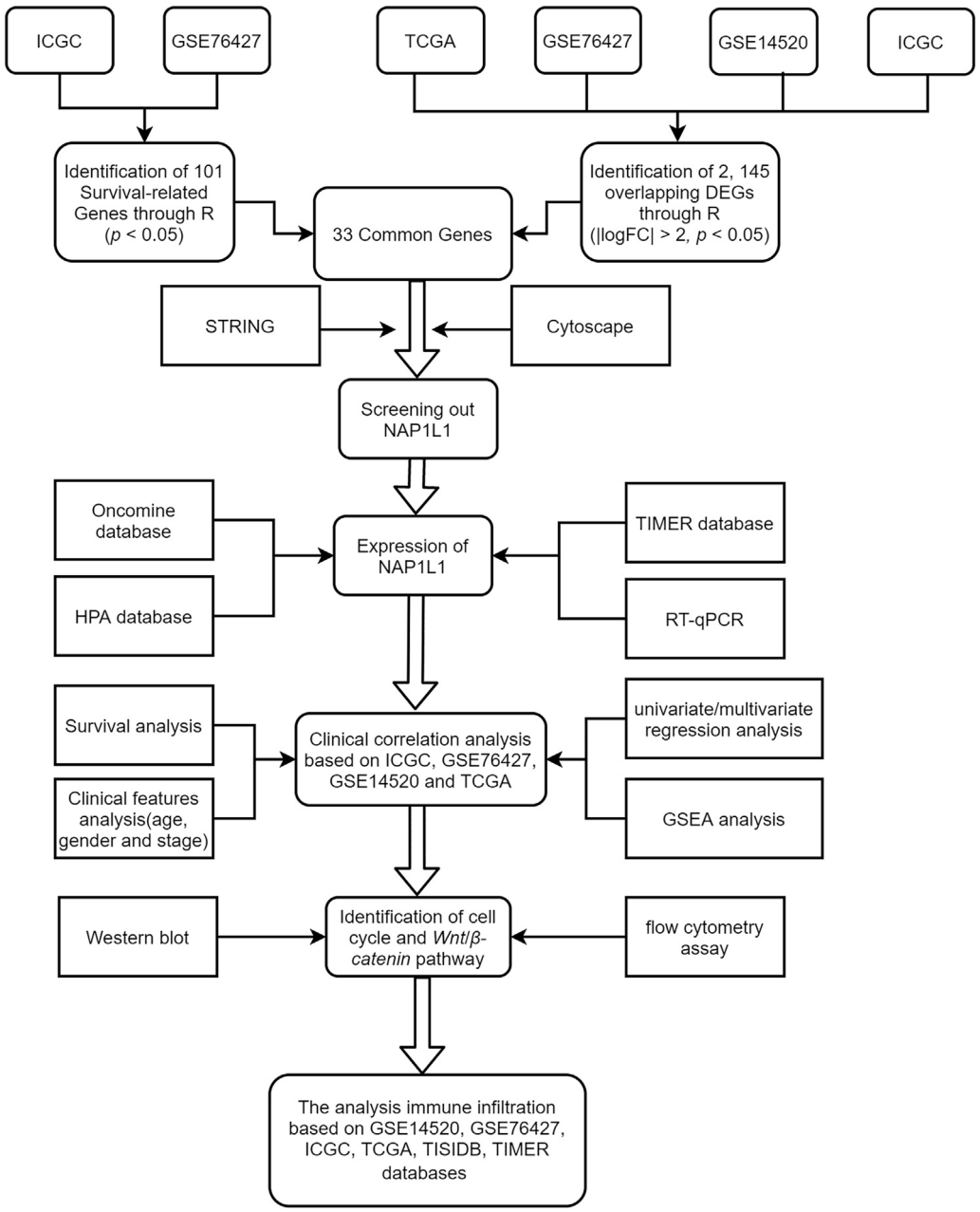


FIGURE 1 | Flow diagram for research.

the adjusted $p < 0.05$. In addition, the survival package of R studio was also utilized to filtrate survival-related genes from GSE76427 and ICGC, in which the cutoff criterion was p-values < 0.05. The common genes between DEGs and survival-related genes were selected by R studio, painting a Venn diagram.

2.3 PPI Network and Screening out Hub Gene

The STRING database (http://string-db.org)(Szklarczyk et al., 2015) formed the PPI network where the MCODE plug-in chose the essential modules in the Cytoscape application. In addition, the standards of choosing the essential modules are k-core = 2, degree cutoff = 2, node score cutoff = 0.2, and maximum depth = 100. CytoHubba plug-in was another plug-in in Cytoscape software in which the top 10 hub genes were selected in the PPI network. What is more, the 12 calculating methods in CytoHubba presented as the following: DMNC, ClusteringCoefficient, EPC, Degree, BottleNeck, EcCentricity, MNC, Radiality, Betweenness, MCC, Stress, and Closeness, which generated 12 different outcomes; through comparing with them, the *NAP1L1* was chosen as the objective of our study.

2.4 The Analysis of the Oncomine Database and HPA Database

The Oncomine database (https://www.oncomine.org/resource/login.html)(Rhodes et al., 2007) consists of 86,733 specimens and 715 gene expression data. Moreover, it is a unitary integrated database aimed to accelerate data mining efforts effectively. Therefore, this database evaluated the *NAP1L1* expression in human tumors. Accordingly, HPA (https://www.proteinatlas.org) is deemed a kind of online tool whose goal was to elucidate the level of *NAP1L1* protein exiting in non-tumor tissues and HCC tissues (Pontén et al., 2011). In 2003, the Swedish-based program, the HPA, witnessed the world project of human proteins. Specifically, the cells, tissues, and organs applying the combination of diverse omics technologies containing antibody-based imaging were all taken in human proteins. HPA version 21.0, appearing in November 2021, was used in this study (Uhlén et al., 2015; Thul et al., 2017; Uhlen et al., 2017). The immunohistochemistry pictures originated from HPA. The expression of the *NAP1L1* protein was explained in non-tumor tissues and HCC tissues among HCC patients.

2.5 Survival Analysis, Clinical Feature Analysis, Univariate/Multivariate Regression Analysis, and GSEA

R studio was utilized to achieve survival analysis, clinical feature analysis, and univariate/multivariate regression analysis. Patients' gene expression datasets and clinical data were used in this study, including GSE14520, GSE76427, ICGC, and TCGA datasets. Nine packages are being taken in this research: the beeswarm, Limma, survival, survminer, Ggpubr, plyr, grid, ggplot2, and gridExtra packages. To be more specific, the beeswarm and Limma packages were used to conduct the differential expression analysis, the survival package and survminer package were applied to perform the survival analysis, the survival package also was utilized to do the univariate/multivariate regression analysis, the Ggpubr package was utilized to carry out the analysis of the clinical features, and the plyr, ggplot2, grid, and gridExtra packages were performed to analyze the multiple GSEA.

GSEA4.2.1 application was derived from the Broad Institute (http://www.gsea-msigdb.org/gsea/login.jsp) (Subramanian et al., 2005), which still played a significant role in generating enrichment results. Plus, R studio was visualized well to complete multiple GSEA results.

2.6 Immune Cell Infiltration Analysis and IPS

CIBERSORT was utilized to figure out the enrichment files of immune cells in the HCC sample ground on GSE14520, GSE76427, ICGC, and TCGA (Chen B. et al., 2018). TISIDB (http://cis.hku.hk/TISIDB/index.php) was utilized to detect the interactions between 28 types of TILs and 30 types of cancers in humans to check the tumor–immune system interactions (Ru et al., 2019). The association between *NAP1L1* and TILs was tested through Spearman's test, in which all hypothetical tests were deemed as statistically significant in that they were bilateral and p-value < 0.05. Then, to verify the enrichment of six immune cells (dendritic cells, B cells, $CD4^+$T cells, $CD8^+$T cells, macrophages, and neutrophils) in HCC patients, the TIMER online database was used to verify the relations between *NAP1L1* expression and the level of immune infiltration (Li T. et al., 2017).

IPS is a terminology that comprises MHC molecules, immunomodulators, effector cells, and suppressor cells. According to the expression of representative genes, IPS could be figured out when applying it, and the TCIA (https://tcia.at/home) was applied to acquire the IPS about HCC patients. R studio was utilized to achieve immune cell infiltration analysis. The immune cell infiltration analysis needed Limma, ggExtra, vioplot, ggpubr, ggplot2, and Venn Diagram packages. Specifically, the Limma package was used to acquire the results of CIBERSORT, and the vioplot package was utilized to paint the violin plot. Moreover, ggpubr, ggplot2, ggpubr, and ggExtra were applied to perform the correlation analysis between NAP1L1's expression and the immune cell, and to paint the correlation graph; the Venn Diagram package was used to paint the Venn graphs.

2.7 Cell Culture

As the objective of our study, the normal liver cell lines (MIHA) and the HCC cell lines (MHCC-LM3, PRF-5, Huh-7, MHCC-97H, and Hep-G2) are originated from the cell bank of the Chinese Academy of Sciences (Shanghai, China). Furthermore, they were maintained in a DMEM cell culture (Gibco, New York, United States) supplemented with 10% FBS (Gemini, West Sacramento, United States) in the incubator at 37°C with 5% CO2.

2.8 Cell Transfection

NAP1L1 siRNA was synthesized from RiboBio (Guangzhou, China) and was used to reduce the NAP1L1 expression. MHCC-97H and Huh-7 cells were cultivated into each well of the six-well plates until they were 30–60% confluent. After that, cells were washed, placed in a serum-free medium, and transfected with siRNA using Lipofectamine™2000 according to the manufacturer's instructions (Invitrogen, MA, United States). After 6 h, the medium was changed to a complete medium, and cells were cultured at 37°C in 5% CO_2. Four groups were generated for all experiments, a negative control group (negative control); a siRNA1 group (siRNA1); a siRNA2 group (siRNA2); a siRNA3 group (siRNA3).

2.9 qRT-PCR

The whole RNA was obtained at first by utilizing the TRIzol reagent and then going through the reverse transcription way to reach the cDNA under the guidance of the Thermo Fisher Scientific Reagent Kit. Finally, we did a qRT-PCR reaction of cDNA, taking a Bio-Rad qRT-PCR Checking System.

2.10 Western Blot

First of all, we centrifuged and gathered the cells when cells were transfected after 48 h. Then, the radioimmunoprecipitation assay buffer was applied to acquire the whole-cell protein lysates. Furthermore, detaching by SDS-PAGE, the protein lysates were blotted into PVF membranes (Bio-Rad). The principal antibodies, including *NAP1L1* (14898-1-AP; 45kDa; 1:1000), *CDK1* (19532-1-AP; 34kDa; 1:1000), *β-catenin* (19532-1-AP; 45kDa; 1:1000), and *GAPDH* (10494-1-AP; 36kDa; 1:1000), were attached to the membranes after blocking, which they stored one night at 4°C. Next, after the TBST was

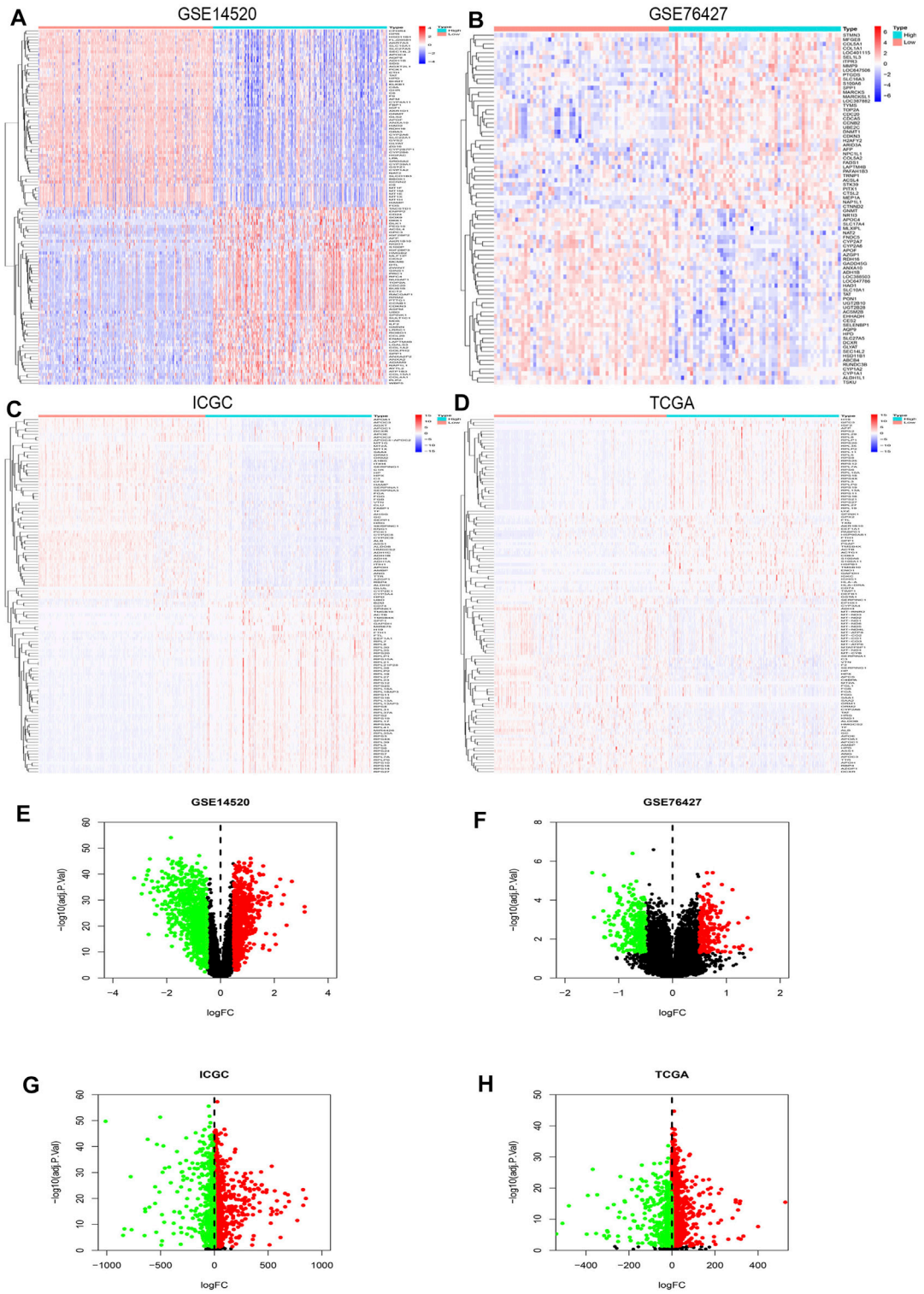


FIGURE 2 | DEGs' heatmap and volcano plot in GSE14520, GSE76427, ICGC, and TCGA: **(A)** The heatmap in GSE14520, **(B)** the heatmap in GSE76427, **(C)** the heatmap in ICGC, **(D)** the heatmap in TCGA, **(E)** the volcano plot in GSE14520, **(F)** the volcano plot in GSE76427, **(G)** the volcano plot in ICGC, and **(H)** the volcano plot in TCGA. In the heatmaps, the high expression represented red, and blue represented the low expression. Red showed the upregulation genes in the volcano plots, and green presented the downregulation genes. In addition, the black dots were not the differential gene in expression in the volcano plots.

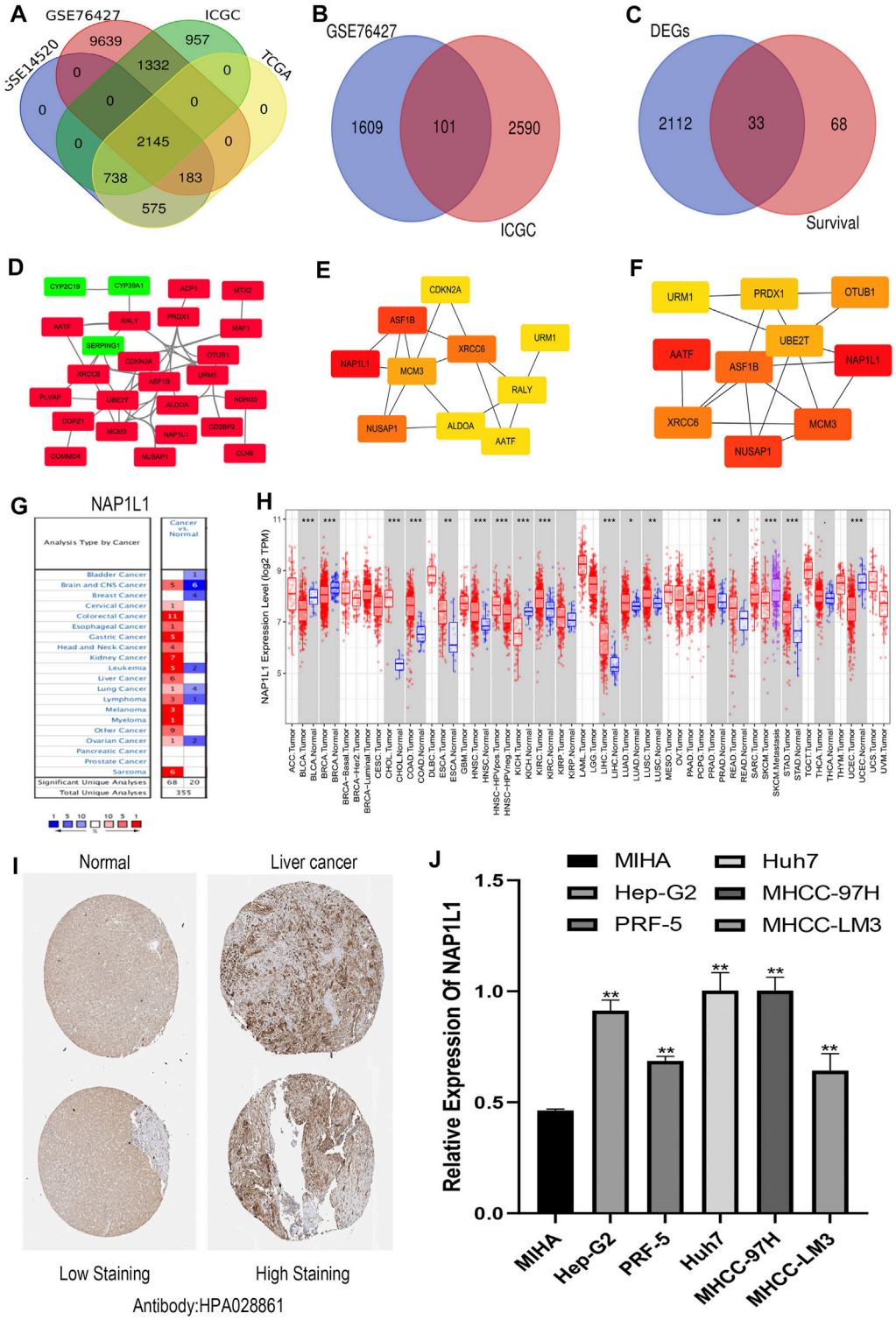


FIGURE 3 | Screening out the hub gene *NAP1L1*, and *NAP1L1* was up-regulated in HCC. **(A)** The identification of the 2,145 overlapping DEGs ground on GSE14520, GSE76427, ICGC, and TCGA. **(B)** The identification of the 101 overlapping survival-related genes ground on GSE76427 and TCGA. **(C)** The identification of the 33 common genes. **(D)** The PPI network was constructed using the 33 common genes. **(E)** The PPI network of the top 10 hub genes *via* DMNC. **(F)** The PPI network of the top 10 hub genes *via* ClusteringCoefficient. **(G)** The expression of *NAP1L1* applying the Oncomine database. **(H)** The expression of *NAP1L1* applying the TIMER database (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(I)** The expression of *NAP1L1* applying the HPA database. **(J)** The mRNA expression of *NAP1L1* applying the RT-qPCR.

washed three times, the second antibody was hatched for 2 h at homeothermy, and then TBST was eluted three times once again. In the end, the Western blot was visualized by imaging systems.

2.11 The Analysis of the Cell Cycle

The day before the transfection, MHCC-97H and Huh-7 cells were planted into the culture plates. When they were transfected successfully, those cells were eluted with PBS and fitted in 75% pre-cooling alcohol. Before applying the flow cytometer, MHCC-97H and Huh-7 cells were hatched in 500 μl specimen buffer and 0.25 mg/ml of RNase A, maintaining thirty minutes at homeothermy. Subsequently, BD FACSCalibur Flow Cytometer was utilized to count the quantities of the 2 cells in each stage, and FlowJo was applied to explain these data.

2.12 Statistical Analysis

In terms of the statistical analysis, we considered the two software: GraphPad Prism (version: 8.2.1) and R studio (version:4.0.5). Data analysis was regarded as meaningful data, while the formula $p < 0.05$ showed the meaning of a statistically significant difference.

3 RESULTS

3.1 Screening out Key Gene *NAP1L1*

R studio demonstrated the DEGs in every database (3,641 in GSE14520, 13,299 in GSE76427, 5,172 in ICGC, and 3,642 in TCGA) and DEGs whose standard was deemed as the $|\log FC| > 2$ as well as the adjusted $p < 0.05$. At the same time, R studio was also used to paint heatmaps and volcano plots about DEGs in every database, presented in figures (**Figures 2A–H**). The 2,145 overlapping DEGs were examined in the aforementioned databases, and the results were displayed in a Venn graph (**Figure 3A**). In addition, we continued to do a survival filter using R studio (1,710 in GSE76427 and 2,691 in ICGC), and the cutoff criterion was p-values < 0.05. Therefore, 101 survival-related genes were screened out. The outcomes were displayed in a Venn graph (**Figure 3B**). Furthermore, we intersected 33 common genes from DEGs and survival-related genes whose outcomes were displayed in a Venn graph (**Figure 3C**).

The STRING was applied to produce a PPI network including 33 common genes taken by the Cytoscape application. It covered 26 high-expression genes labeled with red color and seven low-expression genes labeled with green color. Here, we only showed 25 genes related to each other (**Figure 3D**). CytoHubba plug-in functions as a plug-in in Cytoscape application were selected the top 10 hub genes in the PPI network. The 12 calculating methods in cytoHubba presented as the following: DMNC, ClusteringCoefficient, EPC, Degree, BottleNeck, EcCentricity, MNC, Radiality, Betweenness, MCC, Stress, and Closeness, which generated 12 various results in the top 10 key genes through comparing with them. Moreover, we claimed that *NAP1L1* obtained the highest score in DMNC (**Figure 3E**) and ClusteringCoefficient (**Figure 3F**) so that the *NAP1L1* was chosen as the objective of our study.

3.2 Expression of *NAP1L1* in HCC

Gene expression analyses using the ONCOMINE and the TIMER claimed that *NAP1L1* mRNA levels were evidently higher in HCC than in non-tumor tissues (**Figures 3G,H**). The HPA database presented the expression levels of the *NAP1L1* protein in HCC. The levels of the *NAP1L1* protein were low in normal liver tissues, and the high expression levels of *NAP1L1* were observed in HCC tissues (**Figure 3I**). Subsequently, the qRT-PCR measured the *NAP1L1* mRNA level (**Figure 3J**). Thus, the expression of *NAP1L1* was higher based on the normal liver tissue in HCC tissues.

3.3 Survival Analysis, Clinical Feature Analysis, Univariate/Multivariate Regression Analysis, and Multiple GSEA of *NAP1L1* in GSE14520 Datasets, GSE76427 Datasets, ICGC Datasets, and TCGA Datasets

First, the gene expression profile and the clinical information (the GSE14520, GSE76427 datasets, ICGC datasets, and TCGA datasets) were obtained from the GEO, ICGC, and TCGA databases. From those scatterplots (**Figure 4A**, **Supplementary Figures S1A,2A, 3A**), we concluded that the *NAP1L1* expression of HCC specimens was evidently higher than that of the non-tumor specimens ($p < 0.05$). From those survival curves (**Figure 4B**, **Supplementary Figures S1B,2B, 3B**), we found that the high expression of *NAP1L1* is closely associated with a bad OS ($p < 0.05$).

Furthermore, we also sought the correlations between the expression of *NAP1L1* and the HCC patients' clinical features from age (**Figure 4C**, **Supplementary Figures S1C,S2C, S3C**), gender (**Figure 4D**, **Supplementary Figures S1D,S2D, S3D**), and stage (**Figure 4E**, **Supplementary Figures S1E,S2E, S3E**). Those results showed a significant difference between the expression of *NAP1L1* and the stage of HCC patients ($p < 0.05$). The Conclusion also showed that the significant difference did not exist in age ($p > 0.05$) and gender ($p > 0.05$). Then, the univariate (**Figure 4F**, **Supplementary Figures S1F,S2F, S3F**) and the multivariate regression analyses (**Figure 4G**, **Supplementary Figures S1G,S2G, S3G**) were performed with the clinical data of HCC patients, which reveals that *NAP1L1* was statistically significant in the univariate regression analysis ($p < 0.05$) and the multivariate regression analysis ($p < 0.05$). This result claimed that *NAP1L1* was an independent prognostic factor in HCC patients.

Finally, to deeply investigate how the *NAP1L1* promotes HCC progression, the HCC patients were separated into the high and low expression groups ground on the expression of *NAP1L1*. Furthermore, we took multiple GSEA employing the GO and the KEGG, and the enrichment results of the top five genes of upregulation and the top five genes of downregulation were displayed in GO and KEGG gene sets (**Figures 4H,I**, **Supplementary Figures S1H, S1I, S2H, S2I, S3H, S3I**).

In Conclusion, the aforementioned results in GSE14520 datasets, GSE76427 datasets, ICGC datasets, and TCGA datasets showed that NAP1L1 is a high expression in HCC tissue and is closely associated with poorer prognosis of HCC patients. In addition, we found correlations between the expression of NAP1L1 and clinical features in HCC patients, and NAP1L1 is an independent prognostic factor in HCC patients. which could promote the progression of HCC by some complex mechanisms.

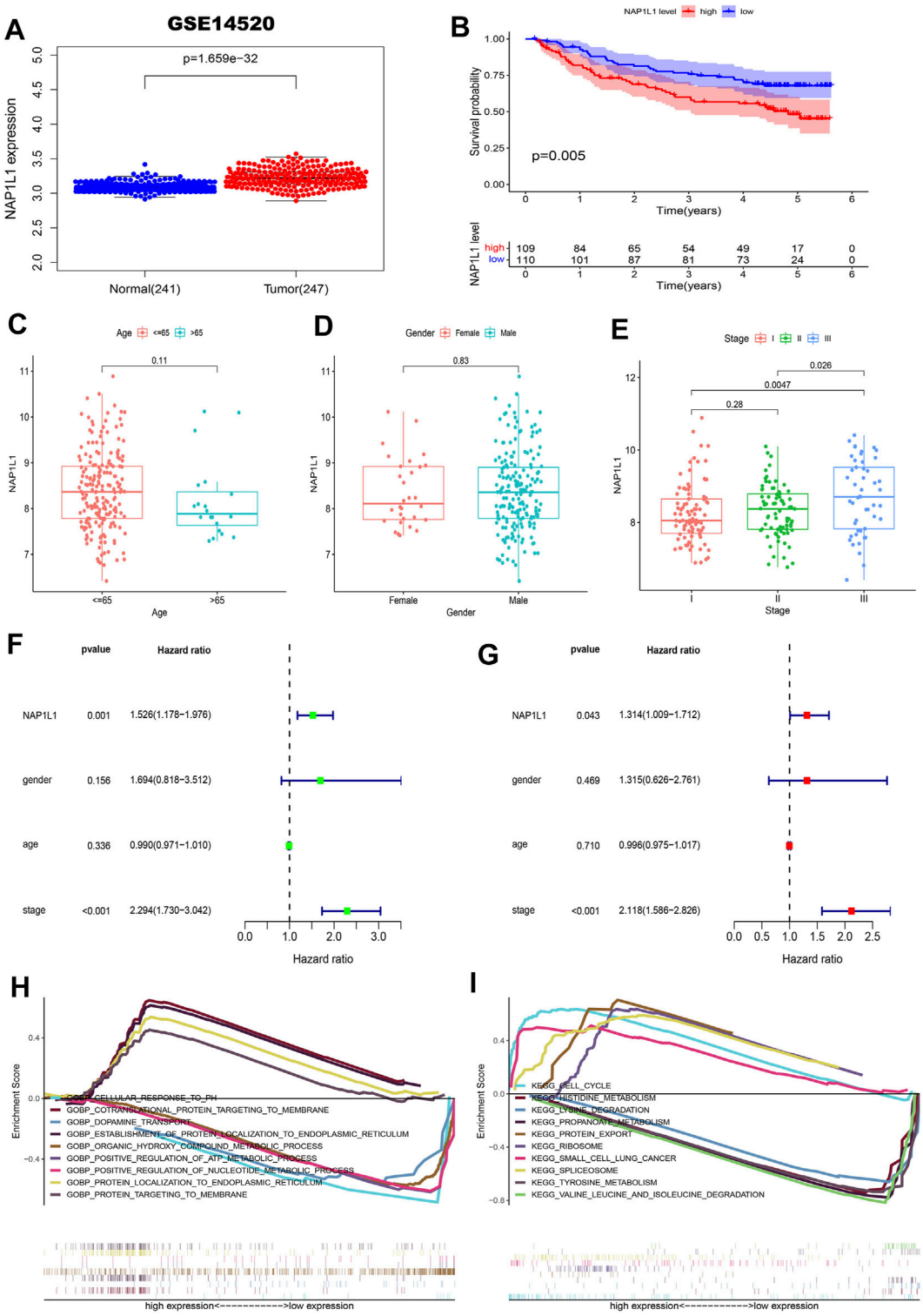


FIGURE 4 | Survival analysis, the clinical feature analysis, the univariate/multivariate regression analysis, and the multiple GSEA in the GSE14520 datasets. **(A)** The scatterplot of the *NAP1L1* expression in the non-tumor and HCC specimens. **(B)** The survival curve of OS in the high and the low *NAP1L1* expression groups. **(C)** The bar graph in age. **(D)** The bar graph in gender. **(E)** The bar graph in stage. **(F)** The univariate regression analysis. **(G)** The multivariate regression analysis. **(H)** The multiple GSEA ground on the GO. **(I)** The multiple GSEA ground on the KEGG.

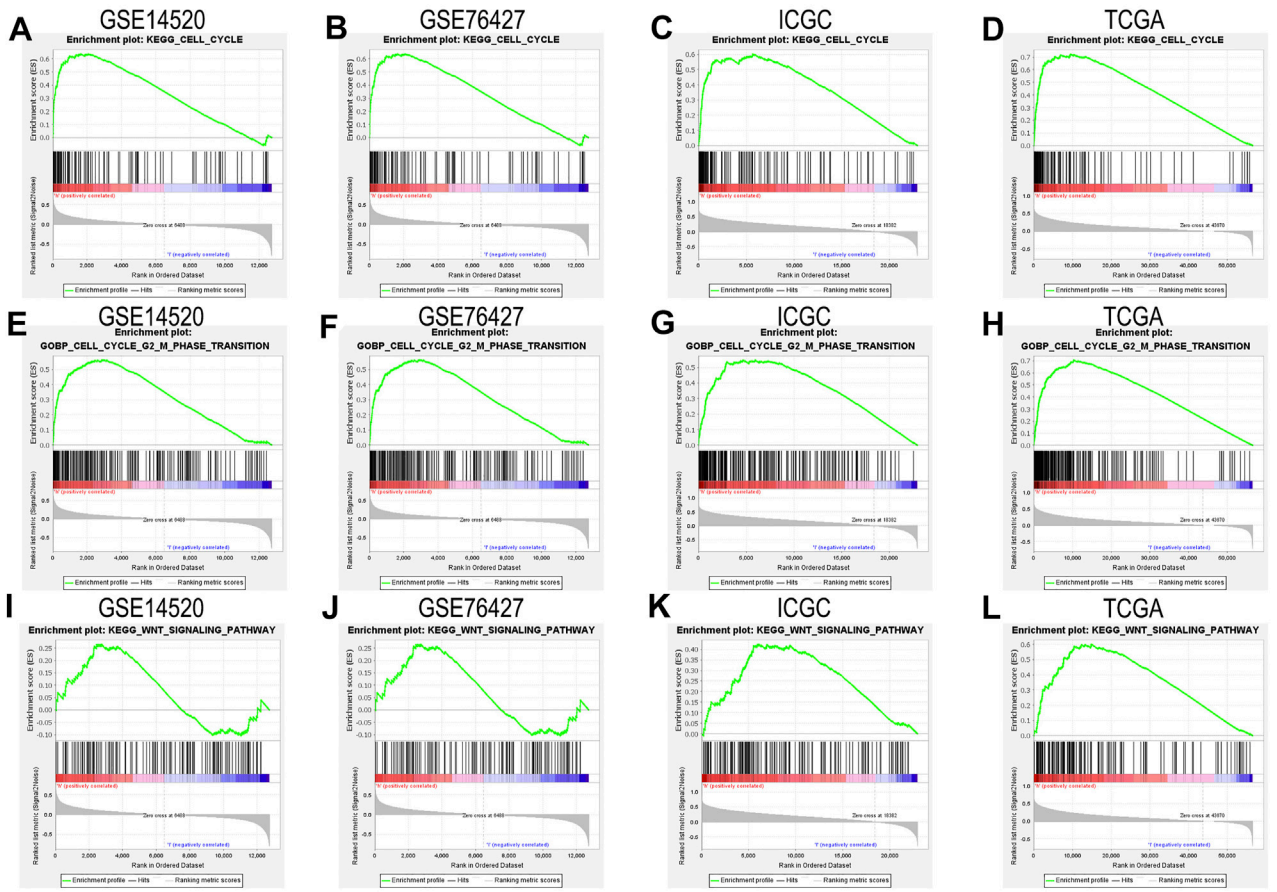


FIGURE 5 | The GSEA: **(A)** the GSEA ground on KEGG in the GSE14520 dataset, **(B)** the GSEA ground on KEGG in the GSE76427 dataset, **(C)** the GSEA ground on KEGG in the ICGC dataset, **(D)** the GSEA ground on KEGG in the TCGA dataset, **(E)** the GSEA ground on GO in the GSE14520 dataset, **(F)** the GSEA ground on GO in the GSE76427 dataset, **(G)** the GSEA ground on GO in the ICGC dataset, **(H)** the GSEA ground on GO in the TCGA dataset, **(I)** the GSEA ground on KEGG in the GSE14520 dataset, **(J)** the GSEA ground on KEGG in the GSE76427 dataset, **(K)** the GSEA ground on KEGG in the ICGC dataset, and **(L)** the GSEA ground on KEGG in the TCGA dataset.

3.4 The Cell Cycle and Wnt Signaling Pathway Through GSEA in GSE14520 Datasets, GSE76427 Datasets, ICGC Datasets, and TCGA Datasets

We closely understood the aforementioned GSEA results to explore complex mechanisms of *NAP1L1* that promoted HCC progression. Hence, we further found that cell cycle and *Wnt* signaling pathway can be found in the GSE14520 dataset, GSE76427 dataset, ICGC dataset, and TCGA dataset, based on GSEA. The result demonstrated that *NAP1L1* might accelerate the progression of HCC, which relies on the cell cycle (**Figures 5A–D**), especially the G2/M phase transition (**Figures 5E–H**) and Wnt signaling pathway (**Figure 5I–L**).

3.5 *NAP1L1* Relevant to the G2/M Phase Transition of the Cell Cycle and Wnt Signaling Pathway in MHCC-97H and Huh7 Cell Lines

The aforementioned results show that the G2/M phase transition of the cell cycle and *Wnt* signaling pathway can be found in the GSE14520 dataset, GSE76427 dataset, ICGC dataset, and TCGA dataset through GSEA. Then, the *NAP1L1* expression level was measured in NC, si-1, si-2, and si-3 groups in MHCC-97H and Huh7 cell lines (**Figure 6A**) *via* the Western blot. The outcomes demonstrated no changed features in NC and si-3 groups, but si-1 and si-2 groups were evidently low in terms of *NAP1L1* expression levels in them.

β-catenin was regarded as a protein to regulate the *Wnt* signaling pathway. *CDK1* was regarded as a protein to regulate G2 to M phase transition of the cell cycle. The expression of *β-catenin* and *CDK1* was measured using the Western blot (**Figure 6B**, **Supplementary Figure S4A**). The outcomes showed that after *NAP1L1* silencing, *β-catenin* and *CDK1* both downregulate accordingly, which meant that the expression of *NAP1L1* had a positive relationship between *β-catenin* and *CDK1*. In addition, after *NAP1L1* silencing, cell numbers in the G2 phase (**Supplementary Figures S4B, C**) were significantly increased in both MHCC-97H and Huh7 cells, measured by the flow cytometry assay.

Therefore, we believed that *NAP1L1* knockdown could inhibit HCC cell proliferation by inhibiting *Wnt*/*β-catenin* pathway activation and the G2/M phase of cell cycle transition.

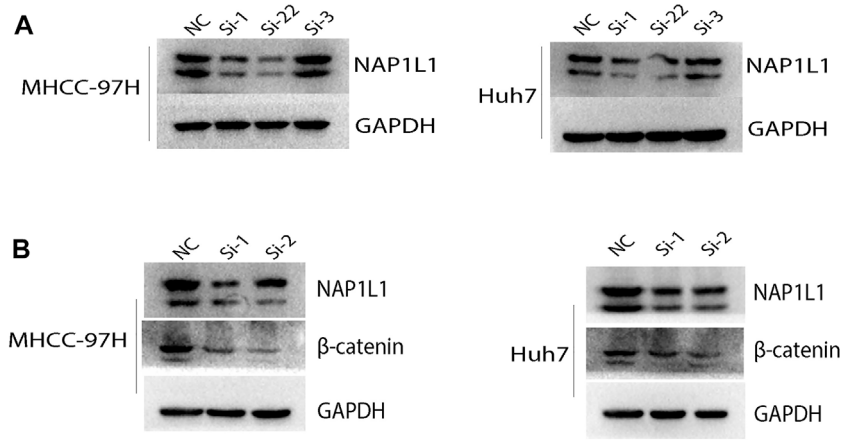


FIGURE 6 | The Western blot: **(A)** the Western blot was utilized to check the expression level of *NAP1L1* among the NC, si-1, si-2, and si-3 groups in MHCC-97H and Huh7 cells, and **(B)** the Western blot was utilized to examine the expression of *NAP1L1*, *β-catenin*, and *GAPDH*.

3.6 The Analysis of Immune Infiltration Ground on GSE14520 Dataset, GSE76427 Dataset, ICGC Dataset, and TCGA Dataset

For exploring the association between *NAP1L1* and the immune cells in HCC patients, we conducted the following analysis:

The CIBERSORT algorithm was taken to acquire the landscape of tumor-infiltrating immune cells. Bar plots were painted through R studio to test the correlation between NAP1L1 expression and 22 immune cells. From those bar plots (**Figure 7A**, **Supplementary Figures S5A, S6A, S7A**), we can observe the percent of 22 immune cells in each HCC sample. The violin plot was utilized to recognize the variety of 22 immune cells between the high and the low expression groups in HCC samples ground on the median expression level of *NAP1L1*. Furthermore, from observing those violin plots (**Figure 7B**, **Supplementary Figures S5B, S6B, S7B**), we found a remarkable difference in activated memory CD4 T cells, activated dendritic cells, gamma delta T cells, regulatory T cells (Tregs), CD8 T cells, activated NK cells, monocytes, M0 macrophages, M1 macrophages, and M2 macrophages.

The correlation graph was utilized to recognize the correlation between *NAP1L1* and 22 immune cells. Then these significant immune cells were selected and presented in the following figures. They activated Mast cells (**Figure 7C**), M0 macrophages (**Figure 7D**, **Supplementary Figures S5C, S7D**), M2 macrophages (**Figure 7E**, **Supplementary Figures S6E**), activated dendritic cells (**Figure 7F**), monocytes (**Figure 7G**), activated memory CD4 T cells (**Figure 7H**), CD8 T cells (**Figure 7I**), M1 macrophages (**Supplementary Figures S6D, S7E**), gamma delta T cells (**Supplementary Figure S5D**), naive B cells (**Supplementary Figure S6C**), activated NK cells (**Supplementary Figure S6F**), and memory B cells (**Supplementary Figure S7C**). From those Venn graphs of meaningful immune cells in the violin plot and correlation graph, common immune cells were selected in GSE14520 datasets (**Figure 7J**), GSE76427 datasets (**Supplementary Figures S5E**), ICGC datasets (**Supplementary Figure S6G**), and TCGA datasets (**Supplementary Figure S7F**). We found CD8 T cells, activated memory CD4 T cells, monocytes, M0 macrophages, M2 macrophages, and activated dendritic cells as the common immune cells in GSE14520 dataset. M0 macrophages and gamma delta T cells were the common immune cells in GSE76427 datasets. Moreover, activated NK cells and M1 macrophages were common immune cells in ICGC dataset. Finally, M0 macrophages and M1 macrophages were the common immune cells in TCGA dataset.

In Conclusion, the aforementioned analysis of immune infiltration in GSE14520 dataset, GSE76427 dataset, ICGC dataset, and TCGA dataset showed that remarkable differences in immune cells existed in HCC patients, and the correlations between the expression of NAP1L1 and some immune cells were statistically significant. Finally, we speculated that NAP1L1 expression is most associated with macrophages in HCC.

3.7 The Analysis of Immune Infiltration Ground on TISIDB and TIMER Online Databases

The association between the enrichment of TILs and the expression of *NAP1L1* were inferred using the TISIDB database, and the correlation between the *NAP1L1* expression and the TILs of people's cancer is displayed in **Figure 8A**. It was a certain association between the enrichment of 28 TIL types and the expression of *NAP1L1*. Specifically speaking, *NAP1L1* expression was closely positively related to the following immune cells: activated B cells (r = 0.104, p = 0.0441), activated CD4 T cells (r = 0.4, p = 0.6.9e-17), activated CD8 T cells (r = 0.266, p = 2e-07), activated dendritic cells (r = 0.104, p = 0.0292), immature B cells (r = 0.105, p = 0.0443), macrophage (r = 0.153, p = 0.00316), mast cells (r = 0.172, p = 0.000864),

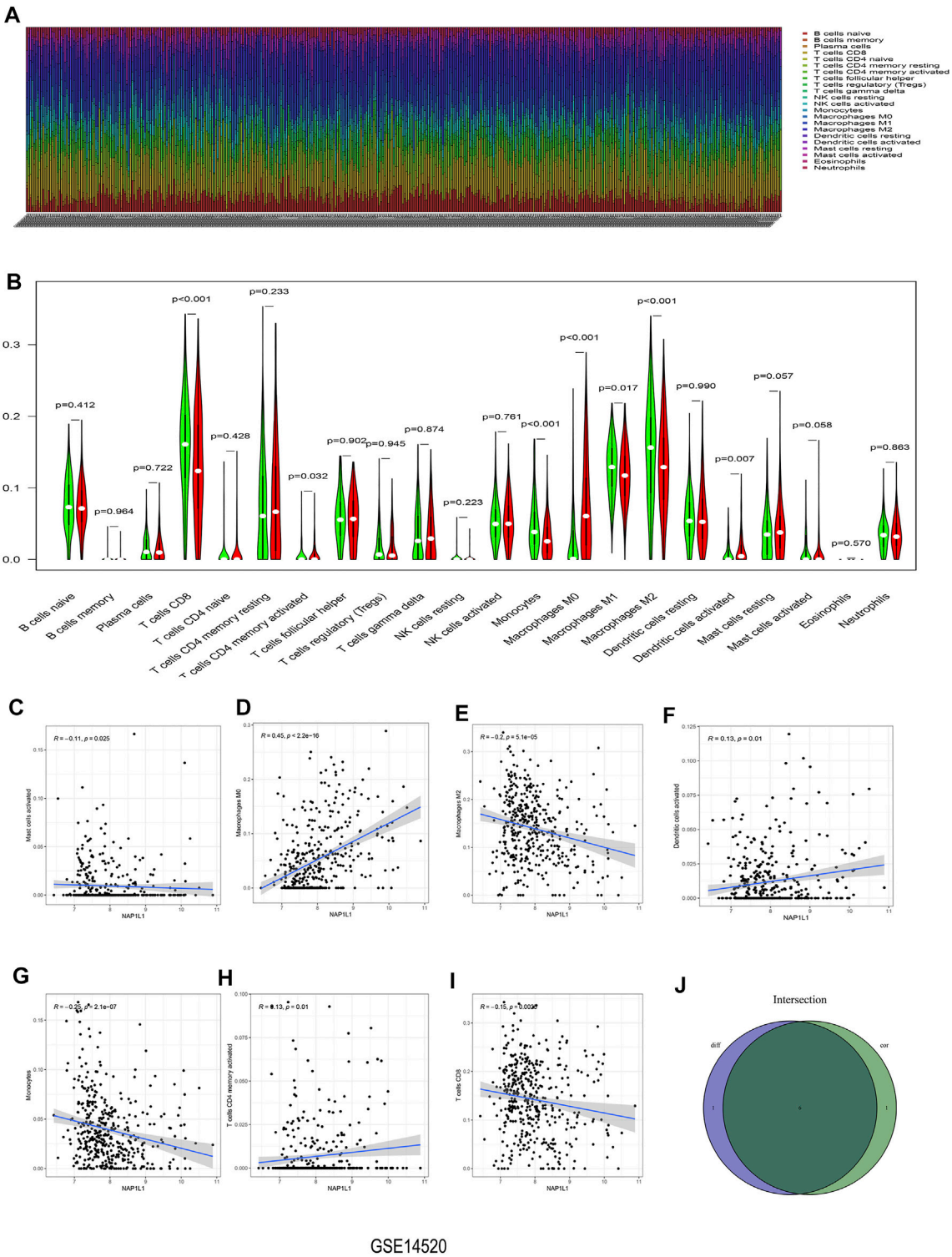


FIGURE 7 | Analysis of immune infiltration in the GSE14520 dataset. **(A)** The bar graph containing the 22 immune cells. **(B)** The violin graph containing the 22 immune cells. **(C)** The relationship between *NAP1L1* and activated mast cells. **(D)** The relationship between *NAP1L1* and M0 macrophages. **(E)** The relationship between *NAP1L1* and M2 macrophages **(F)** The relationship between *NAP1L1* and activated dendritic cells. **(G)** The relationship between *NAP1L1* and monocytes. **(H)** The relationship between *NAP1L1* and activated memory CD4 T cells. **(I)** The relationship between *NAP1L1* and CD8 T cells. **(J)** The Venn graph was utilized to picture the meaningful immune cells between the violin plot and the correlation graph.

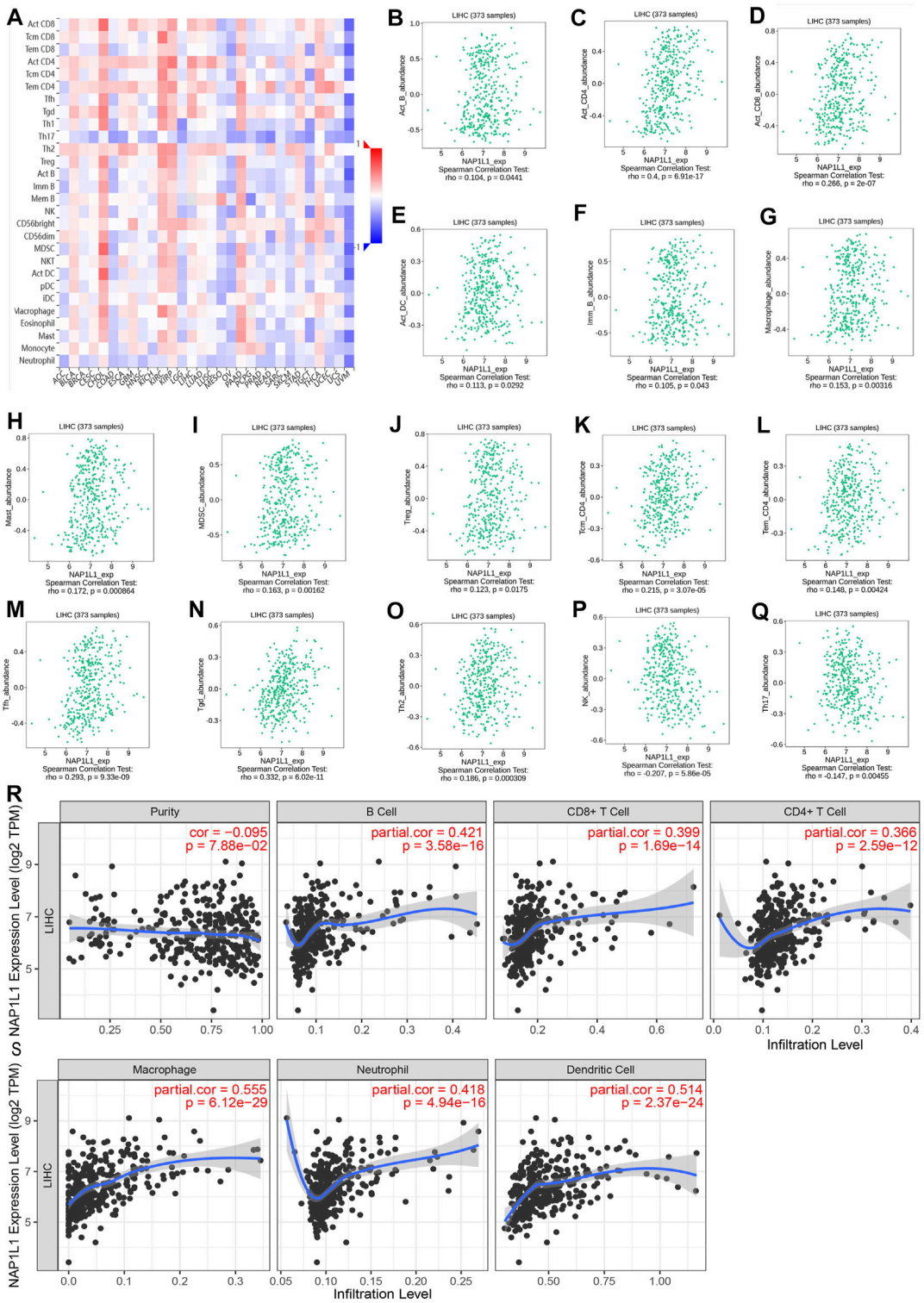


FIGURE 8 | Analysis of immune infiltration in TISIDB and TIMER. **(A)** The panoramic association between the NAP1L1 expression and TILs in human cancer (red represents the positive correlation, and blue represents the negative correlation). **(B–O)** The NAP1*L1* expression was positively related to the infiltrating levels of the immune cell in HCC. **(P,Q)** The *NAP1L1* expression was negatively related to the infiltrating levels of the immune cell in HCC. **(R,S)** The *NAP1L1* expression is correlated with the level of immune infiltration.

myeloid-derived suppressor cells (r = 0.163, p = 0.00162), regulatory T cells (r = 0.123, p = 0.0175), central memory CD4 T cells (r = 0.215, p = 3.07e-05), effector memory CD4 T cells (r = 0.148, p = 0.004), T follicular helper cells (r = 0.293, p = 9.33e-09), gamma delta T cells (r = 0.332, p = 6.02e-11), and Type-2 T helper cells (r = 0.186, p = 0.000309) (**Figure 8B**-O). In addition, the expression of *NAP1L1* was deeply negatively relevant to the following immune cells: natural killer cells (r = −c0.207, p = 5.86e-05) and Type-17 helper cells (r = −0.147, p = 0.00455) (**Figures 8P, Q**). Last, the TIMER database was utilized to further verify the relations between the *NAP1L1* expression and the immune cell infiltration in HCC (**Figures 8R, S**). The outcomes displayed that the *NAP1L1* expression was closely relevant to the following immune cells: B cells (p = 3.58e-16), $CD8^+$ T cells (p = 1.69e−14), $CD4^+$ T cells (p = 2.59e−12), macrophages (p = 6.12e−29), neutrophils (p = 4.94e−16), and dendritic cells (p = 2.37e−24).

To summarize, through the comprehensive analysis of GSE14520, GSE76427, ICGC, TCGA, TISIDB, and TIMER databases, we had a hypothesis that *NAP1L1* expression was related to immune cell and *NAP1L1* may also accelerate the HCC progression by affecting macrophage through some underlying mechanisms.

4 DISCUSSION

Hepatocellular carcinoma (HCC) is regarded as one of the universal cancers worldwide. As for the world cancer data in 2020, the incidence of HCC ranked the seventh, and the mortality rate related to HCC ranked the fourth in the world (Sung et al., 2021). Although the systemic treatment of HCC has made a significant progress in these years, the extreme unbalance of survival and recurrence of HCC patients is still a significant difficulty. It is an imperious topic to be solved in HCC today. Therefore, it is an urgent need to research the association between the hub gene and the occurrence, progression, or prognosis in HCC patients. This study may lead us to acquire some achievements in genes or diseases that the biomarker, the whole mechanisms, and a novel treatment method in HCC patients can be further promoted in the current and future.

Therefore, we screened out 2,145 DEGs and 101 survival-related genes through bioinformatics analysis from four datasets (GSE14520, GSE76427, ICGC, and TCGA) about HCC 33 common genes among them. Last, the 33 common genes were applied to conduct a PPI network *via* String. The top 10 key genes were chosen utilizing 12 different methods by the cytoHubba plug-in of Cystoscope *via* exploring the 12 outcomes, so we chose our hub gene, *NAP1L1*.

NAP1L1 is a critical member of the Nucleosome Assembly Protein (NAP) family and has been widely portrayed in various tumor progressions (Aydin et al., 2020; Chen et al., 2021; Liu X. et al., 2021). The high expression of *NAP1L1* has existed in the heap of human neoplasms covering nasopharyngeal carcinoma (Liu S. et al., 2021), lung adenocarcinoma (Nagashio et al., 2020), glioblastoma (Zottel et al., 2020), colon cancer (Aydin et al., 2020), colorectal cancer (Queiroz et al., 2020), breast cancer (Liu S. et al., 2021), small-intestinal carcinoid (Kidd et al., 2006), and pancreatic neuroendocrine neoplasm (Schimmack et al., 2014). In addition, previous studies have confirmed the high expression was relevant to the worse survival in nasopharyngeal carcinoma (Liu X. et al., 2021), lung adenocarcinoma (Nagashio et al., 2020), colon cancer (Aydin et al., 2020), and breast cancer (Liu S. et al., 2021). However, we only found seldom studies of HCC. First, the high expression of *lncRNA CDKN2B-AS1* could accelerate the development of HCC regulating the *PI3K/AKT/mTOR* signaling pathway in which *NAP1L1* was a target of *lncRNA CDKN2B-AS1* (Huang et al., 2018). In addition, the *PI3K/AKT/mTOR* signaling pathway was inhibited in *PRDM8* overexpression by the regulation of *NAP1L1* in HCC, which made an antitumor effect (Chen Z. et al., 2018). The upregulation of *NAP1L1* can recruit *HDGF/c-Jun*, boosting the HCC progression (Zhang et al., 2021). Other scholars claimed that the proliferation of HCC cells and chemotherapy resistance could be enhanced after the high expression of *NAP1L1* (Le et al., 2019). The aforementioned studies showed that *NAP1L1* could promote HCC progression *via* activating the *PI3K/AKT/mTOR* signaling pathway and recruiting *HDGF/c-Jun*. Meanwhile, we observed that *NAP1L1* could promote HCC progression through other mechanisms, so the study series was conducted in the following.

On the one hand, survival analysis, clinical feature analysis, univariate/multivariate regression analysis, and GSEA were used to study *NAP1L1* by collecting HCC patients from GSE14520, GSE76427, ICGC, and TCGA. The Conclusion presented that HCC patients with higher expression levels of *NAP1L1* had a poorer prognosis than those with lower expression levels, and *NAP1L1* is an independent prognostic factor for HCC. Although the relation between the expression of *NAP1L1* and the HCC patients' survival and prognostic had been studied, the HCC progression is vague today. In order to shed light on NAP1L1's molecular mechanism promoting the progression of HCC closely, GSEA was applied in GSE14520, GSE76427, ICGC, and TCGA. The results confirmed that *NAP1L1* could promote HCC progression by regulating the G2/M transition of the cell cycle and the *Wnt* signaling pathway. Thus, the hypothesis of HCC progression influenced by *NAP1L1* may be achieved by regulating the G2/M phase and activating the *Wnt/β-catenin* pathway.

On the other hand, the Western blot and the flow cytometry experiments were utilized to prove our summary. The result of the Western blot showed that *NAP1L1* silencing led to downregulation of *CDK1* and *β-catenin* proteins; the result of flow cytometry was that cell numbers in the G2 phase were significantly increased when *NAP1L1* silenced. Some scholars thought that *CDK1* was gathered in the G2 phase of the cell cycle during the late period and *CDK1* was required to regulate it (Bonelli et al., 2014; Roskoski, 2019). Another study has also found that promoting the degradation of *CDK1* can block the G2/M transition of the cell cycle (Lee et al., 2016). Then, the *Wnt/β-catenin* pathway is mainly inactive in a healthy liver. However, it is frequently activated and promotes tumor growth in HCC. It strictly gets command of the embryogenesis,

covering hepatobiliary development, maturation, and zonation (Perugorria et al., 2019). Many previous studies have confirmed that *Wnt/β-catenin* pathway activation could accelerate the progression of HCC (Huang et al., 2020; Tian et al., 2020). Our previous experimental results showed that downregulation of CDK1, β-catenin proteins, and cell numbers in the G2 phase significantly increased when NAP1L1 silences. Therefore, we speculate that upregulation of *NAP1L1* may boost the HCC progression by promoting *Wnt/β-catenin* signaling pathway activation and G2/M phase transition of the cell cycle.

The previous studies have also thought that activating the *Wnt/β-catenin* signaling pathway can promote M2 polarization in macrophages (Feng et al., 2018; Tian et al., 2020). Therefore, ground on GSE14520 and GSE76427 datasets, ICGC and TCGA databases, and TISIDB and TIMER databases, the correlation between NAP1L1 and immune cells was analyzed by applying HCC patients. To sum up, *NAP1L1* expression related to immune cells and *NAP1L1* may also accelerate the HCC progression by affecting macrophages through some underlying mechanisms. Nonetheless, the results are based on the bioinformatics analysis, which needs more exploration and research in the following study. One study has found that the TAM was the immunosuppressive cell, functioned as a kind of suppression antitumor immunity, and promoted the progression of the tumor via expressing the cytokines and chemokines. The TAM is a crucial component in promoting HCC growth and invasion of the tumor microenvironment (Lu et al., 2019). The other study has shown the inhibition of the TAM infiltration. Furthermore, the macrophage M2 polarization could reverse the immunosuppressive state of the tumor microenvironment and activate the responses of the antitumor CD8+ T cells (Li X. et al., 2017). The activation of the *Wnt/β-catenin* pathway could accelerate the macrophage M2 polarization (Yang et al., 2018), so inhibiting the *Wnt/β-catenin* pathway may be a potential strategy for treating HCC patients. We confirmed that high expression of *NAP1L1* can stimulate the *Wnt/β-catenin* pathway to promote HCC progression. Therefore, the study of *NAP1L1* provides some novel thoughts and perspectives for the precise treatment of HCC.

According to the analysis of multiple databases about *NAP1L1*, the Conclusion showed that HCC patients with higher expression levels of NAP1L1 had a poorer prognosis than those with lower expression levels, and *NAP1L1* is an independent prognostic factor for HCC. The study further found that the hypothesis of HCC progression influenced by *NAP1L1* may be achieved by regulating the G2/M phase and activating the *Wnt/β-catenin* signaling pathway. What is more, *NAP1L1* expression related to immune cells and *NAP1L1* also may accelerate the HCC progression by affecting macrophages through some underlying mechanisms. Nonetheless, the results are based on the bioinformatics analysis, which needs more exploration and research in the subsequent study.

5 CONCLUSION

Through the comprehensive analysis of four datasets, we figured out the hub gene (*NAP1L1*) in which the higher expression of *NAP1L1* was connected with HCC patients' shorter survival time and poorer prognosis. *NAP1L1* is an independent prognostic factor in HCC. In addition, the analysis of multiple databases and experiments were carried out to explore the *NAP1L1*-relevant mechanism. The result of GSEA showed that *NAP1L1* might accelerate the progression of HCC, which relied on the cell cycle, especially the G2/M phase transition and Wnt signaling pathway. The result of the Western blot showed that *NAP1L1* silencing led to downregulation of *CDK1* and *β-catenin* proteins; the result of flow cytometry showed that cell numbers in the G2 phase were significantly increased when *NAP1L1* was silenced. Thus, we claimed that *NAP1L1* might promote HCC progression by activating the *Wnt* signaling pathway and promoting cell cycle G2/M transition. In the end, the correlation between *NAP1L1* and immune cells was analyzed by applying HCC patients. Furthermore, we also found that *NAP1L1* expression related to immune cells and *NAP1L1* may accelerate HCC progression by affecting macrophages through some underlying mechanisms. Nonetheless, the outcomes in the research are just based on the bioinformatics analysis, which needs more exploration and research in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

BS and WZ designed the study. WZ and XL collected the literature. BS performed statistical analyses and wrote the manuscript. BS, WZ, XL, and JJ analyzed the data. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 81871965).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.876253/full#supplementary-material

REFERENCES

Aydin, M. A., Gul, G., Kiziltan, R., Algul, S., and Kemik, O. (2020). Nucleosome Assembly Protein 1-Like 1 (NAP1L1) in Colon Cancer Patients: A Potential Biomarker with Diagnostic and Prognostic Utility. *Eur. Rev. Med. Pharmacol. Sci.* 24 (20), 10512–10517. doi:10.26355/eurrev_202010_23403

Bonelli, P., Tuccillo, F. M., Borrelli, A., Schiattarella, A., and Buonaguro, F. M. (2014). CDK/CCN and CDKI Alterations for Cancer Prognosis and Therapeutic Predictivity. *Biomed. Res. Int.* 2014, 361020. doi:10.1155/2014/361020

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, Z., Xie, H., Hu, M., Huang, T., Hu, Y., Sang, N., et al. (2020). Recent Progress in Treatment of Hepatocellular Carcinoma. *Am. J. Cancer Res.* 10 (9), 2993–3036.

Chen, Z., Gao, W., Pu, L., Zhang, L., Han, G., Zuo, X., et al. (2018). PRDM8 Exhibits Antitumor Activities toward Hepatocellular Carcinoma by Targeting NAP1L1. *Hepatology* 68 (3), 994–1009. doi:10.1002/hep.29890

Chen, Z., Xie, Y., Luo, H., Song, Y., Que, T., Hu, R., et al. (2021). NAP1L1 Promotes Proliferation and Chemoresistance in Glioma by Inducing CCND1/CDK4/CDK6 Expression through its Interaction with HDGF and Activation of C-Jun. *Aging* 13 (24), 26180–26200. doi:10.18632/aging.203805

Feng, Y., Ren, J., Gui, Y., Wei, W., Shu, B., Lu, Q., et al. (2018). Wnt/β-Catenin-Promoted Macrophage Alternative Activation Contributes to Kidney Fibrosis. *J. Am. Soc. Nephrol.* 29 (1), 182–193. doi:10.1681/asn.2017040391

Grinchuk, O. V., Yenamandra, S. P., Iyer, R., Singh, M., Lee, H. K., Lim, K. H., et al. (2018). Tumor-Adjacent Tissue Co-Expression Profile Analysis Reveals Pro-Oncogenic Ribosomal Gene Signature for Prognosis of Resectable Hepatocellular Carcinoma. *Mol. Oncol.* 12 (1), 89–113. doi:10.1002/1878-0261.12153

Huang, G., Liang, M., Liu, H., Huang, J., Li, P., Wang, C., et al. (2020). CircRNA hsa_circRNA_104348 Promotes Hepatocellular Carcinoma Progression through Modulating miR-187-3p/RTKN2 axis and Activating Wnt/β-Catenin Pathway. *Cell Death Dis* 11 (12), 1065. doi:10.1038/s41419-020-03276-1

Huang, Y., Xiang, B., Liu, Y., Wang, Y., and Kan, H. (2018). LncRNA CDKN2B-AS1 Promotes Tumor Growth and Metastasis of Human Hepatocellular Carcinoma by Targeting let-7c-5p/NAP1L1 Axis. *Cancer Lett.* 437, 56–66. doi:10.1016/j.canlet.2018.08.024

Kidd, M., Modlin, I. M., Mane, S. M., Camp, R. L., Eick, G., and Latich, I. (2006). The Role of Genetic Markers--NAP1L1, MAGE-D2, and MTA1-Iin Defining Small-Intestinal Carcinoid Neoplasia. *Ann. Surg. Oncol.* 13 (2), 253–262. doi:10.1245/aso.2006.12.011

Le, Y., Kan, A., Li, Q.-J., He, M.-K., Chen, H.-L., and Shi, M. (2019). NAP1L1 Is a Prognostic Biomarker and Contribute to Doxorubicin Chemotherapy Resistance in Human Hepatocellular Carcinoma. *Cancer Cel Int* 19, 228. doi:10.1186/s12935-019-0949-0

Lee, M. H., Cho, Y., Kim, D. H., Woo, H. J., Yang, J. Y., Kwon, H. J., et al. (2016). Menadione Induces G2/M Arrest in Gastric Cancer Cells by Down-Regulation of CDC25C and Proteasome Mediated Degradation of CDK1 and Cyclin B1. *Am. J. Transl Res.* 8 (12), 5246–5255.

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.Can-17-0307

Li, X., Yao, W., Yuan, Y., Chen, P., Li, B., Li, J., et al. (2017). Targeting of Tumour-Infiltrating Macrophages via CCL2/CCR2 Signalling as a Therapeutic Strategy against Hepatocellular Carcinoma. *Gut* 66 (1), 157–167. doi:10.1136/gutjnl-2015-310514

Liu, S., Zhang, Y., Cui, S., Song, D., Li, B., Chen, Q., et al. (2021). NAP1L1 Interacts with Hepatoma-Derived Growth Factor to Recruit C-Jun Inducing Breast Cancer Growth. *Cancer Cel Int* 21 (1), 605. doi:10.1186/s12935-021-02301-3

Liu, Y., Li, X., Zhang, Y., Tang, Y., Fang, W., Liu, X., et al. (2021). NAP1L1 Targeting Suppresses the Proliferation of Nasopharyngeal Carcinoma. *Biomed. Pharmacother.* 143, 112096. doi:10.1016/j.biopha.2021.112096

Lu, C., Rong, D., Zhang, B., Zheng, W., Wang, X., Chen, Z., et al. (2019). Current Perspectives on the Immunosuppressive Tumor Microenvironment in Hepatocellular Carcinoma: Challenges and Opportunities. *Mol. Cancer* 18 (1), 130. doi:10.1186/s12943-019-1047-6

Nagashio, R., Kuchitsu, Y., Igawa, S., Kusuhara, S., Naoki, K., Satoh, Y., et al. (2020). Prognostic Significance of NAP1L1 Expression in Patients with Early Lung Adenocarcinoma. *Biomed. Res.* 41 (3), 149–159. doi:10.2220/biomedres.41.149

Perugorria, M. J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J. J. G., et al. (2019). Wnt-β-Catenin Signalling in Liver Development, Health and Disease. *Nat. Rev. Gastroenterol. Hepatol.* 16 (2), 121–136. doi:10.1038/s41575-018-0075-9

Pontén, F., Schwenk, J. M., Asplund, A., and Edqvist, P.-H. D. (2011). The Human Protein Atlas as a Proteomic Resource for Biomarker Discovery. *J. Intern. Med.* 270 (5), 428–446. doi:10.1111/j.1365-2796.2011.02427.x

Queiroz, C. J. S., Song, F., Reed, K. R., Al-Khafaji, N., Clarke, A. R., Vimalachandran, D., et al. (2020). NAP1L1: A Novel Human Colorectal Cancer Biomarker Derived from Animal Models of Apc Inactivation. *Front. Oncol.* 10, 1565. doi:10.3389/fonc.2020.01565

Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B. B., et al. (2007). Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles. *Neoplasia* 9 (2), 166–180. doi:10.1593/neo.07112

Roessler, S., Jia, H.-L., Budhu, A., Forgues, M., Ye, Q.-H., Lee, J.-S., et al. (2010). A Unique Metastasis Gene Signature Enables Prediction of Tumor Relapse in Early-Stage Hepatocellular Carcinoma Patients. *Cancer Res.* 70 (24), 10202–10212. doi:10.1158/0008-5472.Can-10-2607

Roskoski, R., Jr. (2019). Cyclin-Dependent Protein Serine/Threonine Kinase Inhibitors as Anticancer Drugs. *Pharmacol. Res.* 139, 471–488. doi:10.1016/j.phrs.2018.11.035

Ru, B., Wong, C. N., Tong, Y., Zhong, J. Y., Zhong, S. S. W., Wu, W. C., et al. (2019). TISIDB: An Integrated Repository Portal for Tumor-Immune System Interactions. *Bioinformatics* 35 (20), 4200–4202. doi:10.1093/bioinformatics/btz210

Schimmack, S., Taylor, A., Lawrence, B., Alaimo, D., Schmitz-Winnenthal, H., Büchler, M. W., et al. (2014). A Mechanistic Role for the Chromatin Modulator, NAP1L1, in Pancreatic Neuroendocrine Neoplasm Proliferation and Metastases. *Epigenetics Chromatin* 7, 15. doi:10.1186/1756-8935-7-15

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* 43 (Database issue), D447–D452. doi:10.1093/nar/gku1003

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., et al. (2017). A Subcellular Map of the Human Proteome. *Science* 356 (6340), eaal3321. doi:10.1126/science.aal3321

Tian, X., Wu, Y., Yang, Y., Wang, J., Niu, M., Gao, S., et al. (2020). Long Noncoding RNA LINC00662 Promotes M2 Macrophage Polarization and Hepatocellular Carcinoma Progression via Activating Wnt/β-Catenin Signaling. *Mol. Oncol.* 14 (2), 462–483. doi:10.1002/1878-0261.12606

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-Based Map of the Human Proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A Pathology Atlas of the Human Cancer Transcriptome. *Science* 357 (6352), eaan2507. doi:10.1126/science.aan2507

Yang, Y., Ye, Y.-C., Chen, Y., Zhao, J.-L., Gao, C.-C., Han, H., et al. (2018). Crosstalk between Hepatic Tumor Cells and Macrophages via Wnt/β-Catenin Signaling Promotes M2-Like Macrophage Polarization and Reinforces Tumor Malignant Behaviors. *Cel. Death Dis.* 9 (8), 793. doi:10.1038/s41419-018-0818-0

Zhang, Y.-W., Chen, Q., Li, B., Li, H.-Y., Zhao, X.-K., Xiao, Y.-y., et al. (2021). NAP1L1 Functions as a Tumor Promoter via Recruiting Hepatoma-Derived Growth Factor/c-Jun Signal in Hepatocellular Carcinoma. *Front. Cel. Dev. Biol.* 9, 659680. doi:10.3389/fcell.2021.659680

Zottel, A., Jovčevska, I., Šamec, N., Mlakar, J., Šribar, J., Križaj, I., et al. (2020). Anti-Vimentin, Anti-TUFM, Anti-NAP1L1 and Anti-DPYSL2 Nanobodies Display Cytotoxic Effect and Reduce Glioblastoma Cell Migration. *Ther. Adv. Med. Oncol.* 12, 1758835920915302. doi:10.1177/1758835920915302

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shen, Zhu, Liu and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fgene.2022.836509

Biological and Clinicopathological Characteristics of OPN in Cervical Cancers

Shuhang Qin[1], Li Yi[2], Yanchun Liang[1], Yili Chen[1], Wei Wang[1], Yuandong Liao[1], Chunyu Zhang[1], Hua Huang[1], Jiaming Huang[1] and Shuzhong Yao[1]**

[1]Department of Obstetrics and Gynecology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Department of Neurosurgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Decheng Yin,
University of Erlangen Nuremberg, Germany
Paramita Mandal,
University of Burdwan, India

***Correspondence:**
Jiaming Huang
huangjm29@mail.sysu.edu.cn
Shuzhong Yao
yaoshuzh@mail.sysu.edu.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 15 December 2021
Accepted: 08 April 2022
Published: 20 May 2022

Citation:
Qin S, Yi L, Liang Y, Chen Y, Wang W, Liao Y, Zhang C, Huang H, Huang J and Yao S (2022) Biological and Clinicopathological Characteristics of OPN in Cervical Cancers. Front. Genet. 13:836509. doi: 10.3389/fgene.2022.836509

Background: Cervical cancer (CC) is the most common gynecological malignancy. Recently, an increasing number of studies have indicated that osteopontin (OPN) is a promising diagnostic and prognostic biomarker for CC. However, the biological role and detailed mechanism of OPN in CC remain unclear.

Methods: The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets and a clinical sample microarray were used in our study. To identify the clinicopathological characteristics of OPN in CC, we compared the expression of OPN between normal and CC tissue samples and analyzed the correlations between OPN expression and multiple clinicopathological features. To identify biological processes involving OPN, OPN-associated genes were screened with Pearson correlation analysis and applied in hallmark gene set enrichment analysis (GSEA). Additionally, leukocyte infiltration was evaluated based on OPN expression. Finally, OPN-related signaling pathways were identified by GSEA.

Results: OPN expression was higher in CC samples than in normal tissue samples and positively correlated with age, FIGO stage, tumor size, lymphovascular invasion and an unfavorable prognosis. OPN-associated genes were mainly enriched in the immune response, and increased OPN expression was accompanied by increased M2 macrophage infiltration. Additionally, OPN was correlated with hypoxia, high glycolytic metabolism, apoptosis, angiogenesis, epithelial-mesenchymal transition and multiple signaling pathways (the p53 pathway, the PI3K/Akt pathway, IL6/STAT3 signaling, mTORC1 signaling and KRAS signaling).

Conclusion: Our study showed that OPN is involved in immunological activities and multiple tumor processes, identifying it as a potential therapeutic target and useful prognostic factor in CC patients.

Keywords: osteopontin, cervical cancer, immunosuppression, biological process, biomarker

INTRODUCTION

As the most common gynecological malignancy, uterine cervical cancer (CC) is the fourth most frequent female cancer worldwide. According to global cancer statistics, there were approximately 570,000 new CC cases and 311,000 deaths worldwide in 2018 (Bray et al., 2018). Although major achievements have been made in surgery, chemotherapy, and radiotherapy in recent decades, these therapeutic approaches can hardly prevent metastasis and recurrence in CC patients. As a new therapeutic strategy, molecular targeted therapy is urgently required. It is important to identify key prognostic factors and predictive biomarkers to improve the effectiveness of treatment and develop precise treatment strategies.

Uterine CC is primarily caused by infection with high-risk human papillomavirus (HPV). Moreover, the tumor microenvironment plays a significant role in the tumorigenesis of CC. Numerous studies have identified immunity- or metabolism-related gene signatures as prognostic biomarkers in a wide spectrum of human malignancies (Zhang et al., 2003; Galon et al., 2006; Mahmoud et al., 2011). Osteopontin (OPN), known previously as secreted phosphoprotein 1 (SPP1), is a small integrin-binding ligand N-linked glycoprotein that binds to cell-surface receptors, including integrins and CD44 (Oldberg et al., 1986; Thoms et al., 2012). OPN is commonly found in some mineralized tissues as an extracellular matrix component, and it also behaves as a secreted protein in body fluids, such as milk, blood, urine, saliva, seminal fluid and bile. As a result of its diverse distribution, OPN is involved in multiple characteristics of tumor biology, including cell proliferation, survival, angiogenesis, chemoresistance, stem-like properties, tumor invasion, and metastasis (Wei et al., 2017). Furthermore, recent studies have shown that OPN plays important roles in antitumor immunity and metabolism (Wang and Denhardt, 2008; Hsieh et al., 2014; Shi et al., 2014; Moorman et al., 2020). As an immunological and metabolic biomarker, OPN might participate in the progression of CC.

In previous reports, OPN was found to act as an oncogene during tumorigenesis of CC (Cho et al., 2008; Bao et al., 2015; Xu et al., 2019). High expression of OPN was correlated with a higher FIGO stage, larger tumor size, lymph node metastasis, HPV infection, cisplatin resistance and shorter survival (Bao et al., 2015; Huang et al., 2015; Xu et al., 2019). Nonetheless, the immunological and metabolic features of OPN in CC have not been sufficiently investigated, and clinical characteristic analysis using gene datasets from multiple databases has never been performed. Given that OPN could be a promising target in CC, comprehensive reports of the relationships between OPN gene expression and clinical outcomes or biological features in CC are still required.

In this study, we collected a dataset from The Cancer Genome Atlas (TCGA), eight Gene Expression Omnibus (GEO) datasets and our clinical sample microarray to explore the potential role of OPN in CC. First, we compared OPN gene expression between CC and normal tissues. Second, we analyzed the correlations between OPN expression and clinical characteristics. Finally, we explored the potential biological processes and signaling pathways involving OPN in the progression of CC. This is the first integrative bioinformatic analysis to characterize OPN expression in CC molecularly and clinically and provide novel insight to improve the comprehensive understanding of the molecular mechanism of OPN in CC development.

MATERIALS AND METHODS

Datasets

Normalized TCGA gene expression data and clinicopathological data (304 CC patients) were downloaded from the University of California Santa Cruz (UCSC) Xena project (https://xenabrowser.net/datapages/) (**Supplementary Table S1**). According to the UCSC pipeline, gene expression data is re-computed by log2 (fpkm-uq+1) and the fpkm-uq value is a modified version of the fpkm formula (normalized expression value). The comparison of OPN expression between CC and nontumor tissues was performed by Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/). The microarray datasets (GSE39001, GSE9750, GSE7803, GSE63514, GSE27678, GSE67522, GSE46857 and GSE44001) were downloaded from GEO series_matrix data (https://www.ncbi.nlm.nih.gov/geo/) and were normalized in the R statistical environment using the affy Bioconductor library NormalizeBetweenArrays and Log2 transformation. (**Table 1**, **Supplementary Table S2**).

Clinical Sample Microarray Sequencing

Microarray data for five CC samples and six normal cervical tissue samples from the Department of Gynecology, First Affiliated Hospital of Sun Yat-sen University were used to compare the expression of OPN between CC and nontumor tissues. Sample collection, RNA sequencing and data normalization have been described previously (Huang et al., 2018). Written informed consent was obtained from each patient. All specimens were handled according to legal and ethical standards.

OPN-associated gene identification and Hallmark gene set enrichment analysis (GSEA)

The Pearson correlation coefficients for OPN with all other protein-coding genes were calculated in RStudio Version 1.2.1335 (R Version 3.6.3) with the cor. test algorithm. Genes with an R value >0.2 and a p value < 0.05 were defined as the threshold for OPN-associated genes in each dataset. All the OPN-associated genes were introduced into the Metascape website (http://metascape.org/gp/index.html) for Hallmark GSEA. The hallmark gene set V7.2 from the Broad Institute (http://software.broadinstitute.org/gsea/msigdb/collections.jsp#H) contains specific well-defined biological states and processes and displays coherent expression (Liberzon et al., 2015).

The correlations of OPN with immune checkpoint molecules, hypoxia-related genes and glycolytic enzymes were visualized with Circos plots produced with R software using the "circlize" package. The correlations of OPN with apoptosis, angiogenesis and epithelial-mesenchymal transition (EMT) were visualized with a correlogram using the "corrplot" package. Hypoxia-related genes and biomarkers of apoptosis and angiogenesis were selected from the Hallmark gene set.

TABLE 1 | Clinical properties of the CC patients used in the GEO datasets analysis.

GEO ID	Contributors	GEO Platform	No. of Normal Samples in Dataset	No. of Tumor Samples in Dataset	No. of Squamous Intraepithelial Lesion (HSIL/LSIL)[a]	No. of Cervical Squamous Intraepithelial Lesion (CIN1/CIN2/CIN3)[b]
GSE39001	Espinosa AM et al., 2013	GPL201	12	43	—	—
GSE9750	Murty VV et al., 2008	GPL96	24	33	—	—
GSE7803	Zhai Y et al., 2007	GPL96	10	21	HSIL: 7	—
GSE63514	Den Boon J et al., 2015	GPL570	24	28	—	CIN1: 14; CIN2: 22; CIN3: 40
GSE27678	Karagavriilidou K et al., 2013	GPL571	12	—	HSIL: 21; LSIL: 11	—
GSE67522	Sharma S et al., 2015	GPL10558	22	20	—	—
GSE46857	Mulherkar R et al., 2013	GPL7025	4	25	—	—
GSE44001	Lee YY et al., 2013	GPL14951	—	300	—	—

[a]*HSIL: high-grade squamous intraepithelial lesion; LSIL: low-grade squamous intraepithelial lesion.*
[b]*CIN: cervical squamous intraepithelial lesion.*

Tumor Purity and Leukocyte Infiltration

The ESTIMATE tool (https://bioinformatics.mdanderson.org/estimate/disease.html) was used to analyze tumor purity. The CIBERSORT tool (https://cibersort.stanford.edu/) was used to evaluate leukocyte infiltration. The correlations of OPN with tumor purity and leukocyte infiltration were calculated with Pearson correlation analysis and are shown in heatmaps produced by MORPHEUS (https://software.broadinstitute.org/morpheus/). The samples were ordered according to the expression of OPN.

Pathway Gene Signatures Analyzed Using GSEA

GSEA was used to compare dysfunctional pathways between CC samples with high OPN expression and those with low expression in the TCGA. GSEA software (version: 4.0.3) and the Hallmark gene set V7.2 were applied. The upregulated pathways were defined by a normalized enrichment score (NES) > 0 and are listed in **Supplementary Table S5**. Pathways with a normalized p value <0.05 were considered significantly enriched.

Statistical Analyses

An unpaired t test was used to compare OPN expression between two groups in TCGA, GEO datasets and clinical microarray data. An unpaired t test was used to compare OPN expression between two groups defined by a clinicopathological feature. The median value of a gene of interest was applied as the threshold for the low and high expression groups of CC patients. The log-rank test and Kaplan-Meier survival curves were used to describe survival differences between two groups. All these analyses were conducted with GraphPad Prism 8.0.1.

RESULTS

OPN Was Upregulated in CC and Correlated With Multiple Clinicopathological Parameters

According to GEPIA, higher expression of OPN was observed in CC samples than in nontumor samples ($p < 0.05$) (**Figure 1A**). For further validation, we investigated the difference in OPN expression between normal tissue and CC samples based on the GSE39001, GSE9750, GSE7803, GSE63514, GSE27678, GSE67522, and GSE46857 datasets and our clinical sample microarray (**Figure 1B,C**). All datasets revealed that OPN was upregulated in the cancerous tissue groups.

Furthermore, OPN expression was analyzed according to clinical parameters with data from the TCGA (**Figure 2A**, **Supplementary Figure S1A**). Higher expression of OPN was found in patients who were over 50 years old or who exhibited an advanced FIGO stage (stage III and IV), cervical squamous cell carcinoma, or lymphovascular invasion ($p < 0.05$). However, there were no significant correlations between OPN expression and body mass index (BMI), HPV status, neoplasm histologic grade, lymph node metastasis or corpus uteri involvement (**Supplementary Figure S1A**).

In addition, GSE44001 was analyzed (**Figure 2A**). OPN was more highly expressed in samples with an advanced FIGO stage (stage III and IV) or a larger tumor size (tumor diameter over 4 cm).

These results suggest that higher OPN expression indicates more advanced malignancy in CC.

High expression of OPN was associated with poor survival and therapeutic resistance in CC

Survival analyses were conducted for the whole group or subgroups based on data from the TCGA (**Figure 2B**, **Supplementary Figure S1B**). Patients with CC exhibiting higher OPN expression had significantly shorter survival times than their counterparts in the overall cohort (hazard ratio (HR) = 1.748, $p = 0.020$). In addition, we plotted survival curves for cervical squamous cell carcinoma patients. The overall survival time was shorter in the cervical squamous cell carcinoma patients with high OPN expression than those with low expression (HR = 2.047, $p = 0.007$). There was no significant difference in the overall survival time according to OPN expression in the early-stage (FIGO IA1-IIA2), advanced-stage (FIGO IIB-IVB) cohort or adenocarcinoma patients (**Supplementary Figure S1B**). In addition, the patients with higher OPN expression had

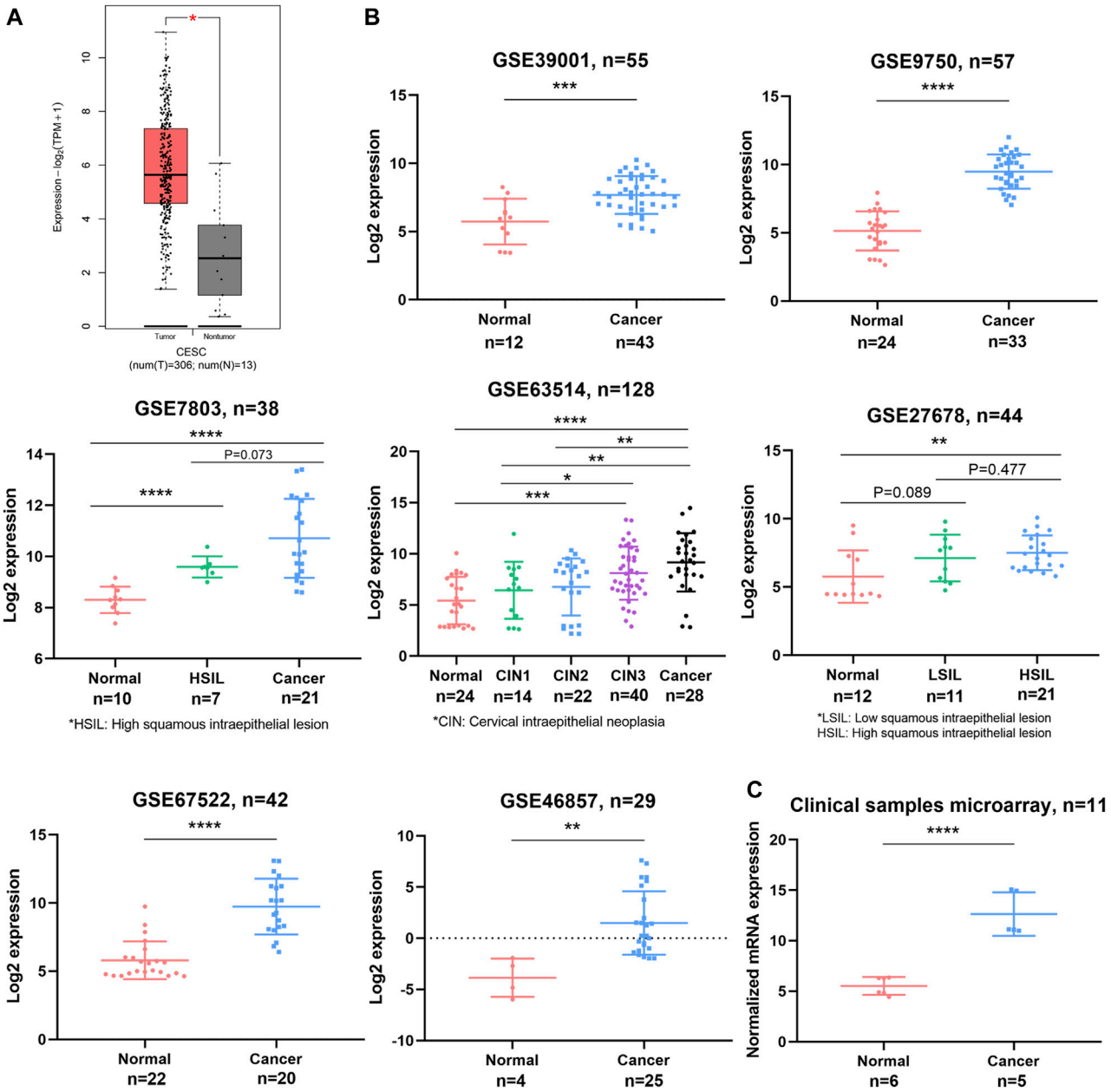


FIGURE 1 | OPN expression was upregulated in CC tissues compared with normal tissues. Comparisons were conducted using data from the TCGA and GTEX databases **(A)**, GEO datasets **(B)**, and a clinical sample microarray **(C)**. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

significantly shorter disease-free survival times than those with lower expression in the GSE44001 dataset (HR = 2.142, $p = 0.025$) (**Figure 2B**).

Moreover, survival analyses of TCGA subgroups defined by treatment were used to evaluate the effectiveness of well-accepted treatments. The CC patients with lower OPN expression showed better responses to chemotherapy and radiotherapy than those with higher expression (Chemotherapy: HR = 1.972, $p = 0.043$, Radiotherapy: HR = 1.897, $p = 0.029$) (**Figure 2B**), while there were no differences in patients who underwent radical hysterectomy or received targeted therapy (**Supplementary Figure S1B**). These results indicate that OPN may contribute to therapeutic resistance to chemotherapy and radiotherapy.

High OPN expression was accompanied by an increased immunosuppressive status in CC

The biological function of OPN in CC has not been fully investigated. Therefore, we aimed to identify the possible

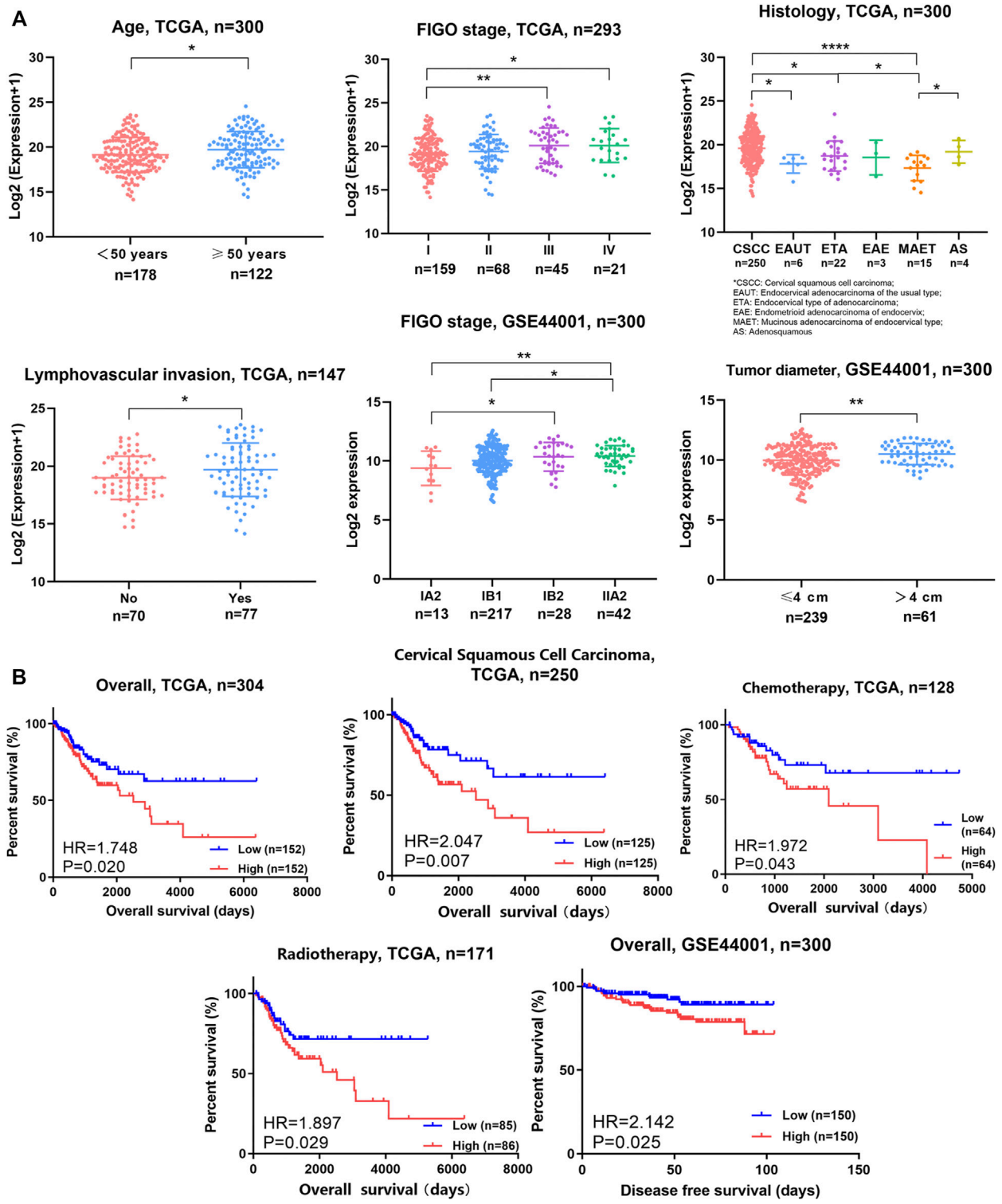


FIGURE 2 | OPN expression was correlated with multiple clinicopathological features, unfavorable survival, and therapeutic resistance in CC **(A)** OPN expression was correlated with clinicopathological features in TCGA data and GSE44001 **(B)** High expression of OPN predicted unfavorable survival and therapeutic resistance in TCGA data and GSE44001. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

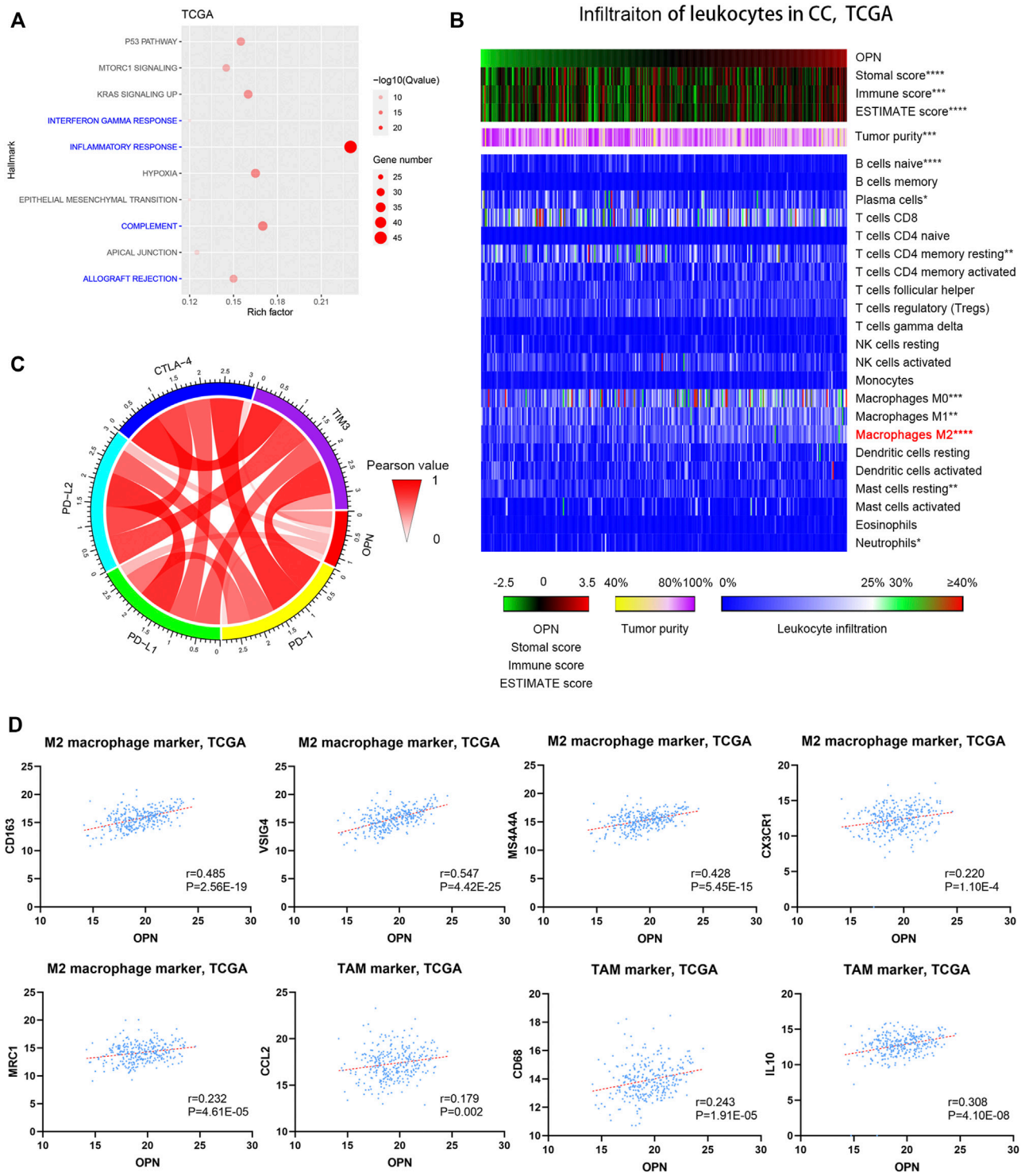


FIGURE 3 | High OPN expression was accompanied by an increased immunosuppressive status in CC data from the TCGA **(A)** Hallmark enrichment analyses showed that OPN-associated genes were mainly enriched in biological processes of the immune response **(B)** The correlations between OPN expression and the infiltration of different kinds of leukocytes were calculated with Pearson correlation analysis and are shown in the heatmap. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$ **(C)** Circos plots show the correlations between OPN and five immune checkpoint molecules **(D)** Pearson correlation analysis showed that OPN was positively associated with M2 macrophage markers and TAM markers.

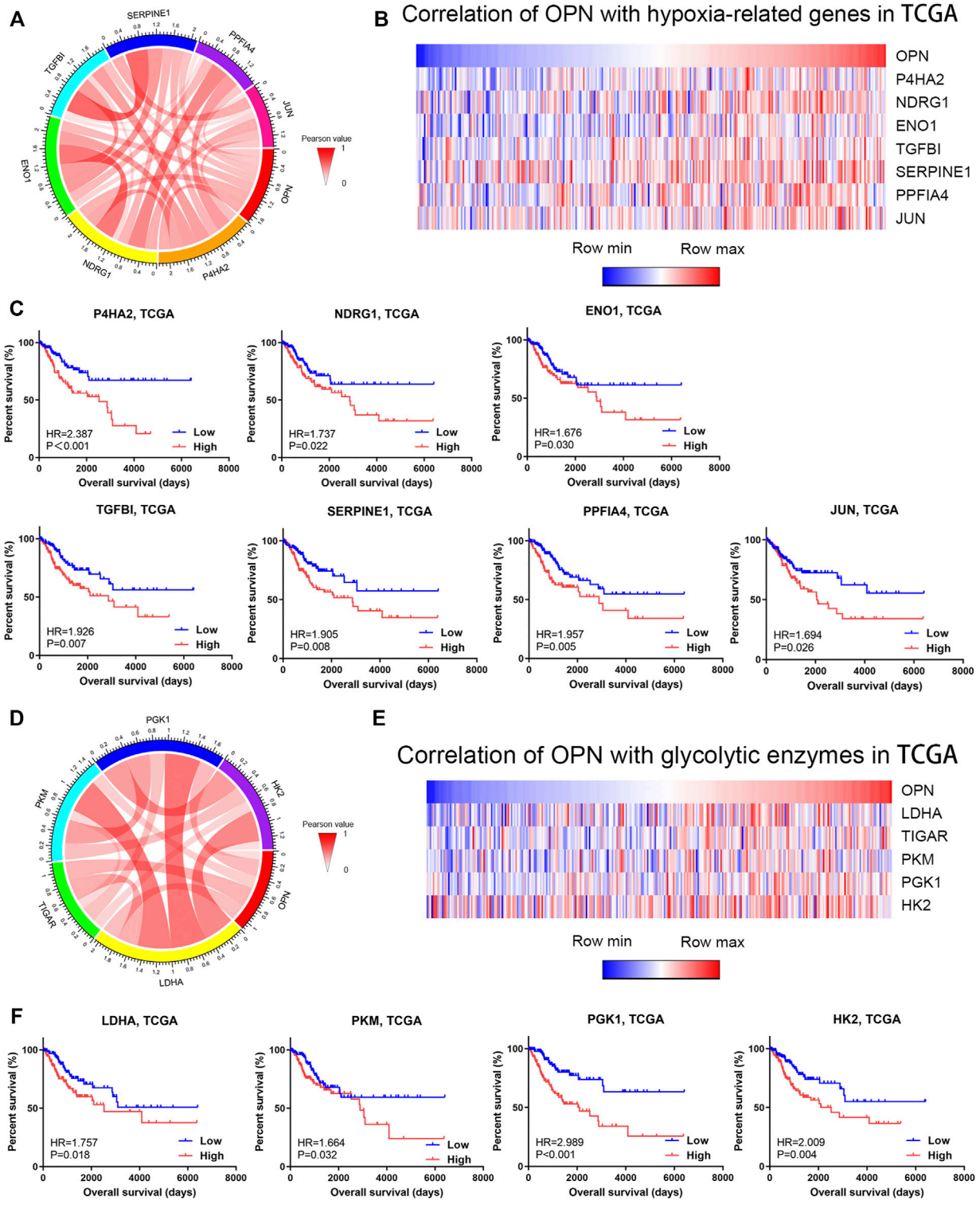


FIGURE 4 | OPN was correlated with hypoxia and high glycolytic levels in CC data from the TCGA **(A,B)** Correlations of OPN with hypoxia-related genes **(C)** Survival plot for OPN-correlated hypoxia-related genes in CC patients **(D,E)** Correlations of OPN with glycolytic enzymes **(F)** Survival plot for OPN-correlated glycolytic enzymes in CC patients.

biological function of OPN through analysis of the biological functions of OPN-associated genes. All the OPN-associated genes from the TCGA are listed in **Supplementary Table S3** and were evaluated by Hallmark GSEA (**Supplementary Table S4**). The top 10 Hallmark terms identified by the Hallmark GSEA are listed in **Figure 3A**; the results indicated that OPN-associated genes were mainly enriched in biological processes of the immune response, including the inflammatory response, IFN-γ response, complement and allograft rejection (**Figure 3A**).

To evaluate the immunological status of CC, both tumor purity and the infiltration of 22 types of leukocytes were assessed for each sample. The samples are displayed in order of their OPN expression level. The results indicated that the immune score and the stromal score exhibited positive correlations with the OPN expression trends (**Figure 3B**, top panels), while tumor purity showed an inverse correlation with the OPN expression trends (**Figure 3B**, middle panels). Moreover, OPN expression was mostly related to the infiltration of M2 macrophages among the 22 types of leukocytes (**Figure 3B**, bottom panels). Pearson correlation analysis further showed that OPN was positively associated with M2 macrophage markers (CD163, VSIG4, MS4A4A, CX3CR1, and MRC1) and tumor-associated macrophage (TAM) markers (CCL2, CD68, and IL10) (**Figure 3D**).

To further explore the role of OPN in the immune microenvironment of CC, the correlations between OPN and five immune checkpoint molecules (PD-1, PD-L1, PD-L2, CTLA-4, and TIM-3) were analyzed (**Figure 3C**), which showed that OPN positively synergized with checkpoint molecules in CC. The results demonstrated that high OPN expression was accompanied by an increased immunosuppressive status.

Furthermore, the OPN-associated genes identified from GSE39001 and GSE9750 were also enriched in biological processes of the immune response (**Supplementary Figure S2A**). OPN was positively associated with some M2 macrophage markers (CD163, VSIG4, MS4A4A, and MRC1) and a TAM marker (CCL2) (**Supplementary Figure S2B**).

High OPN expression was correlated with hypoxic conditions and high glycolytic levels in CC

Hallmark GSEA of OPN-associated genes based on TCGA data showed that OPN was correlated with hypoxia (**Figure 3A**). As shown in the Circos plot and heatmap, OPN expression was positively correlated with multiple hypoxia biomarkers (P4HA2, NDRG1, ENO1, TGFBI, SERPINE1, PPFIA4 and JUN) (**Figure 4A,B**). Survival analysis was applied to these hypoxia-related genes, which showed that all these genes predicted an unfavorable prognosis in CC (**Figure 4C**). These findings indicated that OPN might be involved in CC progression induced by hypoxia.

To further investigate the metabolic role of OPN in CC, we analyzed the correlations between OPN and several glycolytic enzymes. A Circos plot and heatmap showed that the expression of OPN was positively correlated with that of glycolytic enzyme genes (LDHA, TIGAR, PKM, PGK1 and HK2) (**Figure 4D,E**). This indicated that higher OPN expression was correlated with a higher glycolytic level in CC. Except for TIGAR, the glycolytic enzyme genes predicted an unfavorable prognosis in CC (**Figure 4F**). These findings indicated that OPN might synergize with glycolytic enzymes in CC with a poor outcome.

High OPN expression was correlated with apoptosis, angiogenesis, EMT, and multiple signaling pathways in CC

According to hallmark GSEA results based on data from the TCGA, OPN was correlated with apoptosis, angiogenesis, and EMT in CC (**Supplementary Table S4**). Pearson correlation analysis showed that OPN was positively correlated with various apoptosis markers (CD14, HMOX1, CDKN1A, ETF1, IL1B and SMAD7), angiogenesis markers (LPL, OLR1, THBD, NRP1 and POSTN), and EMT markers (VIM/Vimentin, SNAI1/Snail, SNAI2/Slug, CDH2/N-cadherin, ZEB2 and CLDN1/Claudin-1) (**Figure 5A**, **Supplementary Figure S3**).

Moreover, GSEA of TCGA data showed that OPN was involved in several up-regulated signaling pathways, including the p53 pathway, IL6/STAT3 signaling, mTORC1 signaling, PI3K/Akt signaling and KRAS signaling (**Figure 5B**, **Supplementary Table S5**).

These results suggested that OPN might influence the tumor progression of CC through apoptosis, angiogenesis, EMT and the signaling pathways listed above.

DISCUSSION

OPN is a phosphorylated glycoprotein that bridges cell and matrix inorganic substances and is closely associated with the occurrence, progression and metastasis of malignant tumors (Le et al., 2003; Vergis et al., 2008; Raja et al., 2014; Ostheimer et al., 2017; Cao et al., 2019; Nishio et al., 2021). Recently, certain studies have indicated that OPN is a diagnostic and prognostic biomarker for CC (Cho et al., 2008; Leung et al., 2016). OPN expression was found to be positively related to FIGO stage, tumor size and cisplatin resistance (Cho et al., 2008; Chen et al., 2019). Moreover, high OPN expression was shown to predict poor overall survival and disease-free survival in CC patients (Vordermark et al., 2006; Manavi et al., 2007; Huang et al., 2015). By analyzing TCGA and GEO datasets and our tissue microarray, we found that OPN was more highly expressed in CC samples than in normal tissue samples. Additionally, OPN was significantly correlated with age, FIGO stage, tumor size, histology, lymphovascular invasion and poor outcomes. Moreover, CC patients with lower OPN expression showed better responses to chemotherapy and radiotherapy than those with higher expression in a subgroup survival analysis, which indicated that OPN might contribute to therapeutic resistance. All these results were consistent with previous reports. In conclusion, OPN was identified as a good diagnostic and prognostic biomarker for CC.

Only a few studies have investigated how OPN promotes CC progression. Vinit et al. found that OPN regulated the CD44-

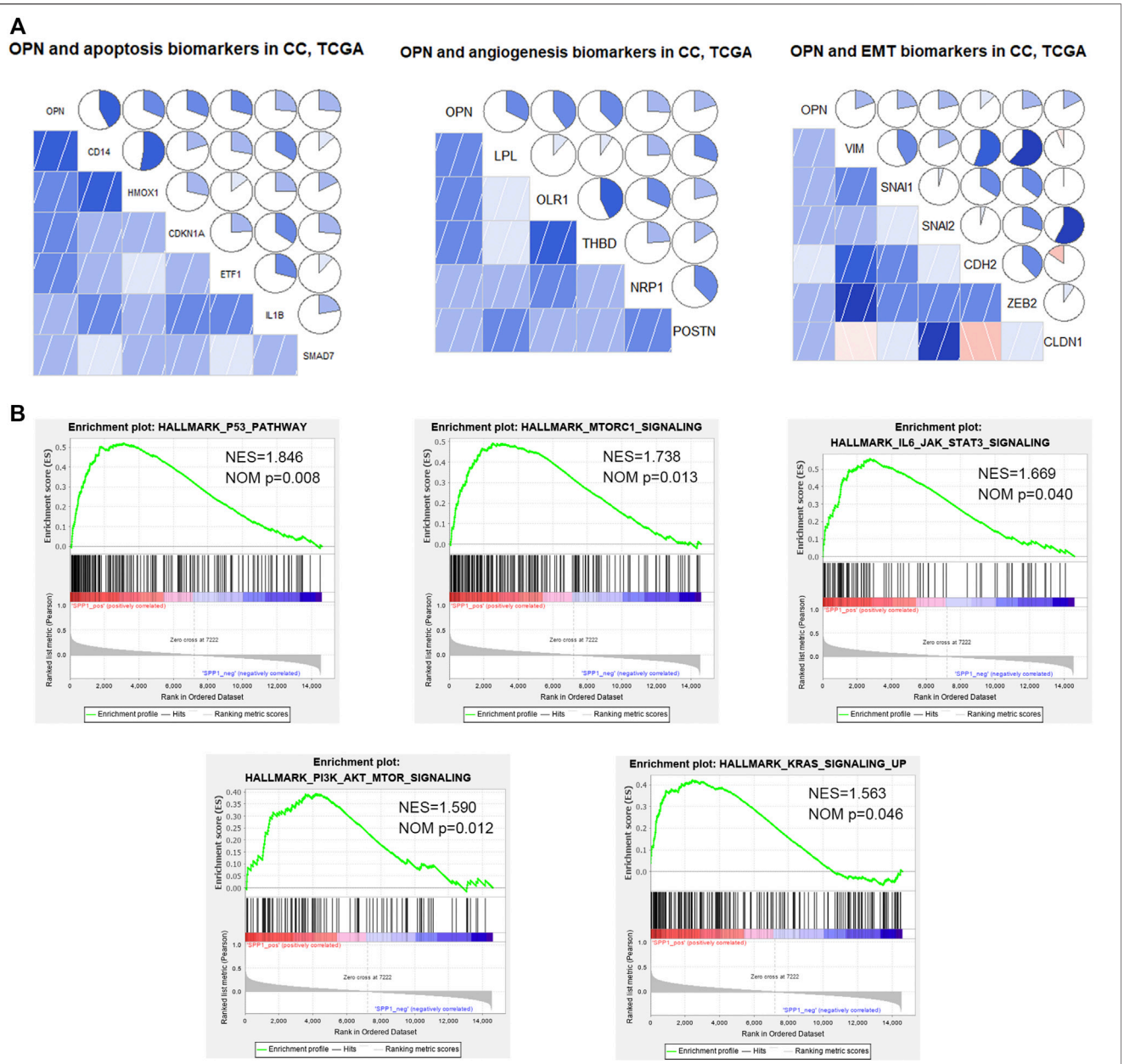


FIGURE 5 | OPN was correlated with apoptosis, angiogenesis, epithelial-mesenchymal transition (EMT), and multiple signaling pathways in CC data from the TCGA **(A)** Relationships between OPN and apoptosis, angiogenesis, or EMT markers **(B)** GSEA showing the up-regulated signaling pathways enriched in OPN-associated genes.

mediated p38 phosphorylation that induces NF-κB activation and the NF-κB–dependent expression of furin, an extracellular protease implicated in HPV processing that enhances CC cell motility (Kumar et al., 2010). Chen et al. found that OPN knockdown resulted in repressed proliferation and enhanced apoptosis in HeLa cells and that downregulation of OPN improved the cisplatin sensitivity of HeLa cells by inhibiting the PI3K/Akt signaling pathway (Vordermark et al., 2006; Manavi et al., 2007; Huang et al., 2015). Xu et al. found that overexpression of miR-181a could inhibit the expression of OPN, induce cell apoptosis, restrain cell proliferation, and reduce cisplatin resistance in CC cells (Vordermark et al., 2006; Manavi et al., 2007; Huang et al., 2015). In our study, we also found that OPN was positively correlated with apoptosis and the PI3K/Akt pathway. In addition, we found that OPN was involved in angiogenesis, EMT and several pathways including the p53 pathway, IL6/STAT3 signaling, mTORC1 signaling and KRAS signaling in CC. The roles of OPN in these pathways have been validated in other cancers (Vordermark et al., 2006; Manavi et al., 2007; Huang et al., 2015).

It is well established that EMT plays important roles in cancer progression, especially in tumor metastasis (Lee et al., 2008; Qureshi et al., 2015; Daugaard et al., 2017). Our previous studies confirmed the close relationship between EMT and CC metastasis (Shang et al., 2018; Liu et al., 2020; Zhang et al., 2020). Emerging reports have shown that OPN regulates the expression of EMT-related transcription factors, including Twist, Snail, Slug and zinc finger E-box-binding homeobox (ZEB), directly or indirectly in various cancers (Kothari et al., 2016). Herein, we showed that OPN was positively correlated with EMT biomarkers in CC, identifying a novel mechanism underlying CC metastasis.

To further investigate the function of OPN in CC, we explored the relationship between OPN and the tumor microenvironment in CC. As a main component of the tumor microenvironment, TAMs play an important role in CC. The polarization of TAMs toward an M2 phenotype is correlated with a poor prognosis, and M2 TAMs can transdifferentiate into lymphatic endothelial cells, inducing lymphangiogenesis and metastasis (Kerjaschki, 2005; Petrillo et al., 2015; Guo et al., 2016). Emerging studies have indicated that OPN induces M2 macrophage polarization, maintains M2 macrophage phenotypes, and acts as a chemoattractant for TAMs (Klement et al., 2021). We analyzed the correlations between OPN and the infiltration of all kinds of leukocytes. OPN expression was mostly related to the infiltration of M2 macrophages and positively related to M2 macrophage markers and TAM markers, indicating that OPN acts as an immune regulator in CC progression. OPN also acts as an immune checkpoint molecule to negatively regulate T cell activation (Kerjaschki, 2005; Petrillo et al., 2015; Guo et al., 2016). We found that OPN synergized well with immune checkpoint molecules (PD-1, PD-L1, PD-L2, CTLA-4, and TIM-3), further validating its immunosuppressive function. As treatments targeting immune checkpoints have shown initial success in CC (Kerjaschki, 2005; Petrillo et al., 2015; Guo et al., 2016), our study highlighted that OPN could be a novel target in immunotherapy.

As hypoxia and intensified glycolytic metabolism are characteristic features of solid tumors, including CC, we investigated the relationships between OPN and hypoxia or glycolysis in CC. Our study suggested that OPN was correlated with a hypoxic microenvironment. Previous reports showed that OPN was upregulated under hypoxic conditions in various cancers (Le et al., 2003; Vergis et al., 2008; Raja et al., 2014; Ostheimer et al., 2017; Cao et al., 2019; Nishio et al., 2021). Raja et al. showed that silencing OPN or its receptor significantly downregulated hypoxia-induced breast cancer cell migration and invasion and that HIF-1α was involved in this process (Le et al., 2003; Vergis et al., 2008; Raja et al., 2014; Ostheimer et al., 2017; Cao et al., 2019; Nishio et al., 2021). Yang et al. showed that hypoxic dendritic cells secreted large amounts of OPN, which were responsible for the enhanced migration of tumor cells (Yang et al., 2009). These findings indicate that OPN could be an important intermediate in hypoxia-driven CC progression.

In addition, our study suggested that high OPN expression was correlated with high glycolytic levels in CC. The relationship between OPN and glycolysis has been investigated previously in some cancers. Hsieh et al. showed that OPN regulated the expression of glucose transporter 1 and glucose transporter 3 in osteosarcoma and enhanced glucose uptake in cells via the integrin αvβ3 (Le et al., 2003; Vergis et al., 2008; Raja et al., 2014; Ostheimer et al., 2017; Cao et al., 2019; Nishio et al., 2021). Lu et al. showed that OPN enhanced hepatocellular carcinoma glycolysis by activating αvβ3-NF-κB signaling (Lu et al., 2020). Shi et al. reported that osteopontin-c, a splice variant of OPN, supported the anchorage independence of invasive breast cancer cells by utilizing glucose to generate energy (Le et al., 2003; Vergis et al., 2008; Raja et al., 2014; Ostheimer et al., 2017; Cao et al., 2019; Nishio et al., 2021). Therefore, OPN might upregulate glycolysis to generate energy to support the survival and expansion of CC cells.

CONCLUSION

Our study is the first study to comprehensively explore the biological and clinicopathological characteristics of OPN in CC by using bioinformatics. We found that the expression of OPN was positively correlated with age, FIGO stage, tumor size, lymphovascular invasion and an unfavorable prognosis. Furthermore, we found that OPN was involved in immune suppression, hypoxia, high glycolytic metabolism, apoptosis, angiogenesis, EMT and multiple signaling pathways (p53 pathway, PI3K/Akt pathway, IL6/STAT3 signaling, mTORC1 signaling and KRAS signaling). Therefore, OPN represents a potential therapeutic target and useful prognostic factor in CC patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

Concept and design: SQ, LY, SY and JH; Development of methodology: SQ, LY and YuL; Acquisition and analysis of data: SQ, YCh, CZ and HH; Writing, review and editing: SQ and YaL; Review and polish: WW and YCh; Supervision: SY and JH; Project administration: SQ, JH and SY; Funding acquisition: SQ, JH and SY.

FUNDING

This study was supported by grants from the Medical Scientific Research Foundation of Guangdong Province (No. A2021030 to SQ), the National Natural and Science Foundation of China (No. 82072874 to SY, No. 81902627 to JH), the Natural and Science

Foundation of Guangdong Province (No. 2018A0303130190 to JH), the Science and Technology Program of Guangzhou (No. 202002020043 to SY) and 5010 Clinical Research Project of Sun Yat-sen University (No. 2018018 to JH).

ACKNOWLEDGMENTS

We thank the Medical Scientific Research Foundation of Guangdong Province, National Natural and Science Foundation of China, Natural and Science Foundation of Guangdong Province and Science and Technology Program of Guangzhou for the grant funding. We also acknowledge the contributions from TCGA and GEO datasets.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.836509/full#supplementary-material

REFERENCES

Bao, L., Si, Q., Jia, L., Ren, X., Ma, R., and Wang, Y. (2015). Detection of Human Papillomavirus and Expression of Osteopontin in Cervical Cancer Specimens. *Mol. Med. Rep.* 11 (1), 447–453. doi:10.3892/mmr.2014.2647

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer J. clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Cao, J., Li, J., Sun, L., Qin, T., Xiao, Y., Chen, K., et al. (2019). Hypoxia-driven Paracrine Osteopontin/integrin αvβ3 Signaling Promotes Pancreatic Cancer Cell Epithelial-Mesenchymal Transition and Cancer Stem Cell-like Properties by Modulating Forkhead Box Protein M1. *Mol. Oncol.* 13 (2), 228–245. doi:10.1002/1878-0261.12399

Chen, X., Xiong, D., Ye, L., Yang, H., Mei, S., Wu, J., et al. (2019). SPP1 Inhibition Improves the Cisplatin Chemo-Sensitivity of Cervical Cancer Cell Lines. *Cancer Chemother. Pharmacol.* 83 (4), 603–613. doi:10.1007/s00280-018-3759-5

Cho, H., Hong, S. W., Oh, Y. J., A Kim, M., Kang, E. S., Lee, J. M., et al. (2008). Clinical Significance of Osteopontin Expression in Cervical Cancer. *J. Cancer Res. Clin. Oncol.* 134 (8), 909–917. doi:10.1007/s00432-007-0351-5

Daugaard, I., Sanders, K. J., Idica, A., Vittayarukskul, K., Hamdorf, M., Krog, J. D., et al. (2017). miR-151a Induces Partial EMT by Regulating E-Cadherin in NSCLC Cells. *Oncogenesis* 6 (7), e366. doi:10.1038/oncsis.2017.66

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., et al. (2006). Type, Density, and Location of Immune Cells within Human Colorectal Tumors Predict Clinical Outcome, *Science*. 313, 1960–1964. doi:10.1126/science.1129139

Guo, Q., Jin, Z., Yuan, Y., Liu, R., Xu, T., Wei, H., et al. (2016). New Mechanisms of Tumor-Associated Macrophages on Promoting Tumor Progression: Recent Research Advances and Potential Targets for Tumor Immunotherapy. *J. Immunol. Res.* 2016, 1–12. doi:10.1155/2016/9720912

Hsieh, I.-S., Yang, R.-S., and Fu, W.-M. (2014). Osteopontin Upregulates the Expression of Glucose Transporters in Osteosarcoma Cells. *PloS one* 9 (10), e109550. doi:10.1371/journal.pone.0109550

Huang, J., Liu, T., Shang, C., Zhao, Y., Wang, W., Liang, Y., et al. (2018). Identification of lncRNAs by Microarray Analysis Reveals the Potential Role of lncRNAs in Cervical Cancer Pathogenesis. *Oncol. Lett.* 15 (4), 5584–5592. doi:10.3892/ol.2018.8037

Huang, X., Qian, Y., Wu, H., Xie, X., Zhou, Q., Wang, Y., et al. (2015). Aberrant Expression of Osteopontin and E-Cadherin Indicates Radiation Resistance and Poor Prognosis for Patients with Cervical Carcinoma. *J. Histochem. Cytochem.* 63 (2), 88–98. doi:10.1369/0022155414561329

Kerjaschki, D. (2005). The Crucial Role of Macrophages in Lymphangiogenesis. *J. Clin. Invest.* 115 (9), 2316–2319. doi:10.1172/jci26354

Klement, J. D., Poschel, D. B., Lu, C., Merting, A. D., Yang, D., Redd, P. S., et al. (2021). Osteopontin Blockade Immunotherapy Increases Cytotoxic T Lymphocyte Lytic Activity and Suppresses Colon Tumor Progression. *Cancers* 13 (5), 1006. doi:10.3390/cancers13051006

Kothari, A., Arffa, M., Chang, V., Blackwell, R., Syn, W.-K., Zhang, J., et al. (2016). Osteopontin-A Master Regulator of Epithelial-Mesenchymal Transition. *Jcm* 5 (4), 39. doi:10.3390/jcm5040039

Kumar, V., Behera, R., Lohite, K., Karnik, S., and Kundu, G. C. (2010). p38 Kinase Is Crucial for Osteopontin-Induced Furin Expression that Supports Cervical Cancer Progression. *Cancer Res.* 70 (24), 10381–10391. doi:10.1158/0008-5472.Can-10-1470

Le, Q. T., Sutphin, P. D., Raychaudhuri, S., Yu, S. C., Terris, D. J., Lin, H. S., et al. (2003). Identification of Osteopontin as a Prognostic Plasma Marker for Head and Neck Squamous Cell Carcinomas. *Clin. Cancer Res.* 9 (1), 59–67.

Lee, M.-Y., Chou, C.-Y., Tang, M.-J., and Shen, M.-R. (2008). Epithelial-mesenchymal Transition in Cervical Cancer: Correlation with Tumor Progression, Epidermal Growth Factor Receptor Overexpression, and Snail Up-Regulation. *Clin. Cancer Res.* 14 (15), 4743–4750. doi:10.1158/1078-0432.Ccr-08-0234

Leung, D. T. M., Lim, P.-L., Cheung, T.-H., Wong, R. R. Y., Yim, S.-F., Ng, M. H. L., et al. (2016). Osteopontin Fragments with Intact Thrombin-Sensitive Site Circulate in Cervical Cancer Patients. *PloS one* 11 (8), e0160412. doi:10.1371/journal.pone.0160412

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, P., Zhang, C., Liao, Y., Liu, J., Huang, J., Xia, M., et al. (2020). High Expression of PTPRM Predicts Poor Prognosis and Promotes Tumor Growth and Lymph Node Metastasis in Cervical Cancer. *Cell Death Dis* 11 (8), 687. doi:10.1038/s41419-020-02826-x

Lu, C., Fang, S., Weng, Q., Lv, X., Meng, M., Zhu, J., et al. (2020). Integrated Analysis Reveals Critical Glycolytic Regulators in Hepatocellular Carcinoma. *Cell Commun Signal* 18 (1), 97. doi:10.1186/s12964-020-00539-4

Mahmoud, S. M. A., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H. S., et al. (2011). Tumor-infiltrating CD8+ Lymphocytes Predict Clinical Outcome in Breast Cancer. *Jco* 29 (15), 1949–1955. doi:10.1200/jco.2010.30.5037

Manavi, M., Hudelist, G., Fink-Retter, A., Gschwandtler-Kaulich, D., Pischinger, K., and Czerwenka, K. (2007). Gene Profiling in Pap-Cell Smears of High-Risk Human Papillomavirus-Positive Squamous Cervical Carcinoma. *Gynecol. Oncol.* 105 (2), 418–426. doi:10.1016/j.ygyno.2006.12.032

Moorman, H. R., Poschel, D., Klement, J. D., Lu, C., Redd, P. S., and Liu, K. (2020). Osteopontin: A Key Regulator of Tumor Progression and Immunomodulation. *Cancers* 12 (11), 3379. doi:10.3390/cancers12113379

Nishio, H., Iwata, T., and Aoki, D. (2021). Current Status of Cancer Immunotherapy for Gynecologic Malignancies. *Jpn. J. Clin. Oncol.* 51 (2), 167–172. doi:10.1093/jjco/hyaa214

Oldberg, A., Franzén, A., and Heinegård, D. (1986). Cloning and Sequence Analysis of Rat Bone Sialoprotein (Osteopontin) cDNA Reveals an Arg-Gly-Asp Cell-Binding Sequence. *Proc. Natl. Acad. Sci. U.S.A.* 83 (23), 8819–8823. doi:10.1073/pnas.83.23.8819

Ostheimer, C., Gunther, S., Bache, M., Vordermark, D., and Multhoff, G. (2017). Dynamics of Heat Shock Protein 70 Serum Levels as a Predictor of Clinical Response in Non-small-cell Lung Cancer and Correlation with the Hypoxia-Related Marker Osteopontin. *Front. Immunol.* 8, 1305. doi:10.3389/fimmu.2017.01305

Petrillo, M., Zannoni, G. F., Martinelli, E., Pedone Anchora, L., Ferrandina, G., Tropeano, G., et al. (2015). Polarisation of Tumor-Associated Macrophages toward M2 Phenotype Correlates with Poor Response to Chemoradiation and Reduced Survival in Patients with Locally Advanced Cervical Cancer. *PloS one* 10 (9), e0136654. doi:10.1371/journal.pone.0136654

Qureshi, R., Arora, H., and Rizvi, M. A. (2015). EMT in Cervical Cancer: its Role in Tumour Progression and Response to Therapy. *Cancer Lett.* 356 (2 Pt B), 321–331. doi:10.1016/j.canlet.2014.09.021

Raja, R., Kale, S., Thorat, D., Soundararajan, G., Lohite, K., Mane, A., et al. (2014). Hypoxia-driven Osteopontin Contributes to Breast Tumor Growth through Modulation of HIF1α-Mediated VEGF-dependent Angiogenesis. *Oncogene* 33 (16), 2053–2064. doi:10.1038/onc.2013.171

Shang, C., Wang, W., Liao, Y., Chen, Y., Liu, T., Du, Q., et al. (2018). LNMICC Promotes Nodal Metastasis of Cervical Cancer by Reprogramming Fatty Acid Metabolism. *Cancer Res.* 78 (4), 877–890. doi:10.1158/0008-5472.Can-17-2356

Shi, Z., Wang, B., Chihanga, T., Kennedy, M. A., and Weber, G. F. (2014). Energy Metabolism during anchorage-independence. Induction by Osteopontin-C. *PloS one* 9 (8), e105675. doi:10.1371/journal.pone.0105675

Thoms, J. W., Dal Pra, A., Anborgh, P. H., Christensen, E., Fleshner, N., Menard, C., et al. (2012). Plasma Osteopontin as a Biomarker of Prostate Cancer Aggression: Relationship to Risk Category and Treatment Response. *Br. J. Cancer* 107 (5), 840–846. doi:10.1038/bjc.2012.345

Vergis, R., Corbishley, C. M., Norman, A. R., Bartlett, J., Jhavar, S., Borre, M., et al. (2008). Intrinsic Markers of Tumour Hypoxia and Angiogenesis in Localised Prostate Cancer and Outcome of Radical Treatment: a Retrospective Analysis of Two Randomised Radiotherapy Trials and One Surgical Cohort Study. *Lancet Oncol.* 9 (4), 342–351. doi:10.1016/s1470-2045(08)70076-7

Vordermark, D., Said, H. M., Katzer, A., Kuhnt, T., Hänsgen, G., Dunst, J., et al. (2006). Plasma Osteopontin Levels in Patients with Head and Neck Cancer and Cervix Cancer Are Critically Dependent on the Choice of ELISA System. *BMC cancer* 6, 207. doi:10.1186/1471-2407-6-207

Wang, K. X., and Denhardt, D. T. (2008). Osteopontin: Role in Immune Regulation and Stress Responses. *Cytokine Growth Factor. Rev.* 19 (5-6), 333–345. doi:10.1016/j.cytogfr.2008.08.001

Wei, R., Wong, J. P. C., and Kwok, H. F. (2017). Osteopontin -- a Promising Biomarker for Cancer Therapy. *J. Cancer* 8 (12), 2173–2183. doi:10.7150/jca.20480

Xu, X., Jiang, X., Chen, L., Zhao, Y., Huang, Z., Zhou, H., et al. (2019). Retracted: MiR-181a Promotes Apoptosis and Reduces Cisplatin Resistance by Inhibiting Osteopontin in Cervical Cancer Cells. *Cancer Biother. Radiopharm.* 34 (9), 559–565. doi:10.1089/cbr.2019.2858

Yang, M., Ma, C., Liu, S., Sun, J., Shao, Q., Gao, W., et al. (2009). Hypoxia Skews Dendritic Cells to a T Helper Type 2-stimulating Phenotype and Promotes Tumour Cell Migration by Dendritic Cell-Derived Osteopontin. *Immunology* 128 (1 Suppl. l), e237–e249. doi:10.1111/j.1365-2567.2008.02954.x

Zhang, C., Liao, Y., Liu, P., Du, Q., Liang, Y., Ooi, S., et al. (2020). FABP5 Promotes Lymph Node Metastasis in Cervical Cancer by Reprogramming Fatty Acid Metabolism. *Theranostics* 10 (15), 6561–6580. doi:10.7150/thno.44868

Zhang, L., Conejo-Garcia, J. R., Katsaros, D., Gimotty, P. A., Massobrio, M., Regnani, G., et al. (2003). Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer. *N. Engl. J. Med.* 348 (3), 203–213. doi:10.1056/NEJMoa020177

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Qin, Yi, Liang, Chen, Wang, Liao, Zhang, Huang, Huang and Yao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fgene.2022.891665

CNTN-1 Upregulation Induced by Low-Dose Cisplatin Promotes Malignant Progression of Lung Adenocarcinoma Cells *via* Activation of Epithelial-Mesenchymal Transition

*Ruijie Zhang, Shengjin Li, Jian Lan, Changyi Li, Xianzhi Du, Weijie Dong, Qian Yu and Daoxin Wang**

Department of Respiratory and Critical Care Medicine, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Shengyang He,
Central South University, China
Wu Di,
Xinqiao Hospital, China
Li Shan,
Hubei University of Medicine, China

***Correspondence:**
Daoxin Wang
304863@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *08 March 2022*
Accepted: *11 April 2022*
Published: *27 May 2022*

Citation:
Zhang R, Li S, Lan J, Li C, Du X, Dong W, Yu Q and Wang D (2022) CNTN-1 Upregulation Induced by Low-Dose Cisplatin Promotes Malignant Progression of Lung Adenocarcinoma Cells via Activation of Epithelial-Mesenchymal Transition. Front. Genet. 13:891665. doi: 10.3389/fgene.2022.891665

Tumor metastasis and invasion are the main impediments to lung adenocarcinoma successful treatment. Previous studies demonstrate that chemotherapeutic agents can elevate the malignancy of cancer cells other than their therapeutic effects. In this study, the effects of transient low-dose cisplatin treatment on the malignant development of lung adenocarcinoma cells (A549) were detected, and the underlying epigenetic mechanisms were investigated. The findings showed that A549 cells exhibited epithelial-mesenchymal transition (EMT)-like phenotype along with malignant progression under the transient low-dose cisplatin treatment. Meanwhile, low-dose cisplatin was found to induce contactin-1 (*CNTN-1*) upregulation in A549 cells. Subsequently, we found that further overexpressing *CNTN*-1 in A549 cells obviously activated the EMT process *in vitro* and *in vivo*, and caused malignant development of A549 cells *in vitro*. Taken together, we conclude that low-dose cisplatin can activate the EMT process and resulting malignant progression through upregulating *CNTN-1* in A549 cells. The findings provided new evidence that a low concentration of chemotherapeutic agents could facilitate the malignancy of carcinoma cells *via* activating the EMT process other than their therapeutic effects.

Keywords: CNTN-1, EMT, cisplatin, A549 cell, metastasis, invasion

INTRODUCTION

Non–small cell lung cancer (NSCLC) is one of the leading causes of mortality in the world, and lung adenocarcinoma is identified as the most common pathological type of NSCLC (Corral de la Fuente et al., 2021; Coleman et al., 2022). Though considerable manpower and research funds are applied to explore the mechanism of tumor formation and progression, the treatment efficiency for advanced lung adenocarcinoma is still unsatisfying. It is well established that tumor metastasis and invasion are the main impediments to tumor successful treatment, which finally lead to a dismal prognosis of lung adenocarcinoma (Gong et al., 2021). A comprehensive understanding of the underlying molecular mechanisms of lung adenocarcinoma progression will help to establish effective therapies.

Epithelial-mesenchymal transition (EMT) is a dynamic cellular reprogramming process through which epithelial cells undergo morphological changes characterized by loss of cell-cell adhesion and acquisition of mobile fibroblast-like phenotype. A previous study demonstrated that EMT

progression participated in embryonic development and tissue remolding (Greaves and Calle, 2022). Subsequent evidence suggested that the aberrant activation of the EMT process promoted metastasis, invasion, and drug resistance of multiple epithelial carcinomas, such as gastric carcinoma, breast cancer, and lung cancer (Menju and Date, 2021; Azimi et al., 2022; Buyuk et al., 2022). The EMT progression is regulated by complicated epigenetic mechanisms including EMT-related transcription factors (e.g., snail, slug, and twist), genes (e.g., FOXC2, DYRK2, and PGRMC1), and signaling pathways (e.g., PI3K/Akt, Wnt/β-catenin, and ERK1/2) (Lin et al., 2021; Verdura et al., 2021; Debnath et al., 2022). Interestingly, several types of research revealed that EMT progression could be activated by chemotherapeutic agents in some cancer cells. For example, ovarian carcinoma cells exhibited EMT phenotype and stem cell properties along with the increased migration and invasion after the transient cisplatin treatment (Sulaiman et al., 2016). Similarly, low-dose doxorubicin was reported to induce the EMT process and result in a malignant enhancement in gastric cancer cells (Han et al., 2013; Han et al., 2014).

Contactin-1 (*CNTN*-1), a neuronal cell adhesion molecular, was not only involved in the nervous system development including axon guidance, synapse formation, and nerve impulse conduction, but also played a vital role in lymphangiogenesis, lymphatic metastasis, and proliferation of many epithelial malignancies including prostate cancer, hepatocellular cancer, breast cancer, and lung cancer (Bamodu et al., 2020; Cao et al., 2021; Chen et al., 2021; Kandasamy et al., 2022). Additionally, in our prior studies, *CNTN*-1 upregulation was revealed to increase metastasis, invasion, and chemoresistance in lung adenocarcinoma cells with cisplatin resistance (Zhang et al., 2015). Moreover, silencing of *CNTN*-1 expression was recently reported to decrease malignancy and improve the prognosis in patients with gastric cancer by inhibiting EMT progression (Umeda et al., 2022).

Chemotherapeutic agents at low concentrations could facilitate the malignancy of carcinoma cells other than their therapeutic effects (Mottaghi and Abbaszadeh, 2021). However, whether and how low-dose cisplatin promotes the malignant progression of lung adenocarcinoma cells is still unclear. Here, we demonstrated that low-dose cisplatin promoted the malignant progression of lung adenocarcinoma cells through CTN-1 up-regulation-induced activation of the EMT process.

MATERIALS AND METHODS

Cell Culture

Cell line A549 of human lung adenocarcinoma was purchased from the Institute of Biochemistry and Cell Biology, Chinese Academy of Science (Shanghai, China) and cultured in RPMI 1640 medium (Hyclone, United States) supplemented with 10% fetal bovine serum (FBS, Gibco, United States) and 1% penicillin-streptomycin (Hyclone, United States) under standard conditions (37°C, 20% O_2 and 5% CO_2). The culture medium was changed every 3 days. When reaching 90% confluence, the cells were passaged using 0.25% trypsin with 0.1% EDTA (Gibco, United States).

Cell Transfection

The *CNTN*-1 overexpressing lentivirus LV5-*CNTN*-1 was purchased from Genepharma (Shanghai, China) and the full sequence of the recombinant lentiviral vector LV5-*CNTN*-1 could be seen in **Supplementary Material S1**. Briefly, after being seeded in a 24-well plate for 48 h, A549 cells were transfected with LV5-*CNTN*-1 lentiviral vector in triplicate according to the manufacturer's protocol, and then fluorescence intensity was observed under the microscope. The successfully transfected cells (A549-*CNTN*-1) were selected *via* 1 g/ml puromycin and cultured for future use.

Cell Proliferation Assay

Cell proliferation was analyzed using the cell counting kit-8 (CCK-8, Beyotime, China). Briefly, A549 cells were plated in 96-well plates (2000 cells/well) and cultured in the humidified incubator overnight. Then, A549 cells were treated with different concentrations (5, 2, 1, 0.5, and 0 μg/ml) of cisplatin for 24, 48, and 72 h, respectively. At each time point, A549 cells were incubated with 100 μL fresh culture medium and 10 μL CCK-8 solution for 1 h. Cell proliferation potency was reflected by the absorbance value at 450 nm wavelength which was measured using an automatic microplate reader (Bio-rad).

Cell Apoptosis Assay

A549 cell apoptosis was analyzed by flow cytometry. Briefly, After A549 cells were seeded in Petri dishes (10-cm diameter, 3×10^4 cells per dish) and grown to 60–70% confluence, A549 cells were incubated with different concentrations (5, 2, 1, 0.5, and 0 μg/ml) of cisplatin for 48 h. After A549 cells were stained with annexin V-FITC/PI stain solution according to the manufacturer's instructions (Beyotime, China), a flow cytometry assay was performed to detect apoptotic cells which were defined as annexin V-FITC-positive but PI-negative cells, or double positive-stained cells.

Cell Migration Measurement

Wound-healing assay was used to detect cell migration. Briefly, after A549 cells were seeded in 24-well plates (5×10^5 cells per well) and cultured overnight, A549 cells were washed with sterile phosphate buffer solution (PBS) and then cultured in the serum-free medium with and without 0.5 μg/ml cisplatin for 24 h. The wound was scratched by sterile pipette tips. Representative images were taken at 0 h and 24 h using a light microscope (Nikon, Japan). Finally, cell migration ability was analyzed using the Image-Pro Plus software (Version 5.1, Media Cybernetics, Inc.).

Invasion Assay

Cell invasion assay was performed using a transwell chamber with 8 μm pore size polycarbonate membrane in a 24-well plate. Briefly, after the Matrigel (BD, Biosciences) coated on the upper chambers was rehydrated for 1 h at 37°C, 3×10^4 cells suspended in 200 μL RPMI-1640 medium with different concentrations (0 and

TABLE 1 | Primers of target genes.

Gene	Accession number	Forward (5′-3′)	Reverse (5′-3′)
GAPDH	NM_001256799.2	CAGCCTCAAGATCATCAGCA	TGTGGTCATGAGTCCTTCCA
CNTN-1	NM_001256063.1	GCCCATGACAAAGAAGAAGC	CGACATGATCCCAGGTGATT
E-cadherin	NM_001317184.1	GAAGTGTCCGAGGACTTTGG	CAGTGTCTCTCCAAATCCTAGA
N-cadherin	NM_001308176.1	GAGTCCACTGAGTACCGGAGAC	TGTAGGTGGCAATCTCAATGTC
Vimentin	NM_003380.3	AAAGTGTGGCTGCCAAGAAC	AGCCTCAGAGAGGTCAGCAA

0.5 μg/ml) of cisplatin were added into the upper chambers whereas 500 μL RPMI-1640 medium containing 10% FBS was added into the lower chambers. After incubation for 24 h, cells that passed through the membranes were sequentially washed with PBS, fixed in 4% paraformaldehyde, and then stained with crystal violet dye. Finally, the invasive A549 cells were observed under a light microscope (Nikon, Japan). Cell invasion ability was analyzed using the Image-Pro Plus software (Version 5.1, Media Cybernetics, Inc.).

In vitro Drug Sensitivity Assay

Briefly, cells were seeded in 96-well plates (3×10^3 cells per well) and incubated with different concentrations of cisplatin (10, 5, 2.5, 1.25, 0.625, and 0 g/ml) for 48 h. Then, cells were incubated with 100 μL fresh culture medium and 10 μL cell counting kit-8 (CCK-8) solution (Beyotime, China) for 1 h. And thereafter, the absorbance value at the wavelength of 450 nm was measured using an automatic microplate reader (Bio-rad). Finally, the 50% inhibitory concentration (IC_{50}) was calculated using GraphPad Prism 5.0 software.

RNA Isolation and Quantitative Real-Time PCR

Briefly, the total RNA was isolated using the Trizol reagent (Invitrogen, United States) according to the manufacturer's instructions and then 1 μg RNA was reverse transcribed into cDNA using a reverse transcription kit (Roche) according to the manufacturer's instructions. Primer pairs used for real-time PCR were shown in **Table 1**. Real-time PCR (qPCR) experiment was performed using a reaction system containing SYBR Green Mix (Agilent Technologies, Santa Clara, United States), cDNA, and primers. The annealing temperature was 62°C and the cycle of gene magnification was 39. Relative expression of target genes was calculated using the $2^{-\Delta\Delta Ct}$ method and normalized to GAPDH.

Protein Extraction and Western Blotting

Briefly, after cells were prepared and then washed with PBS, the cells were lysed using 200μl RIPA buffer containing protease inhibitor (Sigma, United States) on the ice for 30 min. After protein concentration was measured using the BCA Protein Assay Kit (Beyotime, Nanjing, China), proteins in each group were separated by SDS-PAGE and then transferred to the PVDF membranes (Millipore, Germany). After that, the membranes were blocked with 5% bovine serum albumin for 1 h at room temperature and incubated with primary antibodies (*CNTN*-1: Proteintech, diluted 1:1000; E-cadherin: Cell Signaling Technology, diluted 1:1000; N-cadherin: Abcam, diluted 1: 1000) overnight at 4°C, followed by incubation with corresponding HRP-conjugated secondary antibodies (HRP AffiniPure Goat Anti-Rabbit IgG: EARTHOX, diluted 1: 10,000) for 1 h at room temperature. Protein bands on the membrane were detected by SuperSignal West Pico Trial Kit (Thermo, United States) and analyzed using ImageJ software (National Institutes of Health, United States).

Xenograft Tumorigenicity Experiment

To verify the effects of low-dose cisplatin on A549 cells mediated by *CNTN*-1 in lung adenocarcinoma cells *in vivo*, a xenotransplantation experiment was performed on healthy nude mice (4–5-week old females) purchased from the Animal Laboratory of Xinqiao Hospital affiliated with the Third Military Medical University and housed in a climate-control SPF facility. Briefly, A549 cells treated with or without low-dose (0.5 μg/ml) cisplatin was injected into the right flank of nude mice (100 μL, 1×10^7 cells per animal). During the next 30 days, tumor volume was measured every 3 days. At the end of the experiment, all animals were sacrificed and xenograft tumors were excised to detect *CNTN*-1 and EMT biomarkers *via* immunohistochemistry or Western blot. All animal experiments were approved by Zhonghong Boyuan Biotechnology Co., Ltd. [SYXK (GAN) 2020-0001].

Statistical Analysis

All numerical data are expressed as means ± SD and each experiment was performed at least in triplicate in this study. After the homogeneity test for variance was finished, comparisons between two groups were performed by Independent-Samples t test, whereas comparisons between multiple groups were performed by one-way analysis of variance (ANOVA) using SPSS 13.0 software. A significant difference was indicated when the p-value < 0.05.

RESULTS

Effects of Low-Dose Cisplatin on the Viability of A549 Cells

To study the viability of A549 cells affected by low-dose cisplatin, the cells were treated with the gradient cisplatin concentrations (5, 2, 1, 0.5, and 0 μg/ml) and the results showed that the proliferation of A549 cells was attenuated with elevating cisplatin concentration (from 1 to 5 μg/ml) and with extending experiment duration (from 24 to 72 h). However, the proliferation of A549 cells incubated with

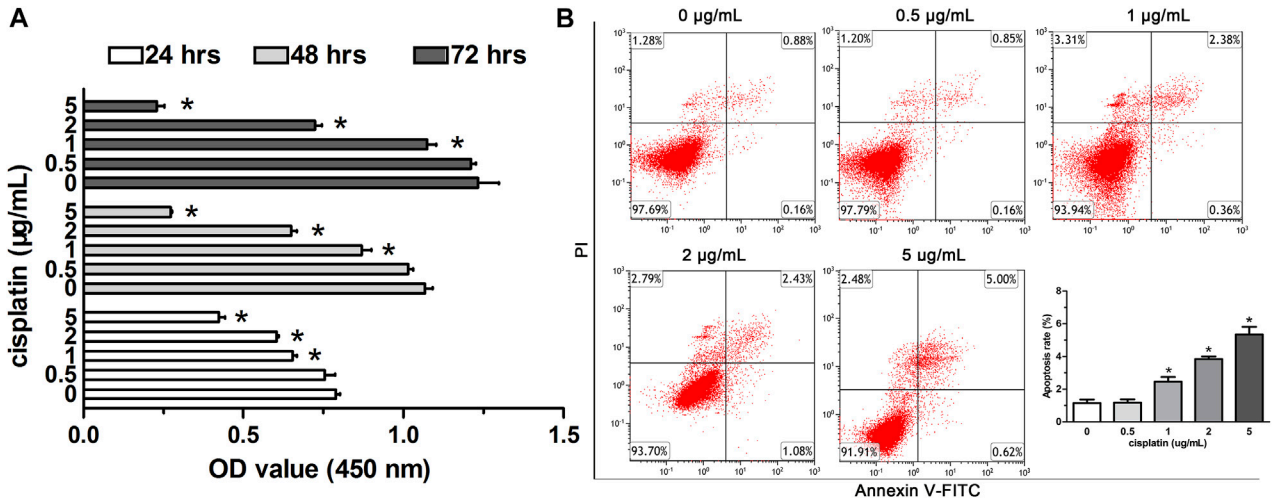


FIGURE 1 | Effects of low-dose cisplatin treatment on the viability of A549 cells. **(A)** Cell proliferation of A549 cells was detected by the CCK-8 method after incubation with different concentrations of cisplatin (5, 2, 1, 0.5, 0 g/mL) for 24, 48, and 72 h, respectively. As shown in **(A)**, the proliferation of A549 cells incubated with low-dose (0.5 μg/ml) cisplatin was similar to that of A549 cells incubated without cisplatin during the whole experiment duration (*$p > 0.05$). **(B)** Apoptosis of A549 cells was assessed with Flow cytometry analysis after being treated with different concentrations of cisplatin (5, 2, 1, 0.5, 0 g/mL). PI staining was indicated as red fluorescence and Annexin V as green fluorescence. As shown in **(B)**, apoptosis of A549 cells was not affected by the low-dose cisplatin (0.5 μg/ml) compared with that of controls (A549 cells treated without cisplatin, *$p > 0.05$). Each experiment was conducted in triplicate and the data were expressed as mean ± SD.

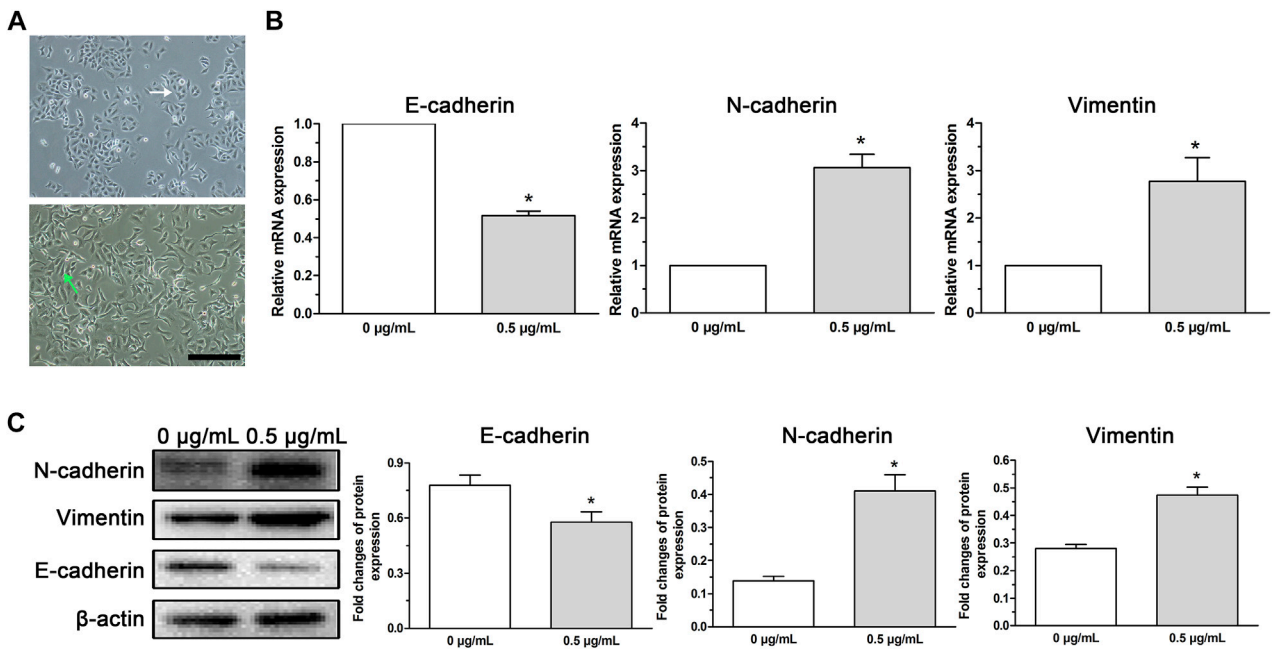


FIGURE 2 | EMT phenotype cells induced in A549 cells by low-dose cisplatin transient treatment. **(A)** Morphological observation of A549 cells under a light microscope. A549 cells incubated with low-dose cisplatin (0.5 μg/ml) presented a spindle-like shape and irregular cell arrangement (magnification: ×100, scale bar = 100 um). **(B,C)** EMT-related markers (E-cadherin, N-cadherin, and vimentin) were evaluated by RT-PCR and Western blot, respectively. As shown in **(B,C)**, mesenchymal markers (N-cadherin and vimentin) were upregulated whereas epithelial marker (E-cadherin) was downregulated in A549 cells treated with low-dose (0.5 μg/ml) cisplatin compared with that of controls both at the mRNA and protein levels (*$p < 0.01$). Each experiment was conducted in triplicate and the data was expressed as mean ± SD.

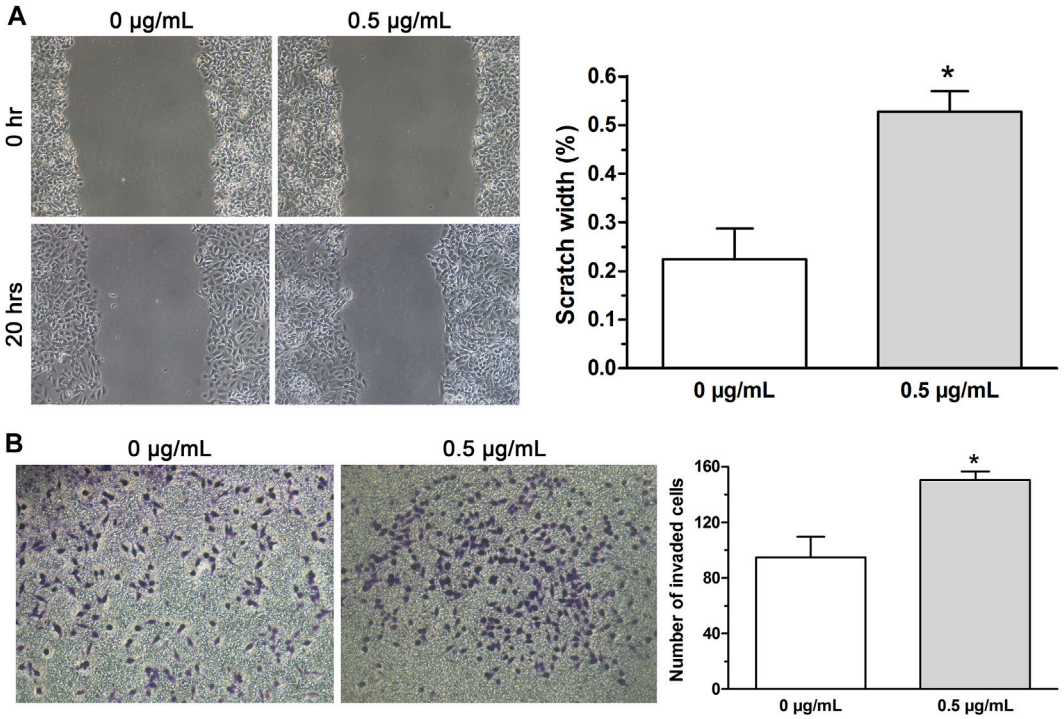


FIGURE 3 | EMT progression of A549 cells promoted by low-dose cisplatin treatment. **(A,B)** Migration and invasion of A549 cells were measured by wound healing assay and transwell invasion assay, respectively. The migration and invasive ability of A549 cells treated with low-dose (0.5 μg/ml) cisplatin was increased compared with that of controls (*$p < 0.01$). All pictures were photographed at a magnification of ×100, scale bar = 100 um. Each experiment was conducted in triplicate and the data were expressed as mean ± SD.

low-dose (0.5 μg/ml) cisplatin was similar to that of A549 cells incubated without cisplatin during the whole experiment duration (**Figure 1A**). Additionally, compared with that of controls (A549 cells incubated with cisplatin-free medium), apoptosis of A549 cells treated with a higher dose of cisplatin (from 1 to 5 μg/ml) was enhanced except for low-dose cisplatin (0.5 μg/ml) which did not have obvious effects on apoptosis of A549 cells (**Figure 1B**). Because there was no significant difference in proliferation and apoptosis between A549 cells treated with 0.5 μg/ml cisplatin and controls, 0.5 μg/ml cisplatin was used as the low dose in the following experiment.

Low-Dose Cisplatin Transient Treatment Induced EMT Phenotype in A549 Cells

After low-dose (0.5 μg/ml) cisplatin transient treatment for 48 h and the following culture for 24 h in cisplatin-free medium, cell death was detected but the survived cells displayed morphological changes characterized by a spindle-like shape and irregular cell arrangement (**Figure 2A**), which were similar to EMT phenotype. The cells with these morphological changes were used to perform cell function experiments and molecular experiments.

To analyze whether the morphological changes of A549 cells induced by low-dose (0.5 μg/ml) cisplatin was associated with EMT phenotype, EMT-related markers (E-cadherin, N-cadherin, and vimentin) were detected. Results showed that mesenchymal markers (N-cadherin and vimentin) were upregulated whereas epithelial marker (E-cadherin) was downregulated in A549 cells treated with low-dose (0.5 μg/ml) cisplatin compared with that of controls both at the mRNA and protein levels (**Figures 2B, C**), suggesting that low-dose cisplatin transient treatment activated EMT process in A549 cells.

Low-Dose Cisplatin Treatment Promoted EMT Progression of A549 Cells

Previous reports revealed that the activation of EMT process renders cytoskeleton remolding of cancer cells followed by the acquirement of new characteristics including the enhanced metastasis and invasion (de Araújo et al., 2022; Meng et al., 2022). Therefore, in this study, wound-healing assay and invasion assay were conducted to examine cell migration and invasion and the data displayed that both migration (**Figure 3A**) and invasive ability (**Figure 3B**) of A549 cells treated with low-dose (0.5 μg/ml) cisplatin were increased compared with that of controls. Together, the findings indicated that low-dose cisplatin treatment could promote EMT progression of A549 cells.

EMT Progression in A549 Cells Induced by Low-Dose Cisplatin Related With Upregulated *CNTN*-1

Our previous study revealed that *CNTN*-1 upregulation in lung adenocarcinoma patients was correlated with lymphatic invasion

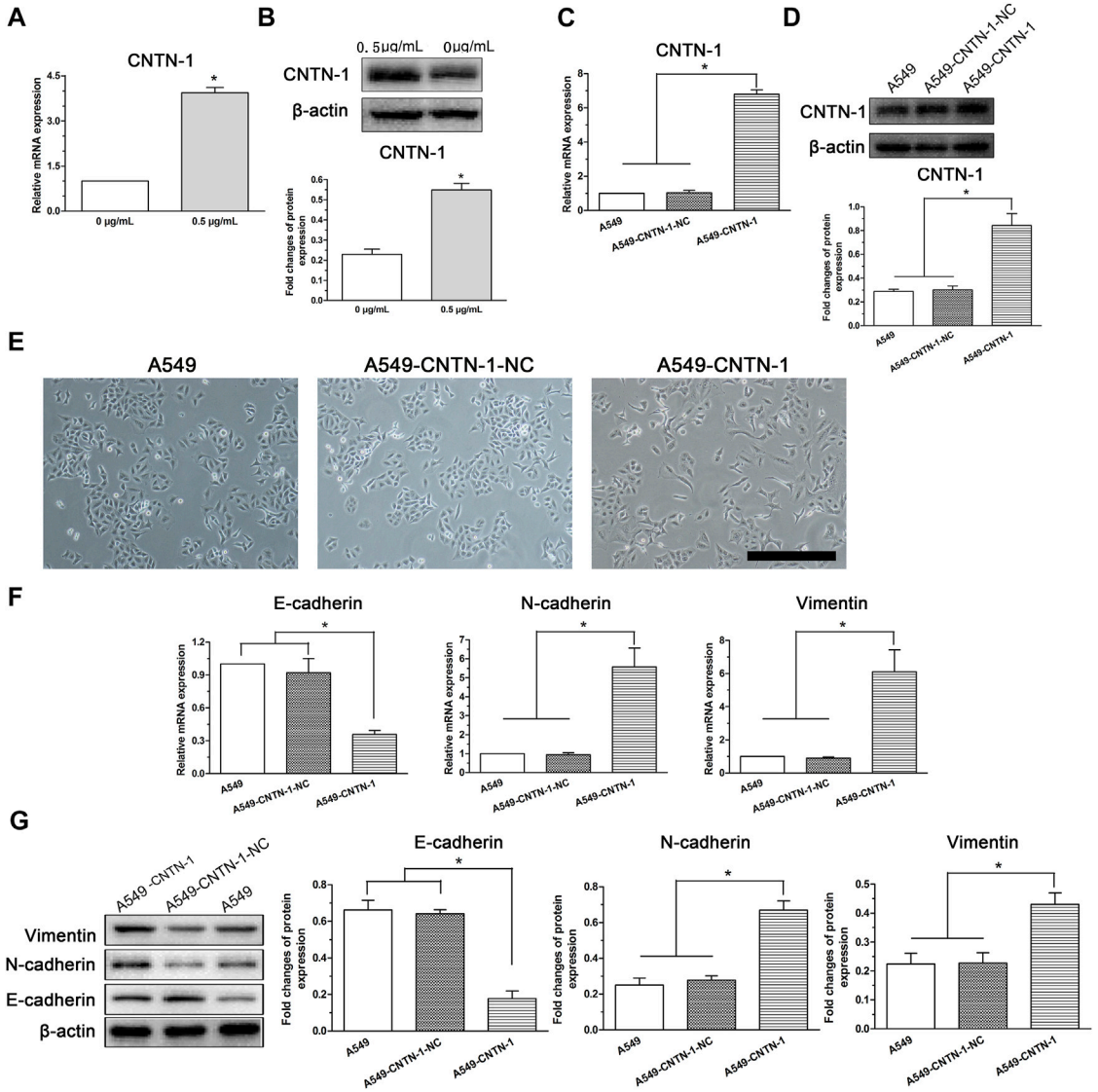


FIGURE 4 | EMT progression of A549 cells induced by low-dose cisplatin treatment related to *CNTN*-1 expression. **(A,B)** *CNTN*-1 was evaluated by RT-PCR and Western blot, respectively. *CNTN*-1 was upregulated in A549 cells treated with low-dose cisplatin (0.5 μg/ml) compared with that of controls (*$p < 0.05$). **(C,D)** *CNTN*-1 overexpression in A549-*CNTN*-1 cells was verified by RT-PCR and Western blot, respectively. *CNTN*-1 was successfully overexpressed in A549-*CNTN*-1 cells both at gene and protein levels (*$p < 0.01$). **(E)** EMT phenotype observed in A549-*CNTN*-1 cells (magnification: ×100, scale bar = 100 um). **(F,G)** EMT-related markers were detected in A549-*CNTN*-1 cells. Both N-cadherin and vimentin were increased whereas E-cadherin was decreased both at gene and protein levels in A549-*CNTN*-1 cells (*$p < 0.01$). Each experiment was conducted in triplicate and the data were expressed as mean ± SD.

during platinum-based chemotherapy (Zhang et al., 2015). Moreover, *CNTN*-1 downregulation was recently reported to inhibit EMT progression in gastric cancer cells (He et al., 2022). Therefore, to investigate whether *CNTN*-1 plays a role in the low-dose cisplatin treatment-induced EMT process and malignant progression in A549 cells, *CNTN*-1 in A549 cells treated with or without low-dose cisplatin was measured. As shown in **Figures 4A,B**, and *CNTN*-1 in A549 cells incubated with low-dose (0.5 μg/ml) cisplatin medium was higher than that of A549 cells incubated without cisplatin.

To determine the role of *CNTN*-1 in mediating the EMT process in A549 cells, after *CNTN*-1 was overexpressed in A549 cells (A549-*CNTN*-1, **Figures 4C,D**), the morphological changes were detected and the data revealed that EMT phenotype was observed in A549-*CNTN*-1 cells compared with that of controls (A549-*CNTN*-1-NC, **Figure 4E**), verified by the increased expression of N-cadherin and vimentin and the decreased expression of E-cadherin both at gene and protein levels (**Figures 4F,G**), which strongly suggests that *CNTN*-1 upregulation induced by low-dose cisplatin treatment promoted EMT phenotype of A549 cells.

Because the aberrant activation of the EMT process enhanced malignant progression (including metastasis, invasion, and chemoresistance) of various kinds of cancer cells (Girisa et al.,

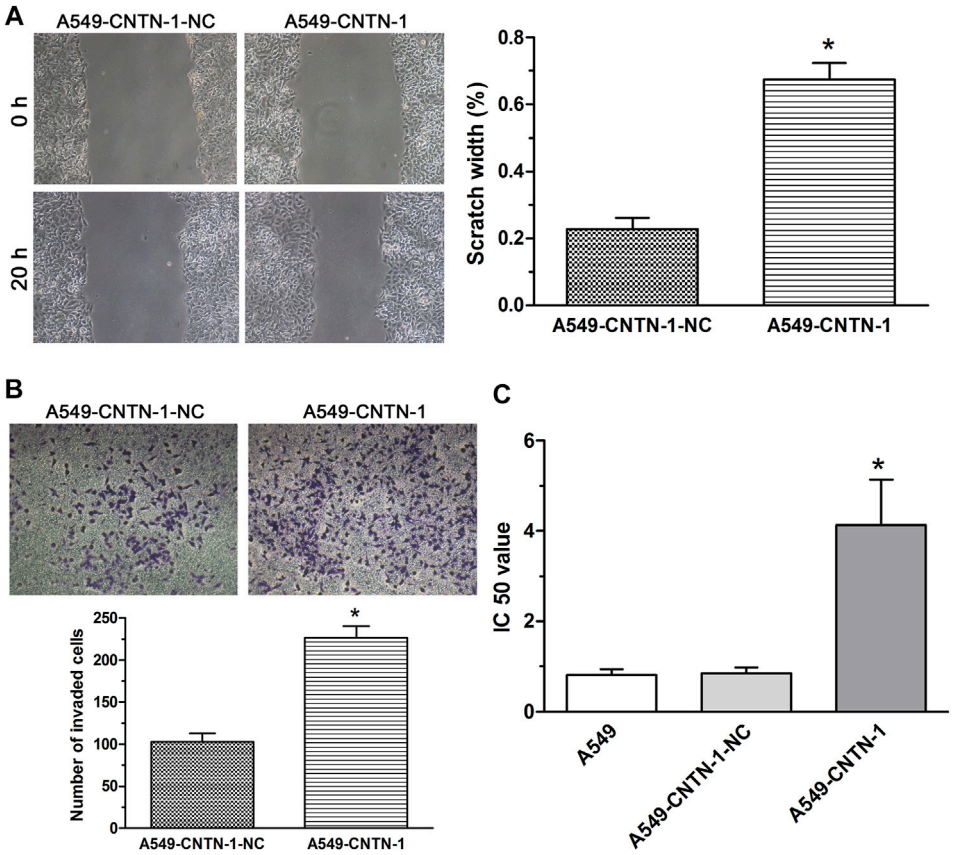


FIGURE 5 | EMT progression promoted by *CNTN*-1 overexpression. **(A,B)** Migration and invasion of A549-*CNTN*-1 cells were measured by wound healing assay and transwell invasion assay, respectively. Both migration and invasion of A549-*CNTN*-1 cells were dramatically enhanced compared with that of controls (*$p < 0.01$). **(C)** Drug sensitivity of A549-*CNTN*-1 cells evaluated by CCK-8 method. A549-*CNTN*-1 cells were more tolerant to cisplatin than A549 cells (*$p < 0.01$). All pictures were photographed at a magnification of ×100, scale bar = 100 um. Each experiment was conducted in triplicate and the data were expressed as mean ± SD.

2021; Cook and Wrana, 2022), to evaluate the association of *CNTN*-1 with malignant progression *via* EMT, the wound-healing assay, invasion assay, and drug sensitivity assay were carried out and the data showed that *CNTN*-1 overexpression dramatically enhanced the migration (**Figure 5A**), invasion (**Figure 5B**), and tolerance to cisplatin (A549-*CNTN*-1 vs. A549-*CNTN*-1-NC, **Figure 5C**). These findings indicated that *CNTN*-1 upregulation promoted EMT phenotype, which enhanced the malignant progression of A549 cells.

CNTN-1 Upregulation Induced by Low-Dose Cisplatin Activated EMT Process *in vivo*

To verify the effects of low-dose cisplatin on *CNTN*-1 expression and activation of EMT progression *in vivo*, a nude mice xenotransplantation model was developed by injection with A549 cells treated with low-dose (0.5 μg/ml) cisplatin or without cisplatin (controls). As shown in **Figures 6A–C**, no significant differences in the volume and weight were observed between tumors developed with A549 cells treated with low-dose cisplatin and controls. Results of immunohistochemistry and Western blot assay showed that *CNTN*-1 was increased in the xenograft tumor of A549 cells treated with low-dose cisplatin compared with that of controls (**Figures 6D,E**). Furthermore, both N-cadherin and vimentin were increased whereas E-cadherin was significantly decreased in the xenograft tumor of A549 cells treated with low-dose cisplatin compared with that of controls (**Figures 6F,G**), which strongly confirmed that EMT was activated by *CNTN*-1 upregulation induced by low-dose cisplatin.

DISCUSSION

Platinum-based chemotherapy is the major treatment of advanced lung adenocarcinoma with epithelial growth factor receptor (EGFR) wide-type and ineffective immunotherapy in clinical practice (Deng et al., 2022). Despite the effectively initial responses to chemotherapy, a majority of patients ultimately succumbed to malignant progression, leading to a dismal prognosis (Giacomini et al., 2020). In recent years, accumulating evidence has shown that chemotherapeutic agents have the potential to promote the malignant

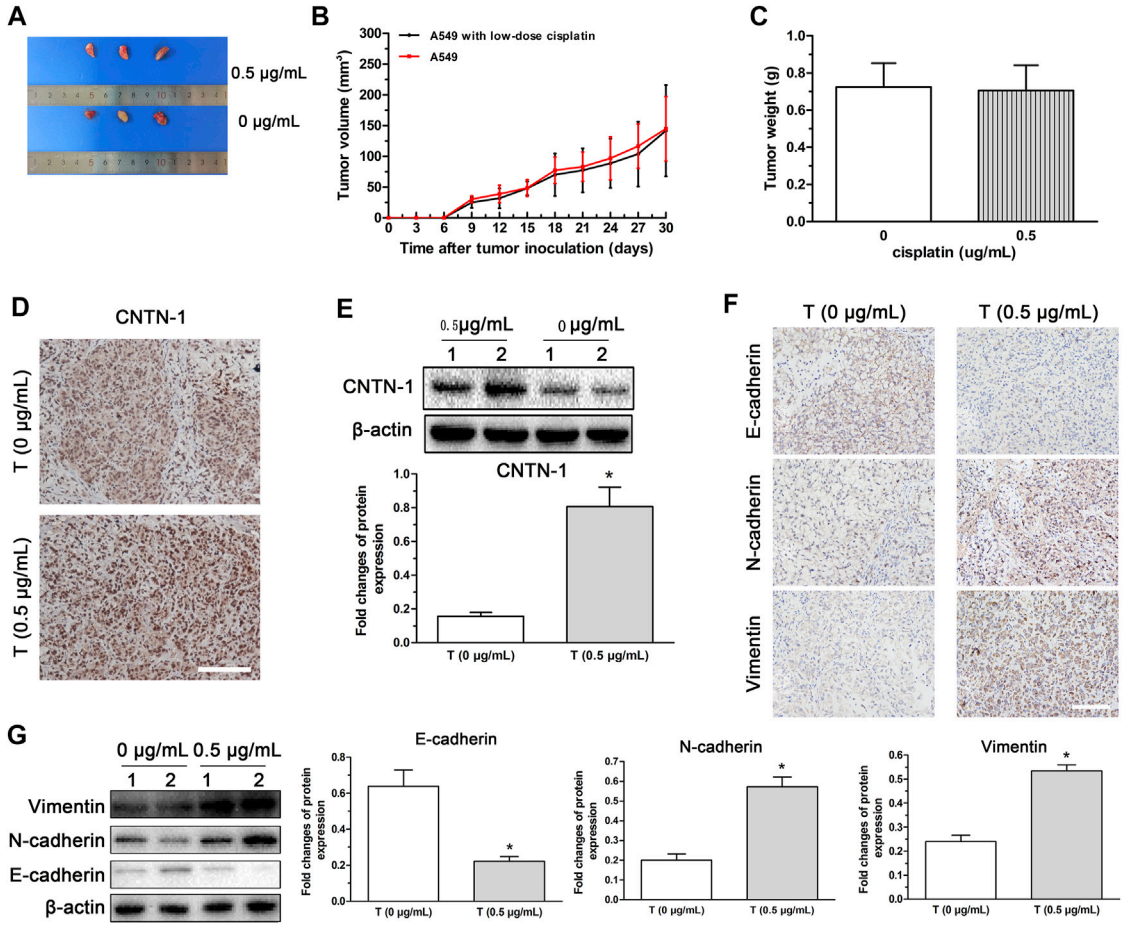


FIGURE 6 | *CNTN-1* upregulation and EMT process induced in xenograft tumors developed with A549 cells treated with low-dose cisplatin. **(A)** Tumors developed with A549 cells treated with low-dose cisplatin (0.5 g/mL) and without cisplatin (control), respectively. **(B)** Growth curves of tumors. No significant difference in volume was observed between tumors developed with A549 cells treated with low-dose cisplatin (0.5 g/mL) and controls. **(C)** Tumor weight. No significant difference of weight was observed between tumors developed by A549 cells treated with low-dose cisplatin (0.5 g/mL) and controls (*$p > 0.05$). **(D–G)** *CNTN-1* and EMT-related markers (E-cadherin, N-cadherin, and vimentin) of tumors measured by Western blot and immunohistochemistry, respectively. Both *CNTN-1* and mesenchymal markers (N-cadherin and vimentin) were significantly upregulated whereas epithelial marker (E-cadherin) was downregulated in the xenograft tumor developed by A549 cells treated with low-dose cisplatin compared with that of controls (*$p < 0.05$). All pictures were photographed at a magnification of ×100, scale bar = 100 um. Each experiment was conducted in triplicate and the data were expressed as mean ± SD.

progression of several carcinoma cells but not their efficiency in eliminating cancer cells (Latifi et al., 2011; Han et al., 2013; Wang et al., 2014; Liu et al., 2015; Lohan-Codeço et al., 2022). Improved understanding of the mechanisms of chemotherapeutic drug-induced tumor progression will be important to provide valuable information to clinical tumor chemotherapy.

The aberrant activation of the EMT process accelerates malignancy and drug tolerance of several epithelial carcinomas (Song et al., 2022). EMT progression is regulated by encoding genes, transcription factors, and Circulating exosomes (Marimuthu et al., 2021). Subsequent research showed that some medicines also had a function on EMT progression instead of their therapeutic action. For example, EGFR tyrosine kinase inhibitors (EGFR-TKIs, e.g., gefitinib and erlotinib) could induce the occurrence of the EMT process and in turn influence their malignant behavior even drug resistance *by* regulating related signaling pathways and transcription factors (Huang and Fu, 2015). Similarly, chemotherapeutics also had the function in tumor progression otherwise therapeutic effects *via* modulating EMT activation. For example, the appropriate dose of doxorubicin could activate the EMT progress and in turn, promote the malignancy in gastric cancer cells *via* β-catenin signaling (Wang et al., 2020). In addition, cisplatin treatment could induce the EMT phenotype and facilitate malignant progression even drug resistance in ovarian carcinoma cells and prostate cancer cells by several transcription factors such as Snail, Slug, and metalloproteinase 9 (Latifi et al., 2011; Liu et al., 2015). To evaluate the association of low-dose cisplatin treatment with the EMT process of lung adenocarcinoma, we first examined the morphology of A549 cells after the transient low-dose cisplatin treatment. Results showed that A549 cells treated with transient low-dose

cisplatin lost their intrinsic epithelial-like cell polarity whereas exhibited spindle-shaped mesenchymal morphology and that A549 cells treated with transient low-dose cisplatin had upregulated mesenchymal markers (N-cadherin and vimentin) expression and downregulated epithelial marker (E-cadherin) expression, indicating that transient low-dose cisplatin can induce EMT process in A549 cells. Because the aberrant activation of the EMT process was reported to promote tumor progression (Aleksakhina et al., 2019; Gu et al., 2021), to analyze whether the EMT process induced by the low-dose cisplatin can further enhance the malignant process of A549 cells, migration and invasion were compared between A549 cells treated with low-dose cisplatin and without cisplatin. The comparison demonstrated that metastasis and invasion of A549 cells treated with cisplatin were increased significantly compared that of A549 cells treat without cisplatin, indicating that transient low-dose cisplatin treatment promoted the malignant progression of A549 cells.

CNTN-1, located on chromosome 12q11-q12 and a neuronal cell adhesion molecules of the immunoglobulin superfamily, participates in nervous system development (Su et al., 2006). Additionally, *CNTN*-1 was also reported to promote the malignant progression of cancer cells as a downstream molecule of the VEGF-C/Flt-4 axis (Zhou et al., 2015; Liang et al., 2020). In our previous studies, *CNTN*-1 expression was found to be positively correlated with lymphatic invasion of lung adenocarcinoma patients who received adjuvant cisplatin- or carboplatin-based treatment after surgery (Chen et al., 2015), suggesting that the *CNTN*-1 upregulation may be induced by platinum-based chemotherapy. To prove the hypothesis, *CNTN*-1 expression was analyzed in A549 cells treated with and without low-dose cisplatin. Importantly, *CNTN*-1 expression was found to be higher in A549 cells treated with cisplatin than in A549 cells treated without cisplatin. Moreover, the upregulated *CNTN*-1 was found to promote EMT progression, verified by changes in cellular morphology, abnormal expression of EMT-related molecules (E-cadherin, N-cadherin, and vimentin), and increased metastasis and invasion of A549-*CNTN*-1 cells. Taken together, these results indicated that low-dose cisplatin-induced EMT and malignant progression of A549 cells were positively regulated by *CNTN*-1 upregulation, which was consistent with the previous report that *CNTN*-1 upregulation enhanced cancer metastasis and invasion *via* EMT alteration in gastric cancer (42).

To prove the effects low-dose of cisplatin on the EMT process *in vivo*, a nude xenotransplantation mice model was established by injection with A549 cells treated with or without low-dose (0.5 μg/ml) cisplatin. Similar to the findings *in vitro*, *CNTN*-1 was increased and the EMT process was activated in xenograft tumor developed with A549 cells treated with low-dose cisplatin compared with that of A549 cells treated without cisplatin, confirming that low-dose cisplatin could upregulate *CNTN*-1 which activated EMT process subsequently. However, because the xenograft mouse model was not suitable to observe tumor metastasis, the *in vivo* tumor metastasis was not investigated in this study.

To our knowledge, this is the first study reporting that *CNTN*-1 upregulation induced by the low-dose cisplatin activated EMT process and thus promoted malignant progression in lung adenocarcinoma cells. The findings provided new evidence that platinum-based chemotherapy could facilitate malignancy of carcinoma cells *via* activation of the EMT process by CNTN-1 overexpression other than their therapeutic effects, indicating that inhibiting the expression of CNTN-1 may reverse the EMT process and in turn, enhance the efficiency of platinum-based chemotherapy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Institutional Ethics Review Board of the Second Affiliated Hospital of CQMU.

AUTHOR CONTRIBUTIONS

RZ: conceptualization, methodology, validation, formal analysis, data curation, writing—original draft, visualization; SL, JL, and CL: methodology, software, data analysis; XD: methodology, formal analysis; WD and QY: methodology; DW: conceptualization, supervision, writing—review editing, project administration, funding acquisition. All authors read and approved the final manuscript.

FUNDING

This work is supported by grants from the National Natural Science Foundation of China (81902341), Postdoctoral Science Foundation of China (2020M683264), Postdoctoral Research Program of Chongqing (2010010005367114), National Natural Science Foundation of Chonging (cstc2021jcyj-msxm0238) and Integrated innovation and application of key technologies for precise prevention and treatment of rimary lung cancer (2019ZX002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.891665/full#supplementary-material

REFERENCES

Aleksakhina, S. N., Kashyap, A., and Imyanitov, E. N. Mechanisms of Acquired Tumor Drug Resistance. *Biochim. Biophys. Acta (Bba) - Rev. Cancer*. 2019 Dec; 1872(2):188310. PubMed PMID: 31442474.doi:10.1016/j.bbcan.2019.188310

Azimi, M., Totonchi, M., and Ebrahimi, M. Determining the Role of MicroRNAs in Self-Renewal, Metastasis and Resistance to Drugs in Human Gastric Cancer Based on Data Mining Approaches: A Systematic Review. *Cell J.* 2022 Jan;24(1): 1–6. PubMed PMID: 35182058.doi:10.22074/cellj.2022.7173

Bamodu, O. A., Chang, H. L., Ong, J. R., Lee, W. H., Yeh, C. T., and Tsai, J. T. Elevated PDK1 Expression Drives PI3K/AKT/MTOR Signaling Promotes Radiation-Resistant and Dedifferentiated Phenotype of Hepatocellular Carcinoma. *Cells*. 2020 Mar 18;9(3). PubMed PMID: 32197467. Pubmed Central PMCID: 7140693.doi:10.3390/cells9030746

Buyuk, B., Jin, S., and Ye, K. Epithelial-to-Mesenchymal Transition Signaling Pathways Responsible for Breast Cancer Metastasis. *Cel. Mol. Bioeng.* 2022Feb; 15(1):1–13. PubMed PMID: 35096183. Pubmed Central PMCID: 8761190.doi:10.1007/s12195-021-00694-9

Cao, H. J., Zhou, W., Xian, X. L., Sun, S. J., Ding, P. J., Tian, C. Y., et al. (2021). A Mixture of Baicalein, Wogonin, and Oroxylin-A Inhibits EMT in the A549 Cell Line *via* the PI3K/AKT-TWIST1-Glycolysis Pathway. *Front. Pharmacol.* 12, 821485. PubMed PMID: 35222014. Pubmed Central PMCID: 8864075. doi:10.3389/fphar.2021.821485

Chen, B., Zhang, Y., Li, C., Xu, P., Gao, Y., and Xu, Y. (2021). CNTN-1 Promotes Docetaxel Resistance and Epithelial-To-Mesenchymal Transition *via* the PI3K/Akt Signaling Pathway in Prostate Cancer. *Arch. Med. Sci.* 17 (1), 152–165. PubMed PMID: 33488868. Pubmed Central PMCID: 7811318. doi:10.5114/aoms.2020.92939

Chen, D.-H., Yu, J.-W., Wu, J.-G., Wang, S.-L., and Jiang, B.-J. Significances of Contactin-1 Expression in Human Gastric Cancer and Knockdown of Contactin-1 Expression Inhibits Invasion and Metastasis of MKN45 Gastric Cancer Cells. *J. Cancer Res. Clin. Oncol.* 2015 Dec;141(12):2109–2120. PubMed PMID: 25952582. Pubmed Central PMCID: 4630258.doi:10.1007/s00432-015-1973-7

Coleman, N., Harbery, A., Heuss, S., Vivanco, I., and Popat, S. Targeting Un-MET Needs in Advanced Non-small Cell Lung Cancer, *Lung Cancer*. 2022 Feb;164: 56–68. PubMed PMID: 35033939.doi:10.1016/j.lungcan.2021.12.016

Cook, D. P., and Wrana, J. L. A Specialist-Generalist Framework for Epithelial-Mesenchymal Plasticity in Cancer. *Trends Cancer*. S2405-8033. 2022 Feb 17. PubMed PMID: 35183479.doi:10.1016/j.trecan.2022.01.014

Corral de la Fuente, E., Olmedo Garcia, M. E., Gomez Rueda, A., Lage, Y., and Garrido, P. (2021). Targeting KRAS in Non-small Cell Lung Cancer. *Front. Oncol.* 11, 792635. PubMed PMID: 35083149. Pubmed Central PMCID: 8784727. doi:10.3389/fonc.2021.792635

de Araújo, R. S. A., Carmo, J. O. D. S., de Omena Silva, S. L., Costa da Silva, C. R. A., Souza, T. P. M., Mélo, N. B., et al. Coumarin Derivatives Exert Anti-lung Cancer Activity by Inhibition of Epithelial-Mesenchymal Transition and Migration in A549 Cells. *Pharmaceuticals (Basel)*, 15. 2022 Jan 17;15(1). PubMed PMID: 35056161. Pubmed Central PMCID: 8782015.doi:10.3390/ph15010104

Debnath, P., Huirem, R. S., Dutta, P., and Palchaudhuri, S. Epithelial-mesenchymal Transition and its Transcription Factors. *Biosci. Rep.*, 42. 2022 Jan 28;42(1). PubMed PMID: 34708244. Pubmed Central PMCID: 8703024.doi:10.1042/BSR20211754

Deng, H., Wang, L., Yang, Y., Wu, J., and Zhou, C. Dual Immunotherapy in Advanced Non-small Cell Lung Cancer: the Progress and Clinical Application. *Zhongguo Fei Ai Za Zhi*. 2022 Feb 20;25(2):102–110. PubMed PMID: 35224963.doi:10.3779/j.issn.1009-3419.2021.102.48

Giacomini, I., Ragazzi, E., Pasut, G., and Montopoli, M. The Pentose Phosphate Pathway and its Involvement in Cisplatin Resistance. *Int. J. Mol. Sci.*, 21. 2020 Jan 31;21(3). PubMed PMID: 32023830. Pubmed Central PMCID: 7036764.doi:10.3390/ijms21030937

Girisa, S., Henamayee, S., Parama, D., Rana, V., Dutta, U., and Kunnumakkara, A. B. Targeting Farnesoid X Receptor (FXR) for Developing Novel Therapeutics against Cancer. *Mol. Biomed.* 2021 Jul 10;2(1):21. PubMed PMID: 35006466. Pubmed Central PMCID: 8607382.doi:10.1186/s43556-021-00035-2

Gong, F., Yang, Y., Wen, L., Wang, C., Li, J., and Dai, J. (2021). An Overview of the Role of Mechanical Stretching in the Progression of Lung Cancer. *Front. Cel Dev. Biol.* 9, 781828. PubMed PMID: 35004682. Pubmed Central PMCID: 8740071. doi:10.3389/fcell.2021.781828

Greaves, D., and Calle, Y. Epithelial Mesenchymal Transition (EMT) and Associated Invasive Adhesions in Solid and Haematological Tumours. *Cells*. 2022 Feb 13;11(4). PubMed PMID: 35203300. Pubmed Central PMCID: 8869945.doi:10.3390/cells11040649

Gu, Y., Xu, H., and Tang, D. Mechanisms of Primary Membranous Nephropathy. *Biomolecules*, 11. 2021 Mar 30;11(4). PubMed PMID: 33808418. Pubmed Central PMCID: 8065962.doi:10.3390/biom11040513

Han, R.-F., Ji, X., Dong, X.-G., Xiao, R.-J., Liu, Y.-P., Xiong, J., et al. (2014). An Epigenetic Mechanism Underlying Doxorubicin Induced EMT in the Human BGC-823 Gastric Cancer Cell. *Asian Pac. J. Cancer Prev.* 15 (10), 4271–4274. PubMed PMID: 24935383. doi:10.7314/apjcp.2014.15.10.4271

Han, R., Xiong, J., Xiao, R., Altaf, E., Wang, J., Liu, Y., et al. Activation of β-catenin Signaling Is Critical for Doxorubicin-Induced Epithelial-Mesenchymal Transition in BGC-823 Gastric Cancer Cell Line. *Tumor Biol.* 2013 Feb; 34(1):277–284. PubMed PMID: 23055201. Epub 2012/10/12. eng.doi:10.1007/s13277-012-0548-3

He, J., Chen, Z., Xue, Q., and Shi, W. (2022). Block of Proliferation 1 Promotes Proliferation, Invasion and Epithelial Mesenchymal Transformation in Gastric Cancer. *Oxid Med. Cel Longev* 2022, 2946989. PubMed PMID: 35222794. Pubmed Central PMCID: 8865985. doi:10.1155/2022/2946989

Huang, L., and Fu, L. (2015). Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors. *Acta pharmaceutica Sinica B* 5 (5), 390–401. PubMed PMID: 26579470. Pubmed Central PMCID: 4629442. doi:10.1016/j.apsb.2015.07.001

Kandasamy, T., Sen, P., and Ghosh, S. Multi-targeted Drug Repurposing Approach for Breast Cancer *via* Integrated Functional Network Analysis. *Molecular Informatics*. 2022 Feb 23. PubMed PMID: 35195941.

Latifi, A., Abubaker, K., Castrechini, N., Ward, A. C., Liongue, C., Dobill, F., et al. Cisplatin Treatment of Primary and Metastatic Epithelial Ovarian Carcinomas Generates Residual Cells with Mesenchymal Stem Cell-like Profile. *J. Cel. Biochem.* 2011Oct;112(10):2850–2864. PubMed PMID: 21618587.doi:10.1002/jcb.23199

Liang, Y., Ma, C., Li, F., Nie, G., and Zhang, H. (2020). The Role of Contactin 1 in Cancers: What We Know So Far. *Front. Oncol.* 10, 574208. PubMed PMID: 33194679. Pubmed Central PMCID: PMC7658624. Epub 2020/11/17. eng.doi:10.3389/fonc.2020.574208

Lin, Z. Y., Yun, Q. Z., Wu, L., Zhang, T. W., and Yao, T. Z. Pharmacological Basis and New Insights of Deguelin Concerning its Anticancer Effects. *Pharmacol. Res.* 2021 Dec;174:105935. PubMed PMID: 34644595.doi:10.1016/j.phrs.2021.105935

Liu, Y.-Q., Zhang, G.-A., Zhang, B.-C., Wang, Y., Liu, Z., Jiao, Y.-L., et al. (2015). Short Low Concentration Cisplatin Treatment Leads to an Epithelial Mesenchymal Transition-like Response in DU145 Prostate Cancer Cells. *Asian Pac. J. Cancer Prev.* 16 (3), 1025–1028. PubMed PMID: 25735324. doi:10.7314/apjcp.2015.16.3.1025

Lohan-Codeço, M., Barambo-Wagner, M. L., Nasciutti, L. E., Ribeiro Pinto, L. F., Meireles Da Costa, N., and Palumbo, A., Jr. Molecular Mechanisms Associated with Chemoresistance in Esophageal Cancer. *Cell. Mol. Life Sci.* 2022 Feb 3; 79(2):116. PubMed PMID: 35113247.doi:10.1007/s00018-022-04131-6

Marimuthu, S., Rauth, S., Ganguly, K., Zhang, C., Lakshmanan, I., Batra, S. K., et al. Mucins Reprogram Stemness, Metabolism and Promote Chemoresistance during Cancer Progression. *Cancer Metastasis Rev.* 2021 Jun;40(2):575–588. PubMed PMID: 33813658.doi:10.1007/s10555-021-09959-1

Meng, C., Liu, K., Cai, X., and Chen, Y. Mechanism of miR-455-3 in Suppressing Epithelial-Mesenchymal Transition and Angiogenesis of Non-small Cell Lung Cancer Cells. *Cell stress & chaperones*. 2022;27 Jan 22. PubMed PMID: 35064898.doi:10.1007/s12192-022-01254-4

Menju, T., and Date, H. Lung Cancer and Epithelial-Mesenchymal Transition. *Gen. Thorac. Cardiovasc. Surg.* 2021 May;69(5):781–789. PubMed PMID: 33754237.doi:10.1007/s11748-021-01595-4

Mottaghi, S., and Abbaszadeh, H. The Anticarcinogenic and Anticancer Effects of the Dietary Flavonoid, Morin: Current Status, Challenges, and Future Perspectives. *Phytotherapy Res.* 2021 Dec;35(12):6843–6861. PubMed PMID: 34498311.doi:10.1002/ptr.7270

Song, J., Sun, H., Zhang, S., and Shan, C. The Multiple Roles of Glucose-6-Phosphate Dehydrogenase in Tumorigenesis and Cancer Chemoresistance. *Life (Basel)*. 2022 Feb 12;12(2). PubMed PMID: 35207558. Pubmed Central PMCID: 8875868.doi:10.3390/life12020271

Su, J.-L., Yang, P.-C., Shih, J.-Y., Yang, C.-Y., Wei, L.-H., Hsieh, C.-Y., et al. The VEGF-C/Flt-4 axis Promotes Invasion and Metastasis of Cancer Cells. *Cancer Cell.* 2006 Mar;9(3):209–223. PubMed PMID: 16530705.doi:10.1016/j.ccr.2006.02.018

Sulaiman, S. A., Ab Mutalib, N.-S., and Jamal, R. (2016). miR-200c Regulation of Metastases in Ovarian Cancer: Potential Role in Epithelial and Mesenchymal Transition. *Front. Pharmacol.* 7, 271. PubMed PMID: 27601996. Pubmed Central PMCID: 4993756. doi:10.3389/fphar.2016.00271

Umeda, S., Kanda, M., Shimizu, D., Nakamura, S., Sawaki, K., Inokawa, Y., et al. Lysosomal-associated Membrane Protein Family Member 5 Promotes the Metastatic Potential of Gastric Cancer Cells. *Gastric Cancer : official J. Int. Gastric Cancer Assoc. Jpn. Gastric Cancer Assoc.* 2022 Feb 28. PubMed PMID: 35226222.doi:10.1007/s10120-022-01284-y

Verdura, S., Cuyàs, E., Ruiz-Torres, V., Micol, V., Joven, J., Bosch-Barrera, J., et al. Lung Cancer Management with Silibinin: A Historical and Translational Perspective. *Pharmaceuticals (Basel).* 2021 Jun 11;14(6). PubMed PMID: 34208282. Pubmed Central PMCID: 8230811.doi:10.3390/ph14060559

Wang, H., Zhang, G., Zhang, H., Zhang, F., Zhou, B., Ning, F., et al. Acquisition of Epithelial-Mesenchymal Transition Phenotype and Cancer Stem Cell-like Properties in Cisplatin-Resistant Lung Cancer Cells through AKT/β-catenin/Snail Signaling Pathway. *Eur. J. Pharmacol.* 2014Jan 15;723:156–166. PubMed PMID: 24333218. Epub 2013/12/18. eng.doi:10.1016/j.ejphar.2013.12.004

Wang, Y., Dong, C., and Zhou, B. P. Metabolic Reprogram Associated with Epithelial-Mesenchymal Transition in Tumor Progression and Metastasis. *Genes Dis.* 2020 Jun;7(2):172–184. PubMed PMID: 32215287. Pubmed Central PMCID: 7083713.doi:10.1016/j.gendis.2019.09.012

Zhang, R., Yao, W., Qian, P., Li, Y., Jiang, C., Ao, Z., et al. Increased Sensitivity of Human Lung Adenocarcinoma Cells to Cisplatin Associated with Downregulated Contactin-1. *Biomed. Pharmacother.* 2015 Apr;71:172–184. PubMed PMID: 25960233.doi:10.1016/j.biopha.2014.11.004

Zhou, Z., Zhang, L., Xie, B., Wang, X., Yang, X., Ding, N., et al. FOXC2 Promotes Chemoresistance in Nasopharyngeal Carcinomas *via* Induction of Epithelial Mesenchymal Transition. *Cancer Lett.* 2015 Jul 28;363(2):137–145. PubMed PMID: 25896630.doi:10.1016/j.canlet.2015.04.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Li, Lan, Li, Du, Dong, Yu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

MINI REVIEW
published: 01 June 2022
doi: 10.3389/fgene.2022.871988

Strategies of LncRNA DLX6-AS1 on Study and Therapeutics

*Yanyan Zhao and Pei Li**

Department of Pathophysiology, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, China

Accumulating evidence has revealed the vital regulatory roles of lncRNA DLX6-AS1 in various tumors at pre-transcriptional, transcriptional, and post-transcriptional levels, which makes it a potential prognosis factor and therapeutic target. In addition, the presence of lncRNA DLX6-AS1 in the exosomes of peripheral blood of patients with tumors may also contribute to it being a possible cancer-related biomarker. However, most literature studies are devoted to studying the effect of lncRNA DLX6-AS1 as a sponging molecule of miRNAs, the research of which is likely to get stuck into a dilemma. Literature studies published already have demonstrated an exciting cell malignant phenotype inhibition with the knockdown of lncRNA DLX6-AS1 in various tumor cell lines. With the comprehensive development of delivery systems, high-throughput sequencing, and aptamers, the problems of finding novel research methods and exploring the therapeutic options which are based on lncRNA DLX6-AS1 *in vivo* could come into a period to deal with. This review aims to summarize the research statuses of lncRNA DLX6-AS1, discuss other study methodologies and therapeutic strategies on it, which might be of help to the deep learning of lncRNA DLX6-AS1 and its application from basic to clinical research.

Keywords: lncRNA DLX6-AS1, regulatory mechanism, tumor, novel research methodologies, therapeutic strategies

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Dongsheng Yan,
Wenzhou Medical University, China
Yinan Jiang,
University of Pittsburgh, United States

***Correspondence:**
Pei Li
Lipeifreemai@zzu.edu.cn

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *09 February 2022*
Accepted: *20 April 2022*
Published: *01 June 2022*

Citation:
Zhao Y and Li P (2022) Strategies of LncRNA DLX6-AS1 on Study and Therapeutics. Front. Genet. 13:871988. doi: 10.3389/fgene.2022.871988

1 INTRODUCTION

Long non-coding RNAs (lncRNAs) are a class of non-coding RNAs with a length longer than 200 nucleotides, which lack the capability of encoding proteins and were transcriptional noises at first. lncRNAs are transcribed from the opposite strand of the protein-coding genes and overlap one-third of them. lncRNA DLX6-AS1 is usually dysregulated in many tumor tissues and cell lines. Besides, the capability of lncRNA DLX6-AS1 to bind with DNAs, RNAs, and proteins makes its regulatory roles represent at pre-transcriptional, transcriptional, and post-transcriptional levels. Increasing evidence has implied the crucial regulatory roles of lncRNA DLX6-AS1 in various tumors, demonstrating the crucial influence of lncRNA DLX6-AS1 on tumorigenesis and development and making it a candidate biomarker to assist treatment and diagnosis of tumors.

2 LNCRNA DLX6-AS1 AND ITS SPLICED FORM EVF2

lncRNA DLX6-AS1, located at the human chromsome7q21.3 (GRCh38/hg38: chr7: 96955141-97014065) with a length of 1990 base pairs, is the reverse transcript of the upstream of the Dlx6 gene. In mouse, lncRNA DLX6-AS1 is also called Evf1 (embryonic ventral forebrain1) and is located at chromosome 6 (Kohtz and Fishell, 2004). Evf1 is transcribed from the upstream of the DLX6 gene,

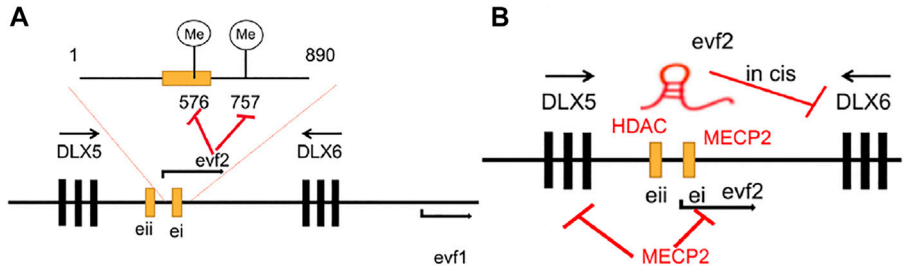


FIGURE 1 | (A) Evf2 induces ^{576}CpG and ^{757}CpG site methylation in the ultra-conserved region of DLX5/6 ei/eii. **(B)** Evf2 recruits MECP2 and further absorbs HDAC1 into the DLX5/6 ei/eii region to repress DLX5 and DLX6 gene expression.

which comprises two exons with about a 37.5 kb intronic region between them. It was first found as one of the target genes of the sonic hedgehog in the ventral forebrain (Kohtz and Fishell, 2004). Its spliced form Evf2 (also known as DLX6OS1) is transcribed from the ultra-conserved region of the DLX5/6 ei enhancer and comprises three parts which can be divided into the Evf2 5′ unique region, the Evf2 3′ unique region, and the common region with Evf1 (Feng et al., 2006). It has been found that the regulatory roles of Evf2 involve DNA methylation, histone deacetylation, and chromosome topological changes.

2.1 The Regulatory Roles of Evf2

2.1.1 DNA Methylation and Histone Deacetylation

Berghoff et al. (2013) found that Evf2 induced site-specific methylation of Dlx5/6 ei CpG^{576} and CpG^{757} in the E13. 5 MGE of the Evf2 transgene mice (Evf2$^{TS/TS}$; R) but did not affect the expression levels of Dlx5 and Dlx6 when the expression levels of evf2 were at 0.38 x wild-type levels. Apart from regulating gene methylation, it was also revealed by Bond et al. (2009) that Evf2 could recruit MECP2 to the conserved region of DLX5/6 ei/eii which decreased the recruitment of HDAC1 in the DLX5/6eii, and thus, may be the reason for the increased expression level of Dlx5 and Dlx6. Besides, Cajigas et al. (2015) also found that Evf2 lncRNAs could form an Evf2-DLX1 ribonucleoprotein (RNP), which contained the SWI/SNF-related chromatin remodelers Brahma-related gene 1 (BRG1, SMARCA4) and Brahma-associated factor (BAF170, SMARCC2) in the DXL5/6 ei region of a developing mouse forebrain, to regulate chromatin remodeling and gene expression. They revealed that Evf2 increased BRG1 binding with key Dlx5/6 enhancers with changes in lysine acetylation of histones H3 and H4, which led to significantly reduced H3AcK9 and H3AcK18 at ei, reduction of H3AcK18 to a lesser extent at eii, and an H4AcK5 decrease at four sites, while total H4AcK decreased at three intergenic sites of Evf2$^{TS/TS}$ E13.5GE when compared with the decrease at those of EVF2$^{+/+}$ E13.5GE (Cajigas et al., 2015) (**Figure 1**).

2.1.2 Chromosome Topological Changes

Chromosome topological changes allow the regulatory effects of enhancers to display over mega-base distances (Mir et al., 2019). Evf2, a cloud-forming DLX5/6 ultra-conserved enhancer ei lncRNA, can directly interact with its targeting genes and regulate the topology across a 27-Mb region by altering the gene–distance relationships among Dlx5/6, Umad1, and Akr1b8 genes (Cajigas et al., 2018). It was also found that evf2 could regulate the Dlx5/6UCE interactions to regulate the Rbmb8 gene repression *via* inducing the formation of protein pools (Sox2, Dlx, and Smarca4) (Cajigas et al., 2015).

2.1.3 Transcriptional Regulation

Berghoff et al. (2013) found that evf2 can regulate evf2, DLX6, and DLX5 expression *via* the competitive binding of MECP2 and DLX1/2 to the DLX5/6 ei/eii region, and the repression of the DLX6 gene is via its antisense transcription. The antagonism of MECP2 and Dlx1/2 comparatively binding to the ultra-conserved region leads to three kinds of activity statuses of the DLX5/6 ei/eii region, which regulates evf2 and Dlx5 differentially: inactive (MECP2 binds two gene loci), low activity (MECP2 and DLX1/2 bind one gene locus), and high activity (Dlx1/2 occupies two gene loci) (Berghoff et al., 2013). Feng et al. (2006) also demonstrated that evf2 could serve as the coactivator of DLX2 to facilitate the expression of DLX5 and DLX6 genes.

3 LNCRNA DLX6-AS1 IN TUMORS

3.1 Pre-Transcriptional Regulation

3.1.1 Mediating Methylation

Zhao et al. (2020) found that overexpression of lncRNA DLX6-AS1 could recruit DNA methyltransferase 1 (DNMT1) to the promoter region of its downstream target gene LARGE, which promoted the progression and lymph node metastasis of prostate cancer. It was also discovered by Zhao and Xu (2020) that lncRNA DLX6-AS1 was upregulated in endometrial cancer tissues and cell lines. It could form an RNA–DNA triplex *via* its triplex-forming oligonucleotide (TFO) sequence and the DLX6 triplex target site (TTS). Besides, lncRNA DLX6-AS1 also recruited a transcription factor E2F1 and histone acetyltransferase p300 to the promoter region of DLX6, which enhanced DLX6 expression and promoted proliferation, invasion, and reduced apoptosis of endometrial cancer cell lines (Zhao and Xu, 2020).

Maja Olsson and his colleagues (Olsson et al., 2016) also reported that lncRNA DLX6-AS1, one of the hyper-methylated genes in NB, was over-expressed and indicated a poorer prognosis. However, it is well known that hyper-methylation

TABLE 1 | miRNAs sponged by lncRNA DLX6-AS1 and miRNA targets in various tumors.

miRNAs	Tumors	mRNAs	References
miR-16-5p	CC	ARPP19	Xie et al. (2020)
miR-199a	CC	—	Wang et al. (2019)
miR-26a	CRC RCC LC	EZH2 PTEN TRPC3	Zeng et al. (2017); Liu Y. et al. (2020); Kong et al. (2020)
miR-27b-3p	NSCLC	GSPT1	Sun et al. (2019)
miR-181a-5p/miR-382-5p	LUSC	CELF1	Zhao et al. (2021)
miR-107	NB	BDNF	Zhang H. Y. et al. (2019)
miR-506-3p	NB	STAT2	Hu et al. (2020)
miR-513c-5p	NB	PLK4	Jia et al. (2020)
miR-124-3p	ES(2)	CDK4	Lei et al. (2019)
miR-129-5p	OS	DLK1	Zhangmou et al. (2018)
miR-144	NSCLC	PRR11	Huang et al. (2019)
miR-16-5p	NSCLC	BMI1	Wu et al. (2021)
miR-181b	PC	ZEB2	An et al. (2018)
miR-497-5p	PC	FZD4/FZD6/Wnt/beta-catenin	Yang J. et al. (2019)
miR-195-5p	OC	FLH2	Kong and Zhang, (2020)
miR-197-5p	Glioma	E2F1	Li X. et al. (2019)
miR-199a-5p	NPC	HIF-1α	Yang et al. (2020)
miR-199b-5p	TNPC	Paxillin	Du et al. (2020)
miR-203a	HCC	MMP-2	Zhang et al. (2017)
miR-15a-5p	HCC	CXCL17	Wang et al. (2021)
miR-513c	HCC	Cul4A/ANXA10	Liu et al. (2021)
miR-424-5p	HCC	WEEK1	Li D. et al. (2019)
miR-204-5p	GCC	OCT-1	Liang et al. (2020)
miR-4290	GCC	PDK1	Qian et al. (2020)
miR-223	BC(1)	HSP90B1	Fang et al. (2019)
miR-195-5p	BC(1)	VEGFA	Wang et al. (2020)
miR-376c	LC	—	Yang Q. et al. (2019)
miR-505-3p	BC(2)	RUNX2	Zhao et al. (2019)
miR-641	OS	HOXA9	Zhang N. et al. (2019)
miR-497-5p	PC(2)	SNCG	Zhu et al. (2021)
miR-193b-3p	TC	HOXA1	Zheng et al. (2021)

NSCLC, non-small cell lung cancer; LUSC, lung squamous cancer; GCC, gastric cancer; CRC, colorectal cancer; ESCC, esophageal squamous carcinoma; HCC, hepatocellular carcinoma; PC, pancreatic cancer; EC(1),esophageal carcinoma; RCC, renal cell carcinoma; BC(1), bladder cancer; CC, cervical cancer; OC, ovarian cancer; LC, laryngeal carcinoma; NPC, nasopharyngeal carcinoma; OS, osteosarcoma; ES(2), Ewing's sarcoma; NB, neuroblastoma; BC(2), breast cancer; PC(2), prostate cancer; TC, thyroid cancer.

reduces transcriptional activity suppression, so the underlying molecular mechanisms need further exploration.

3.1.2 Interaction With Transcription Factors (Self-Regulation)

Lin et al. (2021) revealed that the hyper-methylated status of the promoter region of the DLX6-AS1 gene might be a novel progression-related and prognostic marker for colorectal cancer. Intriguingly, the methylation of the gene promoter region usually leads to transcriptional silencing. However, the expression level of lncRNA DLX6-AS1 in colorectal carcinoma cell lines and tissues is upregulated. Hence, there must be some other underlying mechanism that remains unknown. Besides, Zhao et al. (2021) also demonstrated that H3K4me1 can induce histone methylation around the DLX6-AS1 promoter region, which could upregulate the expression level of lncRNA DLX6-AS1 in lung squamous cell carcinoma cell lines and lead to cisplatin resistance.

3.2 Post-Transcriptional Regulation

3.2.1 Serving as a ceRNA

LncRNA DLX6-AS1 plays key regulatory roles in regulating tumorigenesis and progress via many mechanisms. One of the most important mechanisms is formation of lncRNA DLX6-AS1-miRNAs-mRNAs ceRNAs regulatory networks, in which lncRNA DLX6-AS1 serves as an endogenous competing RNA to competitively sponge miRNAs to further up-regulate the downstream target genes (mRNAs) of miRNAs (Zeng et al., 2017; Zhang et al., 2017; An et al., 2018; Zhangmou et al., 2018; Li X. et al., 2019; Yang Q. et al., 2019; Zhang H. Y. et al., 2019; Li D. et al., 2019; Yang J. et al., 2019; Zhang N. et al., 2019; Fang et al., 2019; Huang et al., 2019; Lei et al., 2019; Sun et al., 2019; Wang et al., 2019; Zhao et al., 2019; Liu Y. et al., 2020; Du et al., 2020; Hu et al., 2020; Jia et al., 2020; Kong et al., 2020; Kong and Zhang, 2020; Liang et al., 2020; Qian et al., 2020; Wang et al., 2020; Xie et al., 2020; Yang et al., 2020; Liu et al., 2021; Wang et al., 2021; Wu et al., 2021; Zhao et al., 2021; Zheng et al., 2021; Zhu et al., 2021). Xue et al. (2021) and Feng et al. (2021) also reviewed the regulatory roles serving as a ceRNA in various tumors which influence the prognosis of patients and which may be the therapeutic and prognosis target in malignant tumors. **Table 1** summarizes the microRNAs sponged by lncRNA DLX6-AS1 in various tumors.

3.2.2 Stabilizing mRNAs

The incidence and mortality of gastric cancer (GC) rank third and fifth, respectively (Fu et al., 2019), and is usually diagnosed at an

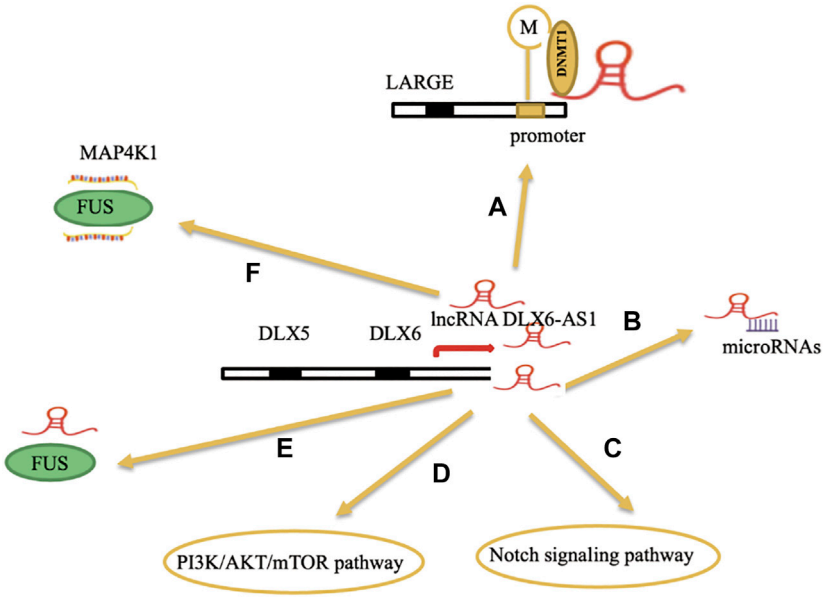


FIGURE 2 | (A) Overexpression of lncRNA DLX6-AS1 could recruit DNA methyltransferase 1 (DNMT1) to the promoter region of its downstream target gene LARGE. **(B)** lncRNA DLX6-AS1 acts as an endogenously competing RNA to sponge miRNAs. **(C,D)** lncRNA DLX6-AS1 can regulate the notch signaling pathway and PI3K/AKT/mTOR pathway in epithelial ovarian cancer and colorectal cancer, respectively. **(E)** lncRNA DLX6-AS1 indirectly binds with MAP4K1(mRNA) *via* FUS(protein) to increase the stability of MAP4K1 in gastric cancer. **(F)** lncRNA DLX6-AS1 interacts with FUS (proteins) and regulate its expression levels in breast cancer.

advanced stage attributing to a lack of typical symptoms. Hence, it is imperative to find molecular biomarkers to assist early prognosis. Wu et al. (2020) found that lncRNA DLX6-AS1 could be a molecular link to stabilize mRNA levels in tumors. In gastric cancer, lncRNA DLX6-AS1 stabilizes MAP4K1 mRNA levels *via* regulating FUS protein expression, forming FUS-MAP4K1 protein-mRNA complexes, by which lncRNA DLX6-AS1 promotes the cell proliferation, migration, and EMT of gastric cancer.

Tian et al. (2020) made a meta-analysis between the expression levels of lncRNA DLX6-AS1 and the clinicopathology and prognosis of various cancers. They found that high expression levels of lncRNA DLX6-AS1 were associated with poor overall survival in tumor patients and overexpression of lncRNA DLX6-AS1 was associated with tumor stage ($p < 0.01$), tumor size ($p < 0.01$), lymph node metastasis ($p < 0.01$), and distant metastasis ($p < 0.01$). Hence, all the aforementioned factors have shown the vital role of research on lncRNA DLX6-AS1 in various tumors (**Figure 2**).

3.2.3 Crosstalk Between Tumor-Associated Macrophages

Wang et al. (2021) found that lncRNA DLX6-AS1 was upregulated in serum exosome derived from patients with hepatocellular carcinoma (HCC) and the density of TAMs in cancer tissues was higher than that in adjacent tissues. They also demonstrated that HCC-exosomes could be delivered to TAMs in a tumor microenvironment by which DLX6-AS1-overexpressed TAMs could further promote the progression of HCC. In the mechanism study, they revealed that after co-culturing HCC-exosomes with monocyte THP-1 cell lines, THP-1 cells are more polarized into M2 macrophages. In addition, after co-injecting hepatocellular cell lines with primed THP-1 cells (THP-1/HCC-exosomes transferred with oe-DLX6-AS1) into the right liver lobe of the mice, they also discovered that primed THP-1 cells (THP-1/HCC-exosomes transferred with oe-DLX6-AS1) could promote HCC lung metastasis *in vivo*. All the above-mentioned observations show that the crosstalk between lncRNA DLX6-AS1 and tumor-associated macrophages (TAMs) could enhance the progression of HCC, which also may provide novel research clues on the regulatory roles of lncRNA DLX6-AS1 in a tumor microenvironment (TME).

3.2.4 Exosomal lncRNA DLX6-AS1

Exosomes (40–100 nm) are key mediators of cell-to-cell communication by transmitting biomolecules such as mRNAs, miRNAs, and lncRNAs (Skog et al., 2008). Several research studies demonstrated that the expression level of exosome-lncRNA DLX6-AS1 in patients with cancer can be a promising prognosis biomarker in tumors.

Zhang et al. reported that the circulating exosome lncRNA DLX6-AS1 expression levels in the serum of patients with NSCLC were significantly higher than those in healthy donors, and its specificity and sensitivity were higher than that of CYFRA21-1 (Ding et al., 2021), which served as a diagnostic marker in NSCLC. They also found that the expression level of exosome lncRNA DLX6-AS1 was positively correlated with tumor differentiation, TNM stage, and lymph vascular invasion, making it a potential early diagnostic and metastatic marker

for NSCLC. Ding et al. enrolled 114 patients with cervical cancer (CC), 60 patients with cervical intraepithelial neoplasia (CIN), and 110 healthy women to their study (Ding et al., 2021). They found that the exosome lncRNA DLX6-AS1 level was elevated in CC patients compared with patients with CIN and normal healthy donors. A high serum exosome lncRNA DLX6-AS1 expression level was positively associated with lymph node metastasis, histopathological differentiation, FIGO stage, and shortened survival of patients with CC (Ding et al., 2021). However, because of the limited number of enrolled patients and high false positive and negative results due to the use of serum exosome lncRNA DLX6-AS1 alone as the prognosis factor of patients with CC, they suggested that a combination with other known tumor biomarkers and clinicopathological parameters was needed to accurately predict the clinical outcome of CC.

3.3 Potential Molecular Biomarker in Tumors: lncRNA DLX6-AS1

As a tumor-promoting gene, lncRNA DLX6-AS1 is often upregulated in many tumors compared with normal tissues. Overexpression of lncRNA DLX6-AS1 could mediate the genesis and progression of many tumors, the knockdown of which could alleviate this phenomenon. Besides, high expression levels of lncRNA DLX6-AS1 are also positively associated with the clinicopathological parameters and negatively with the prognosis of patients with a tumor, which could help with the pathological diagnosis, the survival prediction, and histopathological molecular subtyping. Besides mediating the malignant phenotype of tumor cells, over-expressed lncRNA DLX6-AS1 could also be able to communicate with the tumor microenvironment (TME) such as tumor-associated macrophages (TAMs), which could make it highly expressed in cells in the TME via exosomes and further promote the progress of tumors. In addition, the exosome-lncRNA DLX6-AS1 level could be differently derived from the peripheral blood of patients with a tumor diagnosed at different stages. These features might make it a potential prognosis marker combined with other known tumor molecular markers.

4 LNCRNA DLX6-AS1 IN OTHER DISEASES

4.1 Post-Transcriptional Regulation

lncRNA DLX6-AS1, also named evf1 in the mouse, which involves the development of the mouse's embryonic ventral forebrain, is dysregulated in developing human telencephalon modeling autism spectrum disorder (ASD). Mariani et al. (2015) revealed that lncRNA DLX6-AS1 was one of the top ten up-regulated genes in cerebral organoids derived from members of a family with idiopathic ASD at TD11 and TD31. Wang et al. (2017) also found that DLX6-AS1 and DLX1 were two of three DEGs in $CHD8^{+/-}$ (CHD8 is a highly mutated gene in autism spectrum disorders) cerebral organoids, which increased ~39- and 13-fold, respectively, and were hardly expressed in controls (Wang et al., 2017). In addition, it is well known that evf2 (the spliced form of evf1) forms a complex with DLX1 and DLX2 proteins to facilitate the enhancer activity of DLX5 and DLX6 gene expression. Hence, it supports the potential that lncRNA DLX6-AS1 cooperates with the Dlx family to regulate brain development in humans whose dysregulation is related with nervous system disorders.

lncRNA DLX6-AS1 also serves as a competitively endogenously binding RNA in preeclampsia (PE) and sepsis-induced acute kidney injury (AKI). For instance, Tan et al. (2018) found that high expression levels of lncRNA DLX6-AS1 inhibited proliferation, migration, and invasion of trophoblast JEG3 and HTR-8/SVneo cells *via* directly regulating the miR-376c/GADD45A axis. Interestingly, it has been proven by Liu R. et al. (2020) that the increased lncRNA DLX6-AS1 expression inhibited proliferation, invasion of trophoblast cells of the placenta, and the angiogenesis of HUVEC cells *via* downregulating ERP44 by sponging miR-149-5p in preeclampsia. Tan et al. (2020) reported that the knockout of lncRNA DLX6-AS1 regulated the miR-223-3p/NLRP3 axis in HK-2 cells to induce sepsis-induced acute kidney injury. In addition, Zheng et al. (2022) first demonstrated that over-expressed cAMP-response element binding protein (CREB) induces podocyte injury in diabetic nephropathy via upregulating the lncRNA DLX6-AS1 expression level *in vitro* and *in vivo*.

5 OTHER STUDY METHODOLOGIES OF LNCRNA DLX6-AS1

Most research studies on lncRNA DLX6-AS1 are related to its sponging function to microRNAs as a competitive endogenous RNA via base pairing to each other. MicroRNAs are single-stranded RNAs and generated from endogenous hairpin-shaped transcripts with a length of 19–25 nucleotides (Kim, 2005); they are predicted to regulate the activity of approximately 50% of all protein-coding genes in mammals (Krol et al., 2010). Because of its significant roles in many tumors as a suppressor gene or tumorigenic gene, therapeutics based on microRNAs or anti-microRNAs have been processed to advanced clinical trial stages. Therefore, this may be one of the important reasons for studying the regulatory function of lncRNA DLX6-AS1 in microRNAs.

Due to the biochemical approaches and high-throughput sequencing (Qian et al., 2019), the secondary structure conformations of lncRNAs have been found, and the relationship of lncRNAs between structures and functions should be considered. Hence, the studies of interactions between lncRNA DLX6-AS1 and RNAs, DNAs, or proteins may provide novel research insights into human tumors. We recommend readers to get detailed research clues and inspiration from the articles written by Qian et al. (2019) and Statello et al. (2021).

6 POTENTIAL THERAPEUTIC STRATEGIES OF LNCRNA DLX6-AS1

RNA interference (RNAi) is an endogenous cellular mechanism for regulating gene expression via incorporating siRNAs into

RNA-induced silencing complexes to mediate target genes' cleavage and degradation (Tabernero et al., 2013). lncRNA DLX6-AS1 is usually upregulated in many solid tumor tissues compared to adjacent normal tissues playing a tumorigenic gene role. It is reported that lncRNA DLX6-AS1 is upregulated in many tumor cell lines, and its knockdown can significantly suppress the malignant phenotype of tumor cell lines.

Novel delivery systems can successfully deliver miRNAs to tumor locations without being degraded by RNase in blood and show obvious adverse effects in patients, encapsulating si-lncRNA DLX6-AS1 in lipid nanoparticles to deliver them to tumor sites using the EPR (enhanced permeability and retention) effect on tumor locations (Matsumura et al., 1986). We can also encapsulate si-lncRNA DLX6-AS1 in specific aptamers or target motif-modified nanoparticles to deliver it to tumor locations dependent on ligand–receptor interaction. Simply inhibiting tumor cell malignant phenotypes by delivering si-lncRNA DLX6-AS1 to tumor sites may not show significant tumor inhibitory effects on patients at the advanced stages. Hence, we can encapsulate multiple siRNAs and cooperate with precise photodynamic or photo-thermal therapies to get significant therapeutic effects. However, due to the discontinuous vascular epithelium of the kidneys and spleen similar to immature vascular epithelium in tumor sites, nanoparticles also accumulate easily in these organs and thus may contribute to adverse effects. Nanoparticle drug formulations and pre-clinical trials should be rigorously designed.

7 SUMMARY

Increasing evidence has implied the crucial regulatory roles of lncRNA DLX6-AS1 in various tumors which mainly represent on pre-transcriptional and post-transcriptional levels. One of the most prevalent roles is its sponging function as a ceRNA to form regulatory networks to influence cancer cells' biological characteristics. Besides, LncRNA DLX6-AS1 can also facilitate mRNA stability to influence cell behaviors in tumors and bind with the DNA promoter region to promote methylation of DNAs which further inhibits DNA transcription. Interestingly, in tumors, the upregulation of lncRNA DLX6-AS1 mostly promotes cell proliferation, invasion, and migration, which is the opposite in preeclampsia. Wang et al. (2017) also found that DLX6-AS1 and DLX1 are dysregulated in $CHD8^{+/-}$ cerebral organoids, supporting its developmental regulatory role in human brain development and nervous disorders. The novel biochemical approaches and high-throughput sequencing can provide new insights into the study of the interactions between lncRNA DLX6-AS1 and RNAs, DNAs, or proteins. Along with the rapid development of delivery systems and precisely cooperated therapies, it may provide the possibility of pre-clinical therapeutics based on lncRNA DLX6-AS1 in many tumors.

AUTHOR CONTRIBUTIONS

This review was conceived and drafted by YZ. YZ and PL revised this article. PL supervised the study. All authors agreed to the submission of this article.

FUNDING

This work was supported by the Program for Innovation Research team (in Science and Technology) in Henan University (Grant No. 20IRTSTHN026).

ACKNOWLEDGMENTS

The author thanks PL for her guidance during the completion of this review about lncRNA DLX6-AS1.

REFERENCES

An, Y., Chen, X. M., Yang, Y., Mo, F., Jiang, Y., Sun, D. L., et al. (2018). LncRNA DLX6-AS1 Promoted Cancer Cell Proliferation and Invasion by Attenuating the Endogenous Function of miR-181b in Pancreatic Cancer. *Cancer Cell Int.* 18 (1), 143. doi:10.1186/s12935-018-0643-7

Berghoff, E. G., Clark, M. F., Chen, S., Cajigas, I., Leib, D. E., and Kohtz, J. D. (2013). Evf2 (Dlx6as) lncRNA Regulates Ultraconserved Enhancer Methylation and the Differential Transcriptional Control of Adjacent Genes. *Dev. Camb.* 140 (21), 4407–4416. doi:10.1242/dev.099390

Bond, A. M., Vangompel, M. J. W., Sametsky, E. A., Clark, M. F., Savage, J. C., Disterhoft, J. F., et al. (2009). Balanced Gene Regulation by an Embryonic Brain ncRNA Is Critical for Adult Hippocampal GABA Circuitry. *Nat. Neurosci.* 12 (8), 1020–1027. doi:10.1038/nn.2371

Cajigas, I., Leib, D. E., Cochrane, J., Luo, H., Swyter, K. R., Chen, S., et al. (2015). Evf2 lncRNA/BRG1/DLX1 Interactions Reveal RNA-dependent Inhibition of Chromatin Remodeling. *Development* 142 (15), 2641–2652. doi:10.1242/dev.126318

Cajigas, I., Chakraborty, A., Swyter, K. R., Luo, H., Bastidas, M., Nigro, M., et al. (2018). The Evf2 Ultraconserved Enhancer lncRNA Functionally and Spatially Organizes Megabase Distant Genes in the Developing Forebrain. *Mol. Cell* 71 (6), 956–972. e9. doi:10.1016/j.molcel.2018.07.024

Ding, X. Z., Zhang, S. Q., Deng, X. L., and Qiang, J. H. (2021). Serum Exosomal lncRNA DLX6-AS1 Is a Promising Biomarker for Prognosis Prediction of Cervical Cancer. *Technol. Cancer Res. Treat.* 20 (1128), 1–6. doi:10.1177/15330338219900060

Du, C., Wang, Y., Zhang, Y., Zhang, J., Zhang, L., and Li, J. (2020). LncRNA DLX6-AS1 Contributes to Epithelial-Mesenchymal Transition and Cisplatin Resistance in Triple-Negative Breast Cancer via Modulating Mir-199b-5p/Paxillin Axis. *Cell Transpl.* 29, 963689720929983. doi:10.1177/0963689720929983

Fang, C., Xu, L., He, W., Dai, J., and Sun, F. (2019). Long Noncoding RNA DLX6-AS1 Promotes Cell Growth and Invasiveness in Bladder Cancer via Modulating the miR-223-Hsp90b1 axis. *Cell Cycle* 18 (23), 3288–3299. doi:10.1080/15384101.2019.1673633

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P., and Kohtz, J. D. (2006). The Evf-2 Noncoding RNA Is Transcribed from the Dlx-5/6 Ultraconserved Region and Functions as a Dlx-2 Transcriptional Coactivator. *Genes Dev.* 20 (11), 1470–1484. doi:10.1101/gad.1416106

Feng, L., Wang, R., Wang, Y., Shen, X., Shi, Q., Lian, M., et al. (2021). Silencing Long Non-coding RNA DLX6-AS1 or Restoring microRNA-193b-3p Enhances Thyroid Carcinoma Cell Autophagy and Apoptosis via Depressing HOXA1. *J. Cell Mol. Med.* 25 (19), 9319–9330. doi:10.1111/jcmm.16868

Fu, X., Tian, Y., Kuang, W., Wen, S., and Guo, W. (2019). Long Non-coding RNA DLX6-AS1 Silencing Inhibits Malignant Phenotypes of Gastric Cancer Cells. *Exp. Ther. Med.* 17, 4715–4722. doi:10.3892/etm.2019.7521

Hu, Y., Sun, H., Hu, J., and Zhang, X. (2020). LncRNA DLX6-AS1 Promotes the Progression of Neuroblastoma by Activating STAT2 via Targeting miR-506-3p. *Cancer Manag. Res.* 12, 7451–7463. doi:10.2147/cmar.s252521

Huang, Y., Ni, R., Wang, J., and Liu, Y. (2019). Knockdown of lncRNA DLX6-AS1 Inhibits Cell Proliferation, Migration and Invasion while Promotes Apoptosis by Downregulating PRR11 Expression and Upregulating miR-144 in Non-small Cell Lung Cancer. *Biomed. Pharmacother.* 109, 1851–1859. doi:10.1016/j.biopha.2018.09.151

Jia, P., Wei, E., Liu, H., Wu, T., and Wang, H. (2020). Silencing of Long Non-coding RNA DLX6-AS1 Weakens Neuroblastoma Progression by the miR -513c-5p/PLK4 axis. *IUBMB Life* 72 (12), 2627–2636. doi:10.1002/iub.2392

Kim, V. N. (2005). MicroRNA Biogenesis: Coordinated Cropping and Dicing. *Nat. Rev. Mol. Cell Biol.* 6 (5), 376–385. doi:10.1038/nrm1644

Kohtz, J. D., and Fishell, G. (2004). Developmental Regulation of EVF-1, a Novel Non-coding RNA Transcribed Upstream of the Mouse Dlx6 Gene. *Gene Expr. Patterns* 4 (4), 407–412. doi:10.1016/j.modgep.2004.01.007

Kong, L., and Zhang, C. (2020). LncRNA DLX6-AS1 Aggravates the Development of Ovarian Cancer via Modulating FHL2 by Sponging miR-195-5p. *Cancer Cell Int.* 20 (1), 1–12. doi:10.1186/s12935-020-01452-z

Kong, W-Q., Liang, J-J., Du, J., Ye, Z-X., Gao, P., and Liang, Y-L. (2020). Long Noncoding RNA DLX6-AS1 Regulates the Growth and Aggressiveness of Colorectal Cancer Cells via Mediating miR-26a/EZH2 Axis. *Cancer Biotherapy Radiopharm.* 36 (9), 753–764. doi:10.1089/cbr.2020.3589

Krol, J., Loedige, I., and Filipowicz, W. (2010). The Widespread Regulation of microRNA Biogenesis, Function and Decay. *Nat. Rev. Genet.* 11 (9), 597–610. doi:10.1038/nrg2843

Lei, X., Yang, S., Yang, Y., Zhang, J., Wang, Y., and Cao, M. (2019). Long Noncoding RNA DLX6-AS1 Targets miR-124-3p/CDK4 to Accelerate Ewing's Sarcoma. *Am. J. Transl. Res.* 11 (10), 6569–6576.

Li, D., Tang, X., Li, M., and Zheng, Y. (2019). Long Noncoding RNA DLX6-AS1 Promotes Liver Cancer by Increasing the Expression of WEE1 via Targeting miR-424-5p. *J Cell. Biochem.* 120 (8), 12290–12299. doi:10.1002/jcb.28493

Li, X., Zhang, H., and Wu, X. (2019). Long Noncoding RNA DLX6-AS1 Accelerates the Glioma Carcinogenesis by Competing Endogenous Sponging miR-197-5p to Relieve E2F1. *Gene* 686, 1–7. doi:10.1016/j.gene.2018.10.065

Liang, Y., Zhang, C.-D., Zhang, C., and Dai, D.-Q. (2020). DLX6-AS1/miR-204-5p/OCT1 Positive Feedback Loop Promotes Tumor Progression and Epithelial-Mesenchymal Transition in Gastric Cancer. *Gastric Cancer* 23 (2), 212–227. doi:10.1007/s10120-019-01002-1

Lin, S., Gu, S., Qian, S., Liu, Y., Sheng, J., Li, Q., et al. (2021). Genome-Wide Methylation Profiling of lncRNAs Reveals a Novel Progression-Related and Prognostic Marker for Colorectal Cancer. *Front. Oncol.* 11, 782077. doi:10.3389/fonc.2021.782077

Liu, R., Wang, X., and Yan, Q. (2020). The Regulatory Network of lncRNA DLX6-AS1/miR-149-5p/ERP44 Is Possibly Related to the Progression of Preeclampsia. *Placenta* 93, 34–42. doi:10.1016/j.placenta.2020.02.001

Liu, X., Peng, D., Cao, Y., Zhu, Y., Yin, J., Zhang, G., et al. (2021). Upregulated lncRNA DLX6-AS1 Underpins Hepatocellular Carcinoma Progression via the miR-513c/Cul4A/ANXA10 axis. *Cancer Gene Ther.* 28 (5), 486–501. doi:10.1038/s41417-020-00233-0

Liu, Y., Liu, X., Zhang, X., Deng, J., Zhang, J., and Xing, H. (2020). lncRNA DLX6-AS1 Promotes Proliferation of Laryngeal Cancer Cells by Targeting the miR-26a/TRPC3 Pathway. *Cancer Manag. Res.* 12, 2685–2695. doi:10.2147/cmar.s237181

Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., et al. (2015). FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* 162 (2), 375–390. doi:10.1016/j.cell.2015.06.034

Matsumura, Y., Maeda, H., and Smancs, A. (1986). A New Concept for Macromolecular Therapeutics in Cancer Chemotherapy: Mechanism of Tumoritropic Accumulation of Proteins and the Antitumor Agent Smancs. *Cancer Res.* 46, 6387–6392.

Mir, M., Bickmore, W., Furlong, E. E. M., and Narlikar, G. (2019). Chromatin Topology, Condensates and Gene Regulation: Shifting Paradigms or Just a Phase? *Development* 146 (19), 1–6. doi:10.1242/dev.182766

Olsson, M., Beck, S., Kogner, P., Martinsson, T., and Carén, H. (2016). Genome-wide Methylation Profiling Identifies Novel Methylated Genes in Neuroblastoma Tumors. *Epigenetics* 11 (1), 74–84. doi:10.1080/15592294.2016.1138195

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2019). Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends Biochem. Sci.* 44 (1), 33–52. doi:10.1016/j.tibs.2018.09.012

Qian, Y., Wu, X., Wang, H., Hou, G., Han, X., and Song, W. (2020). MicroRNA-4290 Suppresses PDK1-Mediated Glycolysis to Enhance the Sensitivity of Gastric Cancer Cell to Cisplatin. *Braz J. Med. Biol. Res.* 53 (5), e9330–9. doi:10.1590/1414-431X20209330

Skog, J., Würdinger, T., van Rijn, S., Meijer, D. H., Gainche, L., Sena-esteves, M., et al. (2008). Glioblastoma Microvesicles Transport RNA and Proteins that Promote Tumour Growth and Provide Diagnostic Biomarkers. *Nat. Cell Biol.* 10 (12), 1470–1476. doi:10.1038/ncb1800

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene Regulation by Long Non-Coding RNAs and its Biological Functions. *Nat. Rev. Mol. Cell Biol. [Internet]* 22 (2), 96–118. doi:10.1038/s41580-020-00315-9

Sun, W., Zhang, L., Yan, R., Yang, Y., and Meng, X. (2019). LncRNA DLX6-AS1 Promotes the Proliferation, Invasion, and Migration of Non-small Cell Lung Cancer Cells by Targeting the miR-27b-3p/GSPT1 axis. *Onco Targets Ther.* 12, 3945–3954. doi:10.2147/ott.s196865

Tabernero, J., Shapiro, G. I., LoRusso, P. M., Cervantes, A., Schwartz, G. K., Weiss, G. J., et al. (2013). First-in-humans Trial of an RNA Interference Therapeutic Targeting VEGF and KSP in Cancer Patients with Liver Involvement. *Cancer Discov.* 3 (4), 406–417. doi:10.1158/2159-8290.cd-12-0429

Tan, J., Fan, J., He, J., Zhao, L., and Tang, H. (2020). Knockdown of LncRNA DLX6-AS1 Inhibits HK-2 Cell Pyroptosis via Regulating miR-223-3p/NLRP3 Pathway in Lipopolysaccharide-Induced Acute Kidney Injury. *J. Bioenerg. Biomembr.* 52 (5), 367–376. doi:10.1007/s10863-020-09845-5

Tan, Y., Xiao, D., Xu, Y., and Wang, C. (2018). Long Non-coding RNA DLX6-AS1 Is Upregulated in Preeclampsia and Modulates Migration and Invasion of Trophoblasts through the miR-376c/GADD45A axis. *Exp. Cell Res.* 370 (2), 718–724. doi:10.1016/j.yexcr.2018.07.039

Tian, S., Liu, J., Kong, S., and Peng, L. (2020). LncRNA DLX6-AS1 as a Potential Molecular Biomarker in the Clinicopathology and Prognosis of Various Cancers: a Meta-Analysis. *Biosci. Rep.* 40 (8), BSR20193532. doi:10.1042/BSR20193532

Wang, H., Niu, X., Jiang, H., Mao, F., Zhong, B., Jiang, X., et al. (2020). Long Non-coding RNA DLX6-AS1 Facilitates Bladder Cancer Progression through Modulating miR-195-5p/VEGFA Signaling Pathway. *Aging* 12 (16), 16021–16034. doi:10.18632/aging.103374

Wang, L. P., Lin, J., Ma, X. Q., Xu, D. Y., Shi, C. F., Wang, W., et al. (2021). Exosomal DLX6-AS1 from Hepatocellular Carcinoma Cells Induces M2 Macrophage Polarization to Promote Migration and Invasion in Hepatocellular Carcinoma through microRNA-15a-5p/CXCL17 axis. *J. Exp. Clin. Cancer Res.* 40 (1), 1–16. doi:10.1186/s13046-021-01973-z

Wang, P., Mokhtari, R., Pedrosa, E., Kirschenbaum, M., Bayrak, C., Zheng, D., et al. (2017). CRISPR/Cas9-mediated Heterozygous Knockout of the Autism Gene CHD8 and Characterization of its Transcriptional Networks in Cerebral Organoids Derived from iPS Cells. *Mol. Autism* 8 (1), 11–17. doi:10.1186/s13229-017-0124-1

Wang, X., Lin, Y., and Liu, J. (2019). Long Non-Coding RNA DLX6-AS1 Promotes Proliferation by Acting as a ceRNA Targeting miR-199a in Cervical Cancer. *Mol. Med. Rep.* 19 (2), 1248–1255. doi:10.3892/mmr.2018.9729

Wu, C., Lin, W., and Fu, F. (2021). Long Non-coding RNA DLX6-AS1 Knockdown Suppresses the Tumorigenesis and Progression of Non-small Cell Lung Cancer through microRNA-16-5p/BMI1 axis. *Transl. Cancer Res. TCR* 10 (8), 3772–3787. doi:10.21037/tcr-21-1240

Wu, Q., Ma, J., Meng, W., and Hui, P. (2020). DLX6-AS1 Promotes Cell Proliferation, Migration and EMT of Gastric Cancer through FUS-Regulated MAP4K1. *Cancer Biol. Ther.* 21 (1), 17–25. doi:10.1080/15384047.2019.1647050

Xie, F., Xie, G., and Sun, Q. (2020). Long Noncoding RNA DLX6-AS1 Promotes the Progression in Cervical Cancer by Targeting miR-16-5p/ARPP19 Axis. *Cancer Biotherapy Radiopharm.* 35 (2), 129–136. doi:10.1089/cbr.2019.2960

Xue, C., Lv, L., Jiang, J., and Li, L. (2021). Promising Long Noncoding RNA DLX6-AS1 in Malignant Tumors. *Am. J. Transl. Res.* 12 (12), 7682–7692.

Yang, B., Jia, L., Ren, H., Jin, C., Ren, Q., Zhang, H., et al. (2020). LncRNA DLX6-AS1 Increases the Expression of HIF-1α and Promotes the Malignant

Phenotypes of Nasopharyngeal Carcinoma Cells via Targeting MiR-199a-5p. *Mol. Genet. Genomic Med.* 8 (1), e1017. doi:10.1002/mgg3.1017

Yang, J., Ye, Z., Mei, D., Gu, H., and Zhang, J. (2019). Long Noncoding RNA DLX6-AS1 Promotes Tumorigenesis by Modulating miR-497-5p/FZD4/FZD6/Wnt/β-Catenin Pathway in Pancreatic cancer/FZD6/Wnt/β-Catenin Pathway in Pancreatic Cancer. *Cancer Manag. Res.* 11, 4209–4221. doi:10.2147/cmar.s194453

Yang, Q., Sun, J., Ma, Y., Zhao, C., and Song, J. (2019). LncRNA DLX6-AS1 promotes laryngeal squamous cell carcinoma growth and invasion through regulating miR-376c. *Am. J. Transl. Res.* 11 (11), 7009–7017.

Zeng, X., Hu, Z., Ke, X., Tang, H., Wu, B., Wei, X., et al. (2017). Long Noncoding RNA DLX6-AS1 Promotes Renal Cell Carcinoma Progression via miR-26a/PTEN axis. *Cell Cycle* 16 (22), 2212–2219. doi:10.1080/15384101.2017.1361072

Zhang, H. Y., Xing, M. Q., Guo, J., Zhao, J. C., Chen, X., Jiang, Z., et al. (2019). Long Noncoding RNA DLX6-AS1 Promotes Neuroblastoma Progression by Regulating miR-107/BDNF Pathway. *Cancer Cell Int.* 19 (1), 313. doi:10.1186/s12935-019-0968-x

Zhang, L., He, X., Jin, T., Gang, L., and Jin, Z. (2017). Long Non-coding RNA DLX6-AS1 Aggravates Hepatocellular Carcinoma Carcinogenesis by Modulating miR-203a/MMP-2 Pathway. *Biomed. Pharmacother.* 96, 884–891. doi:10.1016/j.biopha.2017.10.056

Zhang, N., Meng, X., Mei, L., Zhao, C., and Chen, W. (2019). LncRNA DLX6-AS1 Promotes Tumor Proliferation and Metastasis in Osteosarcoma through Modulating miR-641/HOXA9 Signaling Pathway. *J Cell. Biochem.* 120 (7), 11478–11489. doi:10.1002/jcb.28426

Zhangmou, R. M., Tang, T., Yu, H. M., and Yao, X. D. (2018). LncRNA DLX6-AS1/miR-129-5p/DLK1 axis Aggravates Stemness of Osteosarcoma through Wnt Signaling. *Biochem. Biophys. Res. Commun.* 507 (1–4), 260–266. doi:10.1016/j.bbrc.2018.11.019

Zhao, H., and Xu, Q. (2020). Long Non-coding RNA DLX6-AS1 Mediates Proliferation, Invasion and Apoptosis of Endometrial Cancer Cells by Recruiting p300/E2F1 in DLX6 Promoter Region. *J. Cell. Mol. Med.* 24 (21), 12572–12584. doi:10.1111/jcmm.15810

Zhao, P., Guan, H., Dai, Z., Ma, Y., Zhao, Y., and Liu, D. (2019). Long Noncoding RNA DLX6-AS1 Promotes Breast Cancer Progression via miR-505-3p/RUNX2 axis. *Eur. J. Pharmacol.* 865, 172778. doi:10.1016/j.ejphar.2019.172778

Zhao, X., Wang, J., Zhu, R., Zhang, J., and Zhang, Y. (2021). DLX6-AS1 Activated by H3K4me1 Enhanced Secondary Cisplatin Resistance of Lung Squamous Cell Carcinoma through Modulating miR-181a-5p/miR-382-5p/CELF1 axis. *Sci. Rep.* 11 (1), 1–13. doi:10.1038/s41598-021-99555-8

Zhao, Z., Liang, S., and Sun, F. (2020). LncRNA DLX6-AS1 Promotes Malignant Phenotype and Lymph Node Metastasis in Prostate Cancer by Inducing LARGE Methylation. *Front. Oncol.* 10, 1172. doi:10.3389/fonc.2020.01172

Zheng, Q., Gu, X., Yang, Q., Chu, Q., Dai, Y., and Chen, Z. (2021). DLX6-AS1 Is a Potential Biomarker and Therapeutic Target in Cancer Initiation and Progression. *Clin. Chim. Acta* 517 (79), 1–8. doi:10.1016/j.cca.2021.02.006

Zheng, W., Guo, J., Lu, X., Qiao, Y., Liu, D., Pan, S., et al. (2022). cAMP-Response Element Binding Protein Mediates Podocyte Injury in Diabetic Nephropathy by Targeting lncRNA DLX6-AS1. *Metabolism* 129, 155155. doi:10.1016/j.metabol.2022.155155

Zhu, X., Ma, X., Zhao, S., and Cao, Z. (2021). DLX6-AS1 Accelerates Cell Proliferation through Regulating miR -497-5p/SNCG Pathway in Prostate Cancer. *Environ. Toxicol.* 36 (3), 308–319. doi:10.1002/tox.23036

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

BRIEF RESEARCH REPORT
published: 06 June 2022
doi: 10.3389/fgene.2022.863956

Clinical Benefits of Olaparib in Mexican Ovarian Cancer Patients With Founder Mutation *BRCA1*-Del ex9-12

Dolores Gallardo-Rincón[1,2†], Edgar Montes-Servín[1†], Gabriela Alamilla-García[1,2], Elizabeth Montes-Servín[1], Antonio Bahena-González[1,2], Lucely Cetina-Pérez[3,4], Flavia Morales Vásquez[2], Claudia Cano-Blanco[2], Jaime Coronel-Martínez[2], Ernesto González-Ibarra[1], Raquel Espinosa-Romero[1,2], Rosa María Alvarez-Gómez[3,5], Abraham Pedroza-Torres[5,6] and Denisse Castro-Eguiluz[4,6]*

[1]Ovarian and Endometrial Cancer Program (COE), Instituto Nacional de Cancerología (INCan), Mexico City, Mexico, [2]Department of Medical Oncology, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico, [3]Department of Clinical Research and Medical Oncology, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico, [4]Cervical Cancer Program (Micaela), Instituto Nacional de Cancerología (INCan), Mexico City, Mexico, [5]Hereditary Cancer Clinic, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico, [6]Catedrático CONACYT, Instituto Nacional de Cancerología, Mexico City, Mexico

Background: Ovarian cancer (OC) is gynecologic cancer with the highest mortality rate. It is estimated that 13–17% of ovarian cancers are due to heritable mutations in *BRCA1* and *BRCA2*. The *BRCA1* (*BRCA1*-Del ex9-12) Mexican founder mutation is responsible for 28–35% of the cases with ovarian cancer. The aim was to describe the PFS of OC patients treated with olaparib, emphasizing patients carrying the Mexican founder mutation (*BRCA1*-Del ex9-12).

Methods: In this observational study, of 107 patients with *BRCA*m, 35 patients were treated with olaparib from November 2016 to May 2021 at the Ovarian Cancer Program (COE) of Mexico; patient information was extracted from electronic medical records.

Results: Of 311 patients, 107 (34.4%) were with *BRCA*m; 71.9% (77/107) were with *BRCA1*, of which 27.3% (21/77) were with *BRCA1*-Del ex9-12, and 28.1% (30/107) were with *BRCA2* mutations. Only 35 patients received olaparib treatment, and the median follow-up was 12.87 months. The PFS of *BRCA1*-Del ex9-12 was NR (non-reach); however, 73% of the patients received the treatment at 36 vs. 11.59 months (95% CI; 10.43–12.75) in patients with other *BRCA*m ($p = 0.008$). Almost 50% of patients required dose reduction due to toxicity; the most frequent adverse events were hematological in 76.5% and gastrointestinal in 4%.

Conclusion: Mexican OC *BRCA1*-Del ex9-12 patients treated with olaparib had a significant increase in PFS regardless of the line of treatment compared to other mutations in *BRCA*.

Keywords: epithelial ovarian cancer, Mexican founder mutation, large rearrangements, progression-free survival, *BRCA* mutation

OPEN ACCESS

Edited by:
Jian-Guo Zhou, University of Erlangen Nuremberg, Germany

Reviewed by:
Anna Sokolenko, N.N.Petrov Research Institute of Oncology, Russia
Stergios Boussios, King's College London, United Kingdom

****Correspondence:***
Dolores Gallardo-Rincón
dgallardorincon@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *27 January 2022*
Accepted: *02 May 2022*
Published: *06 June 2022*

Citation:
Gallardo-Rincón D, Montes-Servín E, Alamilla-García G, Montes-Servín E, Bahena-González A, Cetina-Pérez L, Morales Vásquez F, Cano-Blanco C, Coronel-Martínez J, González-Ibarra E, Espinosa-Romero R, María Alvarez-Gómez R, Pedroza-Torres A and Castro-Eguiluz D (2022) Clinical Benefits of Olaparib in Mexican Ovarian Cancer Patients With Founder Mutation BRCA1-Del ex9-12. Front. Genet. 13:863956. doi: 10.3389/fgene.2022.863956

INTRODUCTION

Among gynecologic cancers, ovarian cancer (OC) has the highest mortality rate. Epithelial ovarian cancer is the most lethal gynecologic malignancy, as it is commonly diagnosed at an advanced stage and only 10% of all OC is non-epithelial (include mainly germ cell tumors, sex cord–stromal tumors, and some rare tumors) (Boussios et al., 2017). According to GLOBOCAN estimates, in 2020, there were 313,959 new cases and 207,252 deaths worldwide. In Mexico, the estimated number of new cases and deaths for the same year were 4,963 and 3,038, respectively (The Global Cancer Observatory, 2020).

The recent addition of poly (ADP-ribose) polymerase (PARP) inhibitor (PARPi) as a treatment option has caused a paradigm shift in the management of OC patients. PARP prevents the repair of single-stranded DNA breaks and, coupled with a deficiency in repair by homologous recombination, causes synthetic lethality and cell death (Weaver and Yang, 2013). Olaparib, niraparib, and rucaparib are novel oral PARPi agents that have become a standard of care in different clinical settings, such as maintenance therapy after platinum-sensitive recurrence with either partial or complete response or after frontline therapy. Although clinical trials have demonstrated the efficacy of PARPi in the absence of homologous recombination deficiency, patients with *BRCA* gene mutations achieve better outcomes (Coleman et al., 2017; Del Campo et al., 2019; González-Martín et al., 2019). However, apart from mutations in the *BRCA1/2* genes, there are other genomic alterations involving genes in homologous recombination pathways like the Fanconi anemia genes (*BRIP1* and *PALB2*), the core RAD genes (*RAD51C* and *RAD51D*), and genes involved directly (*CHEK2*, *BARD1, NBN*, and *ATM*) or indirectly (*CDK12*). The genome-wide association studies identified single-nucleotide polymorphisms associated with susceptibility for epithelial OC, for example, 27 loci are associated with invasive epithelial OC identified so far account for 6.4% of the polygenic risk for epithelial OC (Boussios et al., 2020).

Mutations in *BRCA1/2* occur in 1 out of 300–500 women, increasing their risk of developing various types of cancer, predominantly breast and ovarian cancer (Zhang et al., 2011; Toss et al., 2015). It is estimated that 13–17% of OC are due to heritable mutations in *BRCA1* and *BRCA2* (Hennessy et al., 2010; Cancer Genome Atlas Research Network, 2011). In addition, 3–7% of OC patients harbor a somatic mutation of the *BRCA* genes (Cunningham et al., 2014; Pennington et al., 2014). There are previous reports of *BRCA* mutation frequency in Mexican OC patients. The first study carried out by Villarreal-Garza et al. (2015a) tested *BRCA* mutations (using HISPANEL) in 188 non-related patients (92 with OC and 96 with breast cancer (BC)). This study reported that *BRCA* mutations were detected in 28% of OC patients and most of the mutations were in *BRCA1* (88%). Gallardo-Rincón et al. (2020) studied 179 OC patients for germline *BRCA* mutations through next-generation sequencing and multiplex ligation-dependent probe amplification. In this study, 33% of patients had a germline mutation and 66% of these were found in *BRCA1*. In addition, the most frequent mutation for Mexican *BRCA* mutation carriers was the deletion of exons 9 to 12 in *BRCA1* (*BRCA1*-Del ex9-12) representing the 28% (11/39) of *BRCA1*-mutated patients. Other studies support these data in BC and OC patients combined (Vaca-Paniagua et al., 2012; Quezada Urban et al., 2018; Oliver et al., 2019).

The *BRCA1* Mexican founder mutation (*BRCA1*-Del ex9-12 or NM_007294.3: c.548-?_4,185+?del) is related to a clear founder effect (Weitzel et al., 2005; Weitzel et al., 2007; Weitzel et al., 2013). The previously mentioned epidemiological studies reported that the founder mutation accounts for 28–35% of *BRCA* gene mutations in Mexican OC (Cunningham et al., 2014; Villarreal-Garza et al., 2015a). The Mexican founder mutation is a large rearrangement (exon deletion). Previous reports suggest the possibility that large rearrangements represent a type of *BRCA* gene mutation with greater penetrance for cancer risk, as it correlates with earlier onset age or more aggressive tumors in BC and OC patients (James et al., 2015; Kwong et al., 2015). This molecular feature could have a meaningful clinical impact on screening, prognosis, and treatment in the case of PARPi. The aim of this study was to describe the survival rate of OC patients treated with olaparib, emphasizing patients carrying the Mexican founder mutation (*BRCA1*-Del ex9-12).

MATERIALS AND METHODS

Study Design

In this single-center observational study, data analysis was carried out from retrospectively collected samples with prospectively followed up. A total of 311 OC patients in clinical stages (CS) from IA to IVB were enrolled from October 2015 to May 2021 at the Instituto Nacional de Cancerología (INCan) of Mexico. All patients provided written informed consent before entering the study. Of 311 OC patients, 35 were treated with olaparib at the Ovarian and Endometrial Cancer Program (COE) at INCan, from November 2016 to May 2021.

Patients

Patient inclusion criteria were: 1. histopathology confirmed diagnosis of epithelial ovarian cancer platinum-sensitive, at any clinical stage. 2. *BRCA1/2* germinal mutation. 3. Partial or complete objective response (either according to response evaluation criteria in solid tumors (RECIST) version 1.1 or patients with stable disease with a decreased level of CA-125 4 of olaparib maintenance treatment starting 4–8 weeks after the last chemotherapy cycle. In November 2020, the first-line olaparib maintenance treatment in OC patients was started. Olaparib dose was 800 mg/day, as maintenance therapy until progression; dose adjustment administration was allowed in grade 2 or more adverse events.

Study Endpoints and Assessments

The primary objective was to evaluate the benefit of olaparib in Mexican OC patients by describing their PFS according to *BRCA1* founder mutation. PFS was defined as the time from the beginning of treatment with olaparib to disease progression, death, or the last contact at a cutoff date of 31 May 2021.

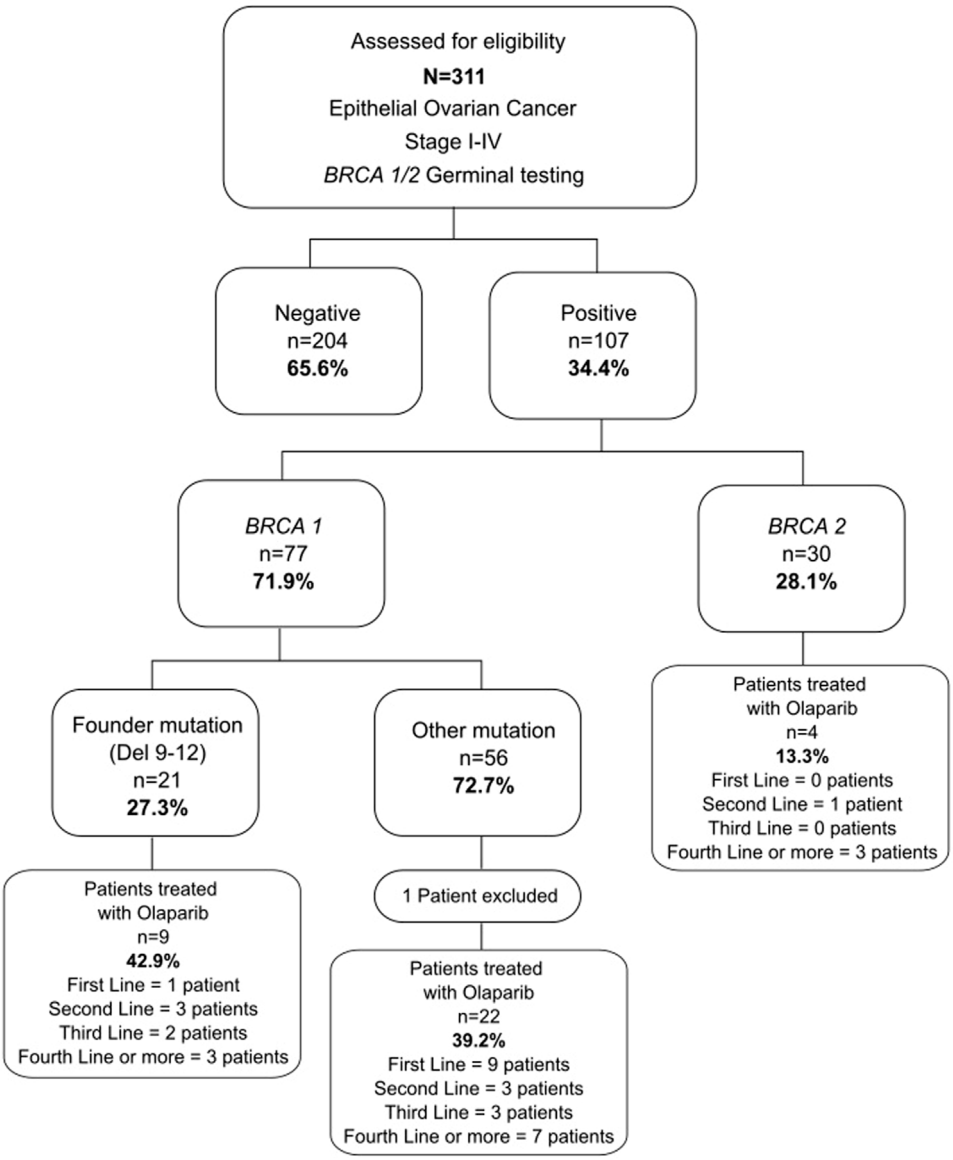


FIGURE 1 | Patient enrollment flowchart. Flowchart summarizes patient enrollment and sub-analysis groups.

Baseline clinicopathological characteristics such as age, histology, stage assigned at diagnosis, and toxicity were extracted from electronic medical records.

Statistical Analyses

Continuous variables were tabulated as medians with ranges or as means with standard deviations (SDs), depending on the data distribution. The distribution was assessed using the Shapiro–Wilk test with a *p*-value greater than 0.05 considered as normally distributed. Two-group comparisons were tested using Student's t-test or Mann–Whitney U test depending on the data distribution. Nominal data were analyzed using the chi-squared (X^2) test. Median PFS curves were estimated using the Kaplan–Meier method, while comparisons among groups were analyzed with log-rank or Breslow tests. Statistical significance was determined as $p \leq 0.05$ with a two-sided test. All data were analyzed using the SPSS software package version 26 (SPSS, Inc., Chicago, Ill, United States) and GraphPad Prism version 9.0 (GraphPad San Diego, CA, United States).

RESULTS

Presence of Germinal *BRCA* Mutations

Of 311 OC patients, 107 (34.4%) had a germinal *BRCA*m, of which 71.9% (77/107) were *BRCA1* and 28.1% (30/107) were *BRCA2* mutations. Among these patients, the most common pathogenic variant detected in 21 (27.3%) was *BRCA1*-Del

TABLE 1 | Baseline characteristics.

Variable	Total % (n = 35)	*BRCA1* (Founder mutation) % (n = 9)	Other *BRCA* mutation % (n = 26)	P[a]
Age at diagnosis (years)				
Mean ± S. D	51.31 ± 7.38	50 ± 7.08	51.77 ± 7.57	0.544
Median (range)	51 (40–69)	52 (40–63)	51 (40–69)	0.677
Histology				
HGS	94.2 (33/35)	100 (9/9)	92.4 (24/26)	0.693
Adenocarcinoma	2.9 (1/35)	0 (0 (9)	3.8 (1/26)	
Others	2.9 (1/35)	0 (0 (9)	3.8 (1/26)	
Stage				
I–II	11.4 (4/35)	0 (0/9)	15.4 (4/26)	0.084
IIIA–B	8.6 (3/35)	22.22 (2/9)	3.8 (1/26)	
IIIC	54.3 (19/35)	33.33 (3/9)	61.5 (16/26)	
IV	25.7 (9/35)	44.44 (4/9)	19.2 (5/26)	
DM				
Negative	88.6 (31/35)	100 (9/9)	84.6 (22/26)	0.211
Positive	11.4 (4/35)	0 (0 (9)	15.4 (4/26)	
SAH				
Negative	82.9 (29/35)	88.9 (8/9)	80.8 (21/26)	0.577
Positive	17.1 (6/35)	11.1 (1/9)	19.2 (5/26)	
CFH				
Negative	11.4 (4/35)	0 (0/9)	15.4 (4/26)	0.211
Positive	88.6 (31/35)	100 (9/9)	84.6 (22/26)	
Ovarian CFH				
Negative	65.7 (23/35)	77.8 (7/9)	61.5 (16/26)	0.376
Positive	34.3 (12/35)	22.2 (2/9)	38.5 (10/26)	
Breast CFH				
Negative	34.3 (12/35)	11.1 (1/9)	42.3 (11/26)	0.089
Positive	65.7 (23/35)	88.9 (8/9)	57.7 (15/26)	
Other CFH				
Negative	57.1 (20/35)	44.4 (4/9)	61.5 (16/26)	0.372
Positive	42.9 (15/35)	55.6 (5/9)	38.5 (10/26)	
Pancreatic	26.7 (4/15)	40 (2/5)	20 (2/10)	0.638
Prostate	33.3 (5/15)	20 (1/5)	40 (4/10)	
Gastrointestinal	40 (6/15)	40 (2/5)	40 (4/10)	
Double primary malignancy				
Negative	77.1 (27/35)	55.6 (5/9)	84.6 (22/26)	0.074
Positive (breast–ovarian)	22.9 (8/35)	44.44 (4/9)	15.4 (4/26)	
CA-125 at starting treatment line (U/mL)				
Mean ± S. D	1009.91 ± 2177.25	2099.08 ± 3968.29	632.88 ± 912.66	0.303
Median (range)	205 (17.40–9244.70)	108.60 (17.40–9244.70)	245.45 (28.00–4,019.40)	0.521
CA-125 at start olaparib (U/mL)				
Mean ± S. D	24.40 ± 32.83	13.80 ± 6.08	21.20 ± 42.63	0.250
Median (range)	14.30 (5.60–194.40)	13.80 (9.50–18.10)	32.47 (6.11–194.40)	0.664
CA-125 at progression to olaparib (U/mL)				
Decrease	133.28 ± 248.76	12.25 ± 5.86	146.02 ± 258.76	**0.037**
Increase	52.10 (8.10–1092.30)	12.25 (8.10–16.40)	56.10 (10.90–1092.30)	0.072
Δ CA-125 (treatment-olaparib)				
Decrease	100 (35/35)	100 (9/9)	100 (26/26)	--
Increase	0 (0/35)	0 (0/9)	0 (0/26)	
Δ CA-125 (olaparib-progression)				
Decrease	19 (4/21)	100 (2/2)	10.5 (2/19)	**0.002**
Increase	81 (17/21)	0 (0/2)	89.5 817/19)	

[a]*Chi-squared (X^2) test.*

Abbreviations: CBP: carboplatin; TXL: taxol; CDDP: cisplatin; GMZ: gemcitabine; BVZ: bevacizumab; HGS: high-grade serous; DM: diabetes mellitus; SAH: systemic arterial hypertension; CFH: cancer family history.

Note: Bold numbers are statistically significant values

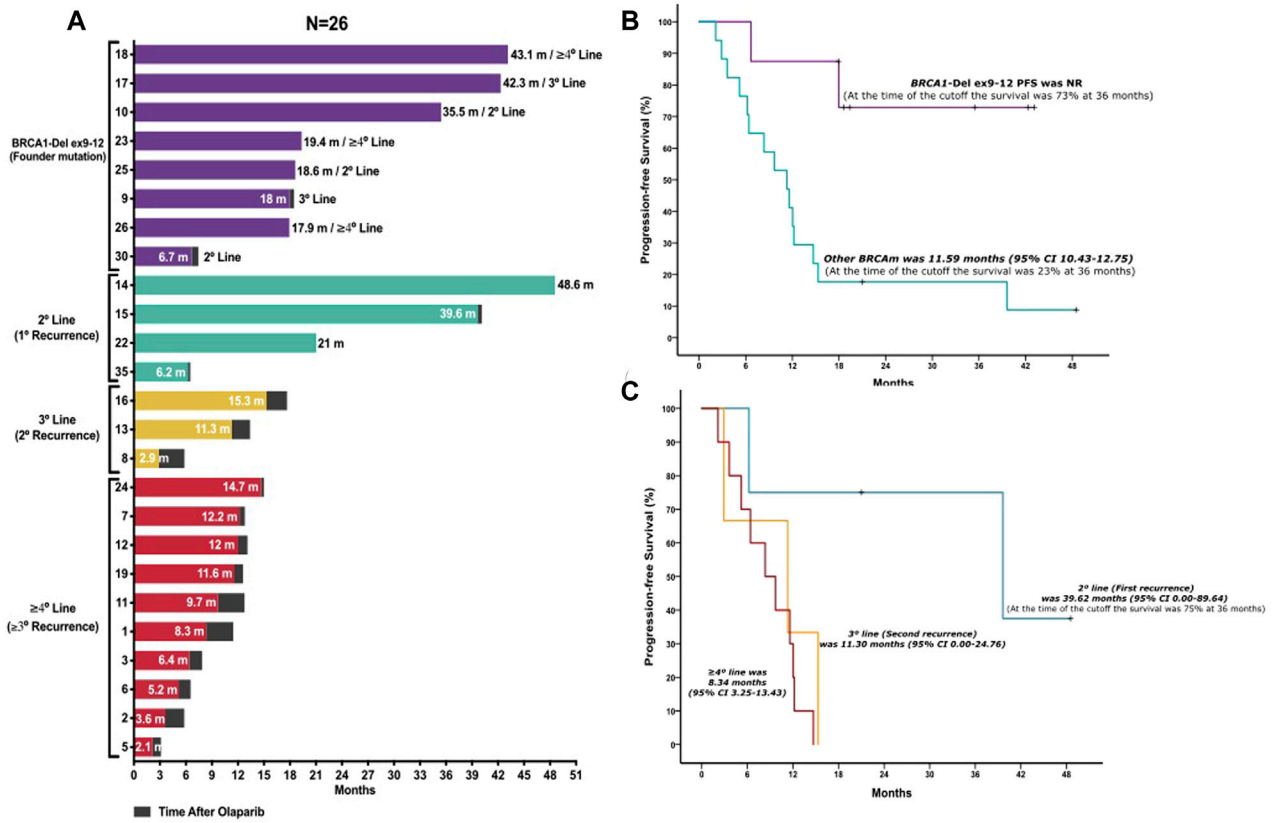


FIGURE 2 | Patient PFS analysis. **(A)** Histogram per patient shows the follow-up of each patient undergoing olaparib treatment and time after progression. Patients are grouped according to the presence of the Mexican founder mutation and other mutations organized by the lines of treatment received before olaparib maintenance therapy. Patients with *BRCA1* founder mutation (*BRCA1*-Del ex9-12) (purple bars); patients with other *BRCA* mutations: second line (turquoise bars); third line (yellow bars); and patients in fourth line or more (red bars). **(B)** Progression-free survival comparison between *BRCA1*-Del ex9-12 and other *BRCA* mutations of patients undergoing olaparib maintenance therapy. Kaplan–Meier curve of PFS. Patients with *BRCA1* founder mutation (*BRCA1*-Del ex9-12) (purple line); patients with other *BRCA* mutations (turquoise line). **(C)** Progression-free survival comparison between lines of treatment of the other *BRCA* mutations of patients undergoing olaparib maintenance therapy. Kaplan–Meier curve of PFS. Patients in second line (turquoise line), third line (yellow line), and patients in fourth line or more (red line).

ex9-12 (Mexican founder mutation). The patient enrollment, testing flowchart, and line of treatment subgroups are summarized in **Figure 1**.

Patient Characteristics

A total of 36 patients received olaparib, and only 35 patients were analyzed in this study; one patient was excluded because she received gemcitabine instead of platinum-based chemotherapy before olaparib treatment (platinum hypersensitivity). The median age was 51 years (range, 40–69); 33 patients were with HGSP histology (94.2%), and 19 patients (54.3%) had a clinical-stage IIIC disease. Most of the patients did not present comorbidities like diabetes mellitus and systemic arterial hypertension (88.6 and 82.9%, respectively). Patients confirmed with the first and second grade of cancer family history (CFH) were 31 (88.6%), 12 patients (34.3%) for OC and 23 patients (65.7%) for BC, and 15 patients (42.9%) were positive for other *BRCA*-associated cancer types (prostate, pancreatic, and gastrointestinal). Almost 23% (8 patients) had double primary malignancy (breast–ovarian) (**Table 1**). The median of follow-up of the 35 patients was 12.87 months. By the time of data analysis, 21 patients (60%) had disease progression to olaparib maintenance therapy, and 14 patients (40%) were still receiving olaparib treatment (**Figure 2A**). There were no statistical differences in the clinical characteristics according to the *BRCA*m, which are visualized in **Table 1**. Genetic variants of *BRCA1/2* were classified according to the prevalence in the cohort of patients, the ovarian cancer cluster regions (OCCRs), and breast cancer cluster regions (BCCRs) in both genes (*BRCA1/2*) **Table 2**.

Olaparib Maintenance Therapy Patient Characteristics

In total, 35 patients received olaparib maintenance therapy after platinum-based chemotherapy; 91.4% had a complete or partial response (n = 32), and 3 patients had stable disease (8.6%) before starting olaparib maintenance therapy. The number of patients treated with olaparib after the first line was 10 (27.8%) with a mean follow-up of 10.55 months. The platinum-sensitive,

TABLE 2 | *BRCA* mutation

Patient ID	Variant/population(s) reported	Cluster region	Gene
36	**Del ex9-12/(AP: Mexican founder mutation)**	**OCCR**	***BRCA1***
30			
10			
25			
17			
9			
26			
23			
18			
32	c.1960A > T (p. Lys654Ter)		
8			
1			
14	**c.1674del (p. Gly559fs)**		
	(AP: Colombian founder mutation)		
34	c.211A > G (p. Arg71Gly)		
21	c.2296_2297del (p.Glu765_Ser766insTer)		
2	c.2611+1G > T		
7	c.3598 C > T (p. Gin1200*)		
28	c.3759_3760del (p. Lys1254fs)		
20	c.4327C > T (p. Arg1443Ter)		
16	c.5123C > A (p. Ala1708Glu)		
33	c.5165C > T (p. Ser1722Phe)		
31	c.5278-1G > C		
27	c.4868 C > G (p. ALA1623GLY)	**BCCR**	
13			
5			
22	**c.68_69del (p. Glu23fs)**		
24	**(AP: Ashkenazi-Jewish)**		
29	c.1504_1508del (p. Leu502fs)		
35	c.68_69dupAG (p. Cys24Serfs)		
11	c.815_824dup (p. Thr276fs)		
6	Del ex18-19		
3	c.5616_5620delAGTAA (p. Lys1872Asnfs)	**OCCR**	***BRCA2***
15	c.5631delC (p. Asn1877Lysfs)		
19	c.6352_6353delGT (p. Val2118Lysfs)	**BCCR**	
12	c.8168A > G (p. Asp2723Gly)		

Abbreviations: OCCRs: ovarian cancer cluster regions; BCCRs: breast cancer cluster regions; AP: associated population.
Note: Bold numbers are statistically significant values

relapsed patients treated with olaparib at the second line (first recurrence) were 7 (20%), 5 patients (14.3%) at the third line (second recurrence), and 13 patients at fourth or more line of treatment (37.1%), all with a mean follow-up of 17.29 months.

Analyzing the patients with recurrent disease (first-line treated patients excluded), a platinum-free interval status was evaluated in 26 patients, 40% (14 patients) had a response of 6–12 months, and 31.4% (11 patients) had a response of 12 months or higher. Most of the mutations (68.6%) are located in the areas known as the OC cluster regions (OCCRs), and 31.4% are located in the BC cluster regions (BCCRs) in both genes (*BRCA1/2*). There were no statistical differences in the clinical characteristics of patients that received olaparib maintenance therapy according to the *BRCA*m which are visualized in S1.

Progression-Free Survival Analysis

The median follow-up of the 35 patients was 12.8 months (95% CI 8.82–16.92). The only baseline characteristic associated with olaparib PFS was breast CFH; these patients had a better survival (11.59 vs. 17.97 months $p = 0.036$). There were no statistical differences in the baseline characteristics associated with *BRCA*m as shown in S2. The median PFS of positive founder mutation *BRCA1*-Del ex9-12 was NR (at the time of cutoff, the survival was 73% at 36 months) vs. 11.59 months (95% CI 10.43–12.75) in those with other *BRCA*m detected ($p = 0.008$) (**Figure 2B** and **Supplementary Figure S3**). The PFS from the patients with positive founder mutation *BRCA1*-Del ex9-12 shows a significant increase regardless of the line of treatment in which they received the treatment compared to other mutations in *BRCA*. The median PFS of other *BRCA*m detected treated with olaparib after the first line was 12.87 months; also, 39.62 months for the patients treated at the second line (first recurrence), 11.30 months for patients at the third line (second recurrence), and 8.34 months for patients at the fourth or more lines of treatment were reported (**Figure 2C** and **Supplementary Figure S3**). The group of other *BRCA* mutations showed that patients with a complete or partial response before olaparib maintenance therapy had a better PFS than patients with

stable disease ($p = 0.008$). Also, multi-treated patients (≥4°L) had the worst PFS among the other lines of treatment ($p = 0.029$) (S3).

Toxicity

Toxicity adverse events were obtained from the 35 analyzed patients. Of all patients, 48.6% (17/35) required dose reduction due to some adverse event of any grade (most of these patients were in the third or more lines of treatment). The most frequent adverse events in the patients with dose reduction were hematological in 76.5% (13 patients) and gastrointestinal in 23.5% (4 patients) (S1).

The use of olaparib was associated with neutropenia in 1 case with grade 3 and anemia in 15 patients (42.8%). Grade 2 anemia in 33.4% (5/15 patients), grade 3 in 46.6% (7/15 patients), and grade 4 in 20% (3/15 patients) were developed. On the other hand, 75% (3/4) of the patients had nausea grade 2 (2 patients), grade 3 in 1 patient, and 1 patient with dysgeusia grade 2. Other adverse event recorded in this cohort was pneumonitis in only 1 case associated with previous breast radiotherapy treatment. Expected adverse events related to the use of olaparib, such as myelodysplastic syndrome (MDS), occurred in 1 patient (toxicity events by subgroups are summarized in S4).

DISCUSSION

The *BRCA1*-Del ex9-12 mutation is related to a founder effect in the Mexican population. Epidemiological studies reported that this founder mutation represents the 28–39% of *BRCA1* gene mutations in Mexican OC patients. In addition, another frequent mutation was present in the OCCR, *BRCA1* c.1970A > T (p.Lys654Ter) at 8.6%. This mutation predicts loss of normal protein function through either protein truncation or nonsense-mediated mRNA decay (Judkins et al., 2005); Weitzel et al. (2005) reported that this mutation is associated with a high risk of developing cancer and is considered a frequent mutation in the Mexican population. The most common mutation in the BCCR of the *BRCA1* gene is c.4868C > G (p.Ala1623Gly) at 8.6%, which is associated with a partial deletion in the exon 15, which is a rare mechanism of splicing alteration (Byers et al., 2016). This specific mutation is associated with a risk of more aggressive breast cancer in men, but its effect in ovarian cancer patients is unknown (Alsop et al., 2012).

Other founder effects have been reported in Latin American countries, such as Brazil (*BRCA1* 5382insC and *BRCA2* c.156_157insAlu) and Colombia (*BRCA1* 3450del4, *BRCA1* A1708E, and *BRCA2* 3034del4) (Ossa and Torres, 2016). Of these, *BRCA1* 3450del4 mutation has also been reported in Brazil and Chile, whereas mutation *BRCA2* 3034del4 has been reported in Argentina and Peru. These data imply that Hispanic (Latin American) populations share common genetic ancestry components from Europe, Africa, and Native Americans which are also genetically heterogeneous (Bryc et al., 2010). To our knowledge, this is the first report of the association of these specific mutations with survival and other outcomes in OC patients.

In our populations, most of the mutations detected in *BRCA* genes were point mutations. The *BRCA1*-Del ex9-12 mutation and *BRCA1*-Del ex18-19 represent the only cases of large rearrangement (exon deletion). Large gene rearrangements (LGRs) represent less than 10% of *BRCA1* pathogenic variants (Sluiter and van Rensburg, 2011). Latin American patients report a prevalence of nearly 21%, similar to Dutch (27%) and Italian (20%) populations (Judkins et al., 2012). We identified that *BRCA1*-Del ex18-19 was detected in a single patient. LGRs in the *BRCA* gene are associated with greater penetrance for cancer risk and correlate with an earlier onset age of cancer or more aggressive tumors (James et al., 2015). Due to the large-scale sequencing efforts, there is currently a better understanding of the genomic landscape of several malignancies, for example, the incidence of germline *BRCA* mutations in newly diagnosed prostate cancer patients is 1.2–2%, and the *BRCA1/2* carriers can have around 4- and 8-fold risk of developing prostate cancer, respectively. So, the importance of detection and the identification of defects in DNA repair genes have led to clinical studies that provide a strong rationale for developing PARPi and DNA-damaging agents in this molecularly defined subset of patients (Ghose et al., 2021).

It has been demonstrated that triple-negative breast cancer patients with the Mexican founder mutation have the worst outcome (Villarreal-Garza et al., 2015b). In agreement with this report; of eight patients that present a double primary malignant neoplasm, four patients were founder-mutated with triple-negative breast cancer and two patients had progressive disease. Probably, the main reason regarding the difference in survival between patients with small-scale *BRCA* mutations and LGRs relies on the resistance mechanisms to PARPi, such as olaparib. Among the most important is mutational reversion, restoring homologous recombination repair of DNA double-strand breaks (Banda et al., 2018).

In this way, *BRCA1* 9–12 exon deletion represents the loss of more than 60% of the gene-coding region, so it would be complex to opt for a mutational reversion resistance mechanism to restore the wild-type allele and thus correlate with the better response and survival. Similar observations have been reported in patients with *BRCA* LGRs and their response to platinum and PARPi (Randall et al., 2020; Wang et al., 2022).

The median PFS reported in the SOLO2 trial in first platinum-sensitive, relapsed *BRCA*-mutated OC patients treated with olaparib was higher (19.1 months [95% CI 16.3–25.7]) than that in the placebo (5.5 months [5.2–5.8]; hazard ratio [HR] 0.30 [95% CI 0.22–0.41], $p < 0.0001$) with 22.1 months of follow-up (Pujade-Lauraine et al., 2017); in our study, the median PFS for the first recurrent platinum-sensitive OC patients like the SOLO2 trial patient characteristics (n = 25) was 39.68 months with a median follow-up of 17.2 months.

The median PFS reported in the SOLO1 trial in first-line maintenance therapy OC patients with *BRCA1/2* mutation was 60% at 3 years of follow-up (Moore et al., 2018); in our study, the median PFS for positive Mexican founder mutation in *BRCA1* was NR, but 73% at 3 years vs. 11.59 months (95% CI 10.43–12.75) in those with other *BRCA*m were detected. (p =

0.004). Our results are consistent and confirmed the preliminary results that were reported by Gallardo-Rincón et al. (2019).

Regarding the use of olaparib in the first line of treatment, our data are immature, and we are still recruiting patients with the founder mutation to compare their survival with other *BRCA* mutations. At the time of the cut-off, 10 patients were receiving maintenance treatment after the first line, and the median PFS was 12.87 months with a short follow-up of 10.5 months. *BRCA1*-Del ex9-12-mutated patients that required dose reduction reported fewer adverse events associated with olaparib treatment than other *BRCA*m patients (33 vs. 52%, respectively). Multi-treated patients (≥4°L) reported more and higher toxicity effects.

Despite the limited number of patients that received olaparib treatment (N = 35), the obtained results are precise on its clinical benefit for patients with PARPi, especially for patients with *BRCA1*-Del ex9-12 Mexican founder mutation. We consider that a new prospective study would be feasible and essential because it may provide more evidence on the efficacy of PARPi in this patient population. Therefore, we recommend the detection of the founder mutation in patients susceptible to treatment with PARPi since the patients in our study benefited from olaparib.

In Mexico, mutation screening in OC patients with and without cancer family history is limited (Martínez-Treviño et al., 2018). Very few research and medical oncology care centers provide this multidisciplinary care service, which allows the identification of patients that may benefit from new therapies for treatment with PARPi (Fragoso-Ontiveros et al., 2019). We highlighted the need to include genetic risk assessment and molecular testing in medical oncology centers that also allows genetic counseling to detect this Mexican founder mutation at diagnosis due to its prevalence in the OC patient population. Therefore, based on our results, we propose that the mutation status (*BRCA1*-Del ex9-12) should be an additional stratification factor in the standard treatment of patients.

The Mexican OC patients with the founder mutation (*BRCA1*-Del ex9-12), treated with PARPi maintenance therapy (olaparib), show a significant increase in PFS regardless of the line of treatment compared to other mutations in *BRCA*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

This study was performed following the provisions of the Declaration of Helsinki and Good Clinical Practice guidelines, under a protocol approved by the Ethics Committee and Institutional Review Board (IRB) of the Instituto Nacional de Cancerología (Mexico City, Mexico) [INCAN (020/019/OMI-CEI/1459/19)]. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization—DG-R and EdM-S. Methodology—DG-R, EdM-S, and ElM-S. Validation—EdM-S and EGI. Formal analysis—EdM-S. Investigation—EdM-S, ElM-S, and RMA-G. Resources—DG-R, GA-G, AB-G, ElM-S, LC-P, FM-V, CC-B, JC-M, and DG-R. Data curation—EdM-S, GA-G, and ElMS. Writing—original draft—DG-R, EdM-S, GA-G, and ElS. Writing–review and editing—DG-R, EdM-S, AB-G, LC-P, RMA-G, AP-T, and DC-E. Visualization—EdM-S and ElM-S. Supervision—EdM-S and RE-R. Project administration management—EG-I. Funding acquisition.—DG-R.

FUNDING

Olaparib treatments were financed through a federal funding scheme awarded to the Ovarian and Endometrial Cancer Program (COE) at the "Instituto Nacional de Cancerología (INCan)" by the lower house of the Congress of the Union of Mexico. The *BRCA* testing was financially supported by AstraZeneca. However, the sponsor had no role in the design, conduct, or oversight of the study. Data collection, management, analysis, and interpretation were carried out independently as well as the preparation, review, and final approval of the manuscript.

ACKNOWLEDGMENTS

• The authors are grateful to Yazmin Bastida-García for her support with project management of patient support and follow-up at the Ovarian Cancer Program, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico.
• The authors also thank Yuliana Sánchez Contreras and Paulina Nuñez Martínez for patient follow-up at the Hereditary Cancer Clinic, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.863956/full#supplementary-material

REFERENCES

Alsop, K., Fereday, S., Meldrum, C., deFazio, A., Emmanuel, C., George, J., et al. (2012). BRCA Mutation Frequency and Patterns of Treatment Response inBRCA Mutation-Positive Women with Ovarian Cancer: A Report from the Australian Ovarian Cancer Study Group. *Jco* 30 (21), 2654–2663. doi:10.1200/jco.2011.39.8545

Banda, K., Swisher, E. M., Wu, D., Pritchard, C. C., and Gadi, V. K. (2018). Somatic Reversion of Germline BRCA2 Mutation Confers Resistance to Poly(ADP-Ribose) Polymerase Inhibitor Therapy. *JCO Precis. Oncol.* 2, 1–6. doi:10.1200/po.17.00044

Boussios, S., Moschetta, M., Zarkavelis, G., Papadaki, A., Kefas, A., and Tatsi, K. (2017). Ovarian Sex-Cord Stromal Tumours and Small Cell Tumours: Pathological, Genetic and Management Aspects. *Crit. Rev. Oncology/Hematology* 120, 43–51. doi:10.1016/j.critrevonc.2017.10.007

Boussios, S., Mikropoulos, C., Samartzis, E., Karihtala, P., Moschetta, M., Sheriff, M., et al. (2020). Wise Management of Ovarian Cancer: On the Cutting Edge. *Jpm* 10 (2), 41. doi:10.3390/jpm10020041

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., et al. (2010). Colloquium Paper: Genome-wide Patterns of Population Structure and Admixture Among Hispanic/Latino Populations. *Proc. Natl. Acad. Sci. U. S. A.* 107 (Suppl. 2), 8954–8961. doi:10.1073/pnas.0914618107

Byers, H., Wallis, Y., van Veen, E. M., Lalloo, F., Reay, K., Smith, P., et al. (2016). Sensitivity of BRCA1/2 Testing in High-Risk Breast/ovarian/male Breast Cancer Families: Little Contribution of Comprehensive RNA/NGS Panel Testing. *Eur. J. Hum. Genet.* 24 (11), 1591–1597. doi:10.1038/ejhg.2016.57

Cancer Genome Atlas Research Network (2011). Integrated Genomic Analyses of Ovarian Carcinoma. *Nature* 474 (7353), 609–615. doi:10.1038/nature10166

Coleman, R. L., Oza, A. M., Lorusso, D., Aghajanian, C., Oaknin, A., Dean, A., et al. (2017). Rucaparib Maintenance Treatment for Recurrent Ovarian Carcinoma after Response to Platinum Therapy (ARIEL3): a Randomised, Double-Blind, Placebo-Controlled, Phase 3 Trial. *Lancet* 390 (10106), 1949–1961. doi:10.1016/S0140-6736(17)32440-6

Cunningham, J. M., Cicek, M. S., Larson, N. B., Davila, J., Wang, C., Larson, M. C., et al. (2014). Clinical Characteristics of Ovarian Cancer Classified by BRCA1, BRCA2 and RAD51C Status. *Sci. Rep.* 4, 4026. doi:10.1038/srep04026

Del Campo, J. M., Matulonis, U. A., Malander, S., Provencher, D., Mahner, S., Follana, P., et al. (2019). Niraparib Maintenance Therapy in Patients with Recurrent Ovarian Cancer after a Partial Response to the Last Platinum-Based Chemotherapy in the ENGOT-OV16/NOVA Trial. *Jco* 37 (32), 2968–2973. doi:10.1200/jco.18.02238

Fragoso-Ontiveros, V., Velázquez-Aragón, J. A., Nuñez-Martínez, P. M., de la Luz Mejía-Aguayo, M., Vidal-Millán, S., Pedroza-Torres, A., et al. (2019). Mexican BRCA1 Founder Mutation: Shortening the Gap in Genetic Assessment for Hereditary Breast and Ovarian Cancer Patients. *PLoS One* 14 (9), e0222709. doi:10.1371/journal.pone.0222709

Gallardo-Rincón, D., Alamilla-García, G., Montes-Servín, E., Morales-Vázquez, F., Cano-Blanco, C., Coronel-Martínez, J., et al. (2019). Experiencia con el uso de olaparib en pacientes con cáncer de ovario. *Gac. Med. Mex.* 155 (6), 585–589. doi:10.24875/GMM.19005494

Gallardo-Rincón, D., Álvarez-Gómez, R. M., Montes-Servín, E., Toledo-Leyva, A., Montes-Servín, E., Michel-Tello, D., et al. (2020). Clinical Evaluation of BRCA1/2 Mutation in Mexican Ovarian Cancer Patients. *Transl. Oncol.* 13 (2), 212–220. doi:10.1016/j.tranon.2019.11.003

Ghose, A., Moschetta, M., Pappas-Gogos, G., Sheriff, M., and Boussios, S. (2021). Genetic Aberrations of DNA Repair Pathways in Prostate Cancer: Translation to the Clinic. *Ijms* 22 (18), 9783. doi:10.3390/ijms22189783

González-Martín, A., Pothuri, B., Vergote, I., DePont Christensen, R., Graybill, W., Mirza, M. R., et al. (2019). Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* 381 (25), 2391–2402. doi:10.1056/nejmoa1910962

Hennessy, B. T. J., Timms, K. M., Carey, M. S., Gutin, A., Meyer, L. A., Flake, D. D., 2nd, et al. (2010). Somatic Mutations in BRCA1 and BRCA2 Could Expand the Number of Patients that Benefit from Poly (ADP Ribose) Polymerase Inhibitors in Ovarian Cancer. *Jco* 28 (22), 3570–3576. doi:10.1200/jco.2009.27.2997

James, P. A., Sawyer, S., Boyle, S., Young, M.-A., Kovalenko, S., Doherty, R., et al. (2015). Large Genomic Rearrangements in the Familial Breast and Ovarian Cancer Gene BRCA1 Are Associated with an Increased Frequency of High Risk Features. *Fam. Cancer* 14 (2), 287–295. doi:10.1007/s10689-015-9785-0

Judkins, T., Hendrickson, B. C., Deffenbaugh, A. M., Eliason, K., Leclair, B., Norton, M. J., et al. (2005). Application of Embryonic Lethal or Other Obvious Phenotypes to Characterize the Clinical Significance of Genetic Variants Found in Trans with Known Deleterious Mutations. *Cancer Res.* 65 (21), 10096–10103. doi:10.1158/0008-5472.can-05-1241

Judkins, T., Rosenthal, E., Arnell, C., Burbidge, L. A., Geary, W., Barrus, T., et al. (2012). Clinical Significance of Large Rearrangements in BRCA1 and BRCA2. *Cancer* 118 (21), 5210–5216. doi:10.1002/cncr.27556

Kwong, A., Chen, J., Shin, V. Y., Ho, J. C. W., Law, F. B. F., Au, C. H., et al. (2015). The Importance of Analysis of Long-Range Rearrangement of BRCA1 and BRCA2 in Genetic Diagnosis of Familial Breast Cancer. *Cancer Genet.* 208 (9), 448–454. doi:10.1016/j.cancergen.2015.05.031

Martínez-Treviño, D. A., León-Cachón, R. B. R., Villarreal-Garza, C., Aguilar Y Méndez, D., Aguilar-Martínez, E., and Barrera-Saldaña, H. A. (2018). A novel method to detect the Mexican founder mutation BRCA1 ex9-12del associated with breast and ovarian cancer using quantitative polymerase chain reaction and TaqMan® probes. *Mol. Med. Rep.* 18 (2), 1531–1537. doi:10.3892/mmr.2018.9141

Moore, K., Colombo, N., Scambia, G., Kim, B.-G., Oaknin, A., Friedlander, M., et al. (2018). Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* 379 (26), 2495–2505. doi:10.1056/nejmoa1810858

Oliver, J., Quezada Urban, R., Franco Cortés, C. A., Díaz Velásquez, C. E., Montealegre Paez, A. L., Pacheco-Orozco, R. A., et al. (2019). Latin American Study of Hereditary Breast and Ovarian Cancer LACAM: A Genomic Epidemiology Approach. *Front. Oncol.* 9, 1429. doi:10.3389/fonc.2019.01429

Ossa, C. A., and Torres, D. (2016). Founder and Recurrent Mutations in BRCA1 and BRCA2 Genes in Latin American Countries: State of the Art and Literature Review. *Oncologist* 21 (7), 832–839. doi:10.1634/theoncologist.2015-0416

Pennington, K. P., Walsh, T., Harrell, M. I., Lee, M. K., Pennil, C. C., Rendi, M. H., et al. (2014). Germline and Somatic Mutations in Homologous Recombination Genes Predict Platinum Response and Survival in Ovarian, Fallopian Tube, and Peritoneal Carcinomas. *Clin. Cancer Res.* 20 (3), 764–775. doi:10.1158/1078-0432.ccr-13-2287

Pujade-Lauraine, E., Ledermann, J. A., Selle, F., Gebski, V., Penson, R. T., Oza, A. M., et al. (2017). Olaparib Tablets as Maintenance Therapy in Patients with Platinum-Sensitive, Relapsed Ovarian Cancer and a BRCA1/2 Mutation (SOLO2/ENGOT-Ov21): a Double-Blind, Randomised, Placebo-Controlled, Phase 3 Trial. *Lancet Oncol.* 18 (9), 1274–1284. doi:10.1016/S1470-2045(17)30469-2

Quezada Urban, R., Díaz Velásquez, C. E., Gitler, R., Rojo Castillo, M. P., Sirota Toporek, M., Figueroa Morales, A., et al. (2018). Comprehensive Analysis of Germline Variants in Mexican Patients with Hereditary Breast and Ovarian Cancer Susceptibility. *Cancers (Basel)* 10 (10), 361. doi:10.3390/cancers10100361

Randall, M., Burgess, K., Buckingham, L., and Usha, L. (2020). Exceptional Response to Olaparib in a Patient with Recurrent Ovarian Cancer and an Entire BRCA1 Germline Gene Deletion. *J. Natl. Compr. Canc Netw.* 18 (3), 223–228. doi:10.6004/jnccn.2019.7378

Sluiter, M. D., and van Rensburg, E. J. (2011). Large Genomic Rearrangements of the BRCA1 and BRCA2 Genes: Review of the Literature and Report of a Novel BRCA1 Mutation. *Breast Cancer Res. Treat.* 125, 325–349. doi:10.1007/s10549-010-0817-z

The Global Cancer Observatory (2020). GLOBOCAN 2020. [Internet]. Lyon CEDEX 08, France. The International Agency for Research on Cancer (IARC) Is the Specialized Cancer Agency of the World Health Organization. The Objective of the IARC Is to Promote International Collaboration in Cancer Research 2020. Available at: https://gco.iarc.fr/ (Accessed July 30, 2021).

Toss, A., Tomasello, C., Razzaboni, E., Contu, G., Grandi, G., Cagnacci, A., et al. (2015). Hereditary Ovarian Cancer: Not Only BRCA 1 and 2 Genes. *Biomed. Res. Int.* 2015, 341723. doi:10.1155/2015/341723

Vaca-Paniagua, F., Alvarez-Gomez, R. M., Fragoso-Ontiveros, V., Vidal-Millan, S., Herrera, L. A., Cantú, D., et al. (2012). Full-exon Pyrosequencing Screening of BRCA Germline Mutations in Mexican Women with Inherited Breast and Ovarian Cancer. *PLoS One* 7 (5), e37432. doi:10.1371/journal.pone.0037432

Villarreal-Garza, C., Alvarez-Gómez, R. M., Pérez-Plasencia, C., Herrera, L. A., Herzog, J., Castillo, D., et al. (2015). Significant Clinical Impact of recurrentBRCA1andBRCA2mutations in Mexico. *Cancer* 121 (3), 372–378. doi:10.1002/cncr.29058

Villarreal-Garza, C., Weitzel, J. N., Llacuachaqui, M., Sifuentes, E., Magallanes-Hoyos, M. C., Gallardo, L., et al. (2015). The Prevalence of BRCA1 and BRCA2 Mutations Among Young Mexican Women with Triple-Negative Breast Cancer. *Breast Cancer Res. Treat.* 150 (2), 389–394. doi:10.1007/s10549-015-3312-8

Wang, X., Hu, N., Cui, L., Si, Y., Yue, J., Zheng, F., et al. (2022). Durable Disease-free Survival in a Patient with Metastatic Triple-Negative Breast Cancer Treated with Olaparib Monotherapy. *Curr. Cancer Drug Targets* [Epub ahead of print]. doi:10.2174/1568009622666220214092207

Weaver, A. N., and Yang, E. S. (2013). Beyond DNA Repair: Additional Functions of PARP-1 in Cancer. *Front. Oncol.* 3, 290. doi:10.3389/fonc.2013.00290

Weitzel, J. N., Lagos, V., Blazer, K. R., Nelson, R., Ricker, C., Herzog, J., et al. (2005). Prevalence of BRCA Mutations and Founder Effect in High-Risk Hispanic Families. *Cancer Epidemiol. Biomarkers Prev.* 14 (7), 1666–1671. doi:10.1158/1055-9965.epi-05-0072

Weitzel, J. N., Lagos, V. I., Herzog, J. S., Judkins, T., Hendrickson, B., Ho, J. S., et al. (2007). Evidence for Common Ancestral Origin of a Recurring BRCA1 Genomic Rearrangement Identified in High-Risk Hispanic Families. *Cancer Epidemiol. Biomarkers Prev.* 16 (8), 1615–1620. doi:10.1158/1055-9965.epi-07-0198

Weitzel, J. N., Clague, J., Martir-Negron, A., Ogaz, R., Herzog, J., Ricker, C., et al. (2013). Prevalence and Type ofBRCAMutations in Hispanics Undergoing Genetic Cancer Risk Assessment in the Southwestern United States: A Report from the Clinical Cancer Genetics Community Research Network. *Jco* 31 (2), 210–216. doi:10.1200/jco.2011.41.0027

Zhang, S., Royer, R., Li, S., McLaughlin, J. R., Rosen, B., Risch, H. A., et al. (2011). Frequencies of BRCA1 and BRCA2 Mutations Among 1,342 Unselected Patients with Invasive Ovarian Cancer. *Gynecol. Oncol.* 121 (2), 353–357. doi:10.1016/j.ygyno.2011.01.020

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gallardo-Rincón, Montes-Servín, Alamilla-García, Montes-Servín, Bahena-González, Cetina-Pérez, Morales Vásquez, Cano-Blanco, Coronel-Martínez, González-Ibarra, Espinosa-Romero, María Alvarez-Gómez, Pedroza-Torres and Castro-Eguiluz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fgene.2022.869967

Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment

Fazhan Li[1,2#], *Huijuan Wen*[1,2#], *Ihtisham Bukhari*[1#], *Bin Liu*[1,2], *Chenxu Guo*[1,2], *FeiFei Ren*[1,2], *Youcai Tang*[3], *Yang Mi*[1,2,4*] and *Pengyuan Zheng*[1,2,4*]

[1]Henan Key Laboratory of Helicobacter Pylori, Microbiota and Gastrointestinal Cancer, Marshall Medical Research Center, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Academy of Medical Science, Zhengzhou University, Zhengzhou, China, [3]Department of Pediatrics, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [4]Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Gastric cancer (GC) is a highly fatal and common malignancy of the digestive system. Recent therapeutic advancements have significantly improved the clinical outcomes in GC, but due to the unavailability of suitable molecular targets, a large number of patients do not respond to the immune checkpoint inhibitors (ICI) therapy. To identify and validate potential therapeutic and prognostic targets of gastric cancer, we used the "inferCNV" R package for analyzing single-cell sequencing data (GSE112302) of GC and normal epithelial cells. First, by using LASSO, we screened genes that were highly correlated with copy number variations (CNVs). Therefrom, five gene signature (*CPVL*, *DDC*, *GRTP1*, *ONECUT2*, and *PRSS21*) was selected by cross-validating the prognosis and risk management with the GC RNA-seq data obtained from GEO and TCGA. Moreover, the correlation analyses between CNVs of these genes and immune cell infiltration in gastric cancer identified *CPVL* as a potential prognostic marker. Finally, *CPVL* showed high expression in gastric cancer samples and cell lines, then siRNA-mediated silencing of *CPVL* expression in gastric cancer cells showed significant proliferation arrest in MGC803 cells. Here, we conclude that CNVs are key regulators of the immune cells infiltration in gastric TME as well as cancer development, and *CPVL* could potentially be used as a prognostic and therapeutic marker in gastric cancer.

Keywords: gastric cancer, cancer immunotherapy, immune infiltration, copy number variation, tumor microenvironment

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Zehua Song,
Henan Provincial People's Hospital, China
Niaz Muhammad,
Minhaj University Lahore, Pakistan

***Correspondence:**
Yang Mi
yangmi198@zzu.edu.cn
Pengyuan Zheng
pyzheng@zzu.edu.cn

#: These authors contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *15 February 2022*
Accepted: *23 May 2022*
Published: *08 June 2022*

Citation:
Li F, Wen H, Bukhari I, Liu B, Guo C, Ren F, Tang Y, Mi Y and Zheng P (2022) Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment. Front. Genet. 13:869967. doi: 10.3389/fgene.2022.869967

INTRODUCTION

Gastric cancer is categorized as the fourth most common and third cause of cancer-related mortality worldwide (Bray et al., 2018). In recent years, the survival rate in many cancers has been improved due to the application of neoadjuvant chemotherapy, but in GC it is still worst because of its diagnosis at advanced stages, drug resistance, recurrence (Pernot et al., 2015; Ilson, 2017; Zurleni et al., 2018), high tumor heterogeneity, and poor immune cell infiltration in TME, which seriously hinder the

Abbreviations: AUC, Area under the curve; CNVs, copy number variations; DEGs, Differentially expressed genes; EBV+, Epstein-Barr virus infection positive; GEO, Gene Expression Omnibus; ICI, Immune checkpoint inhibitor; LASSO, Least Absolute Shrinkage and Selection Operator; MSI, Microsatellite Instability; PCA, Principal component analysis; ROC, Receiver operating characteristic curve; TCGA, The cancer genome atlas project; TME, Tumor microenvironment.

prognosis and therapeutic outcomes (Meacham and Morrison, 2013; Weinberg, 2014; Van Cutsem et al., 2016). Therefore, it is crucial to thoroughly explore the heterogeneity of the GC, the mechanism of immune cell infiltration, and the new therapeutic targets for gastric cancer.

Currently, transcriptomic analyses revealed a variety of biomarkers that provide a base to determine the complexity and heterogeneity of tumors and identify new therapeutic targets (Cieslik and Chinnaiyan, 2018). The large-scale copy number variations (CNVs) through single-cell RNA-seq (scRNA-seq) have distinguished malignant cells from the normal cells (Patel et al., 2014; Müller et al., 2016; Tirosh et al., 2016). In general, the CNVs may regulate the function of somatic cells thus it is useful to use as the potential tumorigenesis markers (Beroukhim et al., 2010) and may affect immunotherapy (Kacew et al., 2019). On the other hand, the regulation of immune cell infiltration in TME plays an important role in the occurrence, progression, therapeutics and prognosis of cancers including GC (Siegel et al., 2017; Waniczek et al., 2017). However, the role of CNVs in immune cell infiltration and tumor development specifically in gastric cancer is not well illustrated.

Certainly, the scRNA-seq is an effective method for analyzing the heterogeneity of the complex biological systems such as TME (Chung et al., 2017; Cochain et al., 2018). Therefore, we quantified CNVs from scRNA-seq data of the gastric cancer and surrounding normal tissues to differentiate malignant cells from the normal epithelial cells. The differentially expressed genes (DEGs) in GC tissues were screened by the least absolute shrinkage and selection operator (LASSO) Cox regression model (McEligot et al., 2020). The expression of the hub genes and their CNVs in gastric cancer were further assessed for their relationship with immune cell infiltration and prognosis in the gastric tumor. Lastly, we determined the expression of levels of all hub genes in gastric cancer tissues and cell lines, then based on the expression pattern of the genes we selected CPVL to explore its role in cellular proliferation in gastric cancer cell lines.

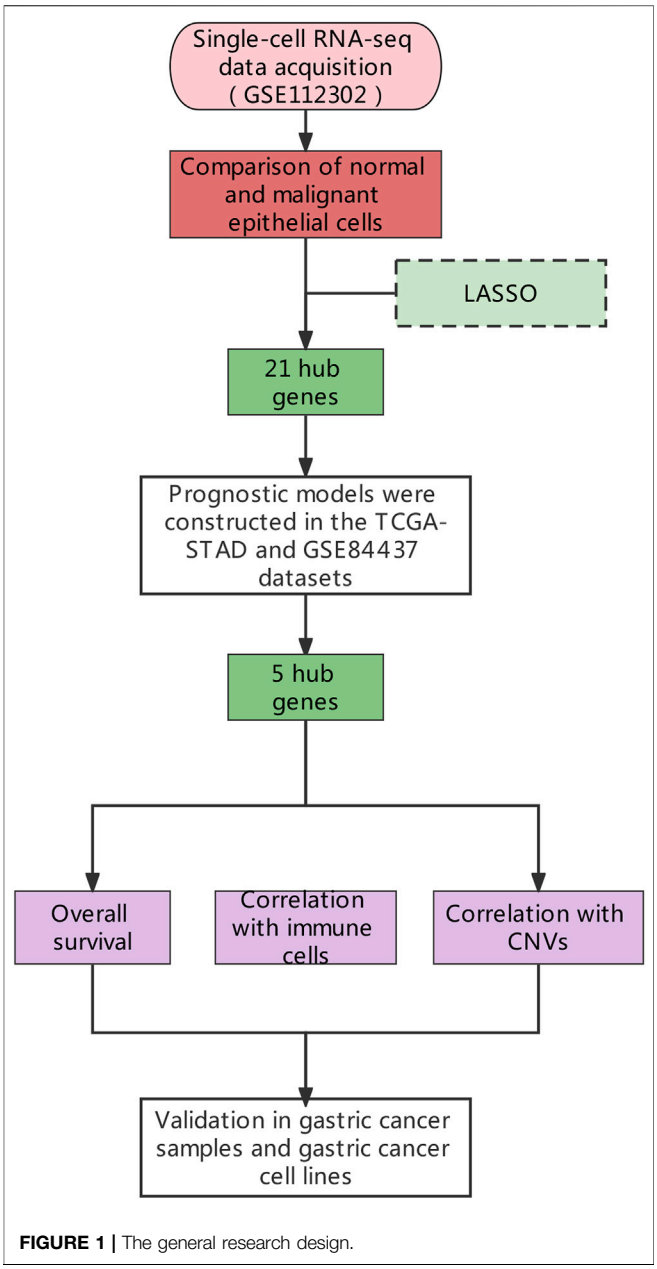


FIGURE 1 | The general research design.

METHODS AND MATERIALS

Data Mining and Processing

The scRNA-seq data of a total of 707 tumors and paired normal cells (GSE112302), and RNA-seq data (GSE84437) of 357 gastric cancer tissues were downloaded from Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/geo/) database. The transcriptome sequencing data of 407 individuals (32 normal and 375 gastric cancer tissue samples) along with complete clinical information were downloaded from The Cancer Genome Atlas (TCGA: https://portal.gdc.cancer.gov/). The complete study design and sample information are shown in **Figure 1**. The Seurat (Loraine et al., 2015) in RStudio (Stuart et al., 2019) was used to filter and standardize the tumor samples (355 tumor cells in total) in GSE112302 single-cell sequencing data. Genes expressed in less than three samples, and mitochondrial (MT) genes expressing >5% samples were filtered out. Principal component analysis (PCA) and tSNE were used for the dimensional reduction (Satija et al., 2015), and different cells' clusters were identified and annotated according to the marker genes (resolution = 0.5). The raw data of GSE84437 and the gastric cancer data from TCGA were standardized using the "limma" package (Smyth et al., 2005).

Epithelial Cell Reanalysis

The CNVs in macrophages (n = 23) and epithelium (n = 332) were evaluated by "infercnv" R package (Patel et al., 2014). The genomic locations of the differential genes with a high rate of copy number variations (CNVs) were determined and compared their average expression level with their relative expression. The "limma" package was used to find the differentially expressed genes among tumor epithelial cells with high and low CNVs ($|\log FC| > 2$, $p < 0.05$).

LASSO Regression Analysis

We constructed a prediction model based on the CNV scores of five signature genes by performing a LASSO regression analysis using the "glmnet" package in R. LASSO regression was applied to the module(s) with *p*-value of cnv-scores less than 0.05 to determine the signature gene, and the analyses were repeated 1,000 times to cross-validate and to select the best lambda value of the hub gene.

Construction of the Prognostic Model

The risk scores of the potential genes from both TCGA transcriptome and GSE84437 data of gastric cancer were calculated by the formula Risk score = βgene1 × exprgene1 + βgene2 × exprgene2 + · ···· + βgene *n* × exprgene *n*. By considering the median risk score as a critical value, the samples were divided into high-risk groups and low-risk groups. Overall survival was determined by using R packages "survival" and "survminer" while R package "survival ROC" was used to draw the receiver operating characteristic curve (ROC), and Area Under Curve (AUC) to show the accuracy of the prognostic model. Single-factor COX analysis and multi-factor COX analysis were used to analyze the relationship between risk scores and clinical characteristics. R packages "rms" were used to build a nomogram model to estimate the factors related to overall survival and to calculate the risk scores. The sum of risk scores was used to predict 3-years and 5-years survival rates.

Immune Relevance of Hub Genes

Kaplan-Meier plotter (http://kmplot.com/analysis/) is an online database for analyzing the survival rate of mRNA or miRNA in various cancers. Using this, we obtained the survival curve of the hub genes in gastric cancer, and their correlation with the infiltration of the immune cells (neutrophils, macrophages, dendritic cells, B cells, and CD4/CD8 T cells) in the tumor microenvironment was analyzed by TIMER (https://cistrome.shinyapps.io/timer/) (Li et al., 2017). The expression levels of the hub genes were also analyzed in the pan-cancer to find the difference between normal tissues and various cancers. The differences in the genes' expression and the relationship of hub genes with immune cell infiltration and immune checkpoints (*PDCD1*, *CD274*, *PDCD1LG2*, and *CTLA4*) were also determined.

Gastric Cancer Tissue Sampling, RNA Extraction and Q-RT-PCR Analyses

A total of seven gastric cancer tissue and adjacent normal tissue samples were obtained from patients undergoing partial gastrectomy at the Fifth Affiliated Hospital of Zhengzhou University. Informed written consent was obtained from all patients before recruitment into the study, and all sampling and experimental procedures were approved by the ethics committee of the fifth affiliated hospital of Zhengzhou University. Fresh tissues were washed thrice with PBS, crushed and treated with TRIzol (Invitrogen, Carlsbad, United States) for RNA extraction. The cDNA samples were prepared by using the Rever Tra Ace qPCR RT Kit (Osaka, Japan) and mRNA expression was performed using SYBR Premix Ex Taq (Tokyo, Japan) and specially designed primers (**Table 1**). GAPDH was used as internal control and the relative transcriptional expression level was calculated according to the 2^{Δ}Ct approach.

TABLE 1 | List of the primers used in the study.

Primer Name	Sequence
CPVL-F	TGACCTTGCGTGACAGAGAC
CPVL-R	CCGTGCACCGCAAAAAGTTA
DDC-F	TGGGGACCACAACATGCTG
DDC-R	TCAGGGCAGATGAATGCACTG
GRTP1-F	GGCTACTACCACCAGCTTCTCCAG
GRTP1-R	AAGGTCCGGTTCAGGTCTGTCC
PRSS21-F	GCAACCACCTCTTCCTCAAGTACAG
PRSS21-R	CTGAGTCACCGAAGCAGGCATC
ONECUT2-F	TACTGCCACACCCATACGAGAGAG
ONECUT2-R	TGCTGAGGAGAGGTCTGCCAAG
GAPDH-F	GAGTCAACGGATTTGGTCGT
GAPDH-R	TTGATTTTGGAGGGATCTCG
CPVL-SiRNA1- sense	5′-CGCUCUCCAUGCUUUACAUTT-3′
CPVL-SiRNA1- antisense	5′-AUGUAAAGCAUGGAGAGCGTT -3′
CPVL-SiRNA2- sense	5′-CAGCUUUACUAUGUGAAAUTT-3′
CPVL-SiRNA2- antisense	5′-AUUUCACAUAGUAAAGCUGTT-3′
hsa-miR-196b-5p	GTCGCATCAAGGATCTTAAACTTTGCC
hsa-miR-561-5p	ACGCTAGGTAGTTTCCTGTTGTTGG
hsa-miR-7-5p	ACGCGTGGAAGACTAGTGATTTTGTTG
hsa-miR-196a-5p	ACGCTAGGTAGTTTCATGTTGTTGGG
LINC01678-F	CCCTCCTGCAACAATCGGG
LINC01678-R	GGACCAGGGTAGTGAGGTAGT
TRG-AS1-F	GCCTGAATCCAGTGTTCCTGAGTG
TRG-AS1-R	TGGGCATTGGTCTTGGCTTTGTC
LINC00472-F	TTCAGGAATATGGCAGGCTCAACAC
LINC00472-R	GGGTGGTCTGAGAGGAGGCATC
AL158207.2-F	ACCACTCATCAAATCCTGCCA
AL158207.2-R	GGAGCACTTAGCACCGCATA
AL122035.2-F	CAGGAATATGGCAGGCTCAA
AL122035.2-R	GGGCATTGGTCTTGGCTTTGT

Correlation of *CPVL* With Immune and Matrix Scores

The "Estimate" package was used to evaluate the immune score and matrix score in pan-cancer. The correlation between *CPVL*, immune, and matrix scores was calculated by spearman text and the "ggplot2" package was used to draw scatter plots.

siRNA-Mediated Gene Knockdown

The siRNA targeting *CPVL* and scrambled negative control siRNA were provided by Shanghai GenePharma (Shanghai, China), the sequences of siRNA are listed in **Table 1**. The MGC803 cells were transfected with siRNA using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, United States) at a final concentration of 80 nM. The efficiency of siRNA knockdown was subsequently confirmed using qPCR.

Cell Proliferation Assay

Cell Counting Kit-8 (DOJINDO, Kumamoto, Japan) was used according to the manufacturer's instructions. After 24–48 h of transfecting cells with siRNA-CPVL, ~5,000 cells were seeded into each well of the 96-well plate, followed by adding 10 μL of

CCK-8 reagent to each well after 24 h and incubating at 37°C for 2 h, then measuring absorbance at 450 nm (ELx800, Bio-Tek, United States). Starting from siRNA transfection to the proliferation test, all experiments were repeated thrice.

The Prediction of *CPVL* Related miRNA and Upstream lncRNA

The StarBase (http://starbase.sysu.edu.cn/) is an online database for studying correlation and predicting the regulatory relationships between mRNA, miRNA, and lncRNA in tumors (Li et al., 2014). First, we predicted the candidate miRNAs associated with *CPVL* by taking |R|> 0.1 and *p*-value <0.05 as a baseline, however, miRNAs showing higher significance were chosen and validated by qPCR in gastric cancer tissues and cell lines.

Statistical Analysis

R package "Seurat" and "infercnv" were used to perform quality control on single-cell sequencing data, and to filter and calculate cell cnv-scores. The expression levels of genes in transcriptome data were normalized by log2 transformation. COX regression analysis was used to establish the relationship between gene expression and risk score, R package "survival ROC" for survival prediction models, and the "rms" package was used to build nomogram models. The "Estimate" package was used to evaluate the immune score and matrix score of pan-cancer, and the Spearman test was used to calculate the correlation between *CPVL* and the immune score and matrix score. $p < 0.05$ was considered as significant.

The experimental data were statistically analyzed and plotted by GraphPad 8.0 software, and other measurement data were described by means ± standard deviation, all using the student t-test.

RESULTS

CNV and Related Genes Within the Tumor Cell Through scRNA-Seq Analyses

First of all, the quality checks were performed for the scRNA-seq GSE112302 dataset (**Supplementary Figures S1A,C**), and selected the top 1,500 genes fulfilling our basic selection criteria (**Supplementary Figure S1B**). For 355 tumor cells, we performed principal component analysis (PCA) and clustered the cells using graph-based clustering ($n = 20$) (**Supplementary Figures S1D–G**). Furthermore, we annotated the cells including epithelial cells and macrophages showing marker genes (**Figures 2A,B**). It has been known that the transformation of normal cells to cancer cells is closely related to large-scale chromosomal aberrations (Beroukhim et al., 2010). Therefore, the CNVs in different genes were analyzed to differentiate between normal epithelial cells and cancer cells, and we found that cluster2 and cluster3 have comparatively low CNVs than other clusters 0, 1 and 4 (**Figure 2C**). Furthermore, we analyzed the differences between the tumor cells expressing high and low CNVs, showing a large number of differential genes in both groups (**Figures 2D,E**). These results indicate that CNVs are one of the key contributors to GC heterogeneity.

LASSO Regression Analysis

Furthermore, we performed the LASSO regression analysis for the differential genes (**Figures 3A,B**), and obtained 21 genes as potential hub genes (*LRMP*, *SDR42E1*, *ERICH5*, *NPPC*, *KCNJ3*, *CNTNAP2*, *LINC00346*, *CPVL*, *UNC93A*, *ONECUT2*, *MAGEA6*, *SLC19A3*, *DDC*, *VGF*, *LINC00392*, *AQP2*, *PRSS21*, *ETV4, F12*, *RAP2A*, and *GRTP1*), suggesting a strong functional association of these genes with tumor epithelial cells.

Hub Genes Correlated With CNV and Prognosis

To find out the role of hub genes in the overall survival of gastric cancer we performed a multivariate Cox regression analysis, among 21 hub genes, the eight genes (*KCNJ3*, *CPVL*, *ONECUT2*, *SLC19A3*, *DDC*, *PRSS21*, *F12*, and *GRTP1*) were designated as independent indicators of poor prognosis. Based on these findings we constructed a prediction model for the clinical prognosis of GC patients and validated it by the ROC curve [0.625 as an area under the curve (AUC)] (**Figures 4A,B**). Next, the survival information of hub genes and clinical features of the patients were combined to construct the nomogram (**Figure 4C**). The total score of each prognostic factor was represented as 1, 3, and 5-years survival rates of the patients. To find out whether the risk signatures were independent predictors of prognosis, the Cox univariate analysis was performed showing that the age ($p = 0.033$), stage ($p = 0.002$), N stage ($p = 0.022$), and risk score ($p < 0.001$) were significantly linked with OS. While in Cox multivariate analysis only age ($p < 0.001$) and risk score ($p < 0.001$) were significantly linked with OS (**Figures 4D,E**).

Using the GSE84437 dataset, we identified five genes (*CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1*) as independent indicators of poor prognosis. Based on that we constructed and validated a prediction model for clinical prognosis, ROC curve (AUC = 0.626) (**Figures 4F,G**), and the relationship between multiple factors and survival rate was analyzed by nomogram (**Figure 4H**). The single factor analysis in GSE84437 showed that age ($p = 0.041$), T stage ($p < 0.001$), N stage ($p < 0.001$), and risk score ($p < 0.001$) were significantly associated with OS. However, the Cox multivariate analysis showed a significant association between the T3-4 stage ($p < 0.001$), N1-3 stage ($p = 0.002$), and risk score ($p < 0.001$) with OS (**Figures 4I,J**). A venn diagram was drawn for these five survival-related genes (**Supplementary Figure S2A**), expressed in all cell clusters (**Supplementary Figure S2B**). Next, we plotted the relationship between the top five hub genes and survival in gastric cancer on the Kaplan-Meier plotter (http://kmplot.com/analysis/) (**Figure 5A**) showing a significant correlation between low expression of *CPVL*, *ONECUT2*, and *PRSS21* with prolonged survival of the patients. On the other hand, the low expression of *GRTP1* showed an association with poor prognosis, and the expression of DDC did not affect the prognosis of GC patients.

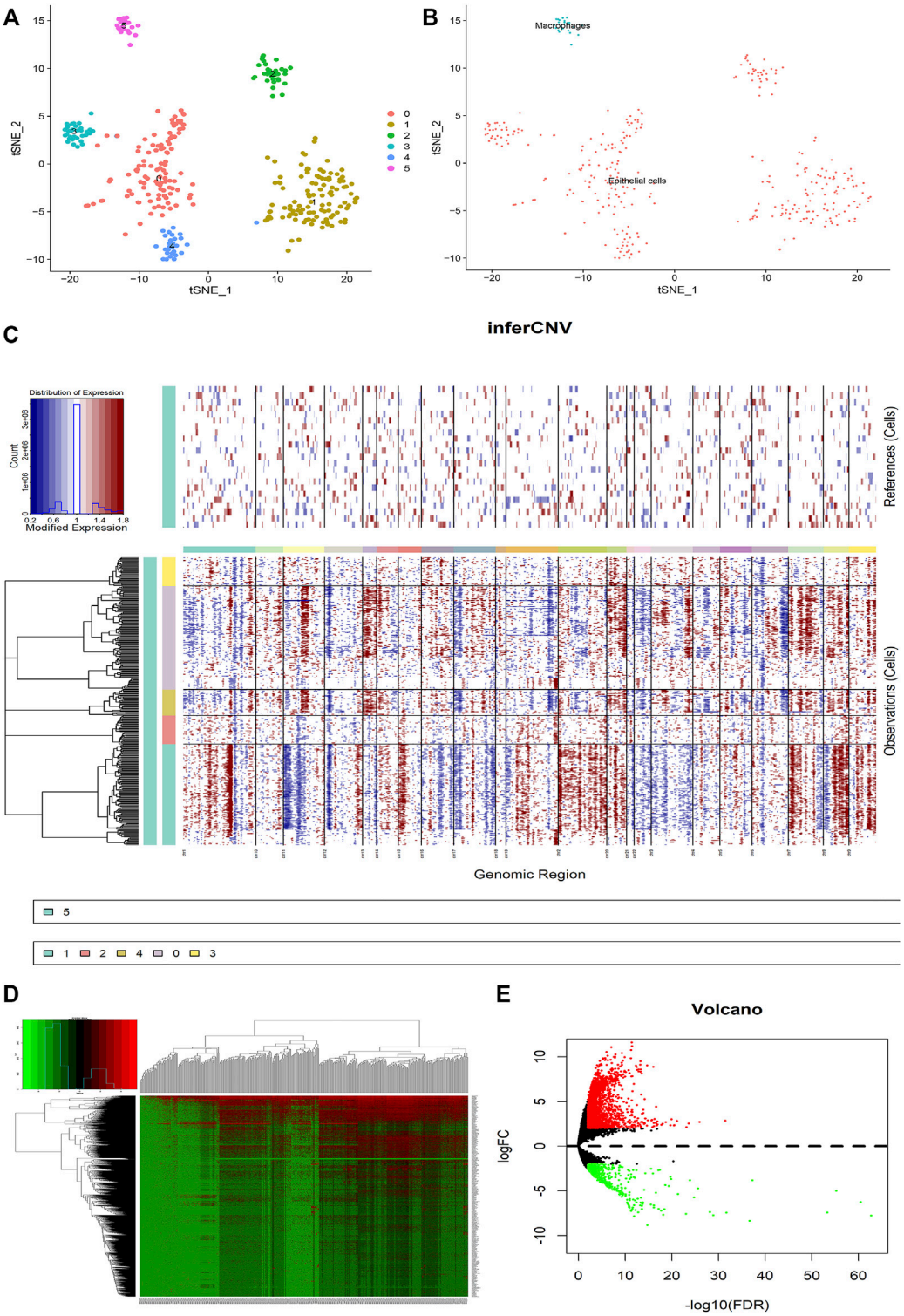


FIGURE 2 | Analysis of the difference between malignant cells and normal epithelial cells in GC tumors. **(A)** The tumor cells are clustered into six clusters including 0–5, clusters 0–4 are annotated as epithelial cells, and cluster 5 is annotated as macrophages. **(B)** The cells are divided into epithelial cells and macrophages. **(C)** CNV heat map of other epithelial cells with cluster 5 macrophages as a control. **(D,E)** Heat map and volcano map for differential analysis. Red represents upregulation and green downregulation, |logFC|>2, $p < 0.05$.

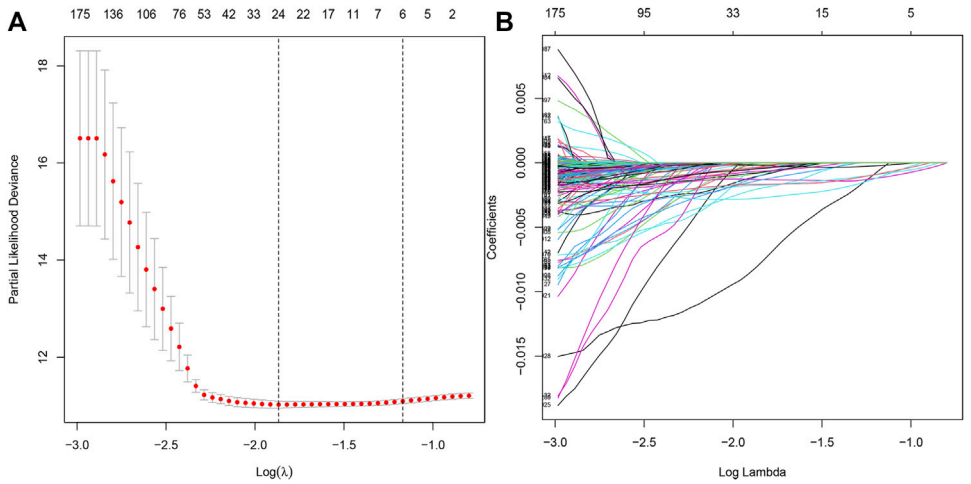


FIGURE 3 | LASSO regression analysis model construction. **(A)** LASSO coefficient distribution of differentially expressed genes. **(B)** Partial likelihood bias of the LASSO coefficient distribution.

Correlation Between Hub Genes and Immune Cells Infiltration

The copy number variations are one of the well-known factors affecting the infiltration of immune cells in tumor microenvironment (Campos-Carrillo et al., 2020). Therefore, we studied the correlation between five hub genes and the infiltration of immune cells such as neutrophils, macrophages, dendritic cells, B cells and CD4/CD8 T cells in GC TME. Among all genes, *CPVL* and *GRTP1* showed a significant positive and negative correlation with immune cell infiltration, respectively. Other genes did not affect the infiltration of immune cells in GC TME (**Figures 5B–F**).

PD1/PD-L1 and *CTLA-4* are important immune checkpoints for tumor immune escape, playing significant roles in tumor immunotherapy (Guo et al., 2021). Among hub genes, only *CPVL* was found to have a positive and *ONECUT2* negative correlation with the immune checkpoints (**Supplementary Figure S3**).

The Relationship Between Hub Genes, Copy Number Variation and Immune Cell Infiltration

Next, we explored the correlation between the expression of the genes and their CNVs with immune cell infiltration in gastric TME. For this purpose, we analyzed the somatic mutations of the hub genes through TIMER [deep deletion (-2), arm-level deletion (-1), diploid/normal (0), arm-level gain (1), and high amplification (2)] and the infiltration of the multiple immune cells. Interestingly, we observed that the copy number variations (CNVs) in hub genes significantly affect the infiltration of immune cells (**Figures 6A–E**). Apart from *DDC*, all other genes including *CPVL*, *ONECUT2*, *PRSS21*, and *GRTP1* positively regulate the infiltration of the immune cells. These findings revealed that CNVs are also major players in the immune regulation of the gastric cancer TME.

Effect of CPVL on Proliferation of Gastric Cancer Cells

We performed the expression of five hub genes in gastric cancers and normal tissue samples which showed significantly high expression of *CPVL* in cancer samples, and cell lines (**Figures 7A–C**) which was consistent with the RNA sequences obtained from TCGA. To further explore the role of *CPVL* in gastric cancer, two different siRNAs were used to silence the expression of *CPVL* in MGC803 cells (**Figure 7D**). Subsequently, we examined the effect of the *CPVL*-knockdown on the proliferation of MGC803 cells. The results showed a significant reduction in the proliferation of MGC803 cells after silencing *CPVL* (**Figure 7E**).

Relationship of *CPVL* With Immunity in Pan-Cancer

Among five hub genes, we focused on *CPVL* because it showed the highest degree of correlation with immune cell infiltration and immune checkpoints *PD-1/PD-L1* and *CTLA-4* in gastric cancer. Thus, we comprehensively explored the *CPVL* in pan-cancer, and its high expression showed a strong association with the poor prognosis in a variety of cancers (**Figure 8**). Total out of 33 pan-cancers, *CPVL* was found to have high expression in 29 cancers (**Supplementary Figure S4**), also showing a significant correlation with the immune and matrix scores. These findings show that *CPVL* has a universal effect on the immune cell infiltration and immune function in TME of various cancers.

The Prediction of *CPVL* Related miRNA and Upstream lncRNA

In starBase (http://starbase.sysu.edu.cn/), we have obtained four potential upstream miRNAs (*hsa-miR-196a-5p*, *hsa-miR-7-5p*,

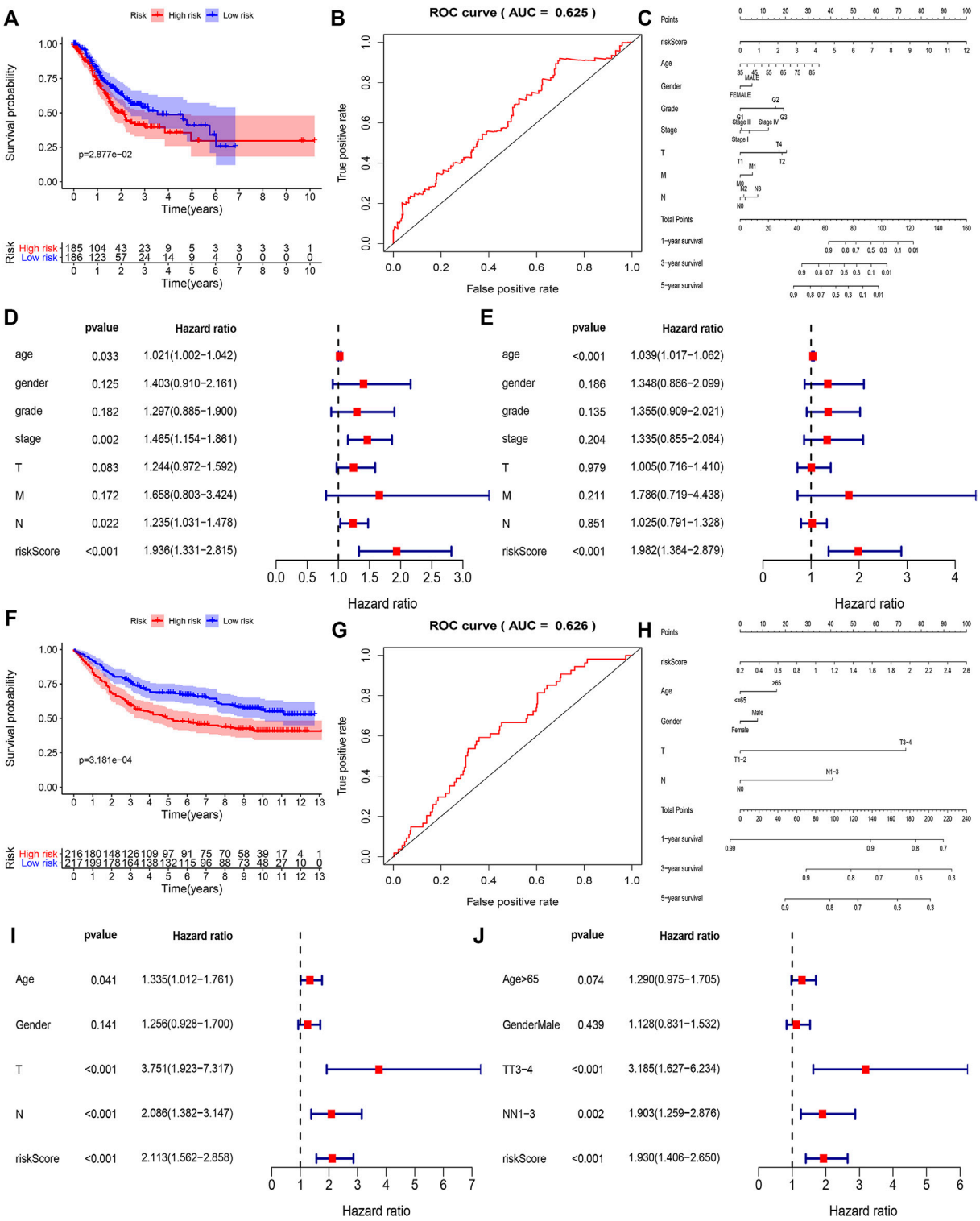


FIGURE 4 | The correlation between the construction of the prognostic model and the clinical characteristics of the selected genes in the TCGA and GEO databases. **(A,B)** Based on COX regression analysis to screen out high-risk genes in TCGA, ROC curve showing accuracy (AUC = 0.625) of the clinical prognosis model. **(C)** Nomogram of TCGA risk score and clinical characteristics of gastric cancer patients. Each factor corresponds to its score, and each score is added to obtain a total score. **(D,E)** Single-factor COX regression analysis and multi-factor COX regression analysis of TCGA risk score. **(F,G)** Based on COX regression analysis a clinical prognostic model was constructed, and the ROC curve is used to detect the accuracy of the model (AUC = 0.625). **(H)** Nomogram of GSE84437 risk score and clinical characteristics of patients with gastric cancer. **(I,J)** Univariate Cox regression analysis and multivariate COX regression analysis of GEO risk score.

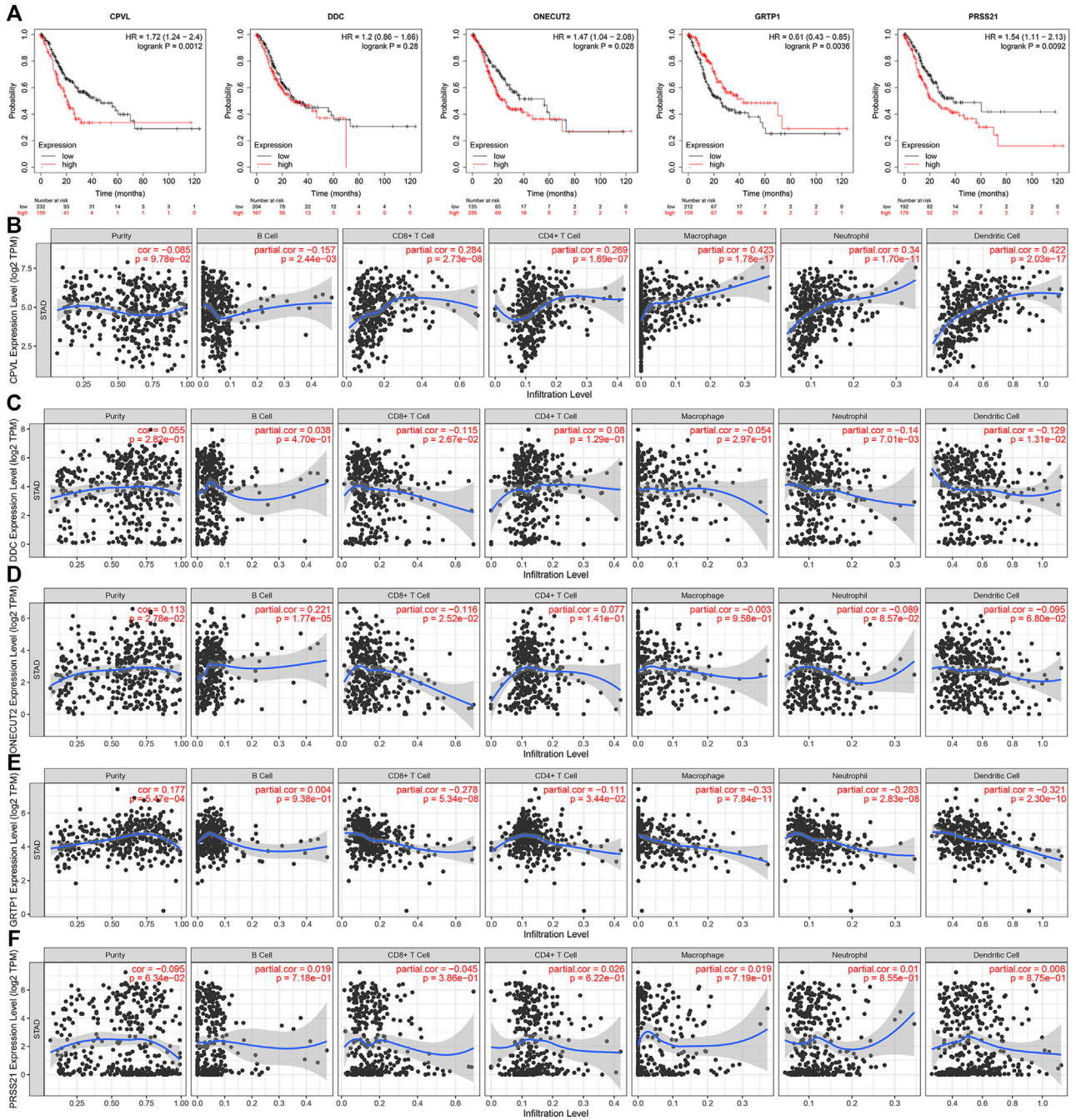


FIGURE 5 | The survival curve of the Hub gene and its relationship with immune cell infiltration. **(A)** The survival curve of *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* in patients with gastric cancer. **(B–F)** The correlation of *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* with neutrophils, macrophages, dendritic cells, B cells, and CD4/CD8 T cells in gastric cancer immune infiltration.

hsa-miR-196b-5p, and *hsa-miR-561-5p*) correlated with *CPVL* (**Table 2**) and among them, we selected *hsa-miR-196b-5p* as a potential target because of its highest degree of negative correlation with *CPVL*. Moreover, we predicted the upstream lncRNAs of *hsa-miR-196b-5p*, and obtained 68 possible lncRNAs, and five lncRNAs showing a negative correlation with *hsa-miR-196b-5p* and positive correlation with *CPVL* in gastric cancer were filtered out (**Table 3**). Furthermore, the expression levels of selected miRNAs and lncRNAs were determined in gastric cancer cell lines. The results showed that lncRNAs AL158207.2 and

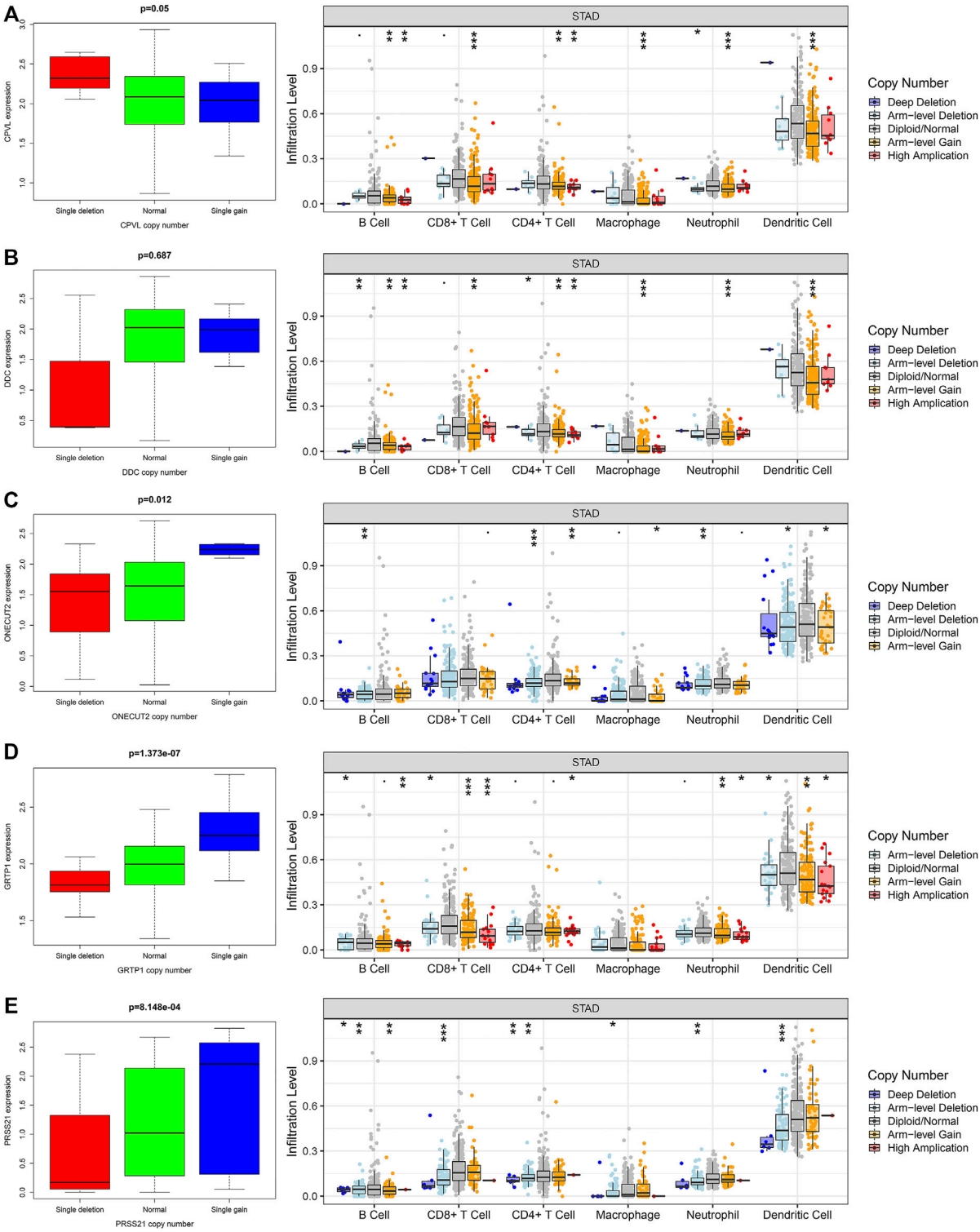


FIGURE 6 | The relationship between the hub genes, CNVs and level of immune infiltration in gastric cancer. The expression and CNVs of **(A)** CPVL, **(B)** ONECUT2, **(C)** DDC, **(D)** PRSS21 and **(E)** GRTP1 showed association with the immune cell infiltration in gastric cancer. *p <0.05, **p <0.01, ***p <0.001.

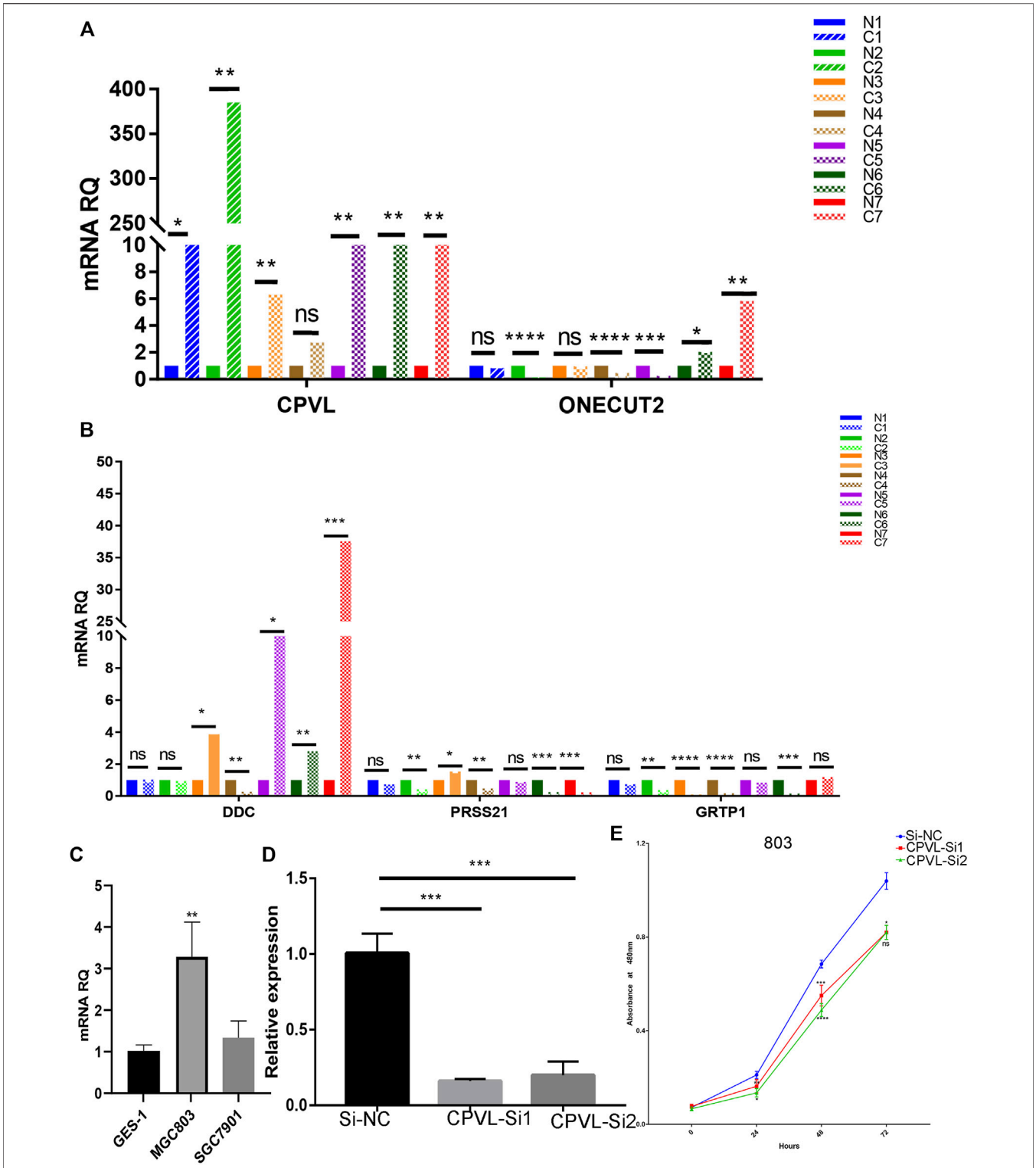


FIGURE 7 | The effect of CPVL on cellular proliferation of Gastric cancer. **(A,B)** Q-PCR analysis of the relative expression levels of *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* in gastric cancer and corresponding adjacent tissues. Where "C" represents the tumor samples and "N" represents the corresponding paracancerous samples. **(C)** Relative expression of CPVL in gastric cancer cell lines. **(D)** Knockdown efficiency of CPVL in MGC803. **(E)** Effects of CPVL knockdown on MGC803 proliferation. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

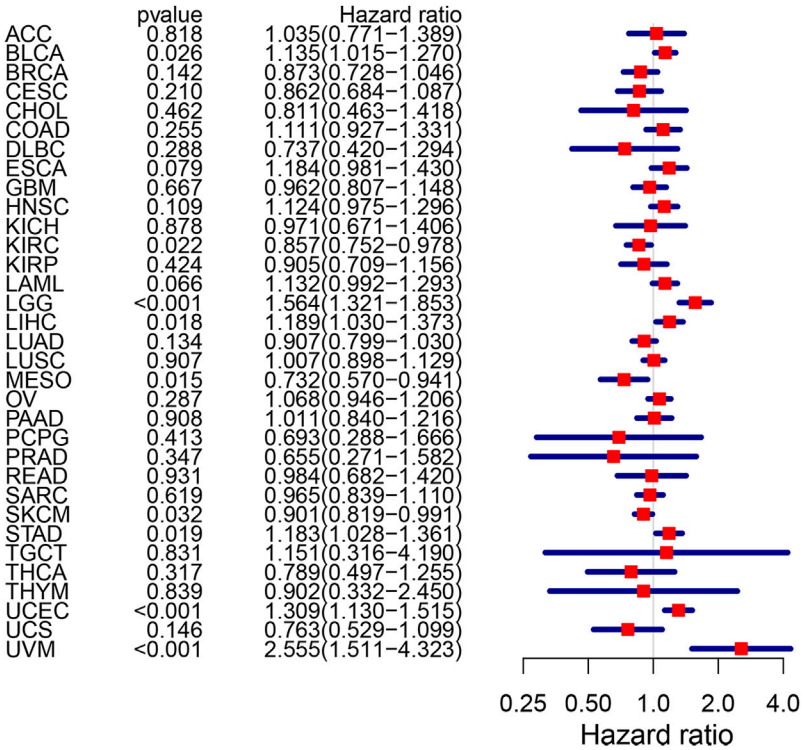


FIGURE 8 | The risk rate of *CPVL* gene in pan-cancer. CPVL is associated with patients' survival in BLCA, KIRC, LCC, LIHC, MESO, SKCM, STAD, UCEC, and UVM ($p < 0.05$). The hazard ratio greater than 1 indicates a high-risk gene(s), meaning that patients with high expression of such gene(s) have a worse prognosis; the hazard ratio less than 1 indicates a low-risk gene(s) for particular cancer, meaning that patients with high expression have a better prognosis.

TABLE 2 | Correlation between *CPVL* and miRNA.

geneName	Coefficient-R		*p*-value
hsa-miR-196a-5p	*CPVL*	−0.09	8.33E-02
hsa-miR-7-5p	*CPVL*	−0.131	1.15E-02
hsa-miR-196b-5p	*CPVL*	−0.185	3.43E-04
hsa-miR-561-5p	*CPVL*	−0.088	8.97E-02

AL122035.2 were significantly reduced in gastric cancer samples, while the expression levels of the four miRNAs in gastric cancer cell lines were significantly higher than the normal controls (**Supplementary Figure S5**). Thus, based on these results, we believe that these lncRNAs and miRNAs could be used to construct ceRNA networks and studying other functions.

DISCUSSION

Gastric cancer is considered as one of the most deadly cancers, ranked as the third leading cause of cancer-associated deaths worldwide (Chen et al., 2014). Exploring the heterogeneity within the tumor can help to deeply understand the microenvironment of gastric cancer (Yan et al., 2018). Recently, it has been validated that CNVs (deletion, insertion or SNPs) in genes are one of the important factors affecting cellular functions (Nakamura, 2009). In a variety of tumors, including gastric cancer, a large number of CNVs are potentially involved in the occurrence and development of tumors (Leary et al., 2008; Despierre et al., 2014; Horpaopan et al., 2015; Xu et al., 2015).

A study using 183 gastric cancer samples identified drugs' targeted genes exhibit a high ratio of copy number gains (CNG) (Labots et al., 2014). It is well-known fact that the CNVs play integral roles in the screening of new prognostic

TABLE 3 | Correlation analysis between lncRNA and hsa-miR-196b-5p or lncRNA and *CPVL* in GC determined by starBase database.

geneName	miRNAname	R	*p*-value	mRNA	R	*p*-value
LINC00472	hsa-miR-196b-5p	−0.381	2.71E-14	*CPVL*	0.166	1.24E-03
TRG-AS1	hsa-miR-196b-5p	−0.304	2.03E-09	*CPVL*	0.286	1.72E-08
LINC01678	hsa-miR-196b-5p	−0.28	4.03E-08	*CPVL*	0.124	1.59E-02
AL158207.2	hsa-miR-196b-5p	−0.226	1.07E-05	*CPVL*	0.147	4.40E-03
AL122035.2	hsa-miR-196b-5p	−0.205	6.79E-05	*CPVL*	0.082	1.15E-01

targets and individualized treatment for patients based on the heterogeneity of their tumors (Hou et al., 2016). Believing that, single-cell sequencing of tumors could serve as a useful tool for exploring tumor heterogeneity (Hou et al., 2016). In the current study, we considered the CNV-score of the cells to differentiate the malignant cells from epithelial cells in GC. Abnormally high expressing genes with CNV-score in the tumor cells were screened out, and top genes were analyzed for their roles in overall survival in gastric cancer patients. Furthermore, we also verified that the level of CNVs and high-risk scores related to the poor prognosis of gastric cancer patients.

The tumor immune process and the host's immune protection mechanism are divided into three phases: elimination, equilibrium, and escape phases (Lazar et al., 2018). Specifically, at the early stages of tumorigenesis and development, the immune system can kill tumor cells to inhibit the development and progression of cancers. With the progression of the tumor, the surviving cancer cells become smarter and can change some of their characteristics to avoid killers' immune cells thus proliferating and metastasizing. We found that high expression of *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* increase the risk of gastric cancer. Interestingly, the expression of *CPVL*, *ONECUT2*, *PRSS21*, and *GRTP1* was linked with the higher ratio of CNVs which further suppressed the level of immune cell infiltration. These results partially explain that the low degree of immune cell infiltration in GC may affect the efficiency of ICIs. High expression of ONECUT2 and PRSS21 have been reported to be associated with poorer prognosis in various cancers including prostate cancer, hepatocellular carcinoma, and gastric cancer (Liu et al., 2021; Pollan et al., 2021; Shen et al., 2021; Sun et al., 2021).

In the current study, *CPVL* showed higher expression than the other genes, thus we selected it for further analyses. *CPVL* is a serine carboxypeptidase that was first characterized in human macrophages (Mahoney et al., 2001). Though, the function of *CPVL* remains unclear in a variety of tumors. So far *CPVL* has not been deeply studied in gastric cancer. However, *CPVL* induced apoptosis in glioma cells through the IFN-γ/STAT1 signaling pathway (Yang et al., 2021). In gastric cancer cell lines, we showed for the first time that *CPVL* on cell proliferation, suggesting that CPVL may be a new prognostic target for gastric cancer. Further comprehensive studies are required to explore its deep function and underlying mechanisms in regulation of gstric cancer.

CONCLUSION

The high risk-scores of *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* could be used to determine the degree of malignancy and prognosis in gastric cancer patients. The increased expression of *CPVL*, *ONECUT2*, *PRSS21*, and *GRTP1* may indicate the advanced stage of GC, as well as a low level of immune cell infiltration in gastric TME. In gastric cancer cell lines, we determined that *CPVL* regulate the cellular proliferation, showing that *CPVL* could be an important predictor in gastric cancer development.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

The sampling and data collection procedures from human participants were reviewed and approved by The ethical committee of the Fifth Affiliated Hospital of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FL: Conceptualization of the study and prepared first draft. HW: Analyzed the data, and revised the first draft. IB: Revised and reviewed manuscript, logical reasoning and language corrections. CG, FR, YT, and BL: Drafted the manuscript. PZ and YM: Conceived idea of the study, supervised overall work and reviewed and revised final draft. All authors read and approved the submission of the final manuscript.

FUNDING

This work was supported by Zhengzhou Major Collaborative Innovation Project (No. 18XTZX12003); Key projects of discipline construction in Zhengzhou University (No. XKZDJC202001); National Key Research and development program in China (No. 2020YFC2006100); Excellent Foreign Scientist Studio of Henan Province in China (No. GZS2018001); Medical service capacity improvement project of Henan Province in China [grant number Yu Wei Medicine (2017) No.66].

ACKNOWLEDGMENTS

We greatly acknowledge the participation of the patients into our study, and doctors and paramedic staff of the fifth affiliated hospital of zhengzhou university for their help in provision of samples. We are also highly thankful to our colleagues who are not listed as authors of this manuscript, for their support in the completion of this study. We also acknowledge the role of institutes and departments involved in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.869967/full#supplementary-material

Supplementary Figure S1 | Quality control and screening of single cells of gastric cancer. **(A)** Filter and normalize the tumor samples (355 tumor cells in total) in the GSE112302 single-cell sequencing data. **(B)** Red indicates the first 1,500 differential genes. **(C)** Sequencing depth and mitochondrial gene distribution. **(D–G)** Principal component analysis to screen marker genes (20 PCs in total).

Supplementary Figure S2 | Take the intersection of the Venn diagram to obtain the hub gene and its expression in a single cell population.

Supplementary Figure S3 | The scatter plot shows the correlation between *CPVL*, *ONECUT2*, *DDC*, *PRSS21*, and *GRTP1* genes and immune checkpoints (*PDCD1*, *CD274*, *PDCD1LG2*, and *CTLA4*) in gastric cancer.

Supplementary Figure S4 | The scatter plot shows the correlation of *CPVL* with immune score and matrix score in pan-cancer.

Supplementary Figure S5 | Relative expression levels of lncRNAs and miRNAs in gastric cancer cell lines.

REFERENCES

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., et al. (2010). The Landscape of Somatic Copy-Number Alteration across Human Cancers. *Nature* 463 (7283), 899–905. doi:10.1038/nature08822

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Campos-Carrillo, A., Weitzel, J. N., Sahoo, P., Rockne, R., Mokhnatkin, J. V., Murtaza, M., et al. (2020). Circulating Tumor DNA as an Early Cancer Detection Tool. *Pharmacol. Ther.* 207, 107458. doi:10.1016/j.pharmthera.2019.107458

Chen, W., Zheng, R., Zhang, S., Zhao, P., Zeng, H., and Zou, X. (2014). Report of Cancer Incidence and Mortality in China, 2010. *Ann. Transl. Med.* 2 (7), 61. doi:10.3978/j.issn.2305-5839.2014.04.05

Chung, W., Eum, H. H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., et al. (2017). Single-cell RNA-Seq Enables Comprehensive Tumour and Immune Cell Profiling in Primary Breast Cancer. *Nat. Commun.* 8, 15081. doi:10.1038/ncomms15081

Cieslik, M., and Chinnaiyan, A. M. (2018). Cancer Transcriptome Profiling at the Juncture of Clinical Translation. *Nat. Rev. Genet.* 19 (2), 93–109. doi:10.1038/nrg.2017.96

Cochain, C., Vafadarnejad, E., Arampatzi, P., Pelisek, J., Winkels, H., Ley, K., et al. (2018). Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circ. Res.* 122 (12), 1661–1674. doi:10.1161/circresaha.117.312509

Despierre, E., Moisse, M., Yesilyurt, B., Sehouli, J., Braicu, I., Mahner, S., et al. (2014). Somatic Copy Number Alterations Predict Response to Platinum Therapy in Epithelial Ovarian Cancer. *Gynecol. Oncol.* 135 (3), 415–422. doi:10.1016/j.ygyno.2014.09.014

Guo, X.-J., Lu, J.-C., Zeng, H.-Y., Zhou, R., Sun, Q.-M., Yang, G.-H., et al. (2021). CTLA-4 Synergizes with PD1/PD-L1 in the Inhibitory Tumor Microenvironment of Intrahepatic Cholangiocarcinoma. *Front. Immunol.* 12, 705378. doi:10.3389/fimmu.2021.705378

Horpaopan, S., Spier, I., Zink, A. M., Altmüller, J., Holzapfel, S., Laner, A., et al. (2015). Genome-wide CNV Analysis in 221 Unrelated Patients and Targeted High-Throughput Sequencing Reveal Novel Causative Candidate Genes for Colorectal Adenomatous Polyposis. *Int. J. Cancer* 136 (6), E578–E589. doi:10.1002/ijc.29215

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., et al. (2016). Single-cell Triple Omics Sequencing Reveals Genetic, Epigenetic, and Transcriptomic Heterogeneity in Hepatocellular Carcinomas. *Cell Res.* 26 (3), 304–319. doi:10.1038/cr.2016.23

Ilson, D. H. (2017). Advances in the Treatment of Gastric Cancer. *Curr. Opin. Gastroenterol.* 33 (6), 473–476. doi:10.1097/mog.0000000000000395

Kacew, A. J., Harris, E. J., Lorch, J. H., Haddad, R. I., Chau, N. G., Rabinowits, G., et al. (2019). Chromosome 3q Arm Gain Linked to Immunotherapy Response in Advanced Cutaneous Squamous Cell Carcinoma. *Eur. J. Cancer* 113, 1–9. doi:10.1016/j.ejca.2019.03.004

Labots, M., Buffart, T. E., Haan, J. C., van Grieken, N. C. T., Tijssen, M., van de Velde, C. J. H., et al. (2014). High-level Copy Number Gains of Established and Potential Drug Target Genes in Gastric Cancer as a Lead for Treatment Development and Selection. *Cell Oncol.* 37 (1), 41–52. doi:10.1007/s13402-013-0162-4

Lazar, D. C., Avram, M. F., Romosan, I., Cornianu, M., Taban, S., and Goldis, A. (2018). Prognostic Significance of Tumor Immune Microenvironment and Immunotherapy: Novel Insights and Future Perspectives in Gastric Cancer. *World J. Gastroenterol.* 24 (32), 3583–3616. doi:10.3748/wjg.v24.i32.3583

Leary, R. J., Lin, J. C., Cummins, J., Boca, S., Wood, L. D., Parsons, D. W., et al. (2008). Integrated Analysis of Homozygous Deletions, Focal Amplifications, and Sequence Alterations in Breast and Colorectal Cancers. *Proc. Natl. Acad. Sci. U.S.A.* 105 (42), 16224–16229. doi:10.1073/pnas.0808041105

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). StarBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucleic Acids Res.* 42 (Database issue), D92–D97. doi:10.1093/nar/gkt1248

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.can-17-0307

Liu, D., Zhang, T., Chen, X., Zhang, B., Wang, Y., Xie, M., et al. (2021). ONECUT2 Facilitates Hepatocellular Carcinoma Metastasis by Transcriptionally Upregulating FGF2 and ACLY. *Cell Death Dis.* 12 (12), 1113. doi:10.1038/s41419-021-04410-3

Loraine, A. E., Blakley, I. C., Jagadeesan, S., Harper, J., Miller, G., and Firon, N. (2015). Analysis and Visualization of RNA-Seq Expression Data Using RStudio, Bioconductor, and Integrated Genome Browser. *Methods Mol. Biol.* 1284, 481–501. doi:10.1007/978-1-4939-2444-8_24

Mahoney, J. A., Ntolosi, B., DaSilva, R. P., Gordon, S., and McKnight, A. J. (2001). Cloning and Characterization of CPVL, a Novel Serine Carboxypeptidase, from Human Macrophages. *Genomics* 72 (3), 243–251. doi:10.1006/geno.2000.6484

McEligot, A. J., Poynor, V., Sharma, R., and Panangadan, A. (2020). Logistic LASSO Regression for Dietary Intakes and Breast Cancer. *Nutrients* 12 (9), 2652. doi:10.3390/nu12092652

Meacham, C. E., and Morrison, S. J. (2013). Tumour Heterogeneity and Cancer Cell Plasticity. *Nature* 501 (7467), 328–337. doi:10.1038/nature12624

Müller, S., Liu, S. J., Di Lullo, E., Malatesta, M., Pollen, A. A., Nowakowski, T. J., et al. (2016). Single-cell Sequencing Maps Gene Expression to Mutational Phylogenies in PDGF - and EGF -driven Gliomas. *Mol. Syst. Biol.* 12 (11), 889. doi:10.15252/msb.20166969

Nakamura, Y. (2009). DNA Variations in Human and Medical Genetics: 25 Years of My Experience. *J. Hum. Genet.* 54 (1), 1–8. doi:10.1038/jhg.2008.6

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma. *Science* 344 (6190), 1396–1401. doi:10.1126/science.1254257

Pernot, S., Voron, T., Perkins, G., Lagorce-Pages, C., Berger, A., and Taieb, J. (2015). Signet-ring Cell Carcinoma of the Stomach: Impact on Prognosis and Specific Therapeutic Challenge. *World J. Gastroenterol.* 21 (40), 11428–11438. doi:10.3748/wjg.v21.i40.11428

Pollan, S. G., Teng, P. C., Jan, Y. J., Livingstone, J., Huang, C., Kim, M., et al. (2021). Loss of CDCP1 Triggers FAK Activation in Detached Prostate Cancer Cells. *Am. J. Clin. Exp. Urol.* 9 (4), 350–366.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial Reconstruction of Single-Cell Gene Expression Data. *Nat. Biotechnol.* 33 (5), 495–502. doi:10.1038/nbt.3192

Shen, C., Wang, J., Xu, Z., Zhang, L., Gu, W., and Zhou, X. (2021). ONECUT2 Which Is Targeted by Hsa-miR-15a-5p Enhances Stemness Maintenance of Gastric Cancer Stem Cells. *Exp. Biol. Med. (Maywood)* 246 (24), 2645–2659. doi:10.1177/15353702211038496

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer Statistics, 2017. *CA Cancer J. Clin.* 67 (1), 7–30. doi:10.3322/caac.21387

Smyth, G. K., Michaud, J., and Scott, H. S. (2005). Use of Within-Array Replicate Spots for Assessing Differential Expression in Microarray Experiments. *Bioinformatics* 21 (9), 2067–2075. doi:10.1093/bioinformatics/bti270

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177 (7), 1888–1902. doi:10.1016/j.cell.2019.05.031

Sun, Z., Mao, Y., Zhang, X., Lu, S., Wang, H., Zhang, C., et al. (2021). Identification of ARHGEF38, NETO2, GOLM1, and SAPCD2 Associated with Prostate Cancer Progression by Bioinformatic Analysis and Experimental Validation. *Front. Cell Dev. Biol.* 9, 718638. doi:10.3389/fcell.2021.718638

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., et al. (2016). Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. *Science* 352 (6282), 189–196. doi:10.1126/science.aad0501

Van Cutsem, E., Sagaert, X., Topal, B., Haustermans, K., and Prenen, H. (2016). Gastric Cancer. *Lancet* 388 (10060), 2654–2664. doi:10.1016/s0140-6736(16)30354-3

Waniczek, D., Lorenc, Z., Śnietura, M., Wesecki, M., Kopec, A., and Muc-Wierzgoń, M. (2017). Tumor-Associated Macrophages and Regulatory T Cells Infiltration and the Clinical Outcome in Colorectal Cancer. *Arch. Immunol. Ther. Exp.* 65 (5), 445–454. doi:10.1007/s00005-017-0463-9

Weinberg, R. A. (2014). Coming Full Circle-From Endless Complexity to Simplicity and Back Again. *Cell* 157 (1), 267–271. doi:10.1016/j.cell.2014.03.004

Xu, H., Zhu, X., Xu, Z., Hu, Y., Bo, S., Xing, T., et al. (2015). Non-invasive Analysis of Genomic Copy Number Variation in Patients with Hepatocellular Carcinoma by Next Generation DNA Sequencing. *J. Cancer* 6 (3), 247–253. doi:10.7150/jca.10747

Yan, H. H. N., Siu, H. C., Law, S., Ho, S. L., Yue, S. S. K., Tsui, W. Y., et al. (2018). A Comprehensive Human Gastric Cancer Organoid Biobank Captures Tumor Subtype Heterogeneity and Enables Therapeutic Screening. *Cell Stem Cell* 23 (6), 882–897. doi:10.1016/j.stem.2018.09.016

Yang, H., Liu, X., Zhu, X., Li, X., Jiang, L., Zhong, M., et al. (2021). CPVL Promotes Glioma Progression via STAT1 Pathway Inhibition through Interactions with the BTK/p300 axis. *JCI Insight* 6 (24), e146362. doi:10.1172/jci.insight.146362

Zurleni, T., Gjoni, E., Altomare, M., and Rausei, S. (2018). Conversion Surgery for Gastric Cancer Patients: A Review. *World J. Gastrointest. Oncol.* 10 (11), 398–409. doi:10.4251/wjgo.v10.i11.398

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Wen, Bukhari, Liu, Guo, Ren, Tang, Mi and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fgene.2022.898507

Identification and Validation of Necroptosis-Related LncRNA Signature in Hepatocellular Carcinoma for Prognosis Estimation and Microenvironment Status

Cong Chen[1†], Yumeng Wu[2†], Kang Chen[1], Zicong Xia[1], Xiaokan Liu[1], Chaojie Zhang[1], Hui Zhao[1] and Aiguo Shen[2]**

[1]Department of Interventional Radiology, Affiliated Hospital of Nantong University, Nantong, China, [2]Cancer Research Center Nantong, Affiliated Tumor Hospital of Nantong University, Nantong, China

OPEN ACCESS

Edited by:
Ruowen Zhang,
Jiahehongsheng (Shenzhen) Health Industry Group, China

Reviewed by:
Feng Jiang,
Fudan University, China
Zhigang Jie,
The First Affiliated Hospital of Nanchang University, China

***Correspondence:**
Hui Zhao
zhaohui800@163.com
Aiguo Shen
aiguoshen_nt@126.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 17 March 2022
Accepted: 16 May 2022
Published: 08 June 2022

Citation:
Chen C, Wu Y, Chen K, Xia Z, Liu X, Zhang C, Zhao H and Shen A (2022) Identification and Validation of Necroptosis-Related LncRNA Signature in Hepatocellular Carcinoma for Prognosis Estimation and Microenvironment Status. Front. Genet. 13:898507. doi: 10.3389/fgene.2022.898507

Background: Hepatocellular carcinoma (HCC) is among malignancies with the highest fatality toll globally and minimal therapeutic options. Necroptosis is a programmed form of necrosis or inflammatory cell death, which can affect prognosis and microenvironmental status of HCC. Therefore, we aimed to explore the prognostic value of necroptosis-related lncRNAs (NRLs) in HCC and the role of the tumor microenvironment (TME) in immunotherapy.

Methods: The RNA-sequencing data and clinical information were downloaded from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). NRLs were identified by Pearson correlation analysis. The signature was constructed using the LASSO–Cox regression analysis and evaluated using the receiver operating characteristic curve (ROC) and the area under the Kaplan–Meier curve. The nomogram was built based on clinical information and risk score. Gene set enrichment analysis (GSEA), immunoassay, half-maximum inhibitory concentration (IC_{50}) analysis of the risk group, and the HCC subtype identification based on NRLs were also carried out. Finally, we detected the expression of lncRNAs in HCC tissues and cell lines *in vitro*.

Results: A total of 508 NRLs were screened out, and seven NRLs were constructed as a risk stratification system to classify patients into distinct low- and high-risk groups. Patients in the high-risk group had a significantly lower overall survival (OS) than those in the low-risk group. Using multivariate Cox regression analysis, we found that the risk score was an independent predictor of OS. Functional analysis showed that the immune status of different patients was different. The IC_{50} analysis of chemotherapy demonstrated that patients in the high-risk group were more sensitive to commonly prescribed drugs. qRT-PCR showed that three high-risk lncRNAs were upregulated in drug-resistant cells, and the expression in HCC tissues was higher than that in adjacent tissues.

Conclusion: The prediction signature developed in this study can be used to assess the prognosis and microenvironment of HCC patients, and serve as a new benchmark for HCC treatment selection.

Keywords: hepatocellular carcinoma, prognostic signature, microenvironment, lncRNA, necroptosis, immune infiltrate

INTRODUCTION

According to the 2020 global cancer statistics, primary liver cancer ranks the sixth and the third in the worldwide incidence rate and mortality rate, respectively, and hepatocellular carcinoma (HCC) accounts for the overwhelming majority of liver cancer cases (Sung et al., 2021). Even though early diagnosis of HCC has developed rather rapidly in recent years, in most patients HCC is already in the intermediate or advanced stage at the time of diagnosis, having missed the best time for surgical resection. New treatments, such as transcatheter arterial chemoembolization (TACE), radiofrequency ablation, immunotherapy, and targeted therapy, can bring hope to patients with advanced liver cancer (Anwanwan et al., 2020). However, the overall survival (OS) of patients with HCC remains unsatisfactory, and changes in the immune microenvironment possibly play a pivotal role in immune escape and resistance in HCC. In recent years, immunotherapy—represented by immune checkpoint inhibitors, adoptive cell therapy (ACT), and tumor vaccines—has brought new hope to patients with advanced HCC (Mizukoshi and Kaneko, 2019). However, only a small fraction of patients can benefit from immunotherapy (Fu et al., 2019), which may be due to tumor heterogeneity and changes in immune-related factors in the tumor microenvironment (TME). Therefore, finding an indicator that can predict the effect of immunotherapy and the state of the microenvironment is crucial for improving the prognosis in patients with HCC.

Cell death is a complex process, that is, achieved through pathological and physiological ways (Tang et al., 2019). One form of cell death is necroptosis, which can be acquired in a programmed way during the development of certain organisms. Unlike typical apoptosis and necrosis, necroptosis induces cell death when the apoptotic mechanism fails. Necroptosis is closely related to the immune microenvironment and can induce cell rupture, activate inflammatory response while releasing cellular contents, and promote infiltration of many inflammatory cells (Green, 2019). Receptor-interacting serine/threonine-protein kinase 3 (RIPK3) can act as a critical regulator of necroptosis to affect the function of immune cells by regulating the activation of natural killer T (NKT) cells and dendritic cells (DCs) (Degterev et al., 2008). The pan-caspase inhibitor Z-VAD (OH)-FMK (zVAD) has been reported to induce necroptosis in melanoma *via* reducing tumor infiltration by regulatory T cells (Tregs) while increasing DC and $CD8^+$ T cells to reduce tumor growth (Werthmöller et al., 2015). Necroptosis also plays a vital role in HCC since heparinase can induce necrotic proliferation of microvascular endothelial cells and promote liver cancer metastasis (Chen et al., 2021). Hence, necroptosis may be a potential target for HCC therapy.

Long noncoding RNAs (lncRNAs) are a class of noncoding RNAs. lncRNAs are closely related to HCC. Specifically, downregulation of lncRNA growth arrest-specific 5 (GAS5) in HCC promotes proliferation and drug resistance through the decrease of phosphatase and tensin homolog (PTEN) expression (Wang et al., 2020). lncRNA small nucleolar RNA host gene 3 (SNHG3) induces epithelial–mesenchymal transition (EMT) and sorafenib resistance by regulating the miR-128/cluster of differentiation 151 (CD151) pathway in HCC (Zhang et al., 2019), having the potential to affect necroptosis through different pathways such as H19-derived miR-675 targeting FAS-associated death domain protein (FADD) (Harari-Steinfeld et al., 2021). In addition, lncRNAs can protect tumor cells from necroptosis by suppressing the expression of some related proteins (Tao et al., 2019). There is also a close correlation between lncRNAs and TME. Long intergenic non-protein coding RNA 665 (LINC00665) affects the level of macrophage and DC infiltration, suppresses Tregs, and prevents T cell failure by targeting lncRNA five prime to Xist (FTX) as competing endogenous RNA (ceRNA) (Zhang et al., 2020a). lncRNA T cell leukemia/lymphoma 6 (TCL6) positively correlates with tumor-infiltrating lymphocyte (TIL) infiltration and immune checkpoint molecules such as cytotoxic T-cell lymphocyte-associated protein 4 (CTLA-4), programmed death receptor 1 (PD-1), and its ligand (PD-L1) (Zhang et al., 2020b). Exploring the lncRNA signatures associated with necroptosis and their role in HCC treatment requires special attention.

In this study, we first downloaded lncRNA expression profiles and clinical information from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC); then, we constructed a necroptosis-related lncRNA prognostic signature, which allowed us to analyze TME, immune cell infiltration, immune checkpoints, human leukocyte antigens (HLA), functional enrichment, and drug sensitivity in different risk groups. Finally, we validated the lncRNAs in the signature using tissues and cell lines. This study may provide a new reference for selecting HCC treatment methods and predicting prognosis.

MATERIALS AND METHODS

Datasets and Preprocessing

The RNA-sequencing data (TPM format) used for HCC samples were downloaded from TCGA (https://portal.gdc.cancer.gov/). After excluding the patients from repeated sequencing, those lacking complete follow-up information, and those with 0 survival days, a total of 50 normal samples and 365 tumor

samples were included. Next, bioinformatics analysis, survival analysis, and model building were performed on these samples. Data from 231 HCC patients were additionally downloaded from ICGC (https://dcc.icgc.org/projects/LIRI-JP) for external validation using the same exclusion criteria. The "SVA" R package was used to perform background correction, normalization, and expression estimates for internal and external validation on the genes associated with the modeling (**Supplementary Figure S1**). R software (version 4.0.5) was used to conduct all of the analyses.

Construction and Validation of Prognostic Signature

Necroptosis-related genes (NRGs) were extracted from previous studies (**Supplementary File S1**) (Zhao et al., 2021). Differential expression of NRGs in normal and HCC samples was analyzed using the limma R package, with $p < 0.05$ and $|\log_2 FC| > 0.5$ as thresholds. Having performed the Pearson correlation analysis on all lncRNAs and having identified differentially expressed NRGs ($p < 0.001$, correlation coefficient >0.4), we finally screened necroptosis-related lncRNAs (NRLs) for subsequent bioinformatics analysis.

Univariate Cox proportional-hazard regression analysis filtered lncRNAs linked to survival ($p < 0.05$) in the batch-adjusted cohort. A risk model was then built using Least Absolute Shrinkage and Selection Operator (LASSO) regression with 10-fold cross-validation and run for 1,000 cycles with 1,000 random stimulations to avoid overfitting effects (Tibshirani, 1997; Simon et al., 2011). After integrating the gene expression values weighted by the LASSO–Cox coefficient, the following formula for the risk score was established:

$$risk\ score = \sum \left[Exp\left(IncRNA\right) \times coef\left(IncRNA\right) \right] \quad (1)$$

where Exp (lncRNA) is the expression of survival-related lncRNAs, and coef (lncRNA) is the associated regression coefficient. Patients in the TCGA and ICGC cohorts were divided into high- and low-risk groups based on the median risk score. Kaplan–Meier (K-M) curves were plotted to find differences in OS between the risk groups, and log-rank tests were performed on the results. Likewise, receiver operating characteristic (ROC) curves were plotted using the survival ROC R package. The area under the curve (AUC) was calculated to assess the model's accuracy.

Finally, we analyzed the clinicopathological information in the dataset through univariate and multivariate Cox regression analysis. We used a nomogram that included tumor-node-metastasis (TNM) staging and risk score to predict the survival of HCC patients at 1, 3, and 5 years. The nomogram's accuracy was measured using ROC curves. The *p*-values in analyzing the differentially expressed genes were adjusted.

Immunology and Cluster Analysis

We utilized different algorithms such as TIMER, CIBERSORT, CIBERSORT-abs, QUANTISEQ, MCP-counter, XCELL, and EPIC to estimate the abundance and correlation of immune cells in different risk groups. In addition, the single-sample gene set enrichment analysis (ssGSEA) algorithm was selected to evaluate immune cells and immune-related functions (Rooney et al., 2015). The enrichment fraction of 29 immunological characteristics per sample in TME was calculated using the R package GSVA (version 1.34.0). The estimation of stromal and immune cells in malignant tumor tissues using the expression data (ESTIMATE) algorithm was employed to calculate the immune score, stromal score, and tumor purity to reflect the state of the immune microenvironment (Yoshihara et al., 2013). Regarding the TCGA cohort, we used a nonnegative matrix factorization (NMF) clustering algorithm to cluster the HCC samples from the NRLs. The ICGC cohort was verified using the same candidate genes. The K value refers to the value selected when the size of the correlation coefficient starts to decrease with the optimal number of clusters. The class mapping analysis evaluated the similarity of subtype classification among different datasets. Simultaneously, the dimensionality reduction analysis was performed on the expression data of the candidate genes, and the principal component analysis (PCA) method was adopted to verify the subtype distribution. In addition, the nearest template prediction (NTP) algorithm was applied to predict the different risk groups of genetic signatures in both cohorts. The prediction results were compared with the classification results of the NMF algorithm.

Functional Enrichment Analysis

Gene pathways were annotated with the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) using the "clusterProfiler" software R package. *p*-value < 0.05 and q-value < 0.05 indicated significantly enriched pathways. The Gene Set Enrichment Analysis (GSEA) algorithm is an enrichment method based on expression profiles, and calculates the estimated proportion of a particular pathway or feature in different clusters. We used the gene set (Kegg.v7.4.symbols.gmt) for GSEA analysis (http://www.gsea-msigdb.org/gsea/index.jsp), where $p < 0.05$ and false-discovery rate (FDR) < 0.05 were considered statistically significant. One thousand permutations of gene sets were done for each analysis to provide a normalized enrichment score (NES). The Benjamini–Hochberg (BH) multiple testing correction was used to adjust the *p*-values.

Drug Sensitivity Analysis

Half-maximum inhibitory concentration (IC_{50}) values of chemotherapy drugs were obtained from the Genomics of Cancer Drug Sensitivity (GDSC) database (https://www.cancerrxgene.org/) (Geeleher et al., 2014) and calculated using the "PRrophytic" R package in R software. The difference in the IC_{50} between the different risk groups was analyzed by the Wilcoxon signed-rank test. The results are shown as box plots.

Cell Lines and Culture Conditions

All cell lines were purchased from the National Certified Cell Culture Collection Center (Shanghai, China). Huh7 and HepG2 cells were cultured in DMEM medium (Gibco) supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin. Hep3B cells were cultured in MEM medium (HyClone), supplemented with 10% fetal bovine serum, 1% penicillin–streptomycin, and 1% non-essential amino acids (Gibco, #11140050). SNU-387 and L-02 cells were cultured in RPMI medium (Gibco) supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin. Cell

TABLE 1 | PCR primer sequences.

Gene	Primers
HCG27	F: CAGCCCTGGGTGGAGATTTAAGATG
	R: AGGTGGGTGGGAAGAGGTGTTAC
C2orf27A	F: CATGCGGTCCTCCAGGTTCAAC
	R: CTCTGCCAACCAACTGCCCATC
BACE1-AS	F: TGGCTGTTGCTGAAGAATGTGACTC
	R: CAACCTTCGTTTGCCCAAGAAAGTG
SNHG4	F: AACTCCTGACCTTGCGATTTGCC
	R: GAGGTTGTAGTGAGCCGAGATTGC
MIR210HG	F: AATAACCAAGCCGAGTTGCCTCTG
	R: TCTGGAGCACACAAAGGGAACAAG
SNHG3	F: CAGCCGTTAAGCCATTTGGAACTTG
	R: CAACCCTGACCTCAACACCTTGG
HCG11	F: CTGAGGCAGGAGAATCACTTGAACC
	R: TGAGATGGAGTCTTGCTGTGTTGC

culture took place in a cell incubator at 37°C under 5% carbon dioxide and 10% humidity conditions. None of the cell lines used in this study were tested for mycoplasma contamination.

Cell Viability and Drug Sensitivity

Cells were seeded in 96-well plates at a density of 5,000 cells/well and placed in a 37°C, 5% CO_2 incubator for 24 h. We added doxorubicin (MCE, #HY-15142A), cisplatin (MCE, #HY-17394), and sorafenib (MCE, #HY-10201) to the experimental group according to the concentration gradient. After 48 h, the plates were removed from the incubator and placed in a dark environment to add 10 μl of CCK8 reagent (Vazyme, #A311-02) to each well. Then, the plates were returned to the incubator for 1–2 h. The optical density (OD) value was measured with a microplate reader (Thermo, Multiskan FC), and GraphPad

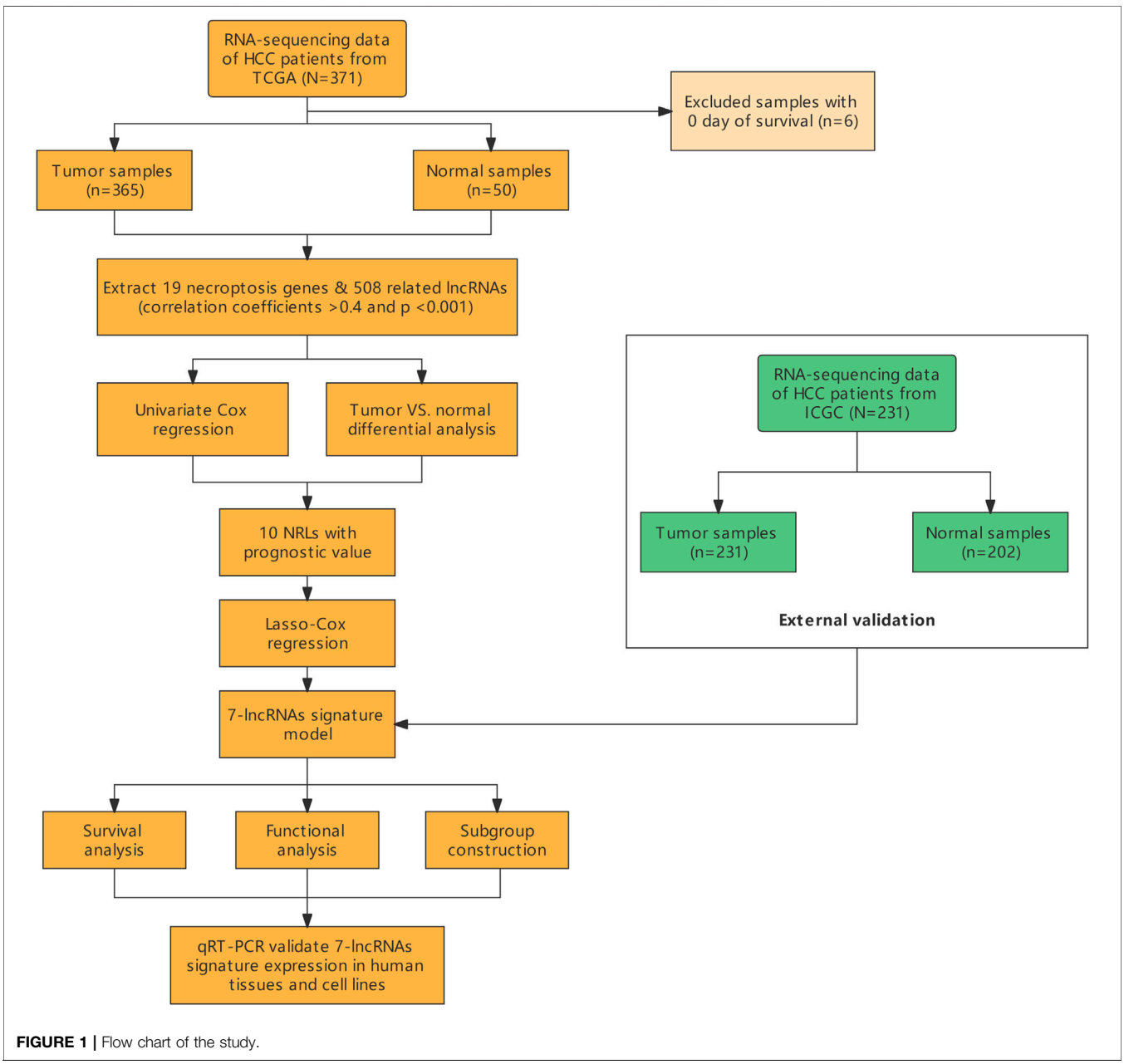


FIGURE 1 | Flow chart of the study.

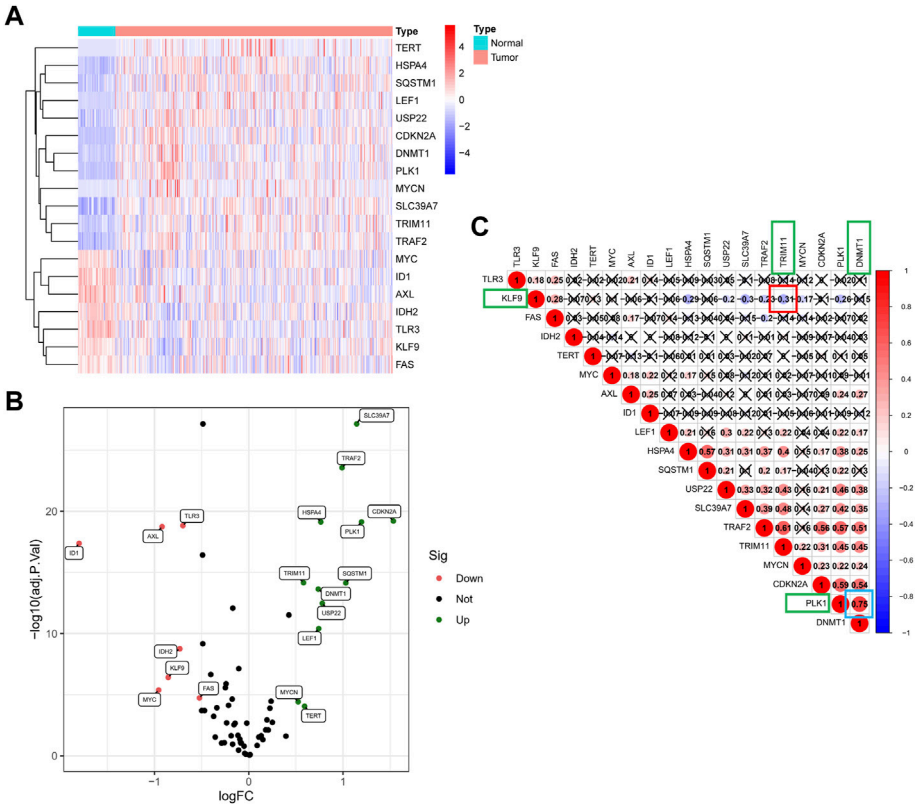


FIGURE 2 | The landscape of NRGs in the TCGA cohort. **(A)** Heat map of the differentially expressed mRNAs in tumor tissues and adjacent normal tissues. **(B)**The volcano plot of 19 differentially expressed genes. **(C)** Correlation analysis of 19 NRGS revealed that DNMT1 has the strongest positive correlation with PLK1 (r = 0.75), and TRIM11 has the strongest negative correlation with KLF9 (r = −0.31).

(Version 9.3.1.471) was used to calculate the IC_{50} value after the exportation of data.

Quantitative Real-Time PCR

We obtained 12 pairs of HCC tissues and paracancerous tissues from the Department of Pathology, the Affiliated Cancer Hospital of Nantong University. Total RNAs from the tissue samples and cell lines were extracted using an RNA isolation kit (Vazyme, #RC112-01), and were then used to synthesize complementary DNA (cDNA) with the help of the cDNA Synthesis Kit (Vazyme, #R233-01) in line with the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) was conducted on the SteponePlus (Applied Biosystems) using SYBR qPCR Master Mix (Vazyme, #Q511-02) and 10 μM primers. Relative expression values were normalized to the control gene (GADPH). The primer pairs used in this study are shown in **Table 1**.

RESULTS

A total of 365 HCC patients from the TCGA cohort and 231 HCC patients from the ICGC (LIRI-JP) cohort were finally enrolled. The overview of this study is presented as a flowchart in **Figure 1**.

The Landscape of Necroptosis-Related Genes in The Cancer Genome Atlas Cohort

In total, 19 of 67 NRGs showed significant differences in expression (**Figure 2A**); specifically, 12 genes were upregulated, and seven genes were downregulated (**Figure 2B**) (**Supplementary File S2**). As shown in **Figure 2C**, the correlation analysis of the 19 NRGs showed that DNA methyltransferase 1 (DNMT1) had the strongest positive correlation with polo-like kinase1 (PLK1) (r = 0.75) and that tripartite motif-containing protein 11 (TRIM11) had the strongest negative correlation with kruppel-like factor 9 (KLF9) (r = −0.31).

Identification and Validation of Necroptosis-Related Long Noncoding RNAs

Having analyzed the correlation between the 19 NRGs and all annotated lncRNAs, we obtained 365 tumor samples and 50 normal samples from the TCGA cohort. Finally, we identified 508 NRLs (correlation coefficient >0.4 and $p < 0.001$), as presented in **Figure 3A**, showing the network diagram between NRGs and lncRNAs. Univariate Cox regression analysis was performed in the batch-adjusted TCGA-HCC

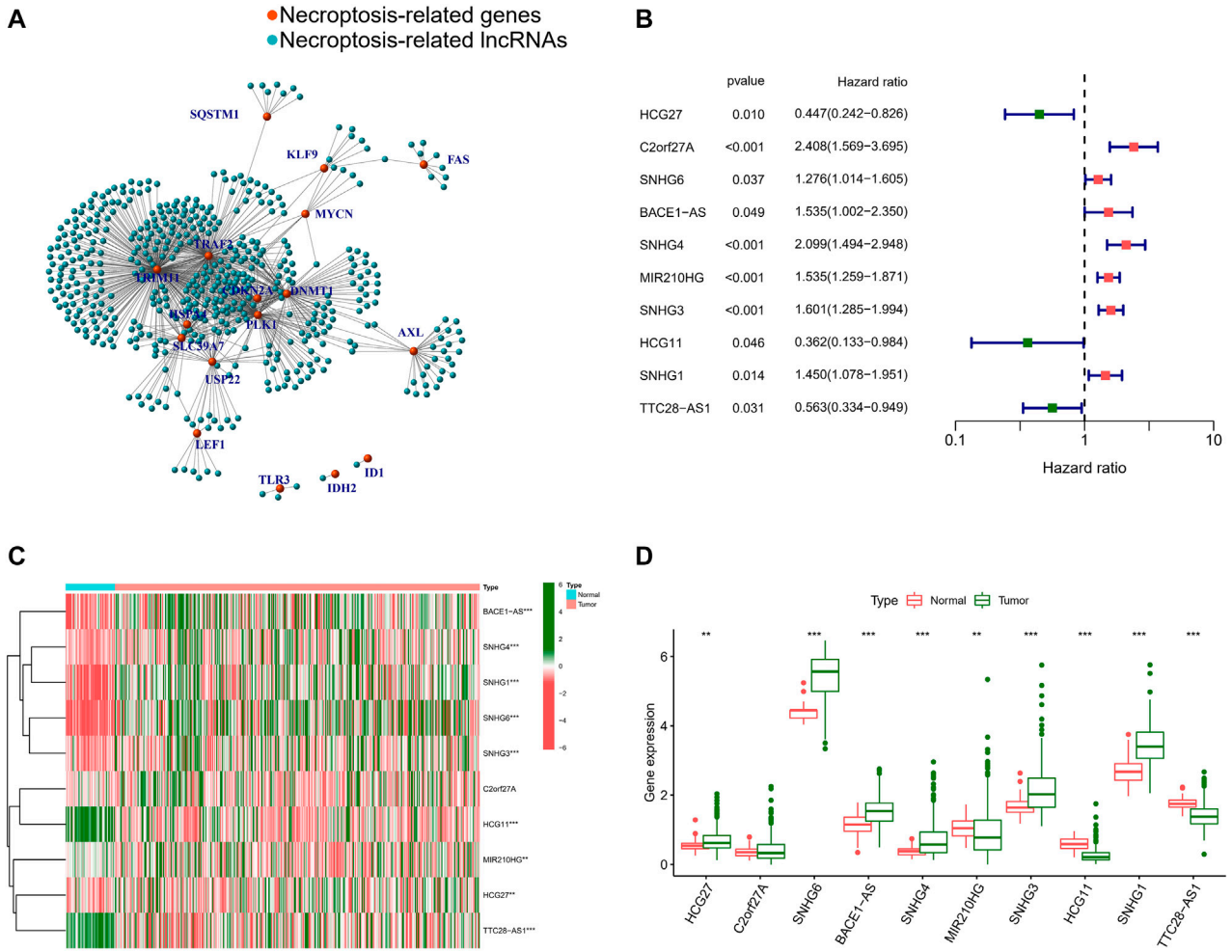


FIGURE 3 | Identification of NRLs. **(A)** The network between NRGs and NRLs (correlation coefficients >0.4 and $p < 0.001$). **(B)** The prognostic NRLs were extracted by univariate Cox regression analysis. **(C)** Heatmap of prognostic NRLs in tumour tissues and adjacent normal tissues. **(D)** The expression of prognostic NRLs in tumour tissues and adjacent normal tissues. (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

cohort to determine the NRLs and their association with survival. Finally, 10 NRLs were screened for subsequent analysis (all $p < 0.05$) (**Figure 3B**). The Wilcoxon test showed that HLA complex group 27 (HCG27), small nucleolar RNA host gene 6 (SNHG6), antisense transcript of BACE1 (BACE1-AS), small nucleolar RNA host gene 4 (SNHG4), small nucleolar RNA host gene 3 (SNHG3), and small nucleolar RNA host gene 1 (SNHG1) were highly expressed in tumors, while MIR210 host gene (MIR210HG), HLA complex group 11 (HCG11), and TTC28 antisense RNA 1 (TTC28-AS1) were highly expressed in normal samples (**Figures 3C,D**).

Construction of the Risk Signature

Based on the optimal value of λ, we performed LASSO regression analysis on these 10 prognosis-related NRLs and screened seven NRLs (**Figures 4A,B**) to avoid overfitting of the prognostic signature. Then, we used multiple Cox regression analysis (ENTER method) to construct a risk stratification system, and showed that HCG27 and HCG11 were moderate-risk genes (**Figure 4C**). Finally, by combining the expression levels and regression coefficients of the seven NRLs (**Figure 4D**), we were able to derive the formula for the risk score of HCC patients: risk score = (−0.7184 × HCG27) + (0.4253 × C2orf27A) + (0.3929 × BACE1-AS) + (0.6010 × SNHG4) + (0.4291 × MIR210HG) + (0.1360 × SNHG3) + (−0.4251 × HCG11).

Validation of the Risk Signature

We calculated the risk scores for patients in the TCGA cohort based on the risk score formula. We selected 231 tumor samples and 202 normal samples from the ICGC cohort as a validation set to test the stability of the signature. We then evaluated the predictive performance for OS using time-dependent ROC curves; the AUC for the TCGA cohort was 0.745, 0.727, and 0.653 at 1, 3, and 5 years,

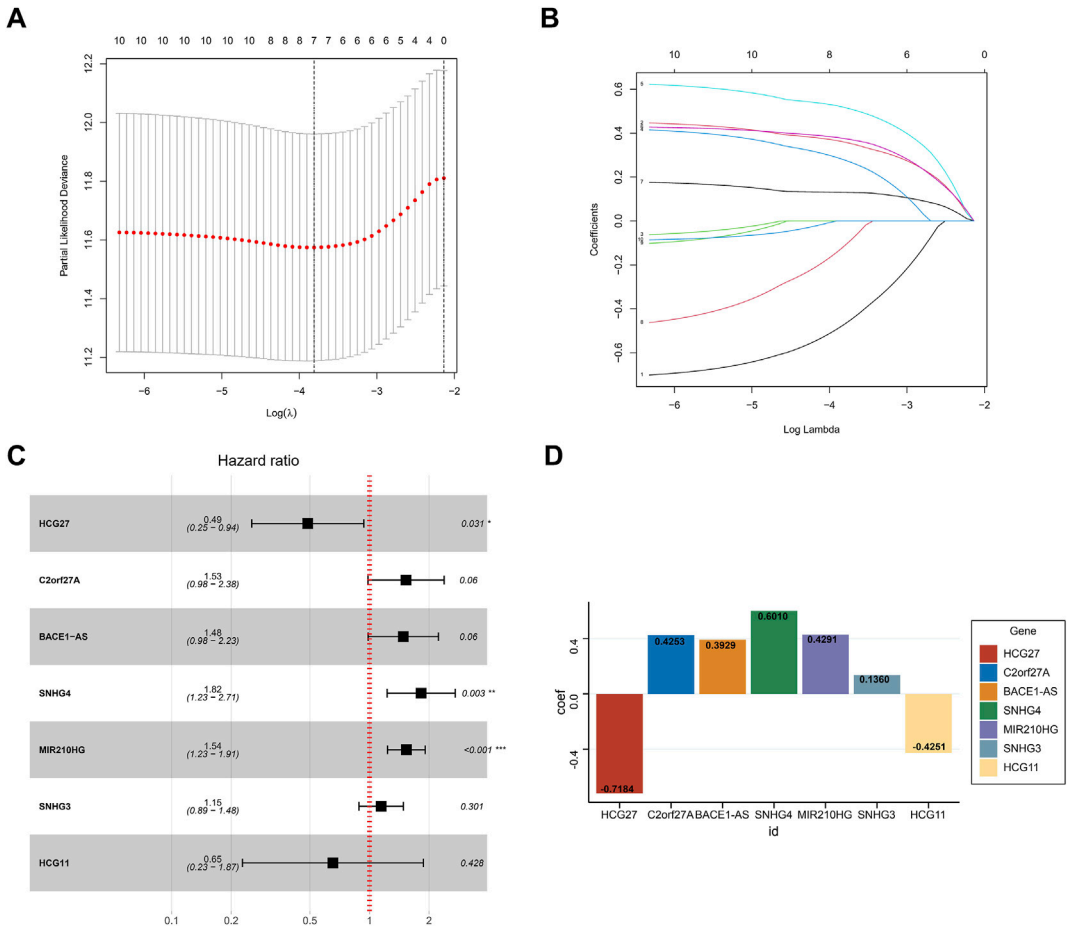


FIGURE 4 | Construction of the prognostic signature. **(A)** The LASSO coefficient profiles of seven NRLs. **(B)** The ten-fold cross-validation for variable selection in the LASSO model. **(C)** Multivariate Cox analysis of the seven NRLs (ENTER method). **(D)** The regression coefficient of the seven NRLs in the signature.

respectively (**Figure 5A**). The AUC of the ICGC cohort was 0.646, 0.632, and 0.613 in the same periods (**Figure 5B**). Kaplan–Meier curves showed that the OS of patients in the high-risk group was significantly lower than that in the low-risk group in both cohorts (all $p < 0.01$) (**Figures 5C,D**). In addition, we compared the risk score distribution and survival status of the high-risk group and the low-risk group using the risk score formula (**Figures 5E,F**).

Afterward, we performed univariate and multivariate Cox regression analysis on clinical characteristics and risk score to determine whether the risk score could serve as an independent prognostic factor for OS in HCC patients. Based on univariate Cox regression analysis, there was a significant association between the risk score and OS (TCGA cohort: HR = 1.462, 95% CI = 1.331–1.607; ICGC cohort: HR = 2.203, 95% CI = 1.519–3.195) (**Figures 6A,C**). After adjusting for other confounding factors, the risk score proved to be an independent predictor of OS in the multivariate Cox regression analysis (TCGA cohort: HR = 1.397, 95% CI = 1.262–1.546; ICGC cohort: HR = 2.296, 95% CI = 1.570–3.359) (**Figures 6B,D**). The hazard ratio (HR) and 95% confidence interval (CI) of the tumor stage in the multivariate Cox regression analysis of the TCGA cohort were 1.508 and 1.216–1.871 ($p < 0.001$), respectively. We believe that the TNM stage can also be considered an independent predictor.

Construction of a Nomogram

Considering the complexity of the risk signature, we visualized the risk signature by constructing a nomogram based on the risk score and TNM stage (**Figure 7A**). We used calibration curves for the TCGA and ICGC cohorts to verify the consistency of the nomogram in predicting the patients' 1-, 3-, and 5-year OS. The prediction curves for both cohorts were close to the standard curve (**Figures 7B,C**), meaning that the nomogram can predict the patients' OS quite well. Finally, we used the ROC curve to evaluate the sensitivity and specificity of the constructed risk signature for prognosis. The results showed that in the TCGA cohort, the areas under the ROC curve were 0.731, 0.728, and 0.677 at 1, 3, and 5 years (**Figure 7D**). In the ICGC cohort, the areas under the ROC curve were 0.670, 0.672, and 0.640 at the

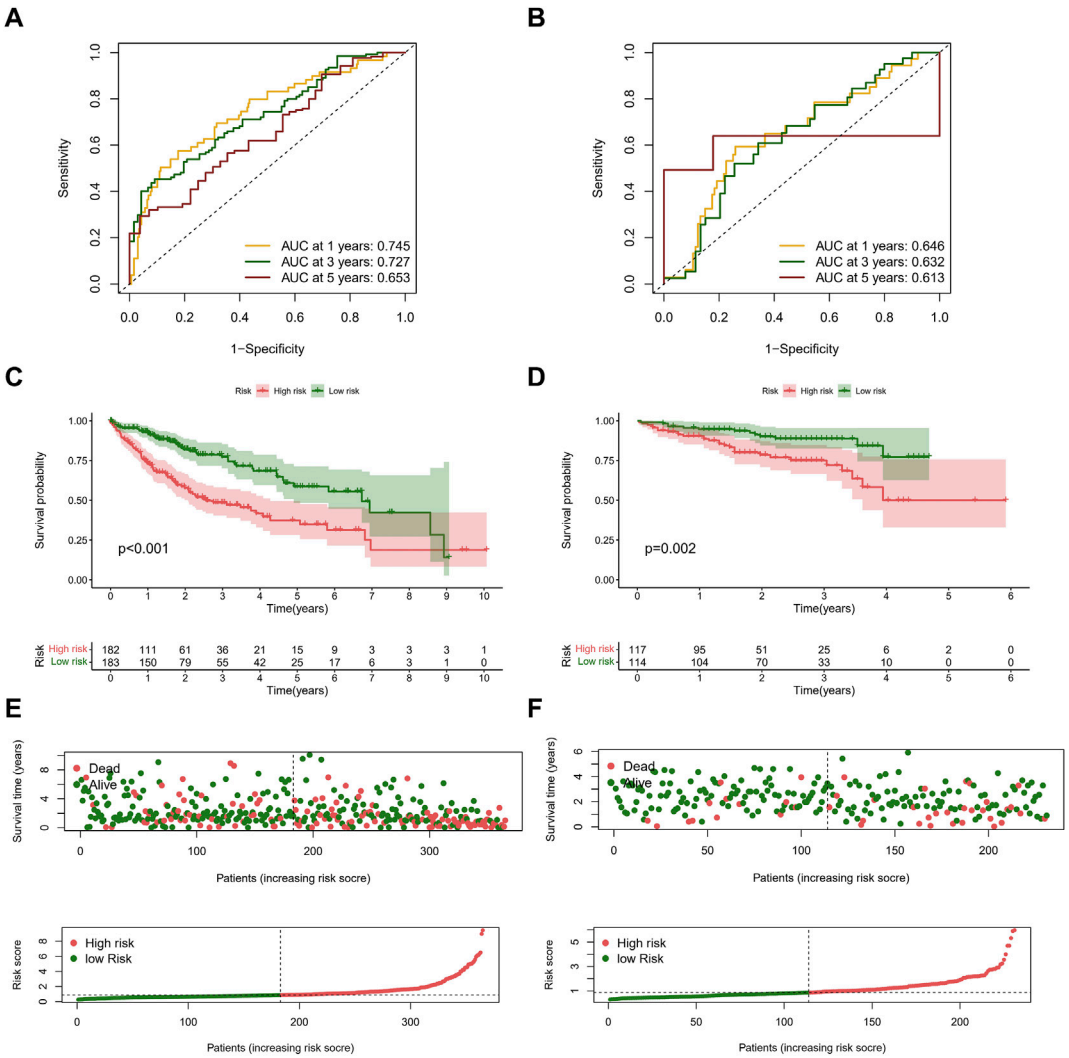


FIGURE 5 | Evaluation of prognostic signature in the TCGA and ICGC cohorts. **(A)** ROC curves of the NRLs signature in the TCGA cohort. The AUCs of 1, 3, and 5 years OS were 0.745, 0.727, and 0.653. **(B)** ROC curves of the NRLs signature in the ICGC cohort. The AUCs of 1, 3, and 5 years OS were 0.646, 0.632, and 0.613. **(C,D)** Kaplan–Meier survival curves of OS (survival probability) of patients between different risk groups in the TCGA **(C)** and ICGC **(D)** cohorts. **(E,F)** Scatter plot (up) and curve plot (down) of risk score in the TCGA **(E)** and ICGC **(F)** cohorts.

same time points (**Figure 7E**). In summary, the risk model showed an excellent predictive potential.

Immune Phenotype Landscape in the Tumor Microenvironment of Hepatocellular Carcinoma

Necroptosis is closely related to the immune signaling of tumor cells. Targeting the necroptotic process has been reported to induce the immune system to kill tumors. RIPK3, which is involved in necroptosis, can drive cells to produce inflammatory chemokines and cytokines during cell death, thereby activating killer T cells (Snyder et al., 2019). To understand the immune cell infiltration of the patients grouped by the predictive model, we used seven algorithms to draw a heat map of immune cell infiltration and found that the high-risk group had a higher immune cell infiltration status (**Figure 8A**). The bubble plot depicting the association of immune cell infiltration with the risk score showed increased immune cell infiltration, including $CD4^+$ memory T cells, mast cells, and B cells at XCELL, $CD4^+$ T cells at TIMER, T cell regulatory at QUANTISEQ, monocytes at MCPCOUNTER, and macrophages M0 at CIBERSORT, in the high-risk group (**Figure 8B**).

We used ssGSEA to quantify the enrichment scores of the immune cell subsets and their associated functions for each sample in the TCGA cohort. The results showed apparent differences in immune cell infiltration among the different risk groups (**Figure 9A**). Antigen-presenting cells such as macrophages were more highly expressed in the high-risk group

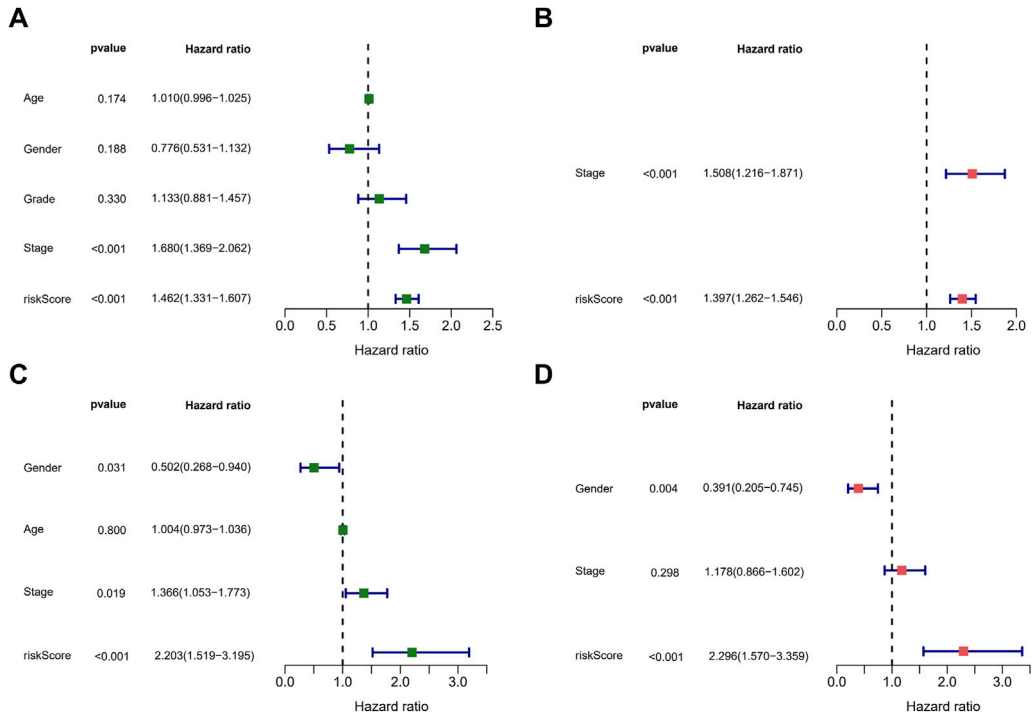


FIGURE 6 | Assessment of the prognostic signature. **(A,C)** Univariate analysis of risk score and clinical characters in the TCGA **(A)** and ICGC **(C)** cohorts. **(B,D)** Multivariate analysis of risk score and clinical characters in the TCGA **(B)** and ICGC **(D)** cohorts.

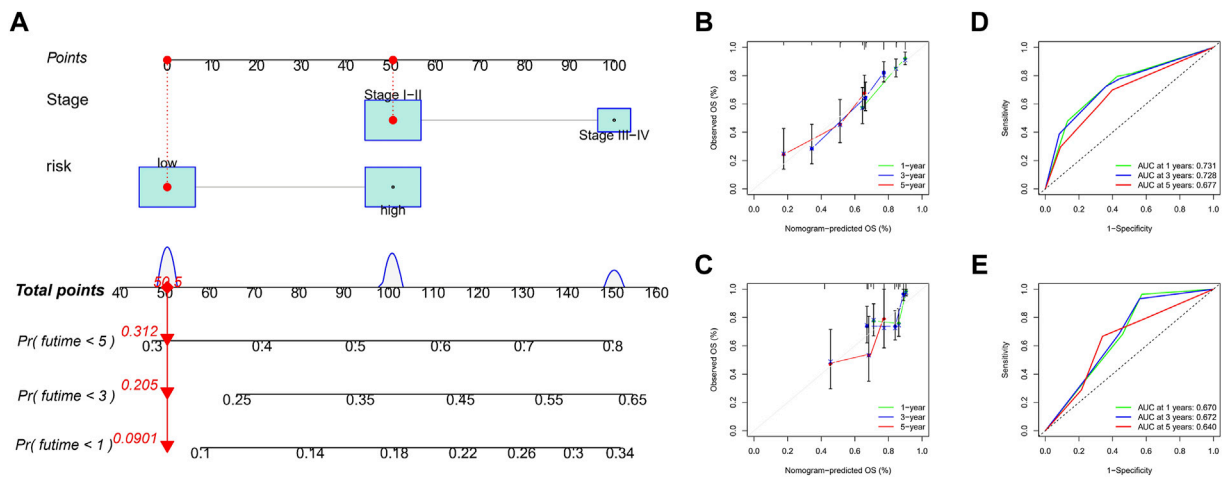


FIGURE 7 | Construction and validation of nomogram. **(A)** The nomogram integrated the risk score and TNM stage to predict the survival rate of the 1, 3, and 5 years. **(B,C)** The 1, 3, and 5 years OS calibration curves for the TCGA **(B)** and ICGC **(C)** cohorts. **(D,E)** The 1, 3, and 5 years ROC curves of the TCGA **(D)** and ICGC **(E)** cohorts.

(**Figure 9C**). ESTIMATE is a tool that uses gene expression data to predict tumor purity and the presence of infiltrating stromal/immune cells in tumor tissue (Yoshihara et al., 2013). We used the ESTIMATE algorithm to evaluate the composition of immune cells in each sample by stromal score, immune score, estimated score, and tumor purity; the results showed that the high-risk group had higher stromal, immune, and estimated scores (**Figure 9B**). We then analyzed the expression of HLA. HLA-C, which belongs to HLA-I and can present endogenous tumor antigens to kill tumor cells effectively, was less expressed in the high-risk group, while HLA-II, such as HLA-DPB2, HLA-DQB2, HLA-DOA, and HLA-DQA2, showed an increase in the high-risk group (**Figure 9D**). HLA-II is mainly expressed on the surface of antigen-presenting cells, and we speculated that the increased expression of HLA-II in high-risk

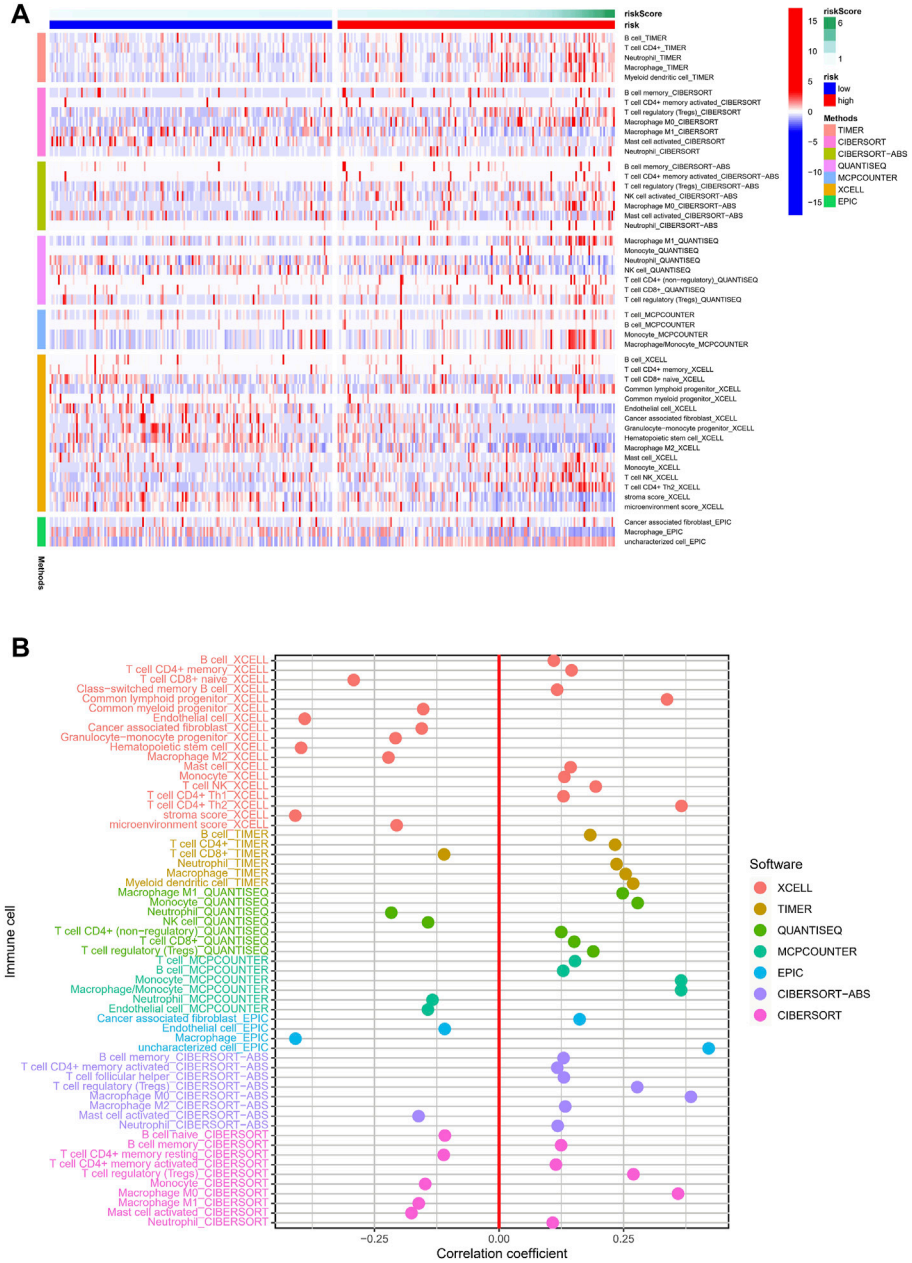


FIGURE 8 | Relationship between immune cells and risk score. **(A)** TIMER, CIBERSORT, CIBERSORT-abs, QUANTISEQ, MCP-counter, XCELL, and EPIC algorithms were used to draw heat maps of immune cell infiltration of patients with different risk scores. **(B)** Correlation coefficient between immune cells and risk score.

patients might be associated with increased immune cell infiltration in a necroptotic environment.

The analysis of HLA reflected the possible differences in the immune status and susceptibility of patients with different risk groups to immune checkpoint blockade (ICB). Given that heterozygosity of HLA-I can reflect the effectiveness of tumor ICB (Chowell et al., 2018), we conducted a differential analysis of immune checkpoints. We found that multiple immune checkpoint proteins, including programmed cell death protein 1 (PDCD1) and CTLA4, were highly expressed in the high-risk group (**Figure 9E**). These findings suggest that patients in the high-risk group may benefit more from ICB therapy.

Functional Enrichment Analysis of Different Risk Groups

To explore the underlying molecular mechanisms of the different risk groups, we performed a differential gene expression analysis of patients in the TCGA cohort and identified 206 genes ($p < 0.05$, $|\log2FC| > 1$) (**Supplementary File S3**). The top 20 differentially

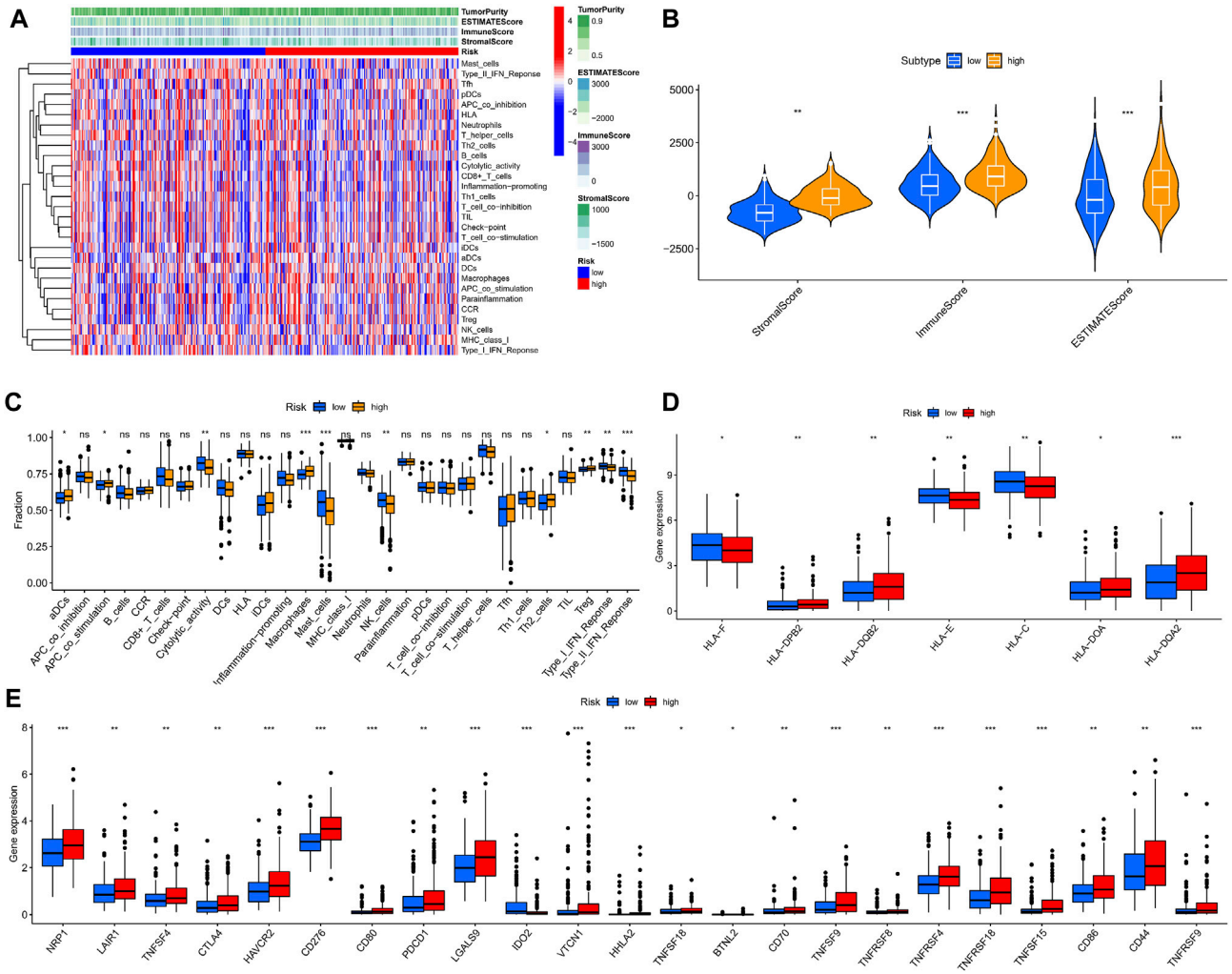


FIGURE 9 | Immune microenvironment analysis in different risk groups. **(A,B)** Stromal, immune, and estimate scores of patients with different risks. **(C)** The ssGSEA scores of immune cells and immune functions. **(D)** Expression of HLAs in different risk groups. **(E)** The comparison of immune checkpoints between high and low-risk groups. (ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

expressed genes are shown in a heat map in **Figure 10A**. The GSEA algorithm was used to detect the main enrichment pathways. Cell cycle and ECM receptor interaction were dominant in the high-risk group (**Figure 10B**), whereas drug metabolism, cytochrome p450, fatty acid metabolism, and peroxisome were critical pathways in the low-risk group (all $p < 0.05$, FDR <0.25, |NES| > 1.5) (**Figure 10C**). Immune- and metabolism-related processes accounted for most of the top 10 results of the GO enrichment analysis, such as xenobiotic metabolic process, fatty acid metabolic process, and steroid metabolic process (**Figure 10D**). As shown in the circle diagram of the KEGG analysis results, the top five enriched pathways mainly involved cell metabolism, drug metabolism, and drug sensitivity, including metabolism of xenobiotics by cytochrome P450, drug metabolism–cytochrome P450, retinol metabolism, bile secretion, and chemical carcinogenesis–DNA adducts (**Figure 10E**).

Drug Effectiveness Analysis

Due to the limitations of systemic chemotherapy, most patients with advanced HCC can choose local therapy based on TACE, which delivers chemotherapy drugs to the vicinity of the tumor (Raoul et al., 2019). The enrichment analysis presented above showed that patients in the different risk groups may differ in drug metabolism and sensitivity. We quantified the IC_{50} values of six drugs commonly used for HCC and found that cisplatin, doxorubicin, etoposide, sorafenib, and vinblastine had lower IC_{50} in the high-risk group (all $p < 0.05$) (**Figure 11**). Cisplatin, doxorubicin, and sorafenib are the first-line drugs recommended for treating HCC in China's standard for diagnosis and treatment of primary liver cancer (2022 edition) (National Health Commission of the people's Republic of China, 2022). The predictive model identified in this study could be a potential predictor of chemosensitivity.

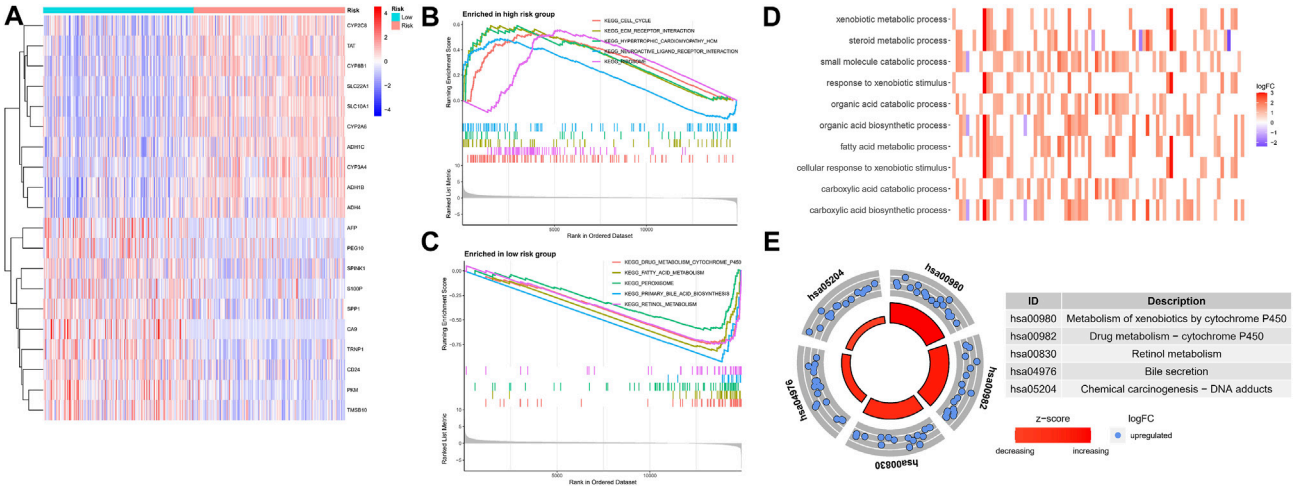


FIGURE 10 | Functional enrichment analysis in different risk groups. **(A)** The heatmap of differentially expressed genes. **(B,C)** GSEA of the top 5 pathways significantly enriched in both high **(B)** and low-risk groups **(C)**. **(D,E)** GO and KEGG enrichment analysis of differentially expressed genes.

Construction of Molecular Subtypes

We divided the TCGA cohort patients into different subtypes based on the NMF algorithm to further explore the role of NRLs in HCC progression. The optimal number of clusters k was established by calculating the cluster correlation coefficient, with k = 3 being the optimal number of clusters (**Figure 12A**). A consistent NMF was performed again to define three clusters, C1 (n = 141), C2 (n = 83), and C3 (n = 141), with an average silhouette width of 0.84 (**Figure 12C**).

Consistent NMF was also performed on the validation set (ICGC cohort), and k = 3 was the optimal number of clusters (**Figure 12B**). We identified three clusters, C1 (n = 82), C2 (n = 67), and C3 (n = 82), with an average silhouette width of 0.86 (**Figure 12D**). The PCA analysis showed obvious distinctions in the different two-dimensional distribution maps of the three clusters (**Figure 12E**). The subtype matching model of the TCGA and ICGC cohorts was identified through the subclass algorithm; we determined the following: TCGA-C1 = ICGC-C2; TCGA-C2 = ICGC-C3; and TCGA-C3 = ICGC-C1. Moreover, the NTP algorithm suggested that the high-risk subgroup in the TCGA cohort (k = 0.774, $p < 0.001$) and in the ICGC cohort (k = 0.679, $p < 0.001$) could better correspond to the cluster C2, with the OS of the C2 being significantly lower than that of C1 and C3 in both cohorts (log-rank test $p < 0.05$, **Figures 12F,G**). The clusters constructed based on the NMF algorithm and prognosis-related NRLs showed better prediction skills concerning survival. We believe that different molecular subtypes can provide another insight into distinguishing patients.

Validation of the Necroptosis-Related Long Noncoding RNAs in Hepatocellular Carcinoma Tissues and Cell Lines

We collected 12 pairs of HCC tissues and paracancerous tissues from the Affiliated Tumor Hospital of Nantong University to verify the expression of the seven NRLs in the signature. We performed RT-PCR after the extraction of total RNA from tissues and found that three out of five high-risk (HR > 1, **Figure 3B**) NRLs (BACE1-AS, SNHG3, SNHG4) were more expressed in HCC tissues than in paracancerous tissues (all $p < 0.01$, **Figures 13D–F**). As a protective factor, the expression of HCG11 in HCC was lower than that in adjacent tissues ($p < 0.05$, **Figure 13G**), which was consistent with the results from the TCGA database. The three remaining lncRNAs (HCG27, C2orf27A, MIR210HG) were not significantly different (**Figures 13H–J**).

Next, we examined the expression of these four NRLs (BACE1-AS, SNHG3, SNHG4, HCG11) in different drug-sensitive HCC cell lines and liver cell lines. Among the five drugs analyzed above (**Figure 11**), cisplatin, doxorubicin, and sorafenib are commonly used for treating HCC. We selected SNU-387, Huh7, Hep3B, and HepG2 cell lines to detect the cell viability at different drug concentrations (**Figures 13A–C**). The IC_{50} values are shown in **Table 2**. The IC_{50} values of the three drugs of SNU-387 were the highest among the four cell lines, indicating that this cell line had apparent resistance to the commonly used drugs. RT-PCR results showed that the expression levels of SNHG3, SNHG4, and BACE1-AS in SNU-387 were higher than those in the other four cell lines (all $p <$ 0.001), and the expression in normal hepatocytes was the lowest (**Figure 13K–M**). The expression of HCG11 as a protective factor was the highest in HepG2 cells, but the expression level in other HCC cell lines, including SNU-387, was not significantly different from that in normal hepatocytes (**Figure 13N**).

DISCUSSION

Despite the variety of treatment options, OS in patients with HCC remains poor. It is important to understand what hinders the

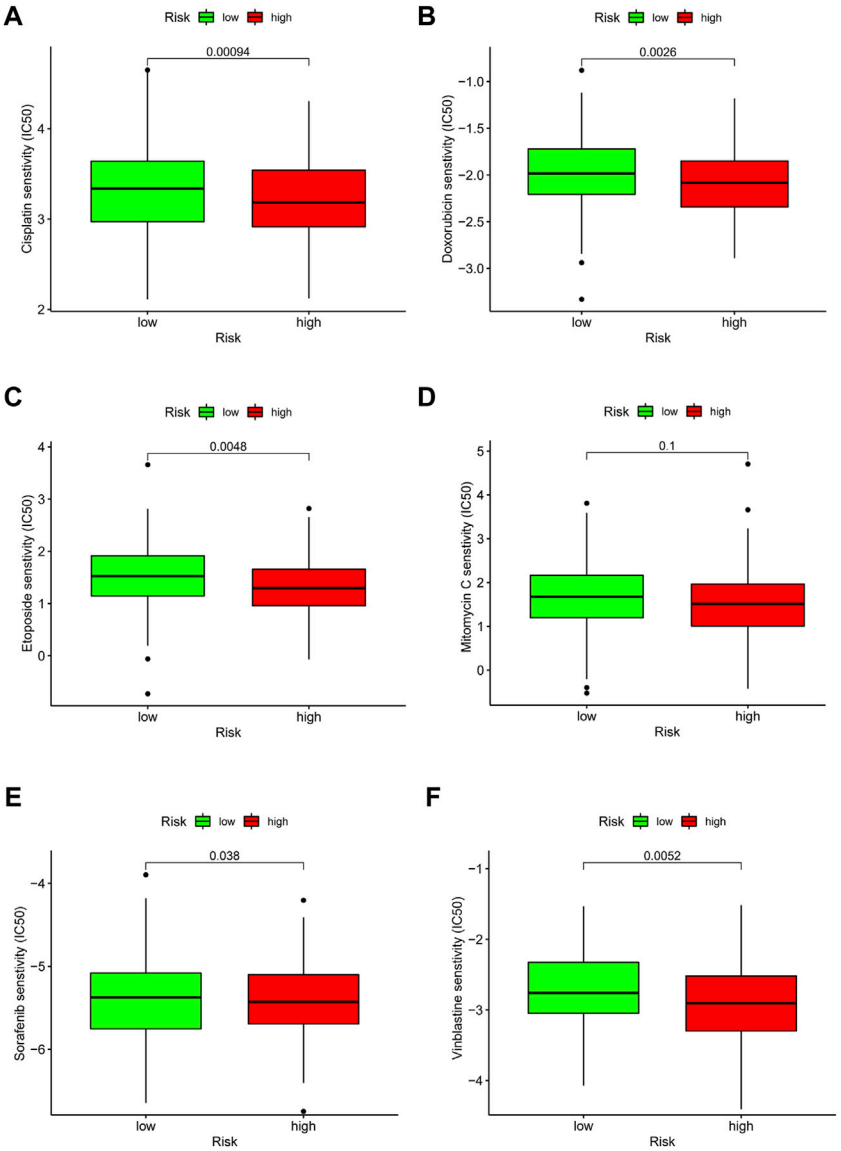


FIGURE 11 | Drug effectiveness of different risk groups. **(A)** Cisplatin. **(B)** Doxorubicin. **(C)** Etoposide. **(D)** Mitomycin C. **(E)** Sorafenib. **(F)** Vinblastine. Five of the six drugs showed IC_{50} differences ($p < 0.05$).

progress of HCC treatment. TME may be an essential factor in the occurrence and development of HCC (Krishnan et al., 1985). Some traditional grading systems, such as TNM grading and Barcelona staging, neither reflect the TME of HCC patients nor accurately predict the patient's prognosis. As an important part of TME, the infiltration by immune cells also affects the benefits of HCC immunotherapy to a certain extent (Yu et al., 2020), with many clinical trials of immunotherapy related to HCC having been conducted globally (Sangro et al., 2013; Kudo et al., 2021). However, the TME may cause an inadequate response and limited therapeutic efficacy when using immunotherapy (El-Khoueiry et al., 2017; Kudo et al., 2021). It is therefore necessary to continue exploring the role of immune factors in the treatment of HCC. It has been reported that necroptosis may alter the TME, thereby affecting the type of liver cancer (Saeed and Jun, 2019). However, its specific role and impact on the prognosis of HCC patients remain unclear. In this study, a prognostic model was constructed based on necroptosis-related lncRNAs, and the patients were grouped into high- and low-risk groups. We systematically investigated differences in immune cell infiltration, immune checkpoints, HLA, and drug sensitivity among the different subgroups and constructed HCC molecular subtypes based on the NMF algorithm. Finally, we used RT-PCR to verify the expression levels of NRLs in tissues and cells.

We analyzed the expression of 67 NRGs in HCC, of which 12 of 19 screened were found to be upregulated and seven were downregulated. Nineteen NRGs and all annotated

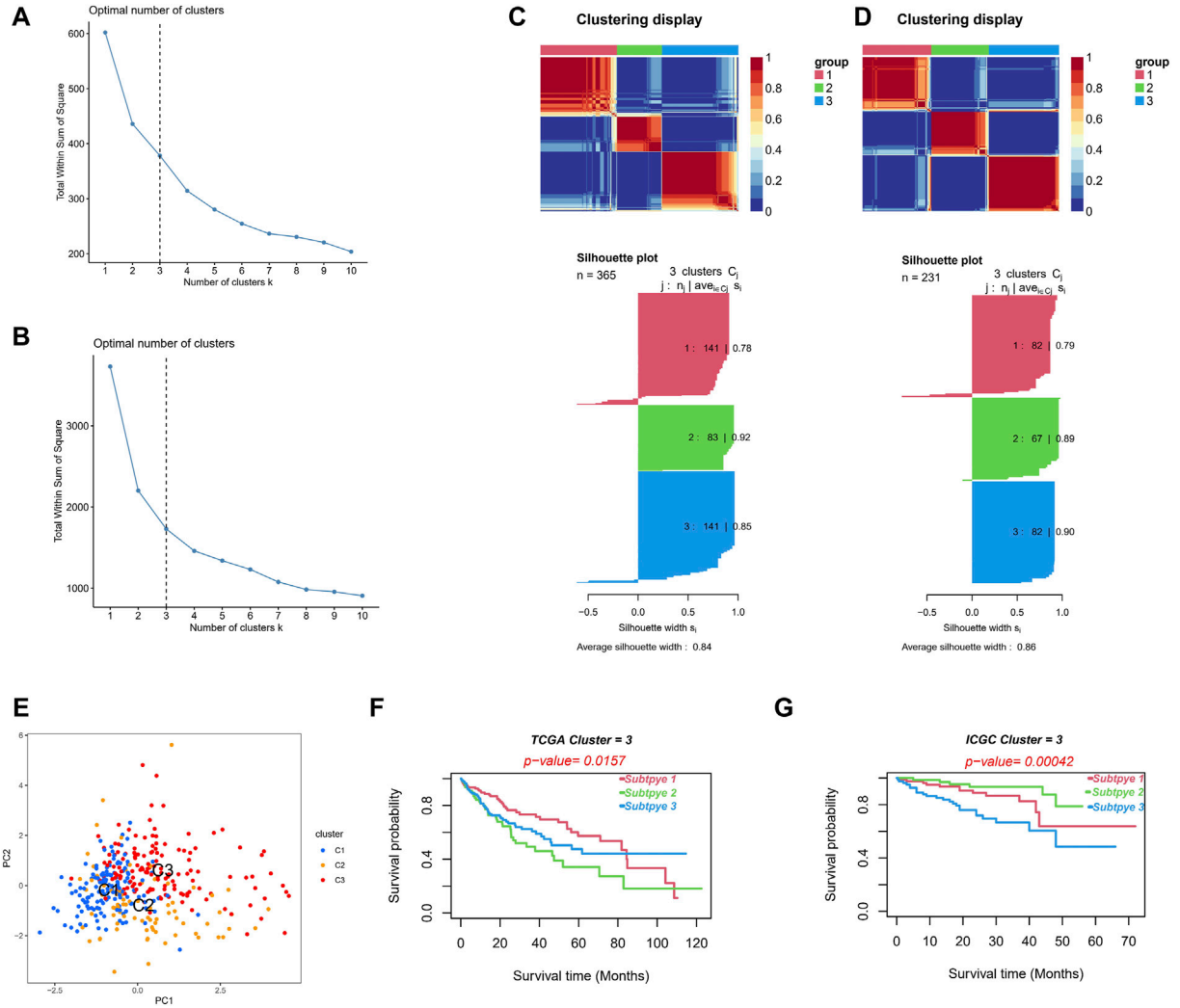


FIGURE 12 | Different molecular subtypes identified by risk signature. **(A,B)** Establish the optimal cluster number k value in the TCGA **(A)** and ICGC **(B)** cohorts. **(C,D)** Patients in the TCGA **(C)** and ICGC **(D)** cohorts were divided into three clusters using the NMF clustering algorithm. **(E)** The PCA of clusters. **(F,G)** Kaplan–Meier survival curves of OS in three clusters.

lncRNAs were subsequently analyzed, and 508 NRLs were identified. We constructed the prognostic signature containing seven NRLs and validated it using an external cohort. The biological function of these NRLs is associated with the progression of HCC. BACE1-AS can promote abnormal proliferation, cell cycle progression, migration, invasion, and apoptosis of HCC through the miR-214-3p/APLN axis (Tian et al., 2021). In HCC cell lines, SNHG3 overexpression promotes the proliferation, migration, and EMT, and inhibits apoptosis (Zhao et al., 2019), while higher levels of SNHG4 are more likely to indicate poor prognosis in liver cancer (Jiao et al., 2020). Through transcriptomic analysis, some studies have suggested that C2orf27A can affect the resistance of HCC cells to sorafenib through immune infiltration (Yuan et al., 2021), which is consistent with our findings. MIR210HG can be used as a glycolysis-related lncRNA to influence the progression of HCC (Xia et al., 2021). We plotted nomograms to predict 1-, 3-, and 5-year OS in HCC patients to intuitively use this predictive model. It was clear from the ROC curves that the predictive model built with NRLs was accurate and reliable.

After measuring the single-cell transcriptomic profiles of HCC biological samples from 19 patients, Wang's team found that the heterogeneity of HCC TME significantly affected the treatment response and prognosis (Ma et al., 2019). Therefore, it is necessary to deeply understand the role of TME in HCC. It has been shown that necroptosis is involved in $CD4^+$ T cell–mediated endothelial cell death (Kwok et al., 2017). Our results showed that the risk score positively correlated with the $CD4^+$ T cell infiltration level, given that the high-risk group had more Th2 and Tregs. We speculate that necroptosis may promote increased $CD4^+$ T cells in the TME. Moreover, the risk score positively correlated with macrophage levels, and the high-risk group had more macrophages, which may have been caused by

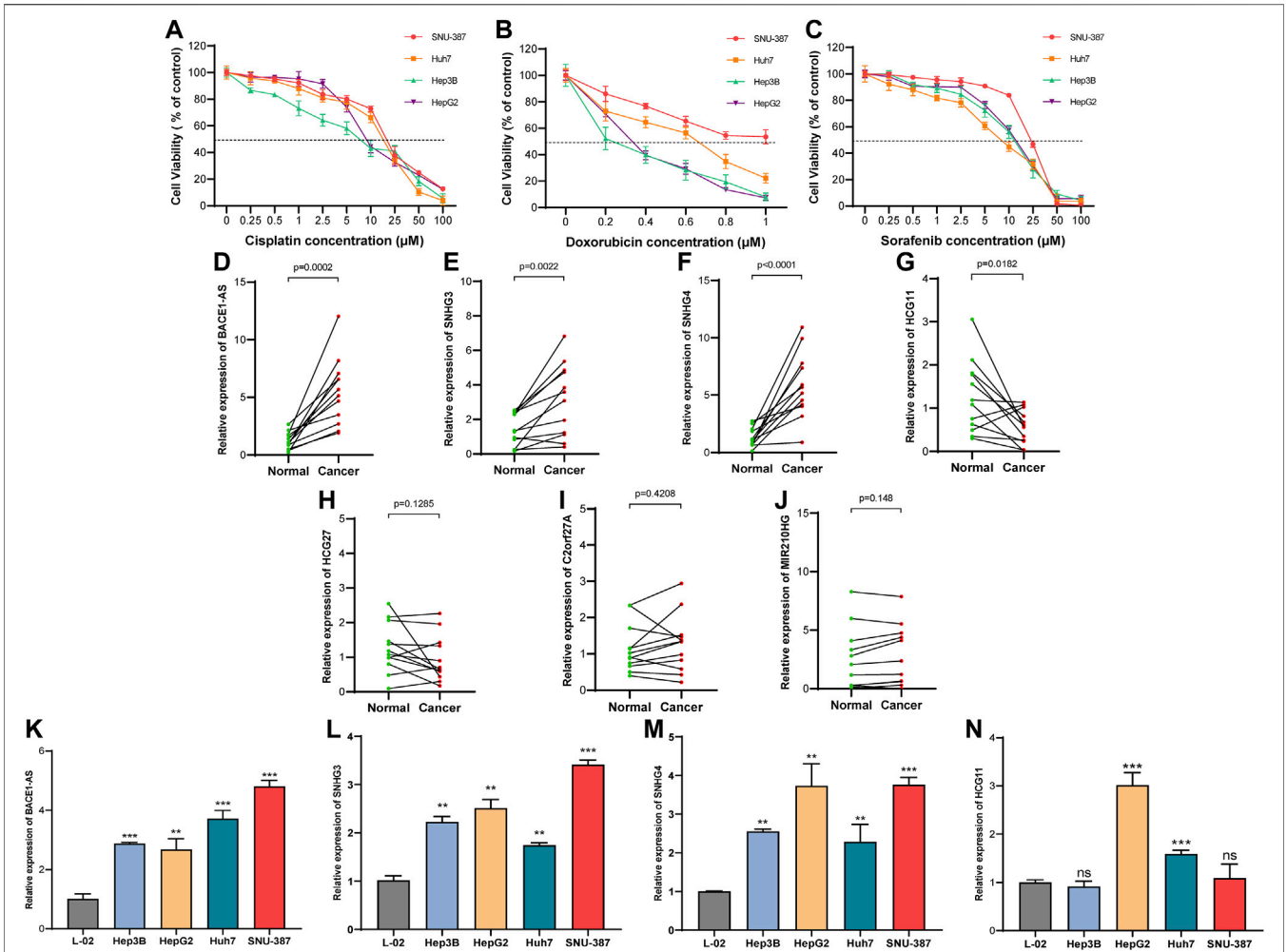


FIGURE 13 | Expression of NRLs in HCC patients and different drug-sensitive cell lines. **(A–C)** Cell viability in the four HCC cell lines after cisplatin, doxorubicin, and sorafenib treatment for 48 h. Data are presented as the mean ± standard deviation ($n = 5$). **(D–J)** Relative expression of seven NRLs in HCC patients. **(K–N)** Expression of BACE1-AS, SNHG3, SNHG4, and HCG11 in cell lines. (ns, $p > 0.05$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; student t-test).

TABLE 2 | IC_{50} values for cisplatin, doxorubicin, and sorafenib treatment in HCC cell lines.

Cell line	IC_{50} values of drugs (μM)[a]		
	Cisplatin	Doxorubicin	Sorafenib
SNU-387	17.047 (15.25–18.844)	1.055 (0.941–1.17)	21.496 (19.884–23.108)
Huh7	13.409 (11.857–14.962)	0.547 (0.488–0.606)	7.571 (6.577–8.566)
Hep3B	6.7 (5.631–7.768)	0.241 (0.197–0.284)	10.878 (9.846–11.909)
HepG2	12.174 (10.89–13.457)	0.326 (0.309–0.343)	11.971 (10.959–12.984)

[a]*IC_{50} values indicate the cisplatin, doxorubicin, and sorafenib concentration [μM; mean (95% CI)].*

increased activation of RIPK3 in the inflammatory environment formed by necrotizing apoptosis (Hao et al., 2021). There were fewer NK cells and mast cells in the high-risk group, and the mechanism is unknown, which needs further exploration. Immune cells and stromal cells are two major non-tumor components of TME that can modulate the sensitivity of immunotherapy by affecting tumor purity. Low purity may be linked to increased immune evasion and poor prognosis (Gong et al., 2020). We found that the high-risk group had higher immune, stromal, and lower tumor purity, which indicates that patients in the high-risk group may benefit more from immunotherapy. HLA-I is plays an essential role in the cytotoxic T-lymphocyte–mediated response, presenting antigens to $CD8^+$ T cells (Durgeau et al., 2018). The ability of

HLA-I class molecules to present antigens is related to the degree of heterozygosity of HLA alleles (Chowell et al., 2019). We found an interesting phenomenon where HLA-DPB2, HLA-DQB2, HLA-DOA, and HLA-DQA2, which belong to HLA-II, showed increased expression in the high-risk group, while HLA-C, which belongs to HLA-I, showed decreased expression. In this study, we only found that HLA-C expression was decreased in high-risk patients; we cannot speculate whether high-risk patients have a decrease in ICB treatment sensitivity, and this topic needs further research. However, most of the immune checkpoint proteins, including PDCD1 and CTLA4, were highly expressed in the high-risk group, which may suggest that the high-risk group may have better ICB treatment effects.

GSEA analysis showed that the high-risk group was mainly enriched in pathways such as cell cycle and ECM receptor interaction. Necroptosis is a specialized form of cell death, which interacts with the cell cycle *via* interferons (Frank et al., 2019). The ECM–receptor interaction pathway regulates the processes of tumor shedding, adhesion, movement, and hyperplasia (Bao et al., 2019). Through specific key mediators (Gong et al., 2019), necroptosis has been identified to promote tumor metastasis and progression (Gong et al., 2019). These findings demonstrate the credibility of GSEA analysis. Significant enrichment in cellular metabolic pathways, including fatty acid metabolism, steroid metabolism, and drug metabolism, was found in the GO and KEGG analyses. It has been reported that ceramides and very long-chain fatty acids accumulate during necroptosis (Parisi et al., 2017), which is consistent with our pathway analysis results. Cisplatin belongs to platinum, which can covalently bind with DNA, inhibit DNA replication, and promote cell cycle arrest. Both cisplatin and doxorubicin upregulate RIPK3, which binds and phosphorylates calmodulin kinase II (CaMKII), thereby regulating the opening of the mitochondrial permeability transition pore (mPTP) and leading to necroptosis (Christidi and Brunham, 2021; Sazonova et al., 2021). Combined with our results, patients in the high-risk group may be more sensitive to cisplatin and doxorubicin. Sorafenib is a protein kinase inhibitor with activity inhibition of many protein kinases, including vascular endothelial growth factor receptor (VEGFR), platelet-derived growth factor receptor (PDGFR), and Raf protein kinase. Heat shock protein 90α (HSP90α) promotes sorafenib resistance in HCC by inhibiting necroptosis under hypoxia (Liao et al., 2021). Finally, we detected NRLs in 12 pairs of tissues. Contrary to our results and TCGA analysis, some studies have reported higher expression of HCG11 in HCC than in adjacent tissues (Xu et al., 2017; Li et al., 2019), which may be caused by differences between different regions and ethnic groups. The expression levels of three high-risk lncRNAs, BACE1-AS, SNHG3, and SNHG4, were significantly higher in SNU-387 than in other cell lines, including normal hepatocytes. It was confirmed from the cellular and tissue levels that the lncRNAs included in the signature may be related to the occurrence, development, and drug resistance of HCC.

This study analyzed the model's predictive performance from various perspectives, including immunotherapy sensitivity, chemotherapy drug sensitivity, and OS. lncRNA was for the first time combined with necroptosis in HCC, with consensus clustering being used for HCC with the assistance of the NMF algorithm. Compared with other necroptosis-related prediction models (Wang and Liu, 2021; Zhao et al., 2021), we used ICGC data for external validation. HCC cells with different drug sensitivity were used for *in vitro* validation. We compared the C-index values of several latest prediction signatures in HCC with ours (Lin et al., 2022; Miao et al., 2022; Ye et al., 2022; Zhao et al., 2022; Zhou et al., 2022) (**Supplementary Figure S2**). Genes in the prediction signature are listed in **Supplementary File S4**. Our signature has the highest C-index of the five signatures included, indicating that it performs better in terms of prediction.

There are still several limitations in our study. The primary datasets were obtained from public databases, and more real-world data are needed to validate the clinical value of the signature. Moreover, we only compared the IC_{50} between the high-risk and low-risk groups on several commonly used drugs due to insufficient data on GDSC. This study has not yet elucidated how lncRNA regulates necroptosis in HCC, which requires further research.

CONCLUSION

We constructed a necroptosis-related prognostic signature that can be used to assess the prognosis and TME status of HCC patients. Combined with preliminary validation at the tissue and cellular levels, the signature could provide an option for individualized patient treatment and prognostic assessment. The potential relationship between necroptosis and lncRNA may be a key to immunotherapy for HCC, but the mechanisms deserve further investigation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Nantong Tumor Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CC and YW contributed equally to this study. CC and YW contributed to the conceptualization; KC and ZX performed the data analyses; CC wrote the manuscript; XL and CZ reviewed and edited the manuscript; HZ and AS contributed to the project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the Scientific Research Project of "333 Project" in Jiangsu Province (BRA2019030) and Nantong Science and Technology Foundation grant (MS22019008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.898507/full#supplementary-material

REFERENCES

Anwanwan, D., Singh, S. K., Singh, S., Saikam, V., and Singh, R. (2020). Challenges in Liver Cancer and Possible Treatment Approaches. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1873 (1), 188314. doi:10.1016/j.bbcan.2019.188314

Bao, Y., Wang, L., Shi, L., Yun, F., Liu, X., Chen, Y., et al. (2019). Transcriptome Profiling Revealed Multiple Genes and ECM-Receptor Interaction Pathways that May Be Associated with Breast Cancer. *Cell Mol. Biol. Lett.* 24, 38. doi:10.1186/s11658-019-0162-0

Chen, X., Cheng, B., Dai, D., Wu, Y., Feng, Z., Tong, C., et al. (2021). Heparanase Induces Necroptosis of Microvascular Endothelial Cells to Promote the Metastasis of Hepatocellular Carcinoma. *Cell Death Discov.* 7 (1), 33. doi:10.1038/s41420-021-00411-5

Chowell, D., Krishna, C., Pierini, F., Makarov, V., Rizvi, N. A., Kuo, F., et al. (2019). Evolutionary Divergence of HLA Class I Genotype Impacts Efficacy of Cancer Immunotherapy. *Nat. Med.* 25 (11), 1715–1720. doi:10.1038/s41591-019-0639-4

Chowell, D., Morris, L. G. T., Grigg, C. M., Weber, J. K., Samstein, R. M., Makarov, V., et al. (2018). Patient HLA Class I Genotype Influences Cancer Response to Checkpoint Blockade Immunotherapy. *Science* 359 (6375), 582–587. doi:10.1126/science.aao4572

Christidi, E., and Brunham, L. R. (2021). Regulated Cell Death Pathways in Doxorubicin-Induced Cardiotoxicity. *Cell Death Dis.* 12 (4), 339. doi:10.1038/s41419-021-03614-x

Degterev, A., Hitomi, J., Germscheid, M., Ch'en, I. L., Korkina, O., Teng, X., et al. (2008). Identification of RIP1 Kinase as a Specific Cellular Target of Necrostatins. *Nat. Chem. Biol.* 4 (5), 313–321. doi:10.1038/nchembio.83

Durgeau, A., Virk, Y., Corgnac, S., and Mami-Chouaib, F. (2018). Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front. Immunol.* 9, 14. doi:10.3389/fimmu.2018.00014

El-Khoueiry, A. B., Sangro, B., Yau, T., Crocenzi, T. S., Kudo, M., Hsu, C., et al. (2017). Nivolumab in Patients with Advanced Hepatocellular Carcinoma (CheckMate 040): an Open-Label, Non-comparative, Phase 1/2 Dose Escalation and Expansion Trial. *Lancet* 389 (10088), 2492–2502. doi:10.1016/s0140-6736(17)31046-2

Frank, T., Tuppi, M., Hugle, M., Dötsch, V., van Wijk, S. J. L., and Fulda, S. (2019). Cell Cycle Arrest in Mitosis Promotes Interferon-Induced Necroptosis. *Cell Death Differ.* 26 (10), 2046–2060. doi:10.1038/s41418-019-0298-5

Fu, Y., Liu, S., Zeng, S., and Shen, H. (2019). From Bench to Bed: the Tumor Immune Microenvironment and Current Immunotherapeutic Strategies for Hepatocellular Carcinoma. *J. Exp. Clin. Cancer Res.* 38 (1), 396. doi:10.1186/s13046-019-1396-4

Geeleher, P., Cox, N. J., and Huang, R. (2014). Clinical Drug Response Can Be Predicted Using Baseline Gene Expression Levels and *In Vitro* Drug Sensitivity in Cell Lines. *Genome Biol.* 15 (3), R47. doi:10.1186/gb-2014-15-3-r47

Gong, Y., Fan, Z., Luo, G., Yang, C., Huang, Q., Fan, K., et al. (2019). The Role of Necroptosis in Cancer Biology and Therapy. *Mol. Cancer* 18 (1), 100. doi:10.1186/s12943-019-1029-8

Gong, Z., Zhang, J., and Guo, W. (2020). Tumor Purity as a Prognosis and Immunotherapy Relevant Feature in Gastric Cancer. *Cancer Med.* 9 (23), 9052–9063. doi:10.1002/cam4.3505

Green, D. R. (2019). The Coming Decade of Cell Death Research: Five Riddles. *Cell* 177 (5), 1094–1107. doi:10.1016/j.cell.2019.04.024

Hao, Q., Idell, S., and Tang, H. (2021). M1 Macrophages Are More Susceptible to Necroptosis. *J. Cell Immunol.* 3 (2), 97–102. doi:10.33696/immunology.3.084

Harari-Steinfeld, R., Gefen, M., Simerzin, A., Zorde-Khvalevsky, E., Rivkin, M., Ella, E., et al. (2021). The lncRNA H19-Derived MicroRNA-675 Promotes Liver Necroptosis by Targeting FADD. *Cancers (Basel)* 13 (3). doi:10.3390/cancers13030411

Jiao, Y., Li, Y., Jia, B., Chen, Q., Pan, G., Hua, F., et al. (2020). The Prognostic Value of lncRNA SNHG4 and its Potential Mechanism in Liver Cancer. *Biosci. Rep.* 40 (1). doi:10.1042/BSR20190729

Krishnan, K. R. R., France, R. D., and Houpt, J. L. (1985). Chronic Low Back Pain and Depression. *Psychosomatics* 26 (4), 299–302. doi:10.1016/s0033-3182(85)72861-7

Kudo, M., Lim, H. Y., Cheng, A.-L., Chao, Y., Yau, T., Ogasawara, S., et al. (2021). Pembrolizumab as Second-Line Therapy for Advanced Hepatocellular Carcinoma: A Subgroup Analysis of Asian Patients in the Phase 3 KEYNOTE-240 Trial. *Liver Cancer* 10 (3), 275–284. doi:10.1159/000515553

Kwok, C., Pavlosky, A., Lian, D., Jiang, J., Huang, X., Yin, Z., et al. (2017). Necroptosis Is Involved in CD4+ T Cell-Mediated Microvascular Endothelial Cell Death and Chronic Cardiac Allograft Rejection. *Transplantation* 101 (9), 2026–2037. doi:10.1097/tp.0000000000001578

Li, M. L., Zhang, Y., and Ma, L. T. (2019). LncRNA HCG11 Accelerates the Progression of Hepatocellular Carcinoma via miR-26a-5p/ATG12 axis. *Eur. Rev. Med. Pharmacol. Sci.* 23 (24), 10708–10720. doi:10.26355/eurrev_201912_19771

Liao, Y., Yang, Y., Pan, D., Ding, Y., Zhang, H., Ye, Y., et al. (2021). HSP90alpha Mediates Sorafenib Resistance in Human Hepatocellular Carcinoma by Necroptosis Inhibition under Hypoxia. *Cancers (Basel)* 13 (2). doi:10.3390/cancers13020243

Lin, Z., Xu, Q., Song, X., Zeng, Y., Zeng, L., Zhao, L., et al. (2022). Comprehensive Analysis Identified Mutation-Gene Signature Impacts the Prognosis through Immune Function in Hepatocellular Carcinoma. *Front. Oncol.* 12, 748557. doi:10.3389/fonc.2022.748557

Ma, L., Hernandez, M. O., Zhao, Y., Mehta, M., Tran, B., Kelly, M., et al. (2019). Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* 36 (4), 418–430. doi:10.1016/j.ccell.2019.08.007

Miao, D., Xu, Q., Zeng, Y., Zhao, R., Song, X., Chen, Z., et al. (2022). Establishment and Validation of a Peroxisome-Related Gene Signature for Prognostic Prediction and Immune Distinction in Hepatocellular Carcinoma. *J. Cancer* 13 (5), 1418–1435. doi:10.7150/jca.65080

Mizukoshi, E., and Kaneko, S. (2019). Immune Cell Therapy for Hepatocellular Carcinoma. *J. Hematol. Oncol.* 12 (1), 52. doi:10.1186/s13045-019-0742-5

National Health Commission of the people's Republic of China (2022). Standard for Diagnosis and Treatment of Primary Liver Cancer (2022 Edition). *J. Clin. Hepatology* 38 (2), 288–303.

Parisi, L. R., Li, N., and Atilla-Gokcumen, G. E. (2017). Very Long Chain Fatty Acids Are Functionally Involved in Necroptosis. *Cell Chem. Biol.* 24 (12), 1445–1454. doi:10.1016/j.chembiol.2017.08.026

Raoul, J.-L., Forner, A., Bolondi, L., Cheung, T. T., Kloeckner, R., and de Baere, T. (2019). Updated Use of TACE for Hepatocellular Carcinoma Treatment: How and when to Use it Based on Clinical Evidence. *Cancer Treat. Rev.* 72, 28–36. doi:10.1016/j.ctrv.2018.11.002

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Saeed, W. K., and Jun, D. W. (2019). Viewpoint: Necroptosis Influences the Type of Liver Cancer via Changes of Hepatic Microenvironment. *Hepatobiliary Surg. Nutr.* 8 (5), 549–551. doi:10.21037/hbsn.2019.04.15

Sangro, B., Gomez-Martin, C., de la Mata, M., Iñarrairaegui, M., Garralda, E., Barrera, P., et al. (2013). A Clinical Trial of CTLA-4 Blockade with Tremelimumab in Patients with Hepatocellular Carcinoma and Chronic Hepatitis C. *J. Hepatology* 59 (1), 81–88. doi:10.1016/j.jhep.2013.02.022

Sazonova, E. V., Kopeina, G. S., Imyanitov, E. N., and Zhivotovsky, B. (2021). Platinum Drugs and Taxanes: Can We Overcome Resistance? *Cell Death Discov.* 7 (1), 155. doi:10.1038/s41420-021-00554-5

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* 39 (5), 1–13. doi:10.18637/jss.v039.i05

Snyder, A. G., Hubbard, N. W., Messmer, M. N., Kofman, S. B., Hagan, C. E., Orozco, S. L., et al. (2019). Intratumoral Activation of the Necroptotic Pathway Components RIPK1 and RIPK3 Potentiates Antitumor Immunity. *Sci. Immunol.* 4 (36). doi:10.1126/sciimmunol.aaw2004

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tang, D., Kang, R., Berghe, T. V., Vandenabeele, P., and Kroemer, G. (2019). The Molecular Machinery of Regulated Cell Death. *Cell Res.* 29 (5), 347–364. doi:10.1038/s41422-019-0164-5

Tao, H., Liu, X., Liu, X., Liu, W., Wu, D., Wang, R., et al. (2019). LncRNA MEG3 Inhibits Trophoblast Invasion and Trophoblast-mediated VSMC Loss in Uterine Spiral Artery Remodeling. *Mol. Reprod. Dev.* 86 (6), 686–695. doi:10.1002/mrd.23147

Tian, Q., Yan, X., Yang, L., Liu, Z., Yuan, Z., and Zhang, Y. (2021). Long Noncoding RNA BACE1-AS Plays an Oncogenic Role in Hepatocellular Carcinoma Cells through miR-214-3p/APLN axis. *Acta Biochim. Biophys. Sin. (Shanghai).* 53 (11), 1538–1546. doi:10.1093/abbs/gmab134

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385:aid-sim380>3.0.co;2-3

Wang, C., Ke, S., Li, M., Lin, C., Liu, X., and Pan, Q. (2020). Downregulation of LncRNA GAS5 Promotes Liver Cancer Proliferation and Drug Resistance by Decreasing PTEN Expression. *Mol. Genet. Genomics* 295 (1), 251–260. doi:10.1007/s00438-019-01620-5

Wang, N., and Liu, D. (2021). Identification and Validation a Necroptosis-related Prognostic Signature and Associated Regulatory Axis in Stomach Adenocarcinoma. *Ott* Vol. 14, 5373–5383. doi:10.2147/ott.s342613

Werthmöller, N., Frey, B., Wunderlich, R., Fietkau, R., and Gaipl, U. S. (2015). Modulation of Radiochemoimmunotherapy-Induced B16 Melanoma Cell Death by the Pan-Caspase Inhibitor zVAD-Fmk Induces Anti-tumor Immunity in a HMGB1-, Nucleotide- and T-cell-dependent Manner. *Cell Death Dis.* 6, e1761. doi:10.1038/cddis.2015.129

Xia, X., Zhang, H., Xia, P., Zhu, Y., Liu, J., Xu, K., et al. (2021). Identification of Glycolysis-Related lncRNAs and the Novel lncRNA WAC-AS1 Promotes Glycolysis and Tumor Progression in Hepatocellular Carcinoma. *Front. Oncol.* 11, 733595. doi:10.3389/fonc.2021.733595

Xu, Y., Zheng, Y., Liu, H., and Li, T. (2017). Modulation of IGF2BP1 by Long Noncoding RNA HCG11 Suppresses Apoptosis of Hepatocellular Carcinoma Cells via MAPK Signaling Transduction. *Int. J. Oncol.* 51 (3), 791–800. doi:10.3892/ijo.2017.4066

Ye, W., Shi, Z., Zhou, Y., Zhang, Z., Zhou, Y., Chen, B., et al. (2022). Autophagy-Related Signatures as Prognostic Indicators for Hepatocellular Carcinoma. *Front. Oncol.* 12, 654449. doi:10.3389/fonc.2022.654449

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yu, S., Wang, Y., Hou, J., Li, W., Wang, X., Xiang, L., et al. (2020). Tumor-infiltrating Immune Cells in Hepatocellular Carcinoma: Tregs Is Correlated with Poor Overall Survival. *PLoS One* 15 (4), e0231003. doi:10.1371/journal.pone.0231003

Yuan, W., Tao, R., Huang, D., Yan, W., Shen, G., and Ning, Q. (2021). Transcriptomic Characterization Reveals Prognostic Molecular Signatures of Sorafenib Resistance in Hepatocellular Carcinoma. *Aging* 13 (3), 3969–3993. doi:10.18632/aging.202365

Zhang, P. F., Wang, F., Wu, J., Wu, Y., Huang, W., Liu, D., et al. (2019). LncRNA SNHG3 Induces EMT and Sorafenib Resistance by Modulating the miR-128/CD151 Pathway in Hepatocellular Carcinoma. *J. Cell. Physiology* 234 (3), 2788–2794. doi:10.1002/jcp.27095

Zhang, Y., Li, Z., Chen, M., Chen, H., Zhong, Q., Liang, L., et al. (2020). lncRNA TCL6 Correlates with Immune Cell Infiltration and Indicates Worse Survival in Breast Cancer. *Breast Cancer* 27 (4), 573–585. doi:10.1007/s12282-020-01048-5

Zhang, Y., Liu, Q., and Liao, Q. (2020). Long Noncoding RNA: a Dazzling Dancer in Tumor Immune Microenvironment. *J. Exp. Clin. Cancer Res.* 39 (1), 231. doi:10.1186/s13046-020-01727-3

Zhao, Q., Wu, C., Wang, J., Li, X., Fan, Y., Gao, S., et al. (2019). LncRNA SNHG3 Promotes Hepatocellular Tumorigenesis by Targeting miR-326. *Tohoku J. Exp. Med.* 249 (1), 43–56. doi:10.1620/tjem.249.43

Zhao, Y., Song, Q., Xu, F., Zhou, Y., Zuo, X., and Zhang, Z. (2022). Pyroptosis-Related Risk Signature Exhibits Distinct Prognostic, Immune, and Therapeutic Landscapes in Hepatocellular Carcinoma. *Front. Genet.* 13, 823443. doi:10.3389/fgene.2022.823443

Zhao, Z., Liu, H., Zhou, X., Fang, D., Ou, X., Ye, J., et al. (2021). Necroptosis-Related lncRNAs: Predicting Prognosis and the Distinction between the Cold and Hot Tumors in Gastric Cancer. *J. Oncol.* 2021, 6718443. doi:10.1155/2021/6718443

Zhou, Z., Wang, T., Du, Y., Deng, J., Gao, G., and Zhang, J. (2022). Identification of a Novel Glycosyltransferase Prognostic Signature in Hepatocellular Carcinoma Based on LASSO Algorithm. *Front. Genet.* 13, 823728. doi:10.3389/fgene.2022.823728

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Wu, Chen, Xia, Liu, Zhang, Zhao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

SYSTEMATIC REVIEW
published: 08 June 2022
doi: 10.3389/fgene.2022.878700

Circular RNAs as Prognostic Biomarkers in Renal Cell Carcinoma: A Systematic Review and Meta-Analysis

Dan Liao[1]†, Qiu Lin[1], Huan Xiao[2], Fenglian Zhang[1] and Qin Han[3]†*

[1]Department of Nephrology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China, [2]Department of Radiology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China, [3]School of Public Health, Chengdu Medical College, Chengdu, China

Background: Recently, several studies have shown that circRNAs play critical roles in renal cell carcinoma (RCC) oncogenesis and development. However, whether the level of circRNA expression in RCC is correlated with prognosis remains unclear. Hence, we conducted a meta-analysis to explore the association between circRNA expression levels and the prognosis of RCC patients.

Methods: We systematically searched Ovid, Embase, PubMed, and Web of Science from January 1950 to June 2021 for the literature published in English. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, we conducted a meta-analysis of 21 selected studies to confirm the association between the circRNA expression level and prognosis of RCC.

Results: This meta-analysis included 20 articles and 1,559 RCC patients. The results showed that the high expression of oncogenic circRNAs (OS: HR = 2.04, 95% CI: 1.63–2.56, p = 0.20; PFS: HR = 2.82, 95% CI: 0.82–9.72, p = 0.03) and low expression of tumor-suppressor circRNAs (OS: HR: 1.92, 95% CI: 1.61–2.30, p < 0.05; PFS: HR: 2.40, 95% CI: 1.76–3.28, p = 0.36) were closely related to poor survival outcomes.

Conclusion: The meta-analysis verifies that circRNAs can be potential prognostic biomarkers of RCC.

Keywords: circular RNAs, renal cell carcinoma, prognostic biomarkers, systematic review, meta-analysis

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Rui'Hao Zhou,
Sichuan University, China
Xiaofei Chen,
Sanofi U.S., United States

***Correspondence:**
Dan Liao
liaodanyu@163.com

†*These authors have contributed equally to this work and share senior authorship*

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *18 February 2022*
Accepted: *25 April 2022*
Published: *08 June 2022*

Citation:
Liao D, Lin Q, Xiao H, Zhang F and Han Q (2022) Circular RNAs as Prognostic Biomarkers in Renal Cell Carcinoma: A Systematic Review and Meta-Analysis. Front. Genet. 13:878700. doi: 10.3389/fgene.2022.878700

INTRODUCTION

Worldwide, kidney cancer has become a serious and widespread prevalent problem and is the 16th most frequently diagnosed cancer with the 17th highest mortality, accounting for 2.2% of all oncological diagnoses and 1.8% of all oncological deaths (Capitanio et al., 2019; Sung et al., 2021). Renal cell carcinoma (RCC) is a common cancer that originates in the renal epithelium and accounts for 90% of kidney cancers (Hsieh et al., 2017). Although the diagnosis and treatment (immunotherapy (Xu et al., 2020), targeted agents (Yang and Chen, 2020), and combination

Abbreviations: circRNAs, circular RNAs; DFS, disease-free survival; HRs, hazard ratios; NOS, Newcastle–Ottawa Scale; OS, overall survival; PFS, progression-free survival; RCC, renal cell carcinoma; RFS, relapse-free survival.

therapy (Cerbone et al., 2020)) of RCC have improved in the last 20 years, the overall survival of patients with RCC is still less than satisfactory (Haddad and Margulis, 2015; Xu et al., 2020). To guide clinical decision-making, a great prognostic evaluation of RCC is urgently necessary for both physicians and patients in treatment management. Currently, aside from imaging examination, reliable biomarkers that can be applied in clinical practice are lacking. Overall, more sensitive prognostic biomarkers and more effective therapeutic strategies for cancer need to be found.

CircRNAs are characterized by covalently closed-loop structures with neither 5′–3′ polarity nor polyadenylated tails (Chen and Yang, 2015). CircRNAs were identified as critical molecules in transcriptional regulation, splicing alternatives, interactions with RNA-binding proteins, and microRNA sponges in cellular physiology and disease pathogenesis (Han et al., 2018). CircRNAs have been confirmed as regulators and biomarkers for numerous types of cancers, which can act as either oncogenic or tumor suppressors in cancer and have also been shown to be enriched and stable in extracellular fluid (Chen and Huang, 2018). These findings indicate the potential of circRNAs to be effective biomarkers. Furthermore, as previously reported, circRNAs have a critical role in promoting metastasis in RCC (Yang et al., 2021). CircRNAs can contribute to tumorigenesis in RCC and promote proliferation and differentiation of RCC by regulating tumor-related signaling pathways (Li et al., 2020a) and activating transcription factors (Chen et al., 2020a). Thus, we reasonably predict that circRNAs may be potential effective therapeutic targets. In addition, many studies have identified that different expression levels of circRNAs are associated with survival in RCC patients. We therefore performed a meta-analysis to evaluate the prognostic value of circRNAs in RCC.

MATERIALS AND METHODS

Literature Search

We searched the English medical literature in PubMed, Ovid, Embase, and Web of Science to identify all publications on circRNAs as prognostic biomarkers in human renal cell carcinoma. The database surveys were conducted on June 4, 2021. The following keywords were used in the database search: ("Circular RNA" and "Renal Cell Cancer") (the detailed search terms are listed in Supplementary Table S1). We eliminated all irrelevant literature works by scanning the article title and abstract. We excluded all duplicated publications by using EndNote X9. The selected studies were identified after they were read in full by our reviewers.

Publication Inclusion and Exclusion Criteria

The two investigators (DL and QH) independently used the same multistep process to evaluate whether these studies were suitable for our meta-analysis. A third investigator (QL) resolved any disagreements.

The inclusion criteria for meta-analysis followed the population, intervention, control, and outcome (PICO) criteria: *1*) patients with a pathological diagnosis of RCC, *2*) the expression of circRNAs in the tissue specimens of patients was measured, and *3*) the included studies provided time-to-event data, including overall survival (OS), progression-free survival (PFS), recurrence-free survival (RFS), disease-free survival (DFS), and association with circRNA expression. Considering the similar survival outcomes, RFS and DFS were combined as PFS (Zhou et al., 2015).

Furthermore, the excluded articles were eliminated based on the following criteria: *1*) the publications that were not published in English; *2*) letters, reviews, expert opinions, case reports, conference articles, clinical guidelines, and meeting records; *3*) studies of patient sample size <30; *4*) duplicated studies; *5*) studies of cell lines or animals; and *6*) the survival data shown in the article were not sufficient to calculate the HR value.

Data Extraction

The two investigators (DL and HX) independently extracted relevant data for meta-analysis, and a third investigator (FZ) resolved any disagreements. Finally, we extracted the following items: the first author's name, publication year, publication journal, circRNA type, total number of patients, sex, country, and follow-up period. The prognostic endpoints included hazard ratios (HRs) and 95% confidence intervals (CIs), OS, DFS, RFS, and PFS. With only the survival curve provided, Engauge Digitizer version 12.1 (available at http://sourceforge.net/) was used to extract related data from the survival curve. According to the extracted data and method of Spotswood et al. (Spruance et al., 2004), the HR was calculated.

Quality Assessment

Two researchers (DL and QH) independently assessed the quality of all selected studies according to the Newcastle–Ottawa Scale (NOS) method (Stang, 2010). The two investigators identified all differences through discussion and consensus. Studies with NOS scores ranging from zero to normal and NOS scores ≥ six were considered high quality.

Statistical Analysis

Meta-analysis was performed using a *meta* package (V.4.18–2). To determine heterogeneity between several studies, we used the I^2 test and the chi-based Q-test to assess statistical heterogeneity. If I^2 was equal to or <50%, the heterogeneity between studies was not obvious, so we used a fixed-effect model and instead applied the random-effect model (Borenstein et al., 2010). Finally, publication bias was evaluated with a funnel plot (a two-sided $p < 0.05$ was considered to be statistically significant). All statistical analyses were carried out in R V.3.6.1 (R Foundation for Statistical Computing). $p \leq 0.05$ was considered to be statistically significant.

RESULTS

Research Results

Detailed information on the literature search is shown in **Figure 1**. A total of 77 publications in English were initially retrieved from the database. The most recent publication date was

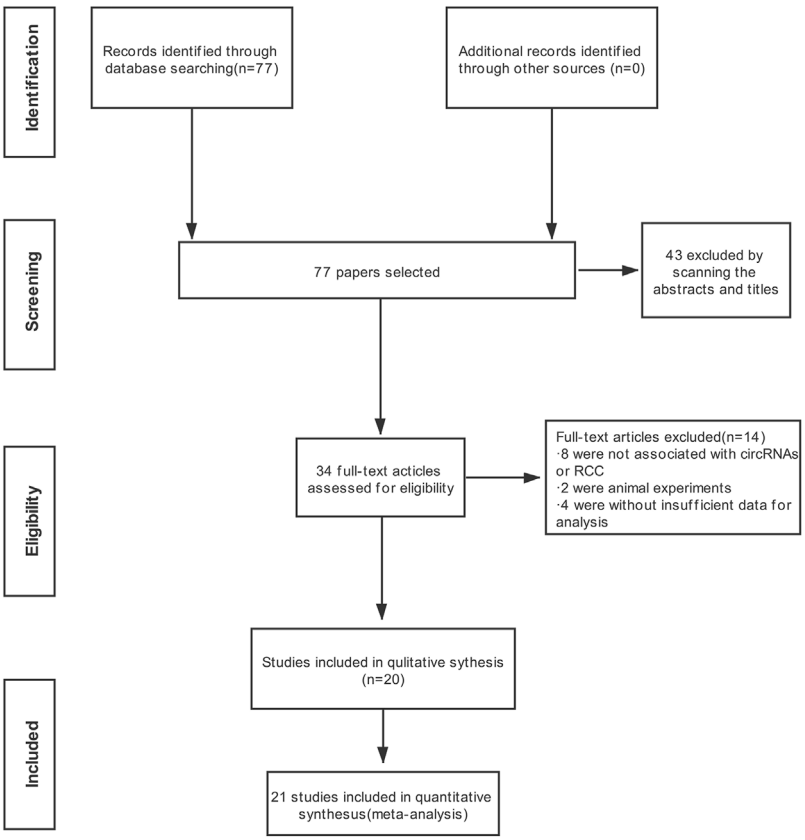


FIGURE 1 | Flow chart demonstrating the study selection process.

June 2021. In total, thirty-three articles were directly excluded after examining the abstract and title. For the remaining 34 publications, after careful reading, 14 studies were eliminated for the following reasons: eight were not associated with circRNAs or RCC, two were animal experiments, and four had insufficient data for analysis. A total of 21 studies were chosen for the meta-analysis based upon the inclusion and exclusion criteria.

Study Characteristics and Quality Assessment

Of the 20 publications identified, 1,559 RCC patients were included, and the average number of patients was 77.95 (range 90.5–193.5). The main characteristics of the 10 studies are summarized in **Table 1**. One study was from Germany, two studies were from The Cancer Genome Atlas (TCGA) database, and other studies were from China. Together, 20 types of circRNAs were included in the meta-analysis (circHIAT1 (Wang et al., 2017; Wang et al., 2019), circNUP98 (Yu et al., 2020), circMYLK (Li et al., 2020a), hsa_circ_001895 (Chen et al., 2020a), circ-EGLN3 (Lin and Cai, 2020), circ_001842 (Zeng et al., 2020), circ_0001368 (Chen et al., 2020b), cRAPGEF5 (Chen et al., 2020c), circ-ABCB10 (Huang et al., 2019), circCSNK1G3 (Li et al., 2021), circAKT1 (Zhu et al., 2020), circAGAP1 (Lv et al., 2021), circ_101341 (Yue et al., 2020), circEHD2 (Frey et al., 2021), circNETO2 (Frey et al., 2021), ciRS-7 (Zhao et al., 2020), hsa-hsa_circ_0085576 (Liu et al., 2020a), circTLK1 (Li et al., 2020b), circPTCH1 (Liu et al., 2020b), and circHIPK3 (Han et al., 2020)). Four circRNAs (circHIAT1, circ_0001368, cRAPGEF5, and circEHD2) were downregulated, and 16 circRNAs (circMYLK, hsa_circ_001895, circ-ABCB10, circAGAP1, circEHD2, hsa_circ_0085576, circTLK1, circPTCH1, circHIPK3, circNUP98, circ-EGLN3, circ_001842, circCSNK1G3, circAKT1, and circ_101341) were upregulated.

The NOS for quality evaluation of the included studies varied from 6 to 8, indicating that all included studies were available and were high-quality studies. Among the 20 identified articles, NOS assessment included 4 documents with an NOS score of 6, 6 articles with an NOS score of 7, and 10 articles with an NOS score of 8. The value of the k-statistic was 0.87, indicating excellent agreement between the two reviewers. Therefore, all 20 eligible studies underwent meta-analysis (**Table 2**).

The Relationship Between circRNA Expression and Survival Outcomes

The Relationship Between circRNA Expression and OS

The meta-analysis defined OS as the primary endpoint. Among the 20 publications, all studies verified the relationship between

TABLE 1 | Characteristics of the included studies.

Study	Year	circRNA	No. of patients	Outcome	Expression regulation	Median OS	HR	Country	Multivariate analysis	Indirect
Kefeng Wang	2017	circHIAT1	40	OS	Downregulation	Low vs. high	3.75	China	No	Yes
Rui Yu	2020	circNUP98	65	OS	Upregulation	Low vs. high	0.48	China	No	Yes
Rui Yu	2020	circNUP98	65	DFS	Upregulation	Low vs. high	0.52	China	No	Yes
Jianfa Li	2020	circMYLK	71	OS	Upregulation	Low vs. high	1.25	China	No	Yes
Zhuangfei Chen	2019	hsa_circ_001895	60	OS	Upregulation	Low vs. high	0.63	China	No	Yes
Zhengmiao Wang	2019	circHIAT1	80	OS	Downregulation	Low vs. high.	1.90	China	No	Yes
Ling Lin	2019	circ-EGLN3	80	OS	Upregulation	Low vs. high	0.59	China	No	Yes
Jiawei Zeng	2020	circ_001842	97	OS	Upregulation	Low vs. high	0.37	China	No	Yes
Lin Chen	2020	circ_0001368	64	OS	Downregulation	Low vs. high	1.92	China	Yes	No
Qiong Chen	2020	cRAPGEF5	245	OS	Downregulation	Low vs. high	1.79	China	Yes	No
Qiong Chen	2020	cRAPGEF5	245	RFS	Downregulation	Low vs. high	1.64	China	Yes	No
Yunfang Huang	2021	Circ-ABCB10	120	OS	Upregulation	High vs. low	5.29	China	Yes	No
Wen Li	2020	circCSNK1G3	64	OS	Upregulation	Low vs. high	0.34	TCGA-KICH	No	Yes
Qingliang Zhu	2021	CircAKT1	70	OS	Upregulation	Low vs. high	0.56	China	No	Yes
Qi Lv	2021	CircAGAP1		OS	Upregulation	High vs. low	1.68	TCGA-KICH	No	No
Yongjun Yue	2020	Circ_101341	60	OS	Upregulation	High vs. low	0.42	China	No	Yes
Lisa Frey	2021	circEHD2	121	PFS	Downregulation	High vs. low	3.58	Germany	Yes	No
Lisa Frey	2021	circNETO2	121	PFS	Upregulation	High vs. low	0.17	Germany	Yes	No
Lisa Frey	2021	circEHD2	121	CSS	Downregulation	High vs. low	2.67	Germany	Yes	No
Lisa Frey	2021	circNETO2	121	CSS	Upregulation	High vs. low	0.14	Germany	Yes	No
Lisa Frey	2021	circEHD2	121	OS	Downregulation	High vs. low	3.91	Germany	Yes	No
Lisa Frey	2021	circNETO2	121	OS	Upregulation	High vs. low	0.15	Germany	Yes	No
Yanhui Zhao	2020	ciRS-7	87	PFS	Upregulation	Low vs. high	0.53	China	No	Yes
Guanghua Liu	2020	has-hsa_circ_0085576	86	OS	Upregulation	High vs. low	1.37	China	Yes	No
Guanghua Liu	2020	has-hsa_circ_0085576	86	DFS	Upregulation	High vs. low	2.14	China	No	Yes
Jianfa Li	2020	CircTLK1	60	OS	Upregulation	Low vs. high	0.45	China	No	Yes
Jianfa Li	2020	CircTLK1	60	DFS	Upregulation	Low vs. high	0.86	China	No	Yes
Huan Liu	2020	circPTCH1	39	OS	Upregulation	Low vs. high	0.65	China	No	Yes
Bin Han	2020	CircHIPK3	50	OS	Upregulation	Low vs. high	0.42	China	No	Yes

PS: indirect: we indirectly calculated HR from the plot; multivariate analysis: multivariate analysis was used to adjust HR.

TABLE 2 | Quality assessment was based on the Newcastle–Ottawa Scale (NOS).

First author	Year	Selection	Comparability	Outcome	Total score
Wang et al. (2017)	2017	4	2	1	7
Yu et al. (2020)	2020	3	2	1	6
Li et al. (2020a)	2020	4	2	1	7
Chen et al. (2020a)	2019	4	2	2	8
Wang et al. (2019)	2019	3	2	2	7
Lin and Cai (2020)	2019	4	1	1	6
Zeng et al. (2020)	2020	4	2	1	7
Chen et al. (2020b)	2020	4	2	2	8
Chen et al. (2020c)	2020	3	2	2	7
Huang et al. (2019)	2019	4	1	1	6
Li et al. (2021)	2020	3	2	2	7
Zhu et al. (2020)	2020	4	2	2	8
Lv et al. (2021)	2021	4	2	2	8
Yue et al. (2020)	2020	3	1	2	6
Frey et al. (2021)	2021	4	1	1	6
Zhao et al. (2020)	2020	4	2	2	8
Liu et al. (2020a)	2020	3	2	2	7
Li et al. (2020b)	2020	4	2	1	7
Liu et al. (2020b)	2020	4	1	2	7
Han et al. (2020)	2020	4	2	2	8

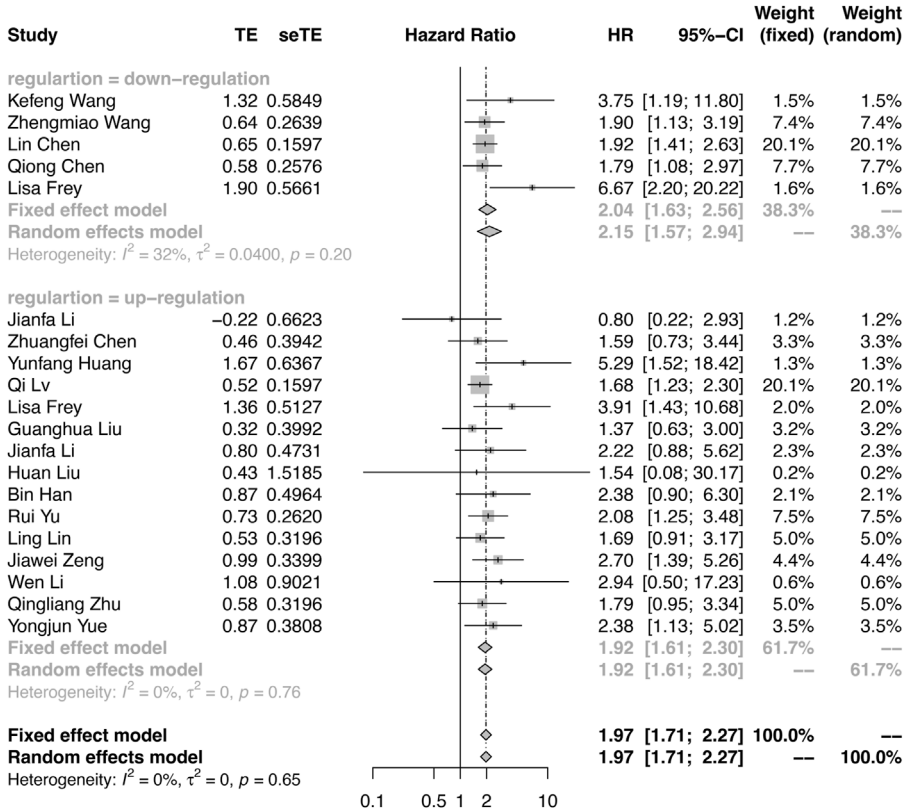


FIGURE 2 | Forest plots verify the association between the expression of circRNAs and overall survival (OS). High expression of oncogenic circRNAs and low expression of tumor-suppressor circRNAs were associated with poor OS.

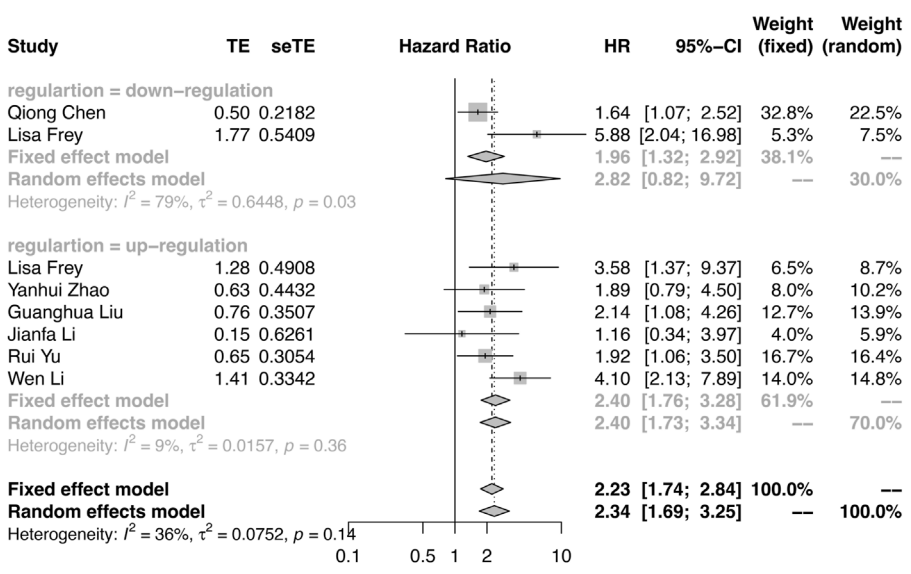


FIGURE 3 | Forest plots verify the association between the expression of circRNAs and progression-free survival (PFS). High expression of oncogenic circRNAs and low expression of tumor-suppressor circRNAs were associated with poor PFS.

TABLE 3 | Meta-regression analysis of the included studies.

Factor	Univariate meta-regression *p*-value	95% CI
Year	0.659	–0.2339 to 0.1479
Regulation	0.682	–0.3491 to 0.9419
Direct vs. indirect	0.768	–0.2401 to 0.3253
Multi	0.531	–0.1992 to 0.3861
Country	0.912	–0.3458 to 0.3087
Factor	**Multivariate meta-regression analysis *p*-value**	**95% CI**
Year	0.611	–0.4241 to 0.2495
Regulation	0.966	–0.4606 to 0.4413
Indirect	0.019	0.1613 to 1.8256
Multi	0.013	0.2048 to 1.7720
Country	0.016	–1.7281 to –0.1788

PS: 1) Year: publication year; 2) Regulation: upregulation vs. downregulation; 3) Direct: extract HR from the manuscript, indirect: calculate HR from the plot; 4) Multivariate analysis: multivariate analysis was used to adjust HR; 5) Country: China vs. TCGA or Germany.

OS and circRNA expression levels (Supplementary Table S2). As shown in **Figure 2**, because of the few heterogeneities among the included studies (OS: downregulation: I^2 = 32%, p = 0.2; upregulation: I^2 = 0%, p = 0.75), the pooled HR and corresponding 95% CI were estimated by applying the fixed-effect model. Overall, the results demonstrated that the differential expression of circRNAs is closely related to poor survival in RCC (HR = 1.97, 95% CI: 1.71–2.27, p < 0.0001). The results indicated that patients with oncogenic circRNA overexpression had worse OS than those with low expression (OS: HR = 2.04, 95% CI: 1.63–2.56, p < 0.0001). Furthermore, the results demonstrated that the lower expression of some tumor-suppressor circRNAs was associated with poor OS (HR: 1.92, 95% CI: 1.61–2.30, p < 0.0001) (**Figure 2**).

The Relationship Between circRNA Expression and PFS

Among the included studies, PFS was reported in three studies, RFS was reported in one study, and DFS was reported in four studies (Supplementary Table S3). Considering the similar survival outcomes, RFS and DFS were combined as PFS. As shown in **Figure 3**, there were large heterogeneities (I^2 = 79%, p = 0.03) in the downregulation group and a few heterogeneities (I^2 = 9%, p = 0.36) in the upregulation group. A random effect model was applied in the downregulation group, and a fixed effect model was applied in the upregulation group. The results also demonstrated that the oncogenic circRNA overexpression was associated with worse PFS than low expression (PFS: HR = 2.82, 95% CI: 0.82–9.72, p < 0.0001), and lower expression of tumor-suppressor circRNAs was correlated with poor PFS (HR: 2.40, 95% CI: 1.76–3.28, p = 0.0009) (**Figure 3**).

Meta-Regression

Meta-regression was performed based on OS. We performed the univariate meta-regression and multivariate meta-regression to further assess the heterogeneity (**Table 3**). The results revealed that plot and country of multivariate analysis may significantly influence the variation in HR (univariate meta-regression: p-value for year = 0.659, p-value for expression regulation = 0.682, p-value for indirect = 0.768, p-value for multivariate analysis = 0.531, and p-value for country; multivariate meta-regression: p-value for year = 0.611, p-value for regulation = 0.966, p-value for indirect = 0.019, p-value for multivariate analysis = 0.013, and p-value for country = 0.016).

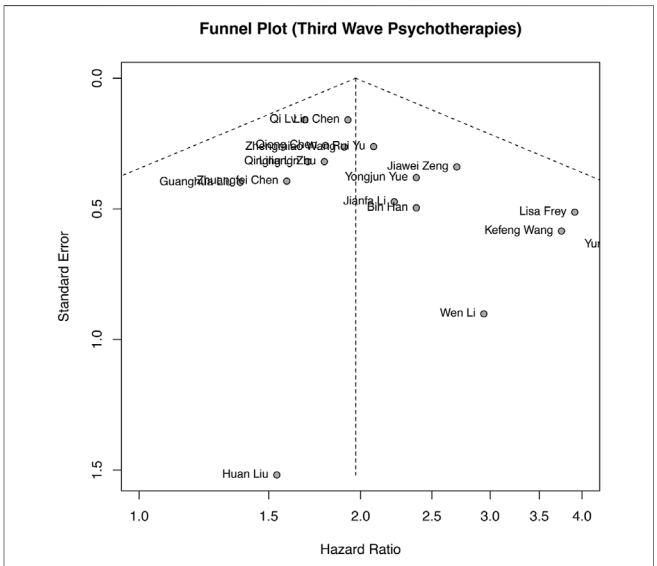


FIGURE 4 | Funnel plots of studies included in the meta-analyses.

Publication Bias

Publication bias analysis was performed based on OS, as shown in **Figure 4**, and the funnel plot was approximately symmetrical. Publication bias was also assessed by Begg's and Egger's tests in the meta-analysis. The results (Begg's p = 0.0693 and Egger's p = 0.0798) and funnel plot all indicated no obvious publication bias among the included studies.

DISCUSSION

RCC is characterized by both high mortality and morbidity (Jonasch et al., 2014). Metastasis and recurrence are some of the leading causes of death. Therefore, more accurate biomarkers are needed for predicting the prognosis of RCC patients to monitor the patient's condition. Here, we summarized the prognostic value of circRNAs in RCC. This meta-analysis, including 21 studies and 1,559 RCC patients, is the first to investigate the relationship between circRNAs and the prognosis of RCC patients. According to the included studies, we found that oncogenic circRNAs with high expression and tumor-suppressor circRNAs with lower expression were correlated with worse survival, which indicated that circRNAs may play important roles in tumor initiation and progression.

CircRNAs play both oncogenic and tumor-suppressor roles in RCC (Yang et al., 2021). CircRNAs function predominantly by acting as sponges of microRNAs. In this way, circRNAs can regulate tumor-related signaling pathways. Li et al. (2020a) found that circMYLK is notably upregulated in RCC and that circMYLK upregulation can promote tumor growth. This role was achieved by circular RNA MYLK regulating miR-513a-5p/VEGFC signaling. Furthermore, Zhang et al. also found that circular RNA hsa_circ_0054537 can regulate the cMet pathway to promote the progression of RCC by sponging miR-130a-3p (Li et al., 2020c). In addition to its oncogenic role, Chen et al. (2019) also found that circular RNA hsa-circ-0072309 can play an antitumor role by deactivating the PI3K/AKT and mTOR pathways by sponging miR-100. Sun et al. (2020) found that circUBAP2 can regulate the miR-148a-3p/FOXK2 pathway to inhibit the proliferation and metastasis of RCC. Furthermore, circRNAs can also regulate transcription factors to impact tumor initiation and evolution. For example, circular RNA hsa_circ_001895 regulates SRY-box transcription factor 12 (SOX12) by sponging microRNA-296-5p to promote RCC progression (Chen et al., 2020a). These studies indicate that circRNAs play an important role in tumorigenesis, tumor development, and metastasis, which demonstrates that circRNAs have the potential to act as biomarkers for the prognosis of RCC.

However, we tried to ensure the authenticity and reliability of the meta-analysis. Nonetheless, there are still some limitations in this research. Only 1,559 RCC patients were included, and all samples were from China and Germany, which also leads to publication bias. Thus, more studies performed in other parts of the world are needed. Furthermore, because some HR values were not reported in the article, HR was calculated by Kaplan/Meier curves in some studies lacking HR values. Subjective factors may be introduced. Future studies investigating the relationship between the expression of circRNAs and prognosis in RCC need to provide more complete data.

In summary, we performed a meta-analysis to identify the prognostic value of circRNAs in RCC patients. The results demonstrate that the high expression of circRNAs with cancer-promoting effects and the low expression of circRNAs with tumor-suppressing effects are associated with poor prognosis in RCC patients. Furthermore, many circRNAs play significant roles in RCC initiation and progression. Future studies utilizing circRNAs may demonstrate an effective prognostic biomarker in all RCC patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: QH, DL, and QL; methodology: DL, QL, and HX; validation: FZ, QH, and QL; data curation: DL, QL, HX, and QH; writing—original draft preparation: DL, QH, and HX; writing—review and editing: DL, QL, HX, and FZ; visualization: DL, QL, and QH; supervision: DL and QH; project administration: QL, FZ, and QH; and funding acquisition: DL and QH. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Special Fund for Hypertension, Sichuan Medical Association (No. 2019TG26).

ACKNOWLEDGMENTS

The authors would like to thank all of the patients, investigators, and staff involved in the included studies.

REFERENCES

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2010). A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis. *Res. Synth. Method* 1 (2), 97–111. doi:10.1002/jrsm.12

Capitanio, U., Bensalah, K., Bex, A., Boorjian, S. A., Bray, F., Coleman, J., et al. (2019). Epidemiology of Renal Cell Carcinoma. *Eur. Urol.* 75 (1), 74–84. doi:10.1016/j.eururo.2018.08.036

Cerbone, L., Cattrini, C., Vallome, G., Latocca, M. M., Boccardo, F., and Zanardi, E. (2020). Combination Therapy in Metastatic Renal Cell Carcinoma: Back to the Future? *Seminars Oncol.* 47 (6), 361–366. doi:10.1053/j.seminoncol.2020.10.003

Chen, B., and Huang, S. (2018). Circular RNA: An Emerging Non-coding RNA as a Regulator and Biomarker in Cancer. *Cancer Lett.* 418, 41–50. doi:10.1016/j.canlet.2018.01.011

Chen, L.-L., and Yang, L. (2015). Regulation of circRNA Biogenesis. *RNA Biol.* 12 (4), 381–388. doi:10.1080/15476286.2015.1020271

Chen, L., Wu, D., and Ding, T. (2020). Circular RNA Circ_0001368 Inhibited Growth and Invasion in Renal Cell Carcinoma by Sponging miR-492 and Targeting LATS2. *Gene* 753, 144781. doi:10.1016/j.gene.2020.144781

Chen, Q., Liu, T., Bao, Y., Zhao, T., Wang, J., Wang, H., et al. (2020). CircRNA cRAPGEF5 Inhibits the Growth and Metastasis of Renal Cell Carcinoma via the miR-27a-3p/TXNIP Pathway. *Cancer Lett.* 469, 68–77. doi:10.1016/j.canlet.2019.10.017

Chen, T., Shao, S., Li, W., Liu, Y., and Cao, Y. (2019). RETRACTED ARTICLE: The Circular RNA Hsa-Circ-0072309 Plays Anti-tumour Roles by Sponging miR-100 through the Deactivation of PI3K/AKT and mTOR Pathways in the Renal Carcinoma Cell Lines. *Artif. Cells, Nanomedicine, Biotechnol.* 47 (1), 3638–3648. doi:10.1080/21691401.2019.1657873

Chen, Z., Xiao, K., Chen, S., Huang, Z., Ye, Y., and Chen, T. (2020). Circular RNA Hsa_circ_001895 Serves as a Sponge of microRNA-296-5p to Promote Clear Cell Renal Cell Carcinoma Progression by Regulating SOX12. *Cancer Sci.* 111 (2), 713–726. doi:10.1111/cas.14261

Frey, L., Klümper, N., Schmidt, D., Kristiansen, G., Toma, M., Ritter, M., et al. (2021). CircEHD2, CircNETO2 and CircEGLN3 as Diagnostic and Prognostic Biomarkers for Patients with Renal Cell Carcinoma. *Cancers* 13 (9), 2177. doi:10.3390/cancers13092177

Haddad, A. Q., and Margulis, V. (2015). Tumour and Patient Factors in Renal Cell Carcinoma-Towards Personalized Therapy. *Nat. Rev. Urol.* 12 (5), 253–262. doi:10.1038/nrurol.2015.71

Han, B., Chao, J., and Yao, H. (2018). Circular RNA and its Mechanisms in Disease: From the Bench to the Clinic. *Pharmacol. Ther.* 187, 31–44. doi:10.1016/j.pharmthera.2018.01.010

Han, B., Shaolong, E., Luan, L., Li, N., and Liu, X. (2020). CircHIPK3 Promotes Clear Cell Renal Cell Carcinoma (ccRCC) Cells Proliferation and Metastasis via Altering of miR-508-3p/CXCL13 Signal. *Ott* 13, 6051–6062. doi:10.2147/OTT.S251436

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., et al. (2017). Renal Cell Carcinoma. *Nat. Rev. Dis. Prim.* 3, 17009. doi:10.1038/nrdp.2017.9

Huang, Y., Zhang, Y., Jia, L., Liu, C., and Xu, F. (2019). Circular RNA ABCB10 Promotes Tumor Progression and Correlates with Pejorative Prognosis in Clear Cell Renal Cell Carcinoma. *Int. J. Biol. Markers* 34 (2), 176–183. doi:10.1177/1724600819842279

Jonasch, E., Gao, J., and Rathmell, W. K. (2014). Renal Cell Carcinoma. *BMJ* 349, g4797. doi:10.1136/bmj.g4797

Li, J., Huang, C., Zou, Y., Ye, J., Yu, J., and Gui, Y. (2020). CircTLK1 Promotes the Proliferation and Metastasis of Renal Cell Carcinoma by Sponging miR-136-5p. *Mol. Cancer* 19 (1), 103. doi:10.1186/s12943-020-01225-2

Li, J., Huang, C., Zou, Y., Yu, J., and Gui, Y. (2020). Circular RNA MYLK Promotes Tumour Growth and Metastasis via Modulating miR-513a-5p/VEGFC Signalling in Renal Cell Carcinoma. *J. Cell. Mol. Med.* 24 (12), 6609–6621. doi:10.1111/jcmm.15308

Li, R., Luo, S., and Zhang, D. (2020). Circular RNA Hsa_circ_0054537 Sponges miR-130a-3p to Promote the Progression of Renal Cell Carcinoma through Regulating cMet Pathway. *Gene* 754, 144811. doi:10.1016/j.gene.2020.144811

Li, W., Song, Y. Y. Y., Rao, T., Yu, W. M., Ruan, Y., Ning, J. Z., et al. (2021). CircCSNK1G3 Up-regulates miR-181b to Promote Growth and Metastasis via TIMP3-mediated Epithelial to Mesenchymal Transitions in Renal Cell Carcinoma. *J. Cell. Mol. Medi* 26, 1729–1741. doi:10.1111/jcmm.15911

Lin, L., and Cai, J. (2020). Circular RNA circ-EGLN3 Promotes Renal Cell Carcinoma Proliferation and Aggressiveness via miR-1299-mediated IRF7 Activation. *J. Cell. Biochem.* 121 (11), 4377–4385. doi:10.1002/jcb.29620

Liu, G., Zhou, J., Piao, Y., Zhao, X., Zuo, Y., and Ji, Z. (2020). Hsa_circ_0085576 Promotes Clear Cell Renal Cell Carcinoma Tumorigenesis and Metastasis through the miR-498/YAP1 axis. *Aging* 12 (12), 11530–11549. doi:10.18632/aging.103300

Liu, H., Hu, G., Wang, Z., Liu, Q., Zhang, J., Chen, Y., et al. (2020). circPTCH1 Promotes Invasion and Metastasis in Renal Cell Carcinoma via Regulating miR-485-5p/MMP14 axis. *Theranostics* 10 (23), 10791–10807. doi:10.7150/thno.47239

Lv, Q., Wang, G., Zhang, Y., Shen, A., Tang, J., Sun, Y., et al. (2021). CircAGAP1 Promotes Tumor Progression by Sponging miR-15-5p in Clear Cell Renal Cell Carcinoma. *J. Exp. Clin. Cancer Res.* 40 (1), 76. doi:10.1186/s13046-021-01864-3

Spruance, S. L., Reid, J. E., Grace, M., and Samore, M. (2004). Hazard Ratio in Clinical Trials. *Antimicrob. Agents Chemother.* 48 (8), 2787–2792. doi:10.1128/aac.48.8.2787-2792.2004

Stang, A. (2010). Critical Evaluation of the Newcastle-Ottawa Scale for the Assessment of the Quality of Nonrandomized Studies in Meta-Analyses. *Eur. J. Epidemiol.* 25 (9), 603–605. doi:10.1007/s10654-010-9491-z

Sun, J., Yin, A., Zhang, W., Lv, J., Liang, Y., Li, H., et al. (2020). CircUBAP2 Inhibits Proliferation and Metastasis of Clear Cell Renal Cell Carcinoma via Targeting miR-148a-3p/FOXK2 Pathway. *Cell. Transpl.* 29, 096368972092575. doi:10.1177/0963689720925751

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Wang, K., Sun, Y., Tao, W., Fei, X., and Chang, C. (2017). Androgen Receptor (AR) Promotes Clear Cell Renal Cell Carcinoma (ccRCC) Migration and Invasion via Altering the circHIAT1/miR-195-5p/29a-3p/29c-3p/CDC42 Signals. *Cancer Lett.* 394, 1–12. doi:10.1016/j.canlet.2016.12.036

Wang, Z., Zhao, Y., Wang, Y., and Jin, C. (2019). Circular RNA circHIAT1 Inhibits Cell Growth in Hepatocellular Carcinoma by Regulating miR-3171/PTEN axis. *Biomed. Pharmacother.* 116, 108932. doi:10.1016/j.biopha.2019.108932

Xu, W., Atkins, M. B., and McDermott, D. F. (2020). Checkpoint Inhibitor Immunotherapy in Kidney Cancer. *Nat. Rev. Urol.* 17 (3), 137–150. doi:10.1038/s41585-020-0282-3

Yang, D. C., and Chen, C.-H. (2020). Potential New Therapeutic Approaches for Renal Cell Carcinoma. *Seminars Nephrol.* 40 (1), 86–97. doi:10.1016/j.semnephrol.2019.12.010

Yang, L., Zou, X., Zou, J., and Zhang, G. (2021). Functions of Circular RNAs in Bladder, Prostate and Renal Cell Cancer (Review). *Mol. Med. Rep.* 23 (5). doi:10.3892/mmr.2021.11946

Yu, R., Yao, J., and Ren, Y. (2020). A Novel circRNA, circNUP98, a Potential Biomarker, Acted as an Oncogene via the miR-567/PRDX3 axis in Renal Cell Carcinoma. *J. Cell. Mol. Med.* 24 (17), 10177–10188. doi:10.1111/jcmm.15629

Yue, Y., Cui, J., Zhao, Y., Liu, S., and Niu, W. (2020). Circ_101341 Deteriorates the Progression of Clear Cell Renal Cell Carcinoma through the miR- 411/EGLN3 Axis. *Cmar* 12, 13513–13525. doi:10.2147/CMAR.S272287

Zeng, J., Feng, Q., Wang, Y., Xie, G., Li, Y., Yang, Y., et al. (2020). Circular RNA Circ_001842 Plays an Oncogenic Role in Renal Cell Carcinoma by Disrupting microRNA-502-5p-mediated Inhibition of SLC39A14. *J. Cell. Mol. Med.* 24 (17), 9712–9725. doi:10.1111/jcmm.15529

Zhao, Y.-H., Wang, Z., Zhang, N., Cui, T., and Zhang, Y.-H. (2020). Effect of ciRS-7 Expression on Clear Cell Renal Cell Carcinoma Progression. *Chin. Med. J. Engl.* 133 (17), 2084–2089. doi:10.1097/CM9.0000000000000867

Zhou, J.-G., Tian, X., Wang, X., Tian, J.-H., Wang, Y., Wang, F., et al. (2015). Treatment on Advanced NSCLC: Platinum-Based Chemotherapy Plus Erlotinib or Platinum-Based Chemotherapy Alone? A Systematic Review and Meta-Analysis of Randomised Controlled Trials. *Med. Oncol.* 32 (2), 471. doi:10.1007/s12032-014-0471-0

Zhu, Q., Zhan, D., Zhu, P., Chong, Y., and Yang, Y. (2020). CircAKT1 Acts as a Sponge of miR-338-3p to Facilitate Clear Cell Renal Cell Carcinoma Progression by Up-Regulating CAV1. *Biochem. Biophysical Res. Commun.* 532 (4), 584–590. doi:10.1016/j.bbrc.2020.08.081

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liao, Lin, Xiao, Zhang and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fgene.2022.877656

Systematic Construction and Validation of a Novel Macrophage Differentiation–Associated Prognostic Model for Clear Cell Renal Cell Carcinoma

Chen Liu[1,2], *Xuhui Zhang*[2], *Caoyang Hu*[2], *Xuezhi Liang*[2], *Xiaoming Cao*[2] and *Dongwen Wang*[1,3]*

[1]First Clinical Medical College, Shanxi Medical University, Taiyuan, China, [2]Department of Urology, First Hospital of Shanxi Medical University, Taiyuan, China, [3]Department of Urology, Cancer Hospital Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Jiayi Ji,
Rutgers, The State University of New Jersey, United States
Shaogang Wang,
Huazhong University of Science and Technology, China

***Correspondence:**
Dongwen Wang
urology2007@126.com

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 17 February 2022
Accepted: 09 May 2022
Published: 14 June 2022

Citation:
Liu C, Zhang X, Hu C, Liang X, Cao X and Wang D (2022) Systematic Construction and Validation of a Novel Macrophage Differentiation–Associated Prognostic Model for Clear Cell Renal Cell Carcinoma. Front. Genet. 13:877656. doi: 10.3389/fgene.2022.877656

Background: Clear cell renal cell carcinoma (ccRCC) is a malignant tumor of the human urinary system. Macrophage differentiation is associated with tumorigenesis. Therefore, exploring the prognostic value of macrophage differentiation–associated genes (MDGs) may contribute to better clinical management of ccRCC patients.

Methods: The RNA sequence data of ccRCC were obtained from The Cancer Genome Atlas (TCGA) database. Differentially expressed MDGs were unveiled in ccRCC and normal samples. The prognostic model was established according to the univariate and multivariate Cox regression analyses. By combining clinico-pathological features and prognostic genes, a nomogram was established to predict individual survival probability. The Tumor Immune Estimation Resource (TIMER) database was utilized to analyze the correlation between prognostic genes and immune infiltrating cells. Eventually, the mRNA and protein expression levels of prognostic genes were verified.

Results: A total of 52 differentially expressed prognosis-related MDGs were identified in ccRCC. Afterward, a six-gene prognostic model (ABCG1, KDF1, KITLG, TGFA, HAVCR2, and CD14) was constructed through the Cox analysis. The overall survival in the high-risk group was relatively poor. Moreover, the risk score was identified as an independent prognostic factor. We constructed a prognostic nomogram with a well-fitted calibration curve based on risk score and clinical data. Furthermore, the prognostic genes were significantly related to the level of immune cell infiltration including B cells, CD8+T cells, CD4+T cells, macrophages, neutrophils, and dendritic cells. Finally, the mRNA expression of prognostic genes in clinical ccRCC tissues showed that the ABCG1, HAVCR2, CD14, and TGFA mRNA in tumor samples were increased compared with the adjacent control tissue samples, while KDF1 and KITLG were decreased, which was consistent with the verification results in the GSE53757.

Conclusion: In conclusion, this study identified and validated a macrophage differentiation–associated prognostic model for ccRCC that could be used to predict the outcomes of the ccRCC patients.

Keywords: clear cell renal cell carcinoma, macrophage differentiation, prognostic, TCGA, bioinformatics

INTRODUCTION

Renal cell carcinoma (RCC) is the most common malignant tumor in the kidney, accounting for about 2%–3% of all cancers (Wang et al., 2019). Clear cell renal cell carcinoma (ccRCC) is the most common pathological subtype of RCC. It originates from the proximal ureter and shows certain invasiveness (Liang et al., 2020). ccRCC has a high incidence rate and mortality rate and poor prognosis. Although tumor detection and treatment methods have made some progress and the survival rate of ccRCC patients has also been significantly improved, the overall survival and progression-free survival are still low (Jonasch et al., 2014), and the risk of metastasis and recurrence is also high (Cohen and McGovern, 2005). In addition, the heterogeneity of tumors makes great differences in the survival rate of patients. In recent years, many molecular biomarkers of ccRCC have been found. C1q/tumor necrosis factor (C1QTNF) (Lin W et al., 2020) and 6-snoRNA characteristics (Zhao et al., 2020) can be used as independent prognostic indicators of ccRCC. Due to the lack of accurate and effective ccRCC therapeutic molecular targets, it is still of great significance to explore new prognostic molecular markers or therapeutic targets in ccRCC.

Macrophages are a kind of innate immune cells, which have the functions of chemotaxis, phagocytosis, regulating inflammatory response, and killing microorganisms, and are an important part of the nonspecific immunity of the body. Recently, it has been found that macrophages can change their phenotypes according to the changing microenvironment, and thus have diverse functions (Shapouri-Moghaddam et al., 2018). There are two major macrophage phenotypes, classically activated macrophages (M1) and selectively activated macrophages (M2), which represent the two extremes of macrophage polarization. M1 macrophages are the main effector cells for the host to destroy pathogens, while M2 macrophages can suppress inflammatory responses and promote angiogenesis and tissue remodeling and repair (Sica and Mantovani, 2012). Research indicates that macrophage differentiation is related to tumorigenesis (Cao et al., 2021). M2-polarized tumor-associated macrophages (M2-TAMs) can help tumor cells change the microenvironment; promote tumor growth, angiogenesis, invasion, and metastasis; and inhibit antitumor immune response (Yuan et al., 2020). M2 macrophages play an important role in the progression and infiltration of lung cancer (Pritchard et al., 2020) and gastric and breast cancer (Chen et al., 2017). Wang Y et al. (2021) study showed that M2 macrophages correlated with the immune microenvironment of clear cell renal cell carcinoma. The key genes involved in macrophage differentiation are of great significance for improving the prognosis and immunotherapeutic effect of patients with hepatocellular carcinoma (Cao et al., 2021). Macrophage differentiation–associated genes (MDGs) may be of great value in screening tumor markers and predicting tumor prognosis. However, the prognostic value and immunotherapy of MDGs in patients with ccRCC have not been reported.

This study mainly constructs a prognostic model based on MDGs and establishes a nomogram to predict individual survival probability by combining clinicopathological features and prognostic genes. The Tumor Immune Estimation Resource (TIMER) database was used to analyze the correlation between prognostic genes and immune infiltrating cells. Moreover, the expression levels of the prognostic genes were verified in clinical ccRCC tissues. It provides a basis and new reference for the treatment of ccRCC and also provides a basis for the exploration of molecular mechanisms related to the prognosis of ccRCC.

MATERIALS AND METHODS

Data Source

The RNA-sequencing data set and clinical data of 530 ccRCC and 72 normal tissue samples were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) database. Transcriptomic and clinical information of 91 ccRCC tissue samples was obtained from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/) as a validation set for the model.

A total of 453 macrophage differentiation–associated genomes (MDGs) were identified using the Gene Cards database (http://www.genecards.org/) and the Gene Set Enrichment Analysis (GSEA) gene set (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). To verify the robustness of the results, we used the Gene Expression Omnibus (GEO) database to perform a validation of hub gene expression level, among which 72 ccRCC and 72 normal tissue samples were selected in the GSE53757.

Identification of Differentially Expressed Macrophage Differentiation–Associated Genomes

The "limma" package was used to identify differentially expressed MDGs in the TCGA—ccRCC and normal tissue samples, with $|\log 2$ (fold change, FC)$| > 1$ and $p < 0.05$ as the cut-off values, and the result was shown as a volcano map.

Identification of Differentially Expressed Prognosis-Related Macrophage Differentiation–Associated Genomes

In order to obtain the MDGs associated with the prognosis of ccRCC, the univariate Cox regression analysis was performed on

438 MDGs that were present in ccRCC, and $p < 0.05$ was identified as significant. Differentially expressed MDGs and prognosis-related MDGs then were taken to intersect to obtain prognostic differentially expressed MDGs. The STRING (https://string-db.org) website was utilized for building a protein–protein interaction (PPI) network, and the "corrr" package in R was used to construct the correlation network of the obtained prognostic differentially expressed MDGs.

Construction and Verification of the Prognostic Model

To better predict the outcome of ccRCC patients, we divide the TCGA–ccRCC sample into a training set and test set at a ratio of 6:4. The univariate Cox analysis and the multivariate Cox regression analysis were performed in the TCGA–ccRCC training set, and then the prognostic model was obtained and validated. In brief, we considered key MDGs that are significantly related to prognosis by the univariate Cox analysis as influencing factors. After bringing them into the multivariate Cox proportional hazard model, significant MDGs will be retained during multiple computing. Multiple genes with a p value less than 0.05 in the univariate Cox regression analysis were subjected to multivariate Cox analysis. Subsequently, the stepwise regression function was used, and the parameter direction was set as "both" to adjust the multivariate regression model, and finally, the optimal model was obtained. The weighted coefficients based on individual gene expression levels were used to calculate the risk score as follows: risk score = $\sum$ regression coefficient (genei) × expression value (genei). The patients in the training set were then stratified into the low- and high-risk groups according to median risk score values, and survival was analyzed using the Kaplan–Meier (K-M) analysis. The K-M analysis and log-rank test were used to assess differences in survival between different groups using the package "survival" in R, and the package "survminer" was used for visualizing the results. The 1-, 3-, and 5-year survival was determined by the receiver operating characteristic (ROC) curve analysis using the "survivalROC" package in R, and the areas under the curve (AUC) were calculated. The prognostic model was further validated in the TCGA—ccRCC test set and the ICGC external validation set.

Association of the Prognostic Model and Clinicopathological Features

To further investigate the prognosis of clinicopathological features and risk score, the correlation between clinical factors including gender, age, stage, grade, T/M/N, and risk score was performed in the TCGA–ccRCC training set first, Wilcoxon test was used to compare differences in risk scores between gender, age, M, and N subgroups, and Kruskal–Wallis test was used to analyze the association between clinicopathological features (stage, grade, and T) and risk scores. Then, the independent prognostic factor was analyzed. In brief, clinical factors including gender, age, stage, grade, T/M/N, and risk scores were determined by univariate Cox analysis. Multivariate Cox regression analysis was then used to identify the independent prognostic factors for ccRCC. Prognostic nomograms were generated using the least absolute shrinkage and selection operator (LASSO) with the "glmnet" package in the TCGA—ccRCC training set that includes gender, age, stage, grade, T/M/N, and risk score clinical characteristics.

Analysis of the Kyoto Encyclopedia of Genes and Genomes Signal Pathway in High- and Low-Risk Groups

To analyze the signaling pathway differences between high- and low-risk groups, the "gsva" package in R was used to calculate the score of the Kyoto Encyclopedia of Genes and Genomes (KEGG) signal pathway in each sample. The KEGG pathway of the top 50 was plotted on a heat map.

Analysis of the Correlation Between Prognostic Macrophage Differentiation–Associated Genomes and Tumor-Infiltrating Immune Cells

The correlation analysis between prognostic MDGs and tumor-infiltrating immune cells (including tumor purity, B cells, CD4+T cells, CD8+T cells, macrophages, neutrophils, and dendritic cells) was performed using the TIMER database in ccRCC, and $p < 0.05$ was considered statistically significant. Also, the Spearman correlation analysis was used to detect their correlation.

Verification of Prognostic Macrophage Differentiation–Associated Genome Expression Level

The mRNA and protein expression level of prognostic MDGs in the GSE53757 datset was verified by using the Wilcoxon test paired test method.

RNA Extraction and Quantitative Real-Time Polymerase Chain Reaction

The mRNA expression levels of prognostic genes were detected in 10 pairs of ccRCC tissue samples and para-cancerous control tissue samples from the First Hospital of ShanXi Medical University. This study was allowed by the Ethics Committee of the First Hospital of ShanXi Medical University. All patients had approved for the use of clinical tissues for research purposes. Total RNA was isolated using TRIzol (Genecopoeia). The SureScript-First-strand-cDNA-synthesis-kit (Genecopoeia) was used for first-strand cDNA synthesis. For the analysis of the target gene mRNA levels, qPCR was performed using BlazeTaq™ SYBR ® Green qPCR Mix 2.0 according to the manufacturer's instructions (Genecopoeia). The relative expression of mRNA was calculated by the $2^{-\Delta\Delta Ct}$ method with the normalization to GAPDH. All specific primers were shown in **Table 1**.

TABLE 1 | Primer sets were used for the PCR assay in this study.

Target gene		Sequence (5′→3′)
KDF1	F	TGTCGAGCCGAGAAATTGAT
	R	GGTGGTAGTCCTGCGTGATG
KITLG	F	CTTGTGGAGTGCGTGAAAG
	R	AGGAGTAAAGAGCCTGGGT
TGFA	F	GTATTGTGTTGGCTGCGTG
	R	CTGAGTGTGGGAATCTGGG
CD14	F	GCAACACAGGAATGGAGAC
	R	GAACGACAGATTGAGGGAG
ABCG1	F	GGATGAAGGCAGAAGGGAA
	R	GTGCAAATGATGGAGCGAC
HAVCR2	F	CCTATCTGCCCTGCTTCTAC
	R	CTCCCCAGTGTCTGTGTCTC
GAPDH	F	CCCATCACCATCTTCCAGG
	R	CATCACGCCACAGTTTCCC

RESULTS

Identification of Differentially Expressed Prognostic Related Macrophage Differentiation-Associated Genomes

A total of 120 differentially expressed MDGs were selected where 87 MDGs were upregulated and 33 MDGs were downregulated in ccRCC samples (**Figure 1A**). A total of 438 MDGs in ccRCC were obtained from the Gene Cards database and GSEA gene se and, then the univariate Cox analysis found that 213 MDG were significantly related to the prognosis of ccRCC. Finally, 52 differentially expressed prognostic-related MDGs were obtained by taking the intersection of 120 differentially expressed MDGs and 213 significantly prognostic-related MDGs of ccRCC (**Figure 1B**). The PPI network analysis of 52 MDGs obtained a protein interaction network comprising 52 nodes and 286 edges (**Figure 1C**). In addition, further analysis revealed that there was a positive correlation between multiple MDGs (**Figure 1D**).

Construction of Prognostic Model in Clear Cell Renal Cell Carcinoma

Thirty-four survival-associated MDGs were identified in the training group by univariate Cox analysis, of which ABCG1, KDF1, PECAM1, ERBB2, KITLG, JAK3, ADA, TGFA, CEBPB, CDKN2A, FLT1, MTURN, TLR3, CCND1, BMP6, CDH1, GDF6, MMP9, KIT, NDRG1, RUNX1, WT1, ITGAX, METRNL, HAVCR2, ENO2, FCGR1A, CSF3R, CD68, CD44, GA TA2, S100A8, TNFRSF11B, and CD14 were further screened by the multivariate Cox proportional hazards model. A six-gene (ABCG1, KDF1, KITLG, TGFA, HAVCR2, CD14) prognostic model was constructed using regression coefficients of each gene, and the risk score was calculated (**Figure 2A**). The risk scores were calculated for each patient in the training group, and the patients were assigned to the high-risk or low-risk group based on the median risk score. As shown in **Figure 2B**, patients with high-risk scores had significantly poorer survival outcomes than those with low-risk scores ($p < 0.05$). Furthermore, the AUC of the risk score for 1-year, 3-year, and 5-year overall survival (OS) was 0.792, 0.739, and 0.748, respectively (**Figure 2C**). The survival status, risk scores, and gene expression data of ccRCC patients in the training group are illustrated in **Figure 2D**.

Validation of the Prognostic Model

To further validate the prognostic signature associated with macrophage differentiation, its prognostic accuracy was further assessed in three independent cohorts, including the ICGC validation group and TCGA–ccRCC test set. The OS was significantly longer for patients in the low-risk than in the high-risk group in the TCGA–ccRCC test set (**Figure 3A**), and the predicted 1-year, 3-year, and 5-year OS was 0.630, 0.642, and 0.653, respectively (**Figure 3B**). The ICGC validation cohort also validated the prognostic accuracy of the prognostic model (**Figure 3C**), with respective AUCs of 0.602, 0.604, and 0.637 for 1-year, 3-year, and 5-year OS, respectively (**Figure 3D**). Thus, the prognostic model predicted the OS of ccRCC patients with superior accuracy.

The Prognostic Model Confers Additional Prognostic Power for Clear Cell Renal Cell Carcinoma Patients

Seven clinicopathological factors, including gender, age, stage, grade, and T/M/N, were included in the prognostic model for univariate Cox independent analysis. As shown in **Figure 4A**, clinical factors including risk score, T/M/N, stage, and grade were closely associated with patient survival. Multivariate Cox regression analysis further showed that the risk score was an independent prognostic indicator for OS in the TCGA–ccRCC cohort (**Figure 4B**). Expression profiles of the six MDGs are shown in **Figure 4C** and significant differences were found in the expression among stages, grades, T, M, and N. In addition, the risk score was significantly different in different stages, grades, T, M, and N (**Figure 5**).

Construction and Verification of a Nomogram

In order to establish a clinical method for predicting the survival probability of ccRCC patients, we constructed a nomogram to predict the possibility of 1-year, 3-year, and 5-year OS. As shown in **Figure 6A**, the score assigned to each factor is proportional to its risk contribution to survival. Thus, based on the patient's characteristic scores, we could obtain 1-, 3-, and 5-year survival rates for ccRCC patients. We draw a nomogram correction curve based on the aforementioned prediction model. The results show that the prediction accuracy of the model for patient 1-, 3-, and 5-year survival rates is very high, indicating that the constructed prediction model can be used as an effective model (**Figure 6B**). In addition, the concordance index analysis shows that the risk score, nomogram, and M were all higher than 0.5 of 1–8 years of OS prediction (**Figure 6C**).

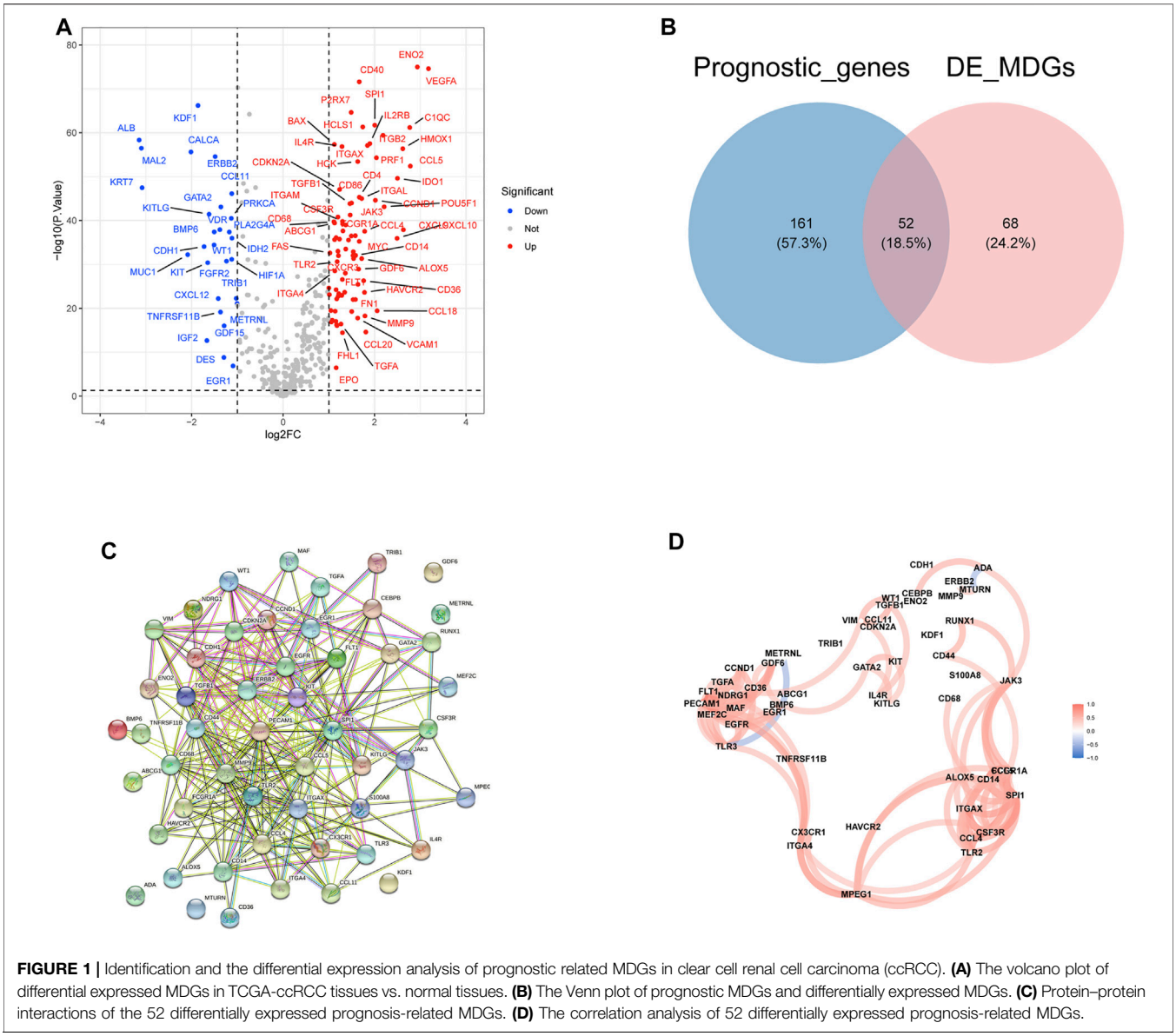


FIGURE 1 | Identification and the differential expression analysis of prognostic related MDGs in clear cell renal cell carcinoma (ccRCC). **(A)** The volcano plot of differential expressed MDGs in TCGA-ccRCC tissues vs. normal tissues. **(B)** The Venn plot of prognostic MDGs and differentially expressed MDGs. **(C)** Protein–protein interactions of the 52 differentially expressed prognosis-related MDGs. **(D)** The correlation analysis of 52 differentially expressed prognosis-related MDGs.

Differential Analysis of Kyoto Encyclopedia of Genes and Genomes Enrichment in High- and Low-Risk Groups

Differential analysis of KEGG pathway enrichment in high- and low-risk groups revealed that the high-risk group was mainly enriched in the systemic lupus erythematosus, RIG l-like receptor signaling pathway, NOD-like receptor signaling pathway, cytoplasmic DNA sensing pathway, antigen processing and presentation, primary immunodeficiency, etc. pathways, while the low-risk group was mainly enriched in the lysine degradation, proximal tubule bicarbonate recovery, glycolytic gluconeogenesis, PPAR signal pathway, renin–angiotensin system, etc. pathways (**Figure 7**).

Analysis of the Correlation Between Prognostic Macrophage Differentiation–Associated Genomes and Immune Infiltration

Correlation analysis of six prognostic MDGs (ABCG1, KDF1, KITLG, TGFA, HAVCR2, and CD14) and six types of immune infiltrating cells, including tumor purity, B cells, CD8+T cells, CD4+T cells, macrophages, neutrophils, and dendritic cells (**Figures 8A–F**) was performed. ABCG1 was significantly correlated with $CD8^+$ T cells, macrophages, neutrophils, and dendritic cells. CD14 was significantly correlated with purity, B cells, $CD8^+$ T cells, CD4+T cells, macrophages, neutrophils, and dendritic cells. HAVCR2 was significantly correlated with B cells, $CD8^+$ T cells, macrophages, neutrophils, and dendritic cells.

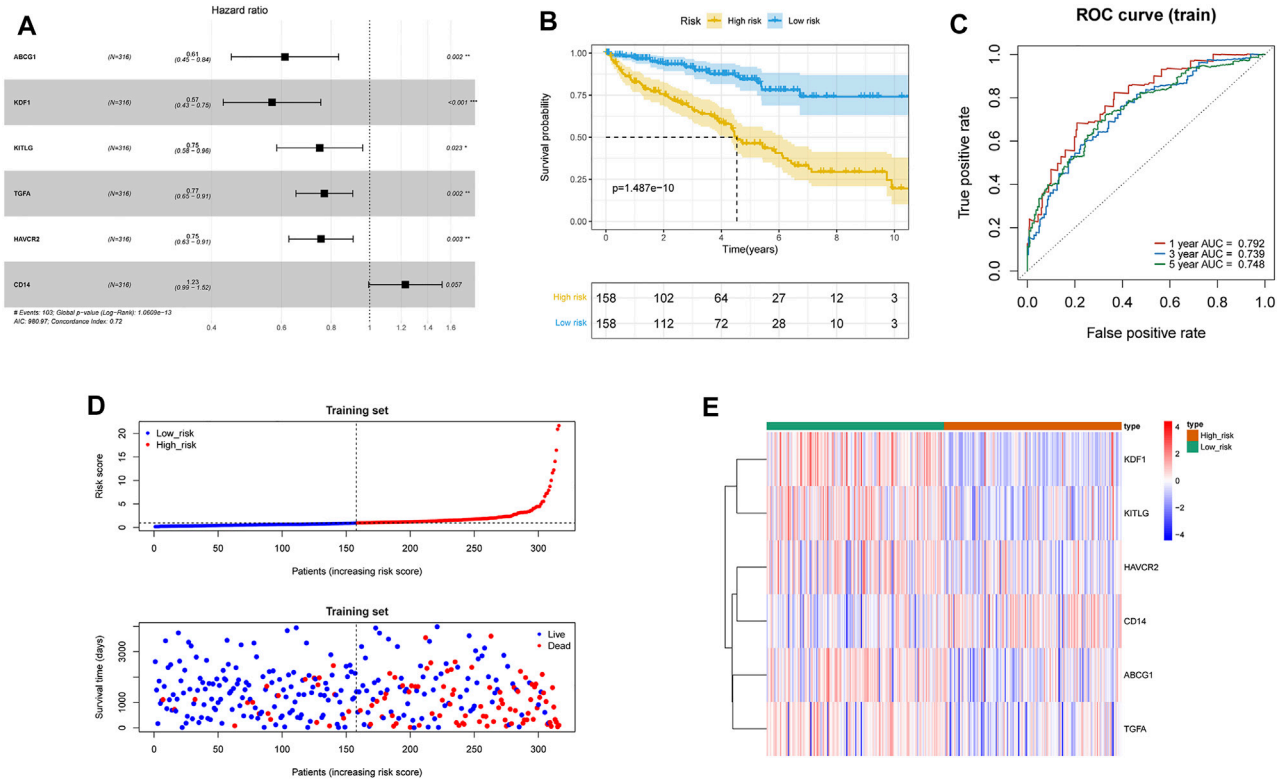


FIGURE 2 | Construction of the macrophage differentiation-related prognostic model in ccRCC. **(A)** Forest plot of MDG Hazard Ratio based on the multivariate Cox regression analysis. **(B)** Survival curves of high- and low-risk groups in the TCGA-ccRCC training set. **(C)** Time-dependent ROC curve of the prognostic model for 1-, 3-, and 5-year overall survival in the training set. **(D)** Distribution of risk score and survival status of the ccRCC patients in the training set, and the dotted line represented the median risk score and divided the patients into low-risk and high-risk groups. **(E)** Heatmap of prognostic gene expression in high- and low-risk groups in the training set.

KITLG was significantly correlated with CD8+T cells, CD4+T cells, macrophages, and neutrophils. TGFA was significantly correlated with B cells, $CD8^+$ T cells, macrophages, neutrophils, and dendritic cells.

Detection and Validation of Prognostic Gene Expression

We tested the mRNA expression differences of six prognostic genes in 10 pairs of ccRCC tissue samples and para-cancerous control tissue samples. The results found that the expression levels of ABCG1, HAVCR2, CD14, and TGFA were significantly increased ($p < 0.05$) in tumor tissue samples compared with para-cancerous control tissue samples, while the expression levels of KDF1 and KITLG were significantly decreased ($p < 0.05$) (**Figure 9A**). The same expression trend was seen in the GSE53757 dataset (**Figure 9B**).

DISCUSSION

Macrophages can be activated and differentiated into macrophage subtypes with different molecular phenotypes and functional characteristics according to different environments, which are regulated by macrophage differentiation–related genes (MDGs). Activated macrophages included the classically activated (M1) subtype macrophages and the selectively activated (M2) subtype macrophages (Ruytinx et al., 2018). It is found that most macrophages in tumor tissues show the characteristics of M2 macrophages, which suggests that there is a special microenvironment that can promote macrophages mainly toward M2 subtype macrophages in tumor tissues (Wang H et al., 2021). Tumor-associated macrophages (TAMs) are the most abundant inflammatory infiltrates in the tumor microenvironment. It is generally believed that TAMs are mainly M2 macrophages. TAMs can promote tumor growth, invasion, metastasis, and drug resistance (Pan et al., 2020). In renal cell carcinoma, TAMs are associated with tumor invasion, poor prognosis, angiogenesis, and immune escape (Kovaleva et al., 2016; Fu et al., 2019); therefore, MDGs as regulators of macrophage differentiation are probably used as biomarkers of ccRCC and potential prognostic indicators of malignant tumors. Herein, to identify the prognostic value of MDGs associated with ccRCC, first, we identified that MDG sets were significantly enriched in ccRCC tissues compared with paired normal tissues, which is consistent with the previous research results (Gottfried et al., 2003). Subsequently, we identified and validated a prognostic model consisting of six gene combinations for ccRCC. Furthermore, risk scores can strongly predict the ccRCC patient outcomes.

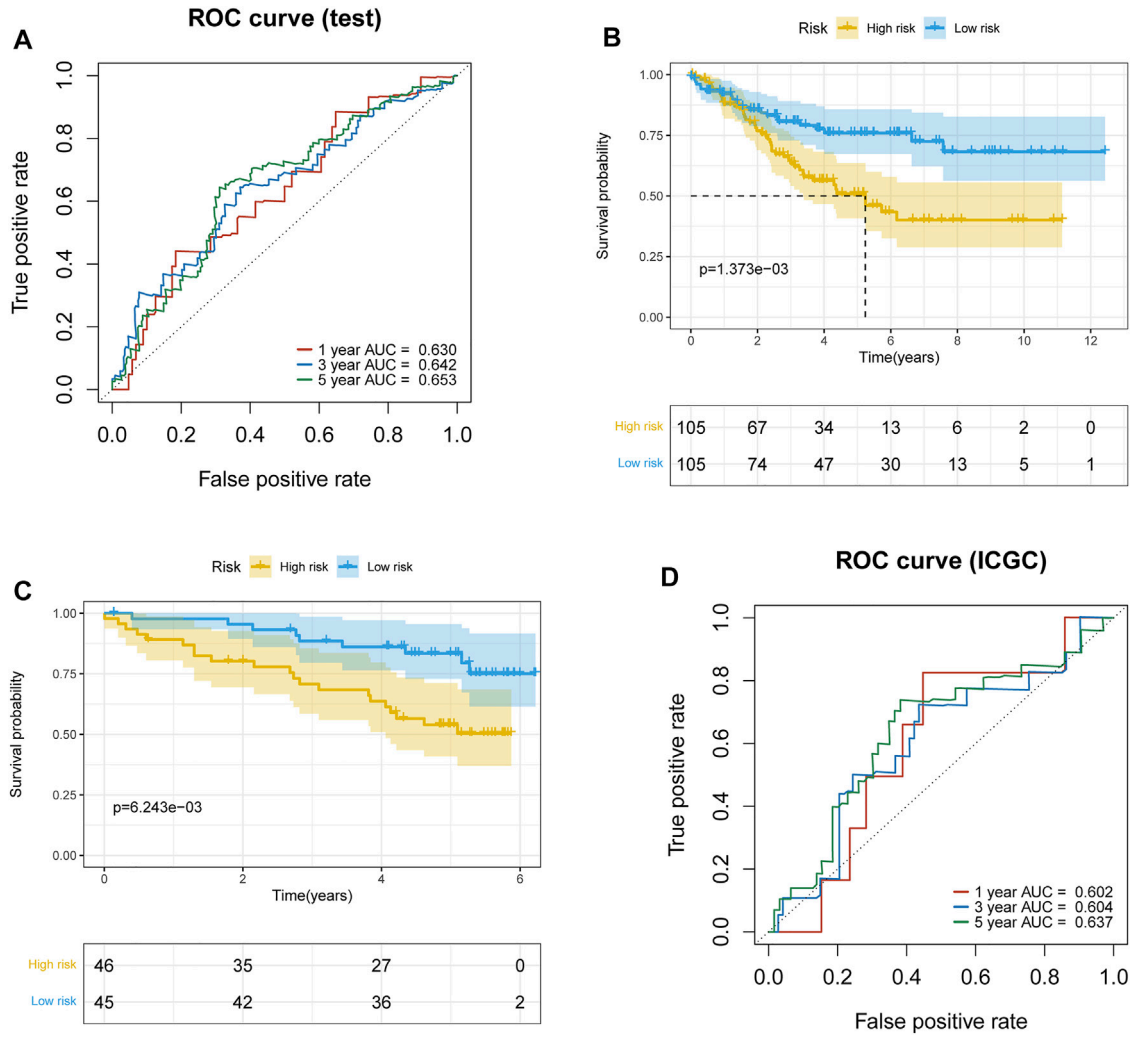


FIGURE 3 | Validation of the prognostic model. **(A)** Time-dependent ROC curve of the prognostic model for 1-, 3-, and 5-year overall survival in the TCGA–ccRCC test set. **(B)** Survival curves of high- and low-risk groups in the TCGA–ccRCC test set. **(C)** Survival curves of high- and low-risk groups in the ICGC test set. **(D)** Time-dependent ROC curve of the prognostic model for 1-, 3-, and 5-year overall survival in the ICGC test set.

The six MDGs (ABCG1, HAVCR2, CD14, TGFA, KDF1, and KITLG) we identified in this study were significantly associated with ccRCC patient outcomes and therefore may play an important role in disease progression. All these genes have been previously reported to be involved in cancers. ABCG1 (ATP-binding cassette G1) belongs to the ATP-binding cassette (ABC) transporter family, which is involved in lipid balance and cholesterol efflux from macrophages (Vasiliou et al., 2009). It was found to be a potential biomarker for lung cancer (Tian et al., 2017), head and neck squamous cell carcinoma (Gonzalez et al., 2003), prostate cancer (Demidenko et al., 2015), and ccRCC (Meng et al., 2021). HAVCR2, also known as TIM-3 (T-cell immunoglobulin domain and mucin domain-3), is involved in the pathogenesis of malignant tumors and the progression of various types of cancer (Pu et al., 2018). Overexpression of TIM-3/HAVCR2 is associated with poor prognosis in squamous cell carcinoma, colorectal, renal cell carcinoma, gastric, and breast cancers (Shen et al., 2016; Wang et al., 2016; Wang et al., 2017; Stenzel, 2018). CD14, a kind of glycoprotein located on the cell surface, was derived mostly from monocytes, macrophages, and neutrophil granulocytes, which mainly acts as a receptor of lipopolysaccharides (LPSs) (Wright et al., 1990). Studies have shown that the CD14 gene is related to macrophage differentiation and regulates immune cell activation (Stenzel, 2018) and is closely related to tumor cell proliferation and the tumor microenvironment in cancers (Wang et al., 2014; Cheah et al., 2015; Guan et al., 2020). TGFA, (transforming growth factor A) as a member of the epidermal growth factor family, is encoded by the TGFA gene. A previous study showed that TGFA is involved in the differentiation of tumor-associated macrophages (TAMs) (Gupta et al., 2018). TGFA is highly expressed in tumor tissues such as

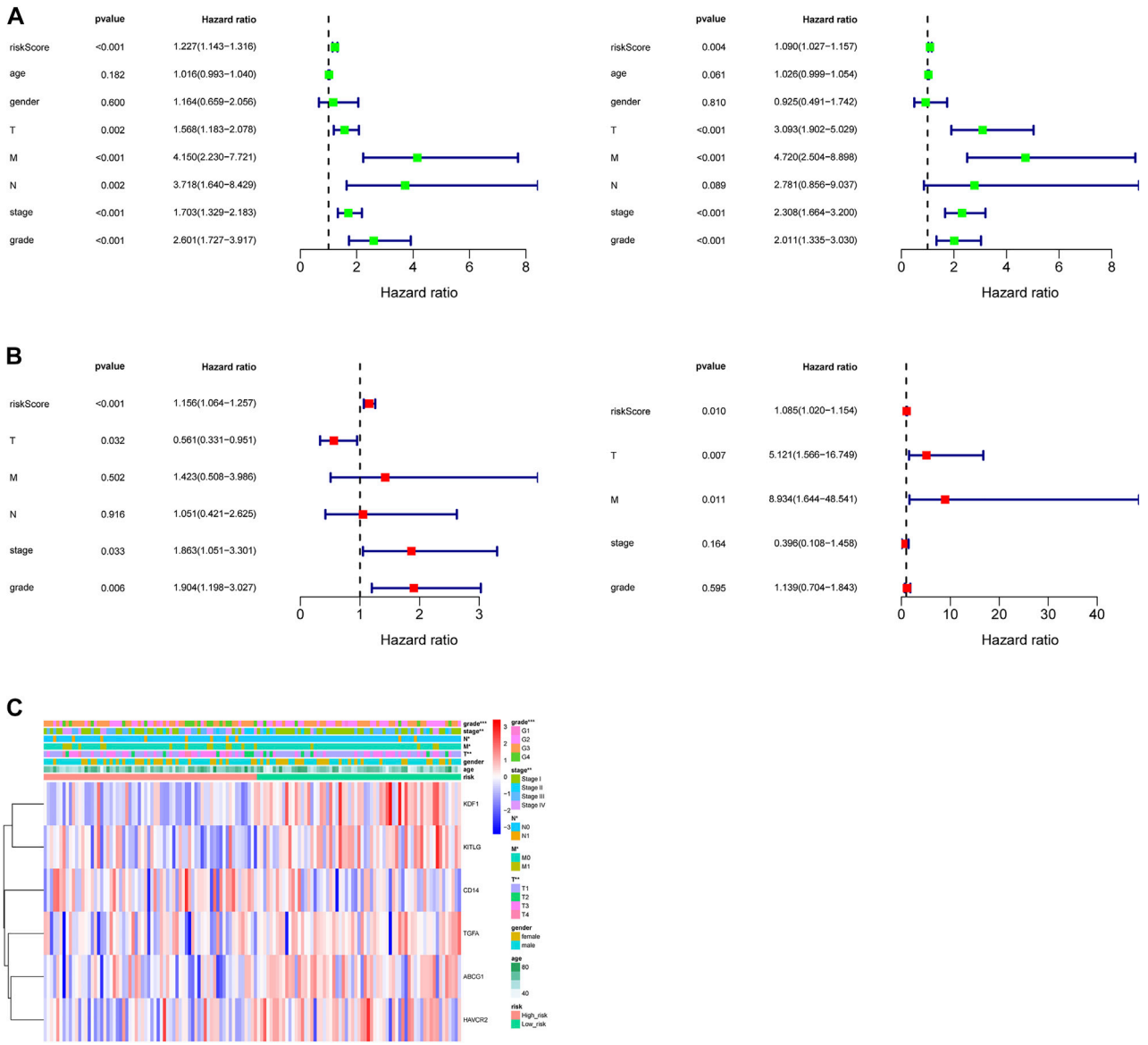


FIGURE 4 | Independent prognostic value of the prognostic model. **(A)** Forest plot of associations between risk factors and the survival of ccRCC based on univariate Cox regression analysis in the training set (left) and test set (right). **(B)** Risk score was an independent predictor of ccRCC based on the multivariate Cox regression analysis in the training set (left) and test set (right). **(C)** Heatmap of the expression profiles of the six prognostic MDGs in different clinicopathological characteristics in the training set.

bladder cancer, oral cancer, pancreatic cancer, and small cell lung cancer, accelerating cell proliferation, invasion, and EMT (epithelial–mesenchymal transition), and is closely related to the prognosis of tumor patients (Kang et al., 2005; Liu et al., 2017; Hao et al., 2018; Zhang et al., 2021). KDF1 (Keratinocyte Differentiation Factor 1) was first reported as a key role in the development of normal epidermis by regulating the proliferation and differentiation of keratinocytes (Lee et al., 2013). Research showed that KDF1 was decreasingly expressed in the cancer cells and correlated negatively with the tumor grade and positively with the survival of the patients, which means KDF1 may function as a tumor suppressor (Zheng et al., 2021). Our study showed that the expression of KDF1 decreased in ccRCC. KITLG is also known as stem cell factor, steel factor, and mast cell growth factor, whose aberrant expression has been implicated in the development of several cancers (Yang et al., 2014; Yang et al., 2015). Our study suggests a decreased expression of KITLG in ccRCC which means KITLG may be a protective gene for ccRCC.

In order to clarify the potential rational mechanism of this prognostic model, we performed GSEA analysis to identify the enriched biological process and pathway in the high-risk

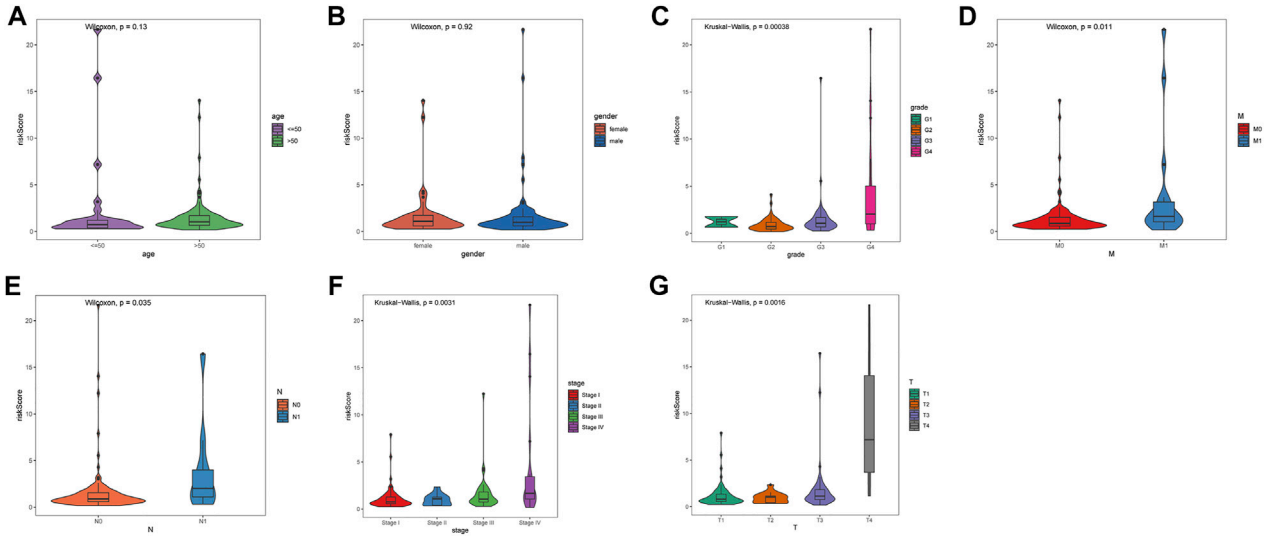


FIGURE 5 | Differences in risk score among seven clinicopathological factors. **(A)** age. **(B)** gender. **(C)** grade. **(D)** pathological M. **(E)** pathological N. **(F)** stage. **(G)** pathological T. Statistical test: Wilcoxon (in two groups) and Kruskal–Wallis (in three or more groups).

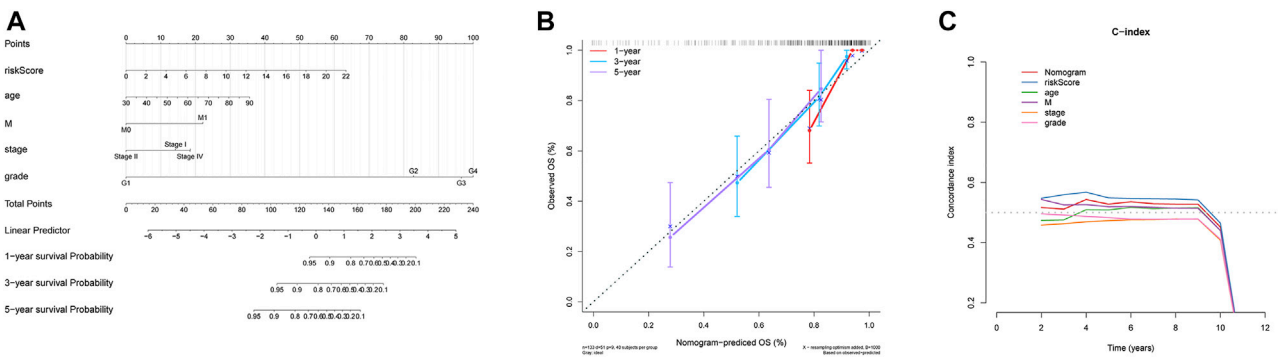


FIGURE 6 | Nomogram can predict the prognosis probability in ccRCC. **(A)** Nomogram to predict the OS possibility of 1 year, 3 year, and 5 year. **(B)** Correction curve of the nomogram. **(C)** Concordance index (C-index) of the nomogram.

cohort compared with the low-risk cohort. Interestingly, the analysis shows that the high-risk group was mainly enriched in the systemic lupus erythematosus. Systemic lupus erythematosus and ccRCC seem to be diseases of distinct patterns. However, there are hidden connections between cancer and systemic lupus erythematosus, which involve multiple of metabolic routes. A previous study showed that systemic lupus erythematosus is highly associated with increased cancer risk compared with the general population (Song et al., 2018). The underlying pathophysiologic mechanisms are still not fully understood, but possible factors include lupus-related medications, inherent immune system abnormalities, viral infections, and/or traditional cancer risk factors (Ladouceur et al., 2020). Other pathways which significantly enriched in a high-risk cohort include the RIG l (retinoic acid inducible-gene 1)-like receptor signaling pathway, the NOD (nucleotide-binding oligomerization domain)-like receptor signaling pathway, the cytoplasmic DNA sensing pathway, antigen processing and presentation, primary immunodeficiency, etc. RIG 1 is the pattern recognition receptor in the cytoplasm that can activate Type I interferon to trigger an antiviral immune response (Ranoa et al., 2016), and Type I IFNs are involved in cancer immunosurveillance and anticancer immune responses widely (Zhang F et al., 2020). NOD-like receptors represent a class of widespread, sophisticated signaling regulators and have been established as crucial regulators in inflammation-associated tumorigenesis, angiogenesis, cancer cell stemness, and chemoresistance (Liu et al., 2019). The cytoplasmic DNA sensing pathway plays an important role in antitumor immunity as well as cancer progression, genomic instability, and the tumor microenvironment (Kwon and Bakhoum, 2020). These aforementioned signaling pathways, including antigen processing and presentation, are all closely related to

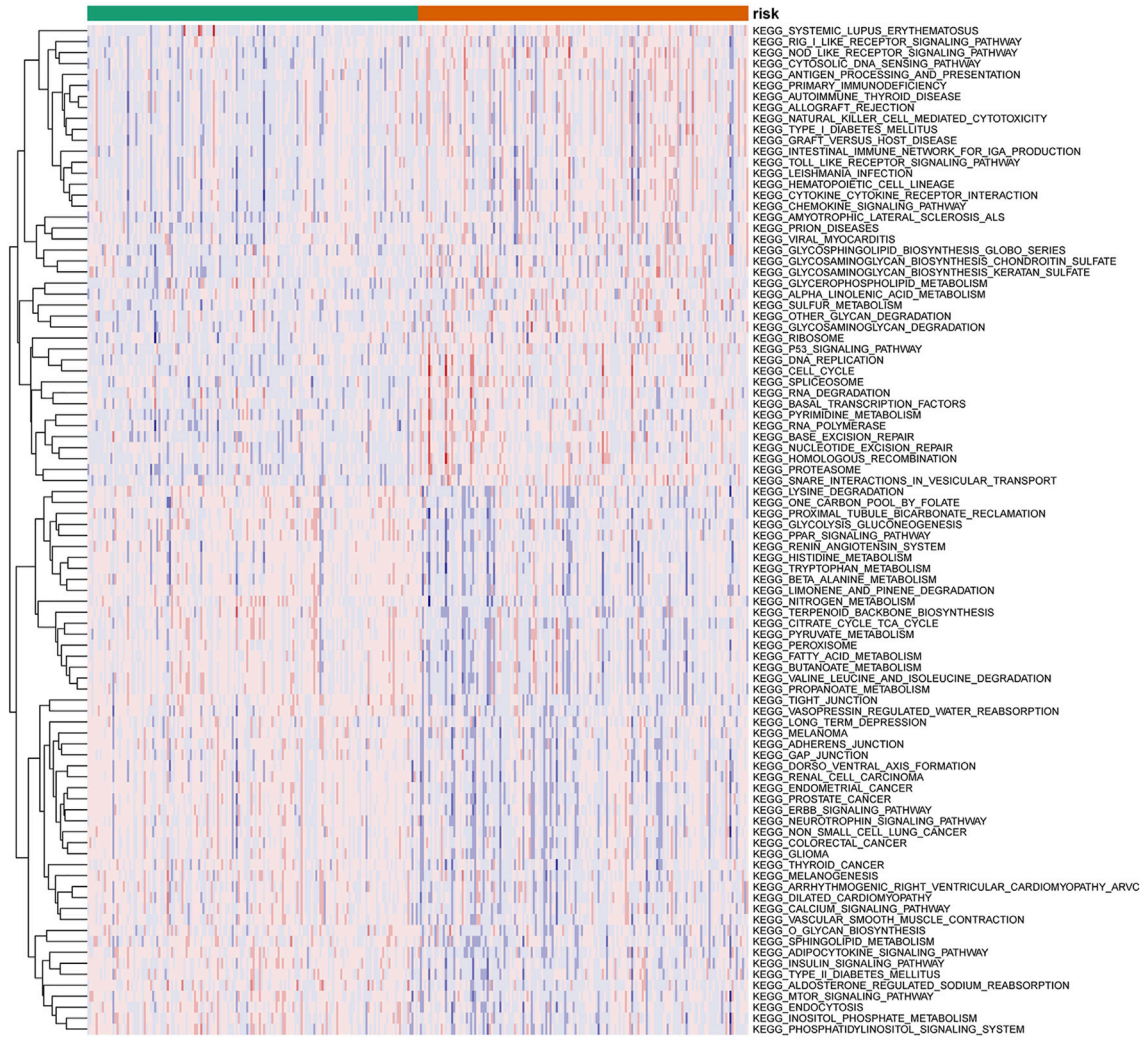


FIGURE 7 | Heatmap of the KEGG pathway in high- and low-risk groups (top 50).

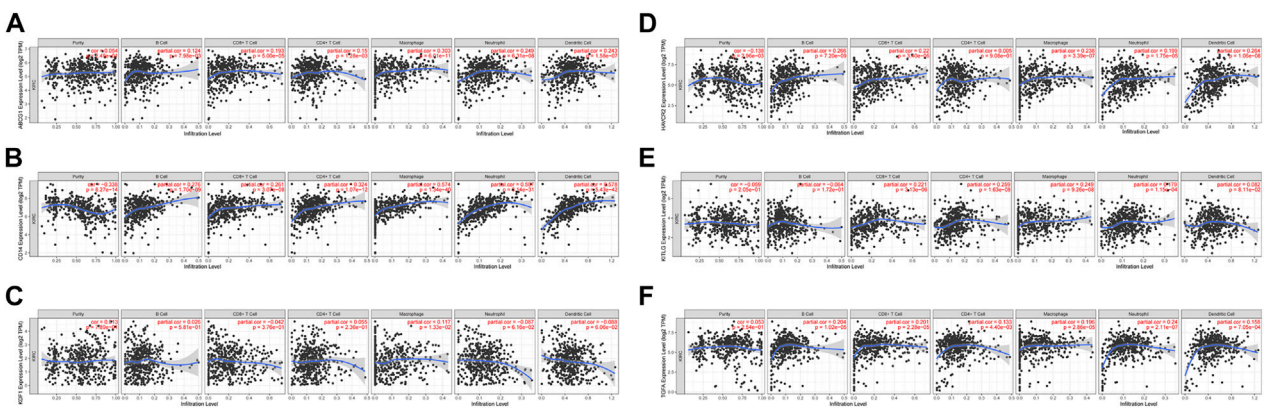


FIGURE 8 | Correlation analysis between the six prognostic genes and infiltration abundances of six types of immune cells (tumor purity, B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells). **(A)** ABCG1. **(B)** CD14. **(C)** KDF1. **(D)** HAVCR2. **(E)** KITLG. **(F)** TGFA.

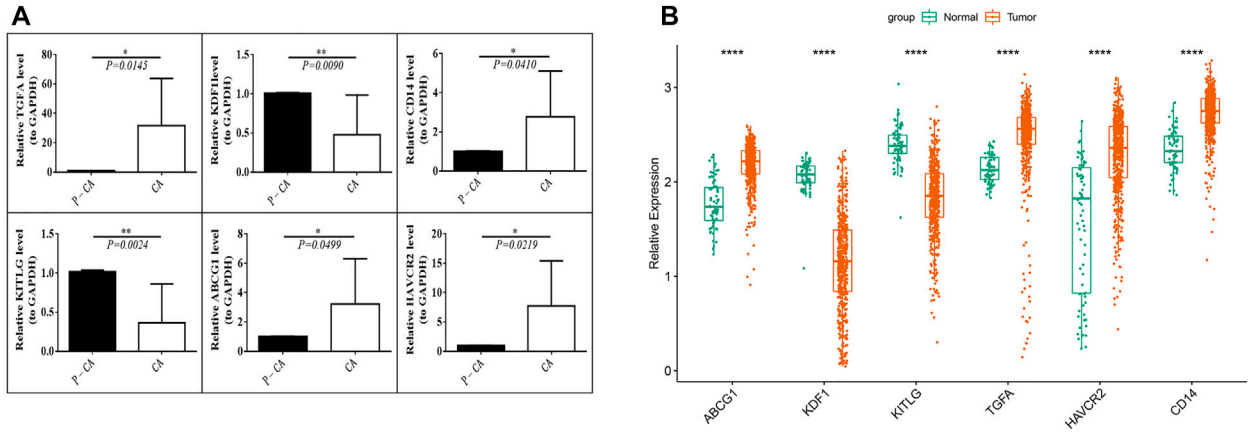


FIGURE 9 | Verification of the prognostic MDG expression level. **(A)** mRNA expression levels of prognostic genes in ccRCC tissue samples (CA) and para-cancerous control tissue samples (P-CA). *$p < 0.05$, **$p < 0.01$. **(B)** mRNA expression levels of the six prognostic MDGs in the GSE53757 data set.

tumor immunity, which means the immune escape of tumor cells and the change of tumor immune microenvironment are important mechanisms of ccRCC. The low-risk group was more associated with lysine degradation, proximal tubule bicarbonate recovery, glycolytic gluconeogenesis, PPAR signal pathway, renin–angiotensin system, etc. Therefore, we speculate that the competition between carcinogenic factors and body resistibility results in disturbed metabolic microenvironments such as amino acids and glucose metabolism and hydroelectrolyte balance. The low-risk group might benefit more from metabolism-related treatment than the high-risk group. These hypotheses need further investigation.

Immune infiltration of tumors is closely associated with clinical outcomes in renal cell carcinoma, tumor-infiltrating immune cells (TIICs) have independent prognostic values in RCC (Zhang et al., 2019). In order to further understand the relationship between these six prognostic MDGs and tumor immunity of ccRCC, we made a correlation analysis between these prognostic MDGs and tumor immune infiltrating cells. The results showed a significant correlation between these six prognostic MDGs and the immune infiltrating cells including $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells. At present, PD-1/PD-L1 immune inhibitor is the first-line treatment of metastatic ccRCC. Despite the encouraging activity and tolerable toxicity of PD-1/PD-L1, there is the additional risk of immune-mediated side effects (Grimm et al., 2020), and its diagnosis and prognosis still lack reliable biomarkers. Therefore, many studies were devoted to finding biomarkers that can predict the response of these therapies. Several studies have found some prognostic-related biomarkers such as CD8 (+) T lymphocytes (Zhang et al., 2019; Lin J et al., 2020) and tumor-infiltrating CD19 (+) B lymphocytes (Lin et al., 2018); however, a single biomarker is not enough to identify the efficacy of immunotherapy accurately. The prognostic model we established by the expression of the six MDGs which showed high accuracy in predicting the 1-, 3-, and 5-year survival rate of ccRCC patients suggested that the prognostic model may be helpful in developing biomarkers and performing prognostic evaluation of metastatic clear cell renal cell carcinoma.

Moreover, we established a nomogram for clinical-decision support. Owing to their intuitive visual presentation and personalized application, nomograms have become a popular tool for oncology prognosis these years. Zhang Z et al. (2020) reported a nomogram based on five deferentially expressed genes to predict 1-, 3- and 5-year OS in ccRCC patients. Peng et al. (2021) constructed a nomogram based on the dysregulated ceRNA network and hope to find a better understanding of the ccRCC tumorigenesis mechanism. The nomogram we established incorporates age, tumor staging, grading, tumor metastasis, and risk score based on clinicopathological factors. By calculating the score of each risk factor, the numerical outcomes are finally generated for the individual patients, and the 1-, 3-, and 5-year possibility survival rate can be predicted. The nomogram correction curve and the concordance index analysis both showed excellent accuracy in our nomogram, which means our nomogram could potentially be used in clinical practice for predicting the individual survival rate and promoting the selection of individual treatment options for ccRCC patients.

However, there are still some limitations to our study. First, our study only focused on the macrophage differentiation–associated genes from TCGA and ICGC platforms. Potential prognostic factors, such as environmental factors, genetic factors, and personal history, are missing in our nomogram. Second, we only tested 10 pairs of ccRCC tissue samples and para-cancerous control tissue samples, and more large-scale molecular experiments are needed to further clarify the mRNA expression of six prognostic genes. Third, our study provides the evidence that the six genes could be potential biomarkers in ccRCC and possibly become therapeutic targets for precision medicine in the future,

and the exact mechanism of the six genes has not been explored in our study. Therefore, the further functional experiment of the six genes in ccRCC based on larger sample size is still valuable and crucial.

CONCLUSION

We developed and validated a six-gene prognostic risk model (ABCG1, HAVCR2, CD14, TGFA, KDF1, and KITLG) associated with macrophage differentiation for ccRCC patients. The higher the risk score, the worse the prognosis. Also, the established nomogram can be used as a novel tool for predicting the clinical outcome of ccRCC patients. However, the model still needs to be validated in prospective clinical trials with large sample sizes in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: https://portal.gdc.cancer.gov/https://dcc.icgc.org/https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53757.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Hospital of ShanXi Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The study was designed by DW. The manuscript was written by CL. The bioinformatics analyses were designed and carried out by XL and XC. The experiment and data analysis were performed by XZ and CH. All authors participated in the reading and critical revision of the final manuscript.

FUNDING

This study was supported by Sanming Project of Medicine in Shenzhen (No. SZSM202111003).

REFERENCES

Cao, J., Zhang, C., Jiang, G.-Q., Jin, S.-J., Wang, Q., Wang, A.-Q., et al. (2021). Identification of Hepatocellular Carcinoma-Related Genes Associated with Macrophage Differentiation Based on Bioinformatics Analyses. *Bioengineered* 12 (1), 296–309. doi:10.1080/21655979.2020.1868119

Cheah, M. T., Chen, J. Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A. K., Scheeren, F. A., et al. (2015). CD14-expressing Cancer Cells Establish the Inflammatory and Proliferative Tumor Microenvironment in Bladder Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 112 (15), 4725–4730. doi:10.1073/pnas.1424795112

Chen, Y., Zhang, S., Wang, Q., and Zhang, X. (2017). Tumor-recruited M2 Macrophages Promote Gastric and Breast Cancer Metastasis *via* M2 Macrophage-Secreted CHI3L1 Protein. *J. Hematol. Oncol.* 10 (1), 36. doi:10.1186/s13045-017-0408-0

Cohen, H. T., and McGovern, F. J. (2005). Renal-cell Carcinoma. *N. Engl. J. Med.* 353 (23), 2477–2490. doi:10.1056/nejmra043172

Demidenko, R., Razanauskas, D., Daniunaite, K., Lazutka, J. R., Jankevicius, F., and Jarmalaite, S. (2015). Frequent Down-Regulation of ABC Transporter Genes in Prostate Cancer. *BMC Cancer* 15, 683. doi:10.1186/s12885-015-1689-8

Fu, Q., Xu, L., Wang, Y., Jiang, Q., Liu, Z., Zhang, J., et al. (2019). Tumor-associated Macrophage-Derived Interleukin-23 Interlinks Kidney Cancer Glutamine Addiction with Immune Evasion. *Eur. Urol.* 75 (5), 752–763. doi:10.1016/j.eururo.2018.09.030

Gonzalez, H. E., Gujrati, M., Frederick, M., Henderson, Y., Arumugam, J., Spring, P. W., et al. (2003). Identification of 9 Genes Differentially Expressed in Head and Neck Squamous Cell Carcinoma. *Arch. Otolaryngol. Head. Neck Surg.* 129 (7), 754–759. doi:10.1001/archotol.129.7.754

Gottfried, E., Faust, S., Fritsche, J., Kunz-Schughart, L. A., Andreesen, R., Miyake, K., et al. (2003). Identification of Genes Expressed in Tumor-Associated Macrophages. *Immunobiology* 207 (5), 351–359. doi:10.1078/0171-2985-00246

Grimm, M. O., Leucht, K., Grünwald, V., and Foller, S. (2020). New First Line Treatment Options of Clear Cell Renal Cell Cancer Patients with PD-1 or PD-L1 Immune-Checkpoint Inhibitor-Based Combination Therapies. *J. Clin. Med.* 9 (2), 565. doi:10.3390/jcm9020565

Guan, Y., Huang, X. F., Li, P. J., Cao, W., Gao, X. H., and Guan, X. (2020). Association of CD14 Gene -260C>T and -561C>T Polymorphisms with Cancer Susceptibility: A Meta-Analysis. *J. Gene Med.* 22 (2), e3151. doi:10.1002/jgm.3151

Gupta, S., Jain, A., Syed, S. N., Snodgrass, R. G., Pflüger-Müller, B., Leisegang, M. S., et al. (2018). IL-6 Augments IL-4-induced Polarization of Primary Human Macrophages through Synergy of STAT3, STAT6 and BATF Transcription Factors. *Oncoimmunology* 7 (10), e1494110. doi:10.1080/2162402x.2018.1494110

Hao, C., Li, Z., Zhang, X., Zhang, H., Shang, H., Bao, J., et al. (2018). Expression and Clinical Significance of EGF and TGF-α in Chronic Pancreatitis and Pancreatic Cancer. *Minerva Endocrinol.* 43 (3), 253–258. doi:10.23736/S0391-1977.17.02721-3

Jonasch, E., Gao, J., and Rathmell, W. K. (2014). Renal Cell Carcinoma. *BMJ* 349, g4797. doi:10.1136/bmj.g4797

Kang, D., Gridley, G., Huang, W.-Y., Engel, L. S., Winn, D. M., Brown, L. M., et al. (2005). Microsatellite Polymorphisms in the Epidermal Growth Factor Receptor (EGFR) Gene and the Transforming Growth Factor-alpha (TGFA) Gene and Risk of Oral Cancer in Puerto Rico. *Pharmacogenetics Genomics* 15 (5), 343–347. doi:10.1097/01213011-200505000-00010

Kovaleva, O. V., Samoilova, D. V., Shitova, M. S., and Gratchev, A. (2016). Tumor Associated Macrophages in Kidney Cancer. *Anal. Cell Pathol. (Amst)* 2016, 9307549. doi:10.1155/2016/9307549

Kwon, J., and Bakhoum, S. F. (2020). The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer Discov.* 10 (1), 26–39. doi:10.1158/2159-8290.cd-19-0761

Ladouceur, A., Tessier-Cloutier, B., Clarke, A. E., Ramsey-Goldman, R., Gordon, C., Hansen, J. E., et al. (2020). Cancer and Systemic Lupus Erythematosus. *Rheumatic Dis. Clin. N. Am.* 46 (3), 533–550. doi:10.1016/j.rdc.2020.05.005

Lee, S., Kong, Y., and Weatherbee, S. D. (2013). Forward Genetics Identifies Kdf1/1810019J16Rik as an Essential Regulator of the Proliferation-Differentiation Decision in Epidermal Progenitor Cells. *Dev. Biol.* 383 (2), 201–213. doi:10.1016/j.ydbio.2013.09.022

Liang, T., Sang, S., Shao, Q., Chen, C., Deng, Z., Wang, T., et al. (2020). Abnormal Expression and Prognostic Significance of EPB41L1 in Kidney Renal Clear Cell Carcinoma Based on Data Mining. *Cancer Cell Int.* 20, 356. doi:10.1186/s12935-020-01449-8

Lin, J., Yu, M., Xu, X., Wang, Y., Xing, H., An, J., et al. (2020). Identification of Biomarkers Related to CD8(+) T Cell Infiltration with Gene Co-expression Network in Clear Cell Renal Cell Carcinoma. *Aging* 12 (4), 3694–3712. doi:10.18632/aging.102841

Lin, W., Chen, X., Chen, T., Liu, J., Ye, Y., Chen, L., et al. (2020). C1QTNF6 as a Novel Diagnostic and Prognostic Biomarker for Clear Cell Renal Cell Carcinoma. *DNA Cell Biol.* 39 (6), 1000–1011. doi:10.1089/dna.2019.5299

Lin, Z., Liu, L., Xia, Y., Chen, X., Xiong, Y., Qu, Y., et al. (2018). Tumor Infiltrating CD19(+) B Lymphocytes Predict Prognostic and Therapeutic Benefits in Metastatic Renal Cell Carcinoma Patients Treated with Tyrosine Kinase Inhibitors. *Oncoimmunology* 7 (10), e1477461. doi:10.1080/2162402X.2018.1477461

Liu, P., Lu, Z., Liu, L., Li, R., Liang, Z., Shen, M., et al. (2019). NOD-like Receptor Signaling in Inflammation-Associated Cancers: From Functions to Targeted Therapies. *Phytomedicine* 64, 152925. doi:10.1016/j.phymed.2019.152925

Liu, X., Chen, L., Tian, X. D., and Zhang, T. (2017). MiR-137 and its Target TGFA Modulate Cell Growth and Tumorigenesis of Non-small Cell Lung Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 21 (3), 511–517.

Meng, F., Xiao, Y., Xie, L., Liu, Q., and Qian, K. (2021). Diagnostic and Prognostic Value of ABC Transporter Family Member ABCG1 Gene in Clear Cell Renal Cell Carcinoma. *Channels* 15 (1), 375–385. doi:10.1080/19336950.2021.1909301

Pan, Y., Yu, Y., Wang, X., and Zhang, T. (2020). Tumor-Associated Macrophages in Tumor Immunity. *Front. Immunol.* 11, 583084. doi:10.3389/fimmu.2020.583084

Peng, Y., Wu, S., Xu, Z., Hou, D., Li, N., Zhang, Z., et al. (2021). A Prognostic Nomogram Based on Competing Endogenous RNA Network for Clear-cell Renal Cell Carcinoma. *Cancer Med.* 10 (16), 5499–5512. doi:10.1002/cam4.4109

Pritchard, A., Tousif, S., Wang, Y., Hough, K., Khan, S., Strenkowski, J., et al. (2020). Lung Tumor Cell-Derived Exosomes Promote M2 Macrophage Polarization. *Cells* 9 (5), 1303. doi:10.3390/cells9051303

Pu, F., Chen, F., Zhang, Z., Feng, J., and Xia, P. (2018). Functional Variants of TIM-3/HAVCR2 3′UTR in Lymphoblastoid Cell Lines. *Future Sci. OA* 4 (5), FSO298. doi:10.4155/fsoa-2017-0121

Ranoa, D. R. E., Parekh, A. D., Pitroda, S. P., Huang, X., Darga, T., Wong, A. C., et al. (2016). Cancer Therapies Activate RIG-I-like Receptor Pathway through Endogenous Non-coding RNAs. *Oncotarget* 7 (18), 26496–26515. doi:10.18632/oncotarget.8420

Ruytinx, P., Proost, P., Van Damme, J., and Struyf, S. (2018). Chemokine-Induced Macrophage Polarization in Inflammatory Conditions. *Front. Immunol.* 9, 1930. doi:10.3389/fimmu.2018.01930

Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S. A., Mardani, F., et al. (2018). Macrophage Plasticity, Polarization, and Function in Health and Disease. *J. Cell Physiol.* 233 (9), 6425–6440. doi:10.1002/jcp.26429

Shen, P., Yue, R., Tang, J., Si, H., Shen, L., Guo, C., et al. (2016). Preferential Tim-3 Expression on Treg and CD8(+) T Cells, Supported by Tumor-Associated Macrophages, Is Associated with Worse Prognosis in Gastric Cancer. *Am. J. Transl. Res.* 8 (8), 3419–3428.

Sica, A., and Mantovani, A. (2012). Macrophage Plasticity and Polarization: *In Vivo* Veritas. *J. Clin. Invest.* 122 (3), 787–795. doi:10.1172/jci59643

Song, L., Wang, Y., Zhang, J., Song, N., Xu, X., and Lu, Y. (2018). The Risks of Cancer Development in Systemic Lupus Erythematosus (SLE) Patients: a Systematic Review and Meta-Analysis. *Arthritis Res. Ther.* 20 (1), 270. doi:10.1186/s13075-018-1760-3

Stenzel, P. (2018). Prognostische Relevanz von TIM3 bei Nierenzellkarzinomen. *Pathologe* 39 (6), 587–588. doi:10.1007/s00292-018-0462-6

Tian, C., Huang, D., Yu, Y., Zhang, J., Fang, Q., and Xie, C. (2017). ABCG1 as a Potential Oncogene in Lung Cancer. *Exp. Ther. Med.* 13 (6), 3189–3194. doi:10.3892/etm.2017.4393

Vasiliou, V., Vasiliou, K., and Nebert, D. W. (2009). Human ATP-Binding Cassette (ABC) Transporter Family. *Hum. Genomics* 3 (3), 281–290. doi:10.1186/1479-7364-3-3-281

Wang, H., Yung, M. M. H., Ngan, H. Y. S., Chan, K. K. L., and Chan, D. W. (2021). The Impact of the Tumor Microenvironment on Macrophage Polarization in Cancer Metastatic Progression. *Int. J. Mol. Sci.* 22 (12), 6560. doi:10.3390/ijms22126560

Wang, J., Guo, X., Yu, S., Song, J., Zhang, J., Cao, Z., et al. (2014). Association between CD14 Gene Polymorphisms and Cancer Risk: a Meta-Analysis. *PLoS One* 9 (6), e100122. doi:10.1371/journal.pone.0100122

Wang, Q., Zhang, H., Chen, Q., Wan, Z., Gao, X., and Qian, W. (2019). Identification of METTL14 in Kidney Renal Clear Cell Carcinoma Using Bioinformatics Analysis. *Dis. Markers* 2019, 5648783. doi:10.1155/2019/5648783

Wang, Y., Sun, J., Gao, W., Song, B., Shao, Q., Zhao, L., et al. (2017). Preoperative Tim-3 Expression on Peripheral NK Cells Is Correlated with Pathologic TNM Staging in Colorectal Cancer. *Mol. Med. Rep.* 15 (6), 3810–3818. doi:10.3892/mmr.2017.6482

Wang, Y., Yan, K., Lin, J., Li, J., and Bi, J. (2021). Macrophage M2 Co-expression Factors Correlate with the Immune Microenvironment and Predict Outcome of Renal Clear Cell Carcinoma. *Front. Genet.* 12, 615655. doi:10.3389/fgene.2021.615655

Wang, Z., Liu, X., Wang, X., Chong, T., Lin, S., Wang, M., et al. (2016). Polymorphisms in TIM-3 and Breast Cancer Susceptibility in Chinese Women: A Case-Control Study. *Oncotarget* 7 (28), 43703–43712. doi:10.18632/oncotarget.9665

Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J., and Mathison, J. C. (1990). CD14, a Receptor for Complexes of Lipopolysaccharide (LPS) and LPS Binding Protein. *Science* 249 (4975), 1431–1433. doi:10.1126/science.1698311

Yang, S., Li, W. s., Dong, F., Sun, H. m., Wu, B., Tan, J., et al. (2014). KITLG Is a Novel Target of miR-34c that Is Associated with the Inhibition of Growth and Invasion in Colorectal Cancer Cells. *J. Cell. Mol. Med.* 18 (10), 2092–2102. doi:10.1111/jcmm.12368

Yang, S., Li, W., Sun, H., Wu, B., Ji, F., Sun, T., et al. (2015). Resveratrol Elicits Anti-colorectal Cancer Effect by Activating miR-34c-KITLG *In Vitro* and *In Vivo*. *BMC Cancer* 15, 969. doi:10.1186/s12885-015-1958-6

Yuan, H., Lin, Z., Liu, Y., Jiang, Y., Liu, K., Tu, M., et al. (2020). Intrahepatic Cholangiocarcinoma Induced M2-Polarized Tumor-Associated Macrophages Facilitate Tumor Growth and Invasiveness. *Cancer Cell Int.* 20 (1), 586. doi:10.1186/s12935-020-01687-w

Zhang, F., Manna, S., Pop, L. M., Chen, Z. J., Fu, Y.-X., and Hannan, R. (2020). Type I Interferon Response in Radiation-Induced Anti-tumor Immunity. *Seminars Radiat. Oncol.* 30 (2), 129–138. doi:10.1016/j.semradonc.2019.12.009

Zhang, S., Zhang, E., Long, J., Hu, Z., Peng, J., Liu, L., et al. (2019). Immune Infiltration in Renal Cell Carcinoma. *Cancer Sci.* 110 (5), 1564–1572. doi:10.1111/cas.13996

Zhang, X., Li, X., Fu, X., Yu, M., Qin, G., Chen, G., et al. (2021). Circular RNA TAF4B Promotes Bladder Cancer Progression by Sponging miR-1298-5p and Regulating TGFA Expression. *Front. Oncol.* 11, 643362. doi:10.3389/fonc.2021.643362

Zhang, Z., Lin, E., Zhuang, H., Xie, L., Feng, X., Liu, J., et al. (2020). Construction of a Novel Gene-Based Model for Prognosis Prediction of Clear Cell Renal Cell Carcinoma. *Cancer Cell Int.* 20, 27. doi:10.1186/s12935-020-1113-6

Zhao, Y., Yan, Y., Ma, R., Lv, X., Zhang, L., Wang, J., et al. (2020). Expression Signature of six-snoRNA Serves as Novel Non-invasive Biomarker for Diagnosis and Prognosis Prediction of Renal Clear Cell Carcinoma. *J. Cell Mol. Med.* 24 (3), 2215–2228. doi:10.1111/jcmm.14886

Zheng, J.-m., Gan, M.-f., Yu, H.-y., Ye, L.-x., Yu, Q.-x., Xia, Y.-h., et al. (2021). KDF1, a Novel Tumor Suppressor in Clear Cell Renal Cell Carcinoma. *Front. Oncol.* 11, 686678. doi:10.3389/fonc.2021.686678

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Zhang, Hu, Liang, Cao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fgene.2022.882307

High Expression of FCRLB Predicts Poor Prognosis in Patients With Colorectal Cancer

Xiaopeng Wang†, Ruirong Lin†, Yi Zeng†, Yi Wang, Shenghong Wei, Zhitao Lin, Shu Chen, Zaisheng Ye and Luchuan Chen**

Department of Gastrointestinal Surgical Oncology, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Xiaotian Chang,
The Affiliated Hospital of Qingdao University, China
Suprabhat Mukherjee,
Kazi Nazrul University, India

***Correspondence:**
Zaisheng Ye
614043465@qq.com
Luchuan Chen
luchuanchen@sina.cn

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *23 February 2022*
Accepted: *12 May 2022*
Published: *16 June 2022*

Citation:
Wang X, Lin R, Zeng Y, Wang Y, Wei S, Lin Z, Chen S, Ye Z and Chen L (2022) High Expression of FCRLB Predicts Poor Prognosis in Patients With Colorectal Cancer. Front. Genet. 13:882307. doi: 10.3389/fgene.2022.882307

Background: Mining the prognostic biomarkers of colorectal cancer (CRC) has important clinical and scientific significance. The role of Fc receptor-like B (FCRLB) in solid tumors has never been reported or studied to our knowledge, and the prognostic role of FCRLB in CRC still awaits characterization.

Methods: The potential prognostic factor FCRLB was screened out through TCGA database analysis. Then, its expression and associations with clinicopathological variables were assessed in the TCGA CRC cohort. The prognostic value of FCRLB was examined with multiple methods, such as the Kaplan-Meier method, ROC curve, time-dependent ROC analysis, and prediction model nomograms. Then, functional enrichment and annotation among the high and low FCRLB groups were achieved utilizing GO and KEGG analyses and GSEA. Fresh CRC tissue samples obtained clinically were used for the preparation of the tissue microarray and for further validation.

Results: FCRLB was highly expressed in CRC tissues compared to normal tissues. Moreover, over-expression of FCRLB correlated with higher CEA levels, advanced T stage, N stage, M stage, AJCC stage, lymphatic invasion, perineural invasion, and incomplete resection (R1 and R2 resection). In addition, high expression of FCRLB was closely correlated to less favorable OS, DSS, and PFI. The analysis of CRC tissue microarray further confirmed the conclusion drawn from the TCGA data analysis.

Conclusion: FCRLB is notably up-regulated in CRC tissues and may serve as a potential biomarker of CRC.

Keywords: FCRLB, colorectal cancer, biomarker, prognosis, TCGA

Abbreviations: FCRL, Fc receptor like; FCRLB, Fc receptor like B; CRC, colorectal cancer; TCGA, The Cancer Genome Atlas; GSEA, gene set enrichment analysis; GO, gene ontology; KEGG, Kyoto Encyclopedia Genes and Genomes; TMA, tissue microarray; IHC, immunohistochemistry; OS, overall survivals, DSS, disease-specific survivals; PFI, progress free interval; CEA, carcinoembryonic antigen; OR, odds ratio; CI, confidence interval; DEGs, differently expressed genes; MAPK, mitogen-activated protein kinases; PI3/AKT, phosphoinositide 3-kinase/AKT; EMT, epithelial-to-mesenchymal transition; FDR, false discovery rate; TME, tumor microenvironment; ssGSEA, single-sample Gene set enrichment analysis; CAFs, cancer-associated fibroblasts; TAMs, tumor-associated macrophages; ROC, receiver operating characteristic; ECM, extracellular matrix; TIICs, tumor-infiltrating immune cells.

INTRODUCTION

Colorectal cancer (CRC) is the third most common type of malignant tumor and the second most frequent leading cause of cancer-related deaths worldwide (Daniel et al., 2017). The mortality rate among patients with CRC has gradually decreased in the past decades, mainly owing to advances in medical technology and treatment. Nevertheless, the prognosis for patients with advanced CRC remains extremely poor, with the 5-year survival rate being less than 14% (Siegel et al., 2018). Thus, determining effective prognostic and predictive markers for early detection and diagnosis of CRC is of great scientific interest and clinical importance.

In recent years, the tumor microenvironment (TME) has gained increasing attention, largely because of its close association with the effectiveness and sensitivity of immunotherapy. Currently, immunotherapy has revolutionized cancer therapy and has rapidly become the mainstay of treatment for a variety of solid cancers, such as malignant melanoma (Komenaka et al., 2004), non-small-cell lung cancer (Nasser et al., 2020), renal cell carcinoma (Deleuze et al., 2020), etc. However, CRC immunotherapy presents enormous challenges to personalized treatment regimens, mainly due to the heterogeneity and dynamics of TME (Ganesh et al., 2019). Tumor-infiltrating immune cells (TIICs) in TME consist of a variety of immune cells, frequently dominated by B cells, as well as other non-tumor cells including T cells, NK cells, and myeloid-lineage cells (e.g., macrophages, mast cells, and neutrophils, etc.). Accumulating evidence indicates that the types and constituents of TIICs in TME not only affect the proliferative and metastatic ability of cancer cells but also affect tumor responses to immunotherapy. For example, it has been reported that tumor-associated macrophages (TAMs) often display characteristics similar to M2 macrophages, which can exert multiple cancer-promoting functions in TME (Cao et al., 2018). Cancer-associated fibroblasts (CAFs), one of the important cellular components of the TME, have been proven to play critical roles in cancer cell proliferation, tumor immunity, extracellular matrix remodeling, and inflammatory responses by secreting cell growth factors, inflammatory factors, and extracellular matrix (Kobayashi et al., 2019). Moreover, clinically, the exhaustion of T cells, the predominant type of immune cell, characterized by the gradual functional deficits of exhausted T cells, has been considered as a cause of the failure of immunotherapy (e.g., chimeric antigen receptor T cell immunotherapy [CAR-T] and immune checkpoint inhibitors) Terranova-Barberio et al., 2020. Accordingly, it is crucial to conduct a comprehensive analysis of tumor-infiltrating immune cell components in TME.

In the present study, we perform bioinformatic analysis of the ribonucleic acid (RNA)-sequencing dataset obtained from The Cancer Genome Atlas (TCGA) project and the Gene Expression Omnibus (GEO) database. Fc receptor-like B (FCRLB) was then screened as a potential prognostic factor for CRC. Furthermore, multifaceted analyses of FCRLB through the TCGA database, such as differential expression, survival analysis, immune infiltration, and potentially related biological pathways, were performed to clarify the clinical utility and prognostic value of FCRLB in CRC and investigate its potential biological mechanisms. Moreover, we use clinically obtained specimens and tissue microarrays (TMA) for further verification of the conclusion drawn from the analysis of the TCGA CRC cohort.

METHODS

Experimental Design and Data Acquisition

The specific experimental process was shown in the flowchart (**Figure 1**). Level 3 RNA-seq V2 data sets and matched clinicopathological information were downloaded from the TCGA CRC cohort (https://portal.gdc.cancer.gov/; 15 December 2018). Gene expression data of GEO GSE331133 was downloaded. Then the transcriptome data were normalized by using the DESeq package as previous literature described. The limma R package and survival R package were used for screening differentially expressed genes (DEGs) and prognosis-correlated genes, respectively.

Patients Specimens and Tissue Microarray

Fresh cancerous tissues and matched adjacent non-tumorous tissues were obtained from each CRC patient who underwent surgery in Fujian Cancer Hospital. The normal colorectal tissue, at a distance of 2 cm from the tumor tissue, was excised as the paracancerous tissue specimen. Then, fresh cancerous tissues and matched adjacent non-tumorous tissues were confirmed by histopathological examination, and then were fixed using 10% formalin overnight, and then embedded in paraffin. Tissue samples were prepared in a tissue microarray (TMA), which contained 40 pairs of CRC tissues and matched adjacent normal tissues, as previously described (Xu et al., 2021). TMA was then used for histopathological validation by performing immunohistochemistry (IHC).

Gene Set Enrichment Analysis

GSEA of microarray data was carried out by using the GSEA v.30 desktop software (Broad Institute) to explore the potential biological mechanisms of FCRLB. If they reached the preset threshold of $p < 0.05$, the biological processes were considered to be apparently enriched.

GO and KEGG Enrichment Analyses

R package was used for GO and KEGG enrichment analyses and visualization as previously described (Yu et al., 2012). ClusterProfiler R package was used for enrichment analysis, Org. Hs. eg.db R package [3.10.0 version] was used for ID conversion, and GO plot R package [1.0.2 version] was used for calculating Zscore.

Immune Cell Infiltration in CRC

GSVA R package (version 1.34.0) was used to explore the enrichment of immune cells in the FCRLB high expression and low expression groups as previously described (Hänzelmann et al., 2013). Tumor purity analysis was performed by using Estimate R package according to a previous study (Aran et al., 2015). All procedures were performed using R software (version 4.1.3).

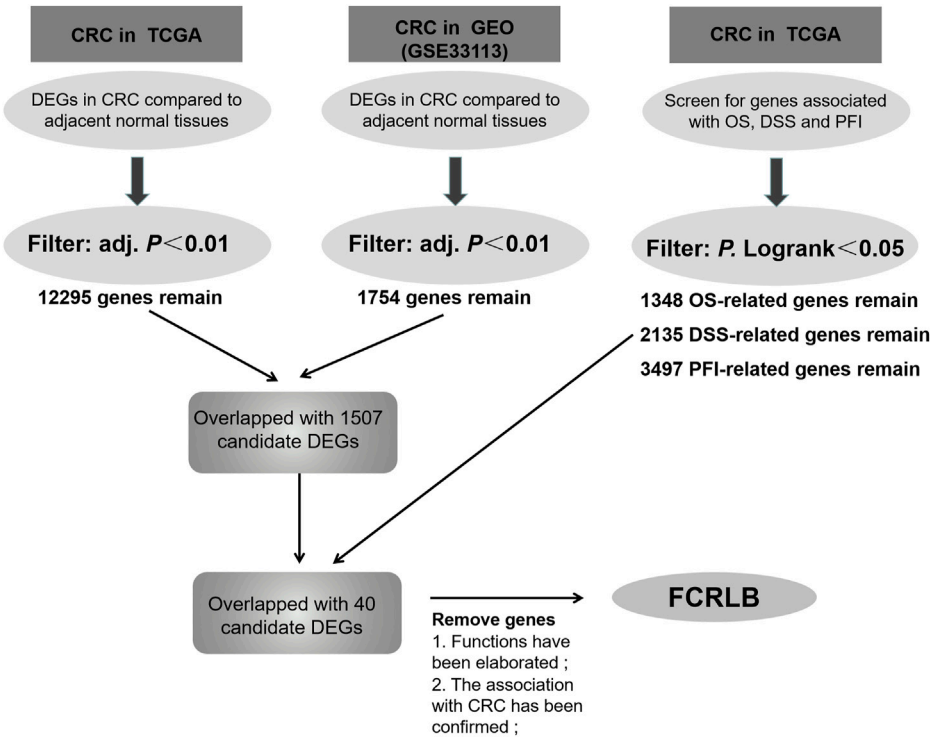


FIGURE 1 | The screening flow chart for selected DEGs associated with prognosis of CRC patients.

Immunohistochemistry

The TMA was used to perform immunohistochemistry according to the methods described in the previous literature (Xu et al., 2021). The expression of FCRLB was detected with a special antibody dilution (proteinech, China, 1:100). Then the stained TMA was examined by two expert pathologists independently, who were both blinded to the detailed clinico-pathological data. The number of positive cells in five high-power fields was randomly counted and then protein expression of FCRLB was determined by evaluating the staining intensity of positive staining, scored as negative (less than 25%), weak (25%–50%), medium (50%–75%), and strong (>75%).

Statistical Analysis

Analysis was performed by using R (version 4.1.3). The data in figures and tables were presented as mean ± SD. The Kaplan-Meier method and the Log-rank test were used to determine the overall survival (OS), disease specific survival (DSS), and progress free interval (PFI) between different groups. Spearman's correlation coefficient was utilized to analyze the correlation between FCRLB expression and the promoter methylation level. Logistic regression analysis was performed to detect the correlation between FCRLB expression and clinico-pathological parameters. Fisher's exact tests were used to analyze the correlation between FCRLB expression in CRC tissues and that in paired non-tumorous tissues. Student's two-tailed unpaired t-test was performed to analyze the FCRLB expression in different groups. Differences at $p < 0.05$ were considered to be significantly different.

RESULTS

FCRLB Was Screened as a Differentially Expressed Prognosis-Related Gene

Data were obtained from the TCGA database and GEO database, and microarray analyses were then performed using the Limma R package to compare the gene expression profiles of CRC and normal tissue groups. A total of 12,295 up-regulated genes were detected in the TCGA database (**Figure 2A**) and 1,754 in the GEO database (GSE331133) (**Figure 2B**), based on the screening parameters set at a probability value of $p < 0.01$. The survival R package was used for screening of OS-related genes, DSS-related genes, and PFI-related genes based on the parameters set with *P*. Logrank <0.05. Then 1,348, 2,135, and 3,497 genes that can affect OS, DSS, and PFI were detected, respectively. To further study the DEGs associated with different prognoses, these high-throughput analysis data overlapped, and only 40 candidate genes were left (**Figure 2C**). Detailed information of candidate genes was upload to the **Supplementary Materials**. Then genes, whose functions have been elaborated previously or whose association with CRC has been confirmed in a previous study, were eventually abandoned. FCRLB was finally found. A schematic of the screening process of DEGs is shown in **Figure 1**.

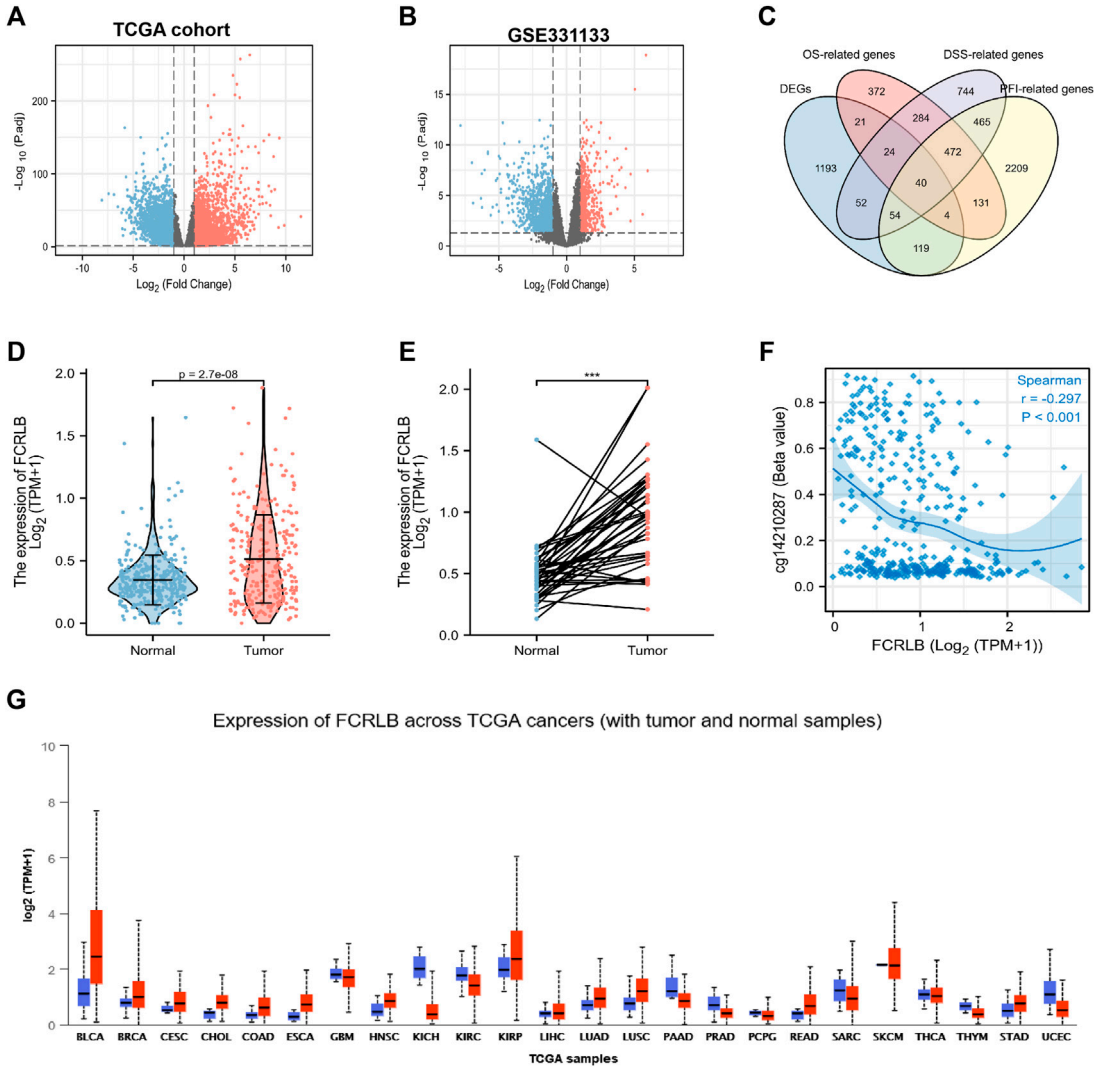


FIGURE 2 | FCRLB was highly expressed in CRC tissues and a variety of tumors. **(A)** The volcano plot showed the DEGs identified between the CRC and normal groups in TCGA cohort. **(B)** The volcano plot showed DEGs identified between CRC and normal groups in GEO database. The red and blue dot represents the up-regulated gene and down-regulated gene with significance, respectively. **(C)** Venn diagram showing genes screened by different methods. **(D)** Expression differences of FCRLB between 647 cancerous tissues and 51 non-tumorous tissues in the TCGA cohort. **(E)** Expression differences of FCRLB between 50 cancerous tissues and 50 matched adjacent normal tissues in the TCGA cohort. ***$p < 0.001$. **(F)** FCRLB expression was negatively correlated with promoter methylation ($r = -0.297$, $p < 0.001$). **(E)** FCRLB was overexpressed in a variety of tumors in the TCGA cohort. **(G)** Expression level of FCRLB in pan-cancer.

Expression of FCRLB Was Remarkably Overexpressed in CRC and Multiple Types of Cancers

The analysis of TCGA CRC data revealed that FCRLB was notably up-regulated in CRC tissues as compared to normal tissues ($p < 0.001$) (**Figure 2D**). Furthermore, FCRLB was highly expressed in CRC tissues compared to matched adjacent normal tissues in the TCGA CRC cohort ($p < 0.001$) (**Figure 2E**). Aberrant DNA methylation (including hyper- and hypo-methylation) at promoter regions has been reported to be closely associated with changes in gene expression. Our results showed that the expression of FCRLB was significantly negatively correlated with its promoter methylation levels (Spearman correlation = -0.297, $p < 0.001$) (**Figure 2F**). In addition, a pan-cancer analysis demonstrated that FCRLB was up-regulated in a variety of cancers, such as bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma (CESC), cholangiocarcinoma (CHOL), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and stomach adenocarcinoma (STAD), as shown in **Figure 2G**.

Association Between FCRLB Expression and Clinicopathological Characteristics

Clinicopathological data obtained from the TCGA CRC cohort, including gender, age, BMI, serum carcinoembryonic antigen

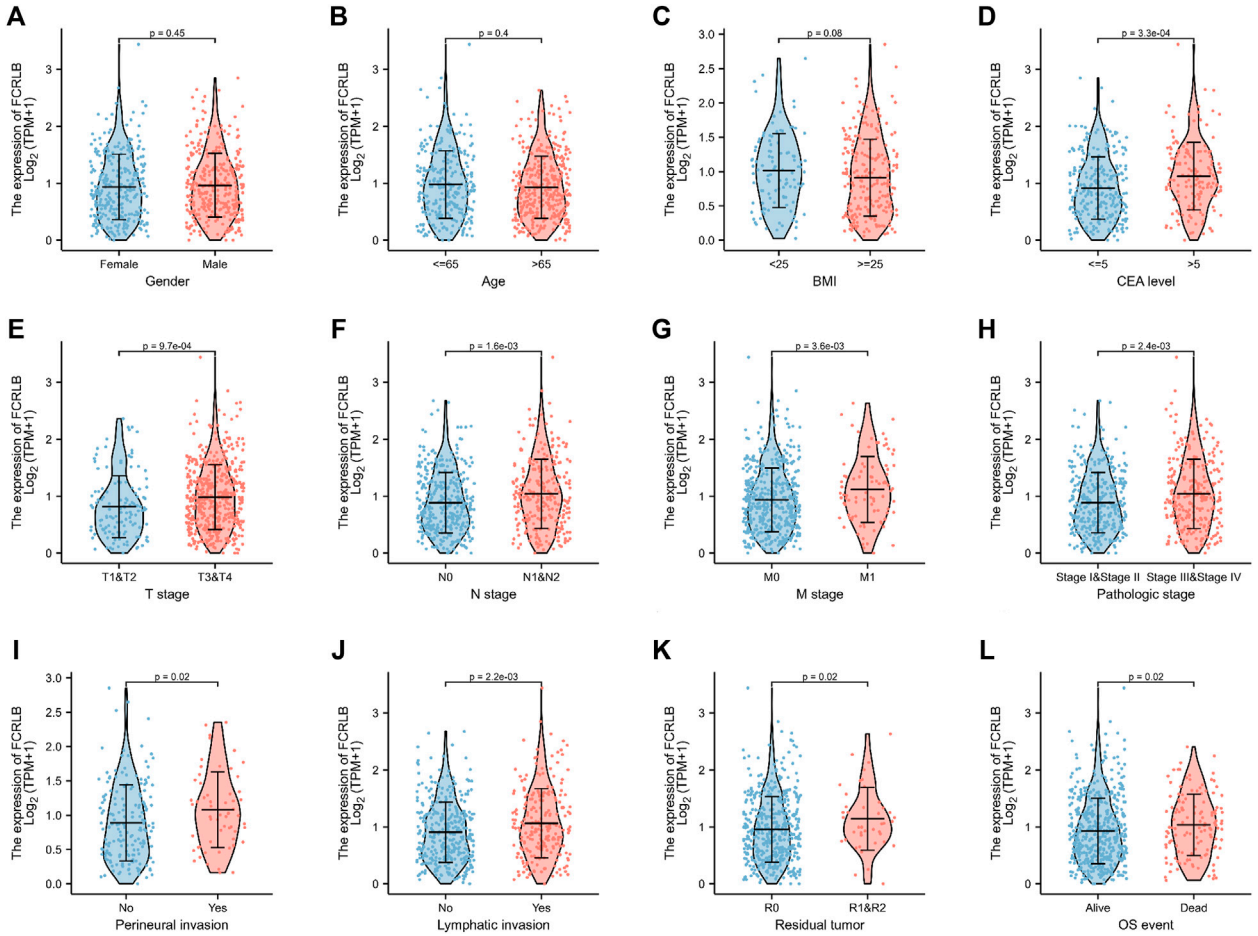


FIGURE 3 | The relationship between FCRLB expression and clinicopathological characteristics in the TCGA CRC cohort. **(A–I)** The FCRLB was highly expressed in CRC groups with higher serum CEA level **(D)**, higher T stage **(E)**, higher N stage group **(F)**, higher N stage group **(G)**, more advanced AJCC stage **(H)** , perineural invasion **(I)**, lymphatic invasion **(J)**, R1/R2 resection **(K)**, more unfavorable OS event **(L)**. However, no significant difference in expression of FCRLB was detected in different gender groups **(A)**, different age groups **(B)** and different BMI groups **(C)**. *p*-values were noted in each figure.

(CEA) level, perineural invasion, lymphatic invasion, residual tumor, TNN classification, and OS event, were analyzed. As shown in **Figures 3A–C**, the FCRLB expression level showed no significant differences between different gender subgroups (**Figure 3A**), age subgroups (**Figure 3B**), and BMI subgroups (**Figure 3C**). In addition, FCRLB expression levels were significantly higher in the subgroup with a higher serum CEA level ($p < 0.001$) (**Figure 3D**). The results revealed that the expression levels of FCRLB in the tissues of patients at the N1 and N2 stages were significantly higher than those in patients with N0 stage ($p < 0.01$) (**Figure 3F**). In addition, the expression levels of FCRLB were significantly up-regulated among patients with distant metastasis (M1 stage) than in those without (M0 stage) ($p < 0.01$) (**Figure 3G**). Moreover, the FCRLB expression levels were much higher among patients in the advanced AJCC stage compared to the early stage ($p < 0.01$) (**Figure 3H**). Moreover, FCRLB expression in the CRC tissues with incomplete resection (R1 or R2 resection) was notably higher than that with RO resection ($p < 0.05$) (**Figure 3K**) . As shown in **Figures 3I,J**, FCRLB expression was strongly associated with perineural invasion ($p < 0.05$) (**Figure 3I**) as well as lymphatic invasion ($p < 0.01$) (**Figure 3J**). Then, logistic regression analysis was performed and it revealed that FCRLB expression level was significantly correlated with CEA level (OR = 2.373 (1.574–3.610), $p < 0.001$), T stage (OR = 1.992 (1.346–2.975), $p < 0.001$), N stage (OR = 1.743 (1.271–2.396), $p < 0.001$, M stage (OR = 2.010 (1.263–3.253), $p = 0.004$), AJCC stage (OR = 1.636 (1.190–2.253), $p = 0.002$), residual tumor (OR = 2.174 (1.125–4.423), $p = 0.025$), perineural invasion (OR = 1.865 (1.033–3.418), $p = 0.040$), and lymphatic invasion (OR = 1.480 (1.061–2.070), $p = 0.021$) (**Table 1**).

High Expression of FCRLB Predicted Unfavorable Prognosis in the TCGA CRC Cohort

We then analyzed the association between FCRLB expression and the prognosis of CRC patients in TCGA. As shown in **Figure 4**,

TABLE 1 | Clinicopathological factors and FCRLB expression in the TCGA CRC cohort.

Characteristics	Total (N)	Odds Ratio (OR)	*p* Value
Age (>65 vs. ≤65)	644	0.837 (0.612–1.144)	0.265
Gender (male vs. female)	644	1.119 (0.821–1.526)	0.477
BMI (≥25 vs. <25)	329	0.689 (0.432–1.095)	0.116
CEA level (>5 vs. ≤5)	415	2.373 (1.574–3.610)	<0.001
T stage (T3 and T4 vs. T1 and T2)	641	1.992 (1.346–2.975)	<0.001
N stage (N1 and N2 vs. N0)	640	1.743 (1.271–2.396)	<0.001
M stage (M1 vs. M0)	564	2.010 (1.263–3.253)	0.004
AJCC stage (Stage III and Stage IV vs. Stage I and Stage II)	623	1.636 (1.190–2.253)	0.002
Residual tumor (R1 and R2 vs. R0)	510	2.174 (1.125–4.423)	0.025
Perineural invasion (yes vs. no)	235	1.865 (1.033–3.418)	0.040
Lymphatic invasion (yes vs. no)	582	1.480 (1.061–2.070)	0.021
Neoplasm type (rectum vs. colon adenocarcinoma)	644	1.176 (0.826–1.677)	0.368

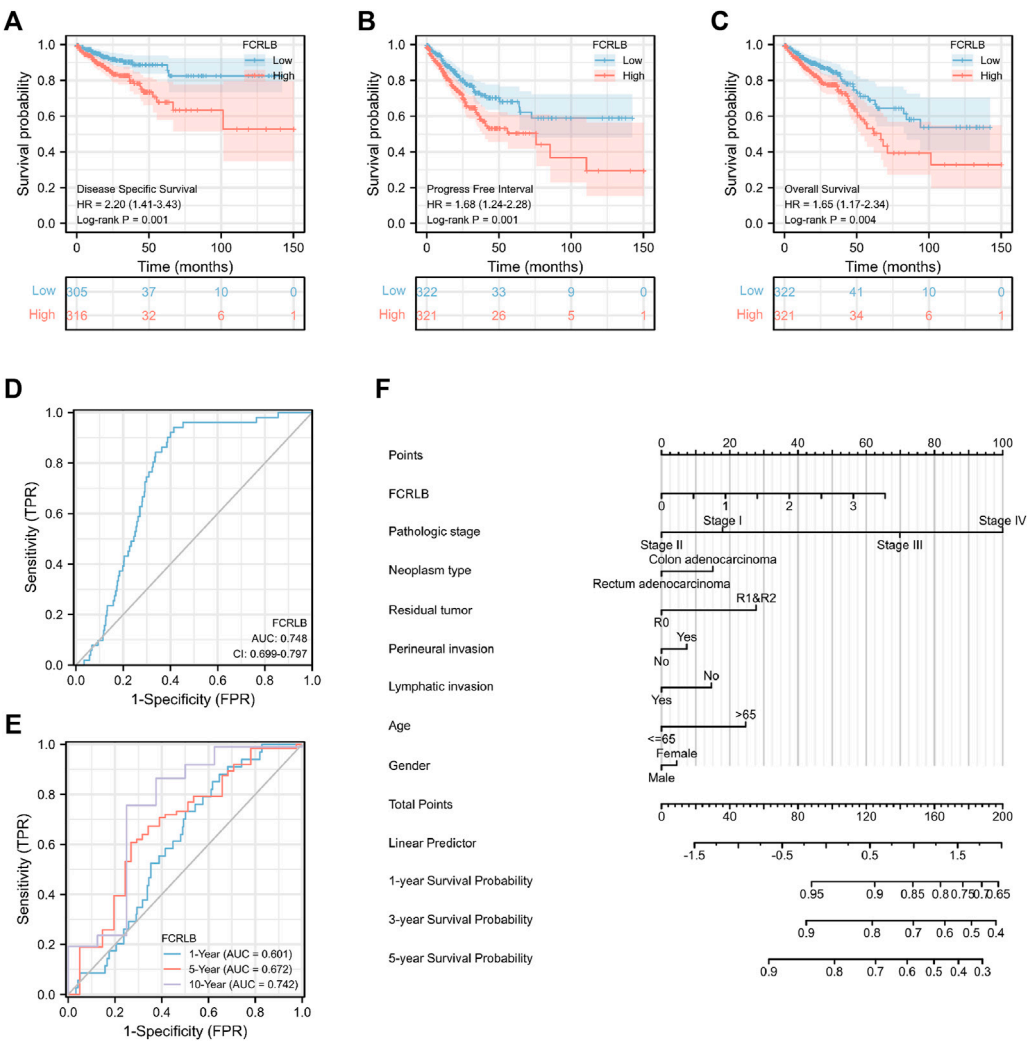


FIGURE 4 | Overexpression of FCRLB was associated with poor prognosis in the TCGA CRC cohort. **(A)** Kaplan-Meier analyses for OS according to FCRLB expression in CRC patients in the TCGA CRC cohort. **(B)** Higher expression of FCRLB was correlated with less favorable DSS in CRC patients in the TCGA CRC cohort. **(C)** Higher FCRLB expression level was correlated with poorer PFI in the TCGA CRC cohort. **(D)** The predictive value of FCRLB was evaluated by ROC curve. **(E)** Time-dependent ROC curves analysis of FCRLB. **(F)** OS Nomogram representation of the multivariate model.

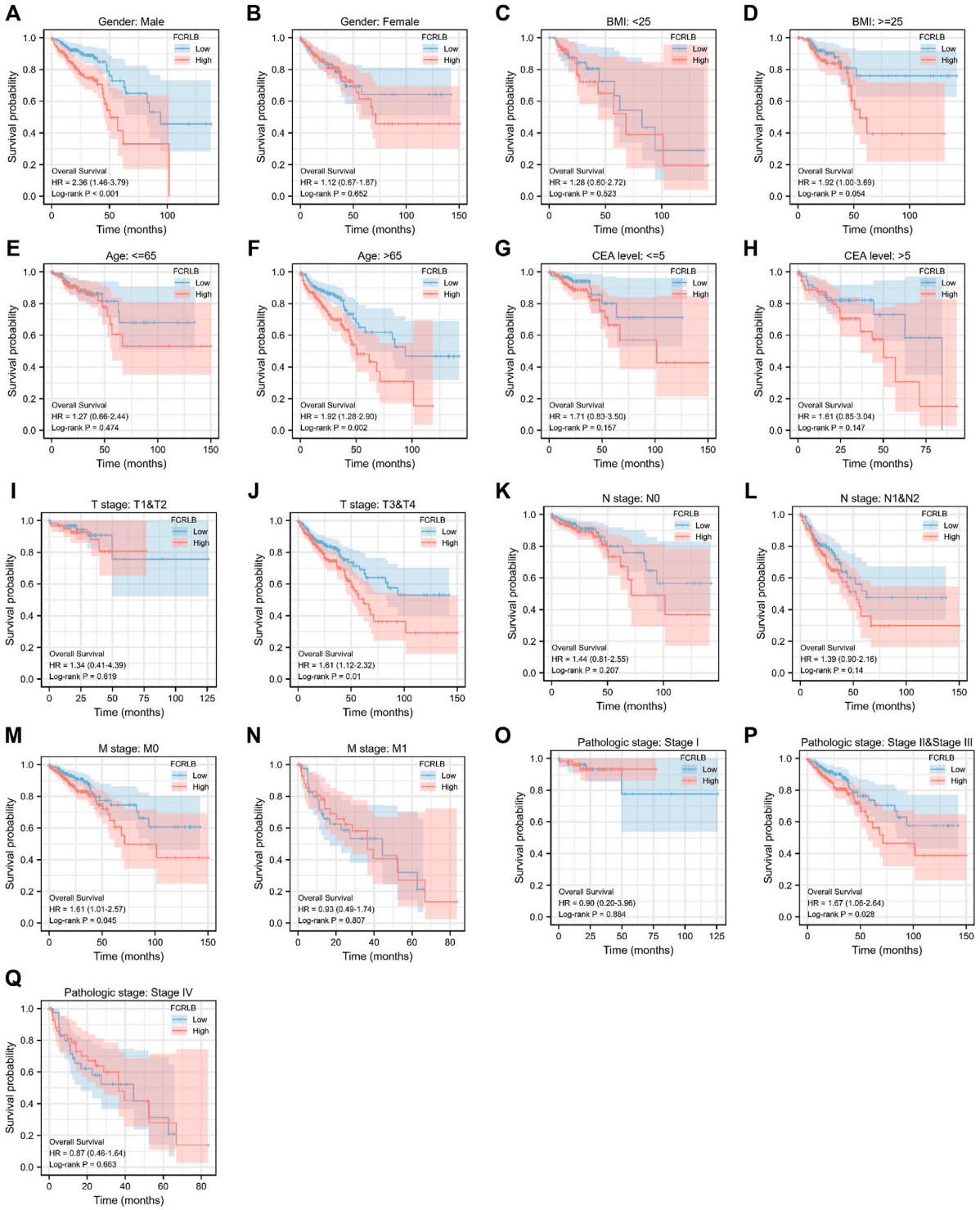


FIGURE 5 | Kaplan-Meier OS curves for FCRLB expression in different clinical subgroups. (A-S) Kaplan-Meier OS curves for FCRLB expression in male subgroup **(A)**, female subgroup **(B)**, BMI < 25 subgroup **(C)**, BMI < 25 subgroup **(D)**, age ≤65 subgroup **(E)**, age < 65 subgroup **(F)**, CEA level ≤5 subgroup **(G)**, CEA level≤ >5 subgroup **(H)**, T1 and T2 stage subgroup **(I)**, T3 and T4 stage subgroup **(J)**, N0 stage subgroup **(K)**, N1 and N2 stage subgroup **(L)**, M0 stage subgroup **(M)**, M1 stage subgroup **(N)**, AJCC pathologic stage I subgroup **(O)**, AJCC pathologic stage II and III subgroup **(P)**, AJCC pathologic stage IV subgroup **(Q)**.

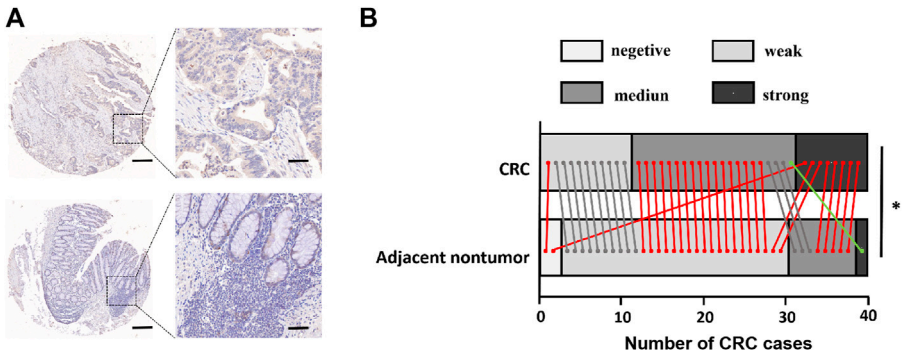


FIGURE 6 | FCRLB protein was highly expressed in CRC. **(A)** Representative image of immunohistochemical staining of FCRLB on a TMA containing 40 pairs of CRC and adjacent normal tissues. Scale bars = 500 μm (right) or 100 μm (left). **(B)** Statistical analysis of IHC staining results for FCRLB from the CRC tissue microarray slide. *$p < 0.05$.

low expression of FCRLB exhibited more favorable OS, PFI, and DSS outcomes ($p < 0.05$) (**Figures 4A–C**). To evaluate the diagnostic value of FCRLB, a receiver operating characteristic (ROC) curve was plotted according to the data obtained from the TCGA cohort. As shown in **Figure 4D**, the area under the ROC curve was 0.748 [95% confidence interval (CI): 0.699–0.797]. Then, time-dependent ROC curve analysis was constructed to further evaluate the diagnostic accuracy, sensitivity, and specificity of the FCRLB in terms of survival time. In the testing set, the 1-year AUC was 0.601, the 5-year AUC was 0.672, and the 10-year AUC was 0.742 (**Figure 4E**). In addition, OS nomograms were created to predict the probability of goal attainment based on clinicopathological characteristics (**Figure 4F**). In sum, the results showed that the overexpression of FCRLB was an unfavorable prognostic factor in CRC patients and presented good predictive performance.

In order to further characterize the role of FCRLB in predicting survival, survival analysis in 17 clinical subgroups was then conducted based on clinic-pathological characteristics such as gender, BMI, age, CEA level, T stage, N stage, M stage, and AJCC pathologic stage. As shown in **Figure 5A–Q**, FCRLB was identified as an independent prognostic and predictive factor in male patients (**Figure 5A**), old patients (age > 65) (**Figure 5F**), T2-3 stage patients (**Figure 5J**), patients without metastasis (M1 stage) (**Figure 5M**), and AJCC stage II–III stage patients (**Figure 5P**).

FCRLB Was Highly Expressed in CRC Tissues and Correlated With Infiltrating and Metastatic Ability in CRC Tissue Microarray (TMA)

First, the expression level of FCRLB in CRC tissues and adjacent normal tissues was compared using constructing and staining tissue microarray analysis. Representative pictures are shown in **Figure 6A**. Moreover, TMA revealed that FCRLB was strongly expressed in 8 of 40, moderately expressed in 20 of 40, and weakly expressed in 11 of 40 CRC tissues, respectively. By contrast, FCRLB was strongly expressed in 1 of 40, moderately expressed in 8 of 40, weakly expressed in 29 of 40, and negatively expressed in 2 of 40 adjacent normal tissues, respectively ($p < 0.001$) (**Figure 6B**). In addition, we analyzed the correlation between the expression level of FCRLB and the clinicopathological parameters of CRC patients. FCRLB was closely related to invasion depth ($X^2 = 5.625$, $p < 0.05$) and lymphatic metastasis ($X^2 = 3.956$, $p < 0.05$)) of CRC (**Table 2.**). Detailed clinicopathological information of CRC TMA was in **Supplementary Materials**.

Functional Annotation Among the High and Low FCRLB Expression Groups

Through the analysis of TCGA data, CRC patients in the TCGA cohort could be divided into FCRLB high and low expression groups according to the FCRLB expression level. Then, differentially expressed genes (DEGs) between the two groups were identified as FCRLB-associated genes. After obtaining the DEGs, through GO/KEGG enrichment analysis or GSEA, it is to further infer the possible functions or pathways involved in FCRLB. The results revealed that 499 DEGs were identified by performing genome-wide co-expression analysis with set cutoff values (p.adj <0.05 and |Log2-fold change|>1) between FCRLB high expression group and the low expression group. A total of 458 up-regulated genes and 41 down-regulated genes were identified and presented in the volcano plot (**Figure 7A**). The top 20 positively correlated genes and the top 20 negatively correlated genes in the TCGA cohort, ordered by the Pearson correlation coefficient, were presented in a gene co-expression heatmap (**Figure 7B**). The results of GO and KEGG pathway analyses were present in a network diagram and bubble diagram, which showed that co-expression of FCRLB was mainly enriched in extracellular matrix organization, T cell activation, positive regulation of cell adhesion, collagen-containing extracellular matrix, cell-cell junction, cell-substrate adherens junction, receptor ligand activity, cell adhesion molecule binding, extracellular matrix structural constituent, phosphoinositide 3-kinase/AKT (PI3/AKT signaling pathway), mitogen-activated

TABLE 2 | Clinicopathological factors and FCRLB expression in TMA.

Characteristics	Low expression (*n* = 20)	High expression (*n* = 20)	X^2	*p* Value
Age (>65)	10	8	0.404	0.525
Gender (male vs. female)	12	9	0.902	0.304
Differentiation (poorly differentiation)	2	5	1.558	0.212
Tumor size (maximal diameter >3 cm)	13	17	2.133	0.144
Invasion depth (T3 and T4 stage)	13	19	5.625	0.018
Tumor metastasis (M1 stage)	1	2	0.360	0.548
Lymphatic metastasis (yes vs. no)	4	10	3.956	0.047

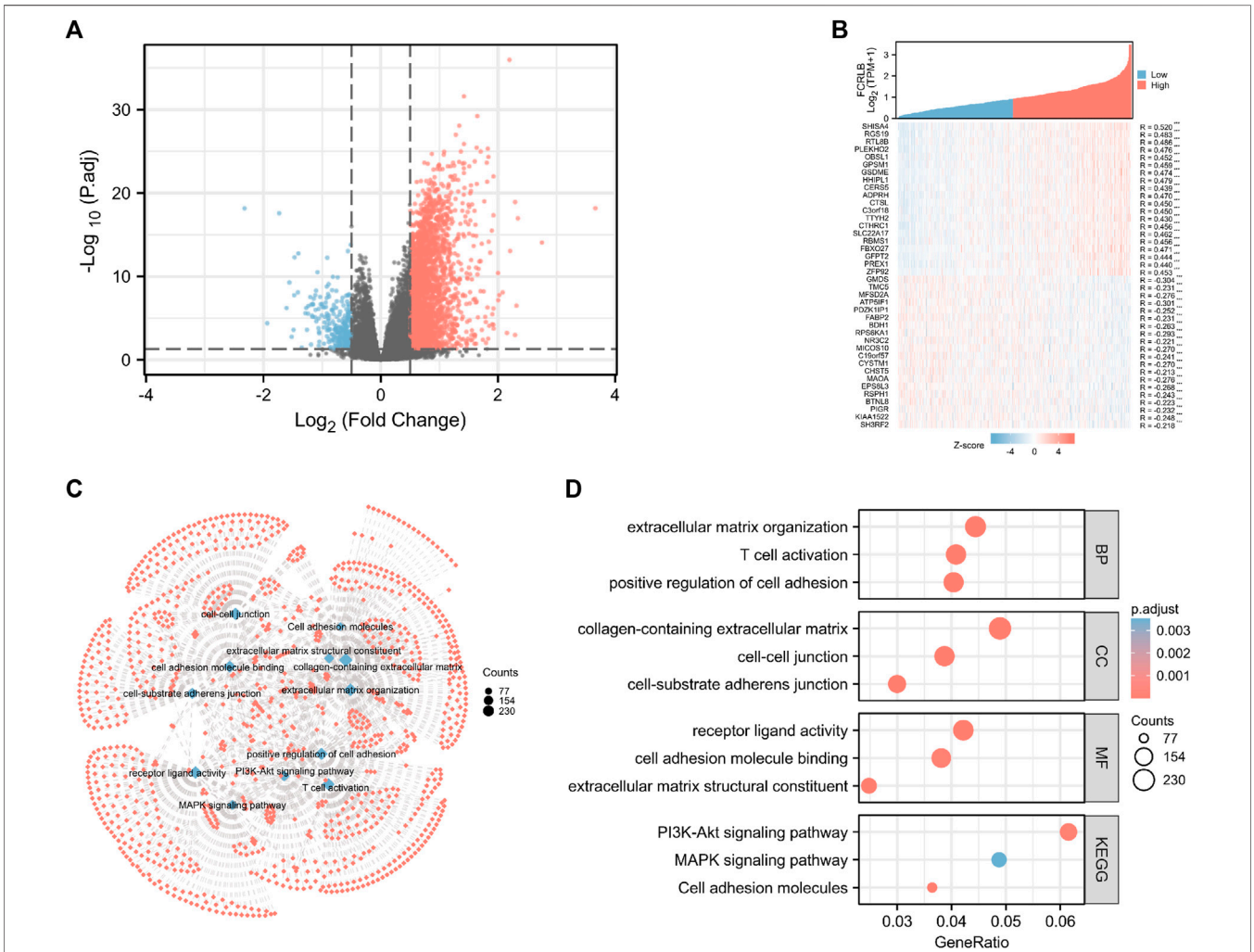


FIGURE 7 | Functional and pathway annotation among the high and low FCRLB groups by GO and KEGG analysis. **(A)** The volcano plot showed the DEGs identified between high and low FCRLB groups in the TCGA CRC cohort. The red and blue dot represents the up-regulated gene and down-regulated gene with significance, respectively, and each grey dot represents genes that did not meet the preset parameters (|log2(FC)|>1 and p. adj<0.05). **(B)** Gene expression heat map and correlations for FCRLB co-expressed genes. **p* < 0.05, ***p* < 0.01**, and **p* < 0.001. **(C,D)** The top 12 significant terms of GO and KEGG pathway analyses. The results were presented in a network diagram **(C)** and a bubble chart **(D)**, respectively.

protein kinase (MAPK) signaling pathway, and cell adhesion molecules (**Figures 7C,D**). In addition, the GSEA was performed to further identify the potential related pathway of FCRLB. The GSEA revealed that FCRLB was enriched in substantial gene sets, including epithelial-to-mesenchymal transition (EMT) in colorectal cancer (NES = 1.345, FDR <0.05) (**Figure 8B**), PI3/

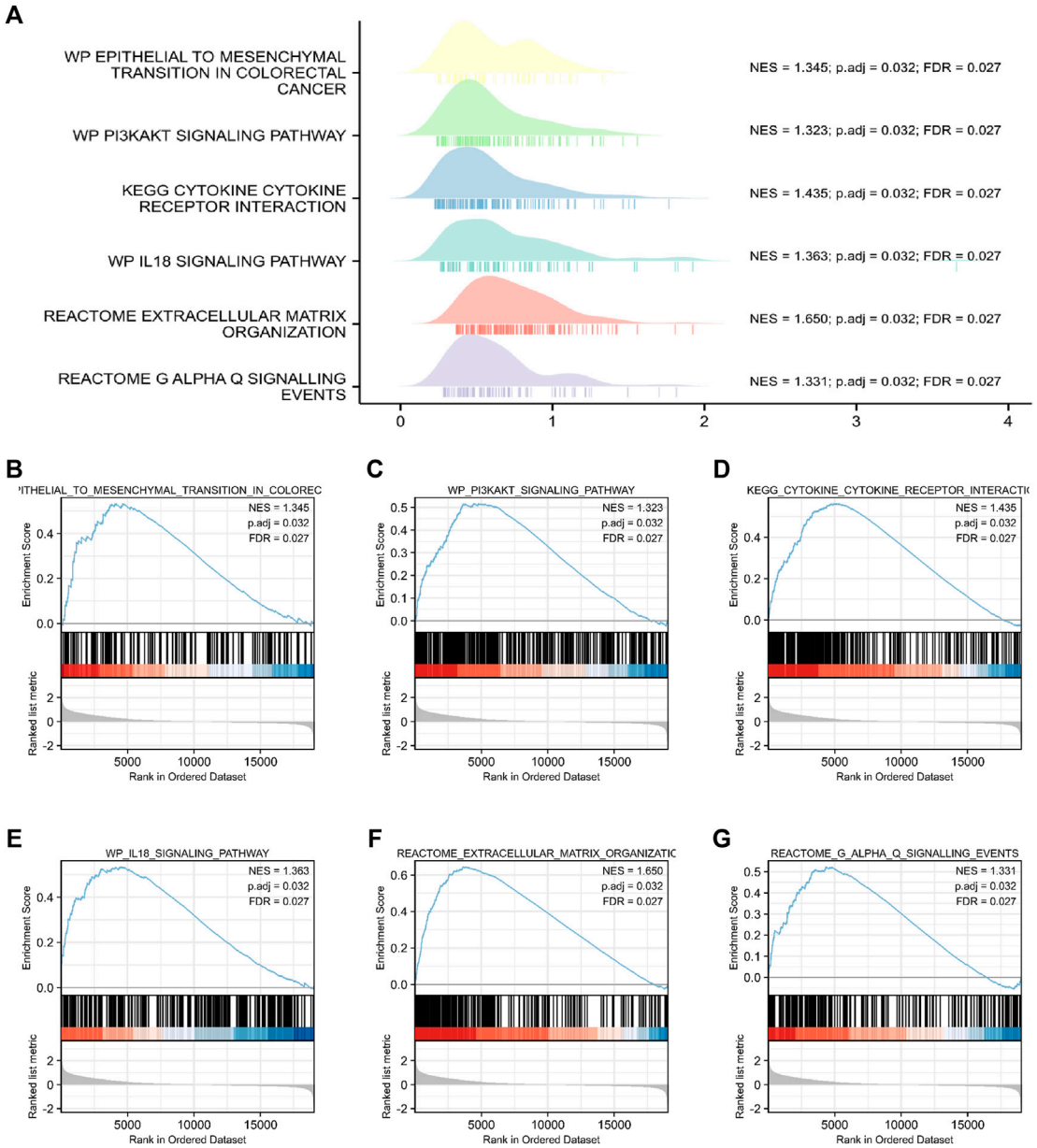


FIGURE 8 | Functional and pathway annotation among the high and low FCRLB groups by GSEA. **(A–G)** The GSEA revealed that EMT **(B)**, PI3K/Akt signaling pathway **(C)**, cytokine and cytokine receptor interaction **(D)**, IL-18 signaling pathway **(E)**, extracellular matrix organization **(F)**, Galphaq signaling **(G)**.

AKT signaling pathway (NES = 1.323, FDR <0.05) (**Figure 8C**), cytokine and cytokine receptor interaction (NES = 1.323, FDR <0.05) (**Figure 8D**), interleukin-18 (IL-18) signaling pathway (NES = 1.363, FDR <0.05) (**Figure 8E**), extracellular matrix organization (NES = 1.650, FDR <0.05) (**Figure 8F**), and Galphaq signal (NES = 1.331, FDR <0.05) (**Figure 8G**).

Correlation Between FCRLB and Tumor Microenvironment in CRC

The tumor microenvironment (TME), including cellular and non-cellular components, is increasingly being recognized to play a critical role in tumorigenesis. According to the previous literature (Rostamzadeh et al., 2018), FCRL family members play a critical role in cell-mediated immunity and tumor immunology. However, the role of FCRLB in modulating TME and tumor immunology has never been addressed before.

In TME, tumor purity, representing the percentage of tumor cells, could be calculated by performing the ESTIMATE method. TumorPurity, StromaScore, and ESTIMATEScore represent the percentage of tumor cells, the percentage of immune cells, and the percentage that merge ImmuneScore and StromaScore, respectively. The association between the ImmuneScore (StromalScore and ESTIMATEScore) and FCRLB expression

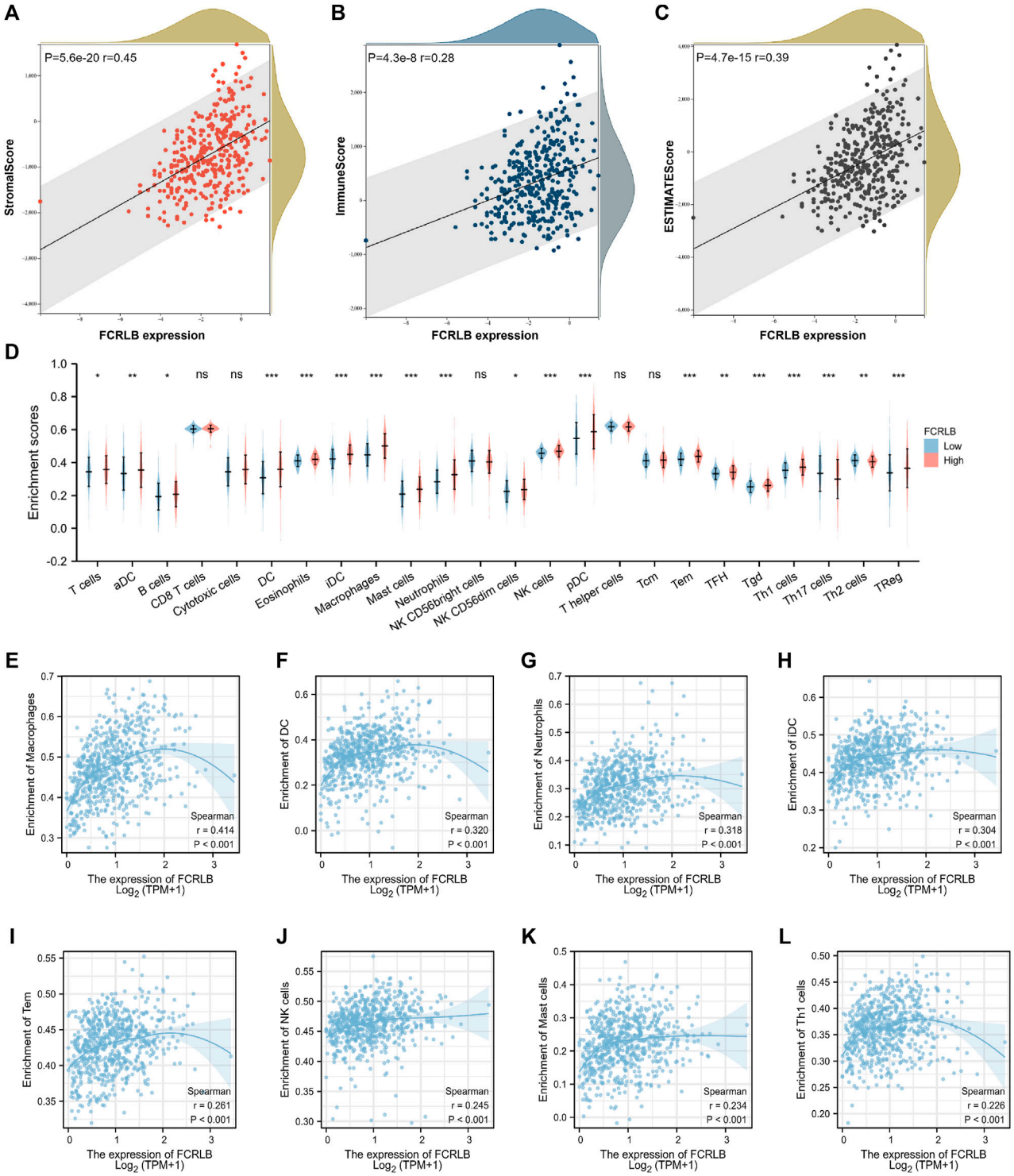


FIGURE 9 | The relationship between FCRLB expression and immune cell infiltration in TCGA CRC cohort. **(A–C)** The relationship between FCRLB expression levels and the StromalScore, ImmuneScore, and ESTIMATEScore. The correlation analysis revealed that FCRLB was positively correlated with the StromalScore **(A)**, ImmuneScore **(B)**, and ESTIMATEScore **(C)**. **(D)** The relationship between FCRLB expression levels and the abundance of different types of immune cells. ns stands for not statistically different. *$P < 0.05$, **$P < 0.01$** and *$P < 0.001$. **(E–L)** FCRLB expression was positively correlated with macrophage infiltration **(E)**, DC infiltration **(F)**, Neutrophil infiltration **(G)**, iDC infiltration **(H)**, Tem infiltration **(I)**, NK cell infiltration **(J)**, Mast cell infiltration **(K)**, Th1 cell infiltration **(L)**, etc. Sperman's coefficient **(R)** and p-value were noted in each figure.

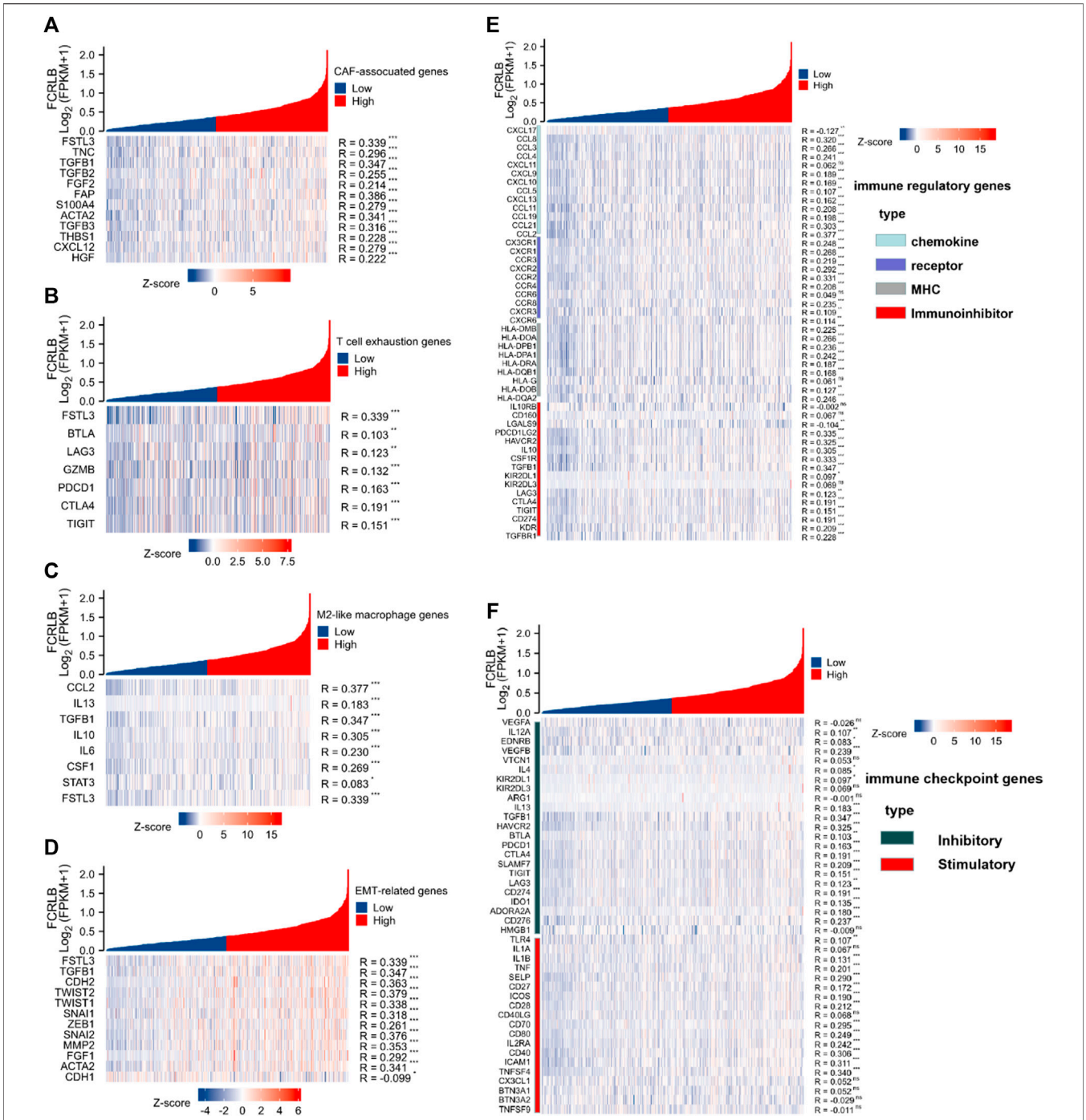


FIGURE 10 | The relationship of FCRLB expression to immune-related modulators and immune checkpoint molecules in the TCGA CRC cohort. **(A–D)** FCRLB was significantly positively correlated with most CAF-related genes **(A)**, T cell exhaustion genes **(B)**, M2 like macrophage genes **(C)**, EMT-related genes **(D)**. **(E)** FCRLB expression level was significantly correlated with majority of the immune regulatory genes. **(F)** FCRLB expression level is notably correlated with the expression of immune checkpoint molecules in CRC. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns: no significant difference.

was assessed as shown in **Figure 9A–C** FCRLB expression was positively correlated with ImmuneScore (r = 0.45, $p < 0.001$), StromalScore (r = 0.28, $p < 0.001$), and ESTIMATEScore (r = 0.39, $p < 0.001$), which demonstrated that FCRLB played a role in remodeling TME.

In order to characterize the immunological role of FCRLB in TME, we further evaluated the relationship between FCRLB and immune cell infiltration in CRC by analyzing the data set in the TCGA cohort and performing a single-sample gene set enrichment analysis (ssGSEA) according to the previous literature (Bindea et al.,

2013). The difference in the abundance of immune/stromal cells between the high FCRLB group and the low FCRLB group was shown in **Figure 9D**. Moreover, the relationship between FCRLB expression and immune cell infiltration was intuitively represented in a bubble diagram (**Supplementary Figure S1**). The analysis revealed that FCRLB expression was positively correlated with the abundance of multiple immune cells, such as macrophages (r = 0.414, $p < 0.001$) (**Figure 9E**), dendritic cells (DC) (r = 0.320, $p < 0.001$) (**Figure 9F**), neutrophils (r = 0.318, $p < 0.001$) (**Figure 9G**), immature dendritic cells (iDC) (r = 0.304, $p < 0.001$) (**Figure 9H**), effective memory T cells (Tem) (r = 0.261, $p < 0.001$) (**Figure 9I**), NK cells (r = 0.245, $p < 0.001$) (**Figure 9J**), mast cells (r = 0.234, $p < 0.001$) (**Figure 9K**), and Th1 cells (r = 0.226, $p < 0.001$) (Fig. 9L).

Correlation Between FCRLB and Immune-Related Modulators in CRC

To better characterize the potential role of FCRLB in modulating TME and tumor immunity, the relationship between FCRLB and immune modulators was further explored by analyzing the association between FCRLB expression and the maker genes of related immune cells.

As shown in **Figure 10A**, significant positive correlations between FCRLB expression and the majority of genes related to CAF-associated genes were clearly observed. Moreover, we found that FCRLB expression was positively correlated with the gene set associated with T cell exhaustion (**Figure 10B**). In addition, FCRLB also showed a significant positive correlation with M2-like macrophage genes (**Figure 10C**), as well as genes associated with epithelial-to-mesenchymal transition (EMT), which is a well-documented transdifferentiation program playing a critical role in cancer metastasis (Ievgenia and Cédric, 2019) and chemoresistance (Zheng et al., 2015) (**Figure 10D**).

Not only that, we also explored the relationship between FCRLB expression and immune regulatory genes (including chemokine, receptor, major histocompatibility complex (MHC), and immunoinhibitor). Notably, FCRLB showed a significant positive correlation with the majority of immune regulatory genes.

Immune checkpoint receptors inhibit the activation of T cells by delivering co-inhibitory signals to modulate the duration and intensity of the immune response (Darvin et al., 2018). In our study, the relationship between FCRLB expression and a variety of immune checkpoint molecules in CRC was explored. The results revealed a strong correlation between FCRLB expression and the majority of immune checkpoint molecules in CRC, including inhibitory and stimulatory genes (**Figure 10F**).

DISCUSSION

Although the prognosis of patients with CRC has dramatically improved over the past decades, early accurate diagnosis and therapy are still considered important prognostic factors. Prognostic and predictive biomarkers are considered to benefit early detection, accurate diagnosis, and precise therapy of CRC. Therefore, from a clinical standpoint, it is imperative to excavate prognostic and predictive biomarkers.

In our research, based on data mining of the TCGA database, FCRLB was screened as a potential prognostic biomarker for CRC patients. We discovered that FCRLB was overexpressed in CRC tissues compared to adjacent normal tissues. Moreover, we found that the high expression of FCRLB was significantly correlated with unfavorable OS, DSS, and PFI in CRC and may serve as a potential prognostic biomarker of CRC. Furthermore, ROC curve, time-dependent ROC curve, and OS nomogram were created and demonstrated the utility and good predictive performance of FCRLB in CRC. According to the analysis of clinicopathological data in the TCGA CRC cohort, FCRLB expression was positively correlated with a high serum CEA level, an advanced TNM stage, incomplete tumor resections, and perineural and lymphatic invasion, all of which are typically considered poor prognostic factors. Subsequently, we further detected the expression level of FCRLB and explored the relationship between FCRLB expression and clinicopathological characteristics using the TMA technique. Additionally, we demonstrated that the strong expression of FCRLB was strongly associated with lymph node metastasis and invasion depth, which confirmed, to some extent, the conclusions drawn from our analysis of the TCGA database.

Fc receptor-like (FCRL) molecules, a family of cellular receptors, are mainly expressed in B cells (Ehrhardt et al., 2007). To date, eight different members of the FCRL family have been identified in humans, including FCRL-6, FCRLA, and FCRLB. FCRL1-6 encode type I transmembrane glycoproteins with similar extracellular immunoglobulin (Ig)-like domains and intracellular regions that contain consensus tyrosine-based motifs (viz., an immunoreceptor tyrosine-based activation motif and/or an immunoreceptor tyrosine-based inhibitory motif). However, FCRLA and FCRLB mainly serve as intracellular proteins rather than transmembrane receptors like other members of the FCR family (Davis, 2020). The FCRL family has been proven to play a critical role in phagocytosis, antibody-dependent cell cytotoxicity, immediate hypersensitivity, and transcytosis of immunoglobulins via their ability to bind immunoglobulin constant regions (Chikaev et al., 2005). Recently, a large number of studies on the FCRL family in tumors and tumor immunity have emerged. For instance, anti-FCRL1 immunotoxin E9 (Fv)-PE38, a well-designed immunotoxin targeting cell-surface receptors, displayed remarkably selective cytotoxicity on FCRL1-positive malignancies (Du et al., 2008). Furthermore, FCRL2 expression predicts clinical progression in chronic lymphocytic leukemia (Li et al., 2008). Genetic polymorphisms of FCRL3 were also associated with the risk of head and neck cancer in a Chinese population (Zhang and Sun, 2021). However, to the best of our knowledge, very few studies regarding the immunological function of FCRLB have been reported previously, and its role in oncology and tumor immunology has never been addressed, partially due to it being a relatively newly defined group within the FCRL family. In this study, to the best of our knowledge we characterized, for the first time, the potential oncological role of FCRLB.

In addition, to further elucidate the underlying biological mechanism of FCRLB in CRC, GSEA, GO, and KEGG analysis were performed. The results demonstrated that some well-documented tumor-related pathways and immune-related

pathways in the FCRLB high expression group were significantly enriched, such as EMT programs, the PI3/AKT pathway, cytokine and cytokine receptor interactions, the IL-18 signaling pathway, extracellular matrix organizations, Galphaq signals, and so on. Accordingly, we speculated that FCRLB may serve as an oncogene that influences the prognosis of patients with CRC through these common biological pathways (Li et al., 2014; Nakamura et al., 2018; Ievgenia and Cédric, 2019; Patra et al., 2021).

In recent years, the tumor immune microenvironment has become an area of intense research interest, mainly due to its close association with the effectiveness of immunotherapy. Most members of the FCRL family ubiquitously serve as transmembrane glycoproteins responsible for the recognition of extracellular ligands by the cellular effector pathways of immune cells. Considering the critical role of the FCRL family in immunoregulation, we explored the potential association between FCRLB and the TME of CRC. GSEA, GO, and KEGG analysis revealed that immune-related pathways were significantly enriched in the FCRLB high expression group, such as T cell activation, cytokine and cytokine receptor interaction, and IL-18 signaling pathways. Moreover, FCRLB expression was significantly positively correlated with the majority of the genes of CAF and M2-like macrophages, which are both immunosuppressive cells. Not only that, a significantly positive correlation was observed between most immunosuppression-associated genes and FCRLB expression. Clinically, T cell exhaustion, characterized by the gradual functional deficits of exhausted T cells, has been considered to be a cause of the failure of immunotherapy. In our research, FCRLB expression correlated positively with T cell exhaustion genes. The immunosuppressive TME is a major barrier to immunotherapy leading to a limited immunotherapy effect (Devalaraja et al., 2020). Thus, we speculated that high FCRLB may lead to the formation of an immunosuppressive TME that can promote immune escape and foster tumor growth. Targeted inhibition of FCRLB may reverse the immunosuppression of TME to improve outcomes for CRC patients. In sum, our study indicated that FCRLB might play an important role in tumor immunity by modulating TME. However, the regulatory mechanism of FCRLB in tumor immunity still needs to be further clarified.

It is undeniable that the current study has some limitations. First, the functional roles of FCRLB in regulating tumorigenesis and the development of colorectal cancer were unclear and need to be addressed with further investigation. Second, the clinical conclusions drawn from the current study were mainly derived from the analysis of public database. Thus, further clinical investigation and validation are urgently needed. Third, in our research, the number of TMA cases was quite small, and the sample size derived from public databases is also limited, which may lead to inaccurate conclusions. Therefore, large samples from multiple datasets are needed to support these conclusions. Moreover, in the data we obtained from the TCGA database, most clinical information about preoperative (neo-adjuvant) or postoperative (adjuvant) therapy was unknown, even as they serve as critical factors affecting prognosis. Therefore, the association between FCRLB expression level and neoadjuvant or adjuvant therapy was not elaborated.

CONCLUSION

In conclusion, we confirmed that FCRLB overexpression is closely related to the poor prognosis of patients with CRC. Moreover, we identified FCRLB as a biomarker for evaluating the prognosis of CRC patients. In light of these findings, further studies on the specific function and potential mechanism of FCRLB are urgently warranted, and the development of an efficient strategy to suppress FCRLB expression may improve the prognosis of patients with CRC in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XW, RL, YZ, and YW completed data analysis wrote the main manuscript text. SW| prepared **Figures 1**, **2**. ZL prepared **Figures 3**, **4**. SC prepared **Figures 5**, **6**. XW prepared **Figures 7–10** and **Tables 1**, **2**. ZY and LC guided other authors in topic selection and data analysis. All authors reviewed the manuscript.

FUNDING

This research was sponsored by the Startup Fund for scientific research, Fujian Medical University (grant number: 200QH1217), in-hospital funding project-talent introduction project, Fujian Cancer Hospital (Grant number: YJ-YJ-03), and in-hospital funding project-talent introduction project, Fujian Cancer Hospital (Grant number: YJ-YJ-01).

ACKNOWLEDGMENTS

Thanks to the TCGA database for the selflessly provided clinicopathological data of CRC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.882307/full#supplementary-material

REFERENCES

Aran, D., Sirota, M., and Butte, A. J. (2015). Systematic Pan-Cancer Analysis of Tumour Purity[J]. *Nat. Commun.* 6, 8971. doi:10.1038/ncomms9971

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Cao, Q., Yan, X., Chen, K., Huang, Q., Melancon, M. P., Lopez, G., et al. (2018). Macrophages as a Potential Tumor-Microenvironment Target for Noninvasive Imaging of Early Response to Anticancer Therapy. *Biomaterials* 152, 63–76. doi:10.1016/j.biomaterials.2017.10.036

Chikaev, N. A., Bykova, E. A., Najakshin, A. M., Mechetina, L. V., Volkova, O. Y., Peklo, M. M., et al. (2005). Cloning and Characterization of the Human FCRL2 Gene. *Genomics* 85 (2), 264–272. doi:10.1016/j.ygeno.2004.10.017

Daniel, S. R., Alison, T., Richard, H., McWilliams, A., Robert, L. R., Christina, G., et al. (2017). Effect of Combined Patient Decision Aid and Patient Navigation vs Usual Care for Colorectal Cancer Screening in a Vulnerable Patient Population: A Randomized Clinical Trial.[J]. *JAMA Intern. Med.* 177 (7), 967–974. doi:10.1001/jamainternmed.2017.1294

Darvin, P., Toor, S. M., Sasidharan Nair, V., and Elkord, E. (2018). Immune Checkpoint Inhibitors: Recent Progress and Potential Biomarkers. *Exp. Mol. Med.* 50 (12), 1–11. doi:10.1038/s12276-018-0191-1

Davis, R. S. (2020). Roles for the FCRL6 Immunoreceptor in Tumor Immunology. *Front. Immunol.* 11, 575175. doi:10.3389/fimmu.2020.575175

Deleuze, A., Saout, J., Dugay, F., Peyronnet, B., Mathieu, R., Verhoest, G., et al. (2020). Immunotherapy in Renal Cell Carcinoma: The Future Is Now. *Int. J. Mol. Sci.* 21 (7). doi:10.3390/ijms21072532

Devalaraja, S., To, T. K., Folkert, I. W., Ramakrishnan, N., Md Zahidul, A., Minghong, L., et al. (2020). Tumor-Derived Retinoic Acid Regulates Intratumoral Monocyte Differentiation to Promote Immune Suppression - ScienceDirect[J]. *Cell* 180 (6), 1098–1114. e16. doi:10.1016/j.cell.2020.02.042

Du, X., Nagata, S., Ise, T., Stetler-Stevenson, M., and Pastan, I. (2008). FCRL1 on Chronic Lymphocytic Leukemia, Hairy Cell Leukemia, and B-Cell Non-hodgkin Lymphoma as a Target of Immunotoxins. *J. Blood* 111 (1), 338–343. doi:10.1182/blood-2007-07-102350

Ehrhardt, G. R., Leu, C. M., Zhang, S., Aksu, G., Jackson, T., Haga, C., et al. (2007). Fc Receptor-like Proteins (FCRL): Immunomodulators of B Cell Function. *Adv. Exp. Med. Biol.* 596 (1), 155–162. doi:10.1007/0-387-46530-8_14

Ganesh, K., Stadler, Z. K., Cercek, A., Robin, B. M., Jinru, S., Neil, H. S., et al. (2019). Immunotherapy in Colorectal Cancer: Rationale, Challenges and Potential[J]. *Nat. Rev. Gastroenterology Hepatology* 16 (6), 361–375. doi:10.1038/s41575-019-0126-x

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data[J]. *Bmc Bioinforma.* 14 (1), 7. doi:10.1186/1471-2105-14-7

Ievgenia, P., and Cédric, B. (2019). EMT Transition States during Tumor Progression and Metastasis [J]. *Trends Cell Biol.* 29 (3), 212–226. doi:10.1016/j.tcb.2018.12.001

Kobayashi, H., Enomoto, A., Woods, S. L., Alastair, D. B., Masahide, T., and Daniel, L. W. (2019). Cancer-associated Fibroblasts in Gastrointestinal Cancer[J]. *Nat. Rev. Gastroenterology Hepatology* 16 (5), 282–295. doi:10.1038/s41575-019-0115-0

Komenaka, I., Hoerig, H., and Kaufman, H. L. (2004). Immunotherapy for Melanoma. *Clin. Dermatology* 22 (3), 251–265. doi:10.1016/j.clindermatol.2003.12.001

Li, F. J., Ding, S., Pan, J., Shakhmatov, M. A., Kashentseva, E., Wu, J., et al. (2008). FCRL2 Expression Predicts IGHV Mutation Status and Clinical Progression in Chronic Lymphocytic Leukemia. *Blood* 112 (1), 179–187. doi:10.1182/blood-2008-01-131359

Li, X., Wang, K., Ren, Y., Zhang, L., Tang, X.-J., Zhang, H.-M., et al. (2014). MAPK Signaling Mediates Sinomenine Hydrochloride-Induced Human Breast Cancer Cell Death via Both Reactive Oxygen Species-dependent and -independent Pathways: an *In Vitro* and *In Vivo* Study. *Cell Death Dis.* 5 (7), e1356. doi:10.1038/cddis.2014.321

Nakamura, K., Kassem, S., Cleynen, A., Chrétien, M.-L., Guillerey, C., Putz, E. M., et al. (2018). Dysregulated IL-18 Is a Key Driver of Immunosuppression and a Possible Therapeutic Target in the Multiple Myeloma Microenvironment. *Cancer Cell* 33 (4), 634–648. e5. doi:10.1016/j.ccell.2018.02.007

Nasser, N. J., Gorenberg, M., and Agbarya, A. (2020). First Line Immunotherapy for Non-small Cell Lung Cancer. *Pharmaceuticals* 13 (11), 373. doi:10.3390/ph13110373

Patra, R., Das, N. C., and Mukherjee, S. (2021). Exploring the Differential Expression and Prognostic Significance of the COL11A1 Gene in Human Colorectal Carcinoma: An Integrated Bioinformatics Approach[J]. *Front. Genet.* 12, 608313. doi:10.3389/fgene.2021.608313

Rostamzadeh, D., Kazemi, T., Amirghofran, Z., and Mahdi, S. (2018). Update on Fc Receptor-like (FCRL) Family: New Immunoregulatory Players in Health and Diseases[J]. *Expert Opin. Ther. Targets* 22 (6), 487–502. doi:10.1080/14728222.2018.1472768

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics, 2018. *CA Cancer J. Clin.* 68 (1), 7–30. doi:10.3322/caac.21442

Terranova-Barberio, M., Pawlowska, N., Dhawan, M., Moasser, M., Chien, A. J., Melisko, M. E., et al. (2020). Exhausted T Cell Signature Predicts Immunotherapy Response in ER-Positive Breast Cancer. *Nat. Commun.* 11, 3584. doi:10.1038/s41467-020-17414-y

Xu, S., Mo, C., Lin, J., Yan, Y., Liu, X., Wu, K., et al. (2021). Loss of ID3 Drives Papillary Thyroid Cancer Metastasis by Targeting E47-Mediated Epithelial to Mesenchymal Transition. *Cell Death Discov.* 7 (1), 226. doi:10.1038/s41420-021-00614-w

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics a J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, Y., and Sun, D. (2021). Genetic Polymorphisms of FCRL3, NLRP3 and IL2 Are Associated with the Risk of Head and Neck Cancer in a Chinese Population. *Pharmgenomics Pers. Med.* 14, 1047–1053. doi:10.2147/pgpm.s324750

Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., et al. (2015). Epithelial-to-mesenchymal Transition Is Dispensable for Metastasis but Induces Chemoresistance in Pancreatic Cancer. *Nature* 527 (7579), 525–530. doi:10.1038/nature16064

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Lin, Zeng, Wang, Wei, Lin, Chen, Ye and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 29 June 2022
doi: 10.3389/fgene.2022.892589

The m7G-Related Long Noncoding RNA Signature Predicts Prognosis and Indicates Tumour Immune Infiltration in Colon Cancer

Li Liu[1†], Yukang Wu[1†], Wenzheng Chen[1], Yebei Li[2], Jiahe Yu[3], Guoyang Zhang[1], Pengcheng Fu[1], Liu Huang[1], Jianbo Xiong[1] and Zhigang Jie[1]**

[1]Department of Gastrointestinal Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, China, [2]Department of Renal Medicine, The Second Affiliated Hospital of Nanchang University, Nanchang, China, [3]College of Clinical Medicine, Hainan Vocational University of Science and Technology, Hainan, China

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Qingguo Li,
Fudan University, China
Siqi Wu,
Soochow University, China

***Correspondence:**
Jianbo Xiong
xiongjianbo2017@foxmail.com
Zhigang Jie
jiezg123@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 09 March 2022
Accepted: 13 June 2022
Published: 29 June 2022

Citation:
Liu L, Wu Y, Chen W, Li Y, Yu J, Zhang G, Fu P, Huang L, Xiong J and Jie Z (2022) The m7G-Related Long Noncoding RNA Signature Predicts Prognosis and Indicates Tumour Immune Infiltration in Colon Cancer. Front. Genet. 13:892589. doi: 10.3389/fgene.2022.892589

With high morbidity and mortality, colon cancer (CC) is considered as one of the most often diagnosed cancers around the world. M7G-related lncRNA may provide a regulatory function in the formation of CC, but the principle of regulation is still unclear. The purpose of this research was to establish a novel signature that may be used to predict survival and tumour immunity in CC patients. Data about CC in TCGA was collected for analysis, coexpression analysis and univariate Cox analysis were used to screen prognostic m7G-related lncRNAs. A consensus clustering analysis based on prognostic m7G-related lncRNAs was applied, and a prognosis model based on least absolute shrinkage and selection operator (LASSO) regression analysis was established. Independent prognostic analysis, nomogram, PCA, clinicopathological correlation analysis, TMB, survival analysis, immune correlation analysis, qRT–PCR and clinical therapeutic compound prediction were also applied. 90 prognostic m7G-related lncRNAs were found, GO and KEGG analysis showed that prognostic m7G-related lncRNAs were mainly related to cell transcription and translation. The results of the consensus clustering analysis revealed substantial disparities in survival prognosis and tumour immune infiltration between two clusters. We built a risk model with 21 signature m7G-related lncRNAs, patients in the high-risk group had a considerably poorer prognosis than those in the low-risk group. Independent prognostic analysis confirmed that patients' prognosis was linked to their tumour stage and risk score. PCA, subgroups with distinct clinicopathological characteristics were studied for survival, multi-index ROC curve, c-index curve, the survival analysis of TMB, and model comparison tested the reliability of risk model. A tumour immunoassay revealed a substantial difference in immune infiltration between high-risk and low-risk individuals. Five chemicals were eliminated, and qRT–PCR indicated that the four lncRNAs were expressed differently. Overall, m7G-related lncRNA is closely related to colon cancer and the 21 signature lncRNAs risk model can efficiently evaluate the prognosis of CC patients, which has a possible positive consequence for the future diagnosis and therapy of CC.

Keywords: m7G, lncRNA, colon cancer, bioinformatics, tumour immunity

INTRODUCTION

Colon cancer (CC) is one of the most prevalent diagnosed tumours worldwide, with the third-highest incidence of occurrence (10%) among all cancers, trailing only lung cancer (11.4%) and breast cancer (11.7%) (Sung et al., 2021). After lung cancer (18%), CC is the second largest cause of mortality (9.4%) (Sung et al., 2021). Colonoscopy have spread worldwide which increases the rate of early detection of CC, and improvements in treatments have decreased the mortality of CC, but CC is still one of the critical causes of threat to health (Miller et al., 2019). Therefore, it is urgent to explore new risk factors and biomarkers to predict prognosis and develop new therapeutic targets.

Noncoding transcripts of more than 200 nucleotides are known as long noncoding RNAs (lncRNAs) (Wang et al., 2011). In previous studies, lncRNAs that lack open reading frames were thought to be "useless impurities" in coding RNAs. It is now clear that these "useless impurities" serve vital cellular functions both in epigenetic gene-regulatory mechanisms and at the transcriptional and posttranscriptional levels (Chen, 2016). Several studies have suggested that the effects of lncRNAs are crucial to the pathogenesis and development of colon cancer (Bhan et al., 2017; Ni et al., 2020; Chen et al., 2021). Furthermore, Sen et al. indicated that lncRNA FEZF1-AS1 is beneficial to immune escape in colon cancer *via* the mediation of T reg cell differentiation (Hong et al., 2021). The above mentioned studies imply that lncRNAs have a strong link to tumour immunity.

Nevertheless, the mechanisms of regulating the expression of lncRNAs are largely unclear. Recent studies have suggested that there is a close relationship between the expression of lncRNAs and RNA methylation (Fazi and Fatica, 2019; He et al., 2020; Yi et al., 2020). This process plays a crucial role in multitudinous functions of cells, including RNA splicing (Zhao et al., 2014; Xiao et al., 2016), DNA damage repair (Xiang et al., 2017), translation (Wang et al., 2015), mRNA stability (Du et al., 2016; Lee et al., 2020), immunogenicity (Karikó et al., 2005), immune reponse (Li et al., 2017), and oncogenesis (Zhang et al., 2017; Wang et al., 2020). Previous studies have confirmed that the methylation of different RNAs directly or indirectly mediates tumour immunity (Li et al., 2017; Liu et al., 2019). N7-methylguanosine (m7G), the methylation of the seventh N of RNA guanine, is type of RNA methylation. Jieyi's study showed that m7G promotes lung cancer progression both in mRNA and tRNA (Ma et al., 2021). Additionally, another study indicated that m7G tRNA modification can enhance the translation of oncogenic mRNA and promote the progression of intrahepatic cholangiocarcinoma (Dai et al., 2021). The above mentioned-studies mentioned that m7G is an oncogenic factor, but the expression and biological function of m7G regulating lncRNAs in cancer are unknown.

In this study, we screened the m7G-related lncRNAs found in CC based on the Cancer Genome Atlas (TCGA) database and constructed an m7G-related prognostic lncRNA model. The purpose was to probe the correlation between m7G-related lncRNAs and CC through bioinformatics analysis and a few experiments.

MATERIALS AND METHODS

Data Collection

Genes related to m7G were obtained prior reviews (Tomikawa, 2018) and GSEA websites (http://www.gsea-msigdb.org/gsea/login.jsp) (Damian and Gorfine, 2004), and genesets were selected from the GSEA website were "GOMF_M7G_5_PPPN_DIPHOSPHATASE_ACTIVITY", "GOMF_RNA_7_METHYLGUANOSINE_CAP_BINDING," and "GOMF_RNA_CAP_BINDING." The TCGA database (https://portal.gdc.cancer.gov/repository) was used to acquire RNA sequencing (RNA-seq) data and clinical characteristics (Liu et al., 2018) on 10 January 2022, including 473 tumour datasets and 41 normal datasets. All data were standardized to fragment per kilobase million (FPKM) (Conesa et al., 2016) values.

Filtering Prognostic m7G-Related lncRNAs

The "limma" (Ritchie et al., 2015), "ggalluvial," "ggplot2" and "dplyr" packages were utilized to filter m7G-related lncRNAs and draw the Sankey relational chart. $|\text{Pearson R}| > 0.4$ and $p < 0.001$ were the criteria for filtering using Pearson's correlation analysis. The "survival" package was used to select prognostic m7G-related lncRNAs with $p < 0.05$ using univariate Cox regression analysis (van Dijk et al., 2008). To show the results more vividly, the heatmap and forest plot were drawn by the "pheatmap" and "limma" packages. The degree of difference is marked: * if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis

For gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, prognostic m7G-related lncRNAs were chosen. GO analysis includes three parts: molecular function (MF), biological process (BP), and cellular component (CC). The "org.Hs.eg.db," "clusterProfiler," "ggplot2" and "enrichplot" packages were used for GO analysis and KEGG analyses, and a bubble chart was drawn. The criteria were FDR <0.05 or $p < 0.05$.

Consensus Clustering and Tumour Immune Analysis

For consensus clustering, prognostic m7G-related lncRNAs were chosen (Cristescu et al., 2015), and the "limma," and "ConsensusClusterPlu" packages (Wilkerson and Hayes, 2010) were used to subdivide all CC data into different groups. Kaplan-Meier curve (Kaplan and Meier, 1958) were used to draw survival disparities between subgroups, and heatmaps were generated to show the correlation between lncRNA expression levels and clinical features. The "survival," "survminer," and "pheatmap" packages were used for drawing.

We probed the differences in the expression of immune cells and immune checkpoints among subgroups. The CIBERSORT algorithm was used for immune scoring, and

the differences between clusters were evaluated. The packages "limma," "estimate," "reshape2," "ggplot2," "ggpubr," and "vioplot" were used for these analyses. The degree of difference was noted: * if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$.

Establishment of the Risk Model

We employed TCGA expression data files and clinical data files to investigate the predictive utility of prognostic m7G-related lncRNAs for clinical prognosis. Least absolute shrinkage and selection operator (LASSO) regression analysis (Bunea et al., 2011) {risk score = Σ [Exp (lncRNA) × coef (lncRNA)]} was used to create a predictive signature for m7G-related lncRNAs, the regression coefficient is coef (lncRNA), and the corresponding expression of the included lncRNAs is Exp (lncRNA). According to a 7:3 ratio, all tumour samples were randomly allocated into training and testing groups. According to the median value of the risk of all patients, the samples are sorted into two categories: high-risk and low-risk. The "survival," "caret," "glmnet," "survminer," and "timeROC" packages were used in this analysis.

After constructing the risk model, we researched the survival risk of the two groups. Survival curves, receiver operating characteristic (ROC) curves (Hanley and McNeil, 1982), heatmaps, and the areas under the time-dependent ROC curves (AUCs) were rendered by the "survival," "timeROC," "survminer," and "pheatmap" packages.

Independent Prognostic Analysis, Nomogram

We researched all of the tumour samples' clinical features. In a Cox regression-based univariate and multivariate independent prognostic analysis, the "survival" package was used to determine whether clinical characteristics (age, sex, and TNM stage) and risk score could be utilized as independent prognostic variables. A nomogram (Iasonos et al., 2008) was used to indicate the prognosis of specific individuals. The efficiency of prediction was assessed using a calibration curve. The terms "survival," "regplot," and "rms" were used in this process.

Principal Component Analysis and Prognostic Analysis of Clinical Features

The total gene expression profile, m7G-related genes, m7G-related lncRNAs, and risk model lncRNAs were all subjected to principal component analysis (PCA) to reduce dimensionality, identify the model, and show the high-dimensional data. We used the "limma" and "scatterplot3d" packages to perform this process. Expression levels and survival analysis of different clinicopathological characteristic subgroups were produced using the "ggpubr" and "limma" packages. This analysis is useful in confirming the accuracy of our risk model.

Tumour Mutational Burden, Model Comparison and Evaluation

Tumour mutational burden (TMB) reflects the frequency of gene mutations in tumour tissue and is assumed to be related to tumour immunity. To display the TMB results of patients in the high-risk and low-risk groups, we analysed the difference and survival of TMB between the high- and low-risk groups in TCGA. The "maftools," "limma," "ggpubr," "survival," and "survminer" packages were utilized.

We drew multi-index ROC curves and concordance index curves (C-index curve) in all tumour samples to examine the model's efficacy (including risk, age, gender, and stage). The "dplyr," "survival," "rms," "pec," "survminer," and "timeROC" packages were implemented for this analysis. At the same time, we selected two published articles that constructed the lncRNA risk model in colon cancer to compare with our model. We compare the ROC values between these models and draw the relevant AUC curves. The terms "limma," "survival," "survminer," and "timeROC" were utilized for this comparison. Of note, this comparison used the AUC values of all samples and they were not divided into training group samples or testing group samples.

Tumour Immune Analysis

The CIBERSORT algorithm was utilized in this study to investigate the relationship between the risk score and immune cells/function. The "reshape2," "ggpubr," and "limma" packages were applied for immune cell analysis; the "GSEABase," "limma," "GSVA," "reshape2," and "ggpubr" packages were applied for immune function analysis, and the relevant boxplots and the bar chart were drawn to present the results more clearly. According to the median of the levels of immune cell enrichment or immune function in all patients, the samples were separated into two groups: those with high immune cell/function and those with low immune cell/function. We analysed the survival differences between the high and low cell/function groups. The "survminer," "survival," and "limma" packages were implemented. The degree of difference was noted: * if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$.

Prediction of Potential Therapeutic Compounds

We evaluated the IC50 values of the compounds retrieved from the GDSC website (https://www.cancerrxgene.org/) to predict probable compounds that may be employed in CC treatment. The "pRRophetic," "limma," "ggpubr," and "ggplot2" packages were used to predict the chemicals that may be employed for CC treatment.

Quantitative Real-Time Polymerase Chain Reaction

The normal human colonic epithelial cell line NCM460, as well as the human CC cell lines HT29 and SW620, were supplied by the Chinese Academy of Sciences' Shanghai Cell Bank. The TransZol

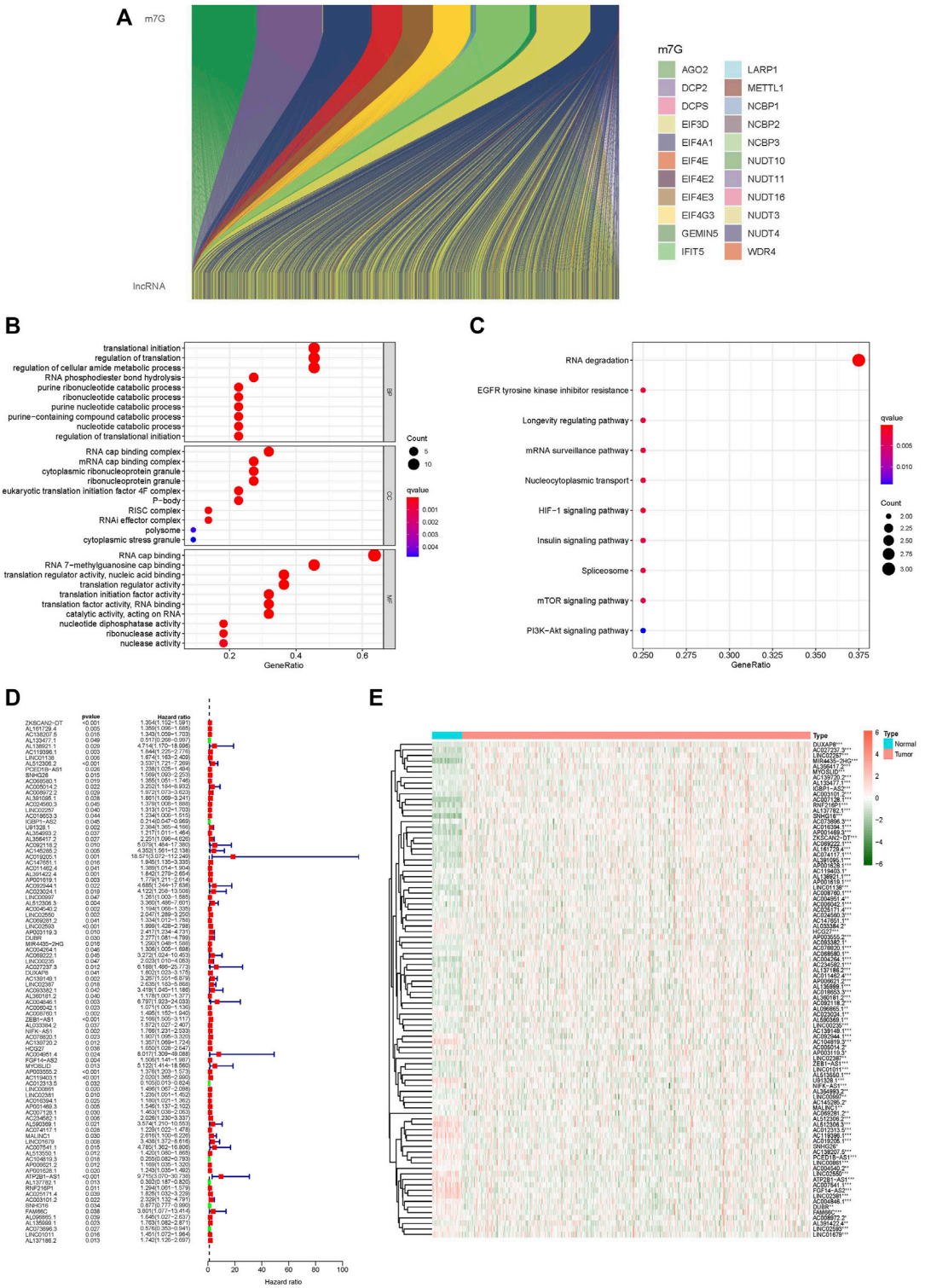


FIGURE 1 | GO and KEGG analysis of m7G-related lncRNAs, differentially expressed lncRNAs. **(A)** Sankey relational diagram for m7G genes and m7G-related lncRNAs, **(B)** GO analysis of m7G-related lncRNAs, **(C)** KEGG analysis of m7G-related lncRNAs, **(D)** Forest plot of differentially expressed m7G-related lncRNAs, **(E)** Heatmap of differentially expressed m7G-related lncRNAs. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Up Plus RNA Kit was used to extract and purify total RNA (Transgen, Beijing, PRC). For reverse transcription, the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen, Beijing, PRC) and T100 Thermal Cycler (BIO-RAD, United States) were employed. Quantitative real-time polymerase chain reaction (qRT–PCR) was performed using PerfectStart Green qPCR SuperMix (Transgen, Beijing, PRC) and CFX Connect Optics Module (BIO-RAD, United States). All experimental procedures were carried out following the product manual's instructions. Considering that the expression of many risk lncRNAs was very low, we chose four lncRNAs that were relatively easy to obtain results for qRT–PCR verification.

For PCR amplification, the primer sequences were as follows:
NIFK-AS1, forward: 5′-TTGGGTCTTCGAAAGTGCTG-3′,
reverse: 5′-ACGCTCCAAAACACTTTCCG-3';
RNF216P1, forward: 5′-GGCCAGCCAAGATGAGACAA-3′,
reverse: 5′-TCAGCAGCTTGGATGAAGCA-3';
ZEB1-AS1, forward: 5′-GGTTTCCTTCCTGCTTCCCA-3′,
reverse: 5′-ACTCCGGTCACGTTTCAGTT-3';
ZKSCAN2-DT, forward: 5′-TCTGGCGGAAGTATCTGTGC-3′,
reverse: 5′-AGCACCAGAAGAGAGCAAGC-3';
GAPDH, forward: 5′-CCCACTCCTCCACCTTTGAC-3′,
reverse: 5′-CCACCACCCTGTTGCTGTAG-3';

GAPDH was utilized as an internal control and each sample was reproduced three times. The relative expression levels were determined using the 2-ΔΔCt method. T-tests were used to compare the expression of NIFK-AS1, RNF216PA, ZEB1-AS1, and ZKSCAN2-DT (mean ± SEM). The graphs were made with GraphPad Prism (version 8.0.2). The degree of difference was noted: * if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$.

Statistical Analysis

R software (version 4.1.0) was used for statistical analysis and result visualization. The differential expression was authenticated using the Benjamini–Hochberg technique. The mRNA levels of pyroptosis-related lncRNAs were determined using the Mann–Whitney U test. The differences between the two groups were determined using student's *t*-test. The chi-square test was used to compare the categorization variables in the training and testing tests. The Pearson correlation test was used to analyse the relationship between subtypes, clinicopathological variables, risk score, immunological check inhibitors, and immune infiltration levels. For survival analysis, the Kaplan–Meier technique was used, along with a two-sided log-rank test.

RESULT

Prognosis-Related lncRNAs With Coexpression of m7G

We identified 1,627 lncRNAs in CC that coexpressed m7G-related genes (|Pearson R| > 0.4 and $p < 0.001$) (**Figure 1A**). Following that, 90 lncRNAs associated with prognosis were uncovered using univariate cox analysis (**Figures 1D,E**): ZKSCAN2-DT, AL161729.4, AC138207.5, AL133477.1, AL138921.1, AC119396.1, LINC01138, AL512306.2, PCED1B-AS1, SNHG26, AC068580.1, AC005014.2, AC008972.2, AL391095.1, AC024560.3, LINC02257, AC018653.3, IGBP1-AS2, U91328.1, AL354993.2, AL356417.2, AC092118.2, AC145285.2, AC019205.1, AC147651.1, AC011462.4, AL391422.4, AP001619.1, AC092944.1, AC023024.1, LINC00997, AL512306.3, AC004540.2, LINC02550, AC069281.2, LINC02593, AP003119.3, DUBR, MIR4435-2HG, AC004264.1, AC069222.1, LINC00235, AC027237.3, DUXAP8, AC139149.1, LINC02387, AC093382.1, AL360181.2, AC004846.1, AC006042.1, AC008760.1, ZEB1-AS1, AL033384.2, NIFK-AS1, AC078820.1, AC139720.2, HCG27, AC004951.4, FGF14-AS2, MYOSLID, AP003555.2, AC119403.1, AC012313.5, LINC00861, LINC02381, AC016394.1, AP001469.3, AC007128.1, AC234582.1, AL590369.1, AC074117.1, MALINC1, LINC01679, AC007541.1, AL513550.1, AC104819.3, AP006621.2, AP001628.1, ATP2B1-AS1, AL137782.1, RNF216P1, AC025171.4, AC003101.2, SNHG16, FAM66C, AL096865.1, AL135999.1, AC073896.3, LINC01011, and AL137186.2.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis

We performed GO and KEGG analyses of m7G-related lncRNAs. GO analysis included CC and MF of BP, and the significant correlations of lncRNAs were analysed at the cellular constituent, molecular function, and biochemical pathway levels in GO analysis. The obtained results revealed that m7G-related lncRNAs were primarily related to a change in the transcription and translation processes in cellular constituents, the function of m7G and the translation process in molecular function, and protein translation and nucleotide metabolism in biological processes (**Figure 1B**). KEGG analysis helped us to understand which pathways m7G-related lncRNAs are mostly associated with in colon cancer, and the outcome revealed that m7G-related lncRNAs were primarily associated with RNA regulation (**Figure 1C**).

Consensus Cluster Analysis of Prognostic m7G-Related lncRNAs

The correlation between 90 prognostic m7G-related lncRNA expression levels and subgroups in all CC samples was investigated using consensus cluster analysis. From the figure of increasing the clustering variable (k) from 2 to 9, there were the highest intragroup correlations and the lowest intergroup correlations when the clustering variable (k) was set to 2. (**Figure 2A**; **Supplementary Figures S2A–C**). The survival analysis revealed a substantial disparity in survival probability between Clusters 1 and 2, with patients in Cluster 2 having a better prognosis than those in Cluster 1. (**Figure 2C**). The expression levels and clinicopathological heatmap between the two clusters revealed that the tumour stage differed markedly between the two clusters (**Figure 2B**). The obtained results revealed that there were significant disparities between the two subgroups by analysing the difference in the expression levels of immune checkpoints

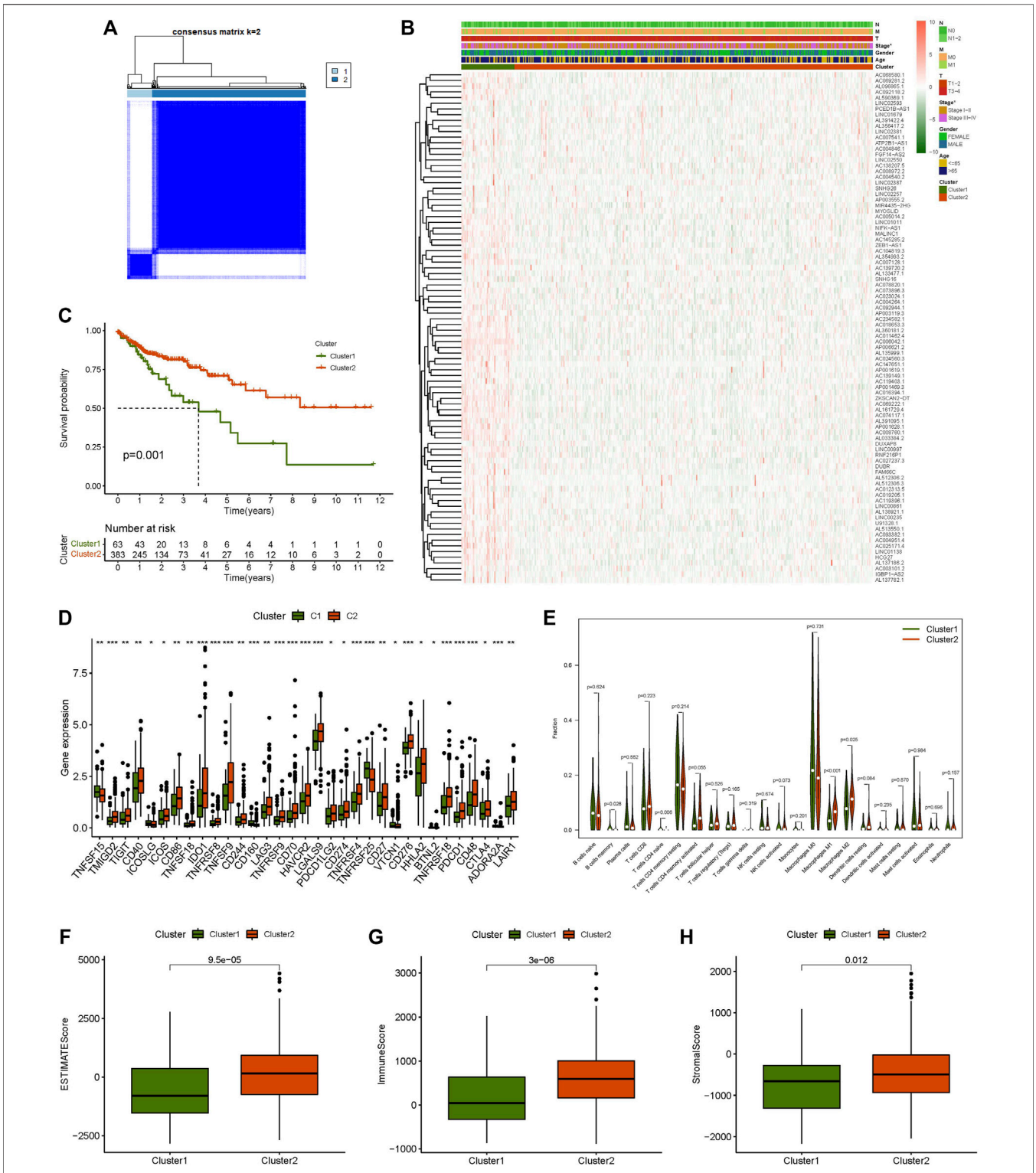


FIGURE 2 | Consensus clustering analysis and immune correlation analysis. **(A)** Consensus clustering matrix for k = 2, **(B)** Heatmap of m7G-related lncRNAs expression and clinicopathologic features in clusters 1 and cluster 2, **(C)** Kaplan-Meier curves of overall survival (OS) in clusters 1 and cluster 2, **(D)** Expression level of immune checkpoint in clusters 1 and cluster 2, **(E)** Relative contents of different immune cells in cluster 1 and cluster 2, **(F)** ESTMATE score in clusters 1 and cluster 2, **(G)** Immune score in clusters 1 and cluster 2, **(H)** Stromal score in clusters 1 and cluster 2. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

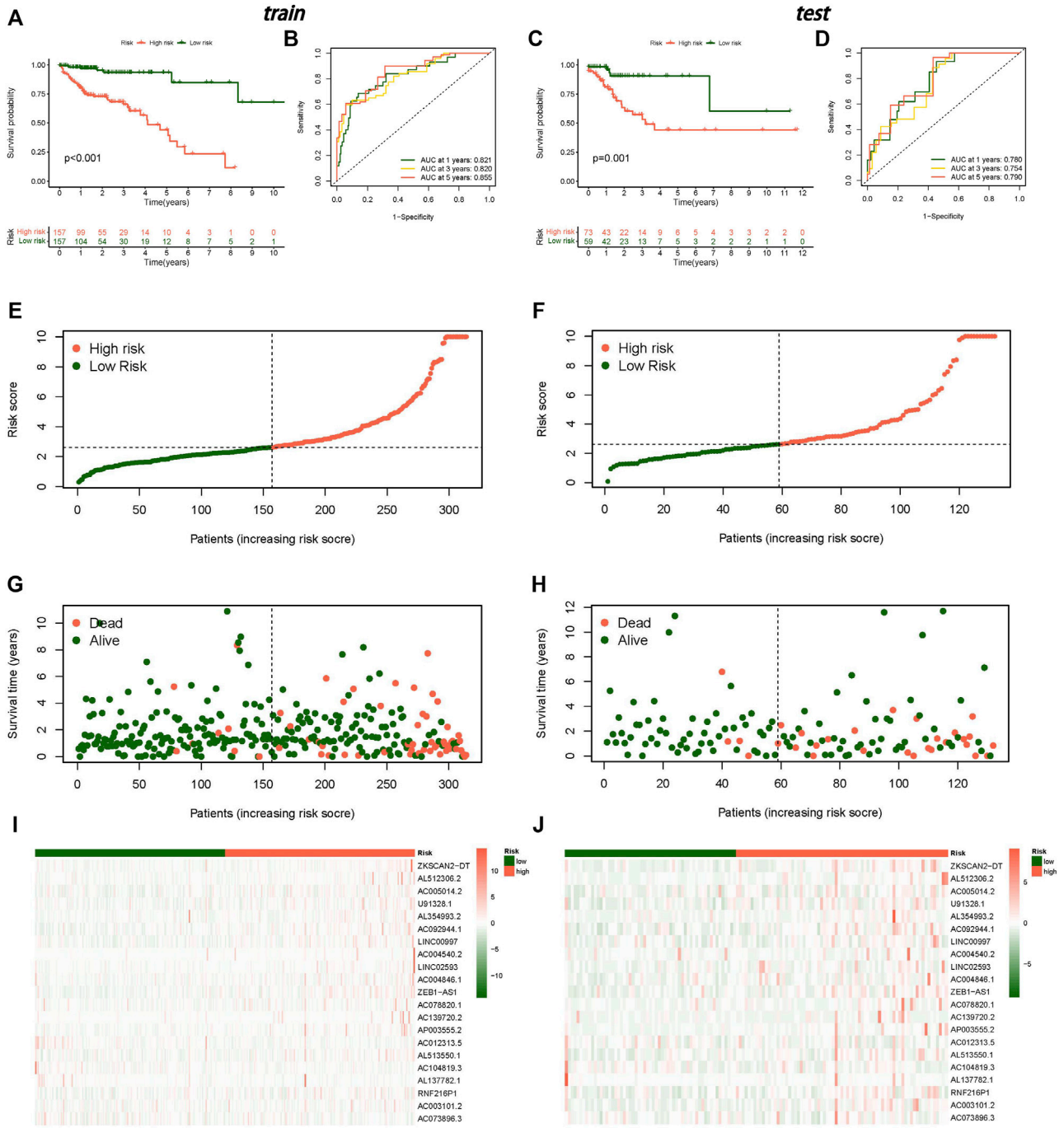


FIGURE 3 | Establishment of the risk model. **(A)** Kaplan-Meier curve for OS in the training group, **(B)** ROC curve in the training group, **(C)** Kaplan-Meier curve for OS in the testing group, **(D)** ROC curve in the testing group, **(E)** Risk score distribution in the training group, **(F)** Risk score distribution in the testing group, **(G)** OS statu in the training group, **(H)** OS statu in the testing group, **(I)** Heatmap of risk lncRNAs expression in the training group, **(J)** Heatmap of risk lncRNAs expression in the testing group.

(including C10orf54, CD200, TNFSF18, CD160, ADORA2A, TNFRSF25, CD28, LGALS9, CD244, TNFSF14, CD40, IDO2, TNFSF15, LAIR1, CD86, ICOSLG, CD70, BTNL2, PDCD1LG2, ICOS, IDO1, TNFSF4, CTLA4, NRP1, LAG3, TNFRSF9, TNFRSF18, CD80, TIGIT, TNFSF9, KIR3DL1, CD200R1, PDCD1, TNFRSF8, TNFRSF14, CD276, TNFRSF4, HHLA2, CD274, CD27, HAVCR2, BTLA, TMIGD2, CD48, CD40LG, VTCN1, and CD44) (**Figure 2D**). The results of multiple analyses of immune cell content indicated that there were remarkable differences

in the content of M1 macrophages, M2 macrophages, and naive CD4 T cells between the two subgroups (**Figure 2E**). Simultaneously, there was a statistically significant difference in immunological score between the two subgroups; Cluster 2 had a significantly higher score (ESTMATE score, stromal score, and immune score) than Cluster 1 (**Figures 2F–H**). These findings suggest that m7G-related lncRNAs have a substantial connection with tumour immunity in CC, implying that m7G-related lncRNAs would provide a novel reference marker for CC immunotherapy.

Establishment of Risk Model

All tumour samples were randomly divided into training and testing groups in a 7:3 ratio. We established a LASSO regression model and selected 21 lncRNAs (including ZKSCAN2-DT, AL512306.2, AC005014.2, U91328.1, AL354993.2, AC092944.1, LINC00997, AC004540.2, LINC02593, AC004846.1, ZEB1-AS1,

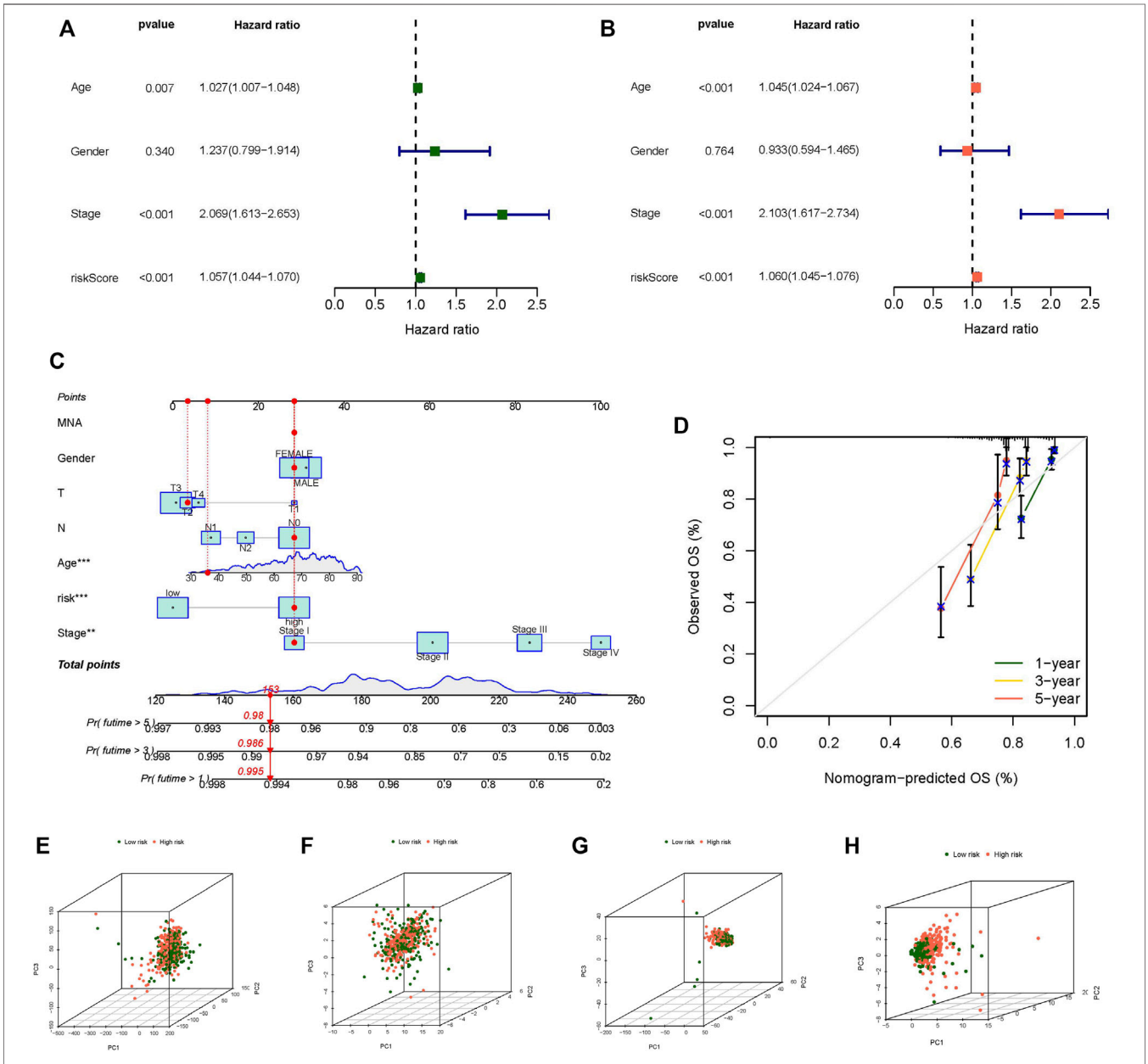


FIGURE 4 | Independent prognostic analysis, Nomogram and principal component analysis. **(A)** Univariate independent prognostic analysis in all tumor samples, **(B)** Univariate independent prognostic analysis in all tumor samples, **(C)** The nomogram predicts the probability of the 1-, 3-, and 5-years OS, **(D)** The calibration plot of the nomogram predicts the probability of the 1-, 3-, and 5-years OS, **(E)** PCA of all genes, **(F)** PCA of m7G genes, **(G)** PCA of m7G-related lncRNAs, **(H)** PCA of risk lncRNAs. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

AC078820.1, AC139720.2, AP003555.2, AC012313.5, AL513550.1, AC104819.3, AL137782.1, RNF216P1, AC003101.2, and AC073896.3) as signature factors from 90 prognostic m7G-related lncRNAs (**Supplementary Figures S2D, E**). The following formula was used to obtain the risk score: risk score = (0.176943613 * ZKSCAN2-DT exp.) + (0.933836642 * AL512306.2 exp.) + (0.326054032 * AC005014.2 exp.) + (0.111829011 *U91328.1 exp.) + (0.064972028 * AL354993.2 exp.) + (1.771428152 * AC092944.1 exp.) + (0.105833265 * LINC00997 exp.) + (0.041148055 * AC004540.2 exp.) + (0.309526196 * LINC02593 exp.) + (0.771493133 * AC004846.1 exp.) + (0.027471189 * ZEB1-AS1 exp.) + (0.430076387 * AC078820.1 exp.) + (0.162469055 * AC139720.2 exp.) + (0.225983414 * AP003555.2 exp.) + (−2.028167065 * AC012313.5 exp.) + (0.255834323 * AL513550.1 exp.) + (−0.568872124 * AC104819.3 exp.) + (−0.402855251 * AL137782.1 exp.) + (0.122413149 * RNF216P1 exp.) + (0.47174027 * AC003101.2 exp.) + (−0.247170433 * AC073896.3 exp.). We classified the risk score as high risk or low risk based on the median risk score. All high-risk samples were defined as high-risk groups, while low risk-score samples were defined as low-risk groups.

Regardless of whether the patients were in the training or testing groups, the survival analysis revealed that the survival probability of patients in the low-risk group was greater than that of patients in the high-risk group, and there was a statistically significant difference. (**Figures 3A,C**). The AUCs of the training and testing cohorts were 0.821 and 0.780 after 1 year, 0.820 and 0.754 at 3 years, and 0.855 and 0.790 after 5 years, respectively (**Figures 3B,D**), which indicates that the risk model performed well in prediction. The risk curve demonstrated that the death rate of patients rose with the rise of the risk score in both the training and testing groups (**Figures 3E–H**). The heatmap revealed a significant variation in the expression levels of 21 lncRNAs between the high-risk and low-risk groups (**Figures 3I,J**). The outcomes of univariate and multivariate independent prognostic analyses revealed that the stage and risk score may be utilized as independent prognostic factors to assess the prognosis of CC patients (**Figures 4A,B**). These analyses indicated that our risk model has a strong capacity to predict patient prognosis and may provied new ideas for the diagnosis and therapy of CC in the future.

Nomogram and PCA Verification

We constructed a nomogram to predict the prognosis of individual patients with CC (**Figure 4C**), and the calibration plot of the nomogram confirmed that our nomogram has good prediction ability (**Figure 4D**). We used PCA to evaluate the differences between the low- and high-risk groups in four expression profiles (m7G-related genes, total gene expression profiles, m7G-related lncRNAs, and the risk model classified by the expression profiles of the 21 m7G-related lncRNAs). Compared to the other three expression profiles, the separation between high-risk samples and low-risk samples was more clear in the 21 m7G-related lncRNA expression profile (**Figures 4E–H**). The findings revealed that 21 m7G-related lncRNAs had the strongest discriminating capacity, distinguishing between low- and high-risk groups rather effectively.

Model Grouping Verification of Clinicopathological Correlation

We discussed the association between the CC risk score and survival. The Kaplan-Meier curve for OS was generated in different clinicopathological characteristic subgroups to demonstrate the difference. Overall, patients in the low-risk group had a considerably greater survival rate than those in the high-risk group (**Figure 5**). The survival analysis of this clinicopathological characteristic subgroup also proved the effect of the risk score on the prognosis of CC patients and reaffirmed the credibility of the risk model established in our research.

Analysis of Tumour Mutational Burden

TMB represents the mutation rate of genes in the coding area, which is correlated with tumour formation and progression. Based on the TMB score produced from the somatic mutation data of TGCA, our study revealed that the TMB of the high-risk group was greater than that of the low-risk group (**Figures 6A,B**). The AUCs of the TMB score were 0.557 after 1 year, 0.582 at 3 years, and 0.491 after 5 years (**Figure 6C**). Compared with the risk score, the AUC value of the TMB score was lower. On the other hand, the difference between the high-risk and low-risk groups was extremely minor, and there was no significant variation in TMB between the two groups (**Figure 6D**). According to the TMB score, we divided all samples into high- and low-mutation groups and conducted a survival analysis on them combined with the grouping of high- and low-risk groups. The results again was confirmed that TMB did not affect the prognosis of patients with CC, but the predictive effect of the risk score was significant (**Figures 6E,F**). These results showed that the model's prediction ability was unquestionably higher than that of TMB.

Verification of the Credibility of the Risk Model

We created a concordance index curve and a multi-index ROC curve to validate the credibility of the risk model developed in our study. By comparing the AUC values among the various indicators (age, gender, stage, and risk score), we discovered that the AUC value of the risk score was the highest, at 0.809, followed by stage at 0.705, age at 0.609, and gender at 0.477 (**Figures 6G,H**). We selected two published articles about lncRNA model prediction in colon cancer and compared the AUC value of the published model and our risk model. The outcome indicated that the AUC value of the model researched by Xing (Xing et al., 2021) and Zhang (Zhang et al., 2021) was lower than that in our study (**Figure 6I**). This result shows that our approach is capable of accurately predicting the prognosis of CC patients.

Immune Correlation Analysis

We examined the correlation between the risk model and immune cells/function. The filtration standard was set at

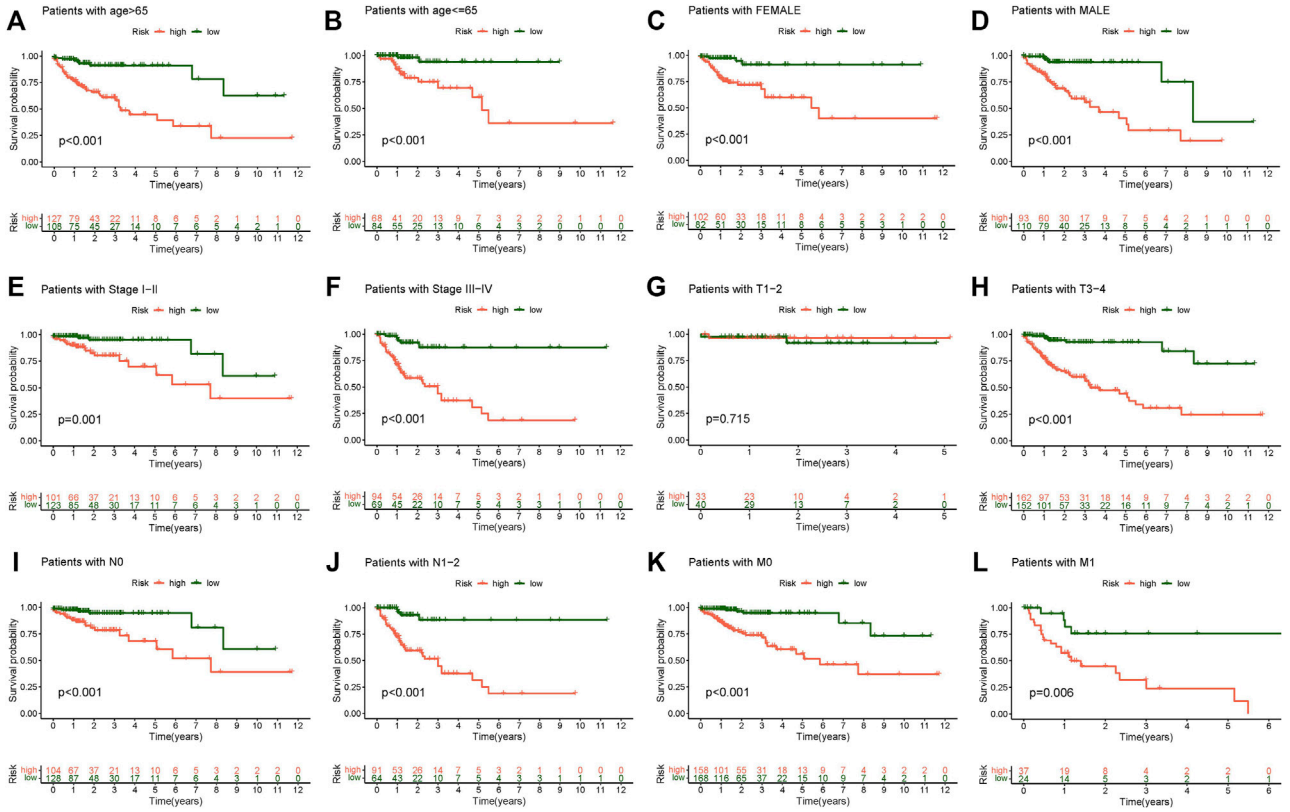


FIGURE 5 | Kaplan-Meier survival subgroup analysis for differential clinicopathological features in high and low-risk scores group. **(A)** Patients with age >65, **(B)** Patients with age ≤65, **(C)** Patients with female, **(D)** Patients with male, **(E)** Patients with stage I-II, **(F)** Patients with stage III-IV, **(G)** Patients with stage T1-2, **(H)** Patients with stage T3-4, **(I)** Patients with stage N0, **(J)** Patients with stage N1-2, **(K)** Patients with stage M0, **(L)** Patients with stage M1.

$p < 0.05$. The obtained result, where significant difference existed in the proportion of both resting dendritic cells and eosinophils between the two risk groups, indicated that the risk score was clearly correlated with resting dendritic cells and eosinophils (**Figure 7A**). A bar graph was also generated to display the proportion of each immune cell in the high- and low-risk categories (**Figure 7B**). The survival analysis of the high and low immune cell groups was researched: resting dendritic cells, naive B cells, M1 macrophages, eosinophils, resting memory T CD4 cells, plasma cells, and regulatory T cells (Tregs) showed significant differences (**Figures 7C–I**). The prognosis of the low-score group was much better than that of the high-score group in naive B cells, M1 macrophages, plasma cells, T regulatory cells (Tregs), and resting memory CD4 T cells. In resting dendritic cells and eosinophils, the high score group had a higher survival probability than the low score group.

In addition, immune function analysis showed that a strong relationship existed between the risk score and immune-related functions of parainflammation and Th2 cells (**Figure 8A**). Survival examination of several immune functions revealed that APC coinhibition, APC costimulation, DCs, iDCs, pDCs, T-cell coinhibition, T-cell costimulation, T helper cells, Th1 cells, Th2 cells, TILLs, Tregs, and Type II IFN reponse differed significantly between the high- and low-score groups (**Figure 8B–N**). The low score group had a greater survival rate than the high score group in the Type II IFN reponse function, while other functions were the opposite. In the future, these investigations may serve as a reference for tailored therapy of CC patients.

Prediction of Potential Therapeutic Drugs

The sensitivity to five chemicals (AG.014699, ABT.263, AS601245, AP.24534, and AZD.0530) in the low- and high-risk groups differed considerably in the prediction of prospective chemical drugs for the treatment of colon cancer (**Figures 9A–E**). The sensitivity of all chemicals in the high-risk group was greater than that of the low-risk group. This discovery may aid in the treatment of CC sufferers in the future.

Analysis of qRT–PCR

Four m7G-related lncRNAs (ZEB1-AS1, NIFK-AS1, RNF216P1, and ZKSCAN2-DT) were chosen. NCM460, HT29, and SW620 cells were used to examine these lncRNAs. The expression levels of these lncRNAs changed dramatically between tumour and normal cell lines, according to experimental evidence (**Figures 9F–I**). This result further validated the accuracy of our risk model.

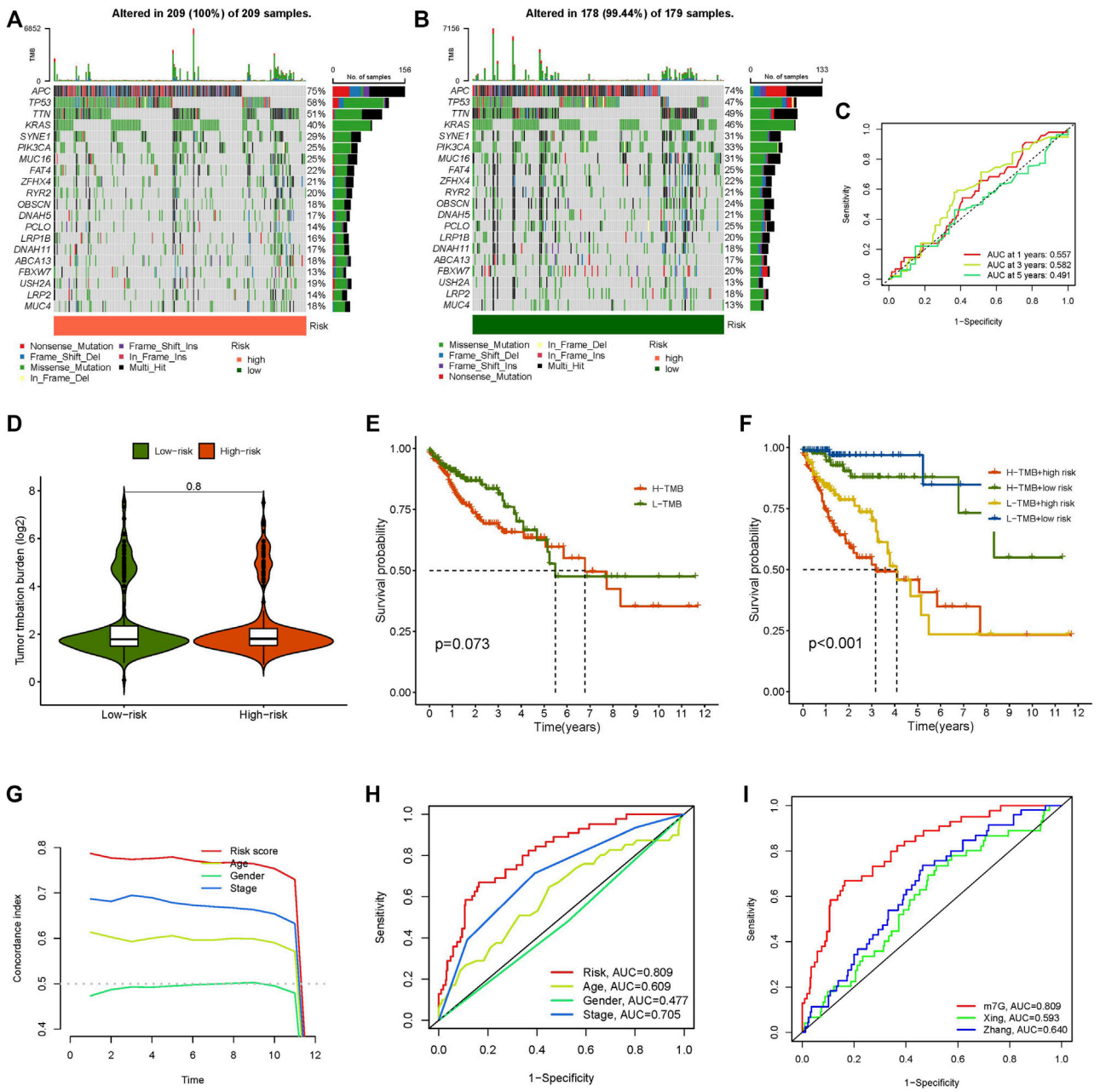


FIGURE 6 | Tumour mutational burden, concordance index curve, multi-index ROC curve and comparison of different lncRNAs models in colon tumors. **(A)** The waterfall plot displays the top 20 genes with high mutation frequency in the high-risk group, **(B)** The waterfall plot displays the top 20 genes with high mutation frequency in the low-risk group, **(C)** ROC curve of TMB score, **(D)** The difference of TMB in high- and low-risk groups, **(E)** Kaplan-Meier curve analysis of OS in high and low TMB score groups, **(F)** Kaplan-Meier curve analysis of OS is shown for patients classified according to the TMB and risk model, **(G)** concordance index curve, **(H)** Multi-index ROC curve, **(I)** ROC curves of different risk models.

DISCUSSION

Colon cancer (CC) is one of the most lethal human tumours. Over the past few years, with the progress of diagnosis and treatment technology, the total incidence of colon cancer has gradually decreased. However, the incidence of colon cancer among young patients is on the rise (Islami et al., 2021), and it is still an illness that human beings need to focus on (Biller and Schrag, 2021). Many lncRNAs have a regulatory function in the onset and progression of CC. Ren's research demonstrated that the lncRNA RPARP-AS1 may increase the migration, proliferation and invasion of CC cells *via* the PARP-AS1/miR-125a-5p axis (Ren et al., 2021). Cen investigated the impact of lncRNA IGFL2-AS1 in colon cancer, and the findings revealed

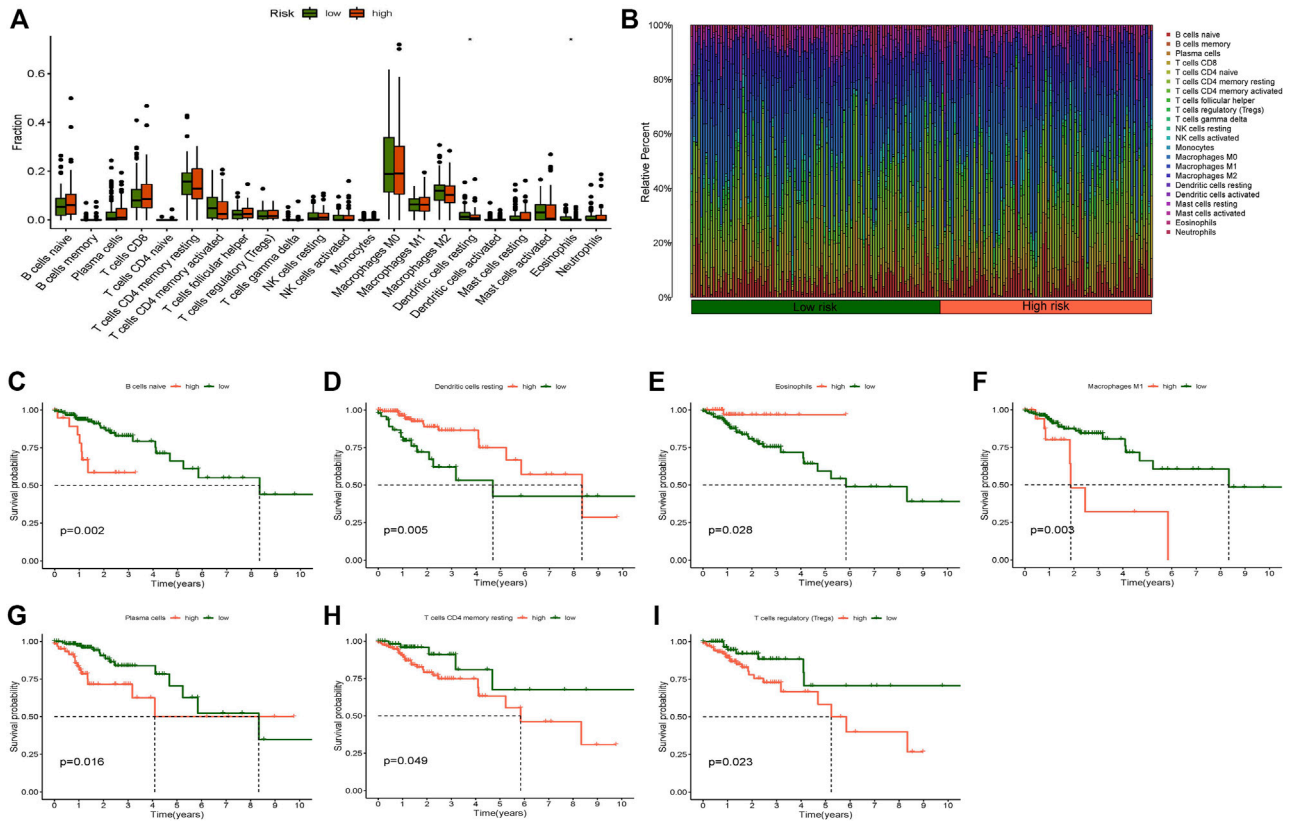


FIGURE 7 | Immune cell infiltration analysis. **(A)** Boxplot of immune cells difference in high- and low-risk group, **(B)** Relative percent of different immune cells in high- and low-risk group, **(C)** Kaplan-Meier survival analysis in naive B cells, **(D)** Kaplan-Meier survival analysis in dendritic cells resting, **(E)** Kaplan-Meier survival analysis in eosinophils, **(F)** Kaplan-Meier survival analysis in macrophages M1, **(G)** Kaplan-Meier survival analysis in plasma cells, **(H)** Kaplan-Meier survival analysis in T cells CD4 memory resting, **(I)** Kaplan-Meier survival analysis in T cells regulatory (Tregs). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

that lncRNA IGFL2-AS1 increased colon cancer cell migration, proliferation, and invasion (Cen et al., 2021). Professor Du confirmed that the lncRNA ELFN1-AS1 promotes colon cancer cell proliferation and invasion by modulating the miR-191-5p/SATB1 axis in colon cells (Du et al., 2020). Progressively increasing research has demonstrated that lncRNAs have a crucial regulatory function in CC; however, the precise mechanism is still being debated.

RNA methylation governs practically all aspects of RNA processing, and it is crucial in controlling gene expression, mRNA stability, and homeostasis. 7-Methylguanosine (m7G) RNA methylation is a recently identified type of RNA methylation that is expected to have a key function in malignancies (Zhou et al., 2021). Orellana found that METTL1-mediated tRNA modification promotes the expression of growth-promoting proteins by reshaping the "translation group" of mRNA, thus driving carcinogenic transformation (Orellana et al., 2021). After the completion of the experiment, Liu confirmed that the overexpression of METTL1 suppressed HMGA2 by upregulating let-7E miRNA, thus inhibiting the progression of CC (Liu et al., 2020). Interestingly, the methylation of RNA is intimately connected to the expression of lncRNAs, which seems to mean that m7G may affect tumorigenesis by regulating the expression of lncRNAs. However, this aspect is rarely mentioned in published studies, and up to this point, the mechanism is not totally definite. In our research, m7G-related lncRNAs were separated into distinct subgroups for the first time, and prognostic markers were constructed to comprehensively investigate the relationship between immune cell infiltration, tumour microenvironment, and m7G-related lncRNAs. We anticipate that these findings will be used to guide future clinical diagnosis and therapy of CC.

We obtained the expression data files and clinical data files for CC from the TCGA database for this study, and 29 genes associated with m7G were gathered from the published literature and GSEA website. Coexpression and univariate Cox analysis were used to identify 90 prognostic m7G-related lncRNAs. We subsequently conducted GO and KEGG analyses, which showed that m7G-related lncRNAs are mainly related to cell transcription and translation. Two subgroups were identified by consensus cluster analysis, and the survival analysis revealed disparities in survival between the two groupings. The difference between the two clusters was validated by immunological score, immune checkpoint expression level, and immune cell analysis. Cluster 2 showed a greater amount

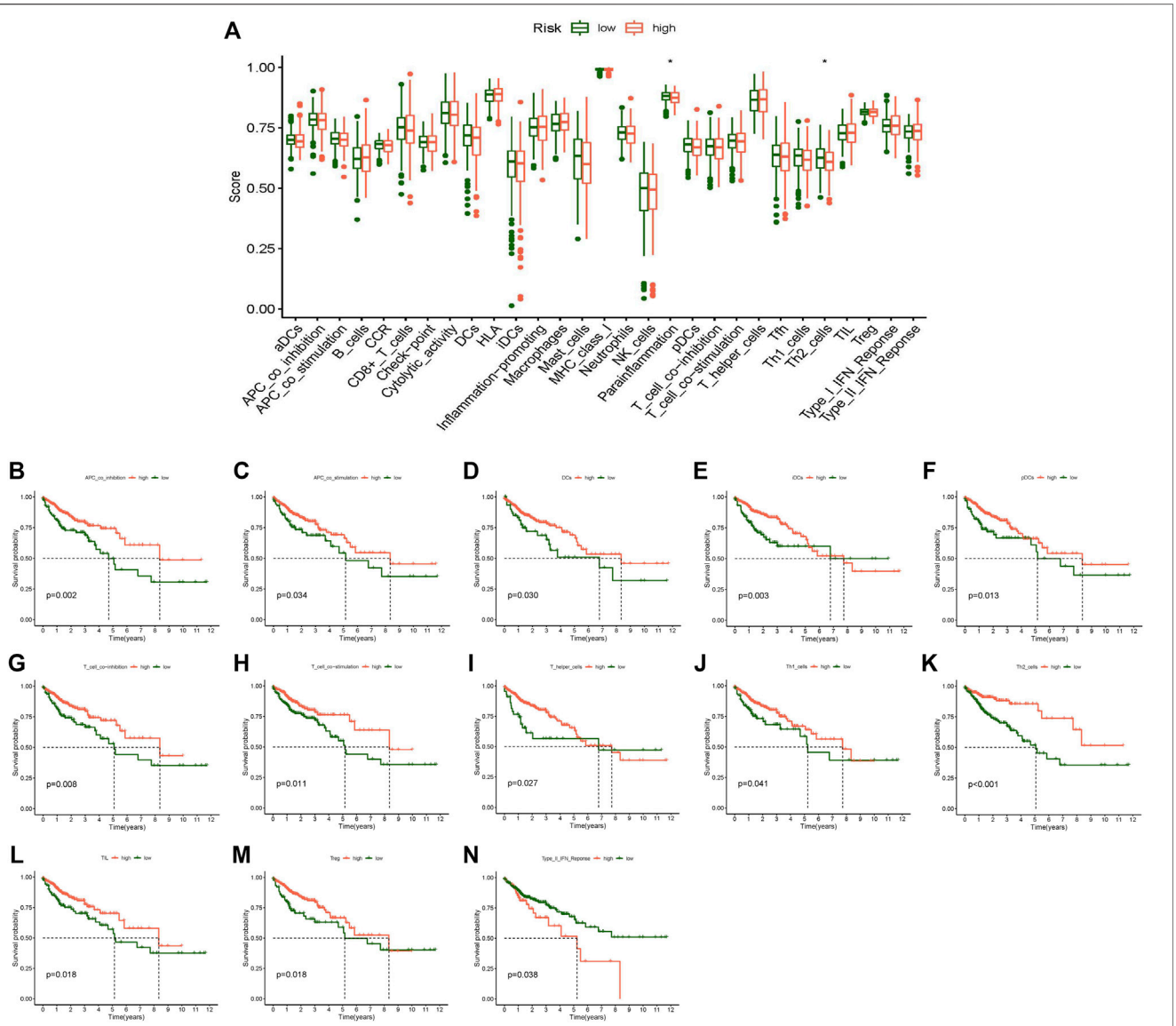


FIGURE 8 | Immune-related function analysis. **(A)** Boxplot of immune-related function score difference in high- and low-risk group, **(B)** Kaplan-Meier survival analysis in APC co-inhibition, **(C)** Kaplan-Meier survival analysis in APC co-stimulation, **(D)** Kaplan-Meier survival analysis in DCs, **(E)** Kaplan-Meier survival analysis in iDCs, **(F)** Kaplan-Meier survival analysis in pDCs, **(G)** Kaplan-Meier survival analysis in T cell co-inhibition, **(H)** Kaplan-Meier survival analysis in T cell co-stimulation, **(I)** Kaplan-Meier survival analysis in T helper cells, **(J)** Kaplan-Meier survival analysis in Th1 cells, **(K)** Kaplan-Meier survival analysis in Th2 cells, **(L)** Kaplan-Meier survival analysis in TIL, **(M)** Kaplan-Meier survival analysis in Treg, **(N)** Kaplan-Meier survival analysis in Type II IFN Reponse. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

of immune checkpoint expression and a higher immunological score, and its prognosis was much better than that of Cluster 1. These findings suggested that consensus clustering is associated with patient prognosis and may be correlated which the immunological microenvironment. To forecast the prognosis of CC patients, we established a risk model with 21 signature m7G-related lncRNAs (ZKSCAN2-DT, AL512306.2, AC005014.2, U91328.1, AL354993.2, AC092944.1, LINC00997, AC004540.2, LINC02593, AC004846.1, ZEB1-AS1, AC078820.1, AC139720.2, AP003555.2, AC012313.5, AL513550.1, AC104819.3, AL137782.1, RNF216P1, AC003101.2, and AC073896.3) by LASSO regression. The risk model is remarkably dependable and related to the prognosis of CC patients, according to ROC curves, Kaplan-Meier curves, and risk curves. PCA, independent prognostic analysis, survival analysis of subgroups with differing clinicopathological characteristics, multi-index ROC curve, survival analysis of TMB, c-index curve, and model comparison again tested the reliability of our model. Immune cell/function analysis showed that m7G-related lncRNAs may be related to tumour immunity. The significant difference in the ratio of resting dendritic cells and eosinophils between the two risk groups may be helpful according to Grisaru-Tal's and Wooster's research. The noteworthy difference of resting dendritic cells and eosinophils in distinct

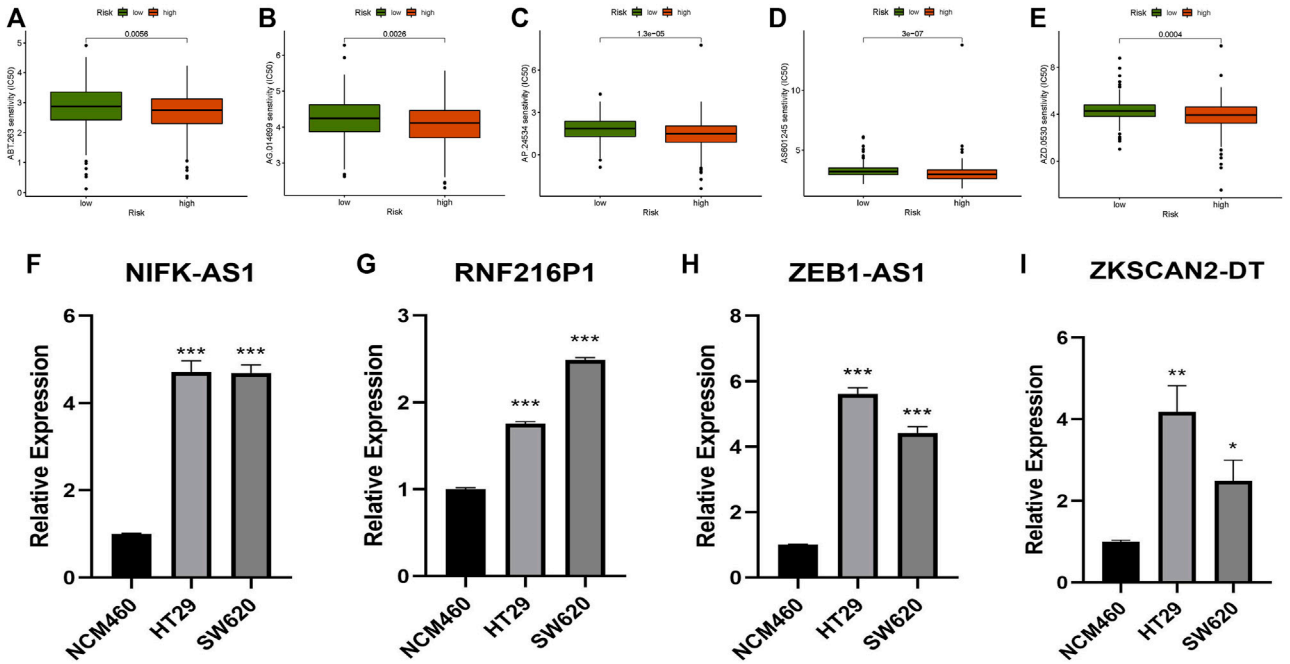


FIGURE 9 | Prediction of potentially therapeutic compounds and qRT–PCR. **(A)** ABT.263, **(B)** AG.014699, **(C)** AP.24534, **(D)** AS601245, **(E)** AZD.0530, **(F–I)** Relative expression of NIFK-AS1, RNF216P1, ZEB1-AS1, and ZKSCAN2-DT in NCM460, HT29, and SW620 cell lines (Mean ± SEM). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

risk groups may possess the potential to direct CC therapy in the future (Grisaru-Tal et al., 2020; Wooster et al., 2021). Furthermore, drug sensitivity analysis also provides a possible reference for the clinical therapy of CC, and the results of qRT–PCR confirm the trustworthiness of our research.

In published research, the lncRNA ZEB1-AS1 was considered to be tightly related to the onset and progression of many types of cancers. According to Jin's research, the positive reciprocal looping of HIF-1/ZEB1-AS1/ZEB1/HDAC1 leads to hypoxia-induced oncogenicity and PC metastasis (Jin et al., 2021). Siena confirmed that lncRNA ZEB1-AS1 can influence the invasiveness and phenotypic transformation of melanoma through epithelial to mesenchymal transition (EMT) (Siena et al., 2019). Wu believes that ZEB1-AS1 may fuel the profitability of colorectal cancer cells by saponifying miR-141-3p (Wu et al., 2020). LncRNA LINC00997 has been discovered to perform a regulatory role in tumours in recent years. Shi's research suggests that LINC00997 promotes colorectal cancer metastasis by targeting miR-512-3p (Shi et al., 2021). Chang believes that the LINC00997-STAT3-S100A11 axis potentially helps KIRC develop (Chang et al., 2019). These tumour-related lncRNAs are closely related to m7G. However, more m7G-related lncRNAs still lack relevant studies to confirm their relationship with tumours, and our findings may serve as a fresh reference for future studies. In our research, we firstly analysed the lncRNA related to m7G in CC by consensus clustering and established a risk model. Second, this work thoroughly investigated the association between m7G-related lncRNA prognostic markers and the tumour microenvironment, immune cell infiltration, and TMB, which may provide a new direction for future research into the predictive significance of m7G-related lncRNA markers in immunotherapy. Third, we projected several potential chemicals that may be effective for CC therapy in the future.

There are also some shortcomings in this research. First, our data source is relatively singlar, and there is a lack of more external data verification. This is mainly due to the lack of datasets with complete lncRNA sequencing data, such as TCGA databases. Second, we lack additional molecular biology experiments to verify our results. These problems are also the direction that we need to work on next.

CONCLUSION

In this study, we investigated the usefulness of m7G-related lncRNAs in predicting survival, the involvement of the tumour microenvironment and immune cell infiltration, the prospective regulation mechanism of m7G-associated lncRNAs, and the prediction of suitable drugs for the treatment of CC. 21 lncRNA characteristics related with m7G may predict CC patient survival and may be beneficial for customized cancer therapy in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

ZJ and JX: funding acquisition and project administration; LL: validation, investigation, resources, and original draft; YW: conceptualization, methodology, software, and original draft; WC: original draft; GZ: visualization and supervision; YL: review and editing; JY, LH, and PF: data curation and formal analysis.

FUNDING

This study was supported by grants from the Natural Science Foundation of Jiangxi Province (No. 20202BABL216051) and the National Natural Science Foundation (Nos. 81960503, 82103165).

ACKNOWLEDGMENTS

Thanks to the scholars who participated in this study for their contributions and the reviewers, thanks to the reviewers and editors for their suggestions on this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.892589/full#supplementary-material

REFERENCES

Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res.* 77 (15), 3965–3981. doi:10.1158/0008-5472.can-16-2634

Biller, L. H., and Schrag, D. (2021). A Review of the Diagnosis and Treatment of Metastatic Colorectal Cancer-Reply. *J. Am. Stat. Assoc.* 325 (23), 2405. doi:10.1001/jama.2021.6027

Bunea, F., She, Y., Ombao, H., Gongvatana, A., Devlin, K., and Cohen, R. (2011). Penalized Least Squares Regression Methods and Applications to Neuroimaging. *Neuroimage* 55 (4), 1519–1527. doi:10.1016/j.neuroimage.2010.12.028

Cen, X., Huang, Y., Lu, Z., Shao, W., Zhuo, C., Bao, C., et al. (2021). LncRNA IGFL2-AS1 Promotes the Proliferation, Migration, and Invasion of Colon Cancer Cells and Is Associated with Patient Prognosis. *Cancer Manag. Res.* Vol. 13, 5957–5968. doi:10.2147/cmar.s313775

Chang, Y., Li, N., Yuan, W., Wang, G., and Wen, J. (2019). LINC00997, a Novel Long Noncoding RNA, Contributes to Metastasis *via* Regulation of S100A11 in Kidney Renal Clear Cell Carcinoma. *Int. J. Biochem. Cell Biol.* 116, 105590. doi:10.1016/j.biocel.2019.105590

Chen, L., He, M., Zhang, M., Sun, Q., Zeng, S., Zhao, H., et al. (2021). The Role of Non-coding RNAs in Colorectal Cancer, with a Focus on its Autophagy. *Pharmacol. Ther.* 226, 107868. doi:10.1016/j.pharmthera.2021.107868

Chen, L.-L. (2016). Linking Long Noncoding RNA Localization and Function. *Trends Biochem. Sci.* 41 (9), 761–772. doi:10.1016/j.tibs.2016.07.003

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A Survey of Best Practices for RNA-Seq Data Analysis. *Genome Biol.* 17, 13. doi:10.1186/s13059-016-0881-8

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.-M., Ting, J. C., Wong, S. S., et al. (2015). Molecular Analysis of Gastric Cancer Identifies Subtypes Associated with Distinct Clinical Outcomes. *Nat. Med.* 21 (5), 449–456. doi:10.1038/nm.3850

Dai, Z., Liu, H., Liao, J., Huang, C., Ren, X., Zhu, W., et al. (2021). N(7)-Methylguanosine tRNA Modification Enhances Oncogenic mRNA Translation and Promotes Intrahepatic Cholangiocarcinoma Progression. *Mol. Cell* 81 (16), 3339–3355.e8. doi:10.1016/j.molcel.2021.07.003

Damian, D., and Gorfine, M. (2004). Statistical Concerns about the GSEA Procedure. *Nat. Genet.* 36 (7), 663. author reply 663. doi:10.1038/ng0704-663a

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., et al. (2016). YTHDF2 Destabilizes m(6)A-Containing RNA through Direct Recruitment of the CCR4-NOT Deadenylase Complex. *Nat. Commun.* 7, 12626. doi:10.1038/ncomms12626

Du, Y., Hou, Y., Shi, Y., Liu, J., and Li, T. (2020). Long Non-Coding RNA ELFN1-AS1 Promoted Colon Cancer Cell Growth and Migration *via* the miR-191-5p/Special AT-Rich Sequence-Binding Protein 1 Axis. *Front. Oncol.* 10, 588360. doi:10.3389/fonc.2020.588360

Fazi, F., and Fatica, A. (2019). Interplay between N (6)-Methyladenosine (m(6)A) and Non-coding RNAs in Cell Development and Cancer. *Front. Cell Dev. Biol.* 7, 116. doi:10.3389/fcell.2019.00116

Grisaru-Tal, S., Itan, M., Grass, D. G., Torres-Roca, J., Eschrich, S. A., Gordon, Y., et al. (2020). Primary Tumors from Mucosal Barrier Organs Drive Unique Eosinophil Infiltration Patterns and Clinical Associations. *Oncoimmunology* 10 (1), 1859732. doi:10.1080/2162402x.2020.1859732

Hanley, J. A., and McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143 (1), 29–36. doi:10.1148/radiology.143.1.7063747

He, R.-Z., Jiang, J., and Luo, D.-X. (2020). The Functions of N6-Methyladenosine Modification in lncRNAs. *Genes Dis.* 7 (4), 598–605. doi:10.1016/j.gendis.2020.03.005

Hong, S., Yan, Z., Song, Y., Bi, M., and Li, S. (2021). Down-regulation of lncRNA FEZF1-AS1 Mediates Regulatory T Cell Differentiation and Further Blocks Immune Escape in Colon Cancer. *Expert Rev. Mol. Diagn.* doi:10.1080/14737159.2022.2012157

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to Build and Interpret a Nomogram for Cancer Prognosis. *J. Clin. Oncol.* 26 (8), 1364–1370. doi:10.1200/jco.2007.12.9791

Islami, F., Ward, E. M., Sung, H., Cronin, K. A., Tangka, F. K. L., Sherman, R. L., et al. (2021). Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. *J. Natl. Cancer Inst.* 113, 1648–1669. doi:10.1093/jnci/djab131

Jin, Y., Zhang, Z., Yu, Q., Zeng, Z., Song, H., Huang, X., et al. (2021). Positive Reciprocal Feedback of lncRNA ZEB1-AS1 and HIF-1alpha Contributes to Hypoxia-Promoted Tumorigenesis and Metastasis of Pancreatic Cancer. *Front. Oncol.* 11, 761979. doi:10.3389/fonc.2021.761979

Kaplan, E. L., and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* 53, 457–481. doi:10.1080/01621459.1958.10501452

Karikó, K., Buckstein, M., Ni, H., and Weissman, D. (2005). Suppression of RNA Recognition by Toll-like Receptors: the Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* 23 (2), 165–175. doi:10.1016/j.immuni.2005.06.008

Lee, Y., Choe, J., Park, O. H., and Kim, Y. K. (2020). Molecular Mechanisms Driving mRNA Degradation by M(6)A Modification. *Trends Genet.* 36 (3), 177–188. doi:10.1016/j.tig.2019.12.007

Li, H.-B., Tong, J., Zhu, S., Batista, P. J., Duffy, E. E., Zhao, J., et al. (2017). m(6)A mRNA Methylation Controls T Cell Homeostasis by Targeting the IL-7/STAT5/SOCS Pathways. *Nature* 548 (7667), 338–342. doi:10.1038/nature23450

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* 173 (2), 400–e11. doi:10.1016/j.cell.2018.02.052

Liu, J., Zhang, X., Chen, K., Cheng, Y., Liu, S., Xia, M., et al. (2019). CCR7 Chemokine Receptor-Inducible Lnc-Dpf3 Restrains Dendritic Cell Migration by Inhibiting HIF-1alpha-Mediated Glycolysis. *Immunity* 50 (3), 600–615. doi:10.1016/j.immuni.2019.01.021

Liu, Y., Zhang, Y., Chi, Q., Wang, Z., and Sun, B. (2020). Methyltransferase-like 1 (METTL1) served as a tumor suppressor in colon cancer by activating 7-

methyguanosine (m7G) regulated let-7e miRNA/HMGA2 axis. *Life Sci.* 249, 117480. doi:10.1016/j.lfs.2020.117480

Ma, J., Han, H., Huang, Y., Yang, C., Zheng, S., Cai, T., et al. (2021). METTL1/WDR4-Mediated M(7)G tRNA Modifications and M(7)G Codon Usage Promote mRNA Translation and Lung Cancer Progression. *Mol. Ther.* 29 (12), 3422–3435. doi:10.1016/j.ymthe.2021.08.005

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (2019). Cancer Treatment and Survivorship Statistics, 2019. *CA A Cancer J. Clin.* 69 (5), 363–385. doi:10.3322/caac.21565

Ni, X., Ding, Y., Yuan, H., Shao, J., Yan, Y., Guo, R., et al. (2020). Long Non-coding RNA ZEB1-AS1 Promotes Colon Adenocarcinoma Malignant Progression via miR-455-3p/PAK2 axis. *Cell Prolif.* 53 (1), e12723. doi:10.1111/cpr.12723

Orellana, E. A., Liu, Q., Yankova, E., Pirouz, M., De Braekeleer, E., Zhang, W., et al. (2021). METTL1-mediated M(7)G Modification of Arg-TCT tRNA Drives Oncogenic Transformation. *Mol. Cell* 81 (16), 3323–3338.e14. doi:10.1016/j.molcel.2021.06.031

Ren, Y., Zhao, C., He, Y., Min, X., Xu, H., and Hu, X. (2021). RPARP-AS1/miR125a-5p Axis Promotes Cell Proliferation, Migration and Invasion in Colon Cancer. *Onco Targets Ther.* 14, 5035–5043. doi:10.2147/ott.s304494

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Shi, Z., Shen, C., Yu, C., Yang, X., Shao, J., Guo, J., et al. (2021). Long Non-coding RNA LINC00997 Silencing Inhibits the Progression and Metastasis of Colorectal Cancer by Sponging miR-512-3p. *Bioengineered* 12 (1), 627–639. doi:10.1080/21655979.2021.1882164

Siena, Á. D. D., Plaça, J. R., Araújo, L. F., de Barros, I. I., Peronni, K., Molfetta, G., et al. (2019). Whole Transcriptome Analysis Reveals Correlation of Long Noncoding RNA ZEB1-AS1 with Invasive Profile in Melanoma. *Sci. Rep.* 9 (1), 11350. doi:10.1038/s41598-019-47363-6

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tomikawa, C. (2018). 7-Methylguanosine Modifications in Transfer RNA (tRNA). *Int. J. Mol. Sci.* 19 (12), 4080. doi:10.3390/ijms19124080

van Dijk, P. C., Jager, K. J., Zwinderman, A. H., Zoccali, C., and Dekker, F. W. (2008). The Analysis of Survival Data in Nephrology: Basic Concepts and Methods of Cox Regression. *Kidney Int.* 74 (6), 705–709. doi:10.1038/ki.2008.294

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., et al. (2011). Reprogramming Transcription by Distinct Classes of Enhancers Functionally Defined by eRNA. *Nature* 474 (7351), 390–394. doi:10.1038/nature10006

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* 161 (6), 1388–1399. doi:10.1016/j.cell.2015.05.014

Wang, Q., Chen, C., Ding, Q., Zhao, Y., Wang, Z., Chen, J., et al. (2020). METTL3-mediated M(6)A Modification of HDGF mRNA Promotes Gastric Cancer Progression and Has Prognostic Significance. *Gut* 69 (7), 1193–1205. doi:10.1136/gutjnl-2019-319639

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wooster, A. L., Girgis, L. H., Brazeale, H., Anderson, T. S., Wood, L. M., and Lowe, D. B. (2021). Dendritic Cell Vaccine Therapy for Colorectal Cancer. *Pharmacol. Res.* 164, 105374. doi:10.1016/j.phrs.2020.105374

Wu, G., Xue, M., Zhao, Y., Han, Y., Li, C., Zhang, S., et al. (2020). Long Noncoding RNA ZEB1-AS1 Acts as a Sponge of miR-141-3p to Inhibit Cell Proliferation in Colorectal Cancer. *Int. J. Med. Sci.* 17 (11), 1589–1597. doi:10.7150/ijms.46698

Xiang, Y., Laurent, B., Hsu, C.-H., Nachtergaele, S., Lu, Z., Sheng, W., et al. (2017). RNA M(6)A Methylation Regulates the Ultraviolet-Induced DNA Damage Response. *Nature* 543 (7646), 573–576. doi:10.1038/nature21671

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., et al. (2016). Nuclear M(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* 61 (4), 507–519. doi:10.1016/j.molcel.2016.01.012

Xing, X.-L., Zhang, T., Yao, Z.-Y., Xing, C., Wang, C., Liu, Y.-W., et al. (2021). Immune-Related Gene Expression Analysis Revealed Three lncRNAs as Prognostic Factors for Colon Cancer. *Front. Genet.* 12, 690053. doi:10.3389/fgene.2021.690053

Yi, Y.-C., Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2020). Novel Insights into the Interplay between M(6)A Modification and Noncoding RNAs in Cancer. *Mol. Cancer* 19 (1), 121. doi:10.1186/s12943-020-01233-2

Zhang, S., Zhao, B. S., Zhou, A., Lin, K., Zheng, S., Lu, Z., et al. (2017). m(6)A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. *Cancer Cell* 31 (4), 591–606.e6. doi:10.1016/j.ccell.2017.02.013

Zhang, W., Fang, D., Li, S., Bao, X., Jiang, L., and Sun, X. (2021). Construction and Validation of a Novel Ferroptosis-Related lncRNA Signature to Predict Prognosis in Colorectal Cancer Patients. *Front. Genet.* 12, 709329. doi:10.3389/fgene.2021.709329

Zhao, X., Yang, Y., Sun, B.-F., Shi, Y., Yang, X., Xiao, W., et al. (2014). FTO-dependent Demethylation of N6-Methyladenosine Regulates mRNA Splicing and Is Required for Adipogenesis. *Cell Res.* 24 (12), 1403–1419. doi:10.1038/cr.2014.151

Zhou, W., Wang, X., Chang, J., Cheng, C., and Miao, C. (2021). The Molecular Structure and Biological Functions of RNA Methylation, with Special Emphasis on the Roles of RNA Methylation in Autoimmune Diseases. *Crit. Rev. Clin. Lab. Sci.* 59, 203–218. doi:10.1080/10408363.2021.2002256

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Wu, Chen, Li, Yu, Zhang, Fu, Huang, Xiong and Jie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fgene.2022.862179

An Autophagy-Related Gene Signature can Better Predict Prognosis and Resistance in Diffuse Large B-Cell Lymphoma

Xuan Zhou[1,2†], Ying-Zhi He[3†], Dan Liu[4†], Chao-Ran Lin[4], Dan Liang[1], Rui Huang[3] and Liang Wang[5*]*

[1]Second Clinical Medical College of Southern Medical University, Zhujiang Hospital of Southern Medical University, Guangzhou, China, [2]Department of Endocrinology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [3]Department of Hematology, Zhujiang Hospital of Southern Medical University, Guangzhou, China, [4]The First School of Clinical Medicine, Guangdong Medical University, Zhanjiang, China, [5]Department of Hematology, Beijing TongRen Hospital, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

Reviewed by:
Jian Yu,
Beihang University, China
Xin Wang,
Shandong Provincial Hospital, China

***Correspondence:**
Rui Huang
rachelchn@163.com
Liang Wang
wangliangtrhos@126.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: 25 January 2022
Accepted: 12 May 2022
Published: 30 June 2022

Citation:
Zhou X, He Y-Z, Liu D, Lin C-R, Liang D, Huang R and Wang L (2022) An Autophagy-Related Gene Signature can Better Predict Prognosis and Resistance in Diffuse Large B-Cell Lymphoma. Front. Genet. 13:862179. doi: 10.3389/fgene.2022.862179

Background: Diffuse large B-cell lymphoma (DLBCL) is a highly heterogeneous disease, and about 30%–40% of patients will develop relapsed/refractory DLBCL. In this study, we aimed to develop a gene signature to predict survival outcomes of DLBCL patients based on the autophagy-related genes (ARGs).

Methods: We sequentially used the univariate, least absolute shrinkage and selector operation (LASSO), and multivariate Cox regression analyses to build a gene signature. The Kaplan–Meier curve and the area under the receiver operating characteristic curve (AUC) were performed to estimate the prognostic capability of the gene signature. GSEA analysis, ESTIMATE and ssGSEA algorithms, and one-class logistic regression were performed to analyze differences in pathways, immune response, and tumor stemness between the high- and low-risk groups.

Results: Both in the training cohort and validation cohorts, high-risk patients had inferior overall survival compared with low-risk patients. The nomogram consisted of the autophagy-related gene signature, and clinical factors had better discrimination of survival outcomes, and it also had a favorable consistency between the predicted and actual survival. GSEA analysis found that patients in the high-risk group were associated with the activation of doxorubicin resistance, NF-κB, cell cycle, and DNA replication pathways. The results of ESTIMATE, ssGSEA, and mRNAsi showed that the high-risk group exhibited lower immune cell infiltration and immune activation responses and had higher similarity to cancer stem cells.

Conclusion: We proposed a novel and reliable autophagy-related gene signature that was capable of predicting the survival and resistance of patients with DLBCL and could guide individualized treatment in future.

Keywords: diffuse large B-cell lymphoma, autophagy-related genes, gene signature, prognosis, autophagy inhibitors

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) accounts for 30% of cases of non-Hodgkin lymphoma (NHL), making it the most common subtype of NHL among adults worldwide (Zhou et al., 2017). Evidence from biological and clinical studies has revealed that there is a striking degree of clinical, phenotypic, and molecular heterogeneity in DLBCL (Reddy et al., 2017). With the advent of the rituximab era, the treatment efficacy and survival status of DLBCL patients have been dramatically improved, and about 2/3 of DLBCL patients could achieve long-term survival. However, the majority of the remaining patients died of disease relapse or drug resistance (Coccaro et al., 2020). To date, the prognostic prediction of DLBCL patients mainly relies on the International Prognostic Index (IPI) and National Comprehensive Cancer Network-International Prognostic Index (NCCN-IPI) (Zhou et al., 2014). As there is lack of information about genes, patients with the same IPI score may still have different prognosis, and the IPI score cannot identify patients with a 5-year overall survival rate of less than 50% (Ruppert et al., 2020). In addition to the commonly used IPI score, the cell of origin according to the Hans model can also predict the patient's prognosis to a certain extent. Most studies have shown that the prognosis of patients with the germinal center B-cell subtype is better than that of the non-germinal center B-cell subtype (Hans et al., 2004). However, some research results are inconsistent with this conclusion, suggesting that the prognostic stratification ability of the Hans model needs to be further verified (Coutinho et al., 2013). With the rapid development of molecular biology in recent years, researchers have been trying to use sequencing and chip technology to stratify risks and optimize chemotherapy strategies for patients with different types of cancer (Tian et al., 2020).

Meanwhile, autophagy is also one of the breakthrough findings in the field of tumors. It is reported that autophagy is involved in several biological functions, such as apoptosis, immune response, maintenance of cancer stem cell, and drug resistance (Cufí et al., 2011; Sun et al., 2011; Jiang et al., 2019). On one hand, autophagy helps tumor cells to sustain cellular growth by degrading and recycling components of damaged or aged organelles (Katheder et al., 2017). On the other hand, autophagy can maintain the normal cell structure and metabolic stability by removing damaged organelles and DNA in the early stage of tumor, thus suppressing tumor development (Hönscheid et al., 2014). Several studies also showed that antitumor drugs can kill

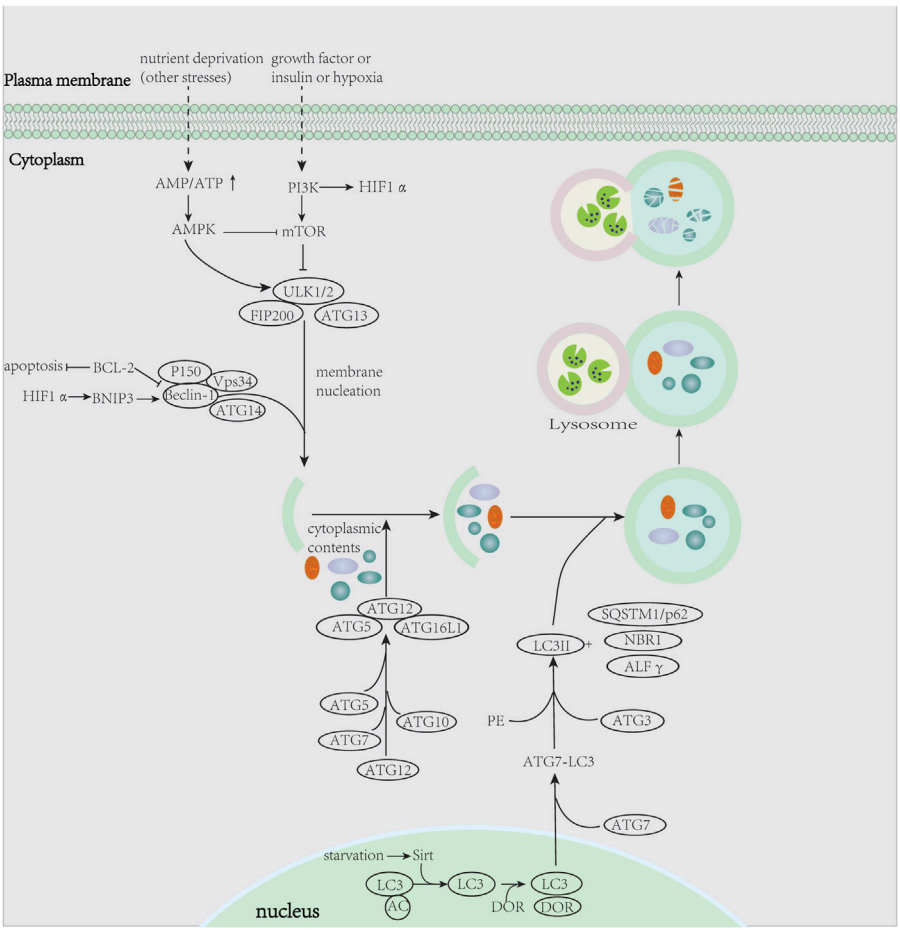


FIGURE 1 | Summary of the process of autophagy.

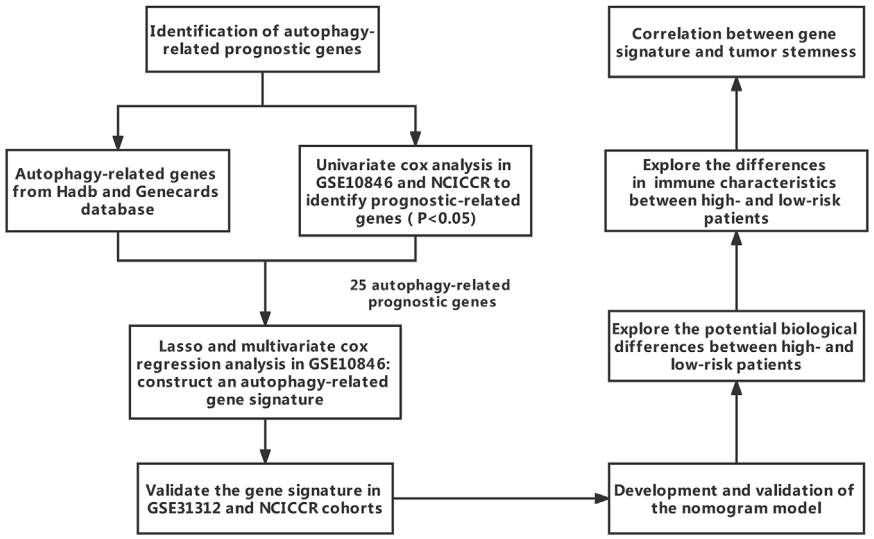


FIGURE 2 | Study flowchart.

TABLE 1 | Summary of DLBCL patients clinical characteristics.

	GSE10846 (*n* = 412)	GSE31312 (*n* = 470)	NCICCR(*n* = 234)
Gender			
Male	222 (53.9%)	271 (57.7%)	139 (59.4%)
Female	172 (41.7%)	199 (42.3%)	95 (40.6%)
Unknown	18 (4.4%)	–	–
Age (years)			
≤60	188 (45.6%)	200 (42.6%)	116 (49.6%)
>60	224 (54.4%)	270 (57.4%)	118 (50.4%)
Ann Arbor Stage			
I/II	188 (45.6%)	220 (46.8%)	109 (46.6%)
III/IV	217 (52.7%)	250 (53.2%)	121 (51.7%)
Unknown	7 (1.7%)	–	4 (1.7%)
ECOG			
<2	295 (71.6%)	374 (79.6%)	–
≥2	93 (22.6%)	96 (20.4%)	–
Unknown	24 (5.8%)	–	–
Subtype			
GCB	182 (44.2%)	248 (52.8%)	–
Non-GCB	230 (55.8%)	222 (47.2%)	–
N_extra			
<2	351 (85.2%)	366 (77.8%)	–
≥2	30 (7.3%)	104 (22.1%)	–
Unknown	31 (7.5%)	–	–
IPI			
≤2	–	274 (58.3%)	126 (53.9%)
>2	–	150 (31.9%)	67 (28.6%)
Unknown	–	46 (9.8%)	41 (17.5%)
Regimen			
RCHOP	232 (56.3%)	470 (100%)	–
CHOP	180 (43.7%)	–	–
Therapy response			
CR	–	354 (75.3%)	–
PR	–	72 (15.3%)	–
SD	–	20 (4.3%)	–
PD	–	24 (5.1%)	–

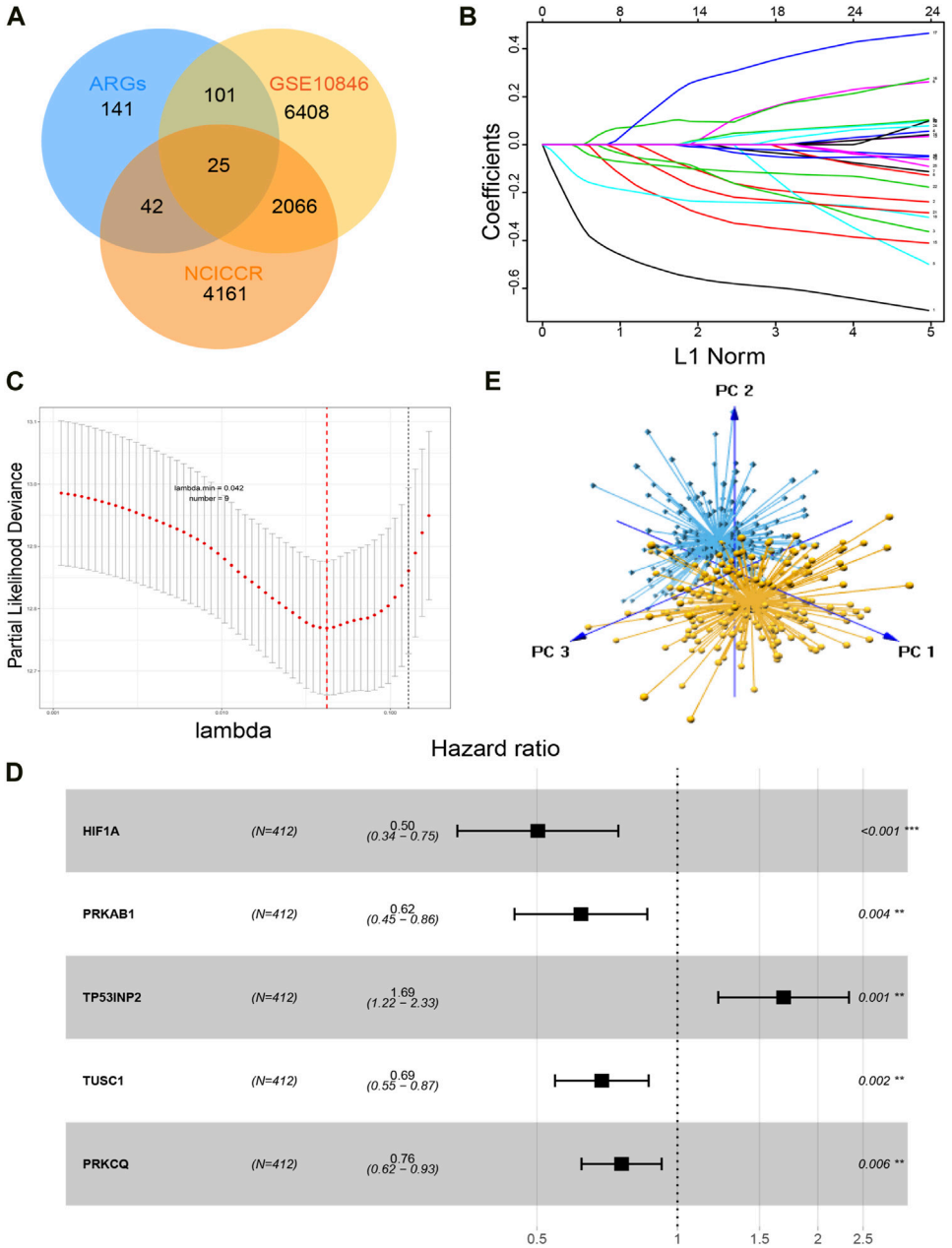


FIGURE 3 | Construction of the gene signature. **(A)** Overlapped genes generated by autophagy-related genes and prognosis-related genes of GSE10846 and NCICCR datasets. **(B)** LASSO coefficient profiles of the 25 overlapped genes. **(C)** Nine genes are selected by LASSO Cox regression analysis. Two dotted vertical lines are drawn at the value of minimum criteria and 1-s.e criteria of cross-validation, respectively. **(D)** Five genes that make up the gene signature and their corresponding hazard ratios resulting from multivariable Cox regression. **(E)** Principal component analysis of the five autophagy-related genes between high-risk and low-risk patients. Low-risk samples are marked in blue, and high-risk samples are marked in yellow.

tumor cells by inducing autophagy (Akar et al., 2008; Chiu et al., 2011). As the link between autophagy and tumor prognosis was strengthened, increasing studies have demonstrated the implication of autophagy in DLBCL. Zhang et al. (2020) found that miR-449a downregulated the expression of ATG4B by binding to the 3′UTR of its mRNA, which subsequently reduced the autophagy of T-cell lymphoma cells and promoted tumor apoptosis. Li et al. (2019) showed that CUL4B activated the protective autophagy to promote the growth of DLBCL cells through the JNK signaling pathway, and interfering with the expression of CUL4B could inhibit autophagy by regulating the JNK signaling pathway, thereby decreasing cell proliferation. These findings indicated that autophagy was tightly associated with the progression of DLBCL, and the autophagy-related genes (ARGs) could serve as promising therapeutic targets for DLBCL patients.

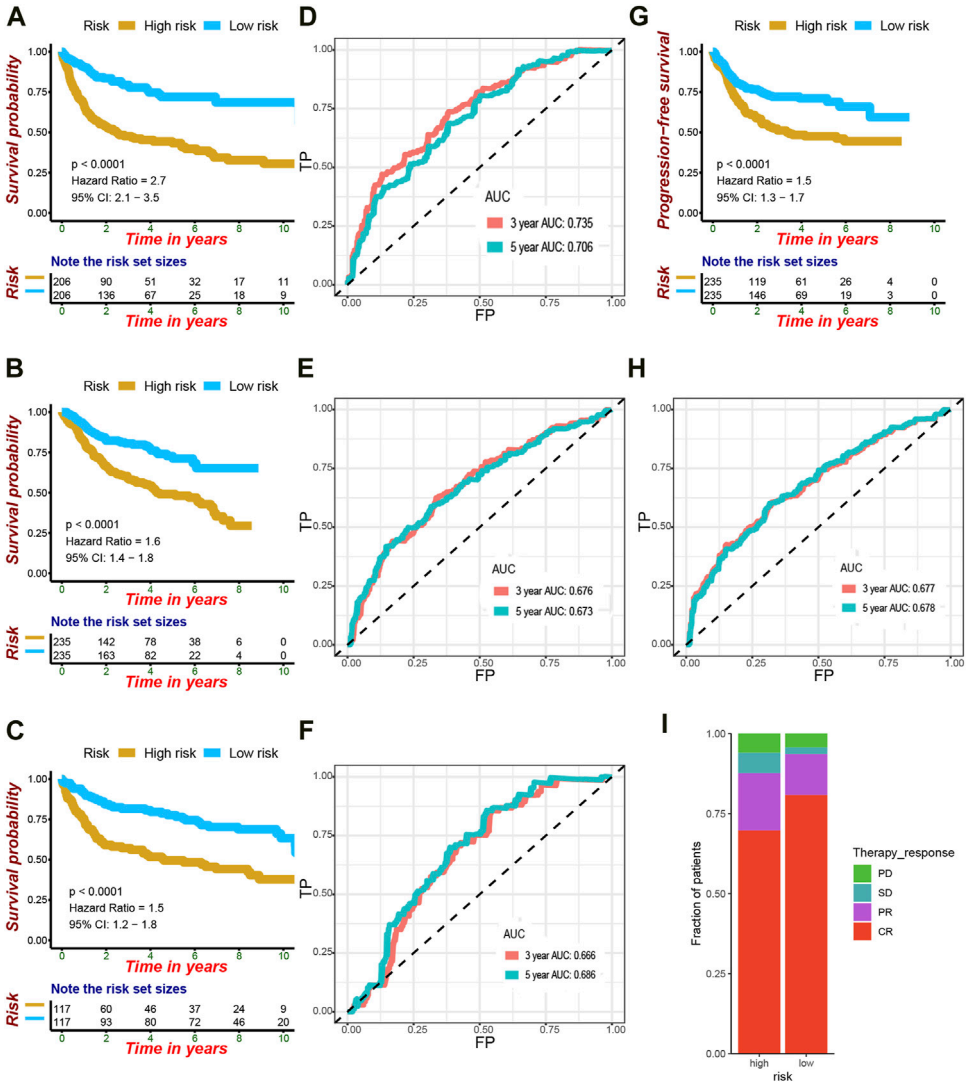


FIGURE 4 | Verification of the efficacy and accuracy of the gene signature. **(A)** Kaplan–Meier analysis of OS for low-risk and high-risk patients in the training cohort. **(B)** Kaplan–Meier analysis of OS for low-risk and high-risk patients in the GSE31312 cohort. **(C)** Kaplan–Meier analysis of OS for low-risk and high-risk patients in the NCICCR cohort. **(D)** Area under the ROC curve at 3 and 5 years to evaluate the accuracy of OS prediction in the training cohort. **(E)** Area under the ROC curve at 3 and 5 years to evaluate the accuracy of OS prediction in the GSE31312 cohort. **(F)** Area under the ROC curve at 3 and 5 years to evaluate the accuracy of OS prediction in the NCICCR cohort. **(G)** Kaplan–Meier curves showing progression-free survival of high-risk and low-risk patients based on the GSE31312 cohort. **(H)** Area under the ROC curve at 3 and 5 years to evaluate the accuracy of PFS prediction in the GSE31312 cohort. **(I)** Difference in the therapy response between high-risk and low-risk patients in the GSE31312 cohort.

The aim of the present study was to construct an autophagy-related gene signature to predict the prognosis of DLBCL patients and explore the differences in pathways, immune response, and tumor stemness between high-risk and low-risk patients.

MATERIALS AND METHODS

Selection of Autophagy-Related Genes

To find genes linked to autophagy, we examined the Human Autophagy Database (http://autophagy.lu/index.html) which included 231 genes reported to be involved in the autophagy process (Moussay et al., 2011). In addition, the term "autophagy" was also searched on the GeneCards website (https://www.genecards.org/) to identify genes that are associated with the autophagy activity. In this study, we defined an association score higher than 7 as autophagy-related genes. After eliminating overlapped genes in the two databases, a total of 309 genes were finally selected for our study.

Patients' Samples

Raw data and corresponding clinical information of the GSE31312 and GSE10846 datasets were retrieved from the Gene Expression Omnibus (GEO) database. Transcriptome

RNA-sequencing and clinical data on 481 DLBCL samples were obtained from the TCGA database (https://cancergenome.nih.gov/). Patients without clinical survival information or with a follow-up time <0 days were removed from the study. Finally, 234 patients from the TCGA-NCICCR dataset and 412 and 470 patients from the GSE10846 and GSE31312 datasets were included in our study. Among them, GSE10846 served as the training cohort, and GSE31312 and NCICCR were used as the validation cohorts.

Data Processing

For GSE10846 and GSE31312 datasets, the robust multi-array average (RMA) algorithm in the "affy" package in Bioconductor was used to perform background correction, quantile normalization, and final summarization of oligonucleotides per transcript using the median polish algorithm for the raw data (Zeng et al., 2019). In addition, the probes were annotated according to the "hgu133plus2.db" package. If multiple probes corresponded to the same gene, the largest average value was calculated as the expression value of this gene. Finally, K Nearest Neighbor (KNN) imputation was used to impute missing expression values in the gene expression profiles (Yasrebi, 2015). For the NCICCR dataset, the IDs were annotated based on the human genome reference (Hg38). In the event of multiple IDs matching to the same gene, the genes with the largest average value were served. Then, the voom algorithm from the "limma" package was used for data normalization (Ducie et al., 2017).

Construction of a Gene Signature Associated With Survival of Diffuse Large B-Cell Lymphoma Patients

Univariate Cox proportional hazard regression analysis was first conducted to screen the genes associated with overall survival (OS). Genes with $p < 0.05$ were considered statistically significant. Then, the intersection of prognosis-related genes and autophagy-related genes was taken to obtain prognostic-related autophagy genes. These genes were further screened by LASSO regression analysis and multivariate Cox regression analysis. We calculated the riskscore for each patient by using the regression coefficients of the individual genes obtained from the multivariate Cox regression model and the expression value of each of the selected genes. The formula was as follows: Riskscore = $\sum_{i=1}^{n}$expi∗βi. Patients in each cohort were classified into high-risk and low-risk groups by using the median riskscore.

Development of a Nomogram

A nomogram was constructed based on the results of multivariate analysis. The performance of the nomogram was measured by area under the ROC curve and the calibration curve, and its predictive ability was further verified in the validation cohort. X-tile software was used to find the best cutoff value of the nomogram score, according to the highest χ^2-value defined by the Kaplan–Meier survival analysis and log-rank test (Camp et al., 2004).

Functional and Pathway Analysis

The correlation test was used to identify genes correlated with the autophagy-related genes of the gene signature under the cutoff value of the absolute value of the correlation coefficient which was higher than 0.4, and the p value was lower than 0.05. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were then performed with these hooked genes by the "clusterProfiler" package to further explore the function of the autophagy-related genes of the gene signature. In addition, we performed Gene Set Enrichment Analysis (GSEA) to uncover the difference in signaling pathways between high-risk and low-risk groups.

Estimation of TME Cell Infiltration

The ESTIMATE algorithm was performed to quantify the tumor microenvironment, including the immune score and stroma score. Moreover, we obtained gene sets of each TME infiltrating immune cell type from the study of Charoentong et al. (2017), which included activated dendritic cells, macrophages, activated $CD8^+$ T cells, regulatory T Cells, and natural killer T cells. Subsequently, ssGSEA in the "GSVA" package based on the deconvolution algorithm was used to estimate the relative infiltration level of each cell population in each DLBCL sample with gene expression data.

Correlation Between the Gene Signature and Immune Activation-Related Genes and Immune Activation Pathways

In order to explore a potential relationship between the gene signature and immune response, we first selected TNF, IFNG, TBX2, GZMB, CD8A, PRF1, GZMA, CXCL9, and CXCL10 that were extracted from the published literature and considered to be associated with immune activation. We further downloaded immune activation gene sets from KEGG, which included antigen processing and presentation, NOD-like receptor, T-cell receptor, and Toll-like receptor (Zeng et al., 2019).

Correlation Between the Gene Signature and Tumor Stemness

To explore the relationship between the gene signature and tumor stemness, we used one-class logistic regression (OCLR) algorithm to calculate the gene expression-based stemness index (mRNAsi) of DLBCL patients, and mRNAsi was then mapped to the range of 0–1, utilizing a linear transformation that was subtracted by the minimum and divided by the maximum (Malta et al., 2018). MRNAsi could describe the similarity between tumor cells and stem cells, and it might be considered a quantitative form of CSCs. Those patients with high mRNAsi scores were associated with active biological processes and a higher level of tumor dedifferentiation.

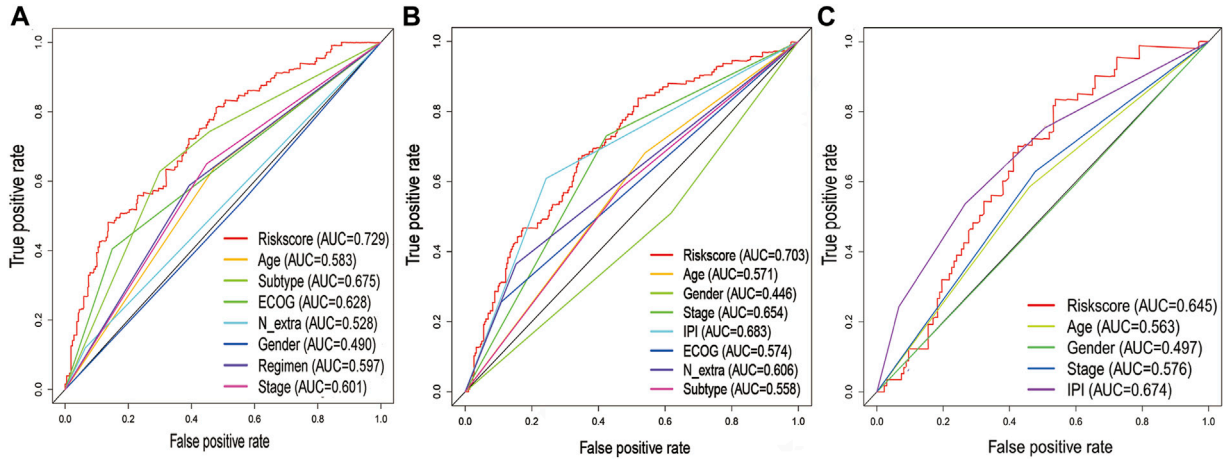


FIGURE 5 | Comparison of 3-year ROC curves with other common clinical characteristics showed the superiority of the gene signature. **(A)** GSE10846 dataset. **(B)** GSE31312 dataset. **(C)** NCICCR dataset.

TABLE 2 | Univariate Cox regression analysis of overall survival in the training and validation cohorts.

Variables	Training cohort		NCICCR cohort		GSE31312 cohort	
	HR (95%CI)	p	HR (95%CI)	p	HR (95%CI)	p
Age						
(>60 vs. ≤60)	2.015 (1.422–2.855)	<0.001	1.963 (1.296–2.972)	0.001	1.850 (1.336–2.560)	<0.001
Gender						
(Men vs. Women)	0.912 (0.654–1.270)	0.585	1.223 (0.804–1.860)	0.348	0.966 (0.712–1.310)	0.824
Subtype						
(Non-GCB vs. GCB)	2.698 (1.858–3.918)	<0.001	–	–	1.512 (1.117–2.047)	0007
Regimen						
(CHOP vs. RCHOP)	1.917 (1.343–2.737)	<0.001	–	–	–	–
ECOG						
(≥2 vs. <2)	2.884 (2.049–4.058)	<0.001	–	–	2.037 (1.460–2.841)	<0.001
Stage						
(III/IV vs. I/II)	1.917 (1.357–2.709)	<0.001	1.500 (0.996–2.260)	0.052	2.337 (1.688–3.238)	<0.001
N_extra						
(≥2 vs. <2)	1.869 (1.089–3.207)	0.023	–	–	2.202 (1.597–3.038)	<0.001
Riskscore						
(High vs. Low)	2.608 (1.824–3.728)	<0.001	2.457 (1.616–3.736)	<0.001	2.326 (1.680–3.220)	<0.001

Statistical Analysis

All analyses were carried out by R version 3.6.1 and corresponding packages. We applied the Wilcoxon test for continuous variables to compare the differences between high-risk and low-risk groups. The Coxph function in the "survival" package was used for univariate and multivariate Cox regression analyses, and the "glmnet" package was performed for Lasso regression analysis. Kaplan–Meier curves were plotted by the "survival" package, and the log-rank test was used to analyze the significant difference in overall survival and progression-free survival of high-risk and low-risk DLBCL patients. The area under the curve (AUC) of the ROC curve was calculated by the "survival ROC" package to evaluate the accuracy of the gene signature. The "rms" package was used to generate a nomogram. A normalized enrichment (NES) and p-adjusted were used to determine the statistical significance of GSEA analysis.

RESULTS

Construction and Validation of the Gene Signature

Figure 1 summarizes the process of autophagy, the key pathways, and autophagy-related genes involved. **Figure 2** showed the study flowchart. The clinical information of DLBCL patients from the GSE31312, GSE10846, and NCICCR datasets is shown in **Table 1**. In order to identify prognosis-related genes, we respectively performed univariate Cox regression analysis in NCICCR and

TABLE 3 | Multivariate Cox regression analysis of overall survival in the training and validation cohorts.

Variable	Training cohort		NCICCR cohort		GSE31312 cohort	
	HR (95%CI)	*p*	HR (95%CI)	*p*	HR (95%CI)	*p*
Age (>60 vs. <=60)	1.801 (1.261–2.573)	0.001	1.945 (1.283–2.949)	0.002	1.496 (1.068–2.096)	0.019
Gender (Men vs. women)	–	–	–	–	–	–
Subtype (Non-GCB vs. GCB)	1.785 (1.188–2.682)	0.005	–	–	1.361 (0.997–1.856)	0.051
Regimen (CHOP vs. RCHOP)	1.757 (1.141–2.705)	0.01	–	–	–	–
ECOG (≥2 vs. < 2)	2.138 (1.486–3.076)	<0.001	–	–	1.745 (1.233–2.467)	0.001
Stage (III/IV vs. I/II)	1.436 (0.995–2.072)	0.053	–	–	1.725 (1.207–2.463)	0.002
N_extra (≥2 vs. < 2)	2.293 (1.227–4.286)	0.009	–	–	1.623 (1.152–2.285)	0.019
Riskscore (High vs. low)	1.830 (1.218–2.750)	0.004	2.444 (1.605–3.723)	<0.001	1.984 (1.417–2.777)	<0.001

TABLE 4 | Stratified analysis for DLBCL patients based on the gene signature.

	Events (n)/patients (N)			HR (95%CI)	*p*-value
	All patients	Low risk	High risk		
Age					
≤60	52/188	13/98	39/90	2.8 (1.8–4.3)	$p < 0.0001$
>60	111/224	35/108	76/116	2.7 (2.0–3.5)	$p < 0.0001$
Subgroup					
GCB	48/182	20/123	28/59	3.9 (2.4–6.4)	$p < 0.0001$
Non_GCB	115/230	28/83	87/147	2.1 (1.5–2.8)	$p = 0.0002$
Stage					
I/II	57/188	18/102	39/86	3.0 (1.9–4.5)	$p = 0.0264$
III/IV	102/217	28/99	74/118	2.6 (1.9–3.4)	$p < 0.0001$
Regimen					
RCHOP	59/232	30/152	29/80	2.6 (1.7–3.9)	$p = 0.0008$
CHOP	104/170	18/54	86/126	2.6 (1.9–3.7)	$p < 0.0001$
ECOG					
<2	97/295	30/151	67/144	2.5 (1.9–3.4)	$p < 0.0001$
≥2	60/93	16/40	44/53	3.1 (2.0–4.8)	$p < 0.0001$
N_extra					
<2	139/351	41/175	98/176	2.9 (2.2–3.8)	$p < 0.0001$
≥2	16/30	5/14	11/16	1.7 (0.8–3.3)	$p = 0.1272$
IPI					
≤2	70/274	16/142	54/132	1.7 (1.4–2.2)	$p < 0.0001$
>2	85/150	26/61	59/89	1.4 (1.2–1.7)	$p = 0.0146$

GSE10846 datasets. Under the cutoff values of $p < 0.05$, 8600 genes in GSE10846 and 6294 genes in the NCICCR dataset were considered as prognosis-related genes. By overlapping the prognosis-related genes with autophagy-related genes, 25 shared genes were retained (**Figure 3A**). These significant genes were afterward entered into LASSO regression analysis and multivariate Cox regression analysis, and the GSE10846 dataset served as the training cohort (**Figures 3B,C**). Finally, TP53INP2, PRKCQ, TUSC1, PRKAB1, and HIF1A were identified as members of the gene signature (**Figure 3D**). TP53INP2 with HR > 1 was regarded as a risk gene, while remaining genes with HR < 1 as protective genes. According to the relative expression level of five genes and the corresponding multivariate Cox regression coefficient, the riskscore of each sample in the training and validation cohorts could be calculated easily. The riskscore was calculated as follows: Riskscore = (0.524 × TP53INP2 expression)—(0.276 × PRKCQ expression)—(0.373 × TUSC1 expression)—(0.476 × PRKAB1 expression)—(0.689 × HIF1A expression). The patients in each cohort were classified into high-risk and low-risk groups according to the median value of the riskscore. Principal component analysis (PCA) based on the five autophagy-related genes confirmed the difference between the two groups (**Figure 3E**). Kaplan–Meier curves demonstrated that

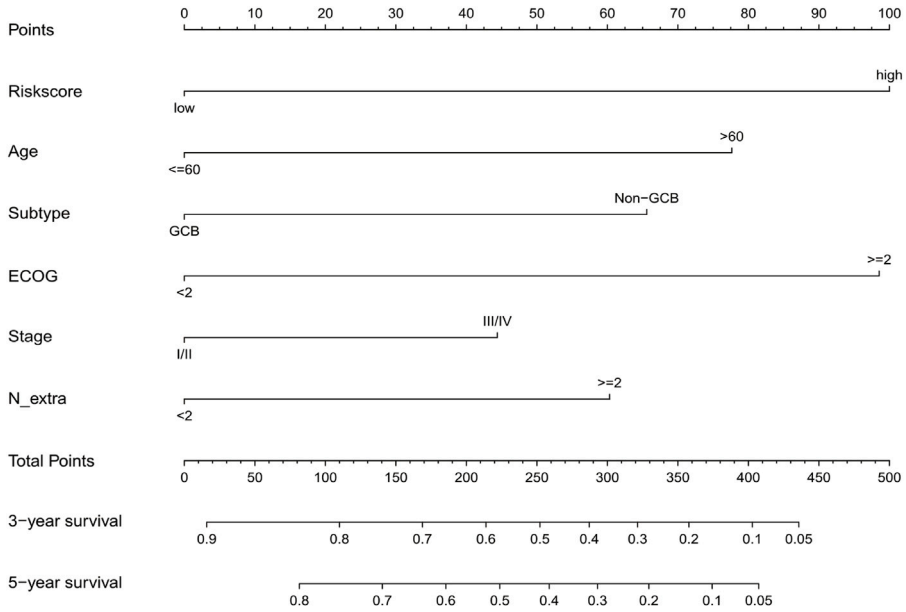


FIGURE 6 | Predictive nomogram was constructed using clinical risk factors and the gene signature.

the patients in the high-risk group had shorter overall survival (OS) than those in the low-risk group (HR: 2.7, 95%CI: 2.1–3.5, and $p < 0.001$; **Figure 4A**). The finding was further validated in external cohorts to evaluate the reproducibility and validity of this gene signature (GSE31312: HR: 1.6, 95% CI: 1.4–1.8, and $p < 0.001$; NCICCR: HR: 1.5, 95% CI: 1.2–1.8, and $p < 0.001$; **Figures 4B,C**). In predicting the 3-year and 5-year OS rate of DLBCL patients, the gene signature achieved AUC values of 0.735 and 0.706 in the GSE10846 cohort, 0.676 and 0.673 in the GSE31312 cohort, and 0.666 and 0.686 in the NCICCR cohort, respectively, showing a substantially effective performance for overall survival prediction (**Figures 4D–F**). In addition, shorter progressive-free survival (PFS) was found in the high-risk patients of the GSE31312 cohort (HR: 1.5, 95%CI: 1.3–1.7; **Figure 4G**). The gene signature also showed good performance in predicting PFS with 3-year AUC and 5-year AUC of 0.677 and 0.678 (**Figure 4H**). In addition, patients with low-risk in the GSE31312 cohort showed a higher rate of remission rate (**Figure 4I**).

After removing patients with incomplete clinical information, we analyzed the correlation between gene signature and clinical factors. The results suggested that the gene signature was related to the ECOG score and subtype, but there was no obvious correlation with other clinical factors (**Supplementary Figure S1**). In addition, the multi-index ROC curve indicated that the predictive accuracy of the gene signature was higher than that of the clinicopathological factors, even higher than the IPI score in the GSE31312 cohort (**Figure 5**).

Gene Signature Is Independent of Other Clinicopathological Factors

We further evaluated the prognostic value of the gene signature and other clinicopathological factors by using univariate and multivariate Cox regression analyses. It suggested that the gene signature may serve as a valuable prognostic parameter, independent of other clinical factors (GSE10846: HR: 1.830, 95% CI: 1.218–2.750, $p = 0.004$; NCICCR: HR: 2.444, 95% CI: 1.605–3.723, $p < 0.001$; GSE 31312: HR: 1.984, 95% I: 1.417–2.777, $p < 0.001$; **Table 2** and **Table 3**). When stratified by age, subtype, stage, regimen, ECOG, and IPI score, the gene signature was still a clinically and statistically significant prognostic model (**Table 4**).

In addition, we further examined the effect of treatment response on survival. The response to treatment was found to be significantly correlated with the survival rate (**Supplementary Figure S2A**). The median OS time was not yet reached in patients with complete response, and the median OS time was 1.429, 0.936, and 0.341 years in patients with partial response, stable disease, and progressive disease, respectively. We also found that the gene signature could divide patients achieving complete response into two groups with significant differences in survival (**Supplementary Figure S2B**). This result indicated that achieving CR did not always mean a favorable prognosis. However, the gene signature did not have the ability to further distinguish the prognosis of patients with partial response, stable disease, and progressive disease (**Supplementary Figure S2C–E**).

Personalized Prognostic Prediction Nomogram

We generated a nomogram to predict the probability of 3-year and 5-year OS by integrating the gene signature and clinical factors (**Figure 6**). In order to better conduct external verification, we converted the gene signature into a binary variable. The AUC at 3 years of the nomogram in the training cohort was 0.771, and

the AUC of the gene signature was 0.658 (**Figure 7A**). The GSE31312 cohort was used to validate the predictive accuracy of the nomogram, and the AUC at 3 years of the nomogram in the GSE31312 cohort was 0.735, which is higher than that of the gene signature and IPI score (**Figure 7B**). In addition, despite the training cohort or the validation cohorts, the calibration plot had a favorable agreement between the prediction by the nomogram and ideal model in the probability of 3-year and 5-year survival (**Figures 7C–F**).

We defined the optimal cutoff value of the nomogram score based on the X-tile plots, and patients in both the training and validation cohort were separately stratified into three subgroups according to the cutoff value (0–143, 144–309, and ≥310). In the training cohort, the 5-year OS rate of the nomogram model was 83.1, 47.9, and 13.6%, respectively, while the 5-year OS rate of the Ann arbor stage I–II and III–IV was 67.6 and 47.6%, respectively (**Figures 7G,H**). In the validation cohort, the 5-year OS rates of the nomogram model were 83.2, 50.7, and 22.5%, respectively, while the 5-year OS rates of the IPI score 0–2 and 3–5 were 72.8 and 36.4%, respectively, and those of Ann arbor stage I–II and III–IV were 74 and 48%, respectively. Therefore, the nomogram displayed better prognostic stratification ability than the IPI score and Ann Arbor staging system (**Figures 7I–K**).

Identification of Involved Functions and Signaling Pathways

To investigate the potential functions and signaling pathways related to the gene signature in DLBCL, we used the five genes as baits to hook 501 highly relevant genes by the correlation test. GO analysis indicated that these genes were associated with cell adhesion, immune cell activation, and differentiation, as well as cytokine and growth factor-binding reaction (**Figure 8A**). KEGG analysis indicated that the 501 genes were involved in protein digestion and absorption, focal adhesion, and ECM-receptor interaction (**Figure 8B**). In addition, the differentially expressed genes between low-risk and high-risk groups were mainly enriched in the PI3K-AKT signaling pathway, focal adhesion, protein digestion and absorption, ECM-receptor interaction, and other pathways (**Supplementary Figure S3**). In the GSEA analysis between low-risk and high-risk groups, we found that the high-risk group was significantly associated with doxorubicin resistance (NES = 2.507 and p.adjust = 0.0064), NF-κB pathway (NES = 1.726 and p.adjust = 0.0145), cell cycle (NES = 2.052 and p.adjust = 0.0157), and DNA replication pathway (NES = 2.446 and p.adjust = 0.0134)(**Figures 9A,B,D,E**), while the low-risk group was associated with the activation of the PI3K-AKT pathway (NES = −1.793 and p.adjust = 0.0101) and adaptive immune response (NES = −1.683 and p.adjust = 0.0262) (**Figures 9C,F**).

Infiltration Characteristics of TME Cells and Immune Response in the High-Risk and Low-Risk Groups

We used the ESTIMATE algorithm to quantify the overall infiltration of immune and stromal cells between high-risk and low-risk patients. The results obtained indicated that there was significant higher immune and stroma cell infiltration in the low-risk patients (**Figure 10A**). To better elaborate the aforementioned findings, subsequent analysis of ssGSEA indicated that the low-risk group was rich in not only innate immune cell infiltration (e.g., dendritic cells, myeloid-derived suppressor cells, macrophages, mast cells, and natural killer cells) but also adaptive immune cell infiltration (e.g., activated CD4 T cells, activated CD8 T cells, gamma delta T cells, regulatory T cells, T follicular helper cells, type 1 T helper cells, type 17 T helper cells, and type 2 T helper cells) (**Figure 10B**). As expected, the low-risk group showed higher expression of mRNA related to immune activation and more obvious activation of immune pathways, including antigen processing and presentation pathways, NOD-like receptors, T-cell receptors, and Toll-like receptor pathways (**Figures 10C,D**). In summary, there was a higher degree of immune cell infiltration and immune activation response in the low-risk group.

Difference in Tumor Stemness Between High-Risk and Low-Risk Groups

We implemented the OCLR algorithm to obtain the mRNAsi of DLBCL. As shown in **Figure 11A**, patients in the high-risk group were more likely to have higher mRNAsi, suggesting that patients in the high-risk group had higher similarity to cancer stem cells, presenting more active biological processes and higher tumor dedifferentiation degree. In addition, the high-mRNAsi group showed a worse prognosis than the low-mRNAsi group (**Figure 11B**).

DISCUSSION

Diffuse large B-cell lymphoma is a highly heterogeneous tumor with different biological and clinical characteristics, so there is a large difference in the survival rate between the high-risk and low-risk patients (Juskevicius et al., 2016). Due to several limits in the present prognostic assessment system, it is urgent to establish a new model containing genetic information to better predict the prognosis of DLBCL patients (Ruppert et al., 2020).

The study of autophagy is a rapidly evolving field with great potential for providing new horizons for the treatment of malignant disease. Autophagy can regulate the components of the immune system, thereby affecting its homeostasis, survival, activation, proliferation, and differentiation (Jiang et al., 2019). At the same time, autophagy is also shown to enhance chemoresistance and helps in maintaining the stemness of tumor stem cells (Jang et al., 2017). Xu et al. (2021)uncovered that ARRDC1-AS1 facilitated the progression of DLBCL and enhanced autophagy of DLBCL by targeting the miR-2355-5p/ATG5 axis. Additionally, Amaravadi et al. (2007)found that inhibition of autophagy with either chloroquine or ATG5 short hairpin RNA (shRNA) enhanced the ability of either p53 activation or alkylating drug therapy to induce lymphoma cell death.

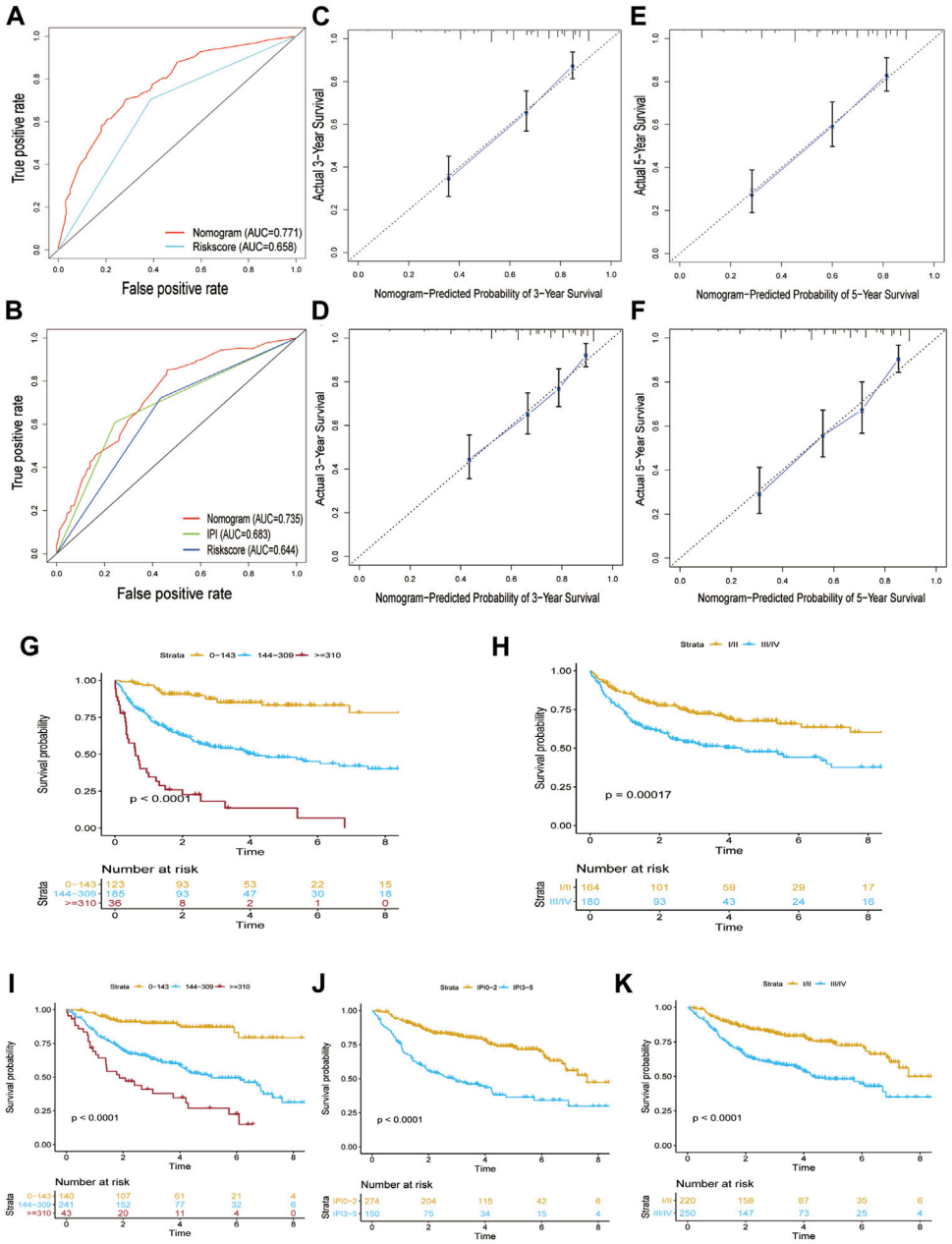


FIGURE 7 | Verification of the performance of the nomogram model. **(A)** Area under the ROC curve at 3 years was used to assess the prognostic accuracy of the nomogram in the GSE10846 dataset. **(B)** Area under the ROC curve at 3 years was used to assess the prognostic accuracy of the nomogram in the GSE31312 dataset. **(C)** Calibration curves of the nomogram in prediction of the 3-year OS in the GSE10846 dataset. **(D)** Calibration curves of the nomogram in prediction of the 3-year OS in the GSE31312 dataset. **(E)** Calibration curves of the nomogram in prediction of the 5-year OS in the GSE10846 dataset. **(F)** Calibration curves of the nomogram in prediction of the 5-year OS in the GSE31312 dataset. **(G)** Kaplan–Meier curves showing the survival difference between different nomogram score groups in the GSE10846 dataset. **(H)** Kaplan–Meier curves showing the survival difference between different stages in the GSE10846 dataset. **(I)** Kaplan–Meier curves showing the survival difference between different nomogram score groups in the GSE31312 dataset. **(J)** Kaplan–Meier curves showing the survival difference between different IPI score groups in the GSE31312 dataset. **(K)** Kaplan–Meier curves showing the survival difference between different stages in the GSE31312 dataset.

In this study, we successively used univariate Cox regression analysis, LASSO Cox regression analysis, and multivariate Cox regression analysis to screen autophagy-related prognostic genes and finally identified five genes to construct gene signature. However, unlike other research, we used two sources of autophagy-related genes from Hadb and GeneCards databases, respectively, to avoid missing some important genes that were fundamental to the prognosis of DLBCL. Meanwhile, to incorporate into genes with higher relevance to autophagy, we defined an association score higher than 7 as autophagy-related genes. Our results showed that this gene signature could effectively classify patients into high-risk and low-risk

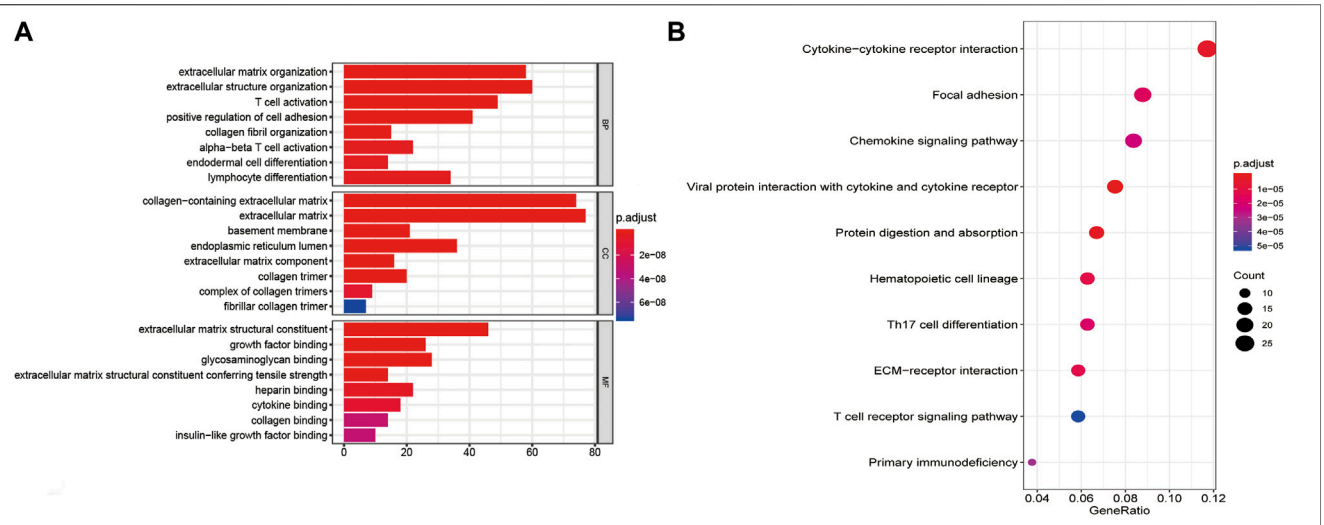


FIGURE 8 | Potential functions and signaling pathways related to the autophagy-related genes of the gene signature. **(A)** GO analysis. **(B)** KEGG analysis.

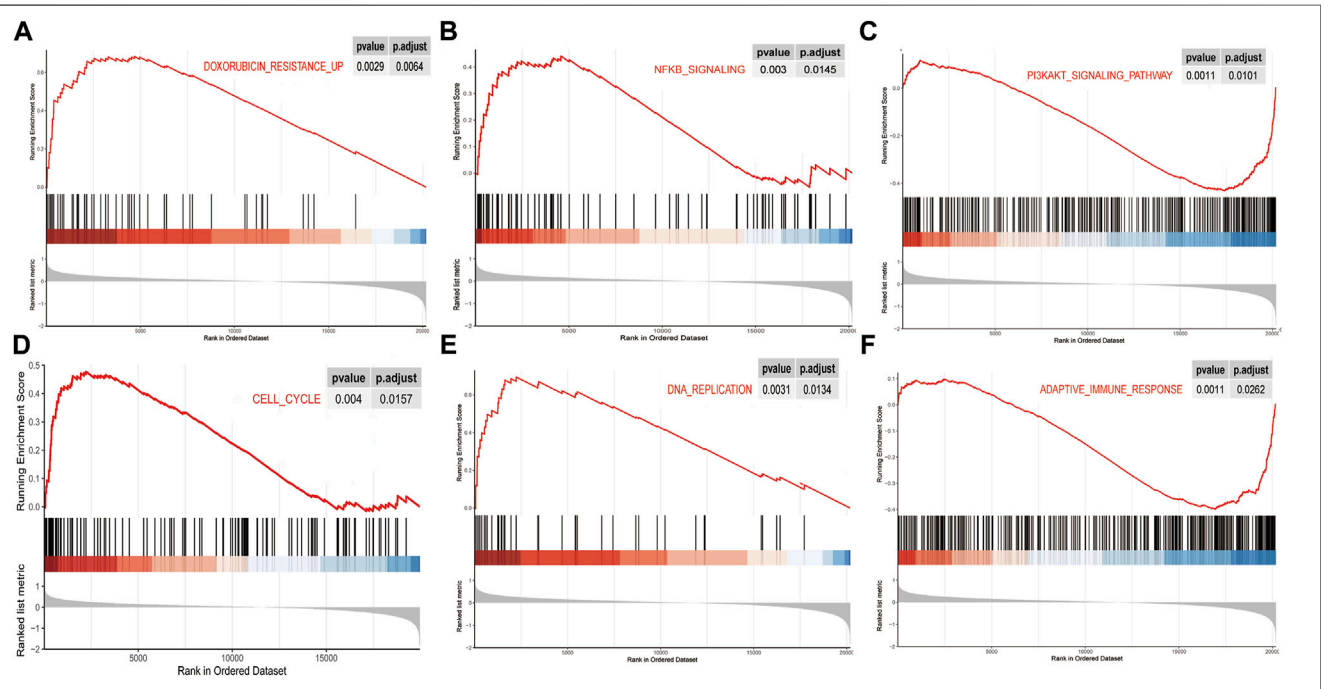


FIGURE 9 | Pathway characteristics of gene signature in DLBCL. **(A-F)** Gene set enrichment analysis revealed pathways enriched in the high-risk group and low-risk group.

groups with significant differences in overall survival and progression-free survival and had favorable prognostic accuracy. The nomogram consisted of the gene signature and clinical factors had better discrimination and prognostic stratification ability than the gene signature and IPI score alone, and it also had a favorable consistency between the predicted and actual survival. To the best of our knowledge, the gene prognostic models have been acknowledged by the majority of researchers, but there were few studies concerning DLBCL. Pan et al. constructed a TME-relevant gene signature that could not only predict prognosis but also explored the relationship between the TME and DLBCL (Pan et al., 2021). The autophagy-related gene signature outlined in our study was the first explanation toward the pathophysiological process of diffuse large B-cell lymphoma with poor prognosis from the perspective of autophagy.

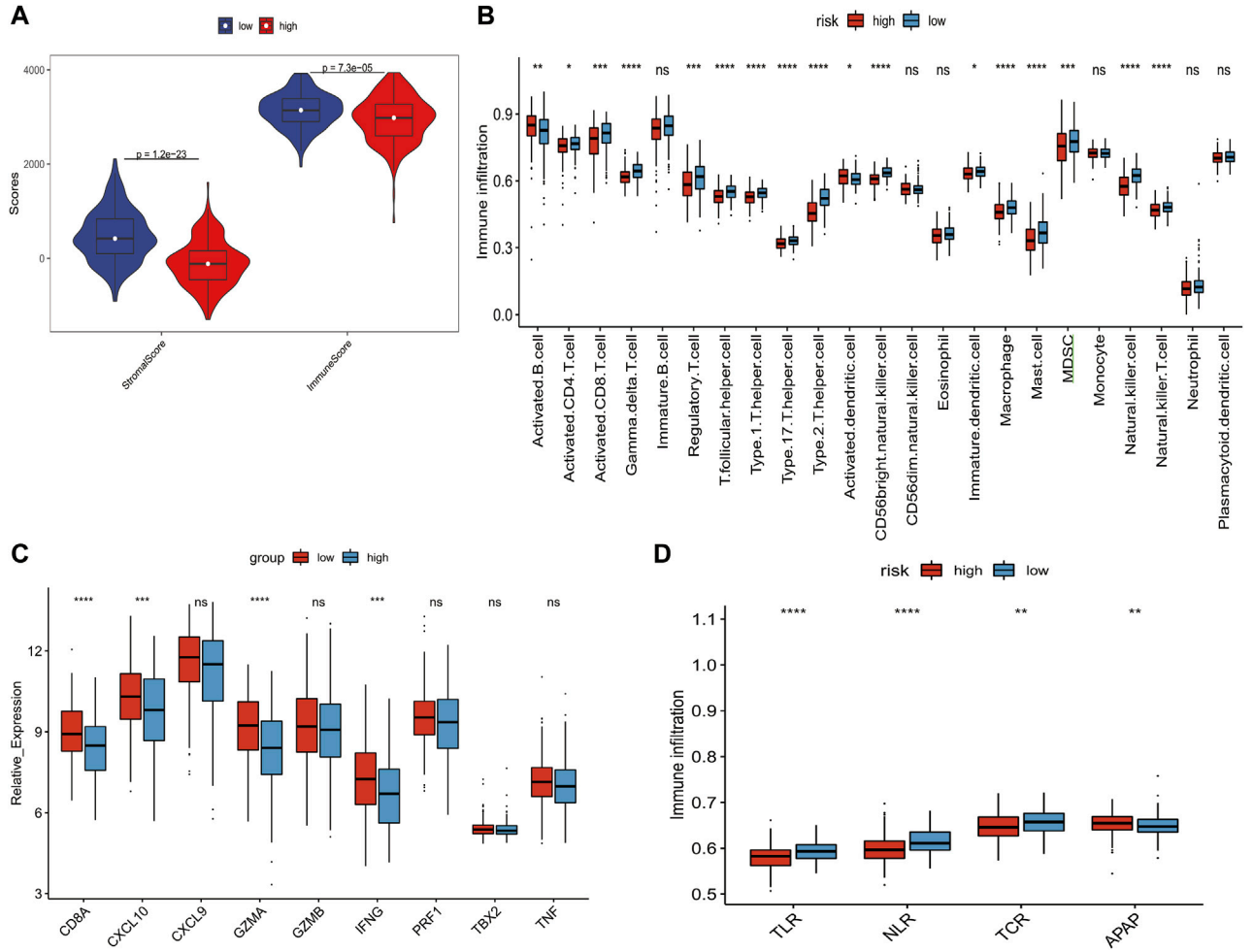


FIGURE 10 | Correlation between gene signature and the tumor microenvironment and immune reaction. **(A)** Differences of immunescore and stromalscore (calculated by the ESTIMATE algorithm) between high-risk and low-risk patients. **(B)** Abundance of each TME cell in high-risk and low-risk groups. **(C)** Difference in the immune-activation–related gene expression between high-risk and low-risk patients. **(D)** Differences in immune-activated pathways between high-risk and low-risk patients. TLR, Toll-like receptor; NLR, NOD-like receptor; TCR, T-cell receptor; APAR, antigen processing and presentation. The statistical difference of two groups was compared by the Wilcoxon test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; and ****$p < 0.0001$.

The results of GSEA analysis revealed the differences in pathways that might lead to different prognosis in high- and low-risk groups. Doxorubicin resistance, NF-κB pathway, cell cycle, and DNA replication pathway were activated in high-risk patients. The PI3K-AKT signaling pathway and adaptive immune response pathway were activated in low-risk patients. It is known that the abnormal regulation of the cell cycle and DNA replication played a crucial role in promoting tumor growth (Song et al., 2021). The previous study showed that constitutive activation of NF-κB was characteristic of most ABC subtypes of DLBCL, and the activation of the NF-κB pathway may be one of the mechanisms resulting in drug resistance of relapsed/refractory DLBCL (Turturro, 2015). For the PI3K-AKT signaling pathway enriched in the low-risk group, it could inhibit the activation of autophagy (M. Martelli et al., 2011). There was a good consistency with the findings of recent studies. Xie and his colleagues constructed an RNA binding protein-based prognostic signature for diffuse large B-cell lymphoma and also found that the activation of autophagy was associated with high-risk patients who had poor outcomes (Xie et al., 2021).

In addition, we also found that the low-risk group exhibited higher immune cell infiltration. The latest research showed that targeting autophagy pathways could reshape the tumor microenvironment by improving antigen processing and presentation to enhance T-cell response or increasing the production of Th1 chemokines to promote the infiltration of effector immune cells (Xia et al., 2021). A phase II clinical study also showed that after hydroxychloroquine was used in combination with the chemotherapy drugs paclitaxel and gemcitabine, patients with pancreatic adenocarcinoma showed increased immune cell infiltration (Zeh et al., 2020). Moreover, there was a higher degree of autophagy in the high-risk group, and patients in the high-risk group might have higher mRNAsi, indicating that these patients were more likely to have characteristics such as chemotherapy resistance and more invasiveness like CSCs.

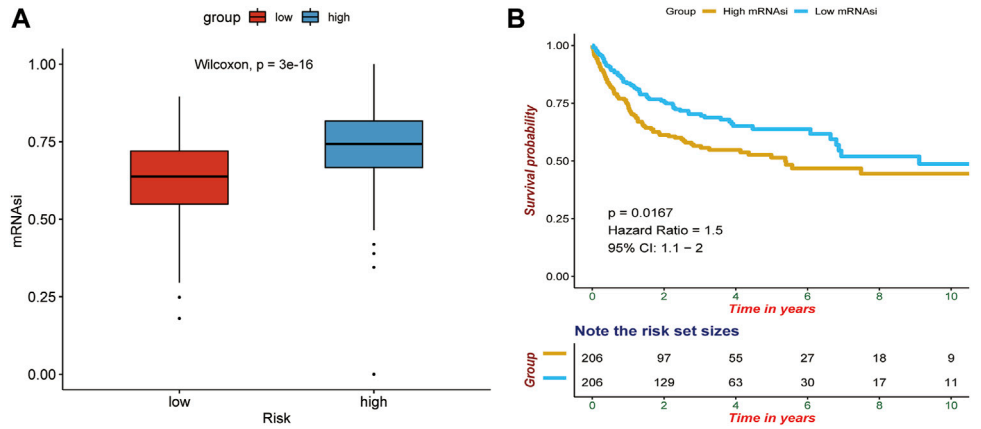


FIGURE 11 | Correlation between the gene signature and tumor stemness. **(A)** Differences of mRNAsi between high-risk and low-risk patients. **(B)** Survival difference between high-mRNAsi and low-mRNAsi. *p*-values were calculated with the log-rank test.

Therefore, the application of autophagy inhibitors might be a potentially important strategy for anti-tumor therapy.

The five genes identified in our study have previously been correlated with the prognosis of tumors. PRKCQ is a member of the novel protein kinase C (PKC) family and has been associated with many types of cancers, such as chromophobe renal cell carcinomas, breast cancer, and Notch-driven T-cell leukemia (Villalba and Altman, 2002; Byerly et al., 2016; Park et al., 2017). According to the GeneCards database, tumor suppressor candidate gene 1 (TUSC1) is located on chromosome 9p and is downregulated in non–small cell lung cancer and small cell lung cancer cell lines, suggesting that it may play a role in lung tumorigenesis. Tumor Protein P53 Inducible Nuclear Protein 2 (TP53INP2), which has about 36% homology with the known tumor protein 53-induced nuclear protein 1, plays a role in carrying LC3 and its homologous proteins out of the nucleus to autophagosome and promoting the interaction of LC3 and its homologous proteins with ATG7 (Liu and Klionsky, 2015). Current research shows that TP53INP2 either can promote the development or inhibit the proliferation of tumor cells depending on the tumor types (He et al., 2015; Hu et al., 2017). AMPK is a heterotrimer consisting of an alpha catalytic subunit and non-catalytic beta and gamma subunits. Among them, the beta1 subunit was encoded by PRKAB1. A finding indicated that the overexpression of AMPK-β1 inhibited the proliferation, migration, and invasion of ovarian cancer cells. When siRNA was used to interfere with AMPK-β1, the invasion ability of tumor cells would be enhanced (Li et al., 2014). As a protein-coding gene, HIF-1A encoded the alpha subunit of transcription factor hypoxia-inducible factor-1 (HIF-1). Overexpression of HIF-1α has been reported in several solid tumors, and elevated HIF-1α protein levels correlate with poor prognosis in majority of tumors (Zhong et al., 1999). However, Evens et al. (2010)revealed that expression of HIF-1α was an independent favorable prognostic factor affecting the overall survival in DLBCL patients receiving RCHOP regimen. Nevertheless, despite the proposed functions of these five ARGs in various types of cancer, the specific role of these genes in patients with DLBCL remains unclear.

Inevitably, there were still some potential limitations that could not be neglected in the present study. First, there are unknown biases because of the retrospective nature of our data. Second, the information of several clinical data of some patients was unavailable in these public datasets, and these patients need to be excluded in some analysis. Third, there were only autophagy-related genes included in our study, which did not represent the entire gene transcription profile associated with DLBCL. Fourth, there was a lack of further experiments to identify the function of these genes. Hence, the value of this gene signature is preliminarily demonstrated, and further verification is expected.

CONCLUSION

Taken together, we identified an autophagy-related gene signature that could efficiently predict the overall survival of DLBCL patients and was independent of other clinical factors. Moreover, the gene signature might serve as a promising marker of therapeutic resistance in DLBCL patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found at: https://cancergenome.nih.gov/ https://www.ncbi.nlm.nih.gov/gds/?term=.

AUTHOR CONTRIBUTIONS

XZ and Y-ZH put forward the concept of the study, designed the study, processed the data, prepared the manuscript, and contributed to the statistical analysis. DL and C-RL contributed to the data acquisition and performed the literature search for the manuscript. DL contributed to the quality control of data. LW and RH put forward the concept

of the study, contributed to the data analysis and interpretation, and reviewed the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the contributors of TCGA and GEO for sharing the diffuse large B-cell lymphoma data. This work was financially supported by the National Natural Science Foundation of China (81873450).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.862179/full#supplementary-material

REFERENCES

Akar, U., Chaves-Reyez, A., Barria, M., Tari, A., Sanguino, A., Kondo, Y., et al. (2008). Silencing of Bcl-2 Expression by Small Interfering RNA Induces Autophagic Cell Death in MCF-7 Breast Cancer Cells. *Autophagy* 4 (5), 669–679. doi:10.4161/auto.6083

Amaravadi, R. K., Yu, D., Lum, J. J., Bui, T., Christophorou, M. A., Evan, G. I., et al. (2007). Autophagy Inhibition Enhances Therapy-Induced Apoptosis in a Myc-Induced Model of Lymphoma. *J. Clin. Invest.* 117 (2), 326–336. doi:10.1172/jci28833

Byerly, J., Halstead-Nussloch, G., Ito, K., Katsyv, I., and Irie, H. Y. (2016). PRKCQ Promotes Oncogenic Growth and Anoikis Resistance of a Subset of Triple-Negative Breast Cancer Cells. *Breast Cancer Res.* 18 (1), 95. doi:10.1186/s13058-016-0749-6

Camp, R. L., Dolled-Filhart, M., and Rimm, D. L. (2004). X-Tile: A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-point Optimization. *Clin. Cancer Res.* 10 (21), 7252–7259. doi:10.1158/1078-0432.ccr-04-0713

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chiu, H.-W., Ho, Y.-S., and Wang, Y.-J. (2011). Arsenic Trioxide Induces Autophagy and Apoptosis in Human Glioma Cells *In Vitro* and *In Vivo* through Downregulation of Survivin. *J. Mol. Med.* 89 (9), 927–941. doi:10.1007/s00109-011-0763-1

Coccaro, N., Anelli, L., Zagaria, A., Perrone, T., Specchia, G., and Albano, F. (2020). Molecular Complexity of Diffuse Large B-Cell Lymphoma: Can it be a Roadmap for Precision Medicine? *Cancers* 12 (1), 185. doi:10.3390/cancers12010185

Coutinho, R., Clear, A. J., Owen, A., Wilson, A., Matthews, J., Lee, A., et al. (2013). Poor Concordance Among Nine Immunohistochemistry Classifiers of Cell-Of-Origin for Diffuse Large B-Cell Lymphoma: Implications for Therapeutic Strategies. *Clin. Cancer Res.* 19 (24), 6686–6695. doi:10.1158/1078-0432.ccr-13-1482

Cufí, S., Vazquez-Martin, A., Oliveras-Ferraros, C., Martin-Castillo, B., Vellon, L., and Menendez, J. A. (2011). Autophagy Positively Regulates the CD44(+) CD24(-/low) Breast Cancer Stem-Like Phenotype. *Cell Cycle* 10 (22), 3871–3885. doi:10.4161/cc.10.22.17976

Ducie, J., Dao, F., Considine, M., Olvera, N., Shaw, P. A., Kurman, R. J., et al. (2017). Molecular Analysis of High-Grade Serous Ovarian Carcinoma with and without Associated Serous Tubal Intra-Epithelial Carcinoma. *Nat. Commun.* 8 (1), 990. doi:10.1038/s41467-017-01217-9

Evens, A. M., Sehn, L. H., Farinha, P., Nelson, B. P., Raji, A., Lu, Y., et al. (2010). Hypoxia-Inducible Factor-1 {alpha} Expression Predicts Superior Survival in Patients with Diffuse Large B-Cell Lymphoma Treated with R-CHOP. *J. Clin. Oncol.* 28 (6), 1017–1024. doi:10.1200/jco.2009.24.1893

Hans, C. P., Weisenburger, D. D., Greiner, T. C., Gascoyne, R. D., Delabie, J., Ott, G., et al. (2004). Confirmation of the Molecular Classification of Diffuse Large B-Cell Lymphoma by Immunohistochemistry Using a Tissue Microarray. *Blood* 103 (1), 275–282. doi:10.1182/blood-2003-05-1545

He, M., Zhao, Y., Yi, H., Sun, H., Liu, X., and Ma, S. (2015). The Combination of TP53INP1, TP53INP2 and AXIN2: Potential Biomarkers in Papillary Thyroid Carcinoma. *Endocrine* 48 (2), 712–717. doi:10.1007/s12020-014-0341-8

Hönscheid, P., Datta, K., and Muders, M. H. (2014). Autophagy: Detection, Regulation and its Role in Cancer and Therapy Response. *Int. J. Radiat. Biol.* 90 (8), 628–635. doi:10.3109/09553002.2014.907932

Hu, Y., Li, X., Xue, W., Pang, J., Meng, Y., Shen, Y., et al. (2017). TP53INP2-Related Basal Autophagy is Involved in the Growth and Malignant Progression in Human Liposarcoma Cells. *Biomed. Pharmacother.* 88, 562–568. doi:10.1016/j.biopha.2017.01.110

Jang, J. E., Eom, J.-I., Jeung, H.-K., Cheong, J.-W., Lee, J. Y., Kim, J. S., et al. (2017). AMPK-ULK1-Mediated Autophagy Confers Resistance to BET Inhibitor JQ1 in Acute Myeloid Leukemia Stem Cells. *Clin. Cancer Res.* 23 (11), 2781–2794. doi:10.1158/1078-0432.ccr-16-1903

Jiang, G. M., Tan, Y., Wang, H., Peng, L., Chen, H. T., Meng, X. J., et al. (2019). The Relationship between Autophagy and the Immune System and its Applications for Tumor Immunotherapy. *Mol. Cancer* 18 (1), 17. doi:10.1186/s12943-019-0944-z

Juskevicius, D., Lorber, T., Gsponer, J., Perrina, V., Ruiz, C., Stenner-Liewen, F., et al. (2016). Distinct Genetic Evolution Patterns of Relapsing Diffuse Large B-Cell Lymphoma Revealed by Genome-Wide Copy Number Aberration and Targeted Sequencing Analysis. *Leukemia* 30 (12), 2385–2395. doi:10.1038/leu.2016.135

Katheder, N. S., Khezri, R., Schultz, F. S. W., Jain, A., Rahman, M. M., Schink, K. O., et al. (2017). Microenvironmental Autophagy Promotes Tumour Growth. *Nature* 541 (7637), 417–420. doi:10.1038/nature20815

Li, C., Liu, V. W., Chiu, P. M., Yao, K.-M., Ngan, H. Y., and Chan, D. W. (2014). Reduced Expression of AMPK-β1 during Tumor Progression Enhances the Oncogenic Capacity of Advanced Ovarian Cancer. *Mol. Cancer* 13, 49. doi:10.1186/1476-4598-13-49

Li, Y., Zhou, X., Zhang, Y., Yang, J., Xu, Y., Zhao, Y., et al. (2019). CUL4B Regulates Autophagy via JNK Signaling in Diffuse Large B-Cell Lymphoma. *Cell Cycle* 18 (4), 379–394. doi:10.1080/15384101.2018.1560718

Liu, X., and Klionsky, D. J. (2015). TP53INP2/DOR Protein Chaperones Deacetylated Nuclear LC3 to the Cytoplasm to Promote Macroautophagy. *Autophagy* 11 (9), 1441–1442. doi:10.1080/15548627.2015.1074373

Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173 (2), 338–354. e315. doi:10.1016/j.cell.2018.03.034

M. Martelli, A., Evangelisti, C., Ramazzotti, M. G., Fini, M., Giardino, R., Manzoli, L., et al. (2011). Targeting the Phosphatidylinositol 3-kinase/Akt/Mammalian Target of Rapamycin Signaling Network in Cancer Stem Cells. *Curr. Med. Chem.* 18 (18), 2715–2726. doi:10.2174/092986711796011201

Moussay, E., Kaoma, T., Baginska, J., Muller, A., Van Moer, K., Nicot, N., et al. (2011). The Acquisition of Resistance to TNFα in Breast Cancer Cells Is Associated with Constitutive Activation of Autophagy as Revealed by a Transcriptome Analysis Using a Custom Microarray. *Autophagy* 7 (7), 760–770. doi:10.4161/auto.7.7.15454

Pan, T., He, Y., Chen, H., Pei, J., Li, Y., Zeng, R., et al. (2021). Identification and Validation of a Prognostic Gene Signature for Diffuse Large B-Cell Lymphoma Based on Tumor Microenvironment-Related Genes. *Front. Oncol.* 11, 614211. doi:10.3389/fonc.2021.614211

Park, C. K., Kim, W. K., and Kim, H. (2017). Clinicopathological Characteristics of KIT and Protein Kinase C-δ Expression in Adenoid Cystic Carcinoma: Comparison with Chromophobe Renal Cell Carcinoma and Gastrointestinal Stromal Tumour. *Histopathology* 71 (4), 529–542. doi:10.1111/his.13270

Reddy, A., Zhang, J., Davis, N. S., Moffitt, A. B., Love, C. L., Waldrop, A., et al. (2017). Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* 171 (2), 481–494. e415. doi:10.1016/j.cell.2017.09.027

Ruppert, A. S., Dixon, J. G., Salles, G., Wall, A., Cunningham, D., Poeschel, V., et al. (2020). International Prognostic Indices in Diffuse Large B-Cell Lymphoma: a Comparison of IPI, R-IPI, and NCCN-IPI. *Blood* 135 (23), 2041–2048. doi:10.1182/blood.2019002729

Song, C., Pan, B., Yang, X., and Tang, W. (2021). Polyphyllin VII Suppresses Cell Proliferation, the Cell Cycle and Cell Migration in Colorectal Cancer. *Oncol. Lett.* 21 (1), 25. doi:10.3892/ol.2020.12286

Sun, W.-L., Chen, J., Wang, Y.-P., and Zheng, H. (2011). Autophagy Protects Breast Cancer Cells from Epirubicin-Induced Apoptosis and Facilitates Epirubicin-Resistance Development. *Autophagy* 7 (9), 1035–1044. doi:10.4161/auto.7.9.16521

Tian, X. P., Xie, D., Huang, W. J., Ma, S.-Y., Wang, L., Liu, Y.-H., et al. (2020). A Gene-Expression-Based Signature Predicts Survival in Adults with T-Cell Lymphoblastic Lymphoma: A Multicenter Study. *Leukemia* 34 (9), 2392–2404. doi:10.1038/s41375-020-0757-5

Turturro, F. (2015). Constitutive NF- κ B Activation Underlines Major Mechanism of Drug Resistance in Relapsed Refractory Diffuse Large B Cell Lymphoma. *Biomed. Res. Int.* 2015, 484537. doi:10.1155/2015/484537

Villalba, M., and Altman, A. (2002). Protein Kinase C-Theta (PKCtheta), a Potential Drug Target for Therapeutic Intervention with Human T Cell Leukemias. *Curr. Cancer Drug Targets* 2 (2), 125–137. doi:10.2174/1568009023333908

Xia, H., Green, D. R., and Zou, W. (2021). Autophagy in Tumour Immunity and Therapy. *Nat. Rev. Cancer* 21 (5), 281–297. doi:10.1038/s41568-021-00344-2

Xie, Y., Luo, X., He, H., Pan, T., and He, Y. (2021). Identification of an Individualized RNA Binding Protein-based Prognostic Signature for Diffuse Large B-Cell Lymphoma. *Cancer Med.* 10 (8), 2703–2713. doi:10.1002/cam4.3859

Xu, H., Yu, X., Yang, Z., Song, Q., Cheng, S., He, Z., et al. (2021). PAX5-Activated lncRNA ARRDC1-AS1 Accelerates the Autophagy and Progression of DLBCL through Sponging miR-2355-5p to Regulate ATG5. *Life Sci.* 286, 119932. doi:10.1016/j.lfs.2021.119932

Yasrebi, H. (2015). Comparative Study of Joint Analysis of Microarray Gene Expression Data in Survival Prediction and Risk Assessment of Breast Cancer Patients. *Brief. Bioinform* 17 (5), 771–785. doi:10.1093/bib/bbv092

Zeh, H. J., Bahary, N., Boone, B. A., Singhi, A. D., Miller-Ocuin, J. L., Normolle, D. P., et al. (2020). A Randomized Phase II Preoperative Study of Autophagy Inhibition with High-Dose Hydroxychloroquine and Gemcitabine/Nab-Paclitaxel in Pancreatic Cancer Patients. *Clin. Cancer Res.* 26 (13), 3126–3134. doi:10.1158/1078-0432.ccr-19-4042

Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., et al. (2019). Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol. Res.* 7 (5), 737–750. doi:10.1158/2326-6066.cir-18-0436

Zhang, N., Qiu, L., Li, T., Wang, X., Deng, R., Yi, H., et al. (2020). MiR-449a Attenuates Autophagy of T-Cell Lymphoma Cells by Downregulating ATG4B Expression. *BMB Rep.* 53 (5), 254–259. doi:10.5483/bmbrep.2020.53.5.219

Zhong, H., De Marzo, A. M., Laughner, E., Lim, M., Hilton, D. A., Zagzag, D., et al. (1999). Overexpression of Hypoxia-Inducible Factor 1alpha in Common Human Cancers and Their Metastases. *Cancer Res.* 59 (22), 5830–5835.

Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L., and Sun, J. (2017). Discovery and Validation of Immune-Associated Long Non-Coding RNA Biomarkers Associated with Clinically Molecular Subtype and Prognosis in Diffuse Large B Cell Lymphoma. *Mol. Cancer* 16 (1), 16. doi:10.1186/s12943-017-0580-4

Zhou, Z., Sehn, L. H., Rademaker, A. W., Gordon, L. I., LaCasce, A. S., Crosby-Thompson, A., et al. (2014). An Enhanced International Prognostic Index (NCCN-IPI) for Patients with Diffuse Large B-Cell Lymphoma Treated in the Rituximab Era. *Blood* 123 (6), 837–842. doi:10.1182/blood-2013-09-524108

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhou, He, Liu, Lin, Liang, Huang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fgene.2022.848476

BTN3A2 Expression Is Connected With Favorable Prognosis and High Infiltrating Immune in Lung Adenocarcinoma

Yuansheng Lin, Hao Zhou*† and Shengjun Li*†*

Suzhou Science and Technology Town Hospital, Gusu School, Nanjing Medical University, Suzhou, China

OPEN ACCESS

Edited by:
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

Reviewed by:
Yinan Jiang,
University of Pittsburgh, United States
Haiyu Zhou,
Guangdong Provincial People's Hospital, China

***Correspondence:**
Yuansheng Lin
1175312943@qq.com
Hao Zhou
5193446@qq.com
Shengjun Li
linys2020@njmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *09 February 2022*
Accepted: *10 June 2022*
Published: *06 July 2022*

Citation:
Lin Y, Zhou H and Li S (2022) BTN3A2 Expression Is Connected With Favorable Prognosis and High Infiltrating Immune in Lung Adenocarcinoma. Front. Genet. 13:848476. doi: 10.3389/fgene.2022.848476

Background: Butyrophilin subfamily 3 member A2 (BTN3A2) is an important mediator in immune activation, and it is reported to be linked to many cancer progresses. However, the relation with infiltrating immune and prognosis of BTN3A2 in lung adenocarcinoma are not clear.

Methods: In our study, we checked the mRNA expression and protein expression profile of BTN3A2 in lung adenocarcinoma (LUAD) and its relation to clinical outcomes using TIMER and UALCAN databases. In addition, we analyzed the survival of BTN3A2 in LUAD using the Kaplan–Meier Plotter database and PrognoScan database. Moreover, we analyzed gene set enrichment analysis (GSEA) of the BTN3A2. Next, we explored the relation of BTN3A2 expression with the immune infiltration by TIMER. At last, in order to enrich the regulatory mechanism of BTN3A2, we used miRarbase, starbase, and miRDB databases to look for miRNA targets of BTN3A2.

Results: The mRNA along with the protein expression of BTN3A2 in the LUAD group was lower than that in the normal group. In addition, high BTN3A2 expression was connected with good first progression (FP) and overall survival (OS) in LUAD. Then, the GSEA analysis demonstrated that T-cell receptor signaling cascade, B-cell receptor signaling cascade, natural killer cell–mediated cytotoxicity, immune receptor activity, immunological synapse, and T-cell activation were enriched differentially in the BTN3A2 high expression phenotype of LUAD. Moreover, BTN3A2 expression is a remarkable positive correlation with invading levels of tumor purity, B cells, neutrophils, CD4+ T cells, dendritic cells, macrophages, and CD8+ T cells in LUAD, and B cells and dendritic cells were linked with a good prognosis of LUAD. To further enrich the possible regulatory mechanisms of BTN3A2, we analyzed the miRNA targets. The results showed that hsa-miR-17-5p may be miRNA targets of BTN3A2.

Conclusion: Taking together, we provide evidence of BTN3A2 as possible prognosis biomarkers of LUAD. In addition, high BTN3A2 expression in LUAD may influence the

Abbreviations: LUAD, lung adenocarcinoma; PFS, progression-free survival, BTN3A2, butyrophilin subfamily 3 member A2; OS, overall survival; CI, confidence intervals; HR, hazard ratio; TCGA, The Cancer Genome Atlas.

prognosis because of immune invasion. Moreover, our findings provide a potential mechanism that hsa-miR-17-5p may be miRNA targets of BTN3A2.

Keywords: BTN3A2, prognosis, immune infiltration, LUAD, TCGA

BACKGROUND

Lung cancer is the most frequent cause of cancer-linked deaths universally. A total of 1.8 million people are diagnosed with lung cancer and 1.6 million people die of the lung cancer each year (Hirsch et al., 2017). In spite of advances in therapy choices, such as irradiation, iatrochemistry, surgery, and targeted treatment, prognosis remains poor due to the presence of metastatic cancers in most patients at the moment of diagnosis (Chen et al., 2014; Altorki et al., 2019). Thus, finding appropriate prognostic biomarker is crucial for LUAD patients.

Butyrophilin (BTN) family members are immunoglobulin-like molecules, which act as immune checkpoint modulators and play a role in self tolerance. BTN3 family members include BTN3A1, BTN3A2, and BTN3A3. The expression of BTN3A2 in neonatal autoimmune and allergic diseases has a strong causal relationship (Huang et al., 2020). BTN3A2 expression is regulated by stress, which is linked to the prognosis of pancreatic cancer, and may constitute an immune escape mechanism to prevent Vγ 9Vδ2 T-cell recognition (Vantourout et al., 2018). Studies suggest that BTN3A2 is differentially expressed in a variety of cancers. BTN3A2 is expressed in triple-negative breast cancer, pancreatic ductal carcinoma, and epithelial ovarian cancer, and is associated with a good prognosis (Le Page et al., 2012; Cai et al., 2020). BTN3A2 is highly expressed and repressed the growth, migration, and infiltration of gastric cancer cells (Han et al., 2017; Zhu et al., 2017). Nevertheless, the role of BTN3A2 in LUAD is unclear.

Herein, we first explored the mRNA along with the protein expression of BTN3A2 in the LUAD, TIMER, and UALCAN databases. Then, we used the Kaplan–Meier Plotter database to analyze the survival of BTN3A2 in LUAD. Next, we utilized the UALCAN database and GSEA software to analyze the BTN3A2 promoter methylation level. Moreover, we used the TIMER database to explore the relationship of BTN3A2 expression with the immune infiltration. At last, we analyzed miRNA targets to enrich the possible regulatory mechanism of BTN3A2. The results showed that high BTN3A2 expression in LUAD may influence the prognosis because of immune invasion. Moreover, our findings provide a potential mechanism that hsa-miR-17-5p may be miRNA targets of BTN3A2.

RESULTS

Butyrophilin Subfamily 3 Member A2 Expression Levels in Different Human Cancers and Lung Adenocarcinoma

The study design is shown as a flowchart for the analysis of the BTN3A2 gene (**Figure 1A**). To detect BTN3A2 expression in cancer and non-malignant tissues, the BTN3A2 mRNA contents in different cancer type tissues were analyzed using the Sangerbox and TIMER data resources. The results illustrated that the BTN3A2 expression was lower in LUAD, lung squamous cell carcinoma, as well as kidney renal clear cell carcinoma than that in the non-malignant tissues (**Figure 1B**). Moreover, data from TCGA demonstrated that the BTN3A2 content was lower in the LUAD group than that in the healthy group (**Figures 1C,D**). Interestingly, the protein expression of BTN3A2 in LUAD was still much lower than in that in non-malignant tissues from CPTAC samples (UALCAN) (**Figure 1E**). To verify the histological level of BTN3A2, we used the Human Protein Atlas database. The results showed that BTN3A2 is upregulated in normal tissue and downregulated in LUAD tissues (**Figure 7F**). The result was found in immunofluorescence staining (**Supplementary Figure S1**). Otherwise, BTN3A2 is also highly expressed in cells with high CD3 expression.

Butyrophilin Subfamily 3 Member A2 Transcription in Subgroups of Patients With Lung Adenocarcinoma, Stratified Based on Gender, Race, Nodal Status Metastasis, and Other Criteria (UALCAN)

Further analysis of multiple clinic features of 515 LUAD samples in TCGA exhibited the low BTN3A2 mRNA content. The BTN3A2 mRNA content was lower in LUAD patients than that in healthy individuals in subgroup analysis on the basis of gender, race, nodal metastasis, smoking, stages, as well as tumor grade (**Figure 2** and **Table 1**). Therefore, BTN3A2 levels may have a promising LUAD diagnostic value.

Prognostic Potential of Butyrophilin Subfamily 3 Member A2 in Lung Adenocarcinoma

To explore whether BTN3A2 expression was related to the prognosis of LUAD patients, we evaluated the influence of BTN3A2 contents on OS and FP through the Kaplan–Meier plotter data resource. The data demonstrated that the high BTN3A2 contents were linked to the good prognosis in LUAD from 204820_s_at probe (OS HR = 0.71, p = 8.5e-08; FP HR = 0.71, p = 5.5e-04) (**Figures 3A,B**). The results were consistent with those in 209846_s_at probe and 212613_s_at probe from the Kaplan–Meier plotter (**Figures 3C–F**). To further verify the prognostic value of BTN3A2 in LUAD, the PrognoScan database was used and the results showed that the low BTN3A2 expression group was linked to good lung cancer prognosis (OS jacob-00182-CANDF HR = −0.72, p = 0.00027; RFS GSE31210 HR = −0.28, p = 0.033) (**Figures 3G,H**). Thus, our

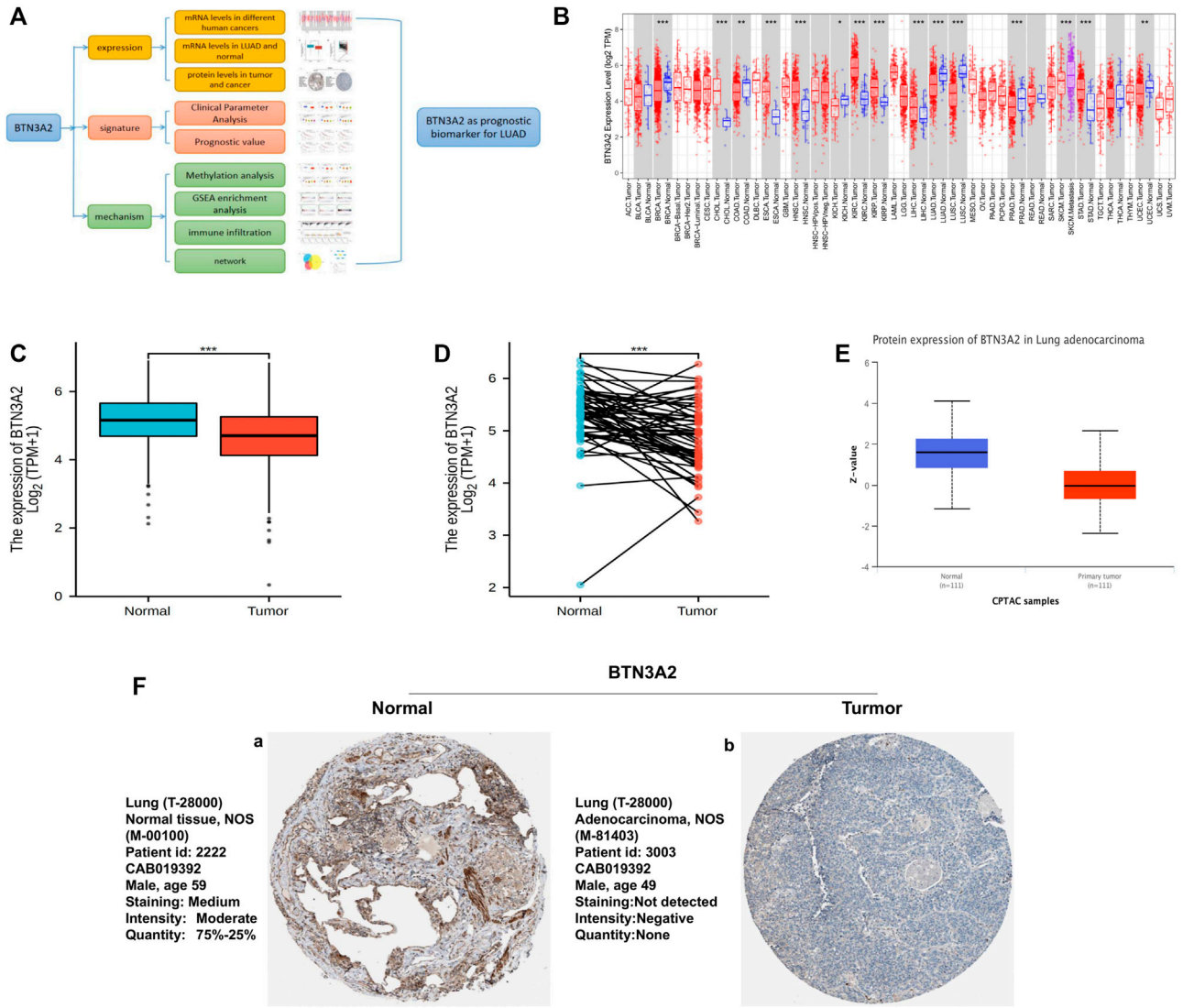


FIGURE 1 | BTN3A2 contents in various human cancers and LUAD. **(A)** Workflow of this study. **(B)** BTN3A2 contents in different cancers from TCGA data resource were analyzed by TIMER. **(C)** BTN3A2 contents in non-malignant and tumor tissues of LUAD patients in the TCGA database. **(D)** BTN3A2 contents in pre-disease and post-disease form the same sample. **(E)** UALCAN database showed the protein expression of BTN3A2 in non-malignant and LUAD groups from CPTAC samples. **(F)** Expression of BTN3A2 in LUAD samples and normal tissues in the Human Protein Atlas.

data suggested that the low BTN3A2 content is a protective factor and results in a good prognosis in LUAD.

Gene Set Enrichment Analysis Showed the Most Significant Pathways and Genes Related With Butyrophilin Subfamily 3 Member A2 Based on The Cancer Genome Atlas

To identify the different signal pathways activated in LUAD, we used GSEA between high and low BTN3A2 expression datasets. A total of six hallmark gene-sets (KEGG_T_CELL_RECEPTOR_SIGNALING_CASCADE, KEGG_B_CELL_RECEPTOR_SIGNALING_CASCADE, KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY, GO_IMMUNE_RECEPTOR_ACTIVITY, GO_IMMUNOLOGICAL_SYNAPSE, and GO_T_CELL_ACTIVATION) were chosen for analysis (**Figures 4A–F**). The results showed that T-cell receptor signaling cascade, B-cell receptor signaling cascade, natural killer cell–mediated cytotoxicity, immune receptor activity, immunological synapse, and T-cell activation are differentially enriched in the BTN3A2 high expression phenotype of LUAD. Moreover, the bar graph shows the top 10 messages of biological processes, cell components, molecular functions, and KEGG, respectively. The GO term annotation indicated that these genes primarily participated in cell proliferation, multicellular organismal

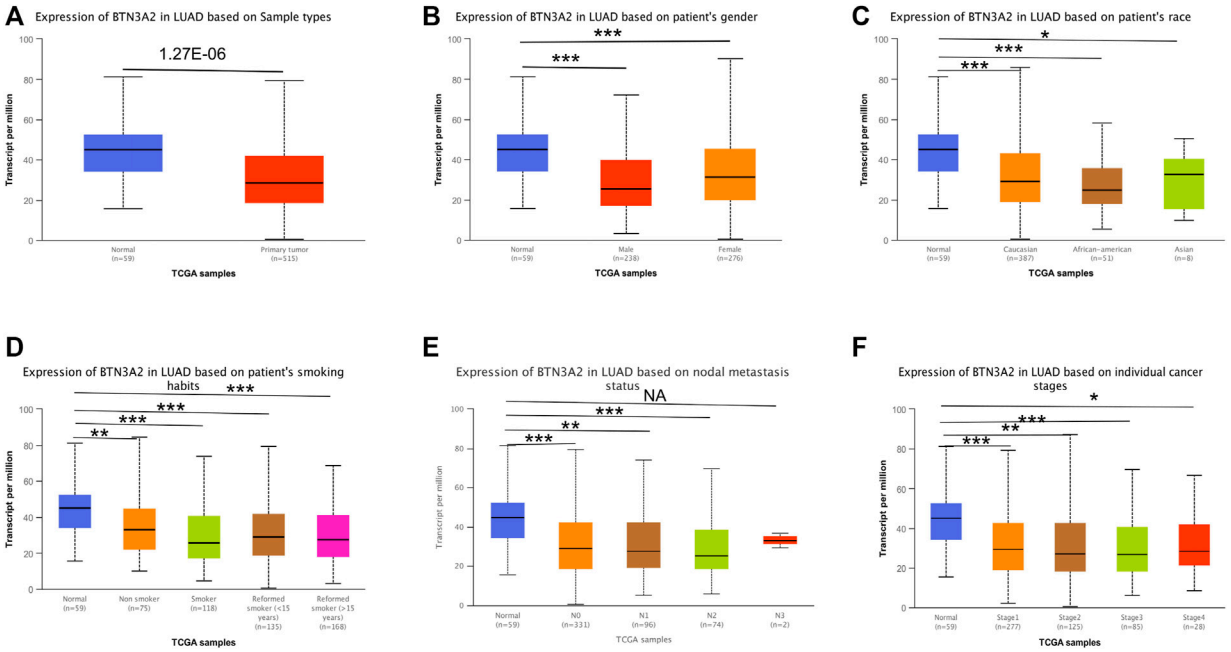


FIGURE 2 | BTN3A2 contents in LUAD (UALCAN). **(A–F)** BTN3A2 contents were low in LUAD. (NA, $p > 0.05$, *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$).

TABLE 1 | Baseline characteristics of patients with LUAD.

Characteristic	Low expression of BTN3A2	High expression of BTN3A2	*p*
N	267	268	
T stage, n (%)			0.069
T1	76 (14.3%)	99 (18.6%)	
T2	151 (28.4%)	138 (25.9%)	
T3	31 (5.8%)	18 (3.4%)	
T4	9 (1.7%)	10 (1.9%)	
N stage, n (%)			0.564
N0	171 (32.9%)	177 (34.1%)	
N1	48 (9.2%)	47 (9.1%)	
N2	43 (8.3%)	31 (6%)	
N3	1 (0.2%)	1 (0.2%)	
M stage, n (%)			0.488
M0	183 (47.4%)	178 (46.1%)	
M1	15 (3.9%)	10 (2.6%)	
Age, median(IQR)	65 (58, 72)	67 (59, 73)	0.104

process, cell communication, protein binding, and structural molecule activity (**Supplementary Figure S2A**). The KEGG pathway analysis indicated that these genes chiefly participated in peroxisome and cell adhesion molecules (**Supplementary Figure S2B**).

Relationship Between Expression of Butyrophilin Subfamily 3 Member A2 and Immune Invasion in Lung Adenocarcinoma

To assess the potential relationship of immune invasion with BTN3A2 expression in LUAD, we used TIMER to conduct the following analysis. First, the "Gene" module analysis revealed that BTN3A2 expression is remarkably positively correlated with invading levels of tumor purity, B cells, dendritic cells, CD8+ T cells, macrophages, neutrophils, and CD4+ T cells in LUAD (**Figure 5A**). Then, the "SCNA" module analysis revealed immune cell infiltration may relevant to altered BTN3A2 gene copy numbers, including B cells, dendritic cells, CD4+ T cells, macrophages, CD8+ T cells, and neutrophils in LUAD (**Figure 5B**). In addition, the "SURVIVAL" module analysis demonstrated that high B-cell and dendritic cell levels were linked to a good prognosis of LUAD (**Figure 5C**, $p < 0.05$, respectively).

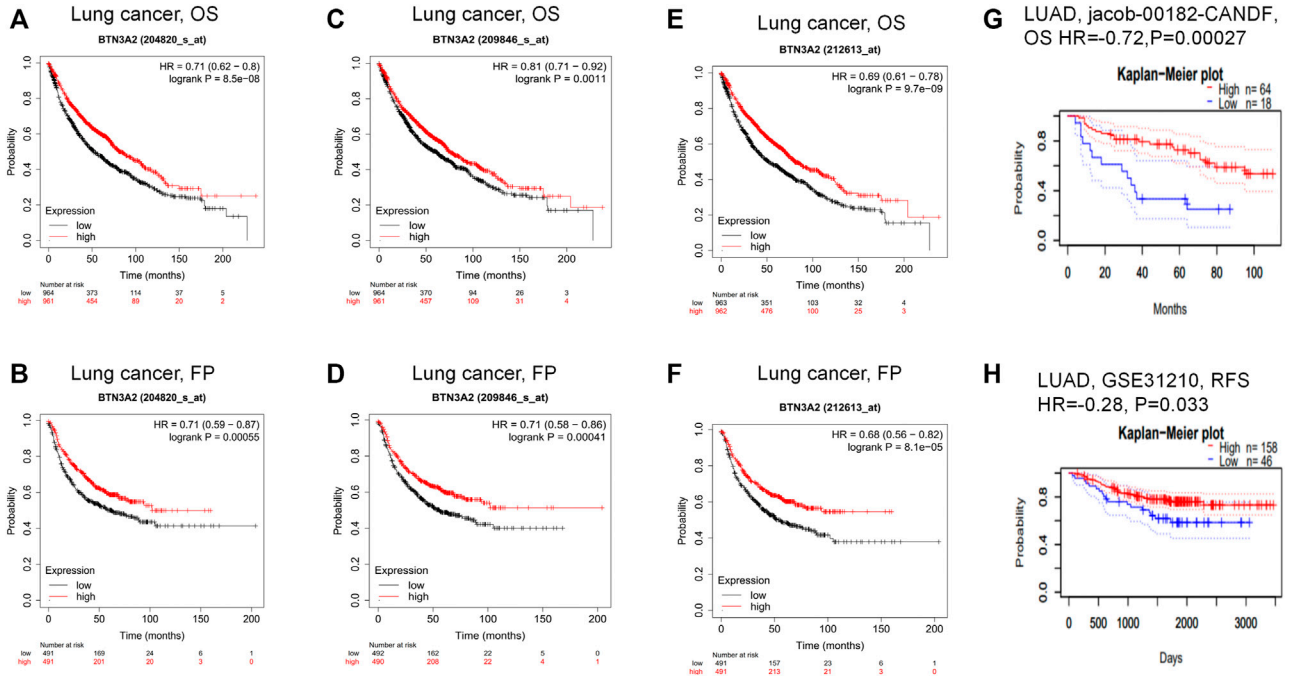


FIGURE 3 | Comparing the high and low expression of BTN3A2 in LUAD by Kaplan–Meier survival curves. **(A–B)** OS and FP survival curves of lung cancer in 204820_s_at probe (n = 1295, n = 982). **(C–D)** OS and FP survival curves of lung cancer in 209846_s_at probe. (n = 1295, n = 982). **(E–F)** FP and OS survival curves of lung cancer in 212613_s_at probe (n = 1295, n = 982). **(G–H)** OS and RFS survival curves of LUAD in jacob-00182-CANDF and GSE31210, respectively. OS, overall survival; FP, first progression; RFS, relapse free survival.

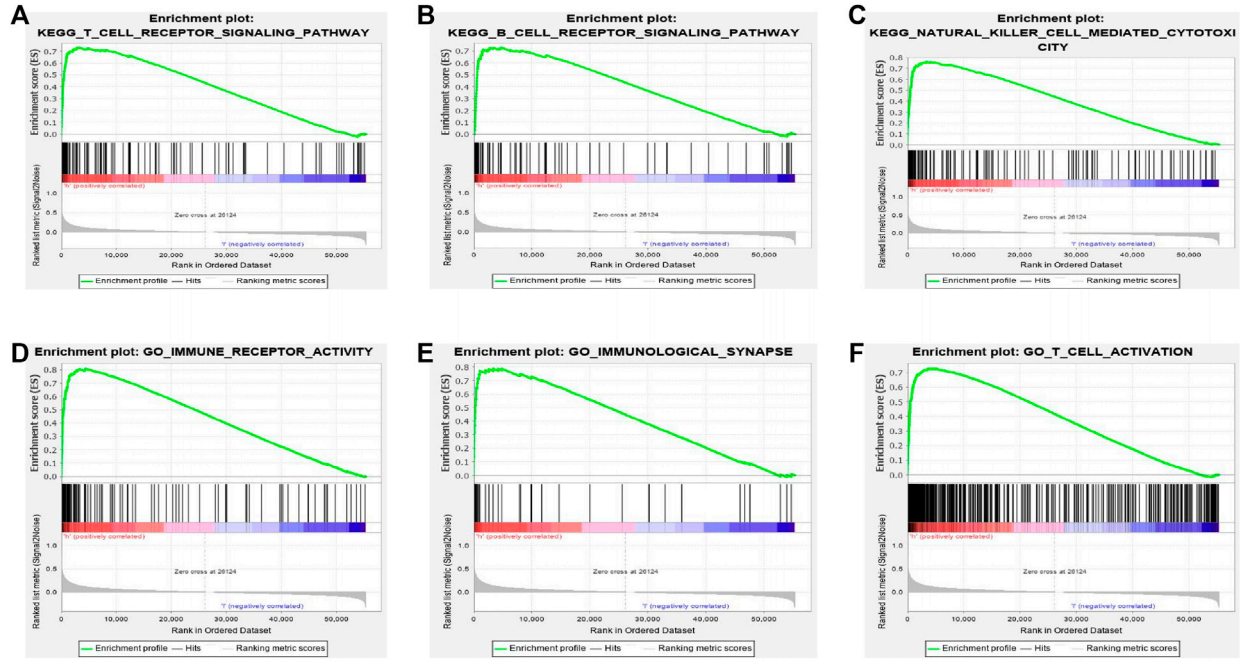


FIGURE 4 | Gene set enrichment analysis (GSEA) indicated the most significant pathways and genes related with BTN3A2 based on TCGA. GSEA results show gene sets six 6 hallmark gene-sets, **(A)** "KEGG_T_CELL_RECEPTOR_SIGNALING_CASCADE," **(B)** "KEGG_B_CELL_RECEPTOR_SIGNALING_CASCADE," **(C)** "KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY," **(D)** "GO_IMMUNE_RECEPTOR_ACTIVITY," **(E)** "GO_IMMUNOLOGICAL_SYNAPSE," and **(F)** "GO_T_CELL_ACTIVATION."

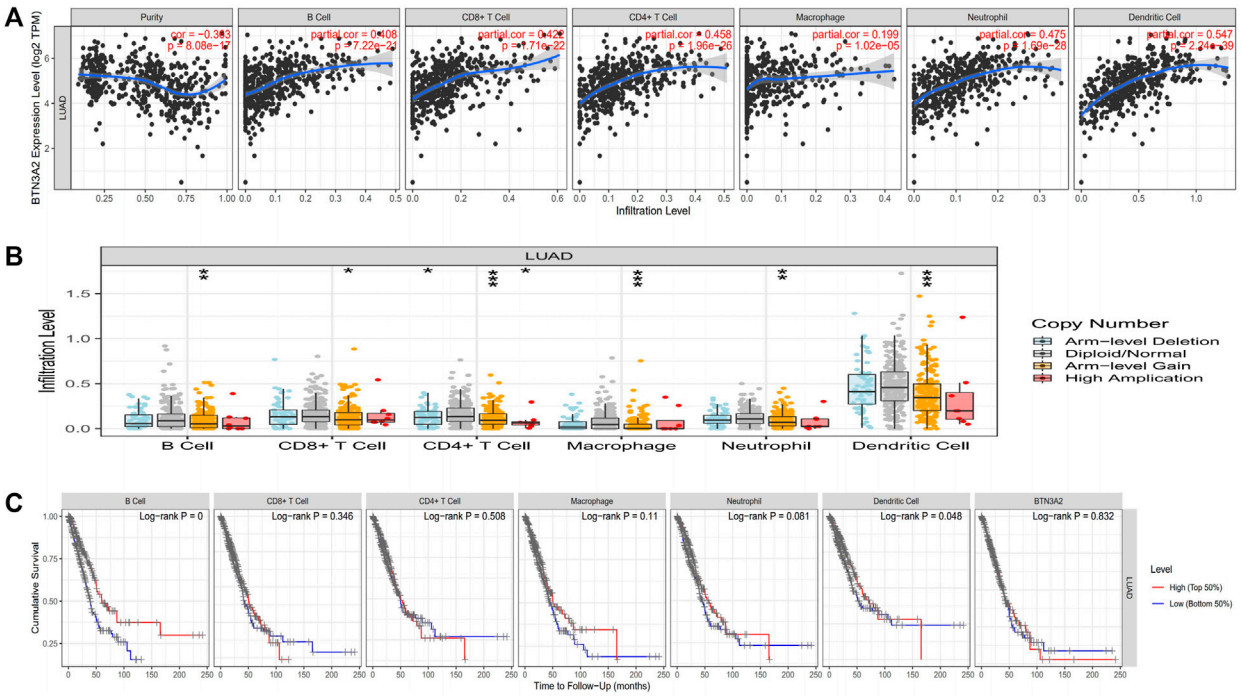


FIGURE 5 | Expression of BTN3A2 in LUAD was linked to immune invasion. **(A)** BTN3A2 expression has positive correlations with the infiltration levels of purity, B cells, neutrophils, CD8+ T cells, dendritic cells, CD4+ T cells, and macrophages in LUAD. **(B)** BTN3A2 CNV influences the infiltration level of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in LUAD. **(C)** Kaplan–Meier plots of the relationship of immune invasion with OS of LUAD. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

To study the relationship between immune cell markers and BTN3A2 expression, we analyzed the markers of B cells, macrophages, T cells, dendritic cells, and neutrophils by the TIMER data resource. Interestingly, we discovered that the expression levels of CD19 along with CD79A in B cells, CD3E, CD2, CD8A, and CD8B in T cells, CEACAM8, ITGAM, and CCR7 in neutrophils, T helper cells (Th1 and Th2), macrophages, and dendritic cell have markedly positively related to the BTN3A2 expression in LUAD (**Table 2**).

The data imply that high BTN3A2 expression in LUAD may influence the prognosis because of immune invasion.

Hsa-miR-17-5p May Be miRNA Targets of Butyrophilin Subfamily 3 Member A2

To further elucidate the miRNA targets of BTN3A2 in LUAD, we used miRarbase, starbase, and miRDB databases to predict miRNA-target genes of BTN3A2 shown in the Venn plot (**Figure 6A**). The result showed that it had six miRNA targets (hsa-miR-17-5p, hsa-miR-20a-5p, hsa-miR-93-5p, hsa-miR-106b-5p, hsa-miR-20b-5p, and hsa-miR-519d-3p) of BTN3A2 and which was visualized in **Figure 6B**. Then, we analyzed the prognostic potential of 6 miRNA-target in lung cancer using the Kaplan–Meier Plotter database. Interestingly, we found that only hsa-miR-17 had differential expression (p = 1.62E-12) and prognostic value (OS, HR = 1.4, p = 0.035) (**Figures 6C,D**). Moreover, the target site in the BTN3A2 3'UTR was predicted to pair with hsa-miR-17-5p by miRDB and RNA22Sites (**Figure 6E**).

We have confirmed that BTN3A2 expression was related to the immune infiltration in LUAD, and the height expression of BTN3A2 was also correlated with the better prognosis of LUAD. Therefore, we hypothesized that hsa-miR-17 expression may be the miRNA targets of BTN3A2. The prognosis analysis on the basis of the expression of hsa-miR-17 of LUAD in linked immune cell subtype was performed using the Kaplan–Meier plotter, and data illustrated that the high hsa-miR-17 content of LUAD in abundant B cells (HR = 0.03), abundant CD4+ memory T cells (HR = 0.042), abundant macrophages (HR = 0.04), abundant regulatory T cells (HR = 0.031), and type 2 T helper cells (HR = 0.0077) had a poor prognosis (**Figures 7A–E**). Thus, the results may provide a potential mechanism that hsa-miR-17-5p may be miRNA targets of BTN3A2.

DISCUSSION

Lung cancer is the primary cause of cancer death due to its poor prognosis and high aggressiveness. Immunotherapy, especially the emergence of immune checkpoint inhibitors (ICIs), has changed the therapeutic prospect of lung cancer (Sun et al., 2007; Peng et al., 2019; Rosner et al., 2019). The development of lung cancer is a complex process involving interactions between stromal fibroblasts, immune cells, and tumor cells.

TABLE 2 | Correlation analysis between BTN3A2 and immune cell type markers in the TIMER database.

Description	Gene marker	COAD					
		None		Tumor		Age	
		COR	*p*	COR	*p*	COR	*p*
$CD8^+$ T cells	CD8A	0.534	<0.001	0.448	9.30e-26	0.542	3.71e-38
	CD8B	0.416	<0.001	0.329	6.62e-14	0.425	1.30e-22
T cells	CD3E	0.62	4.79e-56	0.552	1.27e-40	0.624	1.48e-53
	CD2	0.601	7.67e-52	0.526	2.24e-36	0.604	2.44e-49
B cells	CD19	0.373	1.96e-18	0.246	3.32e-08	0.393	2.98e-19
	CD79A	0.354	1.07e-16	0.23	2.46e-07	0.372	2.61e-17
Macrophages	INOS (NOS2)	0.091	3.99e-2	0.006	8.86e-01	0.093	4.04e-02
	IRF5	0.371	<0.001	0.298	1.36e-11	0.38	4.38e-18
	COX2(PTGS2)	0.089	4.27e-02	-0.099	2.84e-02	-0.086	5.88e-02
Dendritic cells	HLA-DQB1	0.437	1.95e-25	0.363	8.45e-17	0.437	6.30e-24
	HLA-DRA	0.556	<0.001	0.49	3.40e-31	0.546	6.23e-39
	HLA-DPA1	0.597	<0.001	0.544	2.96e-39	0.587	5.14e-46
	BDCA-1(CD1C)	0.288	2.75e-11	0.226	3.82e-07	0.267	2.66e-09
Neutrophils	CD66b (CEACAM8)	0.114	9.6e-03	0.11	1.44e-02	0.085	6.34e-02
	CD11b (ITGAM)	0.485	<0.001	0.418	2.86e-22	0.486	5.44e-30
	CCR7	0.516	<0.001	0.436	3.10e-24	0.508	5.52e-33
Th1	TBX21	0.565	<0.001	0.5	1.52e-32	0.572	2.57e-43
	STAT4	0.519	7.53e-37	0.44	8.39e-35	0.522	3.46e-35
	STAT1	0.554	<0.001	0.496	5.25e-32	0.567	1.75e-42
	IFN-γ(IFNG)	0.431	1.03e-24	0.354	5.23e-16	0.445	8.05e-25
	TNF-α (TNF)	0.312	6.02e-13	0.204	5.10e-06	0.312	2.33e-12
Th2	GATA3	0.552	2.01e-42	0.477	1.93e-29	0.556	1.68e-40
	STAT6	0.181	3.72e-05	0.209	2.93e-06	0.171	1.54e-04
	STAT5A	0.564	1.20e-44	0.497	4.12e-32	0.569	8.13e-43
	IL13	0.166	1.57e-04	0.106	1.87e-02	0.165	2.60e-04

LUAD, lung adenocarcinoma; COR, R value of Spearman's correlation.

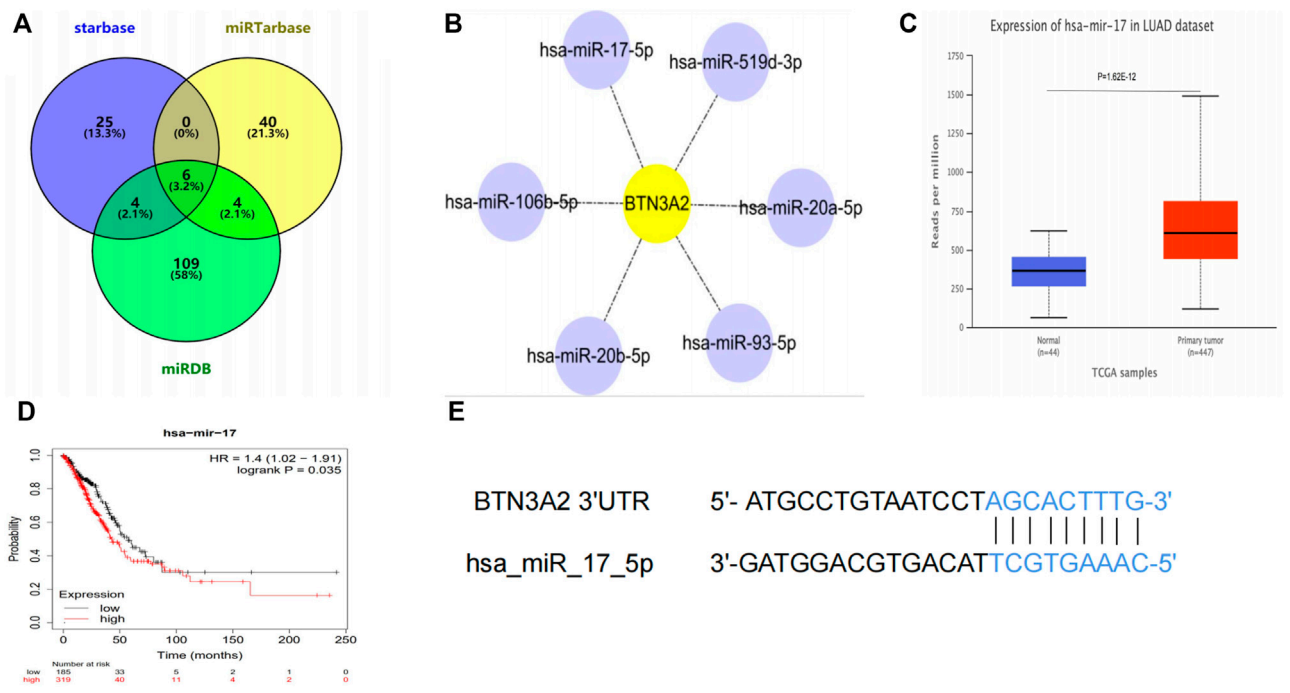


FIGURE 6 | hsa-miR-17-5p may be the target of BTN3A2. **(A)** Venn plot shows miRNA-target genes predicted by miRarbase, starbase, and miRDB. **(B)** miRNA-target regulatory network (Cytoscape). **(C)** Expression of hsa-miR-17 in the UALCAN database. **(D)** Survival curve of hsa-miR-17 in lung cancer by the Kaplan–Meier Plotter database. **(E)** Base pairing between hsa_miR_17_5p and the target site in the BTN3A2 3'UTR predicted by miRDB and RNA22Sites.

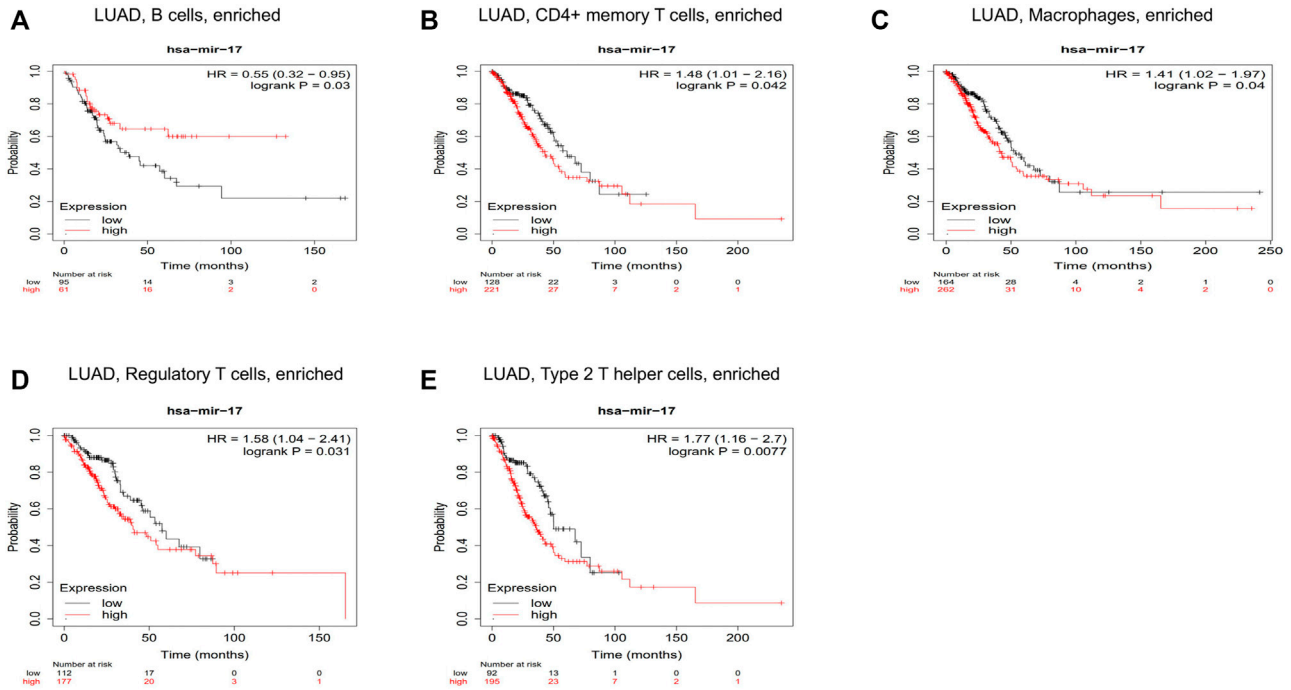


FIGURE 7 | Correlations between hsa-miR-17 expression and immune infiltration in LUAD. Kaplan–Meier survival curves compared between over-expression and down-expression of hsa-miR-17 in LUAD on the basis of immune cells subtypes. Association between LUAD of different immune cells subtype and prognosis **(A–E)**.

Tumor-infiltrating immune cells have an important role in inhibiting or promoting tumor growth (Wang et al., 2019; Öjlert et al., 2019). The immune cell response is related to the clinical prognosis of LUAD, and cancer-linked memory T cells can have anti-tumor immunity of LUAD (Snyder and Farber, 2019). These analyses point out that BTN3A2 may be a possible target for cancer-targeted therapy. Nevertheless, there are few research studies on infiltrating immune and prognosis of BTN3A2 in LUAD.

BTN3A2 is associated with the prognosis of breast cancer (BRCA) and ovarian cancer (OV), and immunoinfiltration in triple-negative breast cancer (TNBC) and OV is highly associated with BTN3A2 (Le Page et al., 2012; Cai et al., 2020). Studies have shown that the prognosis of LUAD is linked to immune invasion (Kadara et al., 2017; Ghorani et al., 2018; Gong et al., 2019). Herein, our study is consistent with previous studies and results show that BTN3A2 expression was remarkably positively linked to invading levels of tumor purity, B cells, neutrophils, CD8+ T cells, macrophages, and dendritic cells, and CD4+ T cells, B cells, and dendritic cells were linked with a good prognosis of LUAD. The immune cell markers in LUAD were further studied, after correction of tumor purity, BTN3A2 in LUAD was significantly positively correlated with gene markers in B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells (**Table 2**). GSEA analysis also suggested that BTN3A2 was associated with immune-related pathways. These strongly confirmed the positive correlation between BTN3A2 and immune infiltration in LUAD. Prognostic analysis of BTN3A2 expression levels in LUAD based on immune cells was performed, the high BTN3A2 expression level in LUAD had a good prognosis in the enriched high B cells and dendritic cells. The analysis indicated that high BTN3A2 expression is connected with favorable prognosis in LUAD and high BTN3A2 expression in LUAD may affect prognoses due to immune infiltration.

Hsa-miR-17 targets transcriptional co-regulators that synergize with Foxp3 and reduce the inhibitory activity of regulatory T cells (Yang et al., 2016). The prognosis analysis on the basis of the expression of hsa-miR-17 of LUAD in linked immune cell subtype was performed using the Kaplan–Meier plotter. Our results demonstrated that the high expression of hsa-miR-17 of LUAD in abundant B cells, $CD4^+$ memory T cells, macrophages, regulatory T cells, and type 2 T helper cells had a poor prognosis (**Figures 7A–E**). Our analysis speculated that hsa-miR-17 inhibited the infiltration of regulatory T cells in LUAD and played an important role in promoting tumor proliferation. These findings indicate that hsa-miR-17 may be an immunotherapy target of LUAD.

BTN3A is a prognostic marker and target of Vγ9Vδ2 T cell immunotherapy for pancreatic ductal adenocarcinoma (PDAC) (Benyamine et al., 2017). To further illustrate whether BTN3A2 had other regulated mechanisms in LUAD. Multiple website tools were used to confirm BTN3A2-target miRNAs, and hsa-miR-17-5p was obtained. The result showed that hsa-miR-17-5p was over-expressed in LUAD tissues, which might be another mechanism for the downregulated expression of BTN3A2. Previous studies have shown that hsa-miR-17-5p was also confirmed to promote the invasion and migration of colorectal cancer (Xi et al., 2016; Yu et al., 2022). In the meantime, hsa-miR-

17-5p was differentially expressed in numerous cancer types and its over-expression could indicate a poorer survival in lung cancer (Chen et al., 2013). The analysis results of hsa-miR-17 are the same as the aforementioned studies. Thus, the results may provide a potential mechanism that hsa-miR-17-5p may be miRNA targets of BTN3A2 in LUAD.

However, there are still some limitations to this study. First, due to the small sample size of this study and the heterogeneity of LUAD, the interpretation of the study results needs to be cautious. In addition, the specific pathophysiological mechanisms of BTN3A2 regulating the initiation and progression of LUAD need to be further studied. Finally, the working mechanism of BTN3A2 is not yet fully understood, so more evidence is needed to discover its biological basis and further exploration is needed to lucubrate the exact mechanism and function of BTN3A2 in LUAD progression.

CONCLUSION

In summary, we provide evidence of BTN3A2 as possible prognosis biomarkers of LUAD. In addition, high BTN3A2 expression in LUAD may influence the prognosis because of immune invasion. Moreover, our findings provide a potential mechanism that hsa-miR-17-5p may be miRNA targets of BTN3A2 in LUAD.

METHODS

UALCAN Analysis

We analyzed the protein expression along with mRNA expression levels of SKA3 in the UALCAN data resource (http://ualcan.path.uab.edu/index.html) (Chandrashekar et al., 2017).

TIMER Analysis

BTN3A2 expression in cancers was determined by the TIMER data resource (https://cistrome.shinyapps.io/timer/) (Li et al., 2017). We identified the expression of BTN3A2 in different tumors by the TIMER data resource. After that, the association of BTN3A2 with immune invasion in LUAD was estimated by the TIMER algorithm. Furthermore, the relationship of ACE2 with the type biomarkers of tumor purity, B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells in LUAD was verified.

Human Protein Atlas

We used the Human Protein Atlas (https://www.proteinatlas.org/) to detect BTN3A2 protein expression in immunohistochemistry.

Kaplan–Meier Plotter Databases

We employed the Kaplan–Meier plotter data resource (http://kmplot.com/) (Győrffy et al., 2010) to explore the value of BTN3A2 expression in LUAD prognosis. This threshold was logrank p value < 0.05 in the Kaplan–Meier plotter data resource.

PrognoScan Database

We analyzed the prognosis in LUAD using the PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) (Mizuno et al., 2009). Cox p-value < 0.05 was considered as the significant difference.

Gene Set Enrichment Analysis

GSEA is the most commonly used statistical method (Subramanian et al., 2005). GSEA was performed to clarify the molecular mechanisms of the prognostic gene signature. GSEA was performed using GSEA v. 3.0 and was searched to determine the enriched biological processes, cellular components, molecular functions, and KEGG pathway associated with survival of the high-risk group. FDR < 0.05 and |NES| > 1 were considered statistically significant. Then, we used the STRING website to analyze co-expression genes of BTN3A2. Also, we analyzed GO and KEGG pathway using WebGestalt.

To further explore the targets of BTN3A2 in LUAD, we used miRarbase, starbase, and miRDB databases to predict miRNA-target genes of BTN3A2. Moreover, the target site in the BTN3A2 3'UTR was predicted to pair with hsa-miR-17-5p by miRDB (http://www.mirdb.org/) (Chen and Wang, 2020) and RNA22Sites (https://cm.jefferson.edu/rna22v2.0/) (Miranda et al., 2006).

Immunofluorescence Staining

The human primary LUAD tissue and corresponding normal tissue were collected from LUAD patients in the Suzhou Science and Technology Town Hospital, Gusu School, Nanjing Medical University. The embedded LUAD and normal clinical sample tissue were sliced and routinely deparafnized to water as described before. The tissue sections were blocked with goat serum after antigen retrieval. These were incubated overnight with anti-BTN3A2 and anti-CD3. The tissue was treated with rhodamine-labeled anti-mouse IgG and fluorescein-labeled anti-rabbit IgG (Qi et al., 2021). DAPI was used for nuclear staining. The images were captured using a confocal microscope (DM6000 CFS; Leica) and processed using LAS AF software.

Statistical Analysis

The results of the survival analysis were acquired from a log-rank test. We used Spearman's correlation to evaluate BTN3A2 with immune invasion and type biomarkers of immune cells. Comparing two independent samples using Student's t test. $p < 0.05$ signified statistical significance.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Medical Ethics Committee of Suzhou Science and;

Technology Town Hospital. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

YL conceived the project. HZ analyzed the data. YL, HZ, and SL contributed toward the interpretation of the data. All authors contributed to the article and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.848476/full#supplementary-material

Supplementary Figure S1 | Immunofluorescence staining for BTN3A2 and CD3. Double immunofluorescence staining was used to examine the expression of BTN3A2 and CD3 in tumor and normal LUAD tissues.

Supplementary Figure S2 | Gene enrichment analysis of BTN3A2 in LUAD cohorts. **(A)** Enriched KEGG pathway of BTN3A2-related genes. **(B)** Enriched GO terms of BTN3A2-related genes.

REFERENCES

Altorki, N. K., Markowitz, G. J., Gao, D., Port, J. L., Saxena, A., Stiles, B., et al. (2019). The Lung Microenvironment: an Important Regulator of Tumour Growth and Metastasis. *Nat. Rev. Cancer* 19 (1), 9–31. doi:10.1038/s41568-018-0081-9

Benyamine, A., Loncle, C., Foucher, E., Blazquez, J.-L., Castanier, C., Chrétien, A.-S., et al. (2017). BTN3A Is a Prognosis Marker and a Promising Target for Vγ9Vδ2 T Cells Based-Immunotherapy in Pancreatic Ductal Adenocarcinoma (PDAC). *Oncoimmunology* 7 (1), e1372080. doi:10.1080/2162402x.2017.1372080

Cai, P., Lu, Z., Wu, J., Qin, X., Wang, Z., Zhang, Z., et al. (2020). BTN3A2 Serves as a Prognostic Marker and Favors Immune Infiltration in Triple-negative Breast Cancer. *J. Cell Biochem.* 121 (3), 2643–2654. doi:10.1002/jcb.29485

Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B. V. S. K., et al. (2017). UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19 (8), 649–658. doi:10.1016/j.neo.2017.05.002

Chen, Q., Si, Q., Xiao, S., Xie, Q., Lin, J., Wang, C., et al. (2013). Prognostic Significance of Serum miR-17-5p in Lung Cancer. *Med. Oncol.* 30 (1), 353. doi:10.1007/s12032-012-0353-2

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48 (D1), D127–D131. doi:10.1093/nar/gkz757

Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F., and Wong, K.-K. (2014). Non-small-cell Lung Cancers: a Heterogeneous Set of Diseases. *Nat. Rev. Cancer* 14 (8), 535–546. doi:10.1038/nrc3775

Ghorani, E., Rosenthal, R., McGranahan, N., Reading, J. L., Lynch, M., Peggs, K. S., et al. (2018). Differential Binding Affinity of Mutated Peptides for MHC Class I Is a Predictor of Survival in Advanced Lung Cancer and Melanoma. *Ann. Oncol.* 29 (1), 271–279. doi:10.1093/annonc/mdx687

Gong, Z., Jia, Q., Chen, J., Diao, X., Gao, J., Wang, X., et al. (2019). Impaired Cytolytic Activity and Loss of Clonal Neoantigens in Elderly Patients with Lung Adenocarcinoma. *J. Thorac. Oncol.* 14 (5), 857–866. doi:10.1016/j.jtho.2019.01.024

Győrffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., et al. (2010). An Online Survival Analysis Tool to Rapidly Assess the Effect of 22,277 Genes on Breast Cancer Prognosis Using Microarray Data of 1,809 Patients. *Breast Cancer Res. Treat.* 123 (3), 725–731. doi:10.1007/s10549-009-0674-9

Han, J., Choi, Y.-L., Kim, H., Choi, J. Y., Lee, S. K., Lee, J. E., et al. (2017). MMP11 and CD2 as Novel Prognostic Factors in Hormone Receptor-Negative, HER2-Positive Breast Cancer. *Breast Cancer Res. Treat.* 164 (1), 41–56. doi:10.1007/s10549-017-4234-4

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Wu, Y.-L., et al. (2017). Lung Cancer: Current Therapies and New Targeted Treatments. *Lancet* 389 (10066), 299–311. doi:10.1016/s0140-6736(16)30958-8

Huang, Q. Q., Tang, H. H. F., Teo, S. M., Mok, D., Ritchie, S. C., Nath, A. P., et al. (2020). Neonatal Genetics of Gene Expression Reveal Potential Origins of Autoimmune and Allergic Disease Risk. *Nat. Commun.* 11 (1), 3761. doi:10.1038/s41467-020-17477-x

Kadara, H., Choi, M., Zhang, J., Parra, E. R., Rodriguez-Canales, J., Gaffney, S. G., et al. (2017). Whole-exome Sequencing and Immune Profiling of Early-Stage Lung Adenocarcinoma with Fully Annotated Clinical Follow-Up. *Ann. Oncol.* 28 (1), 75–82. doi:10.1093/annonc/mdw436

Le Page, C., Marineau, A., Bonza, P. K., Rahimi, K., Cyr, L., Labouba, I., et al. (2012). BTN3A2 Expression in Epithelial Ovarian Cancer Is Associated with Higher Tumor Infiltrating T Cells and a Better Prognosis. *PLoS One* 7 (6), e38541. doi:10.1371/journal.pone.0038541

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.can-17-0307

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A. M., et al. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell* 126 (6), 1203–1217. doi:10.1016/j.cell.2006.07.031

Mizuno, H., Kitada, K., Nakai, K., and Sarai, A. (2009). PrognoScan: a New Database for Meta-Analysis of the Prognostic Value of Genes. *BMC Med. Genomics* 2, 18. doi:10.1186/1755-8794-2-18

Öjlert, Å. K., Halvorsen, A. R., Nebdal, D., Lund-Iversen, M., Solberg, S., Brustugun, O. T., et al. (2019). The Immune Microenvironment in Non-small Cell Lung Cancer Is Predictive of Prognosis after Surgery. *Mol. Oncol.* 13 (5), 1166–1179. doi:10.1002/1878-0261.12475

Peng, S., Wang, R., Zhang, X., Ma, Y., Zhong, L., Li, K., et al. (2019). EGFR-TKI Resistance Promotes Immune Escape in Lung Cancer via Increased PD-L1 Expression. *Mol. Cancer* 18 (1), 165. doi:10.1186/s12943-019-1073-4

Qi, J-C., Yang, Z., Lin, T., Ma, L., Wang, Y-X., Zhang, Y., et al. (2021). CDK13 Upregulation-Induced Formation of the Positive Feedback Loop Among circCDK13, miR-212-5p/miR-449a and E2F5 Contributes to Prostate Carcinogenesis. *J. Exp. Clin. cancer Res. CR* 40 (1), 2. doi:10.1186/s13046-020-01814-5

Rosner, S., Reuss, J. E., and Forde, P. M. (2019). PD-1 Blockade in Early-Stage Lung Cancer. *Annu. Rev. Med.* 70, 425–435. doi:10.1146/annurev-med-050217-025205

Snyder, M. E., and Farber, D. L. (2019). Human Lung Tissue Resident Memory T Cells in Health and Disease. *Curr. Opin. Immunol.* 59, 101–108. doi:10.1016/j.coi.2019.05.011

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, S., Schiller, J. H., and Gazdar, A. F. (2007). Lung Cancer in Never Smokers - a Different Disease. *Nat. Rev. Cancer* 7 (10), 778–790. doi:10.1038/nrc2190

Vantourout, P., Laing, A., Woodward, M. J., Zlatareva, I., Apolonia, L., Jones, A. W., et al. (2018). Heteromeric Interactions Regulate Butyrophilin (BTN) and BTN-like Molecules Governing γδ T Cell Biology. *Proc. Natl. Acad. Sci. U.S.A.* 115 (5), 1039–1044. doi:10.1073/pnas.1701237115

Wang, S.-s., Liu, W., Ly, D., Xu, H., Qu, L., and Zhang, L. (2019). Tumor-infiltrating B Cells: Their Role and Application in Anti-tumor Immunity in Lung Cancer. *Cell Mol. Immunol.* 16 (1), 6–18. doi:10.1038/s41423-018-0027-x

Xi, X.-P., Zhuang, J., Teng, M.-J., Xia, L.-J., Yang, M.-Y., Liu, Q.-G., et al. (2016). MicroRNA-17 Induces Epithelial-Mesenchymal Transition Consistent with the Cancer Stem Cell Phenotype by Regulating CYP7B1 Expression in Colon Cancer. *Int. J. Mol. Med.* 38 (2), 499–506. doi:10.3892/ijmm.2016.2624

Yang, H-Y., Barbi, J., Wu, C-Y., Zheng, Y., Vignali, P. D. A., Wu, X., et al. (2016). MicroRNA-17 Modulates Regulatory T Cell Function by Targeting Co-

regulators of the Foxp3 Transcription Factor. *Immunity* 45 (1), 83–93. doi:10.1016/j.immuni.2016.06.022

Yu, W., Wang, J., Li, C., Xuan, M., Han, S., Zhang, Y., et al. (2022). miR-17-5p Promotes the Invasion and Migration of Colorectal Cancer by Regulating HSPB2. *J. Cancer* 13 (3), 918–931. doi:10.7150/jca.65614

Zhu, M., Yan, C., Ren, C., Huang, X., Zhu, X., Gu, H., et al. (2017). Exome Array Analysis Identifies Variants in SPOCD1 and BTN3A2 that Affect Risk for Gastric Cancer. *Gastroenterology* 152 (8), 2011–2021. doi:10.1053/j.gastro.2017.02.017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lin, Zhou and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 19 July 2022
DOI 10.3389/fgene.2022.888051

Check for updates

OPEN ACCESS

EDITED BY
Changwei Li,
Shanghai Jiao Tong University, China

REVIEWED BY
Yuan Zhou,
Peking University, China
Rafael Medrano,
Oncology Hospital National Medical Center Siglo XXI IMSS, Mexico

*CORRESPONDENCE
Jigang Yang,
yangjigang@ccmu.edu.cn
Jianhua Gong,
ann_gong@imb.pumc.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 02 March 2022
ACCEPTED 27 June 2022
PUBLISHED 19 July 2022

CITATION
Kan Y, Lu X, Feng L, Yang X, Ma H, Gong J and Yang J (2022), RPP30 is a novel diagnostic and prognostic biomarker for gastric cancer.
Front. Genet. 13:888051.
doi: 10.3389/fgene.2022.888051

COPYRIGHT
© 2022 Kan, Lu, Feng, Yang, Ma, Gong and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

RPP30 is a novel diagnostic and prognostic biomarker for gastric cancer

Ying Kan[1†], Xia Lu[1†], Lijuan Feng[1], Xu Yang[1], Huan Ma[1], Jianhua Gong[2*] and Jigang Yang[1*]

[1]Department of Nuclear Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing, China, [2]Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Objective: This study aimed to identify the hub gene in gastric cancer (GC) tumorigenesis. A biomarker prediction model was constructed and analyzed, and protein expression in histopathological samples was verified in a validation cohort.

Methods: Differentially expressed genes (DEGs) were identified from GC projects in The Cancer Genome Atlas (TCGA) database. Functional enrichment analysis of DEGs was performed between the high- and low-Ribonuclease P protein subunit p30 (RPP30) expression groups. ROC analysis was performed to assess RPP30 expression to discriminate GC from normal tissues. Functional enrichment pathways and immune infiltration of DEGs were analyzed using GSEA and ssGSEA. Survival analysis and nomogram construction were performed to predict patient survival. Immunohistochemical staining of GC tissues was performed to validate RPP30 expression in GC and paracancerous samples.

Results: Gene expression data and clinical information of 380 cases (375 GC samples and 32 para-cancerous tissues) were collected from TCGA database. The AUC for RPP30 expression was found to be 0.785. The G alpha S signaling pathway was the most significantly enriched signaling pathway. Primary therapy outcome ($p < 0.001$, HR = 0.243, 95% CI = 0.156–0.379), age ($p = 0.012$, HR = 1.748, 95% CI = 1.133–2.698), and RPP30 expression ($p < 0.001$, HR = 2.069, 95% CI = 1.346–3.181) were identified as independent prognostic factors. As a quantitative approach, a nomogram constructed based on RPP30 expression, age, and primary therapy outcome performed well in predicting patient survival. Nineteen of the 25 tissue samples from the validation cohort showed positive RPP30 expression in GC tissues, whereas 16 cases showed negative RPP30 staining in normal tissues. The difference between the two was statistically significant.

Conclusion: High RPP30 expression was significantly correlated with disease progression and poor survival in GC, promoting tumorigenesis and angiogenesis *via* tRNA dysregulation. This study provides new and promising insights into the molecular pathogenesis of tRNA in GC.

KEYWORDS

gastric cancer, RPP30, early diagnosis, prognosis, bioinformatics analysis

Introduction

Gastric cancer (GC) remains a major challenge in the field of oncology. It is the fifth most frequently occurring cancer and the third leading cause of cancer-related mortality (Global Burden of Disease Cancer et al., 2019). In China, epidemiological studies have shown that GC is the third most common cancer after lung cancer and liver cancer, and is characterized by high mortality and morbidity (Zhou et al., 2019). Although GC treatment protocols have improved markedly, there is still no gold standard therapy, and the 5-year overall survival (OS) rate continues to be less than 30% (Rugge et al., 2015). There are many reasons for the low survival rates, such as late-stage diagnosis, high intra-tumor heterogeneity, and chemotherapy resistance (Stahl et al., 2015). GC is characterized by high molecular and phenotypic heterogeneity. Understanding the underlying mechanism of GC carcinogenesis and progression is pivotal for the early diagnosis and improvement of survival rates.

The use of high-throughput sequencing technology has recently provided new insights into the molecular pathogenesis and prognosis prediction of GC. Hundreds of hub genes have been shown to promote tumorigenesis via different tumor-related pathways, biological processes, and molecular functions. Genetic alterations, such as aberrant DNA methylation and overexpression or downregulation of microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circulating RNAs may play a role in GC initiation and progression. However, all these plasma biomarkers are non-coding RNAs, for which improved extraction techniques, probe enrichment, and validation studies are required for clinical implementation (Smyth et al., 2020).

Ribonuclease P protein subunit p30 (RPP30), a subunit of ribonuclease P (RNase P) with a molecular weight of 30 kDa, cleaves the 5′ leader sequence from transfer RNA (tRNA) precursor molecules (Jarrous et al., 1998). RNase P, a ribonucleoprotein complex with 10 protein components and one catalytic RNA, plays important roles in genome preservation, including gene transcription, replication RNA, DNA repair, and chromatin remodeling (Lemieux et al., 2016; Jarrous, 2017; Wu et al., 2018). The catalytic RNA subunit is responsible for nuclear RNA and tRNA processing (Mondragón, 2013). RPP30 has also been demonstrated to affect the process of RNA modification in protein expression and to promote tumorigenesis in glioblastoma (Li et al., 2020a). Although RPP30 plays a pivotal role in some types of tumors, the expression of RPP30 and its biological effects in GC remain unknown.

Previous studies have confirmed that more cancer-related molecular factors may contribute to the development of GC (Sexton et al., 2020). Identification of novel biomarkers for the early diagnosis and prognosis of GC is crucial for improving treatment efficacy. In this study, we analyzed The Cancer Genome Atlas (TCGA) database to identify possible biomarkers of GC *via* bioinformatics; we further constructed a biomarker prediction model and verified the expression of proteins histopathologically in a validation cohort.

Materials and methods

RNA-sequencing gene expression analysis

The gene expression data (workflow type: HTSeq-FPKM and HTSeq-counts) and clinical information of 380 cases (375 GC samples and 32 para-cancerous tissues) were collected from GC projects in TCGA database (https://portal.gdc.cancer.gov/). Cases with an OS of less than 30 days (n = 29) were excluded. Finally, the 407 RNA-sequencing (RNA-seq) gene expression level 3 HTSeq-FPKM data of patients were transformed into transcripts per million for further analyses. Patient characteristics, including age, sex, race, TNM stage, pathologic stage, histological type, and TP53 and PIK3CA status, were recorded. Some of these data were unavailable and were treated as missing values. The study design fully satisfied the publication guidelines of TCGA.

The high and low RPP30 expression profiles (HTSeq-counts) were compared with each other within DESeq2 package (Love et al., 2014). Genes with adjusted p-value < 0.05 and |log2 (fold-change)| > 1.5 were considered as differentially expressed.

Immunohistochemical assay of the validation cohort of GC patients

Twenty-five GC patients (12 men, 13 women; mean age, 61.8 years) who underwent total or subtotal gastrectomy between September 2018 and December 2019 were recruited at the Beijing Friendship Hospital of Capital Medical University. The ethic of validation cohort was approved by Research Ethics Committee of Beijing Friendship Hospital (NO. 2018-92-045-01). All subjects provided written informed consent in accordance with the Declaration of Helsinki.

Both GC and paracancerous tissues were prepared using formalin-fixed and paraffin-embedded sections. Sodium citrate buffer (pH 6.0) and 3% hydrogen peroxide were used for antigen retrieval and endogenous peroxidase clearance, respectively. The sections were then blocked with 5% bovine serum albumin and incubated overnight (16–18 h) with primary antibody (Anti-RPP30, Solarbio Life Science, Beijing, China). The reaction products were stained with 3,3-diaminobenzidine and counterstained with hematoxylin. Tissue slides were scanned using a Pannoramic MIDI automated slide scanner (3D Histech, Munich, Germany). Immunohistochemical analysis of RPP30 expression was performed by examining three to five random fields under ×400 magnification in the selected area of

each slide. The expression levels of RPP30 protein were classified as positive (RPP30 staining positive cells >10%) or negative (RPP30 staining positive cells <10%). Two investigators who were unaware of the final diagnosis calculated the number of cells.

Gene set enrichment analysis

GSEA was performed using the ClusterProfiler package (Yu et al., 2012) to elucidate the difference between the high and low RPP30 expression groups ($RPP30^{high}$ and $RPP30^{low}$, respectively). The RPP30 expression level was regarded as a phenotype, and all gene set permutations were performed 1,000 times for each analysis. Any function or pathway with adjusted p-value < 0.05, FDR q-value < 0.25, and absolute value of the normalized enrichment score (NES) > 1 were statistically significant.

Correlation between RPP30 expression and immune infiltration

Single-sample GSEA (ssGSEA) from the GSVA package (Barbie et al., 2009) was applied to quantify the relative tumor infiltration levels of 24 immune cells by obtaining the expression levels of genes in published gene lists (Bindea et al., 2013).

The correlation between RPP30 expression and immune cell infiltration was analyzed using Spearman's correlation, and the infiltration of immune cells between the $RPP30^{high}$ and $RPP30^{low}$ groups was analyzed using the Wilcoxon rank-sum test.

Prognostic model generation and prediction

Survival analysis between the $RPP30^{high}$ and $RPP30^{low}$ groups was performed in GC patients. The primary endpoint was OS, and the secondary endpoint was progression-free survival (PFS). The follow-up duration was estimated using the Kaplan-Meier method, and differences between survival curves were examined using the log-rank test. Univariate Cox proportional hazards regression was used to estimate the hazard ratio (HR) for OS and PFS. The significant variables, as determined via univariate analysis, were included in the multivariate analysis. Multivariate Cox regression analysis was used to evaluate the optimal model.

A nomogram was constructed to predict the individualized survival probability of GC patients. The risk score was calculated as the sum of scores for each parameter. A calibration plot was used to evaluate the prediction accuracy of the nomogram based on the prognostic model. All statistical tests were two-tailed, and p-values < 0.05 were considered statistically significant.

Statistical analysis

All statistical analyses were performed using R software (http://www.r-project.org/, version 3.6.2). The cut-off value of RPP30 expression was determined by the median value and used for grouping into the $RPP30^{high}$ and $RPP30^{low}$ groups. The Wilcoxon rank-sum test was used to analyze the expression of RPP30 in normal and tumor samples. Kruskal–Wallis and Spearman's correlation tests were used to evaluate the relationship between RPP30 expression and GC clinicopathological features.

Results

Demographic characteristics and RPP30 expression

The clinical characteristics of GC patients in TCGA database, including sex, age, race, TNM stage, histological type and grade, clinical presentation, and PIK3CA status, were obtained (Supplementary Table S1). A total of 134 female and 241 male patients were included in the study.

Considering the TNM pathologic stage, 53 cases (14.1%) were stage I, 111 (29.6%) were stage II, 150 (40.0%) were stage III, and 38 (10.1%) were stage IV. Regarding histological grade, 10 cases (2.6%) were well differentiated (G1 group), 137 cases (36.5%) were moderately differentiated (G2 group), and 219 cases (58.4%) were poorly differentiated (G3 group). According to PIK3CA status, 59 cases (15.7%) were mutants, and 313 cases (83.4%) were wild-type.

Potential role of RPP30 in regulating the progression of GC

To elucidate whether RPP30 plays a role in prompting GC, TCGA RNA-seq data analysis was performed to compare DEGs between the $RPP30^{high}$ and $RPP30^{low}$ groups. A total of 151 upregulated and 82 downregulated genes were identified in the $RPP30^{high}$ group (using the $RPP30^{low}$ group as reference). DEG expression is shown as a volcano map and heat map in Figure 1.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes functional enrichment analyses were performed to further understand the functional implications of RPP30 in GC. Biological processes were primarily enriched in skin development (GO: 0043588), epidermal development (GO: 0008544), epidermal cell differentiation (GO:0009913), and keratinocyte differentiation (GO: 0030216). The GSEA results showed that several pathways and biological processes were differentially enriched in RPP30 in relation to GC, including the G alpha S signaling pathway, neuronal system, and olfactory

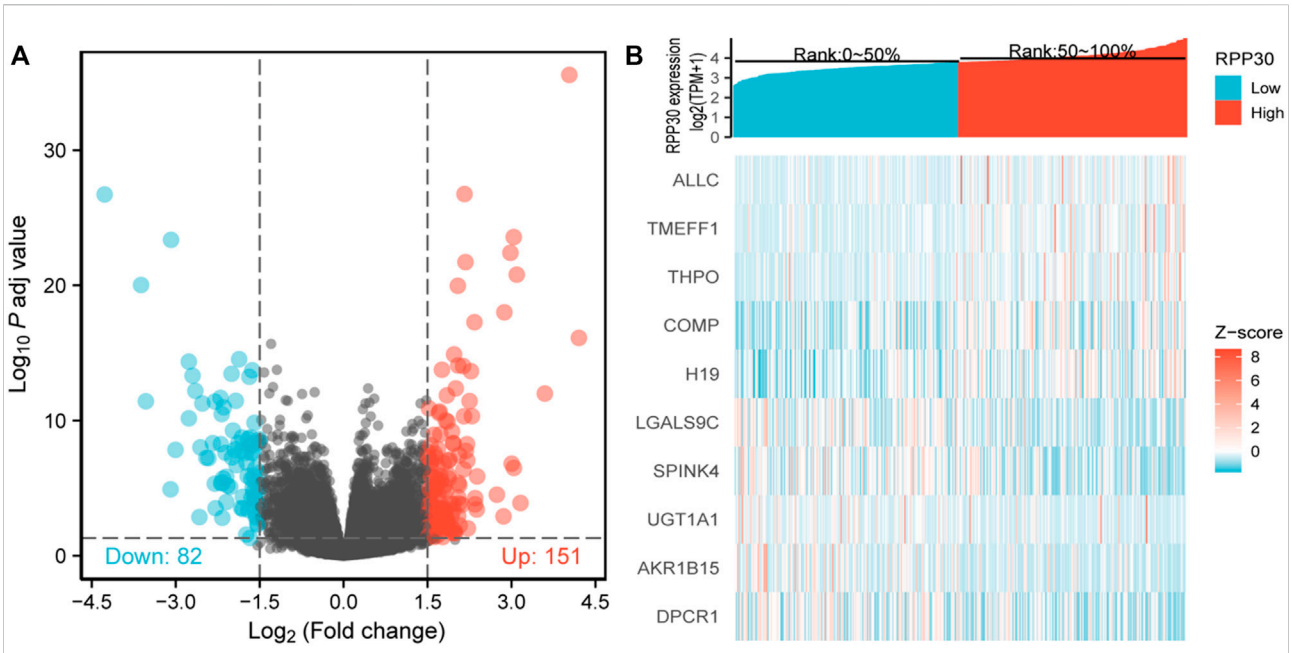


FIGURE 1
Identification of DEGs between RPP30^{high} and RPP30^{low} groups. **(A)** Volcano plot of DEG profiles between RPP30^{high} and RPP30^{low} groups. A total of 233 DEGs were obtained, of which 151 were upregulated and 82 were downregulated. **(B)** Heatmap of GO analysis showing the co-expression of differential gene profiles in TCGA between RPP30^{high} and RPP30^{low} groups. Red indicates upregulated genes; blue indicates downregulated genes; each row indicates each gene expression in different samples, whereas each column indicates the expression of all genes in each sample.

transduction. Based on the NES score, the most significantly enriched signaling pathway was the G alpha S signaling pathway. The results suggested a link between the aberrant *RPP30* gene and immune response (Figure 2).

Correlation between RPP30 expression and immune infiltration

The correlation between the expression level of RPP30 and immune cell infiltration is depicted as a lollipop chart (Figure 3A). RPP30 expression was positively correlated with the abundance of immunocytes (Th2 cells, activated dendritic cells, Th1 cells, and helper T cells) and negatively correlated with that of Th17 cells. In comparison to the low expression of RPP30, the high expression of RPP30 was associated with increased Th2 cell infiltration in the tumor microenvironment ($p < 0.001$; Figure 3B).

Association with RPP30 expression and clinicopathological variables

There was a significant statistical difference in RPP30 expression between tumor and paracancerous normal tissues ($p < 0.001$), and a higher expression of RPP30 was observed in tumor tissues (Figure 4A). Furthermore, we performed ROC analysis to calculate the diagnostic performance of RPP30 expression. The ROC curve showed that the AUC of RPP30 expression for distinguishing tumors from normal tissues was 0.785, indicating a high diagnostic accuracy for GC (Figure 4B). RPP30 expression was significantly correlated with the T stage of GC (Figure 5). These results provided a theoretical basis for the use of RPP30 as a biomarker in the early diagnosis of GC.

A total of 375 samples with RPP30 expression data were analyzed from TCGA database. Logistic regression analysis revealed that the expression of RPP30 was associated with poor clinicopathological characteristics (Supplementary Table S2), including histological grade ($p = 0.018$, OR = 1.66, 95% CI = 1.09–2.54) and PIK3CA status ($p = 0.031$, OR = 1.87, 95% CI = 1.07–3.37).

Survival analysis

Kaplan-Meier survival analysis was performed to evaluate the association between RPP30 expression and GC prognosis. High RPP30 expression was more strongly associated with a

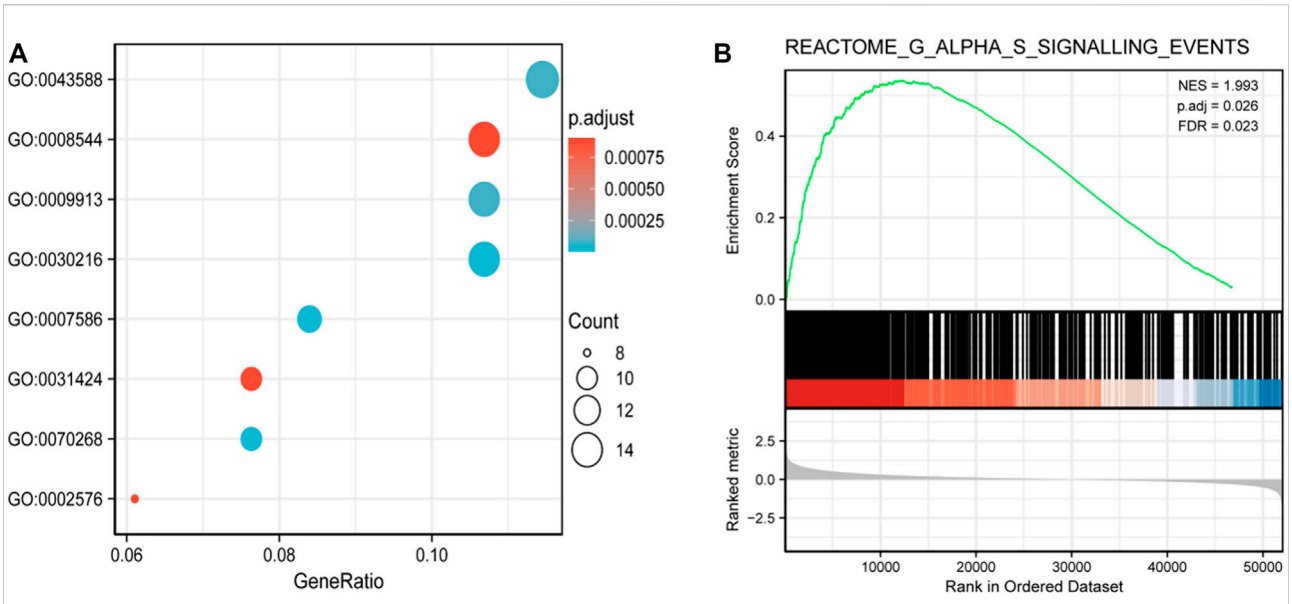


FIGURE 2
Functional enrichment analysis of DEGs between RPP30^{high} and RPP30^{low} groups of GC in TCGA. **(A)** Enriched GO terms in the biological process category. The *x*-axis represents the proportion of DEGs, and the *y*-axis represents different categories. In this category, skin development (GO: 0043588), epidermis development (GO: 0008544), epidermal cell differentiation (GO:0009913), and keratinocyte differentiation (GO: 0030216) were primarily enriched. **(B)** GSEA results showed that the G alpha S signaling pathway was the most enriched. (NES, normalized enriched score; *P*. adj, adjusted *p*-value; FDR, false discovery rate; gene sets with *P*. adj <0.05, FDR q-value < 0.25, and |NES| > 1 are considered as significant).

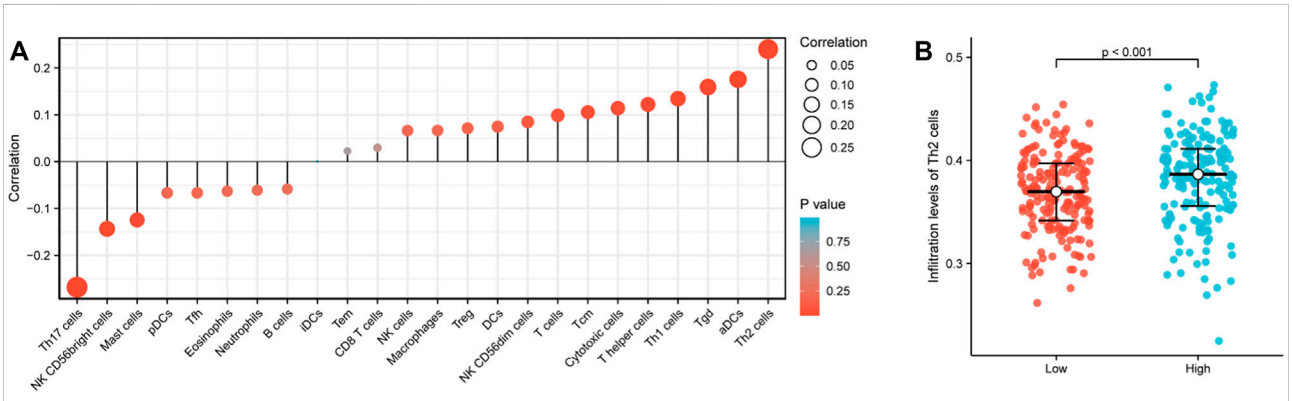


FIGURE 3
RPP30 expression level was associated with immune infiltration in the GC microenvironment. **(A)** Correlation between the marker gene of 24 immune cells and RPP30 expression level is shown in the lollipop chart. The size of the dots shows the absolute value of Spearman's correlation coefficient (r). Larger dots indicate higher correlation coefficients. RPP30 expression was positively correlated with the abundances of immunocytes (Th2 cells, activated dendritic cells, Th1 cells, and helper T cells) and negatively correlated with the presence of Th17 cells. **(B)** Difference between RPP30^{high} and RPP30^{low} groups in terms of Th2 cell infiltration.

worse prognosis than low RPP30 expression. Univariate Cox regression analysis revealed that high RPP30 expression was correlated with shorter OS (HR = 1.53, 95% CI = 1.10–2.14, $p = 0.012$). Univariate analysis also revealed that T stage ($p = 0.011$), N stage ($p = 0.002$), M stage ($p = 0.004$), pathological stage ($p < 0.001$), primary therapy outcome ($p < 0.001$), residual tumor ($p < 0.001$), age ($p = 0.005$), and RPP30 expression ($p = 0.012$) were significantly correlated with OS. These parameters were then included in the multivariate Cox regression model, which revealed that the

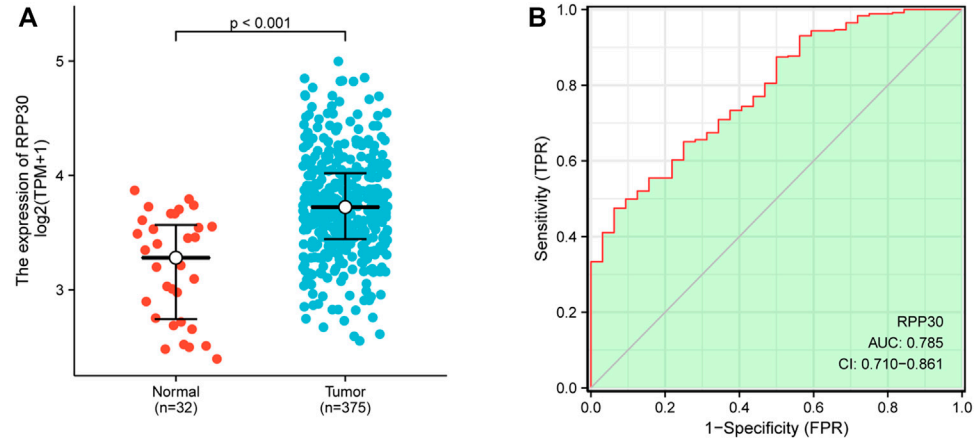


FIGURE 4
RPP30 expression between normal and GC tumor samples in TCGA database. **(A)** RPP30 was significantly upregulated in GC ($p < 0.001$). **(B)** ROC analysis of RPP30 expression showed high diagnostic efficiency in discriminating between tumor and normal tissues. The AUC of RPP30 in the diagnosis of GC was 0.785.

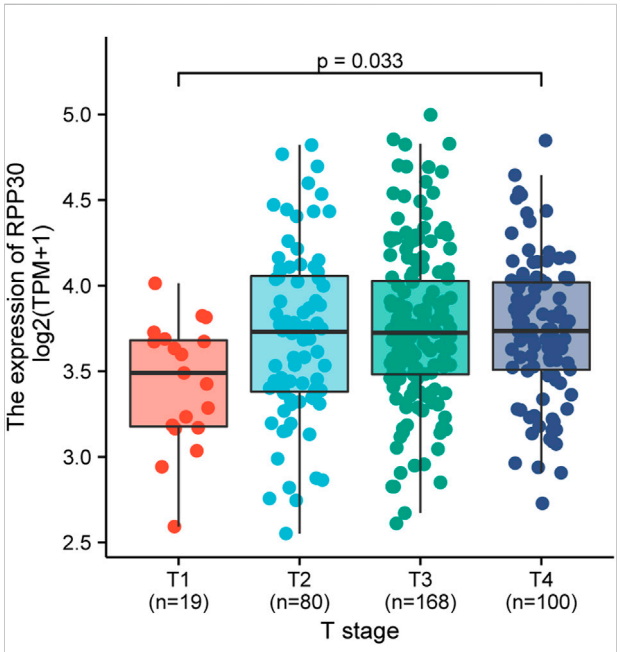


FIGURE 5
Relationship between RPP30 expression and T-stage of TCGA in GC. The results showed that RPP30 expression was significantly correlated with T-stage ($p = 0.033$). Bonferroni correction between each group also showed significance.

primary therapy outcome ($p < 0.001$, HR = 0.243, 95% CI = 0.156–0.379), age ($p = 0.012$, HR = 1.748, 95% CI = 1.133–2.698), and RPP30 expression ($p < 0.001$, HR = 2.069, 95% CI = 1.346–3.181) could serve as independent prognostic factors for GC (Supplementary Table S3).

Construction and validation of the prognostic model based on RPP30 expression and clinicopathological factors

To establish a quantitative approach to GC prognosis, we constructed a prognostic nomogram model based on the results of multivariate Cox regression analysis involving RPP30 and independent clinicopathological risk factors. The concordance index (C-index) for the nomogram was 0.704 (95% CI = 0.680–0.728). The total score was the sum of the scores for each variable. The probability of GC patient survival at 1, 3, and 5 years was determined by drawing a vertical line along the total points (Figure 6). The calibration plot demonstrated that the bias-corrected line was close to the ideal line, which indicated good agreement between the observation and prediction. These findings suggested that the nomogram is a better model for predicting OS in GC patients than individual prognostic factors.

RPP30 expression affects the prognosis of GC at different clinicopathological statuses

To further understand the mechanism of RPP30 expression, we investigated the relationship between RPP30 expression and clinicopathological status. Univariate Cox analysis revealed that RPP30 expression was associated with poor OS in the T and N stages, especially in the T1, T2, and N0 stages. This indicated that in the T1, T2, and N0 stages, high RPP30 expression was

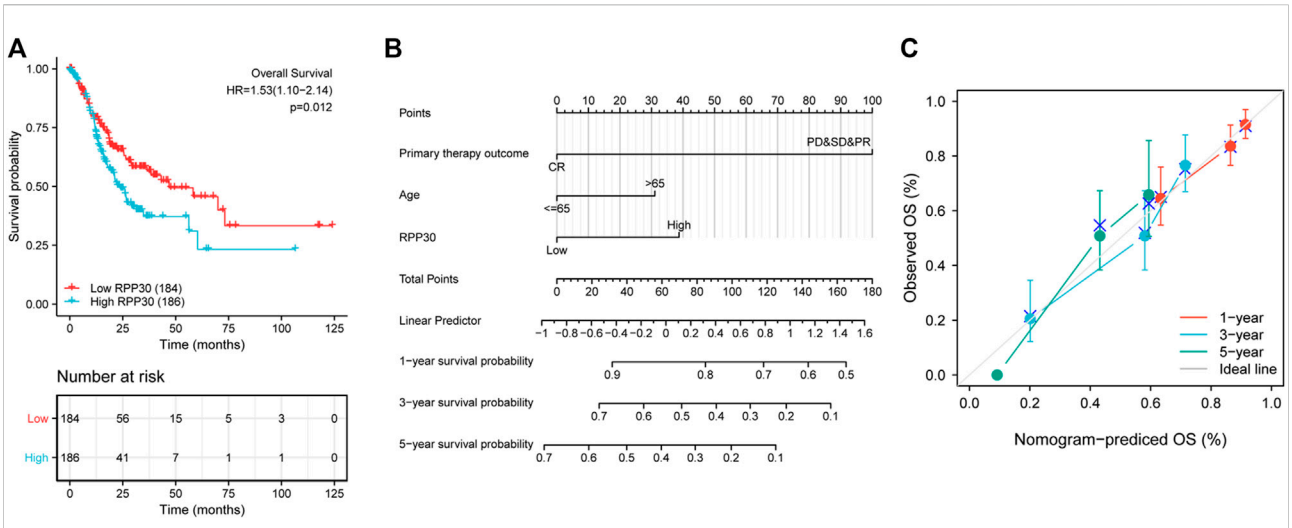


FIGURE 6
Validation of RPP30-based nomogram for GC patients. **(A)** Effect of RPP30 expression on the OS of GC patients in TCGA cohort. The results showed that higher RPP30 expression was associated with poor OS (HR = 1.53 [1.10–2.14], p = 0.012). **(B)** Nomogram for the prediction of 1-, 3-, and 5-years OS of GC patients. The C-index was 0.704 (95% CI = 0.680–0.728). **(C)** The calibration plot of the nomogram indicated a good agreement between the prediction and the ideal line.

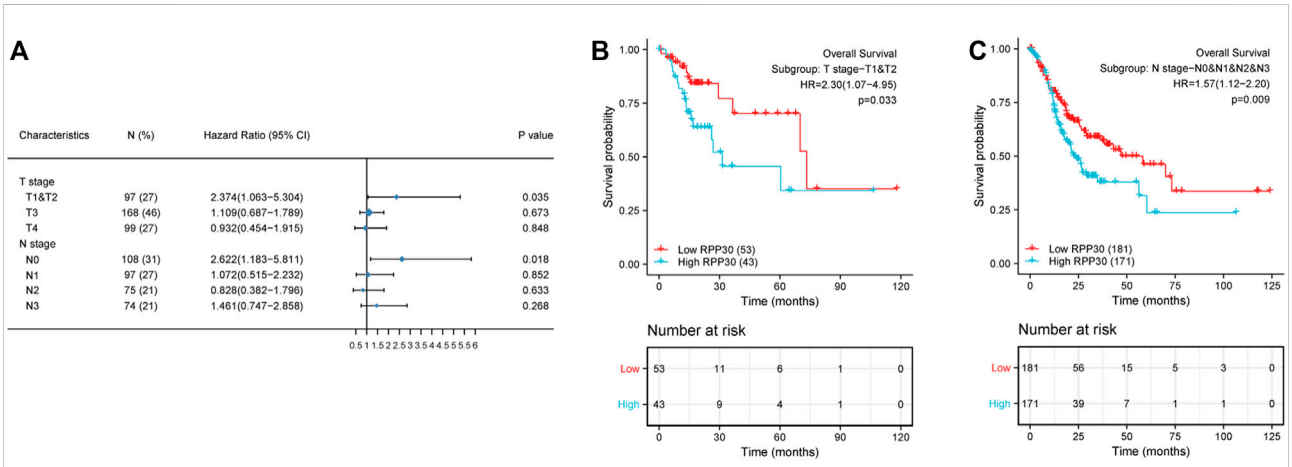


FIGURE 7
Subgroup survival analysis of clinical characteristics of GC. **(A)** High RPP30 expression was correlated with worse OS in the T1 and T2 stages of GC, but not in the T3 and T4 stages. In the N stages, high RPP30 expression was associated with worse OS in the N0, but not in the N1 - N3 stages of GC. **(B)** The K-M plot of OS showed that high RPP30 expression had higher HR value (HR = 2.3, p = 0.033) in subgroups of T1 and T2 stages. **(C)** The hazard ratio in the high-RPP30 expression group was 0.57 times higher than that in the low expression group (p = 0.009).

associated with worse OS (Figure 7). These results suggested that RPP30 expression levels affect the prognosis of GC with different clinicopathological statuses, and that the early detection of RPP30 expression is pivotal for the prognosis of GC patients.

Validation of RPP30 expression in GC tissues

RPP30 expression was investigated using immunohistochemistry in a validation cohort of 25 GC

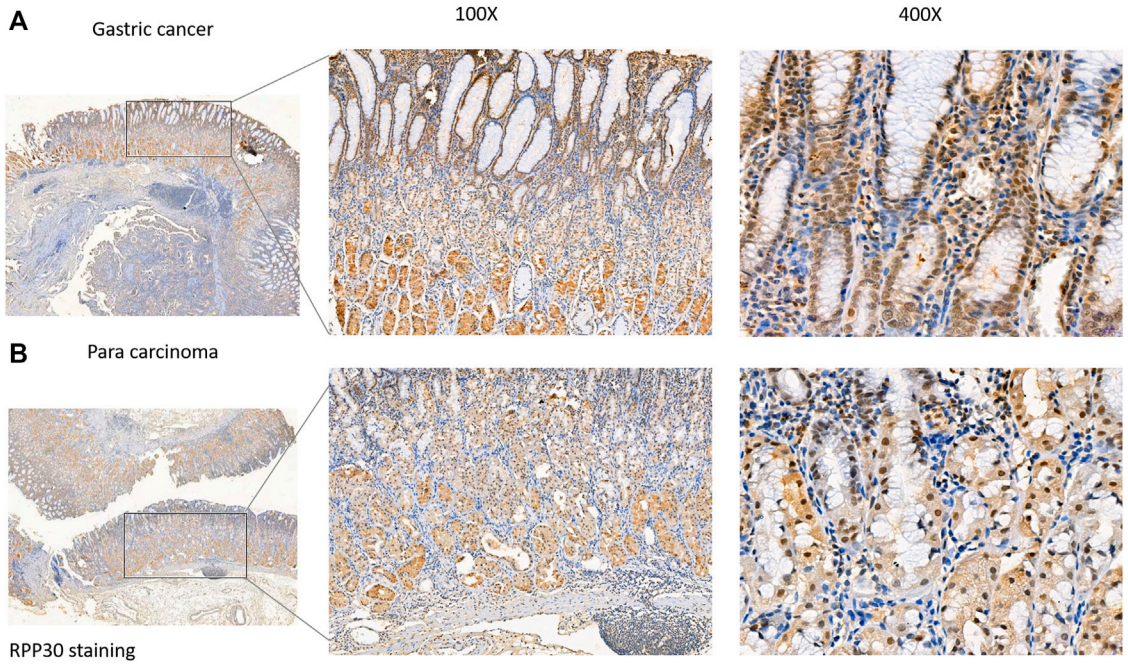


FIGURE 8
RPP30 protein expression in GC and para-cancerous tissues assessed via immunohistochemical staining. **(A)** RPP30 expression was found in GC specimens, especially in the nucleus of the glandular epithelium. **(B)** RPP30 protein expression was negative in para-cancerous tissues.

tissue samples. Of the 25 specimens, 19 showed positive RPP30 expression in GC tissue and 16 showed negative RPP30 expression in normal tissues (Supplementary Table S4). The difference between the two was significant. The RPP30 protein was primarily distributed in the nucleus and cytoplasm (Figure 8). The expression of RPP30 protein was significantly higher in GC tissues than in para-cancerous tissues.

Discussion

GC is the fifth most common malignancy worldwide. It has become the third most common cancer-related cause of death because it is usually diagnosed when the cancer reaches an advanced stage. Therefore, it is essential to explore the molecular mechanisms of GC for its early diagnosis and prognosis. Various GC-related miRNAs, lncRNAs, cytokines, and proteins are disordered and associated with the early diagnosis or prognosis of GC (Wu et al., 2019). However, the sensitivity of these biomarkers for discriminating between early-stage GC and healthy subjects is low (Necula et al., 2019). The expression and function of RPP30 in gliomas has been previously reported (Li et al., 2020a). RPP30-related proteins were primarily enriched in the cancer-related pathways. To our knowledge, the expression of RPP30 and its potential prognostic impact on GC remain unexplored. In this study, bioinformatics analysis of high-throughput RNA-seq data from TCGA demonstrated that high RPP30 expression in GC was associated with advanced clinical pathological characteristics, survival time, and poor prognosis.

RPP30, a subunit of RNase P, a ribozyme involved in pre-tRNA processing, forms mature tRNAs that bind to amino acids and further regulate protein expression. Our results showed that RPP30 was highly expressed in GC tissues compared to normal tissues and showed high diagnostic accuracy for GC. RPP30 was primarily enriched in epidermal development, cell differentiation, and keratinocyte differentiation, and was differentially enriched in the RPP30-high expression phenotype. Previous studies have shown that the differentiation of keratinocytes plays an important role in the differentiation of normal gastric epithelial cells and affects the function of parietal cells (Matsunobu et al., 2006). RPP30 was significantly enriched in the G alpha S signaling pathway and increased cAMP levels (Loffler et al., 2008; Wehbe et al., 2020), and it was related to the histopathology of GC. Our immunohistochemistry results showed that RPP30 expression was significantly higher in GC tissues than in paracancerous tissues and correlated with clinicopathological features. These results provide a theoretical basis for the early diagnosis of GC.

RPP30 affects tRNA processing, transcription replication, DNA repair, and replication fork stalling (Molla-Herman et al., 2015; Wu et al., 2018); regulates protein expression; enriches

cancer-related pathways, leading to tumorigenesis; and eventually promotes the proliferation, metastasis, and invasiveness of cancer cells (Huang et al., 2018). Multiple large-scale genomic studies have shown that altering the co-transcriptional and post-transcriptional regulation of gene expression during RNA processing, including the splicing, transport, editing, and decay of messenger RNA, can initiate tumorigenesis and cancer maintenance (Obeng et al., 2019). Dysregulation of tRNA in tumor pathogenesis has been confirmed in breast cancer, lung cancer, and melanoma; however, relevant studies on GC are still lacking (Huang et al., 2018). In addition, high RPP30 expression was shown to be correlated with the poor prognosis of GC at the T1 to T2 and N0 stages, with the highest HR for poor OS when RPP30 was highly expressed in GC tissues. These findings strongly suggest that RPP30 expression is a powerful indicator of GC prognosis in these subsets. For more accurate prognosis prediction, nomograms have been developed that show better performance than conventional staging systems (Li et al., 2020b; Wang et al., 2020). Our nomogram included three parameters available from clinical records and tissue specimens. As previously reported, age is an independent prognostic factor, and older age is associated with poorer prognosis (Schlesinger-Raab et al., 2016). In the revised RECIST guidelines, response to chemotherapy has been shown to be associated with OS (Eisenhauer et al., 2009). Primary therapy outcome with complete response was associated with a better prognosis than partial response. These results are consistent with our findings. Highly fitted calibration plots demonstrated that the nomogram performed well in predicting the 3- or 5-years survival of GC patients.

Although the present study improved our understanding of the relationship between RPP30 expression and GC, some limitations remain. First, due to the limitations of the study design, we will further examine the additional RPP30-relevant signaling pathways and investigate the mechanism of RPP30 in GC tumorigenesis by experimental studies. Second, although public databases are multicenter, retrospective studies have limitations. In the future, prospective studies should be performed to avoid selection bias.

Our study demonstrated that high RPP30 expression was significantly correlated with tumor progression and poor survival in GC, which might promote tumorigenesis and angiogenesis via tRNA dysregulation. This study provides new insights into the molecular pathogenesis of tRNA in GC.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YK and XL contributed equally to this work. YK, JG, and JY contributed to conception and design of the study. XL and HM contributed the data collection. LF and XY performed the data analysis and statistical analysis. YK and XL wrote the manuscript. All authors contributed to manuscript revision and approved the submitted revision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.888051/full#supplementary-material

References

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., FoRd, R., et al. (2009). New response evaluation criteria in solid tumours: Revised RECIST

guideline (version 1.1). *Eur. J. Cancer* 45 (2), 228–247. doi:10.1016/j.ejca.2008.10.026

Global Burden of Disease Cancer, C., Fitzmaurice, C., Abate, D., et al. (2019). Global, regional, and national cancer incidence, mortality, years of Life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study[J]. *JAMA Oncol.* 5 (12), 1749–1768.

Huang, S. Q., Sun, B., Xiong, Z. P., Shu, Y., Zhou, H. H., Zhang, W., et al. (2018). The dysregulation of tRNAs and tRNA derivatives in cancer. *J. Exp. Clin. Cancer Res.* 37 (1), 101. doi:10.1186/s13046-018-0745-z

Jarrous, N., Eder, P. S., Guerrier-Takada, C., Hoog, C., and Altman, S. (1998). Autoantigenic properties of some protein subunits of catalytically active complexes of human ribonuclease P. *J. Rna.* 4 (4), 407–417.

Jarrous, N. (2017). Roles of RNase P and its subunits. *Trends Genet.* 33 (9), 594–603. doi:10.1016/j.tig.2017.06.006

Lemieux, B., Laterreur, N., Perederina, A., Noel, J. F., Dubois, M. L., Krasilnikov, A. S., et al. (2016). Active yeast telomerase shares subunits with ribonucleoproteins RNase P and RNase MRP. *Cell* 165 (5), 1171–1181. doi:10.1016/j.cell.2016.04.018

Li, G., Zhai, Y., Liu, H., Wang, Z., Huang, R., Jiang, H., et al. (2020). RPP30, a transcriptional regulator, is a potential pathogenic factor in glioblastoma. *Aging (Albany NY)* 12 (16), 16155–16171. doi:10.18632/aging.103596

Li, J., Dong, D., Fang, M., Wang, R., Tian, J., Li, H., et al. (2020). Dual-energy CT-based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. *Eur. Radiol.* 30 (4), 2324–2333. doi:10.1007/s00330-019-06621-x

Loffler, I., Grun, M., Bohmer, F. D., and Rubio, I. (2008). Role of cAMP in the promotion of colorectal cancer cell growth by prostaglandin E2. *BMC Cancer* 8, 380. doi:10.1186/1471-2407-8-380

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Matsunobu, T., Ishiwata, T., Yoshino, M., Watanabe, M., Kudo, M., Matsumoto, K., et al. (2006). Expression of keratinocyte growth factor receptor correlates with expansive growth and early stage of gastric cancer. *Int. J. Oncol.* 28 (2), 307–314.

Molla-Herman, A., Valles, A. M., Ganem-Elbaz, C., Antoniewski, C., and Huynh, J. R. (2015). tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription. *EMBO J.* 34 (24), 3009–3027. doi:10.15252/embj.201591006

Mondragón, A. (2013). Structural studies of RNase P. *Annu. Rev. Biophys.* 42, 537–557. doi:10.1146/annurev-biophys-083012-130406

Necula, L., Matei, L., Dragu, D., Neagu, A. I., Mambet, C., Nedeianu, S., et al. (2019). Recent advances in gastric cancer early diagnosis. *World J. Gastroenterol.* 25 (17), 2029–2044. doi:10.3748/wjg.v25.i17.2029

Obeng, E. A., Stewart, C., and Abdel-Wahab, O. (2019). Altered RNA processing in cancer pathogenesis and therapy. *Cancer Discov.* 9 (11), 1493–1510. doi:10.1158/2159-8290.CD-19-0399

Rugge, M., Fassan, M., and Graham, D. Y. (2015). "Epidemiology of gastric cancer," in *Gastric cancer: Principles and practice*. Editor V. E. Strong (Cham: Springer International Publishing), 23–34. doi:10.1007/978-3-319-15826-6_2

Schlesinger-Raab, A., Mihaljevic, A. L., Egert, S., Emeny, R., Jauch, K. W., Kleeff, J., et al. (2016). Outcome of gastric cancer in the elderly: A population-based evaluation of the Munich cancer registry. *Gastric Cancer* 19 (3), 713–722. doi:10.1007/s10120-015-0527-7

Sexton, R. E., Al Hallak, M. N., Diab, M., and Azmi, A. S. (2020). Gastric cancer: A comprehensive review of current and future treatment strategies. *Cancer Metastasis Rev.* 39 (4), 1179–1203. doi:10.1007/s10555-020-09925-3

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., and Lordick, F. (2020). Gastric cancer. *Lancet* 396 (10251), 635–648. doi:10.1016/S0140-6736(20)31288-5

Stahl, P., Seeschaaf, C., Lebok, P., Kutup, A., Bockhorn, M., Izbicki, J. R., et al. (2015). Heterogeneity of amplification of HER2, EGFR, CCND1 and MYC in gastric cancer. *BMC Gastroenterol.* 15, 7. doi:10.1186/s12876-015-0231-4

Wang, C. Y., Yang, J., Zi, H., Zheng, Z. L., Li, B. H., Wang, Y., et al. (2020). Nomogram for predicting the survival of gastric adenocarcinoma patients who receive surgery and chemotherapy. *BMC Cancer* 20 (1), 10. doi:10.1186/s12885-019-6495-2

Wehbe, N., Slika, H., Mesmar, J., Nasser, S. A., Pintus, G., Baydoun, S., et al. (2020). The role of epac in cancer progression. *Int. J. Mol. Sci.* 21 (18), E6489. doi:10.3390/ijms21186489

Wu, D., Zhang, P., Ma, J., Xu, J., Yang, L., Xu, W., et al. (2019). Serum biomarker panels for the diagnosis of gastric cancer. *Cancer Med.* 8 (4), 1576–1583. doi:10.1002/cam4.2055

Wu, J., Niu, S., Tan, M., Huang, C., Li, M., Song, Y., et al. (2018). Cryo-EM structure of the human ribonuclease P holoenzyme. *Cell* 175 (5), 1393–1404. doi:10.1016/j.cell.2018.10.003

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhou, M., Wang, H., Zeng, X., Yin, P., Zhu, J., Chen, W., et al. (2019). Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: A systematic analysis for the global burden of disease study 2017. *Lancet* 394 (10204), 1145–1158. doi:10.1016/S0140-6736(19)30427-1

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fgene.2022.881948

Common Core Genes Play Vital Roles in Gastric Cancer With Different Stages

Zhiyuan Yu[1,2†], Chen Liang[3†], Huaiyu Tu[2], Shuzhong Qiu[2], Xiaoyu Dong[2], Yonghui Zhang[2], Chao Ma[2] and Peiyu Li[1,2]*

[1]School of Medicine, Nankai University, Tianjin, China, [2]Department of General Surgery, The First Medical Center, Chinese PLA General Hospital, Beijing, China, [3]First Department of Liver Disease / Beijing Municipal Key Laboratory of Liver Failure and Artificial Liver Treatment Research, Beijing You'an Hospital, Capital Medical University, Beijing, China

Background: Owing to complex molecular mechanisms in gastric cancer (GC) oncogenesis and progression, existing biomarkers and therapeutic targets could not significantly improve diagnosis and prognosis. This study aims to identify the key genes and signaling pathways related to GC oncogenesis and progression using bioinformatics and meta-analysis methods.

Methods: Eligible microarray datasets were downloaded and integrated using the meta-analysis method. According to the tumor stage, GC gene chips were classified into three groups. Thereafter, the three groups' differentially expressed genes (DEGs) were identified by comparing the gene data of the tumor groups with those of matched normal specimens. Enrichment analyses were conducted based on common DEGs among the three groups. Then protein–protein interaction (PPI) networks were constructed to identify relevant hub genes and subnetworks. The effects of significant DEGs and hub genes were verified and explored in other datasets. In addition, the analysis of mutated genes was also conducted using gene data from The Cancer Genome Atlas database.

Results: After integration of six microarray datasets, 1,229 common DEGs consisting of 1,065 upregulated and 164 downregulated genes were identified. Alpha-2 collagen type I (COL1A2), tissue inhibitor matrix metalloproteinase 1 (TIMP1), thymus cell antigen 1 (THY1), and biglycan (BGN) were selected as significant DEGs throughout GC development. The low expression of ghrelin (GHRL) is associated with a high lymph node ratio (LNR) and poor survival outcomes. Thereafter, we constructed a PPI network of all identified DEGs and gained 39 subnetworks and the top 20 hub genes. Enrichment analyses were performed for common DEGs, the most related subnetwork, and the top 20 hub genes. We also selected 61 metabolic DEGs to construct PPI networks and acquired the relevant hub genes. Centrosomal protein 55 (CEP55) and POLR1A were identified as hub genes associated with survival outcomes.

Conclusion: The DEGs, hub genes, and enrichment analysis for GC with different stages were comprehensively investigated, which contribute to exploring the new biomarkers and therapeutic targets.

Keywords: gastric cancer, microarray datasets, differentially expressed genes, metabolism, meta-analysis

OPEN ACCESS

Edited by:
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

Reviewed by:
Ehsan Nazemalhosseini-Mojarad, Shahid Beheshti University of Medical Sciences, Iran
Daxiang Cui, Shanghai Jiao Tong University, China
Xianwei Mo, Guangxi Medical University Cancer Hospital, China

***Correspondence:**
Peiyu Li
lipeiyu6301@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *23 February 2022*
Accepted: *31 May 2022*
Published: *22 July 2022*

Citation:
Yu Z, Liang C, Tu H, Qiu S, Dong X, Zhang Y, Ma C and Li P (2022) Common Core Genes Play Vital Roles in Gastric Cancer With Different Stages. Front. Genet. 13:881948. doi: 10.3389/fgene.2022.881948

INTRODUCTION

Gastric cancer (GC) is the third leading cause of cancer deaths and the fifth most frequently diagnosed cancer worldwide, with over 1 million new cases and 700,000 deaths each year (Bray et al., 2018). Most GC patients are already in the advanced stage at the initial visit, and these patients have a poor prognosis and even have to undergo radical surgery (Irino et al., 2021; Manzanedo et al., 2021; Stocker et al., 2021). In the past decades, gene sequencing and molecular targeted therapy have been increasingly widely used in clinical practice. Therefore, it is of great clinical value to explore the core genes and molecular mechanisms in pathogenesis for the diagnosis and treatment of GC.

In an analysis of key circulating exosomal miRNAs, four key miRNAs (hsa-miR-130b-3p, hsa-miR-151a-3p, hsa-miR-15b-3p, and hsa-miR-1246) and the interaction network or enrichments based on their targets (TAOK1, CMTM6, SCN3A, WASF3, IGF1, CNOT7, GABRG1, and PRKD1) may help understand the molecular mechanisms in the GC development (Qian et al., 2021). The study by Mou et al. (2015)demonstrated that four SNP loci (rs2279115, rs804270, rs909253, and rs3765524) showed a potential association with GC risk. In addition, many targets and markers have been applied to the diagnosis and treatment of GC in basic experiments and achieved preliminary results. HER2 monoclonal antibody conjugated RNase-A-associated CdTe quantum dots (Ruan et al., 2012) and BRCAA1 monoclonal antibody conjugated fluorescent magnetic nanoparticles (Wang et al., 2011) both exhibited great potential in applications such as *in situ* GC targeted imaging and selective therapy of GC. The study by Zhang et al. (2015)showed that miR-19b-3p and miR-16-5p were biomarkers that own great potential in applications such as early screening and progression evaluation of GC. Furthermore, the microarray-based prewarning system, which could be applied in the early detection of GC, was developed by Cui et al. (2005).

At present, many bioinformatic studies about GC have been published, but the differentially expressed genes (DEGs) and signaling pathways revealed by different studies were not consistent. The study by Yang and Gong (2021) showed that OLFM4, IGF2BP3, CLDN1, and MMP1 were the most significantly upregulated DEGs, which significantly enriched in negative regulation of growth, fatty acid binding, and cellular response to zinc ions. In a bioinformatics analysis conducted by Xu et al. (2021), the expressions of ITGB1 and alpha-2 collagen type I (COL1A2) were significantly increased in GC tissues, and 63 characteristic DEGs were mainly involved in regulating extracellular matrix (ECM)–receptor interactions and the PI3K-Akt signaling pathway. Xu et al. (2020) also found that SLC1A3 promotes GC progression via the PI3K/AKT signaling pathway.

Most published experimental and bioinformatic studies included GC specimens with unclear tumor stages, making it impossible to accurately analyze the DEGs and signaling pathways throughout GC development (Yang et al., 2021; Xu et al., 2021). In this study, therefore, we retrieved three microarray datasets containing gene data with definite GC stages and then divided them into the early stage (ES) group and the late stage (LS) group. The two groups' DEGs were obtained by comparing GC tissues of the ES and LS groups with adjacent noncancerous gastric tissues. Another three microarray datasets containing gene data with indefinite GC stages were also collected, and DEGs in GC were identified relative to normal tissues. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted on DAVID (https://david.ncifcrf.gov/) and Sangerbox 3.0 (http://vip.sangerbox.com/home.html) using common DEGs from the three groups. Thereafter, protein–protein interaction (PPI) network was constructed using the STRING online tool (https://cn.string-db.org/) and Cytoscape software (Shannon et al., 2003). The effects of significant DEGs and hub genes were verified and explored in other datasets. The related gene expression data and clinical information were also obtained from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/), which were used to carry out overall survival analysis and somatic mutation analysis. Besides this, hub genes associated with metabolic KEGG pathways were also identified. Through this analysis, we identified the key genes and signaling pathways related to GC, aiming to provide the experimental basis and important insight of new biomarkers and prognostic markers.

MATERIALS AND METHODS

Inclusion and Integration of Microarray Datasets

Eligible microarray datasets were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). Datasets in accordance with the following criteria were included and considered for subsequent analysis: upload data were between 2010 to 2021; contained the gene data on GC tissues and adjacent normal tissues and at least three samples per group; tissue samples used were from humans; and detail information on technology and platform were obtainable.

When the staging of GC specimens was definite, gene chips were divided into the ES group and LS group based on tumor stages. The ES group and LS group respectively incorporated stage I–II and stage III–IV GC's microarray data, which were staged according to standards recommended by the American Joint Committee on Cancer. GC gene chips with indefinite tumor stages were classified as the mixed stage (MS) group, incorporating stage I–IV GC's microarray data. DEGs were obtained by comparing the microarray data of GC with those of adjacent noncancerous gastric tissues. Thereafter, meta-analysis was performed to integrate microarray data for the above each group, and the three groups' DEGs would be obtained. DEGs from the three groups were intersected to identify common DEGs. In addition, DEGs were ranked according to adjusted *p*-value (adj.P.Val) from small to large, and the most significant DEGs were also gained by intersecting the top 100 significant DEGs from the three groups. The effects of above significant DEGs were verified and explored in other GEO

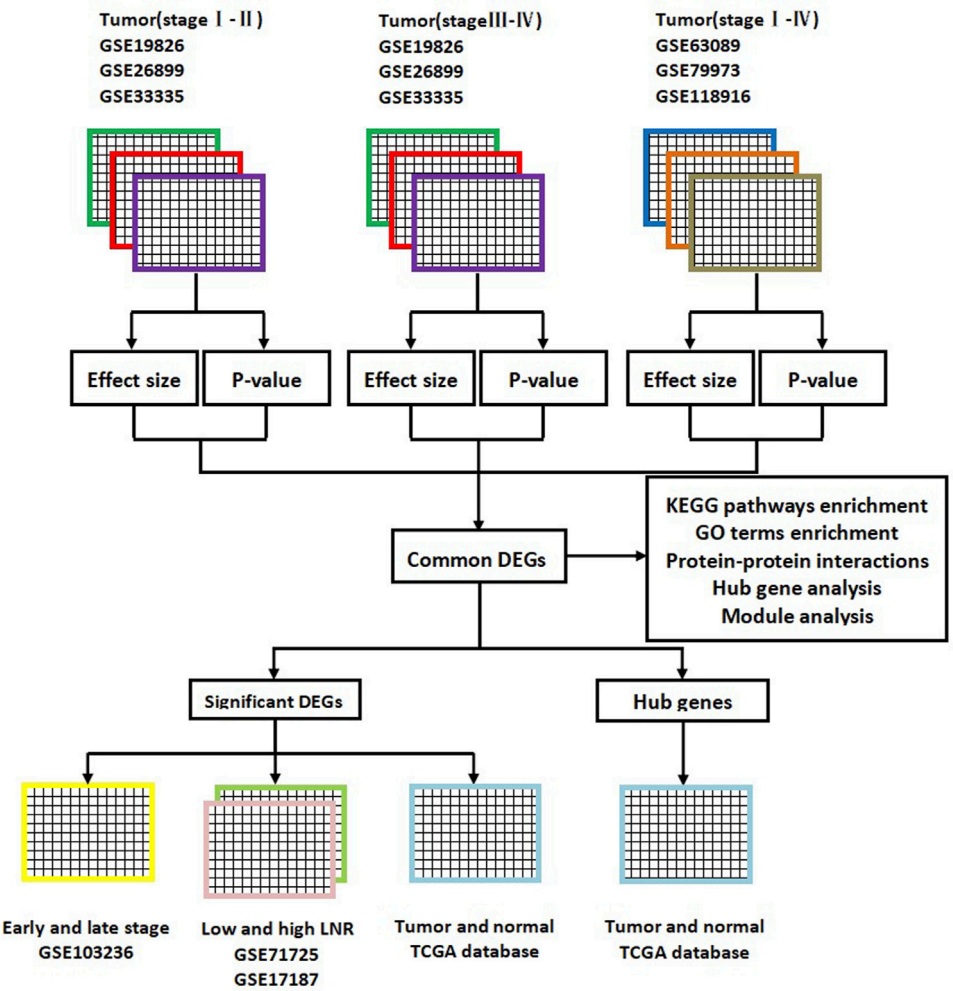


FIGURE 1 | Flow diagram of this bioinformatics and meta-analysis. DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes; GO, gene ontology; LNR, lymph node ratio; TCGA, The Cancer Genome Atlas.

datasets (GSE103236, GSE51725, and GSE17187) and TCGA database.

Enrichment Analysis of Gene Ontology Terms and Kyoto Encyclopedia of Genes and Genomes Pathways

Common DEGs were divided into the upregulated group and the downregulated group. Next, GO terms and KEGG pathways enrichment analyses of upregulated DEGs and downregulated DEGs were respectively performed using the DAVID online tool. GO terms consisted of the following items: biological process (BP), molecular function (MF), and cellular component (CC). KEGG pathway analysis was designed to identify significantly enriched pathways of molecular interactions and reactions. The flow diagram of this bioinformatics and meta-analysis is shown in **Figure 1**.

Establishment and Analysis of Protein–Protein Interaction Network

The PPI network of DEGs was constructed based on the STRING database, with a confidence score set as 0.9 (highest confidence). Then, PPI network files were imported to Cytoscape v3.8 software, in which we constructed the subnetwork using the MCODE plug-in and calculated the top 20 genes based on the Multiscale Curvature Classification (MCC) algorithm. The DAVID, Metascape (https://metascape.org/), and Sangerbox 3.0 online tools were used for enrichment analysis and visualization of hub genes and the first subnetwork. The effects of hub genes were explored and proved in another three GEO microarray datasets (GSE103236, GSE51725, and GSE17187) and TCGA database. The DEGs converged on KEGG pathways associated with metabolism and were also inputted into the STRING database and Cytoscape software to obtain relevant PPI networks and hub genes.

TABLE 1 | Characteristics of included GEO datasets.

Datasets	Platforms	Region	Sample size (Tumors) ES Group	LS Group	MS group	Sample Size (Controls)	Submission Date (year)
GSE19826	GPL570	China	6	6	—	12	2010
GSE26899	GPL6947	United States	23	36	—	12	2011
GSE33335	GPL5175	China	13	11	—	24	2011
GSE63089	GPL5175	China	—	—	45	45	2014
GSE79973	GPL570	China	—	—	10	10	2016
GSE118916	GPL15207	China	—	—	15	15	2018

GEO, Gene Expression Omnibus; GSE, GEO datasets; GPL, GEO platforms; ES, early stage; LS, late stage; MS, mixed stage.

Statistical Analysis

Integrative meta-analysis of GEO data (ImaGEO) (https://imageo.genyo.es/), an online tool, was adopted to integrate microarray data (Toro-Domínguez et al., 2019). What is more, the fixed effect model of the effect size method and Fisher's *p*-value method were both applied to perform a meta-analysis for each group. The adj.P.Val < 0.05 and |Log2 fold change (logFC)| > 1 were set as the parameters to identify DEGs. The Wilcoxon test was employed to compare the statistical differences between the two groups, while the Kruskal–Wallis (KS) test was adopted as a statistical method to compare multiple groups. And beyond that, all cut-off *p*-values were set as $p < 0.05$ in this analysis.

RESULTS

Microarray Dataset Integration and Differentially Expressed Gene Identification

First, the microarray datasets containing GSE19826, GSE26899, GSE33335, GSE63089, GSE79973, and GSE118916 were acquired for further analyses. Inadaptable gene chips were removed from selected microarray datasets. After selection, the expression data of 42 GC tissues from GSE19826, GSE26899, and GSE33335 were classified as the ES group, while the expression data of another 53 GC tissues from the above same microarray datasets were classified as the LS group. In addition, GSE63089, GSE79973, and GSE118916 respectively included the gene data of 45, 10, and 15 GC tissues and were classified as the MS group. The characteristics of selected GEO microarray datasets are shown in **Table 1**.

After intersecting the DEGs obtained using two different meta-analysis methods, 4,125, 3,699, and 3,531 DEGs were gained from the ES, LS, and MS groups, respectively. Then we intersected the above DEGs and identified 1,229 common DEGs, which consisted of 1,065 upregulated genes and 164 downregulated genes. In addition, ATP4A, ATP4B, CPA2, ghrelin (GHRL), KCNE2, GIF, ESRRG, COL1A2, tissue inhibitor matrix metalloproteinase 1 (TIMP1), ADH7, AQP4, thymus cell antigen 1 (THY1), and biglycan (BGN) were identified by intersecting the top 100 significant DEGs from the three groups. GSE103236, GSE51725, and GSE17187 were used to verify the effects of the above 12 DEGs (except AQP4) in tumorigenesis and development. By analyzing the gene data of GSE103236, we found that the expression levels of ESRRG, COL1A2, TIMP1, THY1, and BGN in GC tissues with ES and GC tissues with advanced stage were both significantly different than those in paracancer tissues (**Figure 2A**). GSE51725 and GSE17187 based on the same GEO platform (GPL570) were normalized, and the batch effect was removed. Compared with the expression data before normalizing (**Figure 2B**), those after removing the batch effect were found to be more consistent (**Figure 2C**). In the analysis of integrated gene data, a noticeable correlation was found between the expression level of GHRL and the LNR of the tumor (**Figure 2D**). Moreover, analysis based on TCGA database showed that the expression levels of 12 significant DEGs (except GIF) were significantly different between GC tissues and adjacent normal gastric tissues (**Figure 3A**), and statistical correlation could also be found between the expression levels of COL1A2, TIMP1, THY1, and BGN and tumor stage (**Figure 3B**).

Enrichment Analysis Based on Common Differentially Expressed Genes

Enrichment analyses of KEGG pathways and GO terms were both performed for upregulated and downregulated DEGs using the DAVID online tool. Regarding upregulated DEGs, KEGG pathway analysis showed that DEGs were mainly enriched in DNA replication, ECM–receptor interaction, pyrimidine metabolism, cell cycle, and purine metabolism (**Figure 4A**), while DEGs were mainly enriched in cell division, DNA replication, mitotic nuclear division, sister chromatid cohesion, and G1/S transition of the mitotic cell cycle in GO terms analysis (**Figure 4B**). As for downregulated DEGs, gastric acid secretion, chemical carcinogenesis, retinol metabolism, drug metabolism–cytochrome P450 and metabolism of xenobiotics by cytochrome P450 were KEGG pathways that play important roles (**Figure 4C**), while GO terms analysis showed that DEGs were mainly enriched in digestion, the xenobiotic metabolic process, potassium ion import, and gastric acid secretion (**Figure 4D**). It was also found that most KEGG pathways of downregulated DEGs were related to metabolism, including retinol metabolism, drug metabolism–cytochrome P450, metabolism of xenobiotics by cytochrome P450, metabolic pathways, glycolysis/gluconeogenesis, histidine, alanine, aspartate, glutamate, and tyrosine metabolism. In the analysis of upregulated DEGs, pyrimidine metabolism and purine metabolism were KEGG pathways associated with metabolism. In the end, 31 upregulated DEGs and 30 downregulated DEGs

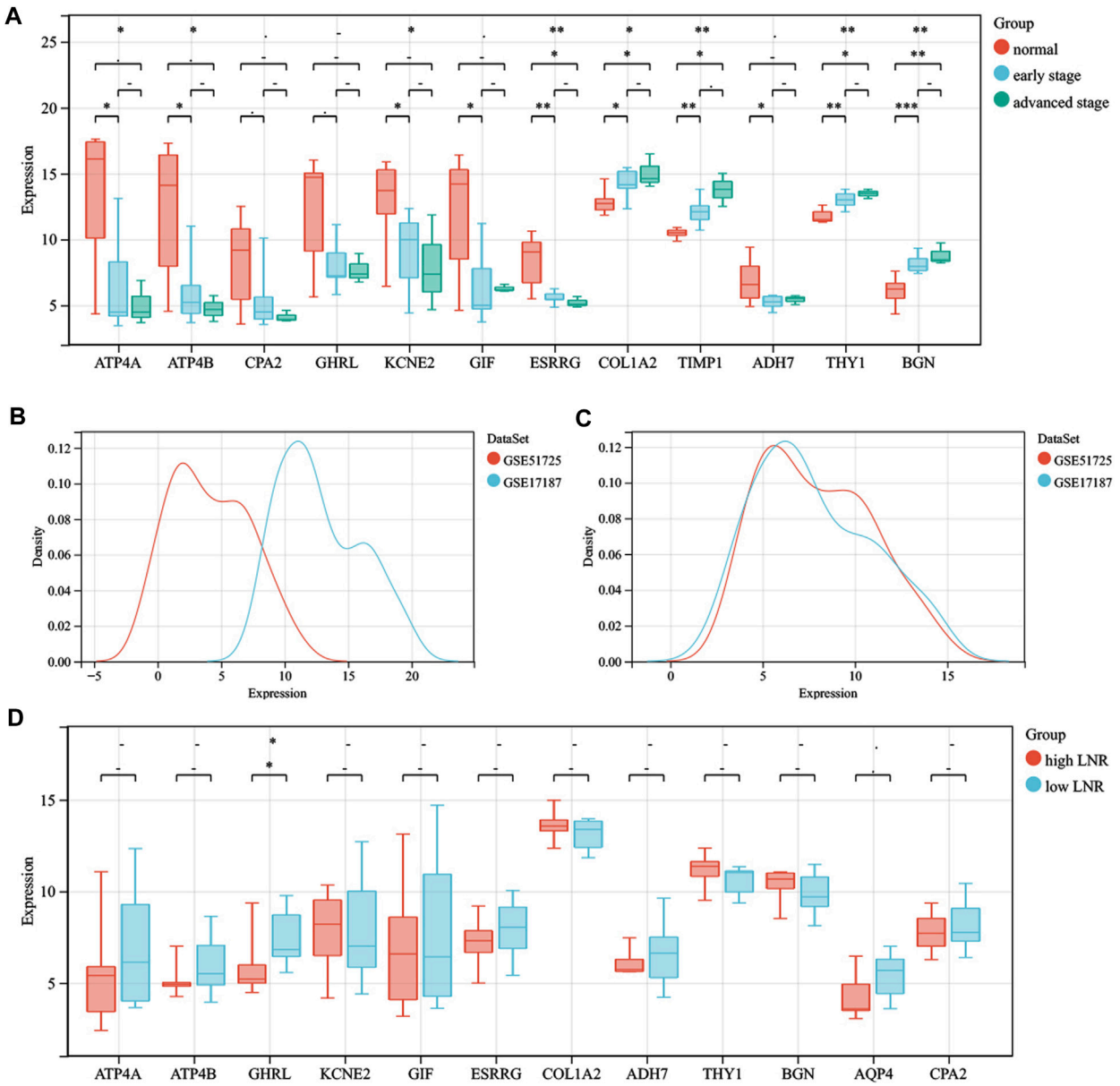


FIGURE 2 | The roles that significant DEGs play in the oncogenesis, progression, and lymph node metastasis. **(A)** comparison of the expression levels of ATP4A, ATP4B, CPA2, GHRL, KCNE2, GIF, ESRRG, COL1A2, TIMP1, ADH7, THY1, and BGN between paracancer tissues, GC tissues with early stage and GC tissues with late stage. **(B)** gene data before normalization. **(C)** gene data after normalization. **(D)** comparison of the expression levels of ATP4A, ATP4B, GHRL, KCNE2, GIF, ESRRG, COL1A2, ADH7, THY1, BGN, AQP4, and CPA2 between GC tissues with high LNR and GC tissues with low LNR. The statistical methods adopted for A and D are the Kruskal–Wallis test and Wilcoxon test, respectively. ˜, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$. DEGs, differentially expressed genes; GC, gastric cancer; LNR, lymph node ratio.

pooled in metabolism-related KEGG pathways were selected for further analysis.

Protein–Protein Interaction Network Construction and Hub Gene Identification

The PPI network of 1,229 common DEGs, which involved 656 nodes (DEGs) and 2,701 edges, was visualized using Cytoscape v3.8 software. As analysis results of the MCODE plug-in showed, 39 subnetworks were identified, and the first cluster (**Figure 5A**) was related to mitotic cell cycle, PID Aurora B pathway, regulation of the cell cycle process and chromosome segregation, PID FOXM1 pathway, and PID PLK1 pathway (**Figure 5B**). Hub genes such as topoisomerase II alpha (TOP2A), cyclin-B2 (CCNB2), KIF11, cyclin-A2 (CCNA2), cell division cycle 20 (CDC20), cell division cycle–associated protein 8 (CDCA8), KIF20A, benzimidazoles 1 homolog beta (BUB1B), targeting protein for Xenopus kinesin-like protein 2

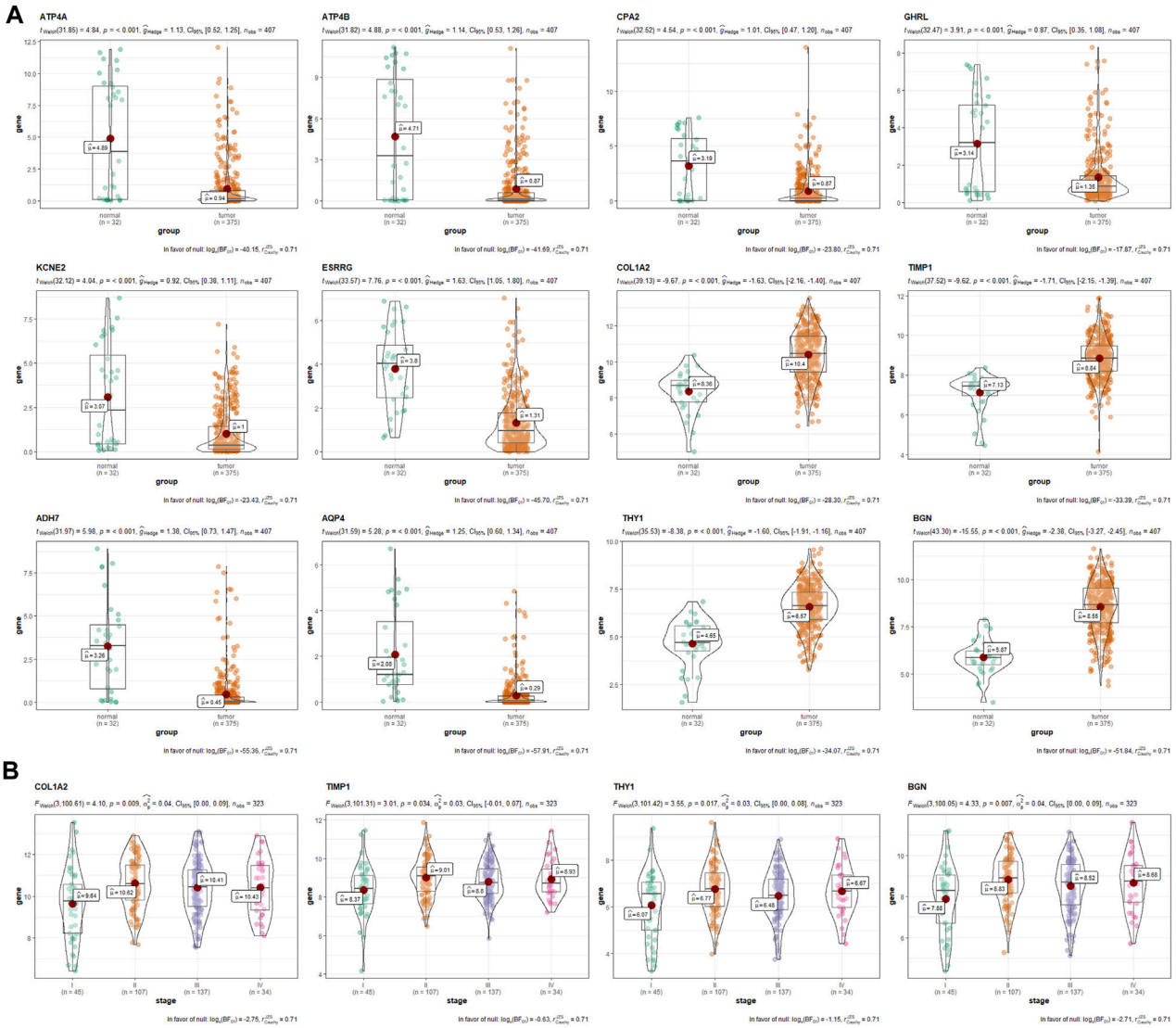


FIGURE 3 | The verification and exploration of significant DEGs in TCGA database. **(A)** comparison of the expression levels of ATP4A, ATP4B, CPA2, GHRL, KCNE2, ESRRG, COL1A2, TIMP1, ADH7, AQP4, THY1, and BGN between GC and paracancer tissues from TCGA database. **(B)** the association between the expression levels of COL1A2, TIMP1, THY1, and BGN with tumor stage. DEGs, differentially expressed genes; TCGA, The Cancer Genome Atlas; GC, gastric cancer.

(TPX2), and KIF2C were identified using an MCC algorithm (**Figure 5C**). Enrichment analysis indicated that the hub genes were significantly enriched in cell division, nuclear division, mitotic cell cycle, microtubule cytoskeleton, PID Aurora B pathway, PID PLK1 pathway, and PID FOXM1 pathway (**Figure 5D**).

In addition, PPI networks and relevant hub genes of 31 upregulated metabolic DEGs (**Figures 5E,F**) and 30 downregulated metabolic DEGs (**Figures 5G,H**) were both constructed and identified. Further, the KM plotters of significant DEGs and hub genes indicated that the expression levels of GHRL, centrosomal protein 55 (CEP55), and POLR1A were associated with survival outcomes in patients with GC (**Figure 5I**).

Mutated Genes Analysis Based on The Cancer Genome Atlas Database

A summary of gene mutation information of GC patients from TCGA database is shown in **Figure 6A**. Somatic mutation profiles for 437 GC patients were retrieved. And we used the waterfall plot to present the mutation data for every gene in every sample (**Figure 6B**). The significant signatures of mutated genes were also explored (**Figure 6C**). Further, we assessed the effects of the top 10 mutated genes (TTN, TP53, MUC16, LRP1B, SYNE1, ARID1A, CSMD3, FAT4, FLG, and PCLO) in tumor progression. As the statistical analysis showed, TP53 and SYNE1 were associated with tumor stage (**Figure 6D**), while TTN, LRP1B, and FAT4 were associated with survival events (**Figure 6E**).

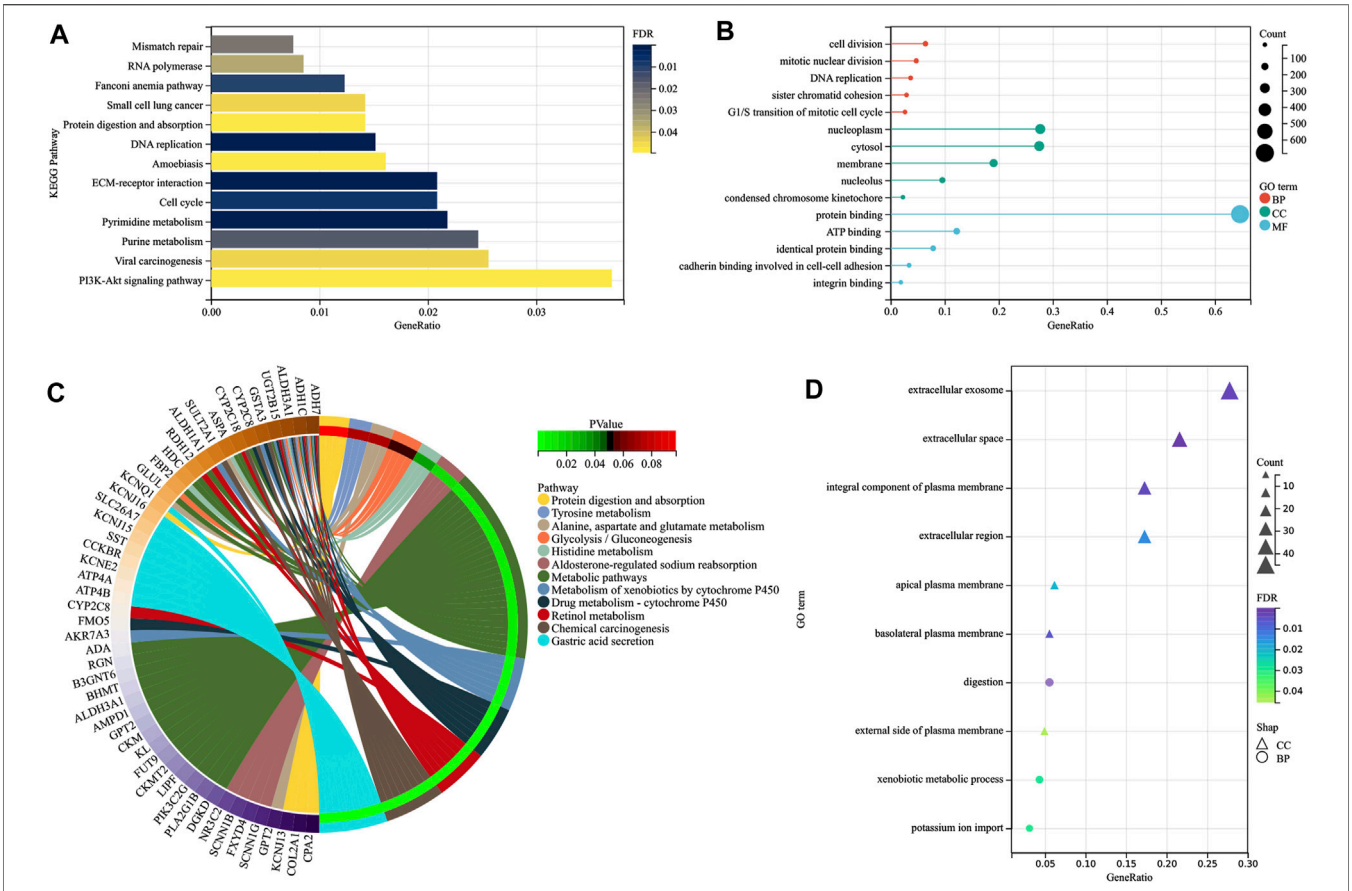


FIGURE 4 | Visualization of the enrichment results. Distribution of the upregulated DEGs in GC for different **(A)** KEGG pathways and **(B)** GO-enriched functions, and distribution of the down -regulated DEGs in GC for different **(C)** KEGG pathways and **(D)** GO-enriched functions. DEGs, differentially expressed genes; GC, gastric cancer; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene ontology; BP, biological process; CC, cellular component; MF, molecular function.

DISCUSSION

As one of the most common malignant tumors with high mortality, GC has always been a serious disease threatening human health. Risk prediction, early diagnosis, and precise therapies are considered the essential measures to improve the prognosis of GC patients. Owing to complex BP in GC occurrence and development, however, existing markers and therapeutic targets could not significantly improve diagnosis or 5-year survival rate (Ajani et al., 2016; Wang et al., 2019). Therefore, a study on the pivotal mechanism of GC occurrence and development is urgently needed, which helps select sensitive and specific biomarkers or therapeutic targets.

In this study, the gene expression data of GSE19826, GSE26899, GSE33335, GSE63089, GSE79973, and GSE118916 were integrated through a meta-analysis method. Then enrichment analysis, PPI network construction, and hub gene identification were performed based on integrated data. In total, 1,065 upregulated and 164 downregulated DEGs throughout GC development were screened out. Enrichment analyses of GO terms and KEGG pathways indicated that DEGs could lead to the incidence and exacerbation of GC by affecting DNA replication, cell and chromosome division, and related metabolic pathways. Furthermore, ECM–receptor interaction and PI3K-Akt were also KEGG pathways that DEGs significantly enriched. The tumor microenvironment (TME), a complex ecosystem composed of ECM, peripheral blood vessels, other non-malignant cells, and signaling molecules, has been proved to be important for tumor invasion, progression, and chemoresistance (Ran et al., 2021). The biological function of PI3K-Akt signaling in GC progression has been well established, which could regulate tumor cell growth, proliferation, apoptosis, and energy metabolism. An activated PI3K-Akt signaling pathway could promote GC progression by enhancing glycolysis, stabilizing mitochondrial membrane potentials, and inhibiting tumor cell apoptosis (Xu et al., 2020). In addition, abundant metabolic pathways were found to be associated with GC, revealing the close connection between the metabolic system and the identified DEGs (Tian et al., 2020). To date, accumulated evidence has suggested that cancer is a metabolic disease, in which cells have lost their normal checks on cell proliferation, resulting in excessive bioenergetic and biosynthetic needs. Therefore, cancer cells must alter their metabolism to sustain such a high demand (Kroemer and

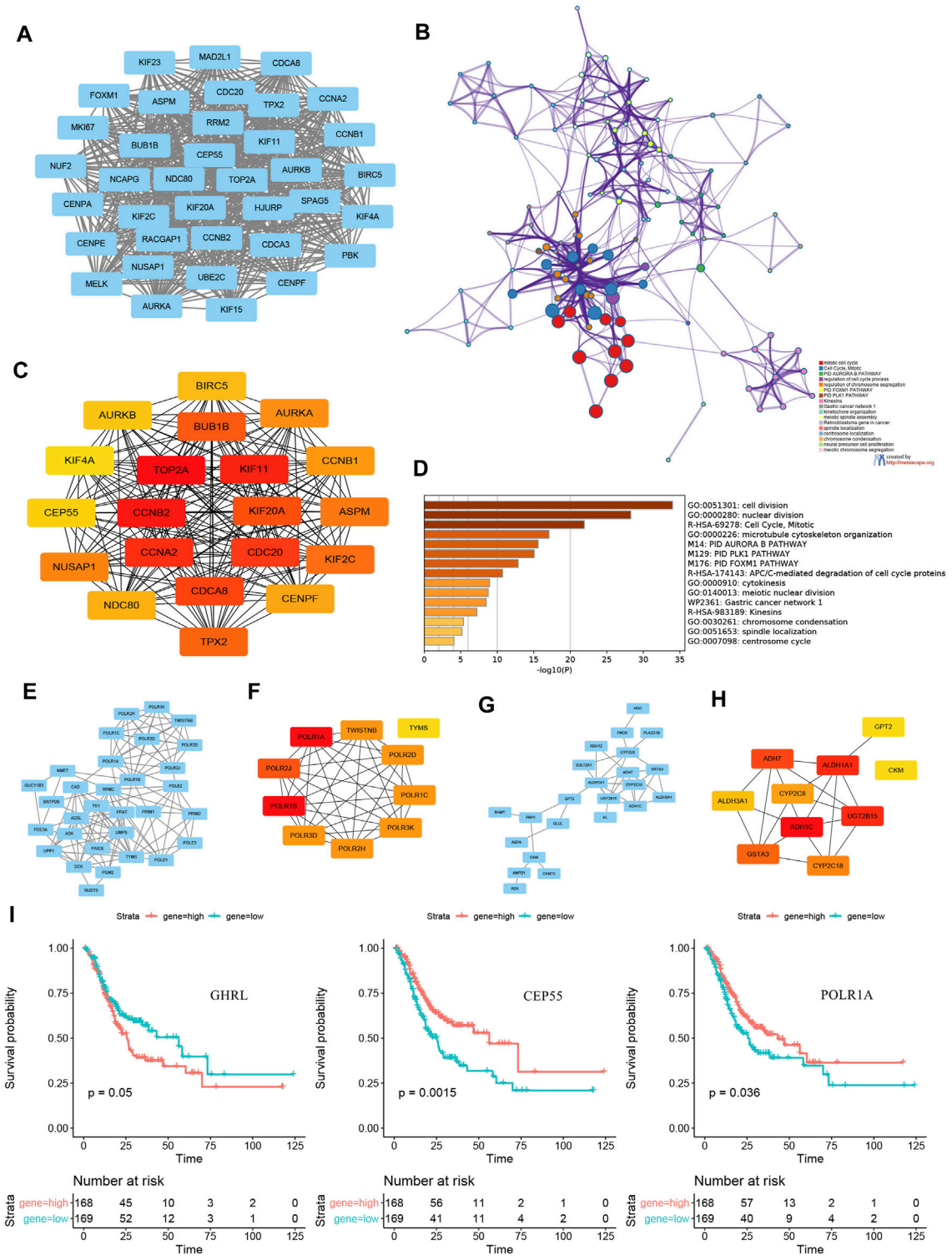


FIGURE 5 | PPI subnetwork construction and hub genes identification. **(A)**the first subnetwork of PPI using MCODE plug-in of Cytoscape. **(B)** enrichment analysis and visualization of first cluster using Metascape online tool. **(C)** top 20 hub genes of all common DEGs identified by an MCC algorithm. **(D)** enrichment analysis and visualization of top 20 hub genes using the Metascape online tool. PPI networks and relevant hub genes of **(E–F)** 31 upregulated metabolic DEGs and **(G–H)** 30 downregulated metabolic DEGs. **(I)** survival curves of GHRL, CEP55, and POLR1A expression. PPI, protein–protein interaction; DEGs, differentially expressed genes; MCC, Multiscale Curvature Classification.

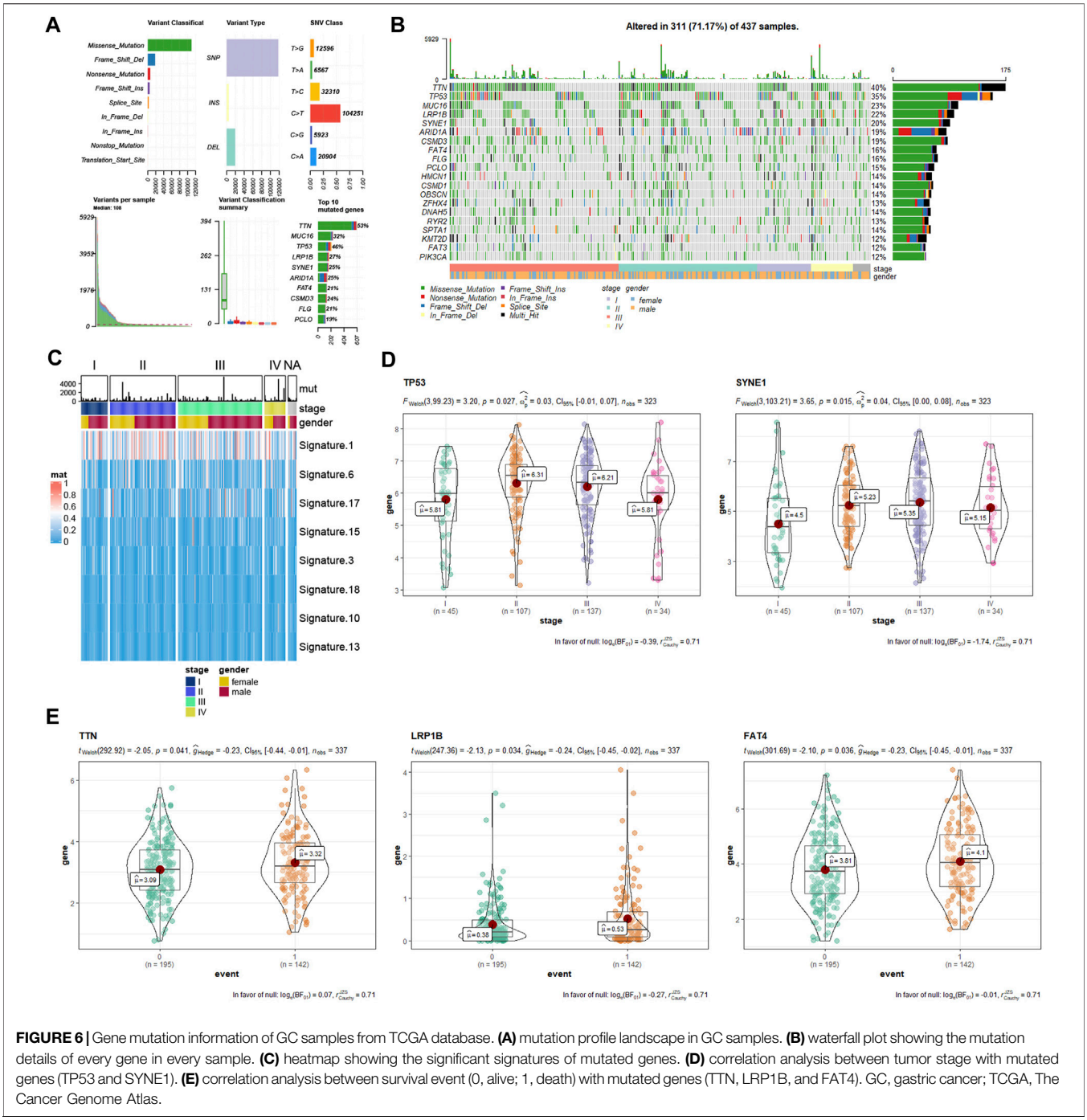


FIGURE 6 | Gene mutation information of GC samples from TCGA database. **(A)** mutation profile landscape in GC samples. **(B)** waterfall plot showing the mutation details of every gene in every sample. **(C)** heatmap showing the significant signatures of mutated genes. **(D)** correlation analysis between tumor stage with mutated genes (TP53 and SYNE1). **(E)** correlation analysis between survival event (0, alive; 1, death) with mutated genes (TTN, LRP1B, and FAT4). GC, gastric cancer; TCGA, The Cancer Genome Atlas.

Pouyssegur, 2008; Li et al., 2013; Duda et al., 2020). Large amounts of purines and pyrimidines are required for cell proliferation, DNA replication, and energy supply of GC tissues; thus, pyrimidine metabolism and purine metabolism are metabolic pathways upregulated DEGs significantly enriched. Moreover, the majority of downregulated DEGs converged on metabolic pathways such as metabolism of retinol, drug–cytochrome P450, glycolysis/gluconeogenesis, histidine, alanine, aspartate, glutamate, and tyrosine. It has long been recognized that tumor metabolism preferentially relies on glycolysis instead of oxidative phosphorylation of glucose despite the status of oxygen supply, and this type of metabolism is known as the Warburg effect (Kroemer and Pouyssegur, 2008). As a result, key enzymes in the oxidative phosphorylation of glucose are downregulated for cancer cell proliferation and tumorigenicity. Fructose-1,6-bisphosphatase-2 (FBP2), one kind of enzyme participating in glycogen synthesis from carbohydrate precursors, can catalyze the hydrolysis of fructose-1,6-bisphosphate to fructose-6-phosphate and inorganic phosphate in glucose metabolism.

FBP2 underexpression may contribute to GC tumor development by stimulating glucose metabolism and inhibiting cell proliferation (Li et al., 2013; Duda et al., 2020). In recent years, glucolipid metabolism therapy of tumors has become a research hotspot, which aims to inhibit the proliferation and metastasis of tumor cells by controlling glycolysis and increasing fat for energy (Hur et al., 2013; Sun et al., 2018). Gastric acid secretion, another important pathway that downregulated DEGs enriched, has been reported by few studies. GC was considered to be associated with gastritis, and intestinal metaplasia developing in atrophic gastritis was believed to be a step in the gastric carcinogenic process. Therefore, patients with GC had reduced gastric acid secretion because of oxyntic atrophic gastritis (Waldum and Rehfeld, 2019; Liu et al., 2020a). Furthermore, decreased GHRL in blood also contributed to this phenomenon.

Up to now, tumor markers such as CEA, CA125, CA199, and CA72-4 have been used as indicators for GC diagnosis. Her2, epidermal growth factor receptor (EGFR), TP53, and PI3K are key oncogenes reported to work (Li et al., 2022; Zhao et al., 2022). Besides these, Xie et al. (2015) established a multi-index prediction model based on the six kinds of biomarkers (CEA, CA199, H.P., P53, PG Ⅰ, and PG Ⅱ), which was designed to achieve early screening and therapeutic evaluation of GC patients. However, the above biomarkers and therapeutic targets had insufficient specificity in the early diagnosis of GC and could not accurately evaluate tumor progression and survival time (Gao and Yang, 2022). In this study, COL1A2, TIMP1, THY1, and BGN were identified as significant DEGs for both the GC with ES and the GC with LS. Further analysis showed that the expression levels of the above four DEGs were associated with tumor stage, and the expression level of GHRL was associated with LNR and survival outcome. Recurrence and metastasis, the primary factors affecting the survival outcome of GC patients, are also the hotspots of current research. The above DEGs may influence tumor stage and LNR by affecting GC recurrence and metastasis. COL1A2 gene can affect cell proliferation, differentiation, adhesion, and metastasis by encoding type I collagen—the most widely expressed collagen among the fibrous collagen family (Pan et al., 2021; Xu et al., 2021). TIMP1 is a natural collagenase inhibitor that can inhibit apoptosis, induce angiogenesis, and stimulate cell proliferation, which may be directly involved in the progression and metastasis of cancers such as GC, breast cancer, and colon cancer (Zhang et al., 2020; Liu et al., 2021). THY1, also known as the CD90 gene, plays an important regulatory role in the cellular interactions between a cell and its matrix. Moreover, previous studies indicated that the THY1 gene can inhibit GC cells' apoptosis by regulating secreted protein acidic and rich in cysteine protein's expression levels (Wu et al., 2019; Wang et al., 2021). As an important component of ECM proteins, BGN seems to play an important role in the oncogenesis and progression of various cancers. According to the research by Hu et al. (2016), BGN may promote cancer progression through the chronic activation of tumor angiogenesis, so high expression of BGN was observed in advanced GC. In addition, tight connections between the above four hub genes could be found; thus, definitive linkages and interaction mechanisms needed to be further explored. Therefore, COL1A2, TIMP1, THY1, BGN, and GHRL have great potential in applications such as early screening of GC patients, prediction of therapeutic outcomes, and real-time dynamic monitoring the progress of GC in near future.

TOP2A, CCNB2, KIF11, CCNA2, CDC20, CDCA8, KIF20A, BUB1B, TPX2, and KIF2C were identified as the top 10 hub genes among all DEGs. TOP2A is one kind of type II DNA topoisomerases that can relax negative and positive supercoiling during replication and transcription. The expression level of the TOP2A gene can reflect tumor proliferation and was reported to be associated with peritoneal and hematogenous recurrence. Thus, TOP2A may be utilized as a specific drug target for malignant tumors such as GC (Hou et al., 2020). As previous studies found, CCNB2 was a cell cycle–related gene that can promote the proliferation and tumor growth of GC cells (Liu et al., 2020b). Similar to CCNB2, CCNA2 is a highly conserved cyclin that plays a critical role in the control of the cell cycle at G1/S and in the G2/M transition (Lee et al., 2020). Kinesin superfamily (KIF), a group of proteins that possess ATPase activity and motion characteristics, participate in numerous cellular biological activities such as mitosis and meiosis. Kruppel (KIF) 11, KIF20A, and KIF2C genes have been illustrated as genes that might function as oncogenes in GC (Imai et al., 2017; Sheng et al., 2018). Several lines of evidence have shown that CDC20 plays a vital role in the correct functioning of the spindle assembly checkpoint (SAC), and overexpression of the CDC20 gene is related to intestinal histology and favorable clinicopathologic parameters in GC (Kim et al., 2019). In addition, multidomain protein kinase budding uninhibited by BUB1B may contribute to the process of SAC, which can delay the separation of sister chromatids to prevent defects in segregation (Hudler et al., 2016). CDCA8, also called Borealin/Dasra B, is a crucial cell cycle–regulated chromosomal passenger protein, and its nuclear accumulation is correlated with a poor prognosis for GC (Chang et al., 2006). TPX2 is a microtubule-associated protein that relates to chromosomal instability and helps format normal bipolar spindles and chromosome segregation (Tomii et al., 2017).

Further, the expression levels of GHRL, CEP55, and POLR1A were proved to be associated with survival outcomes in patients with GC. GHRL, a small peptide characterized as the ligand of the growth hormone secretagogue receptor (GHSR), plays role in stimulating pituitary growth hormone release, the regulation of energy balance, gastric acid release, appetite, insulin secretion, gastric motility, and the turnover of the gastric and intestinal mucosa. Previous studies have shown that GHRL works in several key processes of cancer progression, such as cell proliferation, migration, and invasion. The action mechanism of GHRL in promoting or inhibiting cancer progression, however, is still unclear. The regulation of the GHRL–GHSR axis may play a potential critical role (Lin and Hsiao, 2017). As a member of the centrosomal relative protein family, CEP55 has been reported to participate in cell cycle regulation. Tao's study showed that ectopic overexpression of CEP55 could enhance the cell proliferation, colony formation, and tumorigenicity of GC cells, and CEP55 knockdown induces cell cycle arrest at the G2/M phase in GC cells. Besides, the expression of CEP55 can

affect the PI3K/AKT/p21 signaling pathway and cyclin pathway-related proteins (Tao et al., 2014). Folate receptor gene family has a high affinity for folic acid and several reduced folic acid derivatives, and it mediates delivery of 5-methyltetrahydrofolate to the interior of cells. Moreover, mature FOLR1 is an N-glycosylated protein that is predominantly expressed on epithelial cells and is dramatically upregulated on many carcinomas (Kim et al., 2018).

CONCLUSION

In this study, six eligible microarray datasets were integrated to present the gene expression signatures of GC relative to normal gastric tissues using ImaGEO meta-analysis. COL1A2, TIMP1, THY1, and BGN were identified and verified as significant DEGs throughout GC progression, and the above DEGs are expected to be used as the target molecules in GC diagnosis and therapy. In the future, basal experimentation and clinical tests are needed to verify their roles in early screening and tumor preventing and controlling of GC patients. As enrichment analysis showed, the upregulated DEGs mainly enriched in DNA replication, cell cycle, and ECM–receptor interaction, while most of the downregulated DEGs were related to metabolism. TOP2A, CCNB2, KIF11, CCNA2, CDC20, CDCA8, KIF20A, BUB1B, TPX2, and KIF2C were identified as the top 10 hub genes among all DEGs. Further, the expression levels of GHRL, CEP55, and POLR1A were proved to be associated with survival outcomes in patients with GC. The interaction between hub genes, the intervention mechanism of hub genes on tumors, and the association between hub genes with survival outcomes are the main direction for the next research step.

The identified DEGs, hub genes, and signaling pathways may help us understand the molecular mechanisms of gastric tumor and discover new biomarkers and therapeutic targets for gastric tumor.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The data used in this study are from publicly available datasets. Ethical approval and informed consent for participation were not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YY, LC, and LY designed the study and wrote and amended the original draft. TY and QZ participated in searching microarray datasets and analyzing the gene data. DY, ZH, and MC produced the tables and figures. All authors read and approved the final manuscript.

FUNDING

The current study was supported by a research grant from Capital's funds for health improvement and research (2020–1–1021).

ACKNOWLEDGMENTS

We thank Gene Expression Omnibus and The Cancer Genome Atlas for providing detailed gene data and clinical information.

REFERENCES

Ajani, J. A., D'Amico, T. A., Almhanna, K., Bentrem, D. J., Chao, J., Das, P., Denlinger, C. S., Fanta, P., Farjah, F., Fuchs, C. S., Gerdes, H., Gibson, M., Glasgow, R. E., Hayman, J. A., Hochwald, S., Hofstetter, W. L., Ilson, D. H., Jaroszewski, D., Johung, K. L., Keswani, R. N., Kleinberg, L. R., Korn, W. M., Leong, S., Linn, C., Lockhart, A. C., Ly, Q. P., Mulcahy, M. F., Orringer, M. B., Perry, K. A., Poultsides, G. A., Scott, W. J., Strong, V. E., Washington, M. K., Weksler, B., Willett, C. G., Wright, C. D., Zelman, D., McMillian, N., and Sundar, H. (2016). Gastric Cancer, Version 3.2016, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 14 (10), 1286–1312. doi:10.6004/jnccn.2016.0137

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA a cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Chang, J.-L., Chen, T.-H., Wang, C.-F., Chiang, Y.-H., Huang, Y.-L., Wong, F.-H., Chou, C.-K., and Chen, C.-M. (2006). Borealin/Dasra B Is a Cell Cycle-Regulated Chromosomal Passenger Protein and its Nuclear Accumulation Is Linked to Poor Prognosis for Human Gastric Cancer. *Exp. Cell. Res.* 312 (7), 962–973. doi:10.1016/j.yexcr.2005.12.015

Cui, D.-X., Zhang, L., Yan, X. J., Zhang, L. X., Xu, J. R., Guo, Y. H., et al. (2005). A Microarray-Based Gastric Carcinoma Prewarning System. *Wjg* 11 (9), 1273–1282. doi:10.3748/wjg.v11.i9.1273

Duda, P., Janczara, J., McCubrey, J. A., Gizak, A., and Rakus, D. (2020). The Reverse Warburg Effect Is Associated with Fbp2-dependent Hif1α Regulation in Cancer Cells Stimulated by Fibroblasts. *Cells* 9 (1), 205. doi:10.3390/cells9010205

Gao, W., and Yang, M. (2022). Identification by Bioinformatics Analysis of Potential Key Genes Related to the Progression and Prognosis of Gastric Cancer. *Front. Oncol.* 12, 881015. doi:10.3389/fonc.2022.881015

Hou, G., Deng, J., You, X., Chen, J., Jiang, Y., Qian, T., Bi, Y., Song, B., Xu, Y., and Yang, X. (2020). Mining Topoisomerase Isoforms in Gastric Cancer. *Gene* 754, 144859. doi:10.1016/j.gene.2020.144859

Hu, L., Zang, M.-D., Wang, H.-X., Li, J.-F., Su, L.-P., Yan, M., Li, C., Yang, Q.-M., Liu, B.-Y., and Zhu, Z.-G. (2016). Biglycan Stimulates VEGF Expression in Endothelial Cells by Activating the TLR Signaling Pathway. *Mol. Oncol.* 10 (9), 1473–1484. doi:10.1016/j.molonc.2016.08.002

Hudler, P., Britovsek, N. K., Grazio, S. F., and Komel, R. (2016). Association between Polymorphisms in Segregation Genes BUB1B and TTK and Gastric Cancer Risk. *Radiology Oncol.* 50 (3), 297–307. doi:10.1515/raon-2015-0047

Hur, H., Xuan, Y., Kim, Y. B., Lee, G., Shim, W., Yun, J., Ham, I.-H., and Han, S.-U. (2013). Expression of Pyruvate Dehydrogenase Kinase-1 in Gastric Cancer as a Potential Therapeutic Target. *Int. J. Oncol.* 42 (1), 44–54. doi:10.3892/ijo.2012.1687

Imai, T., Oue, N., Nishioka, M., Mukai, S., Oshima, T., Sakamoto, N., Sentani, K., Matsusaki, K., Yoshida, K., and Yasui, W. (2017). Overexpression of KIF11 in

Gastric Cancer with Intestinal Mucin Phenotype. *Pathobiology* 84 (1), 16–24. doi:10.1159/000447303

Irino, T., Matsuda, S., Wada, N., Kawakubo, H., and Kitagawa, Y. (2021). Essential Updates 2019/2020: Perioperative and Surgical Management of Gastric Cancer. *Ann. Gastroenterol. Surg.* 5 (2), 162–172. doi:10.1002/ags3.12438

Kim, M., Pyo, S., Kang, C. H., Lee, C. O., Lee, H. K., Choi, S. U., and Park, C. H. (2018). Folate Receptor 1 (FOLR1) Targeted Chimeric Antigen Receptor (CAR) T Cells for the Treatment of Gastric Cancer. *PloS one* 13 (6), e0198347. doi:10.1371/journal.pone.0198347

Kim, Y., Choi, J.-W., Lee, J.-H., and Kim, Y.-S. (2019). Spindle Assembly Checkpoint MAD2 and CDC20 Overexpressions and Cell-In-Cell Formation in Gastric Cancer and its Precursor Lesions. *Hum. Pathol.* 85, 174–183. doi:10.1016/j.humpath.2018.10.029

Kroemer, G., and Pouyssegur, J. (2008). Tumor Cell Metabolism: Cancer's Achilles' Heel. *Cancer Cell.* 13 (6), 472–482. doi:10.1016/j.ccr.2008.05.005

Lee, Y., Lee, C. E., Oh, S., Kim, H., Lee, J., Kim, S. B., and Kim, H. S. (2020). Pharmacogenomic Analysis Reveals CCNA2 as a Predictive Biomarker of Sensitivity to Polo-Like Kinase I Inhibitor in Gastric Cancer. *Cancers* 12 (6), 1418. doi:10.3390/cancers12061418

Li, H., Wang, J., Xu, H., Xing, R., Pan, Y., Li, W., Cui, J., Zhang, H., and Lu, Y. (2013). Decreased Fructose-1,6-Bisphosphatase-2 Expression Promotes Glycolysis and Growth in Gastric Cancer Cells. *Mol. Cancer* 12 (1), 110. doi:10.1186/1476-4598-12-110

Li, X., Li, S., Zhang, Z., and Huang, D. (2022). Association of Multiple Tumor Markers with Newly Diagnosed Gastric Cancer Patients: a Retrospective Study. *PeerJ* 10, e13488. doi:10.7717/peerj.13488

Lin, T.-C., and Hsiao, M. (2017). Ghrelin and Cancer Progression. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1868 (1), 51–57. doi:10.1016/j.bbcan.2017.02.002

Liu, A. G., Zhong, J. C., Chen, G., He, R. Q., He, Y. Q., Ma, J., Yang, L. H., Wu, X. J., Huang, J. T., Li, J. J., Mo, W. J., and Qin, X. G. (2020). Upregulated Expression of SAC3D1 Is Associated with Progression in Gastric Cancer. *Int. J. Oncol.* 57 (1), 122–138. doi:10.3892/ijo.2020.5048

Liu, H., Xiang, Y., Zong, Q.-B., Zhang, X.-Y., Wang, Z.-W., Fang, S.-Q., Zhang, T.-C., and Liao, X.-H. (2021). miR-6745-TIMP1 axis Inhibits Cell Growth and Metastasis in Gastric Cancer. *Aging* 13 (21), 24402–24416. doi:10.18632/aging.203688

Liu, M., Li, S., Du, M., Chu, H., Wang, M., and Zhang, Z. (2020). Novel CpG-SNPs in the Gastric Acid Secretion Pathway GNAI3 and Susceptibility to Gastric Cancer. *Gene* 736, 144447. doi:10.1016/j.gene.2020.144447

Manzanedo, I., Pereira, F., Serrano, Á., and Pérez-Viejo, E. (2021). Review of Management and Treatment of Peritoneal Metastases from Gastric Cancer Origin. *J. Gastrointest. Oncol.* 12 (Suppl. 1), S20–S29. doi:10.21037/jgo-20-232

Mou, X., Li, T., Wang, J., Ali, Z., Zhang, Y., Chen, Z., Deng, Y., Li, S., Su, E., Jia, Q., He, N., Ni, J., and Cui, D. (2015). Genetic Variation of BCL2 (Rs2279115), NEIL2 (Rs804270), LTA (Rs909253), PSCA (Rs2294008) and PLCE1 (Rs3765524, Rs10509670) Genes and Their Correlation to Gastric Cancer Risk Based on Universal Tagged Arrays and Fe3O4 Magnetic Nanoparticles. *J. Biomed. Nanotechnol.* 11 (11), 2057–2066. doi:10.1166/jbn.2015.2113

Pan, H., Ding, Y., Jiang, Y., Wang, X., Rao, J., Zhang, X., Yu, H., Hou, Q., and Li, T. (2021). LncRNA LIFR-AS1 Promotes Proliferation and Invasion of Gastric Cancer Cell via miR-29a-3p/COL1A2 axis. *Cancer Cell. Int.* 21 (1), 7. doi:10.1186/s12935-020-01644-7

Qian, X., Xie, F., Wei, H., and Cui, D. (2021). Identification of Key Circulating Exosomal microRNAs in Gastric Cancer. *Front. Oncol.* 11, 693360. doi:10.3389/fonc.2021.693360

Ran, T., Chen, Z., Zhao, L., Ran, W., Fan, J., Hong, S., and Yang, Z. (2021). LAMB1 Is Related to the T Stage and Indicates Poor Prognosis in Gastric Cancer. *Technol. Cancer Res. Treat.* 20, 153303382110049. doi:10.1177/15330338211004944

Ruan, J., Song, H., Qian, Q., Li, C., Wang, K., Bao, C., and Cui, D. (2012). HER2 Monoclonal Antibody Conjugated RNase-A-Associated CdTe Quantum Dots for Targeted Imaging and Therapy of Gastric Cancer. *Biomaterials* 33 (29), 7093–7102. doi:10.1016/j.biomaterials.2012.06.053

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Sheng, Y., Wang, W., Hong, B., Jiang, X., Sun, R., Yan, Q., Zhang, S., Lu, M., Wang, S., Zhang, Z., Lin, W., and Li, Y. (2018). Upregulation of KIF20A Correlates with Poor Prognosis in Gastric Cancer. *Cmar* 10, 6205–6216. doi:10.2147/CMAR.S176147

Stocker, G., Thieme, R., and Lordick, F. (2021). Neoadjuvante und perioperative Therapie des Magenkarzinoms, aktuelle Studien und neue Biomarker. *Chirurg* 92 (6), 499–505. doi:10.1007/s00104-021-01355-3

Sun, K., Hu, P., and Xu, F. (2018). LINC00152/miR-139-5p Regulates Gastric Cancer Cell Aerobic Glycolysis by Targeting PRKAA1. *Biomed. Pharmacother.* 97, 1296–1302. doi:10.1016/j.biopha.2017.11.015

Tao, J., Zhi, X., Tian, Y., Li, Z., Zhu, Y., Wang, W., Xie, K., Tang, J., Zhang, X., Wang, L., and Xu, Z. (2014). CEP55 Contributes to Human Gastric Carcinoma by Regulating Cell Proliferation. *Tumor Biol.* 35 (5), 4389–4399. doi:10.1007/s13277-013-1578-1

Tian, J., Xue, W., Yin, H., Zhang, N., Zhou, J., Long, Z., Wu, C., Liang, Z., Xie, K., Li, S., Li, L., Wu, Z., Daria, V., Zhao, Y., Wang, F., and Wang, M. (2020). Differential Metabolic Alterations and Biomarkers between Gastric Cancer and Colorectal Cancer: A Systematic Review and Meta-Analysis. *Ott* 13, 6093–6108. doi:10.2147/OTT.S247393

Tomii, C., Inokuchi, M., Takagi, Y., Ishikawa, T., Otsuki, S., Uetake, H., Kojima, K., and Kawano, T. (2017). TPX2 Expression Is Associated with Poor Survival in Gastric Cancer. *World J. Surg. Onc* 15 (1), 14. doi:10.1186/s12957-016-1095-y

Toro-Domínguez, D., Martorell-Marugán, J., López-Domínguez, R., García-Moreno, A., González-Rumayor, V., Alarcón-Riquelme, M. E., and Carmona-Sáez, P. (2019). ImaGEO: Integrative Gene Expression Meta-Analysis from GEO Database. *Bioinforma. Oxf. Engl.* 35 (5), 880–882. doi:10.1093/bioinformatics/bty721

Waldum, H. L., and Rehfeld, J. F. (2019). Gastric Cancer and Gastrin: on the Interaction of *Helicobacter pylori* Gastritis and Acid Inhibitory Induced Hypergastrinemia. *Scand. J. gastroenterology* 54 (9), 1118–1123. doi:10.1080/00365521.2019.1663446

Wang, F.-H., Shen, L., Li, J., Zhou, Z.-W., Liang, H., Zhang, X.-T., Tang, L., Xin, Y., Jin, J., Zhang, Y.-J., Yuan, X.-L., Liu, T.-S., Li, G.-X., Wu, Q., Xu, H.-M., Ji, J.-F., Li, Y.-F., Wang, X., Yu, S., Liu, H., Guan, W.-L., and Xu, R.-H. (2019). The Chinese Society of Clinical Oncology (CSCO): Clinical Guidelines for the Diagnosis and Treatment of Gastric Cancer. *Cancer Commun.* 39 (1), 10. doi:10.1186/s40880-019-0349-9

Wang, K., Ruan, J., Qian, Q., Song, H., Bao, C., Zhang, X., Kong, Y., Zhang, C., Hu, G., Ni, J., and Cui, D. (2011). BRCAA1 Monoclonal Antibody Conjugated Fluorescent Magnetic Nanoparticles for *In Vivo* Targeted Magnetofluorescent Imaging of Gastric Cancer. *J. Nanobiotechnol* 9, 23. doi:10.1186/1477-3155-9-23

Wang, Y., Zheng, K., Chen, X., Chen, R., and Zou, Y. (2021). Bioinformatics Analysis Identifies COL1A1, THBS2 and SPP1 as Potential Predictors of Patient Prognosis and Immunotherapy Response in Gastric Cancer. *Biosci. Rep.* 41 (1), BSR20202564. doi:10.1042/BSR20202564

Wu, K., Zou, J., Lin, C., and Jie, Z.-G. (2019). MicroRNA-140-5p Inhibits Cell Proliferation, Migration and Promotes Cell Apoptosis in Gastric Cancer through the Negative Regulation of THY1-Mediated Notch Signaling. *Biosci. Rep.* 39 (7), BSR20181434. doi:10.1042/BSR20181434

Xie, Y., Zhi, X., Su, H., Wang, K., Yan, Z., He, N., Zhang, J., Chen, D., and Cui, D. (2015). A Novel Electrochemical Microfluidic Chip Combined with Multiple Biomarkers for Early Diagnosis of Gastric Cancer. *Nanoscale Res. Lett.* 10 (1), 477. doi:10.1186/s11671-015-1153-3

Xu, J., Wang, X., Ke, Q., Liao, K., Wan, Y., Zhang, K., Zhang, G., and Wang, X. (2021). Combined Bioinformatics Technology to Explore Pivot Genes and

Related Clinical Prognosis in the Development of Gastric Cancer. *Sci. Rep.* 11 (1), 15412. doi:10.1038/s41598-021-94291-5

Xu, L., Chen, J., Jia, L., Chen, X., Awaleh Moumin, F., and Cai, J. (2020). SLC1A3 Promotes Gastric Cancer Progression via the PI3K/AKT Signalling Pathway. *J. Cell. Mol. Med.* 24 (24), 14392–14404. doi:10.1111/jcmm.16060

Yang, C., and Gong, A. (2021). Integrated Bioinformatics Analysis for Differentially Expressed Genes and Signaling Pathways Identification in Gastric Cancer. *Int. J. Med. Sci.* 18 (3), 792–800. doi:10.7150/ijms.47339

Zhang, J., Song, Y., Zhang, C., Zhi, X., Fu, H., Ma, Y., Chen, Y., Pan, F., Wang, K., Ni, J., Jin, W., He, X., Su, H., and Cui, D. (2015). Circulating MiR-16-5p and MiR-19b-3p as Two Novel Potential Biomarkers to Indicate Progression of Gastric Cancer. *Theranostics* 5 (7), 733–745. doi:10.7150/thno.10305

Zhang, Y., Qin, L., Ma, X., Wang, Y., Wu, Y., and Jiang, J. (2020). Coexpression of Matrix Metalloproteinase-7 and Tissue Inhibitor of Metalloproteinase-1 as a Prognostic Biomarker in Gastric Cancer. *Dis. markers* 2020 (10), 1. doi:10.1155/2020/8831466

Zhao, S., Bi, Y., Wang, Z., Zhang, F., Zhang, Y., and Xu, Y. (2022). Accuracy Evaluation of Combining Gastroscopy, Multi-Slice Spiral CT, Her-2, and Tumor Markers in Gastric Cancer Staging Diagnosis. *World J. Surg. Onc* 20 (1), 152. doi:10.1186/s12957-022-02616-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yu, Liang, Tu, Qiu, Dong, Zhang, Ma and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fgene.2022.866696

Check for updates

OPEN ACCESS

EDITED BY
Xiu-Wu Pan,
Shanghai Jiao Tong University, China

REVIEWED BY
Zhenpeng Zhu,
First Hospital, Peking University, China
Yan Kang Cui,
Jinling Hospital, China

*CORRESPONDENCE
Mei-Fu Gan,
ganmf@enzemed.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 31 January 2022
ACCEPTED 12 July 2022
PUBLISHED 11 August 2022

CITATION
Yu Q, Zhang F, Feng D, Li D, Xia Y and Gan M-F (2022), An inflammation-related signature could predict the prognosis of patients with kidney renal clear cell carcinoma.
Front. Genet. 13:866696.
doi: 10.3389/fgene.2022.866696

COPYRIGHT
© 2022 Yu, Zhang, Feng, Li, Xia and Gan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An inflammation-related signature could predict the prognosis of patients with kidney renal clear cell carcinoma

Qingxin Yu[1†], Facai Zhang[2†], Dechao Feng[2], Dengxiong Li[2], Yuhui Xia[1] and Mei-Fu Gan[1*]

[1]Department of Pathology, Taizhou Hospital, Wenzhou Medical University, Linhai, China, [2]Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, China

Background: Kidney renal clear cell carcinoma (KIRC) is an inflammation-related carcinoma, and inflammation has been recognized as an important factor in inducing carcinogenesis. To further explore the role of inflammation in KIRC, we developed an inflammation-related signature and verified its correlation with the tumor micro-environment.

Methods: After the differential inflammation-related prognostic genes were screened by Lasso regression, the inflammation-related signature (IRS) was constructed based on the risk score of multivariate Cox regression. Then, the prognostic value of the IRS was evaluated by Kaplan-Meier analysis, receiver operating characteristic (ROC) curve analysis and multivariate Cox regression. Gene set variation analysis (GSVA) was applied to screen out enriched signaling pathways. Infiltrated immune cells, tumor mutational burden (TMB) and immune checkpoints were explored by CIBERSORTx and maftool.

Results: Four genes (TIMP1, PLAUR, CCL22, and IL15RA) were used to construct the IRS in patients with KIRC. Kaplan-Meier analysis and multivariate Cox regression identified that the IRS could independently predict the prognosis of patients with KIRC in the training and validation groups. The diagnostic value of the nomogram increased from 0.811 to 0.845 after adding the IRS to the multiparameter ROC analysis. The GSVA results indicated that IRS was closely related to primary immunodeficiency and antigen processing and presentation. The immune checkpoint LAG3 was highly expressed in patients with high-risk score ($p < 0.05$), while CD274 (PD-L1) and HAVCR2 were highly expressed in patients with low-risk score ($p < 0.001$). There was a significant positive correlation between the high-risk score group and $CD8^+$ T, activated $CD4^+$ memory T, gamma and delta regulatory T and M0 macrophage cells, while the low-risk score group was negatively associated with B memory, plasma, resting $CD4^+$ memory T, activated NK, M1 macrophages and resting mast cells.

Conclusion: We found that the IRS might serve as a biomarker to predict the survival of KIRC. Moreover, patients with high or low-risk score might be sensitive to immune drugs at different immune checkpoints.

KEYWORDS

inflammation, renal clear cell carcinoma, immune infiltration, signature, immune checkpoint

Introduction

Worldwide, there were 431,288 newly diagnosed kidney cancers in 2020, which represented approximately 2.2% of all cancers (Sung et al., 2021). In United States, according to the data of 2022, the incidence of kidney cancer occupied sixth and ninth in the new cancers of male and female, respectively (Siegel et al., 2022). Of these, renal cell carcinoma (RCC) accounts for approximately 90% of all kidney malignancies. The main pathological type of RCC is kidney renal clear cell carcinoma (KIRC), which accounts for 70%–80% of cases (Nerich et al., 2014). Partial and radical nephrectomy is the optimal therapeutic choice for located KIRC. Unfortunately, approximately one-third of patients with localized RCC inevitably develop metastases, which need systemic treatment to control the disease (Escudier et al., 2019). Given the poor therapeutic outcome, chemotherapy was used to cure patients with KIRC. In recent years, to further improve the prognosis of KIRC, immunotherapy has been applied clinically (Escudier et al., 2019). However, few biomarkers can precisely predict the prognosis and therapeutic outcome of KIRC, which hinders the personalized application of these therapies and the creation of new drugs. This dilemma prompted us to explore the potential mechanism of the occurrence and progression of KIRC.

Chronic inflammation has been recognized as an important factor for carcinogenesis by inducing oxidative and nitrative DNA damage (Ohnishi et al., 2013). Moreover, inflammatory cells are a major component of the tumor microenvironment and an indispensable factor in promoting tumor proliferation, neoplastic processes, survival, and migration (Okada et al., 2021). Tumor-associated inflammation has been widely studied and listed as a hallmark of cancers (Hanahan, 2022). Moreover, Zhao et al. (2019) identified five immune- and inflammation-related core clusters by integrating multiomics data to identify the role of immunity and inflammation in KIRC. To further explore the role of inflammation in KIRC, Marona et al. (2017) identified that low MCPIP1 levels could increase proliferation, tumor outgrowth, and vascularity by upregulating inflammation by degrading mRNAs encoding proinflammatory cytokines in KIRC. Clinically, an included 10-study meta-analysis identified that the systemic immune-inflammation index could independently predict survival outcomes in patients with renal cell carcinoma. KIRC is an inflammation-related carcinoma, which inspired us to clarify the role of inflammation in KIRC. And these evidences suggest that we should construct an inflammation-related signature (IRS) and validate whether it can be used as a potential biomarker for KIRC.

Therefore, this study first constructed an IRS for patients with KIRC and validated the prognostic value of this signature. Furthermore, we explored the correlation between the IRS and immunotherapy by evaluating the tumor mutational burden, immune checkpoint and immune cell infiltration in KIRC.

Materials and methods

Data collection

The clinical and RNA_seq data of KIRC were downloaded from Cancer Genome Atlas (www.gdc.cancer.gov, TCGA) (Wei et al., 2018) database. After excluding patients with postoperative survival times shorter than 30 days, the remaining patients were randomly divided into the TCGA training group (n = 364) and TCGA test groups (n = 156). Then, the TCGA training group was used to develop the IRS, while the TCGA test group was employed to validate the signature. GSE29609 (Edeline et al., 2012) was extracted from the Gene Expression Omnibus (Edgar et al., 2002) to further estimate the diagnostic and prognostic value of the IRS. Before analysis these clinical and RNA_seq data, the removeBatchEffect function in limma package was used to reduce the batch effects of the TCGA and GEO datasets.

Identification of differentially expressed inflammatory genes and biological functional analysis

To identify the differentially expressed genes, the mRNA data of 520 KIRC samples and 72 normal samples were compared by the "limma" package with False Discover 99 Rate (FDR) < 0.05. The inflammation-related genes were provided by hallmark genes of the Molecular Signature Database (Subramanian et al., 2005). In detail, the inflammation-related gene set was generated by a computational methodology based on identifying gene set overlaps and retaining genes that display coordinate expression. Moreover, the Molecular Signatures Databases team provided the microarray data that served for refining and validation of the inflammation-related gene set online (Liberzon et al., 2015). Furthermore, the "Venn Diagram" package in R software was used to screen out the co-expressed inflammation-related genes in the TCGA dataset and GSE29609.

To explore possible biological functions and signaling pathways, the "cluster Profiler" package in R software was used to analyze all differentially expressed genes. In Gene Ontology (GO) enrichment analysis, the results with

p-value < 0.05 and q-value < 0.05 were collected and visualized in bar plots based on molecular function (MF), biological process (BP) and cellular component (CC) categories. Similarly, according to the results of Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, the signaling pathways with p-value < 0.05 and q-value < 0.05 were included and visualized in a bubble plot. Based on a linear regression-based algorithm and a label propagation algorithm, GeneMania (www.genemania.org) (Warde-Farley et al., 2010) was employed to predict interacting genes and generate a visual figure.

Development and validation of an inflammation-related signature

To identify a stable and predictive inflammation-related signature, a lasso regression model was employed to evaluate all differentially expressed inflammation-related genes. In this process, the penalty parameter (λ) of the model was controlled by 10-fold cross-validation. Referring to the lasso results, the IRS was constructed based on the selected genes. Furthermore, risk score was generated in a multivariate Cox regression model with the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} coefficient_i * EXP(mRNA)_i \quad (1)$$

All patients were divided into high- and low-risk groups according to the median risk scores. To validate the prognostic value of the IRS, Kaplan-Meier analysis was performed to compare the survival time and outcomes between the high- and low-risk score KIRC patients in the TCGA training group, TCGA test group and GSE29609 group.

Correlation of the inflammatory signature with clinical parameters

To further validate the relationship between the IRS and clinical parameters, subgroups were generated by classifying age, sex, T stage, metastasis, and AJCC stage and comparing the different risk score in these subgroups. To assess the discernibility of the IRS, we performed principal component analysis and depicted the risk score plot depending on the risk score and survival outcome. The Wilcoxon rank-sum test was used to compare the risk score in each subgroup. Kaplan-Meier analysis was performed to estimate the prognosis of different risk score in each subgroup. Receiver operating characteristic (ROC) curve analysis was used to assess the diagnostic ability of the IRS in the TCGA training, TCGA test and GSE29609 groups. Furthermore, the independent prognostic ability of the IRS and clinical parameters was evaluated by a multivariate Cox regression model.

Gene set variation analysis, tumor mutational burden analysis, and immune cell infiltration

Gene Set Variation Analysis (GSVA) (Hänzelmann et al., 2013) was performed by the "GSVA" package in R. The GSVA results were generated by comparing the high- and low-risk score groups and visualized by heatmap. We uploaded the transcriptome data of high- and low-risk score patients in TCGA training and a collection of KEGG pathways from the Molecular Signature Database (MSigDB, version 7.4) (Liberzon et al., 2015) for GSVA. For the estimation of tumor mutational burden (TMB) in KIRC, tumor mutational burden analysis used the "maftools" package to calculate the mutation rate of each sample downloaded from the TCGA database. After calculating the mutation rate, the samples were divided into high and low TMB groups according to the TMB score, and the survival time and outcome between the groups were compared by Kaplan-Meier analysis.

TMB and immune checkpoint-associated mismatched repair genes in tumor tissue were considered potential biomarkers for predicting immunotherapy response. Thus, we compared the transcriptome data of CD274 (PD-L1), PDCD1LG2 (PD-L2), CTLA4, HAVCR2, LAG3, PDCD1, TIGIT, and SIGLEC15 in both risk score groups. Furthermore, after normalization of transcriptome gene expression data of KIRC patients with the "limma" package, the CIBERSORT algorithm was utilized to evaluate the immune infiltration of 22 leukocyte subtypes (LM22), which were downloaded from the known reference set on the CIBERSORTx website (Newman et al., 2015). The infiltration difference between the high- and low-risk score groups was calculated with the Wilcoxon rank-sum test, and a boxplot was used for visualization. Meanwhile, Tumor Immune Dysfunction and Exclusion (TIDE) algorithm, acquiring data from the TCGA-KIRC cohort, was used to predict and compare the immunotherapy response of patients in high- or low-risk score group (Jiang et al., 2018; Fu et al., 2020).

Development and validation of a predictive nomogram

To quantitatively predict the prognosis of KIRC patients, a nomogram was carefully established based on the risk score and clinical parameters, including age, M stage and AJCC stage. Of these, patients over 60 years old were divided into elderly group and the remaining young group. For the purpose of validation, the concordance index (C-index), multiparameter ROC analysis, decision curve analysis (DCA) and calibration curves were used to validate the reliability and accuracy of the nomogram.

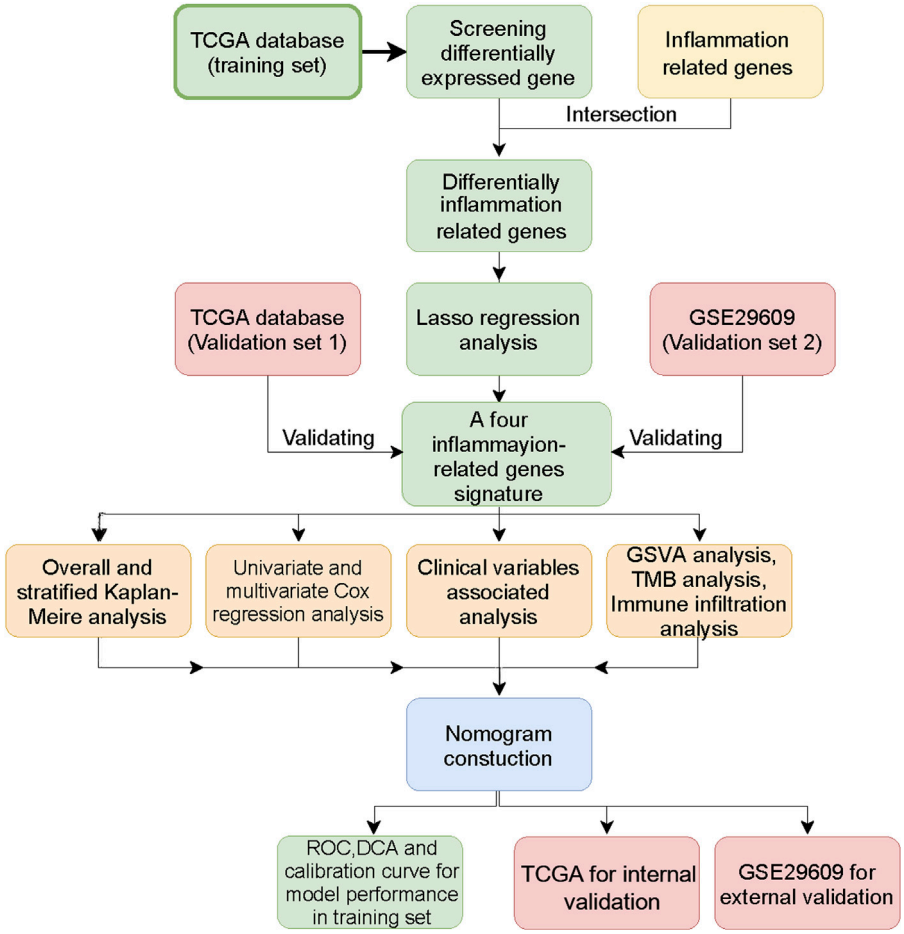


FIGURE 1
The workflow of this study.

Statistical analysis

All analyses were carried out using R software, version 4.1.2. Quantitative data in two groups were compared using Student's *t* test. Quantitative data were compared with one-way analysis of variance (ANOVA) or Welch's test in three or more groups. $P < 0.05$ was regarded as statistically significant.

Results

Patient characteristics, co-expressed inflammation genes and biological functional analysis

Figure 1 shows the workflow of our work. Table 1 presents the clinicopathologic features of all included patients. There were 364, 156, and 39 patients included in the TCGA training, TCGA test and GSE29609 groups, respectively. Figure 2A provides the survival outcomes of the TCGA-train, TCGA-test and GSE29609 groups, and no significant difference existed among these three groups. A total of 15,453 differentially expressed genes were collected with the absolute value of the log2-transformed fold change (FC) > 1 and the adjusted *p* value (adj. *p*) < 0.05 was used as the threshold after normalization and batch effect removal. Then, 108 co-expressed inflammatory genes were selected by Venn Diagram (Figure 2B). Further GO analysis showed that the co-expressed inflammatory genes were positively correlated with cytokine-mediated signaling pathways, leukocyte migration, the external side of plasma and immune receptor activity (Figure 2C). The results of KEGG analysis showed that the co-expressed inflammatory genes were positively associated with cytokine–cytokine receptor interactions (Figure 2D).

TABLE 1 Clinicopathologic characteristics of the included patients.

Characteristic	GSE29609	TCGA test	TCGA train
n	39	156	364
Age, n (%)			
Older	17 (3%)	50 (8.9%)	127 (22.7%)
Young	22 (3.9%)	106 (19%)	237 (42.4%)
Sex, n (%)			
female	NA	60 (11.5%)	117 (22.5%)
Male	NA	96 (18.5%)	247 (47.5%)
T stage, n (%)			
1	11 (2%)	74 (13.2%)	193 (34.5%)
2	5 (0.9%)	25 (4.5%)	45 (8.1%)
3	22 (3.9%)	54 (9.7%)	118 (21.1%)
4	1 (0.2%)	3 (0.5%)	8 (1.4%)
M stage, n (%)			
M0	25 (4.7%)	115 (21.5%)	302 (56.6%)
M1	14 (2.6%)	30 (5.6%)	48 (9%)
AJCC stage, n (%)			
Stage I	15 (2.7%)	72 (12.9%)	189 (33.9%)
Stage II	3 (0.5%)	21 (3.8%)	37 (6.6%)
Stage III	13 (2.3%)	32 (5.7%)	85 (15.3%)
Stage IV	8 (1.4%)	30 (5.4%)	52 (9.3%)
Survival status, n (%)			
Alive	22 (3.9%)	99 (17.7%)	250 (44.7%)
Dead	17 (3%)	57 (10.2%)	114 (20.4%)

TCGA, cancer genome atlas; n, number; NA, no data; M, metastasis; AJCC, american joint committee on Cancer.

As shown in Figure 2E, the lasso regression model selected five inflammatory genes (CX3CL1, TIMP1, PLAUR, CCL22, and IL15RA). To assess the independent prognostic value of these genes, a multivariate Cox regression model was used to calculate the five genes. The results illustrated that TIMP1, PLAUR, CCL22, and IL15RA could independently predict the prognosis of patients with KIRC ($p < 0.05$). According to the GeneMANIA results, we found that TIMP1, PLAUR, CCL22, and IL15RA proteins were associated with some proteins, such as CCL14, PF4V1, ACKR2, and PGAP1 (Figure 2F).

Establishment and validation of the inflammation-related gene signature

Referring to the results of the multivariate Cox regression model above, TIMP1, PLAUR, CCL22 and IL15RA were used to construct the IRS. Patients in the three groups were divided into the high-risk score group and the low-risk score group according to the median risk score of the TCGA training group, TCGA test group and GSE29609 group. As shown in Figures 3A–C, the risk scores clearly differed between the high- and low-risk score groups in the TCGA training group, TCGA test group and GSE29609 group. There were significant differences in overall survival (OS) between the high- and low-risk score groups in the TCGA training group (Figure 3D, $p < 0.001$). Similarly, patients in the high-risk score group of TCGA-test group and GSE29609 group were correlated with worse OS than those in the low-risk group, even GSE29609 provided only 39 patients (Figures 3E,F, $p = 0.002$, $p = 0.049$, respectively). Furthermore, a positive correlation was found between the increased IRS score and the increase in mortality in the TCGA training group, TCGA test group and GSE29609 group (Figures 3G–I).

Correlation between the inflammation-related signature and clinical parameters

As shown in Figures 4A–C, there were significant differences in the risk scores at T stage, M stage and AJCC stage in the TCGA training group (all $p < 0.001$). Similarly, in the TCGA-test group, patients with T3-4 or M1 or AJCC stage III-IV were associated with higher risk scores than those with low stage (Figures 4D–F, all $p < 0.01$). In the GSE29609 group, patients with T3-4 or AJCC stage III-IV were correlated with higher risk scores than patients with low stage, but not M1 (Figures 4G–I).

Prior to analyzing the prognosis of the subgroups, we reported that the IRS score could predict the prognosis of the TCGA training group, TCGA test group and GSE29609 group in the previous section. In the subgroup analysis, high-risk score patients in the TCGA training group were associated with worse OS than those with low-risk scores, including females (Figure 5A, $p < 0.001$), males (Figure 5B, $p < 0.001$), T1-2 (Figure 5C, $p = 0.003$), M0 (Figure 5D, $p < 0.001$), and AJCC stage I-II (Figure 5E, $p = 0.007$). The results of the subgroup analysis in the TCGA-test group were similar to those in the TCGA-training group; patients with high-risk scores were correlated with shorter OS than patients with low-risk scores, including females (Figure 5F, $p = 0.019$), males (Figure 5G, $p = 0.048$), T1-2 (Figure 5H, $p = 0.003$), M0 (Figure 5I, $p = 0.005$), and AJCC stage I-II (Figure 5J, $p = 0.01$). We did not perform Kaplan-Meier analysis in GSE29609 because of the limited number of patients. Figure 5K illustrates that age, M stage, AJCC stage and IRS score were independent predictors of patients in the TCGA training group. To further validate the predictive value of the IRS, we performed a multivariate Cox regression model in the TCGA-test group and identified that M1 stage and high-risk scores were independently associated with worse OS (Figure 5L). As Figure 5M shows, the areas under the curve (AUCs) of the IRS score in the TCGA training group were 0.809, 0.708 and 0.720 at 1, 3, and 6 years, respectively. In the TCGA-test group, the AUC of the IRS score was 0.708, 0.654, and 0.685 at 1, 3, and 6 years, respectively (Figure 5N). Due to the limited number of

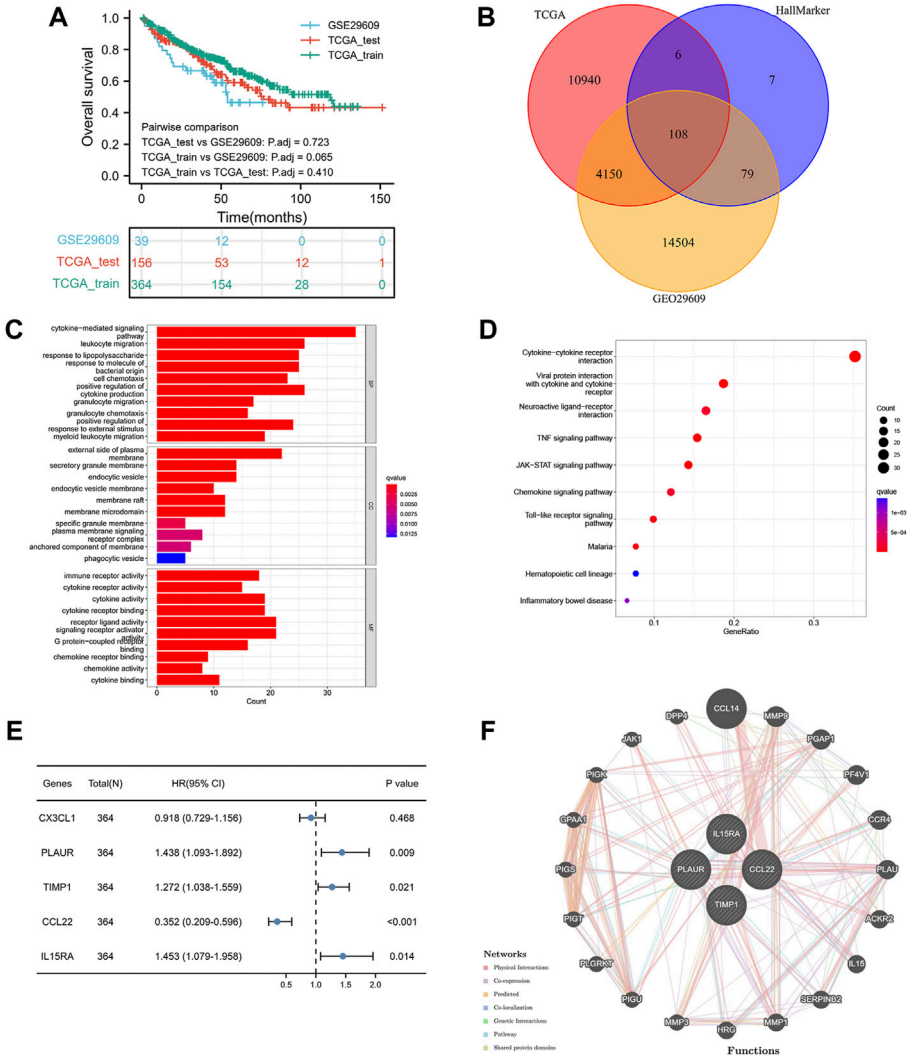


FIGURE 2
Construction of inflammation-related signature: the overall survival of included patients **(A)**, the Venn diagram of differentially coexpressed genes **(B)**, the GO result **(C)**, the KEGG result **(D)**, the results of five genes in multivariate Cox regression model **(E)**, the result of protein–protein interaction network **(F)**.

patients in GSE29609, we only calculated the AUCs at 1, 3, and 4 years, which were 0.527, 0.636, and 0.826, respectively (Figure 5O).

Gene set variation analysis, tumor mutational burden analysis, and immune cell infiltration

As shown in Figure 6A, the enrichment pathways included primary immunodeficiency, antigen processing and presentation and systemic lupus erythematosus, which were mainly correlated with inflammation and immune pathways.

The waterfall presented the top 15 most mutated genes, VHL and PBRM1, making up the majority of the mutations (Figure 6B). Of these, the majority of mutations were nonsense mutations, missense mutations and frame shift del. To further assess the prognostic value of the TMB score, all patients with TMB scores were divided into four groups according to the TMB score and IRS score. The TMB in the high-risk-score group was significantly higher than that in the low-risk-score group in the TCGA training and TCGA test groups (Figures 6C,D). The OS of these four groups was analyzed by Kaplan-Meier analysis, which showed that patients with low-risk scores and low TMB scores were associated with a better OS than patients in other groups

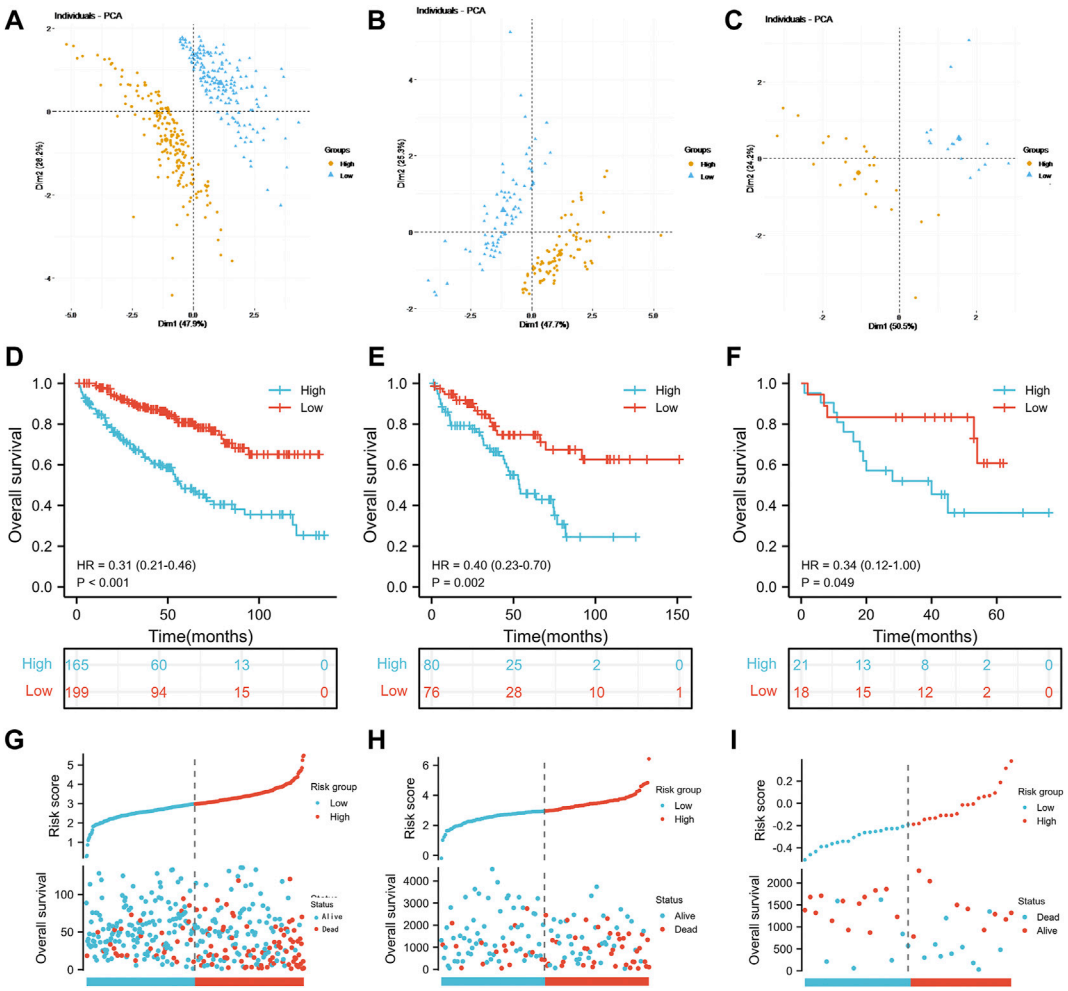


FIGURE 3
Prognostic performance of the inflammation-related signature: Principal component analysis results of the TCGA training group **(A)**, TCGA test **(B)** and GSE29609 **(C)**. Kaplan-Meier analysis results of the TCGA training group **(D)**, TCGA test **(E)** and GSE29609 **(F)**. The risk score plots of TCGA training group **(G)**, TCGA test **(H)** and GSE29609 **(I)**.

(Figure 6E). Interestingly, LAG3 expression in the high-risk score group was higher than that in the low-risk score group, while the expression of mismatch repair genes CD274 (PD-L1) and HAVCR2 in the high-risk score group was lower than that in the low-risk score group (Figure 6F). The Wilcoxon rank-sum test displayed a remarkable discrepancy between the high- and low-risk score groups in 22 immune cell types. There was a significant positive correlation between the high-risk score group and $CD8^+$ T, activated $CD4^+$ memory T, gamma and delta regulatory T and M0 macrophage cells, while the low-risk score group was negatively associated with B memory, plasma, resting $CD4^+$ memory T, activated NK, M1 macrophages and resting mast cells (Figure 6G). Referring the results of TIDE, patients with low-risk score in TCGA training group and test group had higher immunotherapy response rate than patients with high-risk score (Figures 6H–K). Similarly, patients with low-risk score were associated with lower TIDE score than patients with low-risk score (Figures 6L,M).

Development and validation of a predictive nomogram

The correlation between the prognosis of KIRC, the IRS score and clinical parameters was tested using a multivariate Cox regression model based on the TCGA training group in the previous section (Figure 5K). The parameters with $p < 0.05$ were selected for further analysis according to the results of the multivariate Cox regression model. To extend the clinical applicability of the IRS, a nomogram containing parameters with age, AJCC stage, IRS score and M stage was constructed

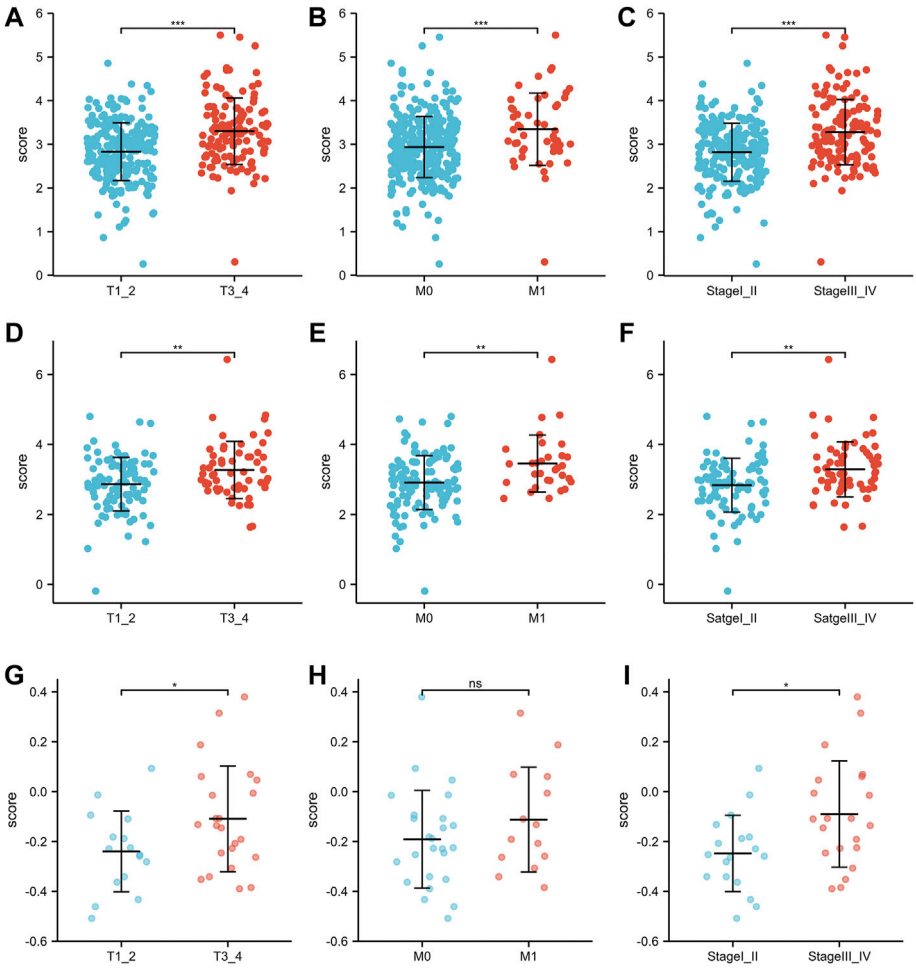


FIGURE 4
The risk scores of patients in different clinicopathological parameters: TCGA training group: T stage **(A)**, M stage **(B)**, AJCC stage **(C)**. TCGA-test group: T stage **(D)**, M stage **(E)**, AJCC stage **(F)**. GSE29609: T stage **(G)**, M stage **(H)**, AJCC stage **(I)**.

based on the TCGA training group (Figure 7A). The concordance index of this module was 0.784 (0.762–0.805). The calibration curves at 1, 3, and 6 years showed that there was good agreement between the predicted value and the true value (Figure 7B). Interestingly, the consistency between the predicted value and the true value increased with time. Furthermore, compared with clinical parameters, the IRS score exhibited potential clinical value (Figure 7C). This nomogram was superior to either the "all positive" or "all negative" model in predicting the prognosis of KIRC (Figure 7D). In detail, module with IRS score was associated with better DCA curve than module without IRS score. To further validate the prognostic value of the IRS, we compared the prognostic value of each factor which built the nomogram. The Figure 7E shown that the IRS score had highest AUC value 0.726. Multiparameter ROC analysis was also used to calculate the AUC of the nomogram with or without the IRS score. As shown in Figures 7F,G, the AUC of the nomogram was obviously increased from 0.811 to 0.845 after adding the IRS score into the multiparameter ROC analysis, which identified the prognostic value of the IRS score.

Discussion

KIRC is an inflammation-related carcinoma with a poor therapeutic outcome. Although surgery can improve the prognosis of patients with early-stage KIRC, 20%–40% of patients will still experience recurrence (Liang, 2020). Furthermore, a previous study reported that cancer cell-intrinsic inflammation can facilitate both KIRC metastasis and the initial progression of KIRC (Nishida et al., 2020). To improve the prognosis of patients with KIRC, many inflammatory genes and cytokines were studied to assess their prognostic and curative

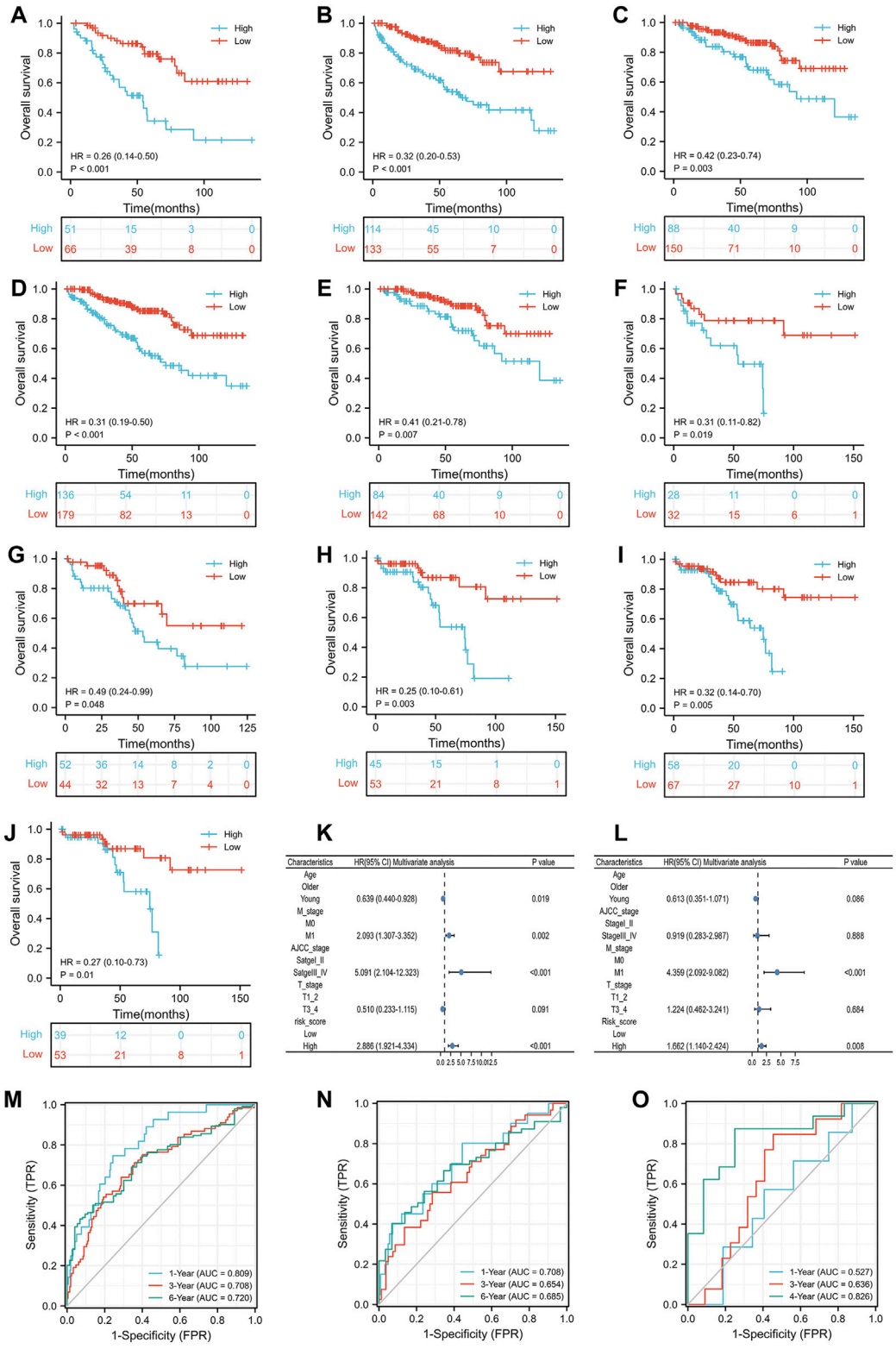


FIGURE 5
The prognostic ability of the risk scores: Kaplan-Meier analysis results of subgroups in TCGA training group: female **(A)**, male **(B)**, T1_2 **(C)**, M0 **(D)** and AJCC stage I-II **(E)**; Kaplan-Meier analysis results of subgroups in TCGA test group: female **(F)**, male **(G)**, T1_2 **(H)**, M0 **(I)** and AJCC stage I_II **(J)**; Results of multivariate Cox regression model for parameters in TCGA training group **(K)**; Results of multivariate Cox regression model for parameters in TCGA test group **(L)**; Receiver operating characteristic curve analysis results of TCGA training group **(M)**, TCGA test group **(N)** and GSE29609 **(O)**.

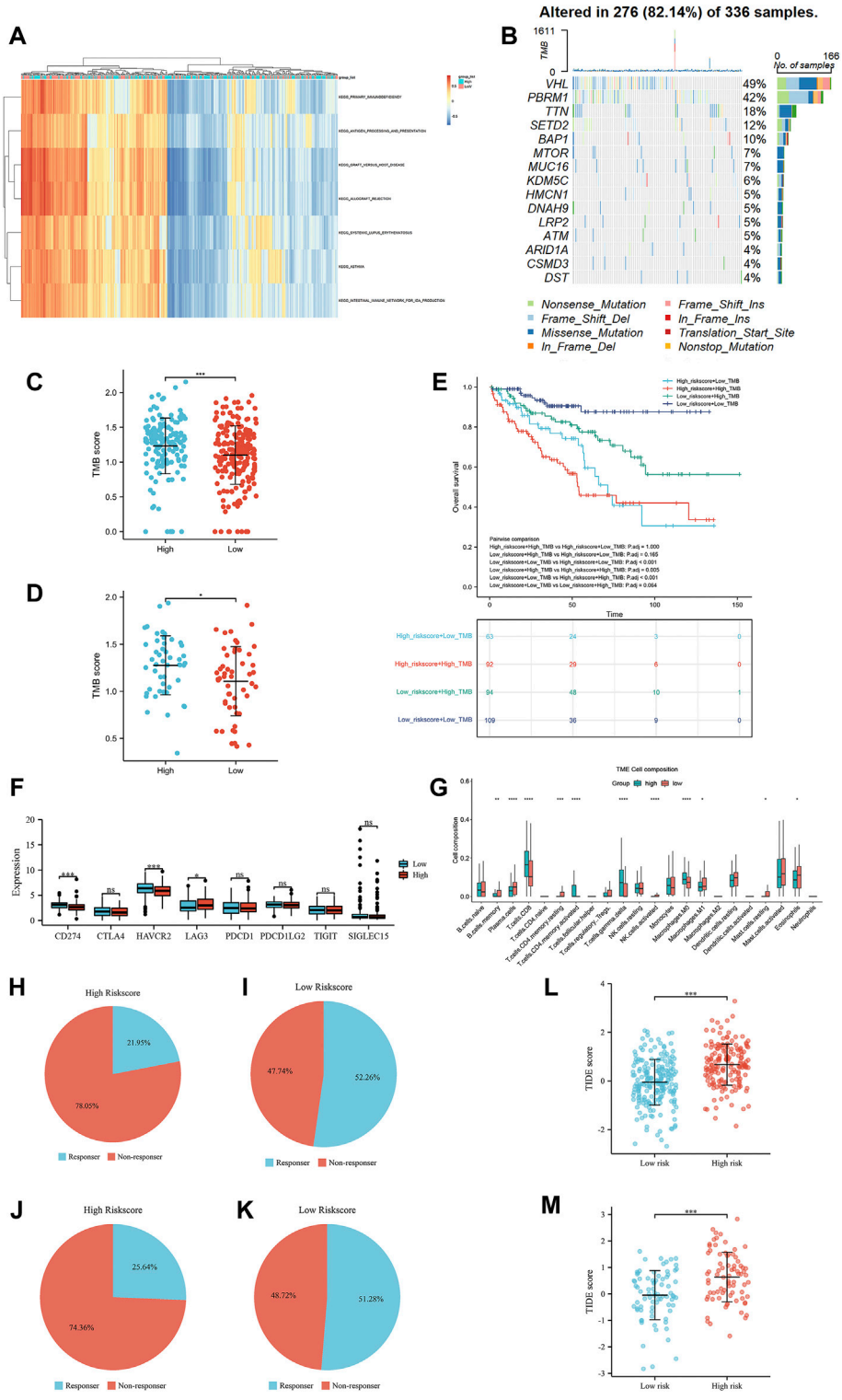


FIGURE 6
The results of gene set variation analysis **(A)**, the results of tumor mutational burden analysis **(B)**, the correlation of tumor mutational burden scores and the risk scores in the TCGA training group **(C)**, the correlation of tumor mutational burden scores and the risk scores in the TCGA test group **(D)**, the OS of the high-risk group compared with that in the low-risk group in the TCGA training group **(E)**, the expression of immune checkpoints in the high-risk group compared with that in the low-risk group in the TCGA training group **(F)**, and the immune cell infiltration in the high-risk group compared with that in the low-risk group in the TCGA training group **(G)**. The TIDE immunotherapy response outcome of high-risk score group **(H)** and low-risk score group **(I)** in the TCGA training group, high-risk score group **(J)** and low-risk score group **(K)** in the TCGA test group. The TIDE score of the TCGA training group **(L)** and the TCGA test group **(M)**.

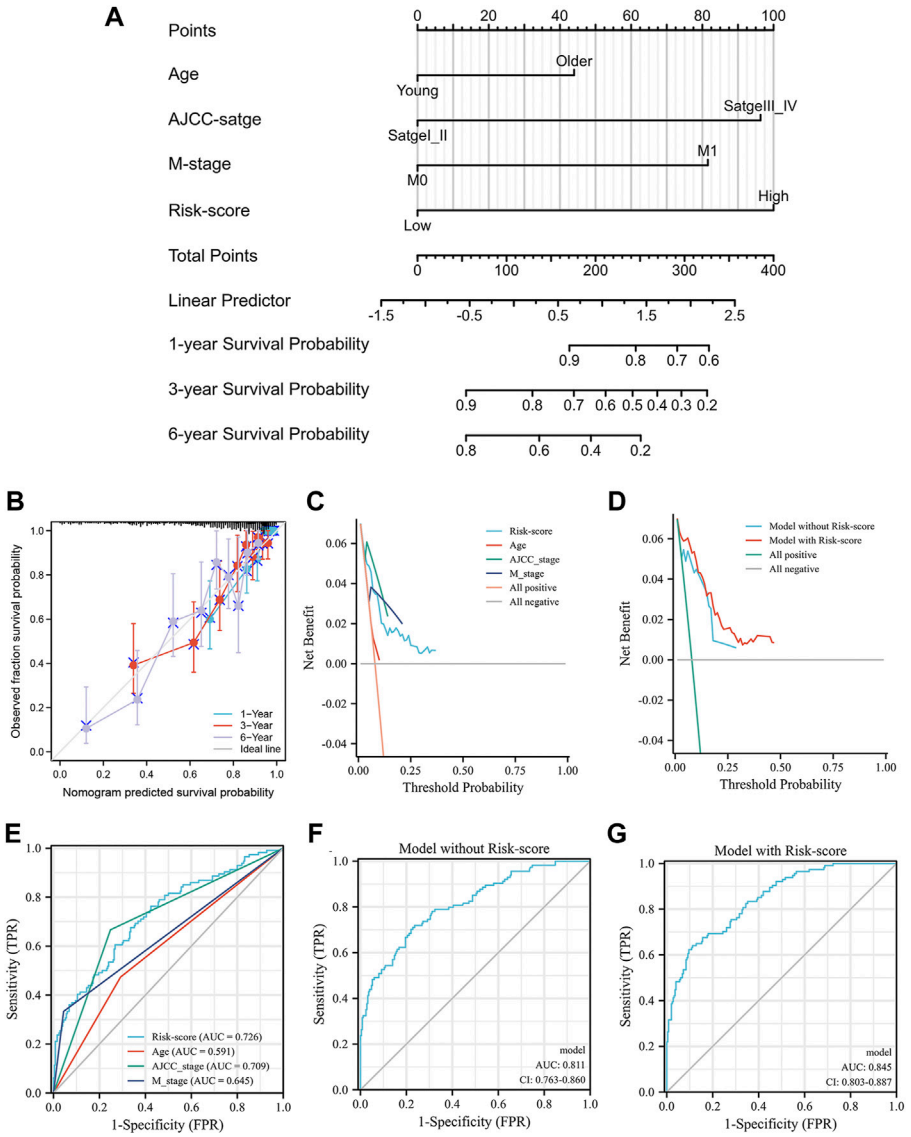


FIGURE 7
The nomogram of the inflammation-related gene signature and its performance: the nomogram **(A)**, the calibration curves of the nomogram **(B)**, the calibration curves of different factors **(C)** and the nomogram **(D)**, the result of ROC analysis with different factors **(E)**, the result of multiparameter ROC analysis without the risk scores **(F)**, the result of multiparameter ROC analysis with the risk scores **(G)**.

value. For example, Yang et al. (2021) identified interferon-induced transmembrane protein 2, an inflammation-related gene, as being associated with lymphatic metastasis and poor clinical outcome of KIRC. Another example of this was interleukin-6 (IL-6), an inflammation-related cytokine that could be an independent early-stage immunologic prognostic factor for KIRC patients. The significant role of inflammation in KIRC enlightened us to construct an IRS to enhance the treatment schedule and prognosis of patients with KIRC. The current study selected four differentially expressed inflammation-related genes (TIMP1, PLAUR, CCL22, and IL15RA) and created an IRS based on the four genes. In further analysis, the AUC of the IRS was 0.809, 0.708, and 0.720 at 1, 3, and 6 years, respectively. Furthermore, the multiparameter ROC analysis illustrated that the AUC increased from 0.785 to 0.829 after adding the IRS score to the analysis. Similarly, the prognostic value was identified in the validation groups, suggesting that the IRS could be considered a potential biomarker for KIRC.

There is a close correlation between inflammation-related genes and the occurrence and progression of KIRC. Therefore, we would like to introduce the role of the four genes in KIRC.

TIMP1, a key tissue inhibitor of metalloproteinase that regulates most matrix metalloproteinases, was usually found to be increased in renal cell carcinoma (RCC) and could affect the efficiency of radiotherapy (Smyth et al., 2007). Moreover, TIMP1 could promote the invasion and metastasis of RCC and predict the prognosis of RCC (Peña et al., 2010; Lu et al., 2014; Feng et al., 2019). PLAUR belongs to the plasminogen activation system and is widely involved in various cancer-specific processes, including inflammation- and immune- and hypoxia-related pathways (Liu et al., 2021). In addition, several studies have identified that PLAUR is associated with the prognosis of patients with RCC using an online database (Shen et al., 2020; Li et al., 2021). Jin et al. (2019) reported that CCL22 was overexpressed in RCC and could promote the progression and metastasis of RCC by downregulating miR-34a-5p. Furthermore, a study found that CCR4, the chemokine receptor for CCL2 and expressed on T cells, could be considered a therapeutic target for cancer immunotherapy (Yoshie, 2021). Interestingly, in the current study, the infiltration of T cells, including $CD8^+$ T cells, gamma and delta regulatory T cells and activated $CD4^+$ memory T cells, was higher in the high-risk score group, which suggested that the high-risk score group might be sensitive to immunotherapy. Although few studies have reported IL15RA in RCC, it could predict the prognosis of some cancers and play a protective role in the progression and treatment of some cancers, including colorectal carcinoma, breast carcinoma and multiple myeloma (Marra et al., 2014; Borrelli et al., 2018; Yang et al., 2019; De Mattia et al., 2021). The mechanism of this phenomenon might be caused by upregulation of IL15RA, which could induce the proliferation and activation of NK cells and activate peripheral blood mononuclear cells upon coculture in a paracrine signaling manner (Marra et al., 2014; Borrelli et al., 2018). In the present study, NK cell infiltration in the low-risk-score group was significantly higher than that in the high-risk-score group, which might be one reason why patients in the low-risk-score group had better OS.

To identify the correlation between the IRS and clinicopathological features, we compared the IRS score in different subgroups. The results indicated that high risk scores were associated with T3-4 stage, AJCC stage III-IV and M1, which was also found in the validation group, except for the M stage of GSE29609. A reasonable explanation was the limited number of patients in GSE29609. Moreover, Kaplan-Meier analysis identified that the IRS score could distinguish the prognosis of patients in the male, female, T1-2, AJCC stage I-II and M0 subgroups, and these results also appeared in the TCGA test. Unfortunately, GSE29609 did not perform subgroup analysis due to the limited samples. Furthermore, the IRS clearly differed the patients in the high- and low-risk score groups in all groups. Therefore, we successfully identified the stability and universality of the IRS.

To explore the related pathways, GSVA was used to determine the enriched KEGG pathways by comparing the transcriptome data of the high- and low-risk score groups. The enrichment pathways included primary immunodeficiency, antigen processing and presentation, allograft rejection and systemic lupus erythematosus, which were mainly correlated with inflammation and immune pathways. In accordance with the present results, Zhang et al. (2021) also constructed an inflammation-related signature for gastric carcinoma, and the enriched KEGG pathways were associated with immune pathways. According to the results of tumor mutational burden analysis, VHL and PBRM1 made up the majority of the mutations. Gong et al. (022) reported that VHL gene expression can significantly inhibit the proliferation ability of RCC and promote its apoptosis. However, the high mutation of VHL suggested that it might be a therapeutic target in KIRC. Furthermore, immune checkpoints are widely focused biomarkers of immunotherapy. PD-L1 and HAVCR2 were significantly higher in the low-risk-score group, which suggested that patients in the low-risk-score group might be sensitive to PD-L1-related immunotherapy. As mentioned in the literature review, previous studies reported that many cancers with high TMB scores were usually associated with better survival after receiving immunotherapy (Valero et al., 2021). However, this outcome was contrary to the effect of TMB in RCC, which suggested that a high TMB score could not bring survival benefit to patients who received immunotherapy (Wood et al., 2020). Similarly, the high-risk score group was correlated with higher TMB but was associated with lower expression of PD-L1 and HAVCR2 in this study. Moreover, in our study, patients with low-risk score were associated with lower TIDE score than patients with high-risk score, which might predict a poor response of patients with high-risk score to anti-PD1 or anti-CTLA4 immunotherapy (Jiang et al., 2018). Therefore, compared with patients in the low-risk score group, those in the high-risk score group might benefit less from immunotherapy with the PD-L1 and HAVCR2 immune checkpoints. However, the high-risk score group was positively correlated with the immune infiltration of $CD8^+$ T cells, activated $CD4^+$ memory T cells, gamma and delta regulatory T cells and macrophages. Therefore, combining the results of TMB, immune checkpoint and immune infiltration, these results might be interpreted cautiously by patients in the high-risk score group, as they might benefit from other immune checkpoints, such as LAG3, rather than PD-L1. Of course, this explanation requires further experiments and studies. Regardless of the explanation, the results of the present study suggest that the IRS is a robust biomarker to predict outcomes and treatment responses in KIRC patients.

This study first identified an inflammation-related signature in KIRC, and we demonstrated its value. It has the potential to become a powerful tool in the management of KIRC patients in clinical practice.

Conclusion

We found that the IRS might serve as a biomarker to predict the survival of KIRC. Moreover, patients with high or low risk scores might be sensitive to immune drugs at different immune checkpoints.

Data availability statement

Raw data used in this paper were obtained from the public database, and further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

QY and FZ designed the study, analysis of the data, drafted the manuscript and gave final approval of the version to be published. DF and YX participated in the design of the study, analysis of the data and gave final approval of the version to be published. DL and M-FG analysis of the data, participated in revision of the manuscript, and gave final approval of the version to be published.

Funding

The project was supported by Basic Public Welfare Research Project of Zhejiang 436 Province, Grant/Award Number: LGF20H160023.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Borrelli, C., Ricci, B., Vulpis, E., Fionda, C., Ricciardi, M. R., Petrucci, M. T., et al. (2018). Drug-induced senescent multiple myeloma cells elicit NK cell proliferation by direct or exosome-mediated IL15 trans-presentation. *Cancer Immunol. Res.* 6 (7), 860–869. doi:10.1158/2326-6066.Cir-17-0604

De Mattia, E., Polesel, J., Roncato, R., Labriet, A., Bignucolo, A., Gagno, S., et al. (2021). IL15RA and SMAD3 genetic variants predict overall survival in metastatic colorectal cancer patients treated with folfiri therapy: A new paradigm. *Cancers (Basel)* 13 (7), 1705. doi:10.3390/cancers13071705

Edeline, J., Mottier, S., Vigneau, C., Jouan, F., Perrin, C., Zerrouki, S., et al. (2012). Description of 2 angiogenic phenotypes in clear cell renal cell carcinoma. *Hum. Pathol.* 43 (11), 1982–1990. doi:10.1016/j.humpath.2012.01.023

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30 (1), 207–210. doi:10.1093/nar/30.1.207

Escudier, B., Porta, C., Schmidinger, M., Rioux-Leclercq, N., Bex, A., Khoo, V., et al. (2019). Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 30 (5), 706–720. doi:10.1093/annonc/mdz056

Feng, X., Yan, N., Sun, W., Zheng, S., Jiang, S., Wang, J., et al. (2019). miR-4521-FAM129A axial regulation on ccRCC progression through TIMP-1/MMP2/MMP9 and MDM2/p53/Bcl2/Bax pathways. *Cell Death Discov.* 5, 89. doi:10.1038/s41420-019-0167-5

Fu, J., Li, K., Zhang, W., Wan, C., Zhang, J., Jiang, P., et al. (2020). Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* 12 (1), 21. doi:10.1186/s13073-020-0721-z

Gong, Y., Wang, D., and Wang, W. (2022). Biostatistics of VHL-gene transfection in the health informatics analysis of renal cell carcinoma. *Comput. Math. Methods Med.* 2022, 5297580. doi:10.1155/2022/5297580

Hanahan, D. (2022). Hallmarks of cancer: New dimensions. *Cancer Discov.* 12 (1), 31–46. doi:10.1158/2159-8290.Cd-21-1059

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Jin, C., Shi, L., Li, Z., Liu, W., Zhao, B., Qiu, Y., et al. (2019). Circ_0039569 promotes renal cell carcinoma growth and metastasis by regulating miR-34a-5p/CCL22. *Am. J. Transl. Res.* 11 (8), 4935–4945.

Li, Z., Du, G., Zhao, R., Yang, W., Li, C., Huang, J., et al. (2021). Identification and validation of a hypoxia-related prognostic signature in clear cell renal cell carcinoma patients. *Med. Baltim.* 100 (39), e27374. doi:10.1097/md.0000000000027374

Liang, F. (2020). Optimising first-line treatment for metastatic renal cell carcinoma. *Lancet* 395 (10219), e8. doi:10.1016/s0140-6736(19)32594-2

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, M., Chen, S., Zhang, A., Zheng, Q., and Fu, J. (2021). PLAUR as a potential biomarker associated with immune infiltration in bladder urothelial carcinoma. *J. Inflamm. Res.* 14, 4629–4641. doi:10.2147/jir.S326559

Lu, H., Cao, X., Zhang, H., Sun, G., Fan, G., Chen, L., et al. (2014). Imbalance between MMP-2, 9 and TIMP-1 promote the invasion and metastasis of renal cell carcinoma via SKP2 signaling pathways. *Tumour Biol.* 35 (10), 9807–9813. doi:10.1007/s13277-014-2256-7

Marona, P., Górka, J., Mazurek, Z., Wilk, W., Rys, J., Majka, M., et al. (2017). MCPIP1 downregulation in clear cell renal cell carcinoma promotes vascularization and metastatic progression. *Cancer Res.* 77 (18), 4905–4920. doi:10.1158/0008-5472.Can-16-3190

Marra, P., Mathew, S., Grigoriadis, A., Wu, Y., Kyle-Cezar, F., Watkins, J., et al. (2014). IL15RA drives antagonistic mechanisms of cancer development and immune control in lymphocyte-enriched triple-negative breast cancers. *Cancer Res.* 74 (17), 4908–4921. doi:10.1158/0008-5472.Can-14-0637

Nerich, V., Hugues, M., Paillard, M. J., Borowski, L., Nai, T., Stein, U., et al. (2014). Clinical impact of targeted therapies in patients with metastatic clear-cell renal cell carcinoma. *Onco. Targets. Ther.* 7, 365–374. doi:10.2147/ott.S56370

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Nishida, J., Momoi, Y., Miyakuni, K., Tamura, Y., Takahashi, K., Koinuma, D., et al. (2020). Epigenetic remodelling shapes inflammatory renal cancer and neutrophil-dependent metastasis. *Nat. Cell Biol.* 22 (4), 465–475. doi:10.1038/s41556-020-0491-2

Ohnishi, S., Ma, N., Thanan, R., Pinlaor, S., Hammam, O., Murata, M., et al. (2013). DNA damage in inflammation-related carcinogenesis and cancer stem cells. *Oxid. Med. Cell. Longev.* 2013, 387014. doi:10.1155/2013/387014

Okada, F., Izutsu, R., Goto, K., and Osaki, M. (2021). Inflammation-related carcinogenesis: Lessons from animal models to clinical aspects. *Cancers (Basel)* 13 (4), 921. doi:10.3390/cancers13040921

Peña, C., Lathia, C., Shan, M., Escudier, B., and Bukowski, R. M. (2010). Biomarkers predicting outcome in patients with advanced renal cell carcinoma: Results from sorafenib phase III Treatment Approaches in Renal Cancer Global Evaluation Trial. *Clin. Cancer Res.* 16 (19), 4853–4863. doi:10.1158/1078-0432.Ccr-09-3343

Shen, C., Liu, J., Wang, J., Zhong, X., Dong, D., Yang, X., et al. (2020). Development and validation of a prognostic immune-associated gene signature in clear cell renal cell carcinoma. *Int. Immunopharmacol.* 81, 106274. doi:10.1016/j.intimp.2020.106274

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2022). Cancer statistics, 2022. *Ca. Cancer J. Clin.* 72 (1), 7–33. doi:10.3322/caac.21708

Smyth, A., Reid, H. M., Baker, A. H., and McGlynn, H. (2007). Modifications of the radiosensitivity of a renal cancer cell line as a consequence of stable TIMP-1 overexpression. *Int. J. Radiat. Biol.* 83 (1), 13–25. doi:10.1080/09553000600983136

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Valero, C., Lee, M., Hoen, D., Wang, J., Nadeem, Z., Patel, N., et al. (2021). The association between tumor mutational burden and prognosis is dependent on treatment context. *Nat. Genet.* 53 (1), 11–15. doi:10.1038/s41588-020-00752-4

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Wei, L., Jin, Z., Yang, S., Xu, Y., Zhu, Y., and Ji, Y. (2018). TCGA-Assembler 2: Software pipeline for retrieval and processing of TCGA/CPTAC data. *Bioinformatics* 34 (9), 1615–1617. doi:10.1093/bioinformatics/btx812

Wood, M. A., Weeder, B. R., David, J. K., Nellore, A., and Thompson, R. F. (2020). Burden of tumor mutations, neoepitopes, and other variants are weak predictors of cancer immunotherapy response and overall survival. *Genome Med.* 12 (1), 33. doi:10.1186/s13073-020-00729-2

Yang, H., Zhou, L., Chen, J., Su, J., Shen, W., Liu, B., et al. (2019). A four-gene signature for prognosis in breast cancer patients with hypermethylated IL15RA. *Oncol. Lett.* 17 (5), 4245–4254. doi:10.3892/ol.2019.10137

Yang, N., Liu, Z., Pang, S., Wu, J., Liang, J., and Sun, L. (2021). Predicative value of IFITM2 in renal clear cell carcinoma: IFITM2 is associated with lymphatic metastasis and poor clinical outcome. *Biochem. Biophys. Res. Commun.* 534, 157–164. doi:10.1016/j.bbrc.2020.11.124

Yoshie, O. (2021). CCR4 as a therapeutic target for cancer immunotherapy. *Cancers (Basel)* 13 (21), 5542. doi:10.3390/cancers13215542

Zhang, S., Li, X., Tang, C., and Kuang, W. (2021). Inflammation-related long non-coding RNA signature predicts the prognosis of gastric carcinoma. *Front. Genet.* 12, 736766. doi:10.3389/fgene.2021.736766

Zhao, E., Li, L., Zhang, W., Wang, W., Chan, Y., You, B., et al. (2019). Comprehensive characterization of immune- and inflammation-associated biomarkers based on multi-omics integration in kidney renal clear cell carcinoma. *J. Transl. Med.* 17 (1), 177. doi:10.1186/s12967-019-1927-y

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fgene.2022.834935

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Ning Yang,
Luoyang Orthopedic Traumatological Hospital, China
Xiao Liang,
Sichuan University, China

*CORRESPONDENCE
Mingming Zhang,
zhangmm@cqmu.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 14 December 2021
ACCEPTED 11 July 2022
PUBLISHED 11 August 2022

CITATION
Tong F, Xu L, Xu S and Zhang M (2022), Identification of an autophagy-related 12-lncRNA signature and evaluation of NFYC-AS1 as a pro-cancer factor in lung adenocarcinoma.
Front. Genet. 13:834935.
doi: 10.3389/fgene.2022.834935

COPYRIGHT
© 2022 Tong, Xu, Xu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of an autophagy-related 12-lncRNA signature and evaluation of NFYC-AS1 as a pro-cancer factor in lung adenocarcinoma

Fang Tong[1,2], Lifa Xu[1], Sheng Xu[3] and Mingming Zhang[1,4]*

[1]Department of Medical Immunology, School of Medicine, Anhui University of Science and Technology, Anhui, China, [2]Anhui Province Engineering Laboratory of Occupational Health and Safety, Anhui University of Science and Technology, Anhui, China, [3]The First Affiliated Hospital, Anhui University of Science and Technology, Anhui, China, [4]Chongqing Key Laboratory of Traditional Chinese Medicine for Prevention and Cure of Metabolic Diseases, College of Traditional Chinese Medicine, Chongqing Medical University, Chongqing, China

Objective: To develop an autophagy-related lncRNA-based risk signature and corresponding nomogram to predict overall survival (OS) for LUAD patients and investigate the possible meaning of screened factors.

Methods: Differentially expressed lncRNAs and autophagy genes were screened between normal and LUAD tumor samples from the TCGA LUAD dataset. Univariate and multivariate Cox regression analyses were performed to construct the lncRNA-based risk signature and nomogram incorporating clinical information. Then, the accuracy and sensitivity were confirmed by the AUC of ROC curves in both training and validation cohorts. qPCR, immunoblot, shRNA, and ectopic expression were used to verify the positive regulation of NFYC-AS1 on BIRC6. CCK-8, immunofluorescence, and flow cytometry were used to confirm the influence of NFYC-AS1 on cell proliferation, autophagy, and apoptosis *via* BIRC6.

Results: A 12-lncRNA risk signature and a nomogram combining related clinical information were constructed. Furthermore, the abnormal increase of NFYC-AS1 may promote LUAD progression through the autophagy-related gene BIRC6.

Conclusion: 12-lncRNA signature may function as a predictive marker for LUAD patients, and NFYC-AS1 along with BIRC6 may function as carcinogenic factors in a combinatorial manner.

KEYWORDS

LUAD, LncRNA, long noncoding RNA, autophagy, risk signature, prognostic indicator, NFYC-AS1, BIRC6

Introduction

Lung cancer accounts for 12.7% of the newly diagnosed cancer patients and up to 22.4% of deaths caused by cancer in 2020, according to the National Cancer Institute of the United States. About 50% of lung cancer cases belong to the lung adenocarcinoma (LUAD) type. Evaluation at early stages may help identify patients with a high-risk re-occurrence rate, develop a specific treatment strategy, and improve the overall survival rate of patients. Thus, the development of meaningful prognostic markers and the construction of related risk evaluation are essential for the diagnosis and treatment of LUAD patients.

Long noncoding RNA (lncRNA) is defined as a class of transcripts longer than 200 nucleotides (bps) and not translated into proteins (Iyer et al., 2015). LncRNA has been recently found to have many biological functions such as transcriptional regulation, translation, epigenetic modification, and cell fate decision (Ulitsky and Bartel, 2013; Li and Izpisua Belmonte, 2015). LncRNA has been demonstrated to be the driver of carcinogenesis, cancer progression, and metastasis in various cancers (Hirata et al., 2015; Yue et al., 2016; Yin et al., 2018; Tamang et al., 2019). For example, lncRNA-DANCR is overexpressed in tumor initiation cells, identifying DANCR as a prognostic biomarker and therapeutic target for hepatocellular carcinoma treatment (Yuan et al., 2016). Regarding LUAD, LncRNAs have been linked to various processes such as epithelial–mesenchymal transition (EMT), angiogenesis, and metastasis (Peng et al., 2018; Guan et al., 2019; Wang et al., 2020a).

Autophagy is an evolutionarily highly conserved catabolic process in eukaryotic cells (Han et al., 2014). It maintains the intracellular homeostasis of the environment by degrading and recycling the cellular components *via* the formation of autophagosome vesicles to engulf dysfunctional cytoplasmic organelles and the formation of autolysosome to degrade the contents of vesicles in the end (Klionsky, 2007; Fulda and Kögel, 2015). Autophagy plays a dual-faced role in tumor progression. On the one hand, it provides essential circulating metabolic substrates for biomolecule synthesis and supports the survival of cancer cells (Fulda and Kögel, 2015). On the other hand, autophagy could interplay with apoptosis to induce the death of cancer cells, including LUAD cells (Gewirtz, 2014; Liu et al., 2017). In addition, it plays an even more complex role in drug resistance, which depends on the activation status of autophagy and cellular environments (Li et al., 2017). Recently, accumulating evidence has shown that lncRNAs regulate the autophagy network *via* transcriptional regulations of autophagy-related genes in LUAD cells (Wang et al., 2019). Given the relevance of autophagy and lncRNAs in LUAD initiation and

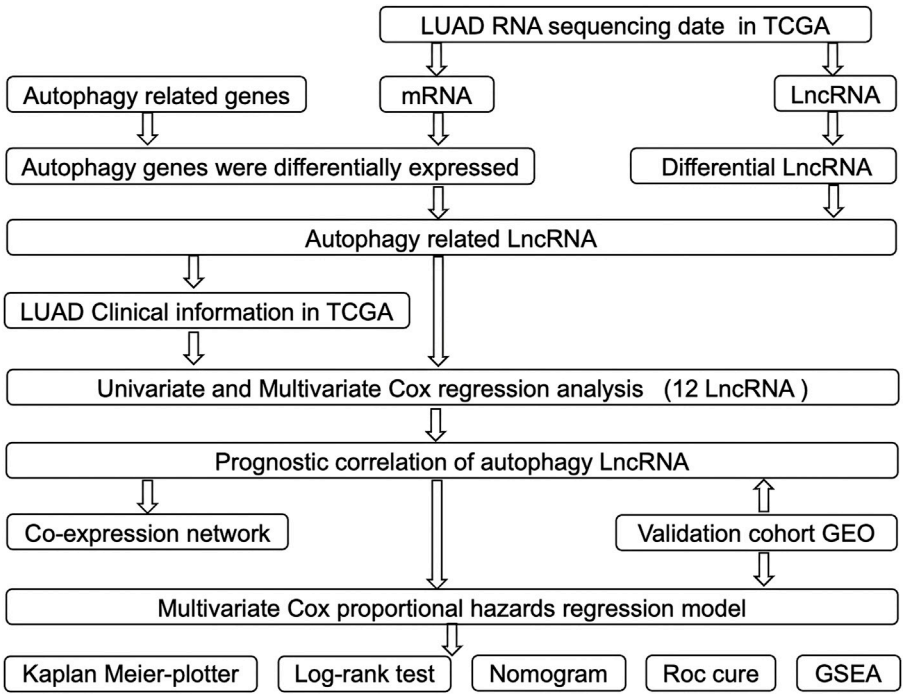


FIGURE 1
Flowchart of construction and validation of lncRNA risk signature and nomogram incorporating clinical parameters.

progression, this study focuses on investigating the expression patterns of autophagy-related lncRNAs from TCGA and Human Autophagy Database (HADb) to construct a useful risk signature to predict the prognosis of LUAD patients.

Materials and methods

Data collection and processing

A detailed workflow for the construction of risk signature was developed, as shown in Figure 1. lncRNA and mRNA FPKM (fragments per kilobase of transcript per million fragments mapped) (level 3) sequencing profiles and associated clinical information of patients with LUAD were obtained from the TCGA data portal (https://tcgadata.nci.nih.gov/tcga/) before 1 November 2020. The GSE31210 microarray expression data were downloaded from the Gene Expression Omnibus (GEO) database, which included data of 226 LUAD samples, 20 normal tissue samples, and their associated clinical information. All autophagy-related genes were downloaded from the Human Autophagy Database (HADb) (http://www.autophagy.lu/). The "Limma" package of R software was used to screen differential autophagy-related gene expression and lncRNAs between adjacent and tumor tissues when $|\log_2\text{fold change (FC)}| \geq 0.5$, and adjusted $p < 0.05$ was used as a filtering threshold. "Limma" package of R software was used to analyze the correlation between each differential lncRNA and every differential autophagy-related mRNA, and autophagy-related lncRNAs were screened when corFilter ≥0.3, and $p < 0.05$ was used as a filtering threshold.

Identification of autophagy-related lncRNA prognostic signature

After the exclusion of LUAD patients with unavailable or insufficient follow-up clinical information, 458 LUAD patients were submitted for univariate Cox regression analysis. The "Survival" package of R software was used to perform the univariate regression analysis for the above lncRNAs, and prognosis-related lncRNAs were initially selected as candidates when $p < 0.05$. Next, the "Survival" package of the R software was used to perform multivariate COX regression analysis, and interdependent lncRNAs were identified as the risk signature, further validated by the Kaplan–Meier survival analysis. The risk score of each patient was based on a linear combination of lncRNA expression level (Exp) multiplied by a regression coefficient (β), represented as the following:

$$\textit{Risk score} = \sum_{i=1}^{n} \textit{Expi}\,\beta i.$$

Then, based on the median risk score value, patients in the cohort were divided into high-risk and low-risk groups.

Visualization of expression correlation network of autophagy-related genes and lncRNAs

After lncRNAs were identified as the risk signature, Cytoscape of version 3.6.1 was used to generate the expression correlations network between autophagy-related genes and lncRNAs when corFilter ≥ 0.3. The relations among high- or low-risk, autophagy-related genes and lncRNAs were also illustrated as a Sankey diagram to show the possible influence of autophagy-related lncRNAs on clinical outcomes.

Risk signature assessment

After the risk signature was constructed, the Kaplan–Meier survival analysis was first used to evaluate the overall survival between high- and low-risk groups. Furthermore, to evaluate the predictive ability of the constructed risk signature, the time-dependent receiver operating characteristic (ROC) analyses were performed by the "Survival ROC" package of R software to test the signature's sensitivity and specificity.

Development and validation of a nomogram incorporating the lncRNA signature with clinical factors

Next, based on univariate and multivariate Cox regression results, a novel nomogram incorporating the 12-lncRNA signatures and related clinical factors was developed by the "rms" package of the R software. The predictive ability of the risk signature was further evaluated with the area under the curve (AUC) in the ROC analysis. "Caret," "foreign," and "survival" packages of R software were also used to randomly select half of the samples from the TCGA LUAD dataset as the internal cohort for validation. All the analyses were performed in the TCGA internal cohort and GSE dataset.

Gene set enrichment analysis

Based on the median risk scores, 551 LUAD samples in the TCGA cohort were divided into two groups (high- and low-risk groups). Then, gene set enrichment analysis (GSEA) was performed to identify the significantly differentially expressed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between the high- and low-risk groups using the Java-based GSEA software. A nominal (NOM) $p < 0.05$ was chosen as the cutoff value.

TABLE 1 Primers for the detection of lncRNAs expression in LUAD specimens.

Primer	Sequence
HCG18-F	TGAAAGTCGACGAAGAGGGC
HCG18-R	ACTAGTCGGAAGTGACGTGC
TMPO-AS1-F	GCAGTTTAAAAGGCGCTGGG
TMPO-AS1-R	CCCGATTGTAGAGGTCTGGC
NFYC-AS1-F	CTTCCCTATCACCCTCGCAC
NFYC-AS1-R	GATCTCGAG CTCCCGAAAGG
LINC00996-F	CTCCCATCTTTTCTGCCGGT
LINC00996-R	AGGGGCATCCAAAGGTCTTG

Quantitative reverse transcription polymerase reaction assays for lung cancer belong to lung adenocarcinoma patient specimens and further validation in GSE40791 datasets

Ten paired LUAD tumor and adjacent normal tissues were obtained from patients of the First Affiliated Hospital of Anhui University of Science and Technology. All these patient tissues were transferred on ice and stored in liquid nitrogen. Total RNA was extracted from tissues with TRIzol reagent (Invitrogen, China) according to the manufacturer's guide manual. Reverse transcription was performed according to the manufacturer's instructions using the ReverTra Ace qPCR RT Kit (Toyobo, Japan). The SYBR Green Realtime PCR Master Mix (Toyobo) was used in the quantitative reverse transcription-polymerase chain reaction (qRT-PCR) experiment. Gene expression level was calculated with the 2-DDCt method. Primers in the qRT-PCR test are listed in Supplementary Table S1. Informed consent was signed by all the participants. The study was approved by the Ethics Committee of Chongqing Medical University. To further validate the expression patterns of these lncRNAs, we use GSE40791 datasets, which included 100 non-neoplastic (N) lung samples, and 69, 12, and 13 stages I, II, and II lung adenocarcinoma (AD) frozen tissues, respectively. The box plot is implemented by the "ggplot2" package of R software based on the Wilcoxon rank sum test.

Cell culture, plasmid construction, and transfection

A549 cells were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in Dulbecco's MEM medium (Hyclone, Los Angeles, United States) supplemented with 10% fetal bovine serum and a 1% penicillin–streptomycin solution. All cells were incubated at 37°C in a humidified incubator containing 5% CO_2. pEGFP-BIRC6 overexpression plasmid was constructed as previously described (Zhang et al., 2020a) by replacing it with the BIRC6 CDS region. For knockdown of NFYC-AS1, the specific shRNAs were synthesized and constructed into the same vector as previously described (Zhang et al., 2020a) by replacing targeted sequences with "CACCTGTAA TCCCAGCACTTT" (sh1) or "CACACCTGTAATCCCAGCATT" (sh2). Assayed cells were transfected with Lipofectamine® 2000 Reagent (Invitrogen) following the supplier's instruction manual.

Cell proliferation and apoptosis assays

Cells were seeded in 96-well plates at a density of 2,000 cells per well. The growth rate of cells was evaluated

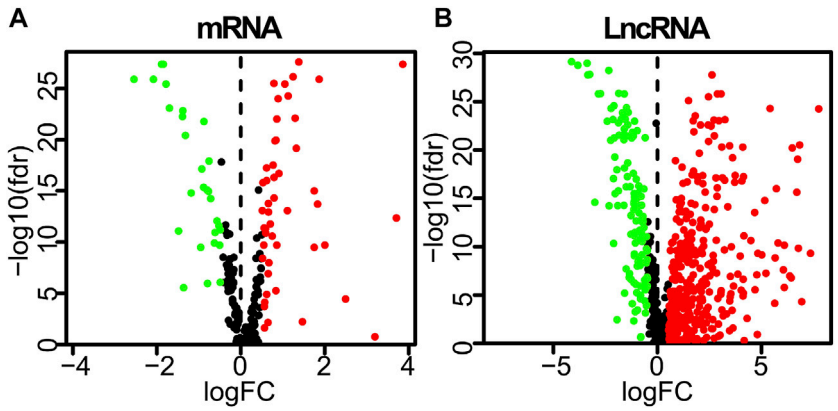


FIGURE 2
Differently expressed autophagy-related genes and lncRNAs based on the correlation coefficient | log2(FC)| ≥ 0.5 and adjusted $p < 0.05$. **(A)** Volcano plot of differently expressed autophagy-related genes. **(B)** Volcano plot of differently expressed lncRNAs between LUAD tumor and normal samples. FC, fold change.

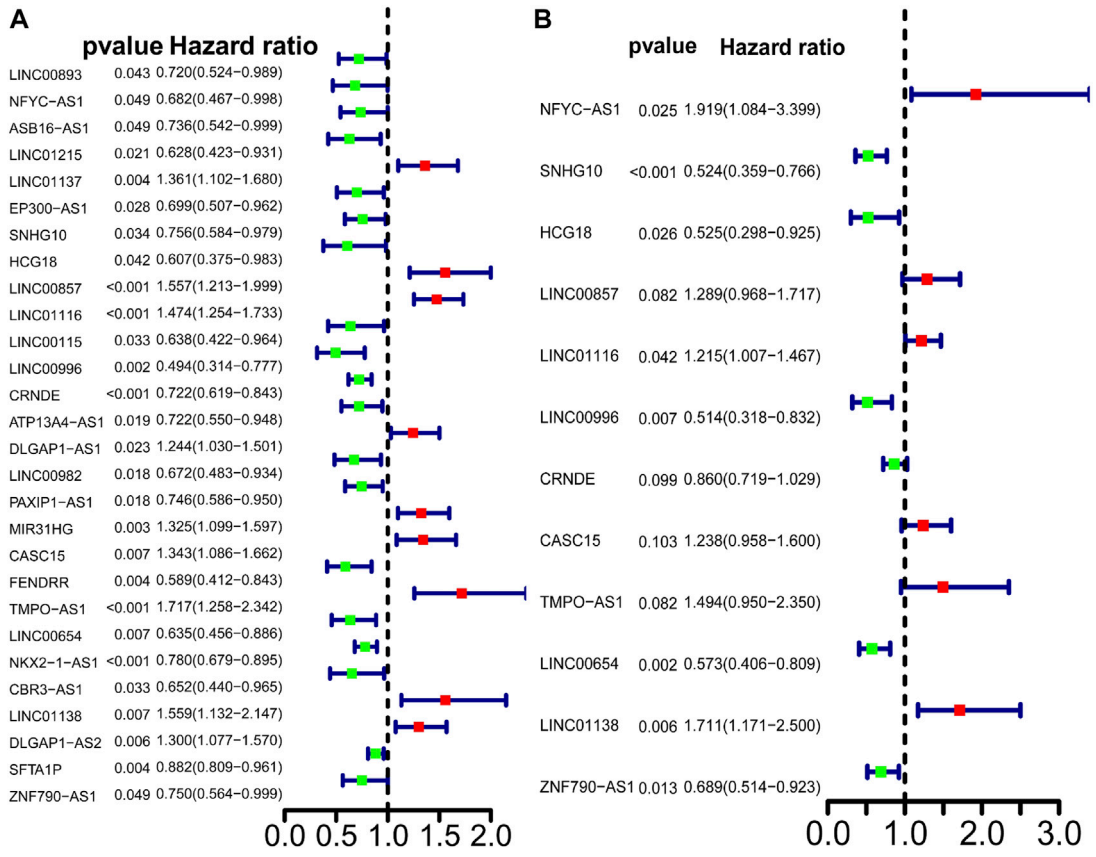


FIGURE 3
Predictive power of the differently expressed lncRNAs in LUAD patients. **(A)** Univariate Cox regression analysis of differently expressed autophagy-related lncRNAs. Cox analysis. **(B)** Multivariate Cox analysis of differently expressed autophagy-related lncRNAs.

using the CCK-8 cell proliferation kit (Dojindo Laboratories, Kumamoto, Japan), according to the manufacturers' instructions. OD detection at 450 nm was carried out by infinite 200Pro (Tecan). Cell apoptosis was analyzed using the Annexin V/7-AAD Apoptosis Detection kit (Keygen Biotech, Nanjing, China) on a CytoFLEX flow cytometer (Beckman, California, United States). Results were further analyzed by Flowjo 10.4.

Western blotting, immunofluorescence assay, and immunohistochemistry

For immunoblotting experiment, cells were lysed in RIPA buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS). An equal amount of protein was loaded and simultaneously subjected to electrophoresis in SDS-polyacrylamide gel and transferred to 0.22 μm pore-sized PVDF membranes (Roche, Basel, Switzerland). Membranes were briefly blocked with 5% skim milk and incubated with the primary antibodies overnight at 4°C, followed by incubation with the species-matched secondary antibody conjugated with HRP (Cell Signaling Technology, Danvers, United States) for 1 h at room temperature prior to chemiluminescence detection. For the immunofluorescence (IF) experiment, cells with less than 50% confluency were seeded and grown on coverslips overnight. Cells were fixed with 4% paraformaldehyde and permeabilized with Triton-100. After blocking in 5% goat serum, samples were incubated with the appropriately diluted primary and secondary antibodies. Eventually, cells were observed and photographed by fluorescence microscopy (Nikon C2). For the immunohistochemistry experiment, all tissues were fixed with 4% paraformaldehyde overnight at 4°C and then embedded in paraffin. The samples were subsequently sectioned into thin slices to be mounted on slides, followed by deparaffinization in xylene and rehydration through a series of ethanol–water solutions. Antigen retrieval was carried out by immersing the sections in citrate acid buffer with heating by a microwave oven. Slides

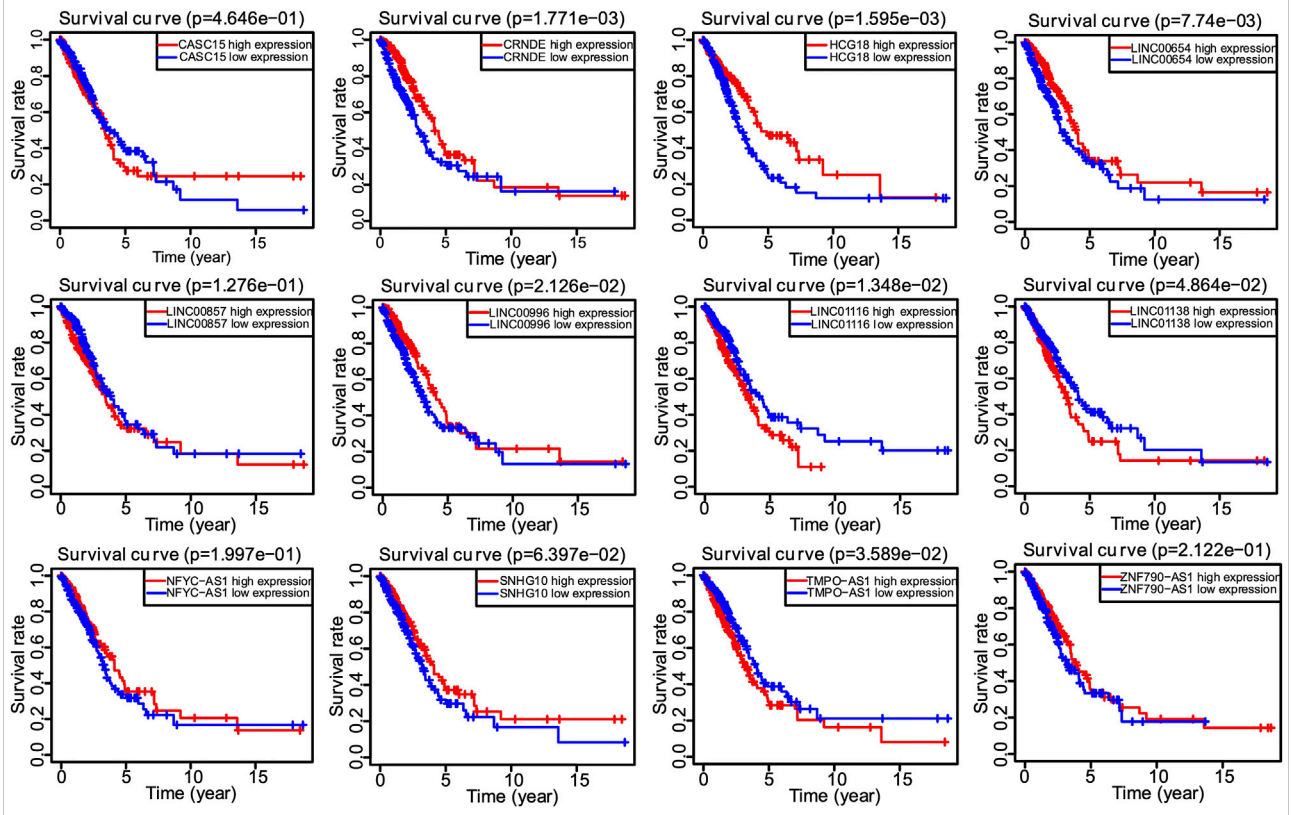


FIGURE 4
The Kaplan–Meier survival analysis of 12 screened lncRNAs.

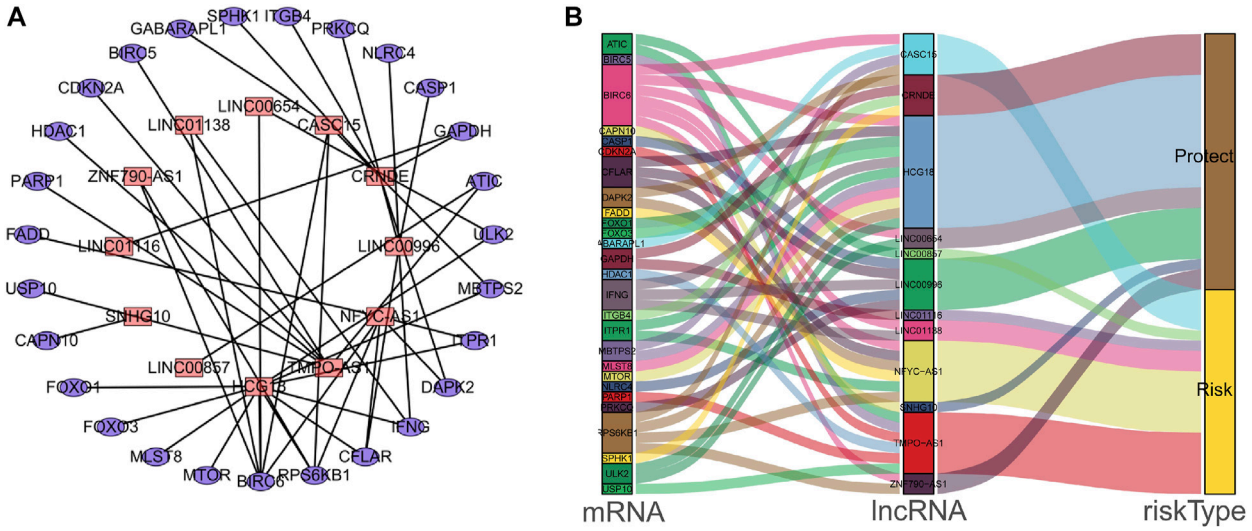


FIGURE 5
The interaction network of autophagy genes and OS-associated lncRNAs in LUAD patients. **(A)** The co-expression network of the autophagy-related mRNAs and lncRNAs was constructed and visualized using Cytoscape. **(B)** The co-expression network with risk-type information was visualized using the Sankey diagram.

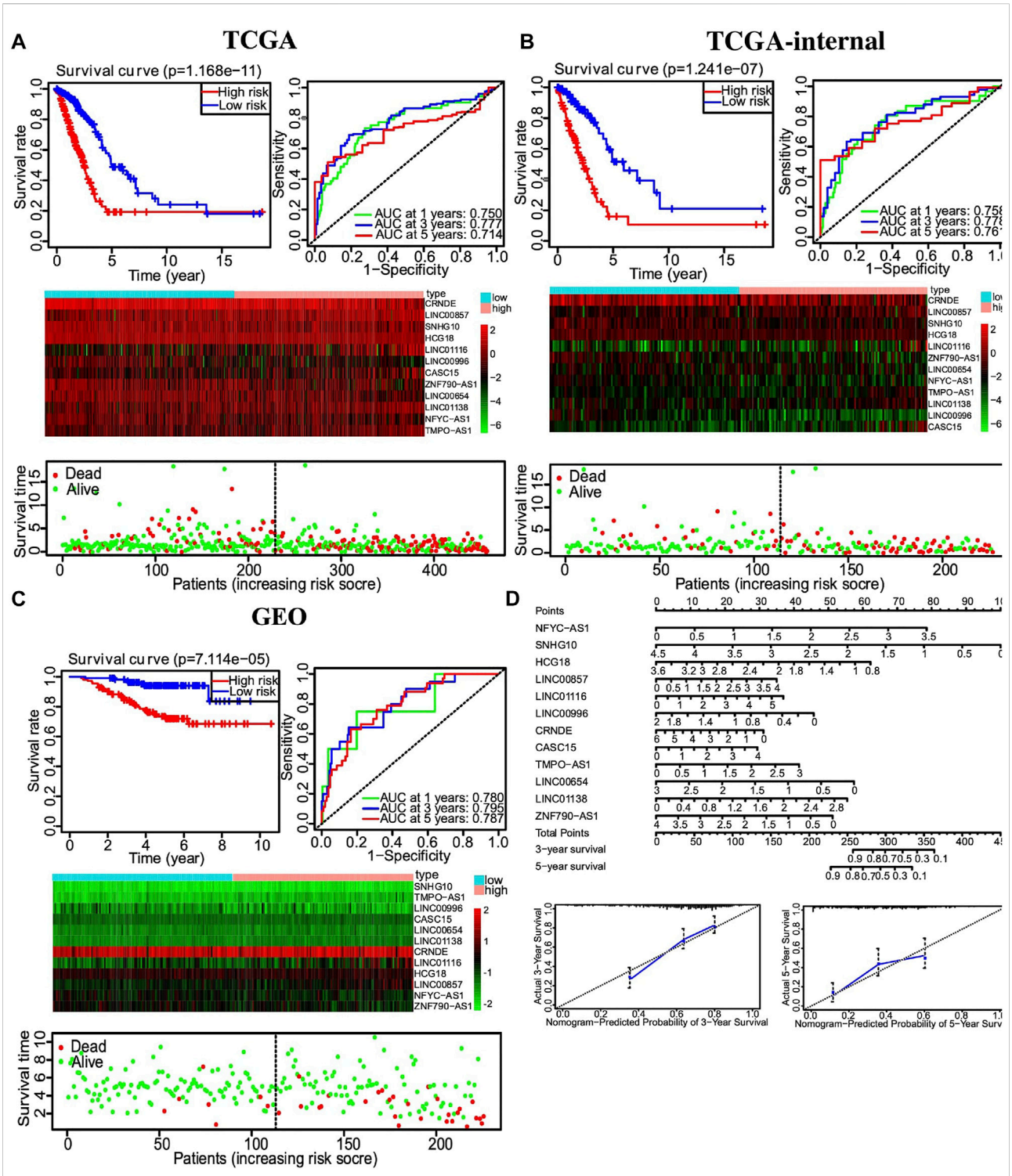


FIGURE 6
Autophagy-related lncRNA risk score analysis of LUAD patients. **(A)** Top left, Kaplan–Meier survival analysis of the high- and low-risk groups based on the median risk score. Top right, 1-, 3-, and 5-year AUC value of ROC analysis. Middle, heatmap of autophagy-related lncRNAs expression profiles in the low- and high-risk groups. Bottom, the scatterplot based on the survival status of each patient. **(B)** Similar results to 6A from the TCGA train dataset. **(C)** Similar results to 6A from the GSE validation dataset. **(D)** Nomogram of 12-lncRNA signature based on multivariate Cox proportional hazards regression analyses.

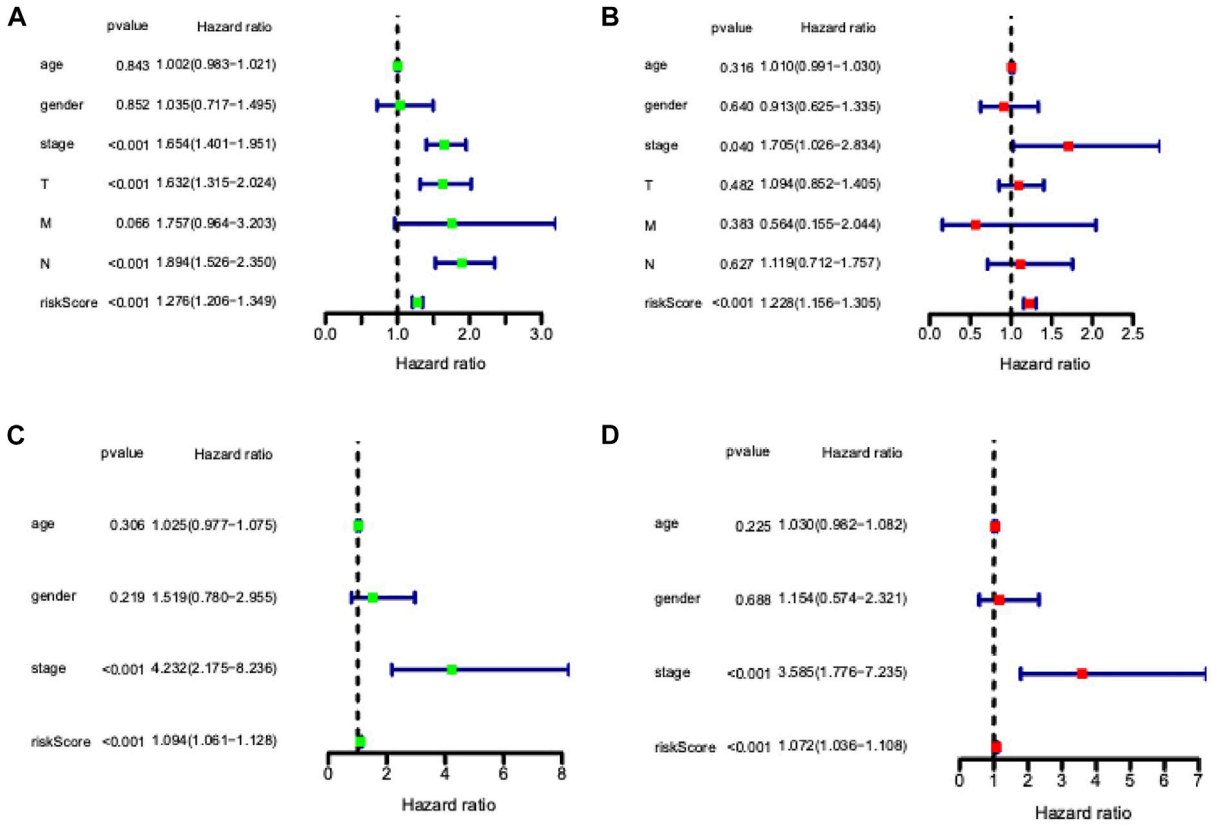


FIGURE 7
Independent prognostic ability evaluation for the constructed prognostic signature in LUAD. **(A)** Univariate Cox regression analysis of risk score and related clinical information regarding prognostic value in TCGA LUAD dataset. **(B)** Multivariate Cox analysis of risk model score and clinical features regarding prognostic value in the TCGA LUAD dataset. **(C)** Univariate Cox regression analysis of risk score and related clinical information regarding prognostic value in GSE dataset. **(D)** Multivariate Cox analysis of risk model score and clinical features regarding prognostic value in GSE dataset.

were then blocked with 3% hydrogen peroxide to block nonspecific activity. After rinsing, slides were blocked with 5% BSA and then incubated with BIRC6 antibody overnight at 4°C. An immunohistochemical staining kit (BOSTER Biological Technology, Wuhan, China) was used for color development. Images were captured and confirmed by a professional pathologist under the microscope (Nikon, Tokyo, Japan). The color intensity of slides was divided into five grades (points) to score the BIRC6 expression level for statistical analysis. Antibodies used in the above experiments were as follows: β-actin (Cat No. 20536-1-AP, Proteintech, 1:1000), BIRC6 (ab19609, Abcam, 1:2000), Beclin 1 (Cat No. 11306-1-AP, Proteintech, 1:1000), SQSTM1/p62 (Cat No. 18420-1-AP, Proteintech, 1:1000), LC3B (Cat. No. NB600-1384, Novus Biologicals, 1:500), and BIRC6 (29760-1-AP, Proteintech, 1:200) for immunohistochemistry.

Immunodeficient xenograft mouse tumor model

BALB/c-nude mice (4–5 weeks of age, 18–20g) were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. (China). All experimental procedures were approved by the Institutional Animal Care and Use Committee of Chongqing Medical University. Briefly, cells were trypsinized, washed, and resuspended in phosphate-buffered saline at a density of 10^7 cells/ml. A total of fifteen mice were inoculated subcutaneously with 5×10^6 cells of A549 Vector cells, A549 NFYC-AS1 knockdown cells, and A549 NFYC-AS1 knockdown with BIRC6 overexpression cells, respectively. Mice were sacrificed 7 days after tumor cell implantation, and the tumors were removed and weighed. GraphPad Prism 8 and unpaired *t*-test were used for statistical analysis.

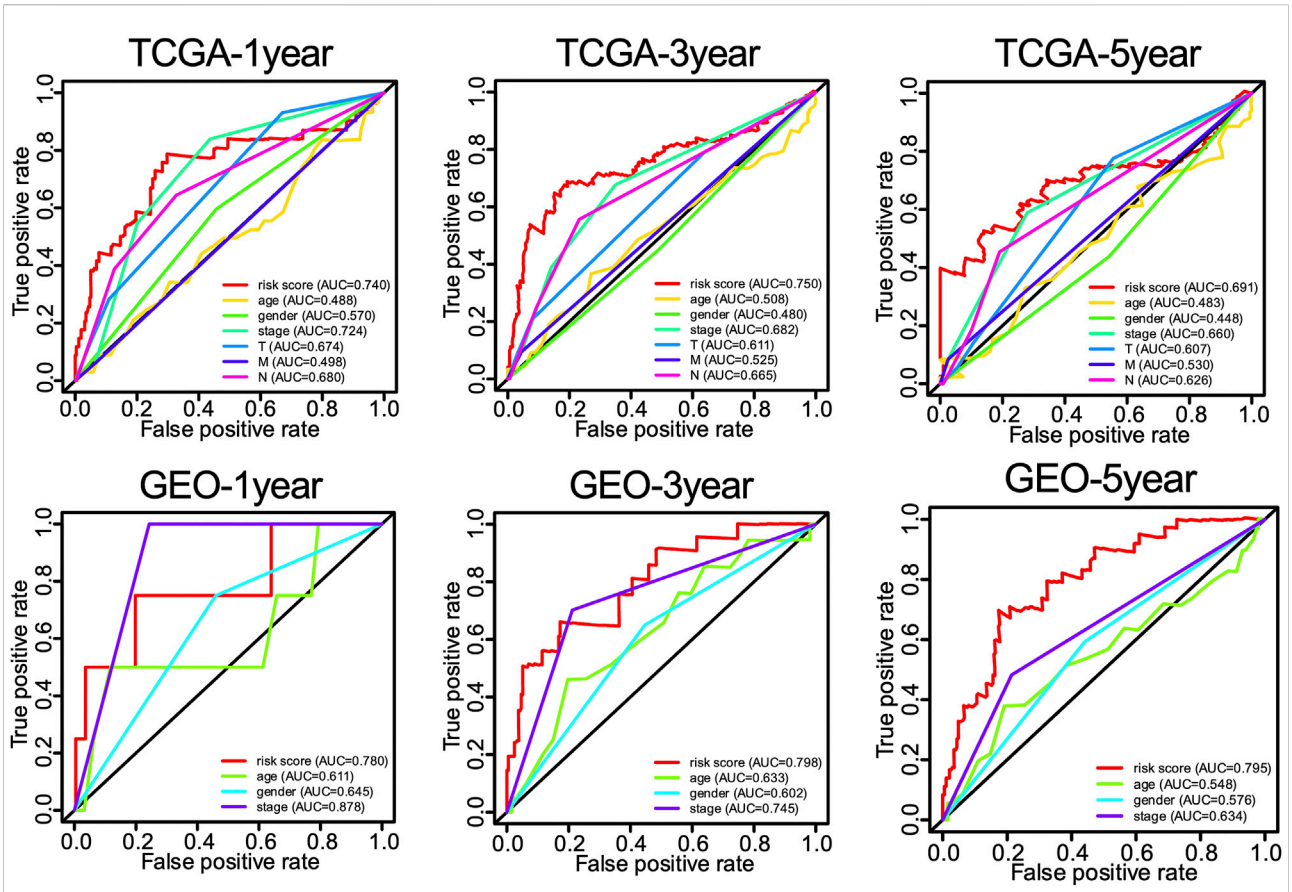


FIGURE 8
The 1-, 3-, and 5-year ROC curves of risk score and related clinical information of LUAD patients.

Results

Differential expression of autophagy-related genes and lncRNAs in lung cancer belong to lung adenocarcinoma patients

A detailed data processing flowchart is shown in Figure 1. Clinical and diagnostic information of LUAD patients in the TCGA and validation cohorts are shown in Table 1. A total of 232 autophagy genes were collected from the HADb database, and we identified 76 differentially expressed autophagy genes between 497 tumor tissues and 54 normal tissues from TCGA at the criteria of $|\log_2(FC)| \geq 0.5$ and an adjusted $p < 0.05$ (Figure 2A). A total of 507 differentially expressed lncRNAs were identified, including 373 upregulated and 134 downregulated lncRNAs (Figure 2B). Then, a total of 107 lncRNAs related to autophagy were screened using the "Limma" package of R software when the criteria were set to corFilte ≥ 0.3 and $p < 0.05$ (Supplementary File S1).

Identification of autophagy-related lncRNA prognostic signature for lung cancer belonging to lung adenocarcinoma

Univariate Cox regression analysis identified 28 out of 107 lncRNAs as prognostic factors (Figure 3A). Next, multivariate Cox regression was performed and 12 lncRNAs were further found to be independent factors in predicting the prognosis of LUAD patients (Figure 3B). Results of the Kaplan–Meier survival analysis further confirmed the relevance between these 12 lncRNAs and OS (Figure 4). These 12 lncRNAs are NFYC-AS1, SNHG10, HCG18, LINC00857, LINC01116, LINC00996, CRNDE, CASC15, TMPO-AS1, LINC00654, LINC01138, and ZNF790-AS1. An expression correlation network was built on the 12 OS-associated lncRNAs and 16 corresponding associated autophagy-related genes, including FOXO1, FOXO3, MLST8, MTOR, BIRC6, RPS6KB1, CFLAR, IFNG, DAPK2, ITPR1, MBTPS2, ULK2, ATIC, GAPDH, CASP1, NLRC4, PRKCQ, ITGB4, SPHK1,

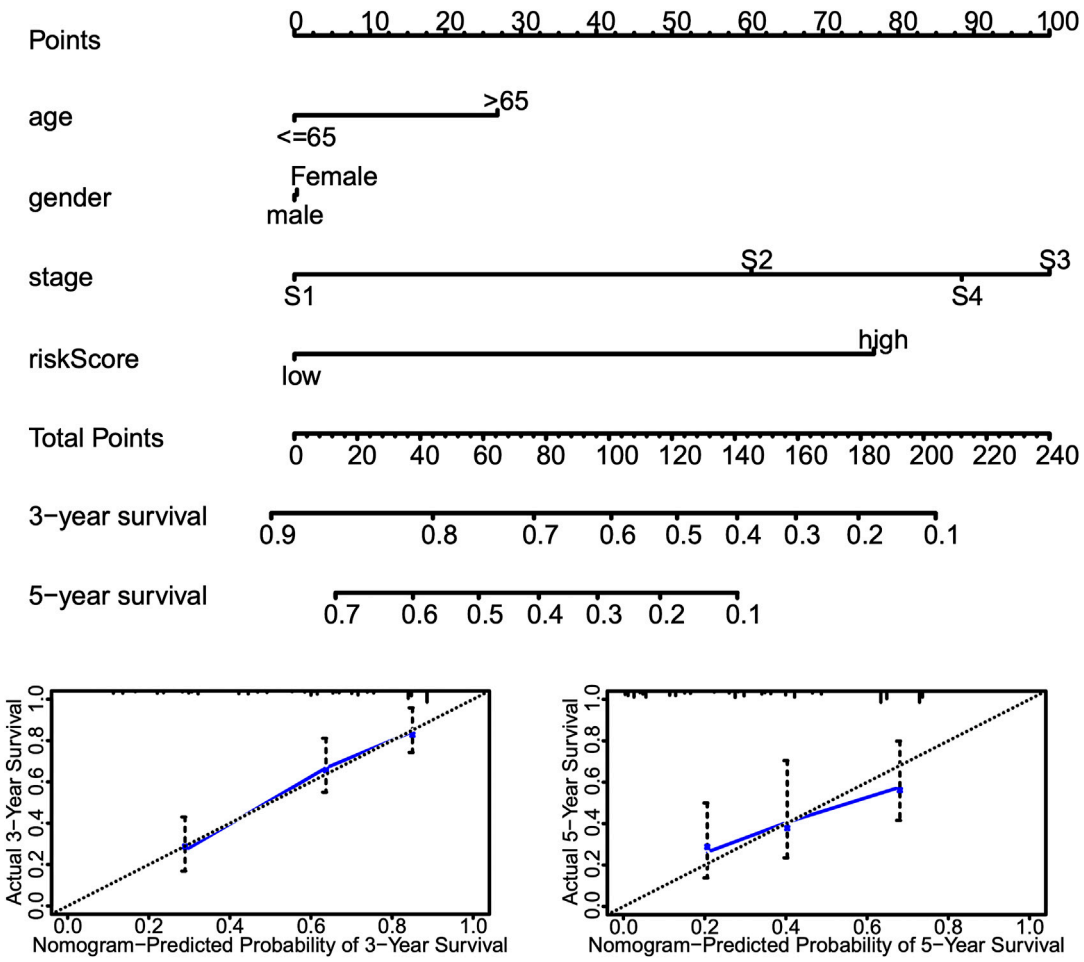


FIGURE 9
Nomogram incorporating risk score and other clinical information. Top, nomograms to predict 3- or 5-year OS of patients with LUAD. Bottom, calibration curves of the nomogram for 3- or 5-year OS prediction.

GABARAPL1, BIRC5, CDKN2A, HDAC1, PARP1, FADD, USP10, and CAPN10 as shown in Figure 5.

Overall survival-associated lncRNA-based risk signature development for lung cancer belonging to lung adenocarcinoma

Multivariate Cox regression was used to develop a risk signature consisting of 12 autophagy-related lncRNAs with the "Survival" package of R to predict the OS of LUAD. Risk score = (0.651*EXP_NFYC-AS1) + (−0.645* EXP_ SNHG10) + (0.644*EXP_HCG18) + (0.254*EXP_LINC00857) + (0.194*EXP_LINC01116) + (−0.664*EXP_LINC00996) + (−0.150*EXP_CRNDE) + (0.213*EXP_CASC15) + (−0.401* EXP_TMPO-AS1) + (0.556*EXP_LINC00654) + (0.537*EXP_LINC01138) + (0.372*EXP_ZNF790-AS1). A total of 458 LUAD patients were divided into high- and low-risk groups based on the median risk score. Kaplan–Meier survival analysis showed a significant OS advantage of the low-risk group over the high-risk group, as shown in Figure 6A ($p < 0.001$). Moreover, more than 0.7 of 1-, 3-, and 5-year area under the curve (AUC) in receiver operating characteristic (ROC) analysis exhibited a satisfactory diagnostic performance (Figure 6A). We further validated this risk signature in both TCGA internal cohort and the GSE validation dataset. Once again, Kaplan–Meier survival analysis showed an improved OS in the low-risk group compared to the high-risk group, as shown in Figures 6B,C. 0.7 of 1-, 3-, and 5-year AUC in ROC analysis also showed a fair sensitivity and specificity of this risk signature. Then, based on the multivariate Cox proportional hazards regression analyses, a nomogram including 12-lncRNA signatures was constructed by the "rms" package in R

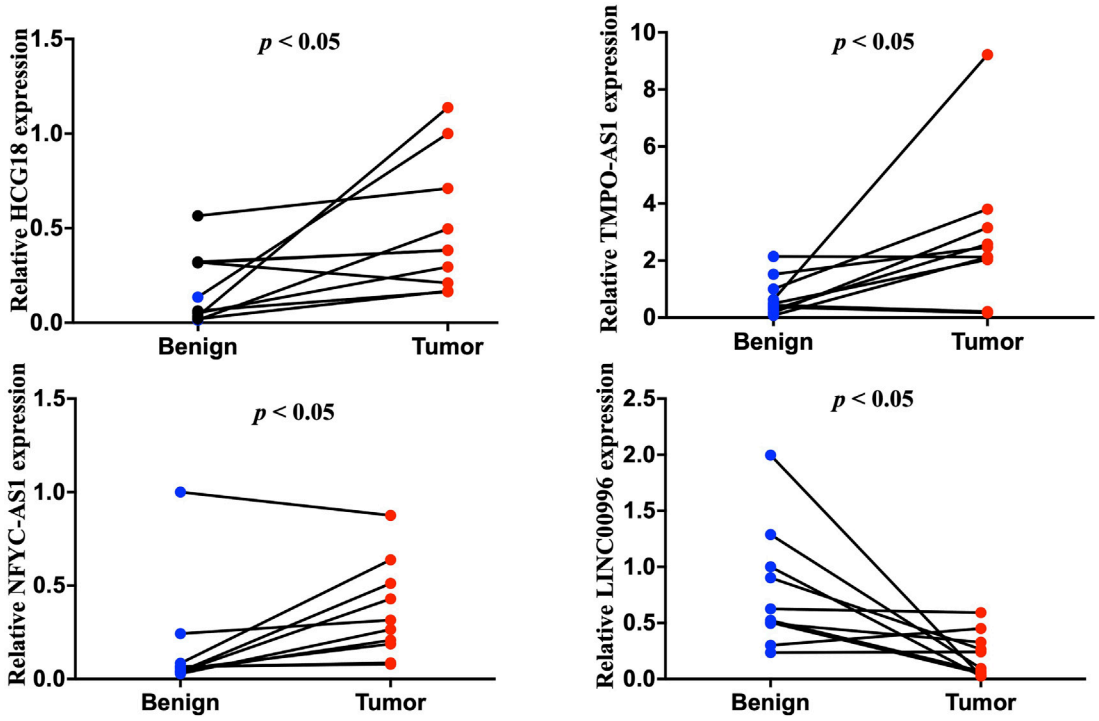


FIGURE 10
Quantitative reverse transcription polymerase reaction results of four lncRNAs, including HCG18, TMPO-AS1, NFYC-AS1, and LNC00996, from ten LUAD patients.

software to help predict the 3- or 5-year survival of LUAD patients (Figure 6D).

Nomogram development and validation

Univariate and multivariate Cox proportional hazards regression analyses were performed to evaluate the predictive value of different clinical parameters. Univariate Cox regression analysis showed that stage, T status, N status, and risk score were significantly associated with OS of LUAD patients in TCGA datasets, as shown in Figure 7A ($p < 0.001$). On the other hand, multivariate Cox regression analysis only showed that stage and risk score are the independent predictive factor of OS for LUAD patients (Figure 7B). In the GSE31210 dataset, univariate Cox regression analysis showed that the stage and risk score are significantly related to the OS of LUAD patients (Figure 7C). Multivariate Cox regression showed that stage and risk score are the independent predictive factors of OS (Figure 7D). Also, risk signature based on 12 lncRNA showed 1-, 3-, and 5-year AUC of above 0.69 in ROC analysis along with other clinical factors in both TCGA and GEO datasets (Figure 8). Then, we constructed a nomogram incorporating 12-lncRNA signatures and other clinical information by the "rms" package in R software to predict the 3- and 5-year survival of LUAD patients. The following calibration curve showed a satisfactory consistency between the predictive and observed results (Figure 9).

Validation of the expression of lncRNAs in lung cancer belong to lung adenocarcinoma patient samples and GSE40791

The expression of selected lncRNAs from our constructed signatures (HCG18, TMPO-AS1, NFYC-AS1, LNC00996) was investigated in 10 paired tumor and adjacent tissues from LUAD patients using qRT-PCR. The results indicated that HCG18, TMPO-AS1, and NFYC-AS1 were significantly upregulated in tumors compared with adjacent samples. In contrast, LNC00996 were significantly downregulated in tumors compared with adjacent samples (Figure 10) ($p < 0.05$). These results were consistent with the expression pattern from the TCGA LUAD dataset (Supplementary Figure S1). Also, results from the GSE40791 dataset, including 100 normal lung and 94 LUAD lung tissues, further validate the expression pattern of these lncRNAs (Supplementary Figure S2).

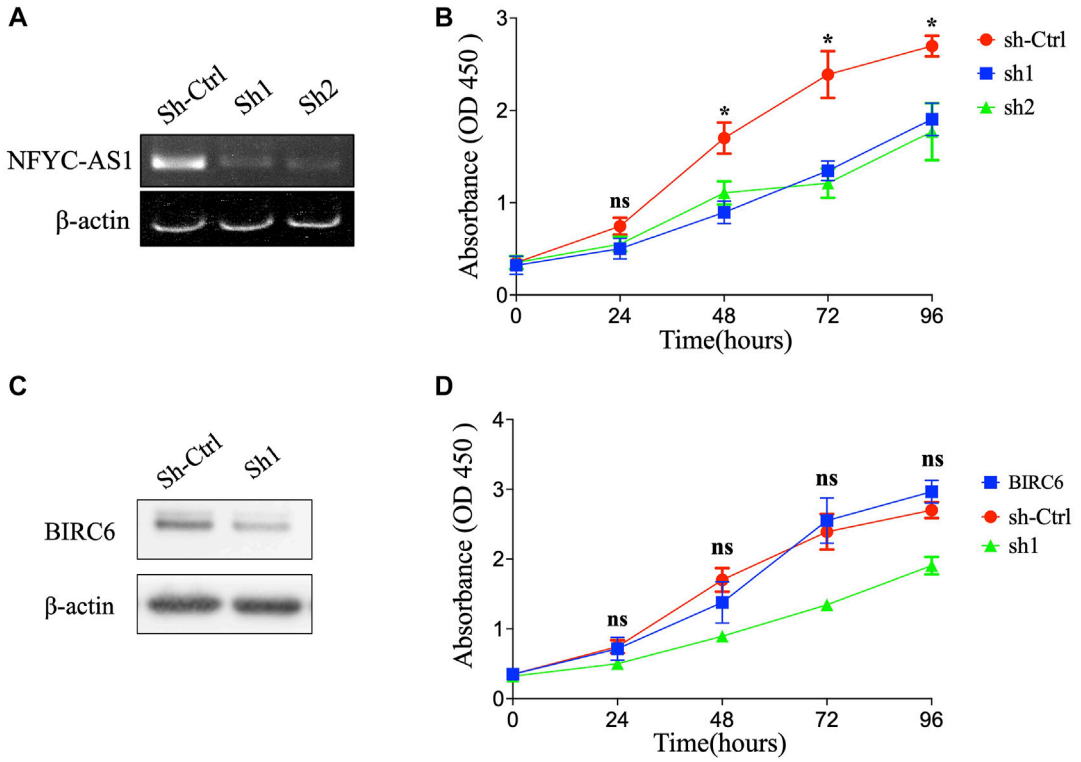


FIGURE 11
The influence of NFYC-AS1 and BIRC6 on the proliferation of LUAD cell lines. **(A)** RT-PCR results of NFYC-AS1 expression in A549 cells with or without shRNA knockdown. **(B)** The cell proliferation of A549 cells with or without knockdown of NFYC-AS1 expression. **(C)** Immunoblot of BIRC6 expression in A549 cells with or without knockdown of NFYC-AS1 expression. **(D)** The cell proliferation of A549 cells with knockdown of NFYC-AS1 expression or the rescue of BIRC6 expression.

lncRNA NFYC-AS1 promotes lung adenocarcinoma cell proliferation through BIRC6

To investigate the possible connection and its influence between lncRNAs and related autophagy genes, we selected two potential players, NFYC-AS1 and BIRC6, based on previous literature (Song et al., 2021) and their high degree of correlation (Figure 5; Supplementary Figure S3A). Correlation analysis from GEPIA confirmed a positive correlation between the NFYC-AS1 and BIRC6 expression in TCGA LUAD dataset ($p < 0.001$, R = 0.35) but a relatively weak correlation in GTEx lung dataset ($p < 0.05$, R = 0.15) (Supplementary Figures S3B,C). Immunoblot results from lung adenocarcinoma cell line showed that knockdown of NFYC-AS1 did reduce the BIRC6 protein level (Figures 11A,C), suggesting a possible regulation of BIRC6 expression by NFYC-AS1. To determine the subsequent effect of this regulation, we explored the influence of NFYC-AS1 and BIRC6 on A549 cells. Results show that the knockdown of NFYC-AS1 clearly inhibits cell growth (Figure 11B), whereas the restoration of BIRC6 expression reverses the inhibitory effect of NFYC-AS1 knockdown (Figure 11D), suggesting that the pro-proliferative effect of NFYC-AS1 is *via* the regulation of BIRC6. More importantly, IHC of BIRC6 from 34 LUAD patients showed significantly lower expression in NFYC-AS1 low or normal lung tissues and a markedly elevated expression in tumor tissues with high NFYC-AS1 expression (Figures 12A,B,C). In addition, mRNA expression pattern of BIRC6 from the TCGA database was consistent with the above results from patient samples (Figure 12D), both indicating that NFYC-AS1 may promote tumor progress and act as a high-risk factor of poor prognosis in our constructed signature through the upregulation of BIRC6.

Given that BIRC6 is known to be the regulator of both autophagy and apoptosis (Ebner et al., 2018), we investigated the influence of this regulation on cell autophagy and apoptosis. Results reveal that the knockdown of NFYC-AS1 indeed alters the expression of some autophagy-related markers, including LC3B, Beclin 1, and SQSTM1/p62 in A549 cells (Figures 13A,B). Furthermore, the knockdown of NFYC-AS1 also induced evident apoptosis in A549 cells (Figures 14A,B), whereas BIRC6 overexpression seemingly offset this effect (Figures 14C,D). More importantly, we performed *in vivo* tumor xenograft experiment in nude mice and found that the knockdown of NFYC-AS1 did exhibit the inhibitory effect on

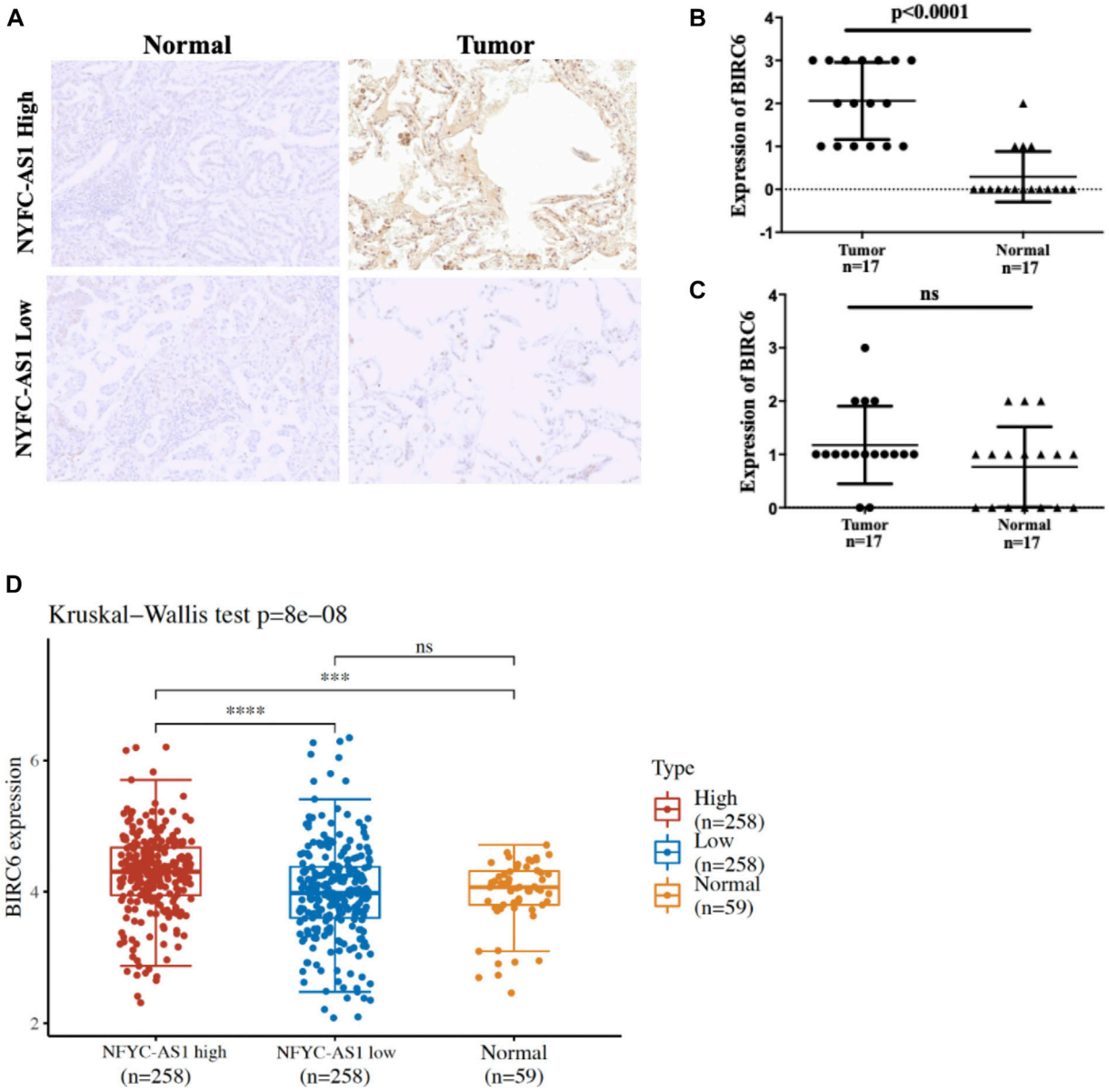


FIGURE 12
Immunohistochemistry results of 34 pair of tissues from LUAD patients. **(A)** Representative results of BIRC6 expression in normal and tumor tissues based on the median of NFYC-AS1 expression level. **(B)** The quantitative scores of above IHC results based on the higher level of NFYC-AS1 expression. **(C)** The quantitative scores of above IHC results based on the lower level of NFYC-AS1 expression. **(D)** BIRC6 expressions from TCGA database based on the median of NFYC-AS1 expression level.

tumor growth, whereas the overexpression of BIRC6 could restore the tumor growth in nude mice (Supplementary Figure S5).

Discussion

As the most common pathological subtype of lung cancer, the proportion of lung adenocarcinoma (LUAD) is up to 40%–70% of all lung cancer patients. Although certain progress has been made in the research and practice of LUAD treatment, it is still one of the most difficult types of cancers to treat due to its highly metastatic and malignant nature. The dysregulation of autophagy is thought to play an important role in lung cancer. For example, in a mouse model of lung cancer, upregulation of autophagy *via* caspase-3 and mTOR inhibition promotes the efficacy of radiotherapy. C-myc/miR-150/EPG5 axis mediated dysfunction of autophagy induced increased cellular ROS levels

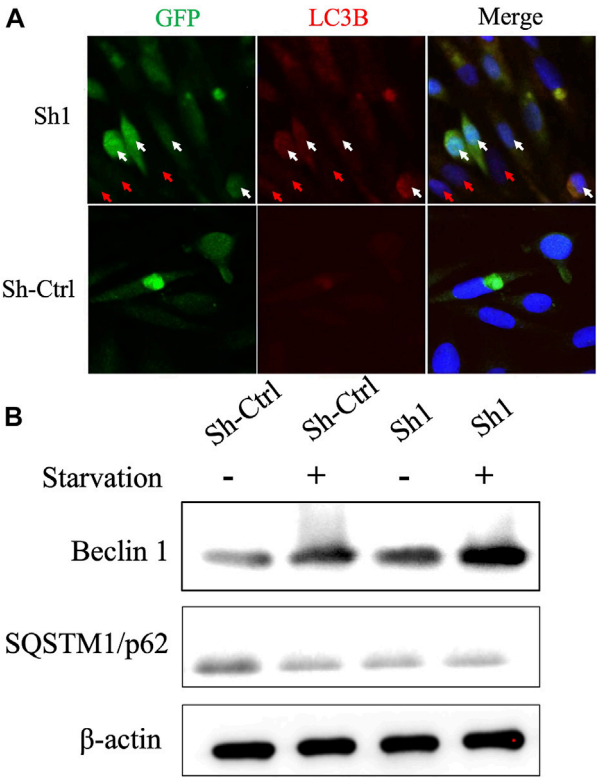


FIGURE 13
The influence of NFYC-AS1 on the autophagy of A549 cells. **(A)** Immunofluorescence analysis of key autophagy marker LC3B in A549 cells with or without knockdown of NFYC-AS1 expression. The white arrow indicates the significant expression of shRNA plasmids, whereas the red arrow indicates a less significant expression. **(B)** Immunoblot of Beclin 1 and SQSTM1/p62 expression in A549 cells with or without knockdown of NFYC-AS1 expression.

and DNA damage response and promoted NSCLC development (Li et al., 2019a). Also, autophagy could protect LUAD cells to Src inhibitors, whereas microRNA-106a could target autophagy kinase ULK1 and compromise this protective effect (Rothschild et al., 2017). In addition, autophagy inhibition of cancer stem cells could promote the efficacy of cisplatin against NSCLC in both A549 cell lines and NOD/SCID mice (Hao et al., 2019).

LncRNA, along with miRNA and mRNA expression, has been used to develop risk signatures to prevent and predict the development of various tumors. Increasing evidence has shown significant correlations between lncRNAs and autophagy in lung cancer. For instance, the lncRNA BLACAT1 promoted ATG7 expression through miR-17, facilitated autophagy, and promoted the chemoresistance of NSCLC cells through the miR-17/ATG7 axis (Huang et al., 2019). lncRNA CASC2 inhibited autophagy and promoted apoptosis in NSCLC cells *via* the miR-214/TRIM16 signaling pathway (Li et al., 2018a). LncRNA MSTO2P promotes lung cancer cell proliferation and autophagy by upregulating EZH2 (Wang et al., 2019). Moreover, lncRNA LCPAT1 mediates smoking or PM 2.5-induced autophagy and epithelial–mesenchymal transition *via* RCC2, implying the possible role of lncRNA in rendering lung cancer cells into a more invasive state (Lin et al., 2018).

In this study, we identified and validated an autophagy-related 12-lncRNA risk signature that was highly associated with the OS of patients with LUAD. The signature showed accuracy and robustness by calculating the AUC in the ROC analysis. Multivariate Cox regression analyses suggested that age, N stage, and the risk signature were independent risk factors for OS in the primary cohort. A novel nomogram incorporating clinical factors was constructed and validated to predict the prognosis for LUAD patients. The expression levels of lncRNA signatures were also validated in the clinical LUAD samples by qRT-PCR. These results suggested that both the 12-lncRNA risk signature and the novel nomogram were effective prognostic indicators in patients with LUAD.

SNHG10 has been identified as oncogenic in bladder cancer (Jiang et al., 2018), hepatocellular carcinoma (Lan et al., 2019), osteosarcoma (He et al., 2020), glioma (Jin et al., 2020), non–small cell lung cancer (Zhang et al., 2020b; Liang et al., 2020), acute myeloid leukemia (Xiao et al., 2021), and gastric cancer (Yuan et al., 2021; Zhang et al., 2021). HCG18 was found to be involved in several tumors (Xi et al., 2017; Li et al., 2018b; Wang et al., 2018; Li et al., 2019b; Li et al., 2020a; Liu et al., 2020; Ma et al., 2020; Zhang and Lou, 2020; Zou et al., 2020; Yang et al., 2021; Zhu et al., 2021) and recognized as the driver of LUAD (Li et al., 2020b). It is worth mentioning that although HCG18 is linked to a better prognosis in our risk signature, its expression is higher in tumors compared to adjacent normal tissues, which may be due to the compensatory effect of this lncRNA as a potential tumor suppressor. It has been shown that LINC00857 enhances BIRC5-dependent radio-sensitivity of cancer cells (Han et al., 2020), regulates apoptosis and glycolysis (Wang et al., 2020b), and is identified as one of the prognostic markers for the early stage lung adenocarcinoma (Mu et al., 2021). LINC01116 is oncogenic in several other kinds of tumors (Hu et al., 2018; Chen et al., 2020; Wu et al., 2020; Ye et al., 2020) and promotes lung adenocarcinoma proliferation and metastasis possibly *via* the Akt pathway (Zeng et al., 2020; Shang et al., 2021; Wu et al., 2021). The prognostic function of LINC00996 is also implicated in colorectal cancer (Ge et al., 2018), head and neck cancer (Ge et al., 2018), and multiple myeloma (Zhou et al., 2020). Overexpressed CRNDE was observed in colorectal cancer (Yu et al., 2017), glioma (Wang et al., 2015), hepatocellular carcinoma (Ji et al., 2019), breast cancer (Huan et al., 2017), and gastric cancer (Hu et al., 2017) and was found to be related to the metastasis and radiosensitivity in lung cancer (Zhang et al., 2018; Jing et al., 2019). CASC15 is another unfavorable prognostic marker found in multiple

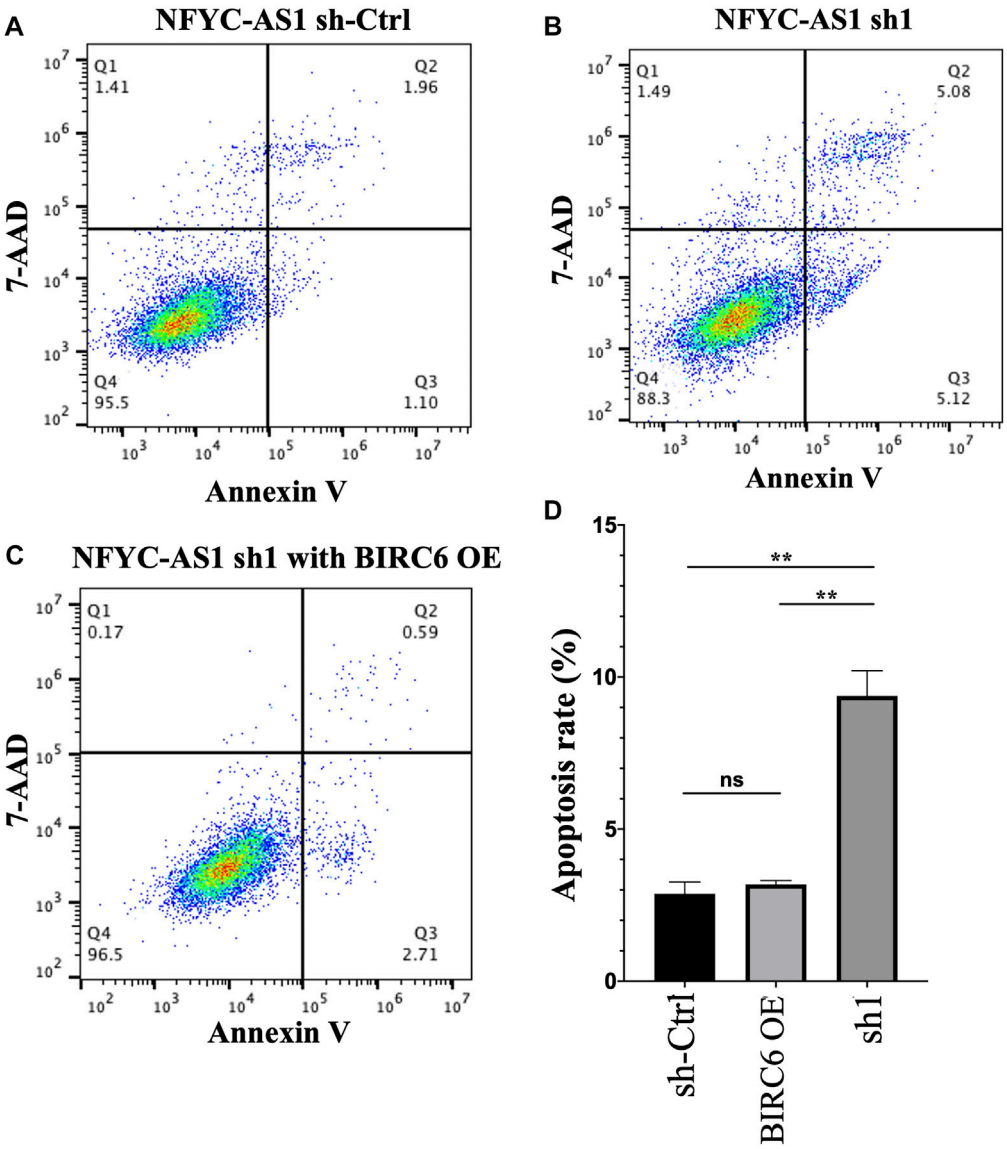


FIGURE 14
The influence of NFYC-AS1 on the apoptosis of A549 cells. **(A)** Representative flow cytometry analysis of apoptosis in A549 cells without the knockdown of NFYC-AS1 expression. **(B)** Representative flow cytometry analysis of apoptosis in A549 cells with the knockdown of NFYC-AS1 expression. **(C)** Representative flow cytometry analysis of apoptosis in A549 cells with the knockdown of NFYC-AS1 expression combined with the rescue of BIRC6 expression. **(D)** Statistic results of apoptosis rate in A549 cells with or without the knockdown of NFYC-AS1 expression or combining the NFYC-AS1 knockdown with BIRC6 overexpression.

tumors, including lung cancer (Lessard et al., 2015; Fernando et al., 2017; He et al., 2017; Yao et al., 2017; Wu et al., 2018; Bai et al., 2019; Li et al., 2019c; Xie and Cheng, 2019; Yu et al., 2019; Liang et al., 2021; Yu et al., 2021). Previous studies have found that TMPO-AS1 is upregulated and correlated with poor prognosis in LUAD patients (Qin et al., 2019). Bioinformatics analysis also revealed that TMPO-AS1 could affect the prognosis of LUAD by regulating cell cycle pathway genes such as CDC25A (Peng et al., 2017).

Compared to other members in the risk signature, NFYC-AS1 received relatively little attention, except that it has been demonstrated as a prognostic biomarker in the miRNA-lncRNA-mRNA regulatory network in LUAD patients (Li et al., 2016). BIRC6 (Baculoviral IAP repeat-containing 6) is an IAP (inhibitor of apoptosis proteins) that not only plays an anti-apoptotic role through the inhibition of pro-apoptotic protein SMAC or caspases (Hauser et al., 1998; Hao et al., 2004; Qiu and Goldberg, 2005; Ren et al., 2005) but also is

associated with other cellular processes; for instance, it regulates autophagosome-lysosome fusion in autophagy (Ikeda, 2018; Jia and Bonifacino, 2019). In addition, overexpression of BIRC6 has been found in various tumors, including colorectal cancer (Bianchini et al., 2006), gastric carcinomas (Salehi et al., 2016), and lung cancer (Dong et al., 2013). In our study, the NFYC-AS1 expression shows a high positive correlation with autophagy genes, especially with BIRC6 ($p < 0.001$, R = 0.35), and the knockdown of NFYC-AS1 decreased the BIRC6 expression and induces both autophagy and apoptosis in A549 cells, whereas the rescue of BIRC6 reverses the inhibitory effect (Figures 13, 14). Although there is no evidence showing a direct regulation of BIRC6 by NFYC-AS1 until now, combining all the above results, it is possible that NFYC-AS1 enhances BIRC6 expression *via* binding to specific proteins or regulating miRNAs, which in turn leads to the inhibition of autophagy, apoptosis, and the progression of tumors.

Gene set enrichment analysis (GSEA) further revealed that the constructed risk signature may regulate several KEGG pathways to influence the progression of LUAD, including the "p53 signaling," "proteasome," "protein export," and "pentose phosphate" pathways. Tumor suppressor p53 acts as the guardian of the genome, which senses the stress in the cellular environment, initiates DNA repair, or leads to the cell cycle arrest. Recently, growing evidence has revealed that p53 can activate the transcription of autophagy-related genes, whereas autophagy could suppress p53 *via* the regulation of ROS or AMPK. In contrast, proteasome is considered responsible for proteolysis and organelle homeostasis by interconnecting with autophagy pathways. As a crucial part of autophagy, lysosomes carry out the degradation of extracellular particles from endocytosis and of intracellular components from autophagy, which contains more than 50 membrane proteins that control the specificity and timing of cargo influx and protein export. The pentose phosphate pathway (PPP) plays a key role in facilitating tumor cells with the glycolytic process and countering the damage of reactive oxidative species. G6PD is the limiting enzyme of the PPP associated with cancer progression and drug resistance, the inhibition of which indirectly leads to the dysfunction of autophagy. Although GESA analysis implicated the involvement of lncRNAs and autophagy in LUAD tumor progression, the detailed association of how autophagy-related tumor progression and lncRNAs are included in our signatures should be further investigated.

This study has some obvious advantages. First, it included 551 samples from the LUAD TCGA dataset and 246 samples from GSE31210, providing a sufficient sample size to avoid statistical bias and make a satisfactory result. Indeed, the constructed signature and incorporated nomogram were further validated in both TCGA train and GSE datasets and showed a good specialty and sensitivity. Second, clinical samples further confirmed that the expressions of several lncRNAs are consistent with the results from the database. Then, a novel nomogram incorporating the risk signature and related clinicopathological factors is constructed and provides a helpful prediction for the prognosis of LUAD patients. Third, we further investigated the possible regulation of BIRC6 by NFYC-AS1 and related influences on lung cancer cells, proving that NFYC-AS1 may act not only as an indicator for poor prognosis but also as a potential driver of carcinogenesis through the negative modulation of autophagy and apoptosis. Although, in current experimental settings, it is difficult to distinguish which effect is dominant or whether BIRC6 is directly responsible for this inhibitory effect through autophagy, all the above results suggest the importance of NFYC-AS1 as a pro-cancer factor in LUAD patients.

In the meantime, there are some limitations to the current study. First, datasets from GEO or TCGA have various kinds of lncRNAs due to their unique sequencing methods. Thus, on the one hand, using both GEO and TCGA datasets may increase the sample size and help the validation of risk signatures. On the other hand, the overlap of two different datasets may neglect some lncRNAs that are significantly deregulated in one database but not sequenced at all in another. Second, the TCGA and GSE31210 databases did not cover important information such as medical history, treatment strategy, and family history, which could alter the prognosis of LUAD patients. Third, although our results showed a possible regulation of NFYC-AS1 on BIRC6 and BIRC6 expression indeed increased in tumor tissues (Supplementary Figure S4), the direct evidence of this regulation still needs to be further discovered. Last, in the current study, we investigated the expression of four lnRNAs in 10 patients' samples. To assure the accuracy of the expression patterns, clearly more clinical samples should be analyzed.

Conclusion

Collectively, our current study analyzed autophagy-related lncRNAs in LUAD patients comprehensively and constructed a lncRNA-based risk signature. Our results revealed that high-risk scores are associated with a worse prognosis in LUAD patients. Our study also suggested that the 12 autophagy-related lncRNAs have significant value in predicting the prognosis of LUAD patients. Furthermore, NFYC-AS1 and BIRC6 may be potential therapeutic targets due to previous research and their significance in our study. More specific experiments and research based on clinical samples should be carried out to validate the results of our study.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Chongqing Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

TF and XL performed most of the research work. SX collected clinical samples and carried out some of the experiments. MZ designed the project and wrote the manuscript. All authors reviewed and approved the final manuscript for submission.

Funding

This work was supported by the Xinglin program of Chongqing TCM/TCM-integrated Key discipline.

Acknowledgments

We acknowledge TCGA and GEO databases for providing their platforms and contributors for uploading their meaningful datasets.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.834935/full#supplementary-material

References

Bai, Y., Zhang, G., Cheng, R., Yang, R., and Chu, H. (2019). CASC15 contributes to proliferation and invasion through regulating miR-766-5p/KLK12 axis in lung cancer. *Cell. Cycle* 18, 2323–2331. doi:10.1080/15384101.2019.1646562

Bianchini, M., Levy, E., Zucchini, C., Pinski, V., Macagno, C., De Sanctis, P., et al. (2006). Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. *Int. J. Oncol.* 29, 83–94. doi:10.3892/ijo.29.1.83

Chen, J., Yuan, Z. H., Hou, X. H., Shi, M. H., and Jiang, R. (2020). LINC01116 promotes the proliferation and inhibits the apoptosis of gastric cancer cells. *Eur. Rev. Med. Pharmacol. Sci.* 24, 1807–1814. doi:10.26355/eurrev_202002_20358

Dong, X., Lin, D., Low, C., Vucic, E. A., English, J. C., Yee, J., et al. (2013). Elevated expression of BIRC6 protein in non-small-cell lung cancers is associated with cancer recurrence and chemoresistance. *J. Thorac. Oncol.* 8, 161–170. doi:10.1097/JTO.0b013e31827d5237

Ebner, P., Poetsch, I., Deszcz, L., Hoffmann, T., Zuber, J., and Ikeda, F. (2018). The IAP family member BRUCE regulates autophagosome-lysosome fusion. *Nat. Commun.* 9, 599. doi:10.1038/s41467-018-02823-x

Fernando, T. R., Contreras, J. R., Zampini, M., Rodriguez-Malave, N. I., Alberti, M. O., Anguiano, J., et al. (2017). The lncRNA CASC15 regulates SOX4 expression in RUNX1-rearranged acute leukemia. *Mol. Cancer* 16, 126. doi:10.1186/s12943-017-0692-x

Fulda, S., and Kögel, D. (2015). Cell death by autophagy: Emerging molecular mechanisms and implications for cancer therapy. *Oncogene* 34, 5105–5113. doi:10.1038/onc.2014.458

Ge, H., Yan, Y., Wu, D., Huang, Y., and Tian, F. (2018). Potential role of LINC00996 in colorectal cancer: A study based on data mining and bioinformatics. *Onco. Targets. Ther.* 11, 4845–4855. doi:10.2147/ott.s173225

Gewirtz, D. A. (2014). The four faces of autophagy: Implications for cancer therapy. *Cancer Res.* 74, 647–651. doi:10.1158/0008-5472.can-13-2966

Guan, H., Zhu, T., Wu, S., Liu, S., Liu, B., Wu, J., et al. (2019). Long noncoding RNA LINC00673-v4 promotes aggressiveness of lung adenocarcinoma via activating WNT/β-catenin signaling. *Proc. Natl. Acad. Sci. U. S. A.* 116, 14019–14028. doi:10.1073/pnas.1900997116

Han, F., Yang, S., Wang, W., Huang, X., Huang, D., and Chen, S. (2020). Silencing of lncRNA LINC00857 enhances BIRC5-dependent radio-sensitivity of lung adenocarcinoma cells by recruiting NF-κB1. *Mol. Ther. Nucleic Acids* 22, 981–993. doi:10.1016/j.omtn.2020.09.020

Han, J., Hou, W., Goldstein, L. A., Stolz, D. B., Watkins, S. C., and Rabinowich, H. (2014). A complex between Atg7 and caspase-9: A novel mechanism of cross-regulation between autophagy and apoptosis. *J. Biol. Chem.* 289, 6485–6497. doi:10.1074/jbc.M113.536854

Hao, C., Liu, G., and Tian, G. (2019). Autophagy inhibition of cancer stem cells promotes the efficacy of cisplatin against non-small cell lung carcinoma. *Ther. Adv. Respir. Dis.* 13, 1753466619866097. doi:10.1177/1753466619866097

Hao, Y., Sekine, K., Kawabata, A., Nakamura, H., Ishioka, T., Ohata, H., et al. (2004). Apollon ubiquitinates SMAC and caspase-9, and has an essential cytoprotection function. *Nat. Cell. Biol.* 6, 849–860. doi:10.1038/ncb1159

Hauser, H-P., Bardroff, M., Pyrowolakis, G., and JentSch, S. (1998). A giant ubiquitin-conjugating enzyme related to IAP apoptosis inhibitors. *J. Cell. Biol.* 141, 1415–1422. doi:10.1083/jcb.141.6.1415

He, P., Xu, Y., and Wang, Z. (2020). LncRNA SNHG10 increases the methylation of miR-218 gene to promote glucose uptake and cell proliferation in osteosarcoma. *J. Orthop. Surg. Res.* 15, 353. doi:10.1186/s13018-020-01865-6

He, T., Zhang, L., Kong, Y., Huang, Y., Zhang, Y., Zhou, D., et al. (2017). Long non-coding RNA CASC15 is upregulated in hepatocellular carcinoma and

facilitates hepatocarcinogenesis. *Int. J. Oncol.* 51, 1722–1730. doi:10.3892/ijo.2017.4175

Hirata, H., Hinoda, Y., Shahryari, V., Deng, G., Nakajima, K., Tabatabai, Z. L., et al. (2015). Long noncoding RNA MALAT1 promotes aggressive renal cell carcinoma through Ezh2 and interacts with miR-205. *Cancer Res.* 75, 1322–1331. doi:10.1158/0008-5472.can-14-2931

Hu, C. E., Du, P. Z., Zhang, H. D., and Huang, G. J. (2017). Long noncoding RNA CRNDE promotes proliferation of gastric cancer cells by targeting miR-145. *Cell. Physiol. biochem.* 42, 13–21. doi:10.1159/000477107

Hu, H. B., Chen, Q., and Ding, S. Q. (2018). LncRNA LINC01116 competes with miR-145 for the regulation of ESR1 expression in breast cancer. *Eur. Rev. Med. Pharmacol. Sci.* 22, 1987–1993. doi:10.26355/eurrev_201804_14726

Huan, J., Xing, L., Lin, Q., Xui, H., and Qin, X. (2017). Long noncoding RNA CRNDE activates Wnt/β-catenin signaling pathway through acting as a molecular sponge of microRNA-136 in human breast cancer. *Am. J. Transl. Res.* 9, 1977–1989.

Huang, F. X., Chen, H. J., Zheng, F. X., Gao, Z. Y., Sun, P. F., Peng, Q., et al. (2019). LncRNA BLACAT1 is involved in chemoresistance of non-small cell lung cancer cells by regulating autophagy. *Int. J. Oncol.* 54, 339–347. doi:10.3892/ijo.2018.4614

Ikeda, F. (2018). The anti-apoptotic ubiquitin conjugating enzyme BIRC6/BRUCE regulates autophagosome-lysosome fusion. *Autophagy* 14, 1283–1284. doi:10.1080/15548627.2018.1471311

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi:10.1038/ng.3192

Ji, D., Jiang, C., Zhang, L., Liang, N., Jiang, T., Yang, B., et al. (2019). LncRNA CRNDE promotes hepatocellular carcinoma cell proliferation, invasion, and migration through regulating miR-203/BCAT1 axis. *J. Cell. Physiol.* 234, 6548–6560. doi:10.1002/jcp.27396

Jia, R., and Bonifacino, J. S. (2019). Negative regulation of autophagy by UBA6-BIRC6–mediated ubiquitination of LC3. *Elife* 8, e50034. doi:10.7554/eLife.50034

Jiang, B., Hailong, S., Yuan, J., Zhao, H., Xia, W., Zha, Z., et al. (2018). Identification of oncogenic long noncoding RNA SNHG12 and DUXAP8 in human bladder cancer through a comprehensive profiling analysis. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 108, 500–507. doi:10.1016/j.biopha.2018.09.025

Jin, L., Huang, S., Guan, C., and Chang, S. (2020). ETS1-activated SNHG10 exerts oncogenic functions in glioma via targeting miR-532-3p/FBXL19 axis. *Cancer Cell. Int.* 20, 589. doi:10.1186/s12935-020-01649-2

Jing, H., Xia, H., Qian, M., and Lv, X. (2019). Long noncoding RNA CRNDE promotes non-small cell lung cancer progression via sponging microRNA-338-3p. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 110, 825–833. doi:10.1016/j.biopha.2018.12.024

Klionsky, D. J. (2007). Autophagy: From phenomenology to molecular understanding in less than a decade. *Nat. Rev. Mol. Cell. Biol.* 8, 931–937. doi:10.1038/nrm2245

Lan, T., Yuan, K., Yan, X., Xu, L., Liao, H., Hao, X., et al. (2019). LncRNA SNHG10 facilitates hepatocarcinogenesis and metastasis by modulating its homolog SCARNA13 via a positive feedback loop. *Cancer Res.* 79, 3220–3234. doi:10.1158/0008-5472.can-18-4044

Lessard, L., Liu, M., Marzese, D. M., Wang, H., Chong, K., Kawas, N., et al. (2015). The CASC15 long intergenic noncoding RNA locus is involved in melanoma progression and phenotype switching. *J. Invest. Dermatol.* 135, 2464–2474. doi:10.1038/jid.2015.200

Li, D. S., Ainiwaer, J. L., Sheyhiding, I., Zhang, Z., and Zhang, L. W. (2016). Identification of key long non-coding RNAs as competing endogenous RNAs for miRNA-mRNA in lung adenocarcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 20, 2285–2295.

Li, H., Liu, J., Cao, W., Xiao, X., Liang, L., Liu-Smith, F., et al. (2019). C-myc/miR-150/EPG5 axis mediated dysfunction of autophagy promotes development of non-small cell lung cancer. *Theranostics* 9, 5134–5148. doi:10.7150/thno.34887

Li, L., Ma, T. T., Ma, Y. H., and Jiang, Y. F. (2019). LncRNA HCG18 contributes to nasopharyngeal carcinoma development by modulating miR-140/CCND1 and Hedgehog signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* 23, 10387–10399. doi:10.26355/eurrev_201912_19678

Li, M., Chen, Y., Zhu, J., Gao, Z., Wang, T., and Zhou, P. (2019). Long noncoding RNA CASC15 predicts unfavorable prognosis and exerts oncogenic functions in non-small cell lung cancer. *Am. J. Transl. Res.* 11, 4303–4314.

Li, M., and Izpisua Belmonte, J. C. (2015). Roles for noncoding RNAs in cell-fate determination and regeneration. *Nat. Struct. Mol. Biol.* 22, 2–4. doi:10.1038/nsmb.2946

Li, Q., Chen, K., Dong, R., and Lu, H. (2018). LncRNA CASC2 inhibits autophagy and promotes apoptosis in non-small cell lung cancer cells via regulating the miR-214/TRIM16 axis. *RSC Adv.* 8, 40846–40855. doi:10.1039/c8ra09573f

Li, S., Wu, T., Zhang, D., Sun, X., and Zhang, X. (2020). The long non-coding RNA HCG18 promotes the growth and invasion of colorectal cancer cells through sponging miR-1271 and upregulating MTDH/Wnt/β-catenin. *Clin. Exp. Pharmacol. Physiol.* 47, 703–712. doi:10.1111/1440-1681.13230

Li, W., Pan, T., Jiang, W., and Zhao, H. (2020). HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung adenocarcinoma. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 129, 110217. doi:10.1016/j.biopha.2020.110217

Li, Y. J., Lei, Y. H., Yao, N., Wang, C. R., Hu, N., Ye, W. C., et al. (2017). Autophagy and multidrug resistance in cancer. *Chin. J. Cancer* 36, 52. doi:10.1186/s40880-017-0219-2

Li, Z., Li, X., Chen, C., Li, S., Shen, J., Tse, G., et al. (2018). Long non-coding RNAs in nucleus pulposus cell function and intervertebral disc degeneration. *Cell. Prolif.* 51, e12483. doi:10.1111/cpr.12483

Liang, C., Wang, J., Liu, A., and Wu, Y. (2021). Tumor promoting long non-coding RNA CASC15 affects HMGB2 expression by sponging miR-582-5p in colorectal cancer. *J. Gene Med.* 24, e3308. doi:10.1002/jgm.3308

Liang, M., Wang, L., Cao, C., Song, S., and Wu, F. (2020). LncRNA SNHG10 is downregulated in non-small cell lung cancer and predicts poor survival. *BMC Pulm. Med.* 20, 273. doi:10.1186/s12890-020-01281-w

Lin, H., Zhang, X., Feng, N., Wang, R., Zhang, W., Deng, X., et al. (2018). LncRNA LCPAT1 mediates smoking/particulate matter 2.5-induced cell autophagy and epithelial-mesenchymal transition in lung cancer cells via RCC2. *Cell. Physiol. biochem.* 47, 1244–1258. doi:10.1159/000490220

Liu, G., Pei, F., Yang, F., Li, L., Amin, A. D., Liu, S., et al. (2017). Role of autophagy and apoptosis in non-small-cell lung cancer. *Int. J. Mol. Sci.* 18 (2), 367. doi:10.3390/ijms18020367

Liu, Y., Lin, W., Dong, Y., Li, X., Lin, Z., Jia, J., et al. (2020). Long noncoding RNA HCG18 up-regulates the expression of WIPF1 and YAP/TAZ by inhibiting miR-141-3p in gastric cancer. *Cancer Med.* 9, 6752–6765. doi:10.1002/cam4.3288

Ma, F., An, K., and Li, Y. (2020). Silencing of long non-coding RNA-HCG18 inhibits the tumorigenesis of gastric cancer through blocking PI3K/Akt pathway. *Onco. Targets. Ther.* 13, 2225–2234. doi:10.2147/ott.s240965

Mu, L., Ding, K., Tu, R., and Yang, W. (2021). Identification of 4 immune cells and a 5-lncRNA risk signature with prognosis for early-stage lung adenocarcinoma. *J. Transl. Med.* 19, 127. doi:10.1186/s12967-021-02800-x

Peng, F., Wang, R., Zhang, Y., Zhao, Z., Zhou, W., Chang, Z., et al. (2017). Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer* 16, 98. doi:10.1186/s12943-017-0666-z

Peng, Z., Wang, J., Shan, B., Li, B., Peng, W., Dong, Y., et al. (2018). The long noncoding RNA LINC00312 induces lung adenocarcinoma migration and vasculogenic mimicry through directly binding YBX1. *Mol. Cancer* 17, 167. doi:10.1186/s12943-018-0920-z

Qin, Z., Zheng, X., and Fang, Y. (2019). Long noncoding RNA TMPO-AS1 promotes progression of non-small cell lung cancer through regulating its natural antisense transcript TMPO. *Biochem. Biophys. Res. Commun.* 516, 486–493. doi:10.1016/j.bbrc.2019.06.088

Qiu, X-B., and Goldberg, A. L. (2005). The membrane-associated inhibitor of apoptosis protein, BRUCE/Apollon, antagonizes both the precursor and mature forms of Smac and caspase-9. *J. Biol. Chem.* 280, 174–182. doi:10.1074/jbc.M411430200

Ren, J., Shi, M., Liu, R., Yang, Q. H., Johnson, T., Skarnes, W. C., et al. (2005). The Birc6 (Bruce) gene regulates p53 and the mitochondrial pathway of apoptosis and is essential for mouse embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* 102, 565–570. doi:10.1073/pnas.0408744102

Rothschild, S. I., Gautschi, O., Batliner, J., Gugger, M., Fey, M. F., and Tschan, M. P. (2017). MicroRNA-106a targets autophagy and enhances sensitivity of lung cancer cells to Src inhibitors. *Lung Cancer* 107, 73–83. doi:10.1016/j.lungcan.2016.06.004

Salehi, S., Jafarian, A. H., Montazer, M., Moghbeli, M., and Forghanifard, M. M. (2016). BRUCE protein, new marker for targeted therapy of gastric carcinoma. *J. Gastrointest. Cancer* 48, 151–155. doi:10.1007/s12029-016-9874-9

Shang, B., Li, Z., Li, M., Jiang, S., Feng, Z., Cao, Z., et al. (2021). Silencing LINC01116 suppresses the development of lung adenocarcinoma via the AKT signaling pathway. *Thorac. Cancer* 12, 2093–2103. doi:10.1111/1759-7714.14042

Song, Y., Du, J., Lu, P., Zou, Q., Zeng, S., Liu, M., et al. (2021). LncRNA NFYC-AS1 promotes the development of lung adenocarcinomas through autophagy, apoptosis, and MET/c-Myc oncogenic proteins. *Ann. Transl. Med.* 9, 1621. doi:10.21037/atm-21-4995

Tamang, S., Acharya, V., Roy, D., Sharma, R., Aryaa, A., Sharma, U., et al. (2019). SNHG12: An LncRNA as a potential therapeutic target and biomarker for human cancer. *Front. Oncol.* 9, 901. doi:10.3389/fonc.2019.00901

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell.* 154, 26–46. doi:10.1016/j.cell.2013.06.020

Wang, J., Gao, J., Chen, Q., Zou, W., Yang, F., Wei, C., et al. (2020). LncRNA LINC01116 contributes to cisplatin resistance in lung adenocarcinoma. *Onco. Targets. Ther.* 13, 9333–9348. doi:10.2147/ott.s244879

Wang, L., Cao, L., Wen, C., Li, J., Yu, G., and Liu, C. (2020). LncRNA LINC00857 regulates lung adenocarcinoma progression, apoptosis and glycolysis by targeting miR-1179/SPAG5 axis. *Hum. Cell.* 33, 195–204. doi:10.1007/s13577-019-00296-8

Wang, L. J., Sun, G. Z., and Chen, Y. F. (2019). LncRNA MSTO2P promotes proliferation and autophagy of lung cancer cells by up-regulating EZH2 expression. *Eur. Rev. Med. Pharmacol. Sci.* 23, 3375–3382. doi:10.26355/eurrev_201904_17701

Wang, W., Zhao, Z., Yang, F., Wang, H., Wu, F., Liang, T., et al. (2018). An immune-related lncRNA signature for patients with anaplastic gliomas. *J. Neurooncol.* 136, 263–271. doi:10.1007/s11060-017-2667-6

Wang, Y., Wang, Y., Li, J., Zhang, Y., Yin, H., and Han, B. (2015). CRNDE, a long-noncoding RNA, promotes glioma cell growth and invasion through mTOR signaling. *Cancer Lett.* 367, 122–128. doi:10.1016/j.canlet.2015.03.027

Wu, J., Chen, Z., Zhang, L., Cao, J., Li, X., Gong, Z., et al. (2020). Knockdown of LINC01116 inhibits cell migration and invasion in head and neck squamous cell carcinoma through epithelial-mesenchymal transition pathway. *J. Cell. Biochem.* 121, 867–875. doi:10.1002/jcb.29331

Wu, L., Wen, Z., Song, Y., and Wang, L. (2021). A novel autophagy-related lncRNA survival model for lung adenocarcinoma. *J. Cell. Mol. Med.* 25, 5681–5690. doi:10.1111/jcmm.16582

Wu, Q., Xiang, S., Ma, J., Hui, P., Wang, T., Meng, W., et al. (2018). Long non-coding RNA CASC15 regulates gastric cancer cell proliferation, migration and epithelial mesenchymal transition by targeting CDKN1A and ZEB1. *Mol. Oncol.* 12, 799–813. doi:10.1002/1878-0261.12187

Xi, Y., Jiang, T., Wang, W., Yu, J., Wang, Y., Wu, X., et al. (2017). Long non-coding HCG18 promotes intervertebral disc degeneration by sponging miR-146a-5p and regulating TRAF6 expression. *Sci. Rep.* 7, 13234. doi:10.1038/s41598-017-13364-6

Xiao, S., Zha, Y., and Zhu, H. (2021). miR-621 may suppress cell proliferation via targeting lncRNA SNHG10 in acute myeloid leukemia. *Cancer Manag. Res.* 13, 2117–2123. doi:10.2147/cmar.s269528

Xie, Y., and Cheng, Y. (2019). Long noncoding RNA CASC15 is upregulated in glioma and facilitates cell proliferation and metastasis via targeting miR-130b-3p. *Eur. Rev. Med. Pharmacol. Sci.* 23, 7475–7481. doi:10.26355/eurrev_201909_18857

Yang, Y., Gong, P., Yao, D., Xue, D., and He, X. (2021). LncRNA HCG18 promotes clear cell renal cell carcinoma progression by targeting miR-152-3p to upregulate RAB14. *Cancer Manag. Res.* 13, 2287–2294. doi:10.2147/cmar.s298649

Yao, X. M., Tang, J. H., Zhu, H., and Jing, Y. (2017). High expression of LncRNA CASC15 is a risk factor for gastric cancer prognosis and promote the proliferation of gastric cancer. *Eur. Rev. Med. Pharmacol. Sci.* 21, 5661–5667. doi:10.26355/eurrev_201712_14010

Ye, J., Zhu, J., Chen, H., Qian, J., Zhang, L., Wan, Z., et al. (2020). A novel lncRNA-LINC01116 regulates tumorigenesis of glioma by targeting VEGFA. *Int. J. Cancer* 146, 248–261. doi:10.1002/ijc.32483

Yin, D., Lu, X., Su, J., He, X., De, W., Yang, J., et al. (2018). Long noncoding RNA AFAP1-AS1 predicts a poor prognosis and regulates non-small cell lung cancer cell proliferation by epigenetically repressing p21 expression. *Mol. Cancer* 17, 92. doi:10.1186/s12943-018-0836-7

Yu, B., Ye, X., Du, Q., Zhu, B., Zhai, Q., and Li, X. X. (2017). The long non-coding RNA CRNDE promotes colorectal carcinoma progression by competitively binding miR-217 with TCF7L2 and enhancing the wnt/β-catenin signaling pathway. *Cell. Physiol. biochem.* 41, 2489–2502. doi:10.1159/000475941

Yu, D. J., Zhong, M., and Wang, W. L. (2021). Long noncoding RNA CASC15 is upregulated in non-small cell lung cancer and facilitates cell proliferation and metastasis via targeting miR-130b-3p. *Eur. Rev. Med. Pharmacol. Sci.* 25, 7943–7949. doi:10.26355/eurrev_201909_19010

Yu, X., Wang, Z. L., Han, C. L., Wang, M. W., Jin, Y., Jin, X. B., et al. (2019). LncRNA CASC15 functions as an oncogene by sponging miR-130b-3p in bladder cancer. *Eur. Rev. Med. Pharmacol. Sci.* 23, 9814–9820. doi:10.26355/eurrev_201911_19544

Yuan, S. X., Wang, J., Yang, F., Tao, Q. f., Zhang, J., Wang, L. l., et al. (2016). Long noncoding RNA DANCR increases stemness features of hepatocellular carcinoma by derepression of CTNNB1. *Hepatology* 63, 499–511. Article. doi:10.1002/hep.27893

Yuan, X., Yang, T., Xu, Y., Ou, S., Shi, P., Cao, M., et al. (2021). SNHG10 promotes cell proliferation and migration in gastric cancer by targeting miR-495-3p/CTNNB1 Axis. *Dig. Dis. Sci.* 66, 2627–2636. doi:10.1007/s10620-020-06576-w

Yue, B., Qiu, S., Zhao, S., Liu, C., Zhang, D., Yu, F., et al. (2016). LncRNA-ATB mediated E-cadherin repression promotes the progression of colon cancer and predicts poor prognosis. *J. Gastroenterol. Hepatol.* 31, 595–603. doi:10.1111/jgh.13206

Zeng, L., Lyu, X., Yuan, J., Wang, W., Zhao, N., Liu, B., et al. (2020). Long non-coding RNA LINC01116 is overexpressed in lung adenocarcinoma and promotes tumor proliferation and metastasis. *Am. J. Transl. Res.* 12, 4302–4313.

Zhang, J., and Lou, W. (2020). A key mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network linked to diagnosis and prognosis of hepatocellular carcinoma. *Front. Oncol.* 10, 340. doi:10.3389/fonc.2020.00340

Zhang, M., Gao, C., Yang, Y., Li, G., Dong, J., Ai, Y., et al. (2018). Long noncoding RNA CRNDE/PRC2 participated in the radiotherapy resistance of human lung adenocarcinoma through targeting p21 expression. *Oncol. Res.* 26, 1245–1255. doi:10.3727/096504017x14944585873668

Zhang, M., Wei, Y., Liu, Y., Guan, W., Zhang, X., Kong, J., et al. (2020). Metastatic phosphatase PRL-3 induces ovarian cancer stem cell sub-population through phosphatase-independent deacetylation modulations. *iScience* 23, 100766. doi:10.1016/j.isci.2019.100766

Zhang, Y., Guo, H., and Zhang, H. (2021). SNHG10/DDX54/PBX3 feedback loop contributes to gastric cancer cell growth. *Dig. Dis. Sci.* 66, 1875–1884. doi:10.1007/s10620-020-06488-9

Zhang, Z., Nong, L., Chen, M. L., Gu, X. L., Zhao, W. W., Liu, M. H., et al. (2020). Long noncoding RNA SNHG10 sponges miR-543 to upregulate tumor suppressive SIRT1 in nonsmall cell lung cancer. *Cancer biother. Radiopharm.* 35, 771–775. doi:10.1089/cbr.2019.3334

Zhou, S., Fang, J., Sun, Y., and Li, H. (2020). Integrated analysis of a risk score system predicting prognosis and a ceRNA network for differentially expressed lncRNAs in multiple myeloma. *Front. Genet.* 11, 934. doi:10.3389/fgene.2020.00934

Zhu, Y., Zhao, J., Tan, L., Lin, S., Long, M., and Peng, X. (2021). LncRNA-HCG18 regulates the viability, apoptosis, migration, invasion and epithelial-mesenchymal transition of papillary thyroid cancer cells via regulating the miR-106a-5p/PPP2R2A axis. *Pathol. Res. Pract.* 221, 153395. doi:10.1016/j.prp.2021.153395

Zou, Y., Sun, Z., and Sun, S. (2020). LncRNA HCG18 contributes to the progression of hepatocellular carcinoma via miR-214-3p/CENPM axis. *J. Biochem.* 168, 535–546. doi:10.1093/jb/mvaa073

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 15 August 2022
DOI 10.3389/fgene.2022.808542

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Shihai Liu,
The Affiliated Hospital of Qingdao University, China
Juan Wang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Shiwei Duan,
Zhejiang University City College, China

*CORRESPONDENCE
Liu Huilin,
liuhuilin202110@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 03 November 2021
ACCEPTED 04 July 2022
PUBLISHED 15 August 2022

CITATION
Yicheng F, Xin L, Tian Y and Huilin L (2022), Association of FLG mutation with tumor mutation load and clinical outcomes in patients with gastric cancer.
Front. Genet. 13:808542.
doi: 10.3389/fgene.2022.808542

COPYRIGHT
© 2022 Yicheng, Xin, Tian and Huilin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Association of *FLG* mutation with tumor mutation load and clinical outcomes in patients with gastric cancer

Fu Yicheng[1†], Liu Xin[2†], Yu Tian[3] and Liu Huilin[1*]

[1]Department of Geriatrics, Peking University Third Hospital, Beijing, China, [2]Department of Gastroenterology, Peking University Third Hospital, Beijing, China, [3]Graduate School, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Background: Stomach adenocarcinoma (STAD) is one of the most frequently diagnosed cancers in the world with a poor prognosis due to genetic heterogeneity. The present study aimed to explore potential prognostic predictors and therapeutic targets that can be used for STAD treatment.

Methods: We collected relevant data of STAD patients from the Cancer Genome Atlas (TCGA), including somatic mutation, transcriptome, and survival data. We performed a series of analyses such as tumor mutational burden (TMB), immune infiltration, and copy number variation (CNV) analysis to evaluate the potential mechanism of filaggrin (*FLG*) mutation in gastric cancer. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, and gene set enrichment analysis (GSEA) were performed for annotation of differentially expressed genes (DEGs). The STRING online database was used to construct the protein–protein interaction (PPI) and ceRNA network and hub genes were identified. Univariate and multivariate Cox regression analyses were used to determine the effect of selected DEGs on tumor prognosis.

Results: The *FLG*-mutant group (*FLG*-MT) showed a higher mutation load and immunogenicity in gastric cancer. GO and KEGG analyses identified and ranked unique biologic processes and immune-related pathway maps that correlated with the *FLG*-mutant target. GSEA analysis showed that several tumorigenesis and metastasis-related pathways were indeed enriched in *FLG*-mutant tumor tissue. Both cell cycle–related pathways and the DNA damage and repair associated pathways were also enriched in the *FLG*-MT group. The *FLG* mutations resulted in increased gastric cancer sensitivity to 24 chemotherapeutic drugs. The ceRNA network was established using Cytoscape and the PPI network was established in the STRING database. The results of the prognostic information further demonstrated that the OS and DFS were significantly higher in *FLG* mutation carriers, and the *FLG* gene mutation might be a protective factor.

Conclusion: The multiple molecular mechanisms of the *FLG* gene in STAD are worthy of further investigation and may reveal novel therapeutic targets and biomarkers for STAD treatment.

KEYWORDS

filaggrin, stomach adenocarcinoma, TCGA, prognostic model, tumor microenvironment FLG mutation in gastric cancer

Introduction

Gastric cancer is one of the most common gastrointestinal malignancies and still constitutes a health burden worldwide which was responsible for over one million new cases of mortality globally (Sung et al., 2021). Stomach adenocarcinoma (STAD) constitutes approximately 95% of GC cases, as a well-recognized heterogeneous type, its rate of 5-year survival is less than 30% for the advanced stage due to lack of effective therapeutic modalities (Ajani et al., 2017). Many DNA alterations have been detected in STAD, and accumulating evidence has demonstrated a crucial role in genetic correlation research (Ramezankhani et al., 2021). Recently, the genomic mechanisms of STAD have been widely studied, and many advanced STAD patients achieved better life expectations through the progression of targeted therapies and the chemotherapeutic drug optimization. The therapeutic backbone of metastatic STAD remains molecularly targeted therapies, which include *HER2*-targeting agents, anti-angiogenic agents, and epidermal growth factor receptor (*EGFR*) inhibitors (Alsina et al., 2019). However, several studies have failed to find the survival benefit of specific new innovative agents due to the marked multidrug-resistant phenotype of STAD, only three targeted therapeutics showed modest OS benefit in phase III trials: trastuzumab, ramucirumab, and lapatinib (Mizrak et al., 2017). However, negative trials with targeted agents have significantly outnumbered the positive trials in STAD in the past few years, and the cause of the genomic alterations in tumor growth and drug resistance are not fully known. For example, in the GATSBY trial, compared with paclitaxel or docetaxel as second-line treatment, trastuzumab showed no OS benefit ($p = 0.86$) and was not recommended as routine second-line treatment for *HER2*-positive advanced GAC (Liguigli et al., 2014). The two-phase III first-line and second-line trials of lapatinib both observed the absence of the primary end-point of OS advantage (Satoh et al., 2014; Hecht et al., 2016). The inherent genomic instability may give rise to a weak and inefficient response to cancer drug therapies, eventually leading to tumor progression and treatment failure (Russi et al., 2019). Thus, developing more specific and reliable biomarkers for clinical trials is essential for understanding the mechanisms of drug action in STAD therapy.

Recently, bioinformatics has become an effective tool for screening significant genetic variations that occur in carcinogenesis and offers a great promise for improved diagnosis, prognosis, treatment selection, and surveillance for cancer patients. A previous study used bioinformatics to predict the DEGs of STAD and its enriched pathways and screened and evaluated some hub genes to provide some ideas and references for the early diagnosis and treatment of STAD at the molecular level. For example, upregulation of *COL3A1*, *COL1A2*, *BGN*, and *THBS2* were found to significantly reduce the survival time of STAD patients (Qiu et al., 2020). *FN1*, *SPARC*, and *SERPINE1* were highly expressed and significantly related to a poor prognosis of STAD (Li et al., 2019). The high expressions of *PER1* and *NR1D1* were not only associated with poor OS, progression-free survival, and disease-free survival, but also associated with immune infiltration in STAD patients (Huang et al., 2021). RNA binding protein genes such as *PTBP1*, *PPIH*, *SMAD5*, *MSI2*, *RBM15*, *MRPS17*, and *ADAT3* were identified to be prognosis-related in STAD patients, the regulatory network and functional study showed *MRPS17* and *PTBP1* could reduce the number of infiltrated immune cells. The complement component 3a receptor 1 (*C3AR1*) was proven to promote the polarization of M2 macrophages and T-cell exhaustion, leading to the immune escape of STAD and high expression of the *C3AR1* gene is correlated with a poor prognosis (Li et al., 2021). In addition, in combination with bioinformatics, a prognostic model has been developed and proven capable of predicting prognosis of STAD patients (Ye et al., 2020). Nevertheless, the consequences of genomic alterations on tumor growth and drug resistance remain largely unexplored. The potential reason is most likely a combination of complex modes of inheritance span of different tumor stages and lack of specific target biomarkers. So far, no biomarker has been shown to be accurate enough to diagnose or predict the prognosis of STAD. Accordingly, as one of the emerging frontiers of STAD carcinogenesis and therapeutic target exploration, in-depth informatics investigation with larger sample sizes and fine-grained understanding of new genetic loci are helpful to identify more robust and reliable genetic biomarkers.

In this study, we systematically analyzed the somatically mutated genes of STAD based on TCGA database and screened out the filaggrin (*FLG*) gene for further investigation. The *FLG* gene could be an important candidate for STAD, which coincides with findings in another study (Wang et al., 2020). Filaggrin (*FLG*) protein, which is known as a filament-aggregating protein, is important for the formation of the stratum corneum and was proven to play a key role in the maintenance of an optimal skin, oral, and cervical mucosa barrier. Its monomer can combine with the keratin filaments as a matrix protein and results in aggregation of the keratinocytes. These keratinocytes act as a matrix in the stratum corneum. A previous study has shown that *FLG* gene mutations possibly bring about a greater susceptibility to Epstein–Barr virus (EBV)–associated gastric carcinoma (Kuang et al., 2016). This may be related to the fact that about 9% of gastric cancers harbor EBV infection. The deletion mutations of the *FLG* gene are also prevalent in

Asian populations, so making related genetic research valuable (Salama et al., 2021). However, the precise mechanisms underlying its development are still unclear as little is known about the role of the *FLG* gene. Thus in this study, we performed an in-depth investigation into the functional roles of the *FLG* gene in STAD. According to the mutation status of the *FLG* gene, patients were divided into *FLG*-mutant (MT) and *FLG*-wild-type (WT) groups. We found that the *FLG*-MT gastric cancer patients had showed a higher mutation load and immunogenicity. In addition, we explored the differences in pathway activation between the *FLG*-MT gastric cancer patients and *FLG*-WT patients through functional enrichment analysis to explain the effect of *FLG* on the tumor microenvironment (TME). Next, we obtained insights into prognosis evaluation, protein and ceRNA interaction networks, immune infiltration, and anticancer drug sensitivities, which may provide valuable references for diagnosis, targeted drug research, and prognosis evaluation of STAD.

Materials and methods

Data downloads

The masked somatic mutation data of STAD patients used in this study were retrieved from the TCGA GDC database (http://portal.gdc.cancer.gov/). Patients with a pathologic diagnosis of stomach adenocarcinoma were included. The data were preprocessed by VARSCAN software, and visualizations of somatic mutations were carried out in R (Foundation for Statistical Computing, Vienna, Austria) using the package Maftools (Mayakonda et al., 2018). The gene expression data (FPKM value) of the patients' RNA sequencing were downloaded and converted to the TPM value, and the lncRNA and mRNA associations were then established. In addition, the clinicopathological features and prognosis of the STAD patients, such as gender, age, malignant stage, TNM stage, and MSI value, all these data were obtained from the UCSC Xena (http://xena.ucsc.edu/). We used the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (http://tide.dfci.harvard.edu) to predict the response to immunotherapy of each sample (Jiang et al., 2018).

Copy number variation analysis

To analyze the copy number variations (CNVs) of *FLG* genes in TCGA-STAD patients, the masked copy number segment data were downloaded using the TCGAbiolinks R package (version 2.6.12). GISTIC 2.0 was used to conduct GenePattern5 analysis of downloaded CNV fragments (Reich et al., 2006). We use default settings in the GISTIC 2.0 analysis with the exception of several parameters (e.g., the confidence coefficient was 0.99; X chromosomes were not excluded prior to analysis). Finally, the results of GISTIC 2.0 analysis were visualized using the Maftools package of R software (Mayakonda et al., 2018).

Calculation and correlation analysis of somatic mutation and tumor mutation load fraction

Tumor mutational burden (TMB) in this study was defined as the number of somatic synonymous mutations per megabase in each tumor sample, with silent mutations excluded. The Wilcoxon matched-pairs signed-rank test was used to compare TMB values between *FLG* mutation and non-mutation groups.

Identification of differentially expressed genes (DEGs) and clinical correlation analysis

To analyze the effect of *FLG* mutation on tumorigenesis in STAD patients, the samples in the TCGA database were divided into a mutation group and a non-mutation group according to the *FLG* mutation situation. The DEGs between the two groups were determined using the DESeq2 package in R (Love et al., 2014). The cut-off criteria for statistical significance were a log-fold change (FC) of greater than 1 and a p-value of less than 0.05. Visualizations of differentially expressed genes such as volcano plots and heatmaps were generated using standard R packages.

Functional enrichment analysis

The GO analysis serves as a bioinformatics tool that provides structured annotations, including biological processes (BPs), molecular functions (MFs), and cellular components (CCs), for genes and gene products. KEGG (http://www.genome.jp/) is a widely used database storing information about genomes, biological pathways, diseases, and drugs. Enrichment plots of gene signatures were generated using the R package clusterProfiler (Yu et al., 2012), and FDR critical value of less than 0.05 was considered to indicate a statistically significant difference. To investigate the differences in biological processes between different groups, the enrichment analysis was performed using GESA (Hanzelmann et al., 2013). GSEA is a statistical method to assess whether a priori defined set of genes shows statistically significant concordant differences between two different biological statuses (Subramanian et al., 2005). GSEA analysis of the gene expression profiling dataset of TCGA-STAD patients was implemented using the clusterProfiler package. C2.all.v6.2.symbols.gmt was selected as the reference gene set. False discovery rate < 0.1, and p-value < 0.05 were set as the cut-off criteria.

Comparison of biological functions and immune estimation scores

We further analyzed the correlation between different subgroups and some biological related processes. The immune and stromal scores were evaluated by applying the ESTIMATE algorithm using the estimate R package (R version 3.5.3) (Yoshihara et al., 2013). The scores were used to reflect the level of immune cell and stromal cell infiltration of tumor tissue. The Mann–Whitney U-test was used to compare the infiltration levels of immune cells between two groups.

Protein–protein interaction network construction

The PPI information available in the STRING network in the STRING database (http://string-db.org, version 10) is useful for predicting physical and functional interactions (Szklarczyk et al., 2019). All DEGs were mapped to the STRING database, and the interactions with reliability scores of more than 0.4 were selected to analyze the relationship of the DEGs. Cytoscape v3.7.2 was used to select the key nodes with the strongest connectivity for visualizing molecular interaction networks (Shannon et al., 2003). The MCODE plugin in Cytoscape 3.7.2 was used to identify the most densely connected region in the PPI based on vertex weights, which could identify hub genes in the PPI network.

Construction of ceRNA networks

We further retrieved experimentally validated miRNA–mRNA interactions from the miRTarBase. Based on core mRNAs obtained from PPI interaction analysis, miRTarBase was utilized to predict possible regulatory miRNA, and further predict related lncRNA. The alluvial diagrams of the co-expression network with the overlapped lncRNA–miRNA–mRNA relationships were generated using the R package ggalluvial.

Sensitivity analysis of anticancer drugs

Genomics of Drug Sensitivity in Cancer (GDSC; https://www.cancerrxgene.org/) is a public database for tumor molecular therapy and mutation exploration (Vanden et al., 2018). The R package pRRophetic was used for downloading cell line gene mutation data and IC50 values of different anticancer drugs and analyzing the correlation between *FLG* gene mutation patients and the sensitivity of different anticancer drugs.

Validation of clinical prediction models

The relationship between clinicopathological and prognostic features (overall survival OS) and *FLG* gene mutations of STAD patients in TCGA was analyzed with the logistic regression and receiver operating characteristic (ROC) methods. The Harrell consistency index (C-index) of *FLG* expression was based on the best separation. The diagnostic ROC curve was used to explore the prognostic or predictive accuracy of each characteristic underlying the area under the curve (AUC). The Kaplan–Meier curve was used to estimate the effects of *FLG* on the overall survival of STAD patients.

Statistical analysis

All statistical analyses were performed as the means ± standard deviation. The R software (version 4.0.2) was utilized to measure the data. Quantitative data that were not normally distributed were evaluated using the Mann–Whitney U-test, and Student's t-test was used for normally distributed data. For analysis of relations between categorical variables, we used the chi-squared test or Fisher's exact test when appropriate. The correlation coefficients among different genes were calculated by Pearson correlation analysis. Survival analysis was estimated using R package survival and survival differences were determined by Kaplan–Meier analysis, the log-rank test was used to compare OS among groups. R software with the package pROC was used to produce the ROC curve, the area under the curve (AUC) as a measure of accuracy. All tests were two-sided with a significance level of $p < 0.05$.

Results

Association between the *FLG* status and clinical characteristics

Somatic genetic alterations data on STAD patients were downloaded from the TCGA database and analyzed as previously described (see Table 1). The waterfall plot for all STAD patients in the study showed that the top ten most mutated genes are *TTN*, *TP53*, *MUC16*, *ARIDIA*, *LRPIB*, *SYNE1*, *FLG*, *CSMD3*, *FAT4*, and *PCLO*, and 19% of the patients carried *FLG* mutation. Missense mutations accounted for the majority of mutation types of STAD patients, with C>T being the most common single-nucleotide variant (SNV) (Figure 1A). Meanwhile, correlation analysis showed that *FLG* had co-occurrence with *LRP1B* and *RYR2* ($p < 0.05$), and tended to be mutually exclusive with *TTN*, *TP53*, and *MUC16* ($p < 0.05$). (Figure 1B). TCGA-STAD patients were allocated into two groups based on the *FLG* mutation status, the *FLG*-mutant

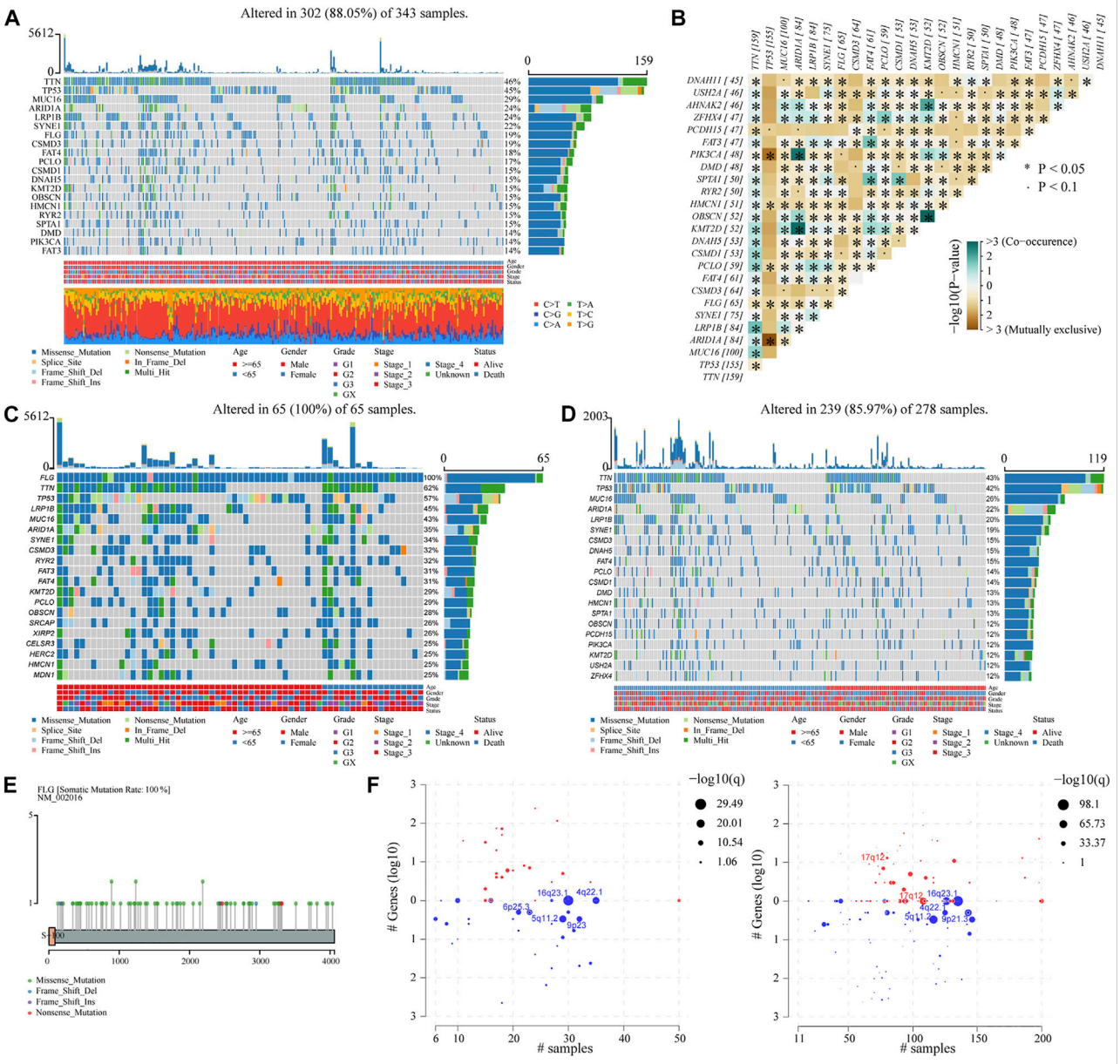


FIGURE 1
Correlation of clinical characteristics of STAD patients with *FLG* gene mutation. **(A)** Waterfall plot for all STAD patients in the study. Types of mutations were classified according to different categories, the majority of which were missense mutations. The C > T mutation is the most common SNV. The corresponding TMB values of specific tumor samples and top 30 ranked mutant genes are also shown. **(B)** Correlation analysis of different mutant genes. **(C,D)** Top 30 most frequently mutated genes in *FLG*-mutant and non-mutant groups. The left side is the *FLG*-mutant group, the right side is the non-mutant group. Genes are sorted by mutational frequency, and samples are sorted and ordered according to non-synonymous mutational load. The legend earlier shows the mutation load. Age, gender, grade, stage, and OS status are noted in order. **(E)** Distribution diagram of amino acid variation of FLG proteins in the TCGA-STAD dataset, missense mutation is the main form. **(F)** TCGA-STAD samples with available CNV data were analyzed using GISTIC 2.0 software and visualized using Maftools package.

group (*FLG*-MT) (Figure 1C) and *FLG*-wild-type group (*FLG*-WT) (Figure 1D). The left side is the *FLG*-MT group, and the right side is the *FLG*-WT group. Genes are sorted by the mutational frequency, and samples are sorted and ordered according to the non-synonymous mutational load. The top 30 most frequently mutated genes of *FLG*-MT and *FLG*-WT groups are shown in the waterfall plot. The amino acid changes of the *FLG* gene are highlighted in Figure 1E, the missense mutation appeared to be the major form of mutations in all STAD patients. Meanwhile, based on *FLG* mutation levels, the CNV data of

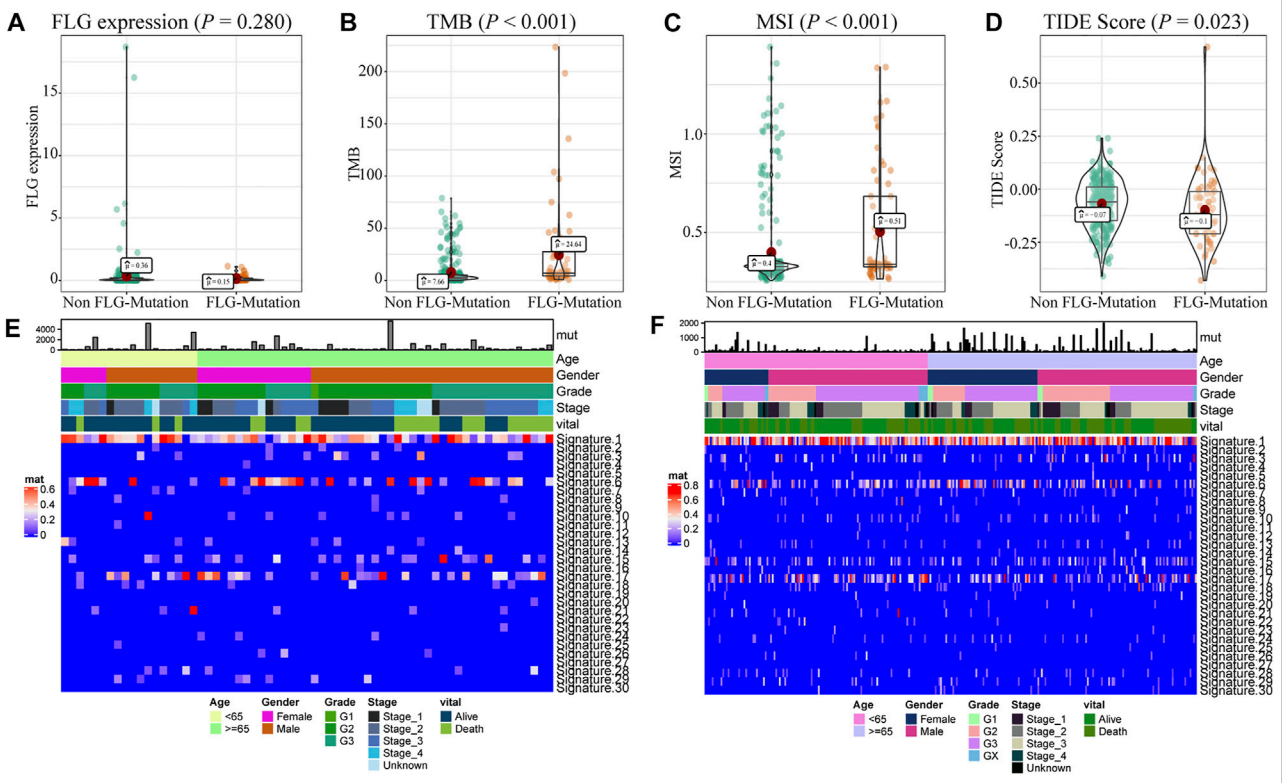


FIGURE 2
Biological characteristics of *FLG* gene mutations in STAD patients. **(A)** There were no significant differences in gene expression levels between patients with and without *FLG* mutations. (p = 0.280). **(B)** TMB was significantly higher in patients with *FLG* mutation (p < 0.001). **(C)** MSI was significantly higher in patients with *FLG* mutation (p < 0.001). **(D)** Patients with *FLG* mutation had a significantly lower TIDE score than those in the non-mutant group (p = 0.023). The lower the TIDE score, the better the effect of immunotherapy. **(E)** Cosmic signature thermogram analysis of *FLG*-mutant patients in the TCGA dataset, the corresponding clinical features of the patients are shown earlier. **(F)** Cosmic signature thermogram analysis of non-mutant patients in the TCGA dataset.

TCGA-STAD patients were divided into the mutant group and the non-mutant group. As shown in Figure 1F, the CNV levels of multiple genes have changed significantly.

Relationships between the *FLG* mutation status and biological characteristics and mutation load in STAD

We analyzed the effects of *FLG* gene mutations on different biological characteristics. The *FLG* gene expression levels did not differ significantly between *FLG*-mutant and non-mutant groups (p = 0.280), as shown in Figure 2A. The TMB value (Figure 2B, p < 0.001) and MSI value (Figure 2C, p < 0.001) were elevated significantly in the *FLG*-mutant group, and the TIDE score (Figure 2D, p = 0.023) was decreased compared with the non-mutant group. These data support that the patients with *FLG* mutation may benefit from targeted therapies and immunotherapy. In addition, combining tumor biological characteristics with somatic mutational signatures, the Sanger signatures decomposed 96 spectrums of mutational signatures into 30 different local areas (Alexandrov et al., 2013). A significant change occurred in signatures 1, 6, and 17 in the *FLG*-mutant group (Figures 2E,F). We further analyzed the effects of *FLG* gene mutations on immunologic characteristics of TCGA-STAD patients. The immune and stromal scores were used to quantify the immune and matrix components in STAD. The results showed that there was no significant difference in the immune (Figure 3A, p = 0.622) and stromal scores (Figure 3B, p = 0.504) among patients with *FLG* gene mutations, compared to patients without *FLG* gene mutations. Meanwhile, based on the TIMER database searches, we observed effects of somatic copy number alterations (SCNAs) on immune cell infiltration in *FLG* mutation tumor samples. The results suggested SCNAs, especially arm-level gain and high amplification, have a differential effect on tumor infiltrating immune cells including B cells, $CD4^+$ T cells, $CD8^+$ T cells, macrophages, neutrophils, and dendritic cells. Then, we further analyzed the correlation between *FLG* mutations and different biological

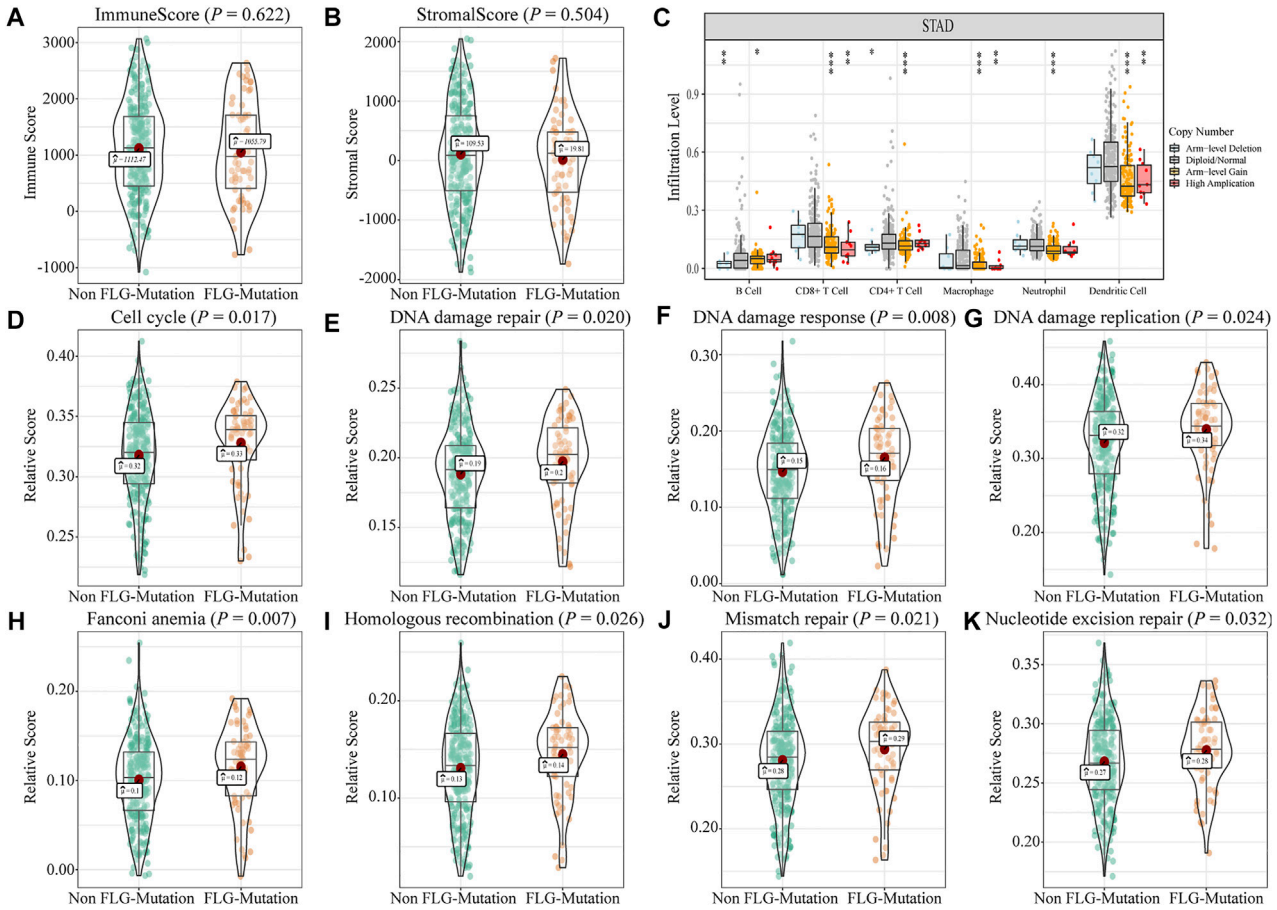


FIGURE 3
Effects of *FLG* gene mutations on immunological characteristics in TCGA-STAD datasets. **(A)** In TCGA-STAD datasets, the immune scores in the *FLG*-mutant group were unremarkable compared to those in the non-mutant group. (p = 0.622). **(B)** Matrix score in the *FLG*-mutant group was unremarkable compared to that in the non-mutant group. (p = 0.504). **(C)** Immune cell infiltration was analyzed using the TIMER website. The significant differences were observed in the immune cell infiltration between *FLG*-mutant and non-mutant groups. **(D–K)** Biological functions were significantly different between mutant and non-mutant groups ($p < 0.05$), including cell cycle–related pathways, DNA injury repairment-related pathways.

pathways. The data demonstrated that changes in biological pathways were mainly enriched in cell cycle, DNA damage repair, DNA damage response, DNA damage replication, Fanconi anemia, homologous recombination, mismatch repair, and nucleotide excision repair (Figures 3D–K, $p < 0.05$).

Drug sensitivity analysis

To detect the drug sensitivity of *FLG* gene mutations in STAD patients, we utilized the Genomics of Drug Sensitivity in Cancer (GDSC) database, which contains drug sensitivity data for 138 chemotherapy drugs and small-molecule drugs. The results showed that the IC50 values of 24 chemotherapeutic drugs and small-molecule anticancer drugs were significantly different between *FLG* and non-mutant patients ($p < 0.001$, Figure 4), especially in LFM. A13, AZD8055, and X17. AAG.

Construction and evaluation of the nomogram model

To further explore the relationships between the *FLG* mutation status and clinical phenotype in bladder cancer, clinical correlation analysis of *FLG* gene mutations was conducted. In the TCGA-STAD dataset, the results of survival analysis showed that the *FLG* mutations suggest better overall survival (OS, log-rank p = 0.074, Figure 5A, Table 2) and disease-free survival (DFS, log-rank p = 0.041, Figure 5B, Table 3) in STAD patients, but have no significant effect in progression-free survival (PFS, log-rank p = 0.163; Figure 5C). To further explore

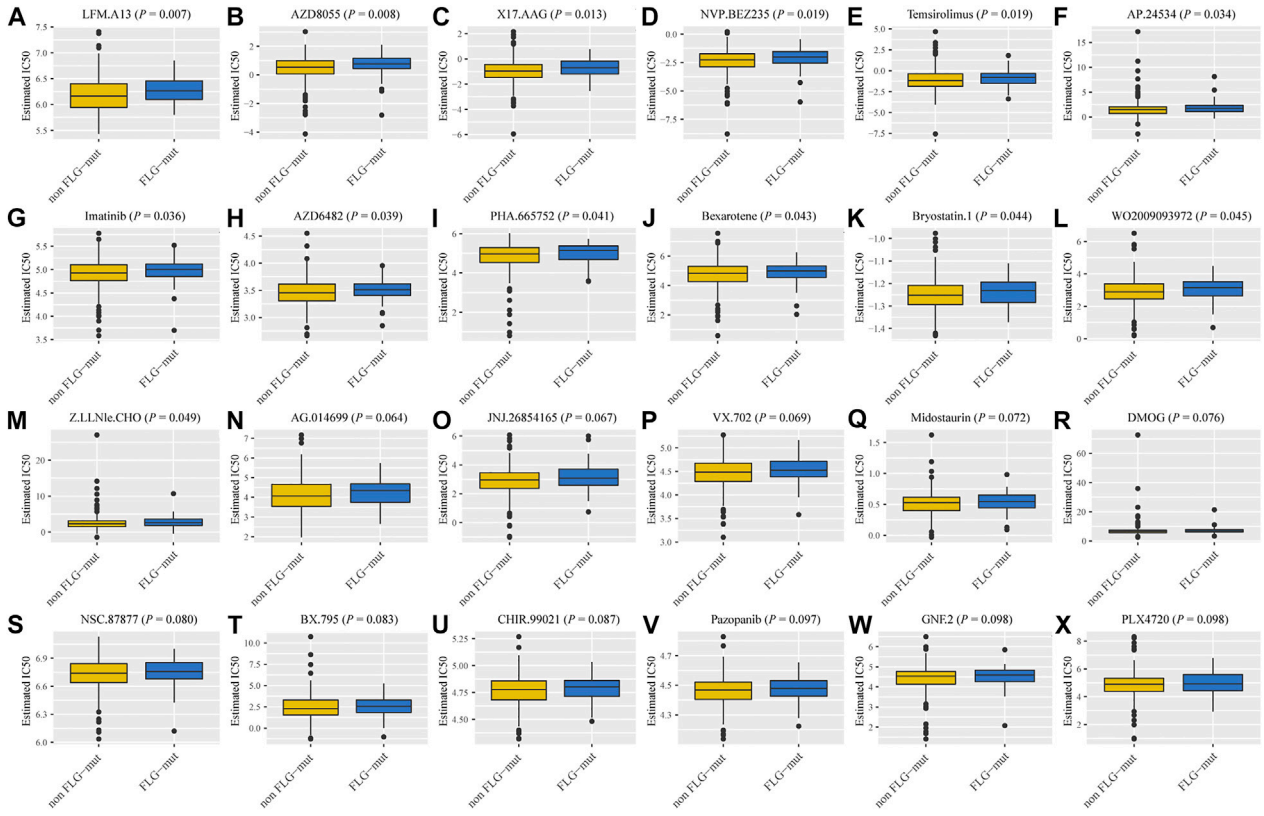


FIGURE 4
Sensitivity of *FLG* mutations on chemotherapeutic drugs and small-molecule anticancer drugs based on GDSC database analysis. **(A–X)** The IC50 values of 24 drugs between *FLG* and non-mutant patients.

the effects of mutations in *FLG* genes on TCGA-STAD patients, combined with the clinicopathological features of the patients, the univariate and multivariate Cox regression analyses showed that the mutation levels of the *FLG* gene are protective factors for STAD patients, but not independent protective factors, suggesting the potential diagnostic roles of *FLG* in STAD. Using clinicopathological parameters including age, gender, tumor stage, depth of invasion, lymph node metastasis, and distant metastasis, a nomogram to prognosticate OS and DFS was proposed and internally validated (Figures 5D,F). Discrimination of the nomogram was measured by calculating the c-index (concordance index), which indicated a high discrimination ability (OS: 0.675, 95% CI, 0.628–0.722; DFS: 0.663, 95% CI, 0.600–0.725). The calibration plot showed excellent concordance for the 1-, 3-, and 5-year predicted and actual OS and DFS probabilities (Figures 5E,G). Therefore, The *FLG* mutation can be used as an independent prognostic indicator in STAD. The prognostic nomogram based on *FLG* mutation may serve as a reliable model for predicting survival of patients.

Differential expression analysis

To analyze the effect of *FLG* gene mutation on tumorigenesis in TCAG-STAD patients, the patients were divided into the *FLG*-mutant group and non-mutant group, and the differential gene expression analysis was further performed. After standardization and removal of batch effects in the microarray results, we found that 100 genes were significantly upregulated and 414 genes were significantly downregulated in the TCGA-STAD dataset (Figures 6A,B). To explore how *FLG* gene mutation may affect the gastric carcinogenesis, we conducted a functional enrichment analysis on the differentially expressed genes. GO analysis revealed a number of biological processes affected by *FLG* gene mutation, such as cornification, keratinocyte differentiation, skin development, endoplasmic reticulum lumen, and cornified envelope (Figure 5C, Table 4). The KEGG pathway analysis showed that the differentially expressed immune genes were related primarily to neuroactive ligand–receptor interaction, protein digestion and absorption, chemical carcinogenesis, and pancreatic secretion pathways (Figure 5D, Table 5). Next, we analyzed the

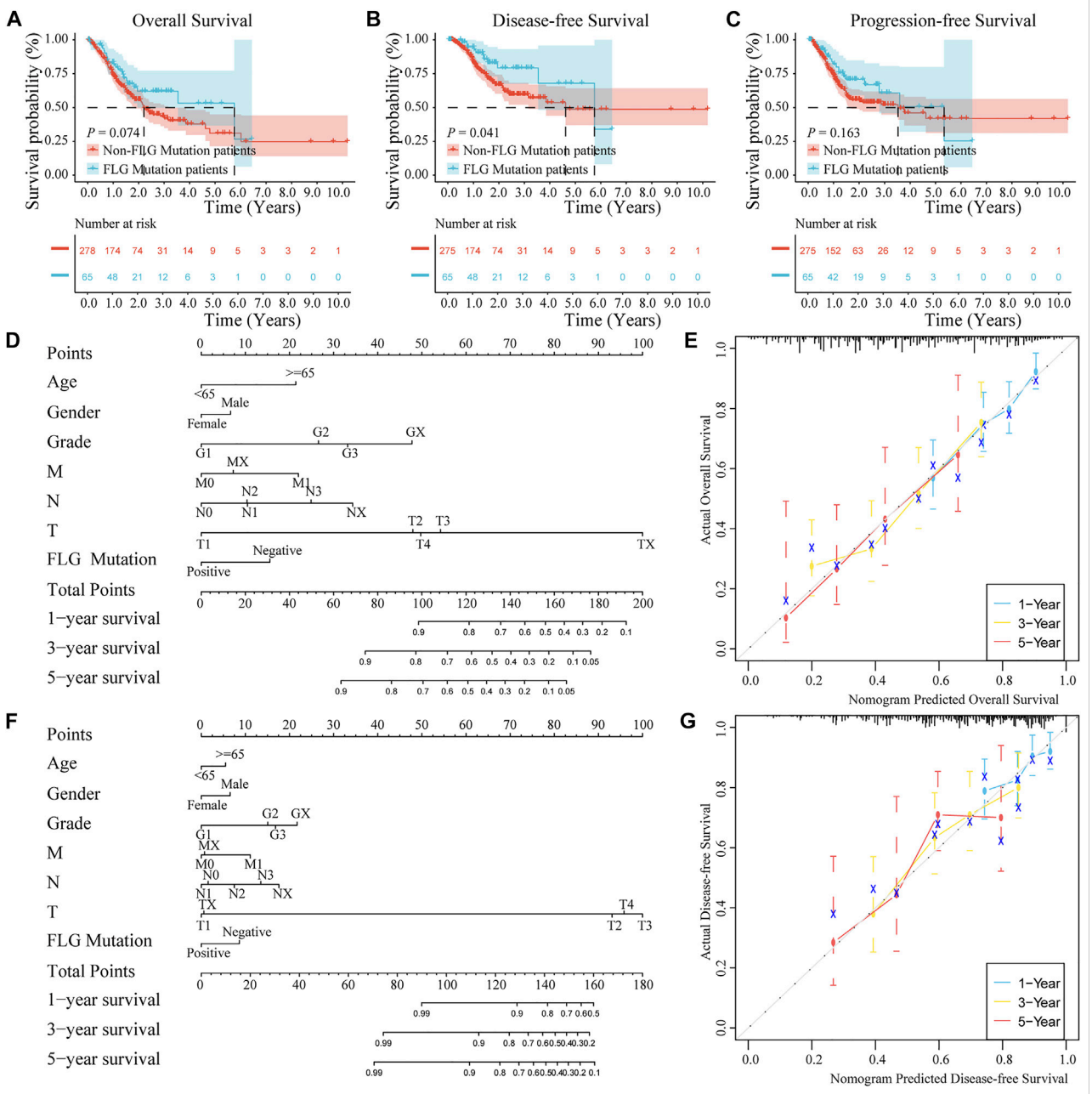


FIGURE 5
Effects of *FLG* gene mutation on clinicopathological features in the TCGA-STAD dataset. **(A–C)** In the TCGA-STAD dataset, the survival analysis showed that OS (log-rank $p = 0.074$) and DFS (log-rank $p = 0.041$) were better in patients with *FLG* gene mutations, but no significant impact on PFS (log-rank $p = 0.163$). **(D)** OS nomogram was constructed by the mutation of *FLG* gene combining with clinicopathological features. **(E)** Calibration curve of the *FLG* gene mutation nomogram. The abscissa axis is predicted survival, the ordinate axis is observed survival, every calculation is repeated 1,000 times. The calibration plots demonstrate good agreement between the predictions made by the nomogram and actual prognosis of patients for 1-,3-, and 5 years. **(F)** Nomogram constructed by the mutation of *FLG* gene combining with clinicopathological features. **(G)** Calibration curve of the nomogram of *FLG* gene mutation. The abscissa axis is predicted survival, the ordinate axis is observed survival, every calculation is repeated 1,000 times. The calibration plots demonstrate good agreement between the predictions made by the nomogram and actual prognosis of patients for 1-,3-, and 5 years.

functional enrichment pathways of the *FLG*-MT and *FLG*-WT groups in the TCGA-gastric cancer by GSEA (Table 6). The results showed that pathways enriched among the differentially expressed genes in *FLG*-mutant tissues, which include ribosomes, focal adhesion, dilated cardiomyopathy, regulation of actin cytoskeleton, and ECM receptor interaction, as is shown in Figure 7B.

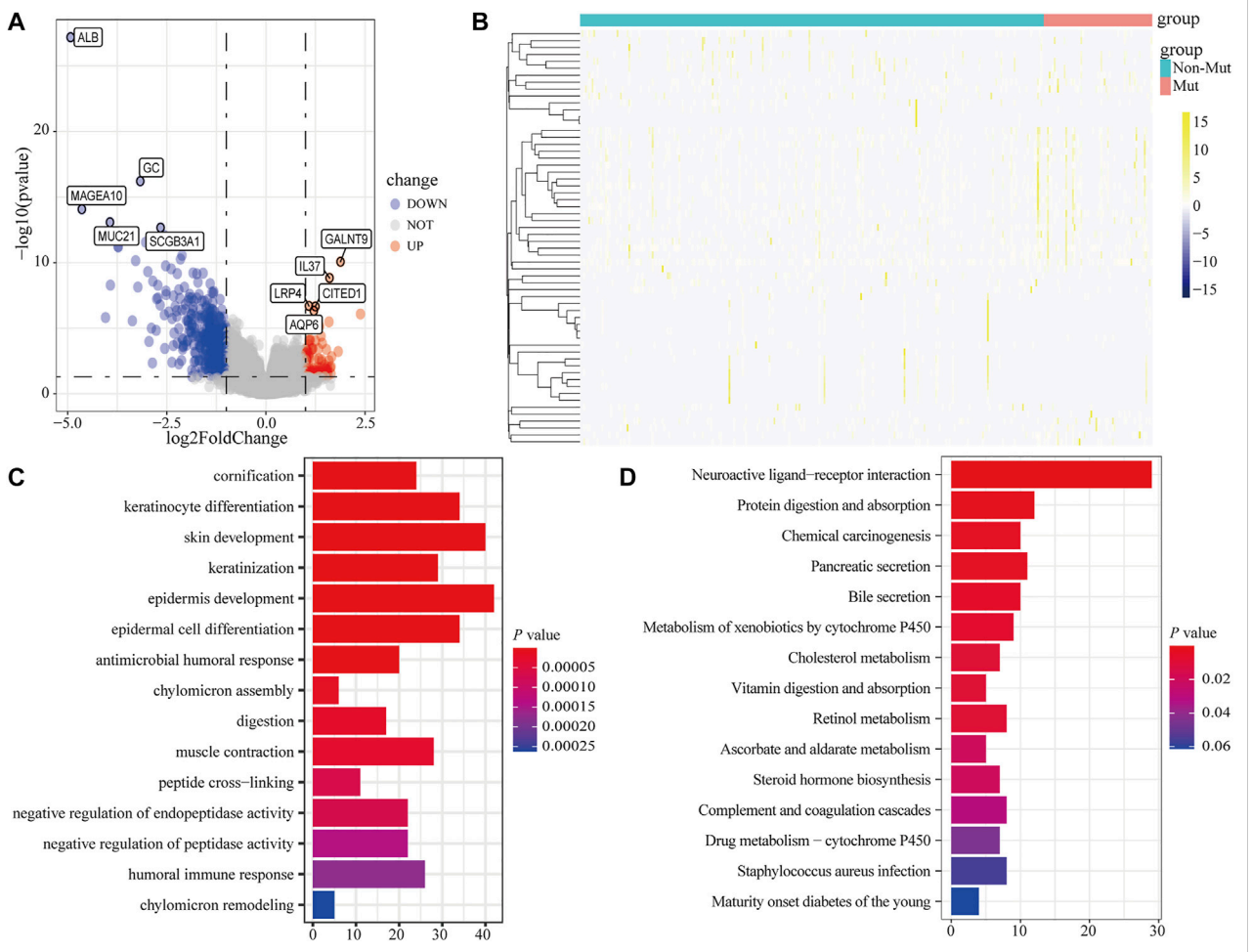


FIGURE 6
Functional enrichment analysis of the differentially expressed genes based on *FLG* gene mutation. **(A,B)** Volcano plot and heatmap show the expression of DEGs between *FLG*-mutant and non-mutant groups. **(C)** Based on CC, BP, and MF levels, GO analysis suggests that differentially expressed genes are closely related to cornification, keratinocyte differentiation, skin development, endoplasmic reticulum lumen, and cornified envelope biological processes. **(D)** KEGG analysis showed that these differentially expressed genes were participated in the neuroactive ligand–receptor interaction, protein digestion and absorption, chemical carcinogenesis, and pancreatic secretion and other biological related signaling pathways.

Construction of PPI and ceRNA networks

The PPI network of the DEGs was constructed using Cytoscape software based on the STRING database (Figure 8A). The upregulation of gene expression was indicated in red, whereas the downregulation of gene expression was indicated in blue. The MCODE plugins were used to choose the local high density as important nodes, and defined as hub genes (Figure 8B). The top six miRNAs with the highest number of target mRNAs are identified, hsa-miR-29c-3p, hsa-miR-409-3p, hsa-miR-548p, hsa-miR-29b-3p, hsa-miR-29a-3p, and hsa-miR-144-3p (Figure 8C). Also, by integrating the miRNA–mRNA and miRNA–lncRNA regulatory relationships, the ceRNA network of miRNA–mRNA–lncRNA interactions was constructed based on the miRTarBase. As shown in Figure 8D, the lncRNA–miRNA–mRNA network comprises six lncRNA, 5 miRNA, and 2 mRNA nodes.

Discussion

Stomach adenocarcinoma (STAD) is among the most lethal human malignancies with both high mortality and high metastatic capacity (Smyth et al., 2020). Certain genes have been shown to play essential roles in STAD development (Wang et al., 2019). Despite the large number of studies carried out to date, our understanding of molecular mechanisms of STAD is still limited due to lack of stable and effective biomarkers. Therefore, there is an urgent need for more

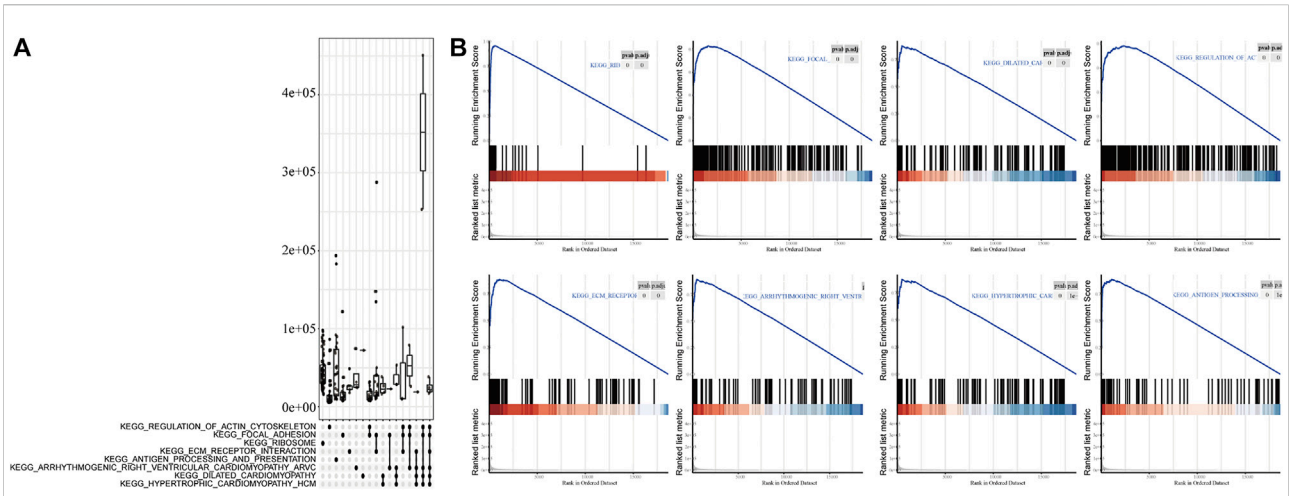


FIGURE 7
GSEA analysis based on differential expression of TCGA-STAD datasets. **(A)** Upper panel showed the GSEA analysis based on the differential expression of TCGA-STAD datasets. **(B)** Results showed tumor tissue with *FLG* mutation is closely related to ribosome, focal adhesion, dilated cardiomyopathy, regulation of action cytoskeleton, and ECM receptor interaction pathways.

reliable molecular biomarkers, early diagnosis, and effective treatments. We collected the TCGA-gastric cancer containing clinical data and mutation data to explore the gene mutation and pathological mechanism of STAD using bioinformatic analysis, and we found that *FLG* gene mutations have an important impact on the clinical and biological characteristics of STAD. These mutations comprise missense and CNV changes, the majority being missense mutations. We further identified 100 upregulated DEGs and 414 downregulated DEGs in the *FLG* mutation group. We applied three pathway analyses, GO, KEGG, and GSEA, to analyze the biological functions of these DEGs. Go analysis revealed that changes in modules were mostly enriched in biological processes such as keratinization, keratinocyte differentiation, and skin development. KEGG pathway enrichment analysis reveals that DEGs were mainly involved in neuroactive ligand–receptor interaction, protein digestion and absorption, and chemical carcinogenesis pathways. GSEA also revealed biologically relevant pathways associated with ribosome, focal adhesion, dilated cardiomyopathy, regulation of actin cytoskeleton, and ECM receptor interaction. Then through PPI network construction, key hub genes and the most significant module were selected. In addition, we found that mutations in the *FLG* gene were associated with immune cell infiltration, and genetic mutation of *FLG* in STAD causes increased sensitivity to anticancer agents. We examined the relationship between *FLG* mutations and prognosis. *FLG* mutations were significantly associated with better disease-free and overall survival, and appeared to be an independent prognostic factor. Based on these findings, we proposed that the *FLG* gene could be regarded as a potential biomarker to further explore the molecular mechanism and the prognostic effects of STAD.

Gastric cancer development involves multiple gene alterations. In this study, we first applied computational algorithms to detect driver genes using somatic mutations of STAD tissues and classified the data into different categories. *TTN*, *TP53*, *MUC16*, *ARIDIA*, *LRPIB*, *SYNE1*, *FLG*, *CSMD3*, *FAT4*, and *PCLO* were the 10 most frequently mutated genes, which is partially in concordance with previously published studies (Wang et al., 2020). For example, *TP53* mutations were more common in gastrointestinal adenocarcinomas with intact DNA mismatch repair protein expression (Krishnamurthy et al., 2022). *MUC16* mutations were found to be potentially associated with GC prognosis, some mutation statuses of *MUC16* and *TTN* were identified with high potential in predicting TMB (Yang et al., 2020). For another example, the somatic mutation rate of the *ARIDIA* gene varies significantly between GC patients of Asian and Caucasian descent (20.7% vs. 32.1%, $p = 0.01$), which might have important implications for precise therapeutics in GC patients (Jia et al., 2017). There are also genes, such as *SYNE1*, the high level of *SYNE1* promoter methylation was associated with poorer chemotherapy efficacy in advanced gastric cancer patients (Qu et al., 2021). Also, *FAT4*, a tumor suppressor gene exerts an important role in cell adhesion. Reduced expression of *FAT4* and increased methylation of its promoter may accelerate the progression of benign tumors to malignant GC (Pilehchian et al., 2017). However, at present, there are still limited studies on the effects of *FLG* mutations on STAD. We, therefore, chose the *FLG* gene for subsequent studies.

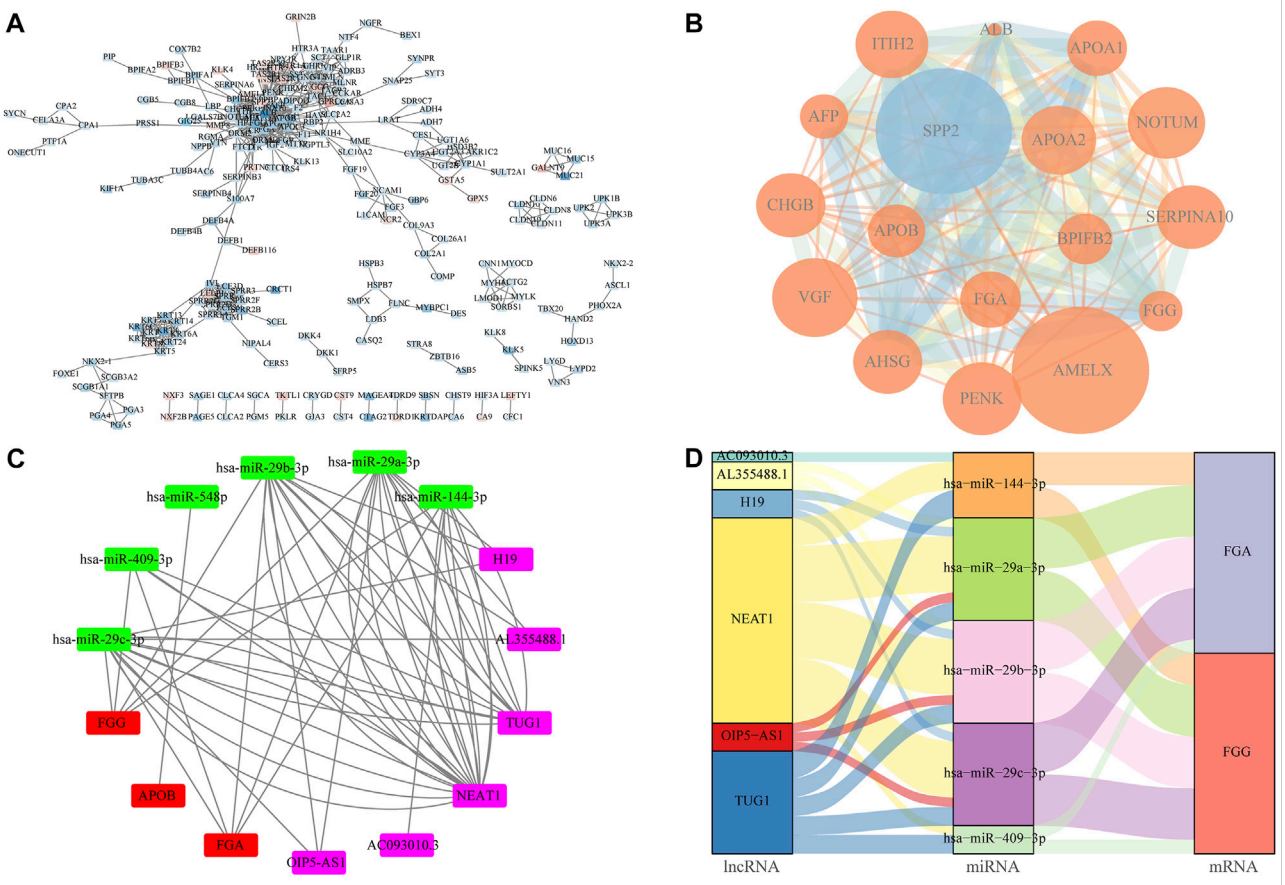


FIGURE 8
Construction of protein–protein interaction (PPI) and ceRNA network. **(A)** Layout in Figure for protein–protein interactions were created by Cytoscape software *via* the link provided by STRING. Values on the lines connecting the spheres indicate upregulated genes in red and downregulated genes in blue. The color shades have a positive relation with logFC. **(B)** MCODE algorithm is used to identify high density regions from the PPI network and the size of the circles is proportionate to the logFC, the color shades are proportional to *p*-value. **(C,D)** ceRNA interaction network was constructed by Cytoscpe and Sankey diagram based on the hub genes.

The protein encoded by the *FLG* gene is an intermediate filament-associated protein that aggregates keratin intermediate filaments in mammalian epidermis. It is initially developed as profilaggrin, which is localized in keratohyalin granules, and is subsequently proteolytically processed into individual functional filaggrin molecules. In humans, the *FLG* gene is located within the epidermal differentiation complex (EDC) on chromosome 1q21, spans~25 kb of DNA and comprises three exons and two introns. Exon 1 is non-coding and protein translation start site begins within exon 2. The majority of the profilaggrin protein is encoded by the exon 3 (Sandilands et al., 2009). The *FLG* gene is composed of tandem repeats with CNVs consisting of 10, 11, or 12 copies of the sequence encoding filaggrin monomers. Tandem repeats are usually present in coding and regulatory regions of the human genomes, so it has a greater chance of mutation and are associated with many genetic diseases (Kim et al., 2020). The loss-of-function mutations in *FLG* are common, and approximately 2–10% of Europeans carried at least one *FLG* null mutation. The *FLG* loss-of-function mutation is associated with ichthyosis vulgaris, atopic dermatitis, inflammatory dermatosis, and inflammation dysregulated diseases such as asthma and allergy (Brown and McLean, 2012; Zhu et al., 2018). One previous study has explored two single-nucleotide polymorphism (SNP) loci of the *FLG* gene, rs3126085 and K4671X, which were associated with EBV-associated gastric carcinoma for (EBVaGC). The genotype AA of rs3126085 (c.3321delA), as a most popular *FLG* mutation in Chinese Han people, was considered as a hazardous sign for EBV-associated gastric carcinoma (Yang et al., 2017). Another study detected the *FLG* rs2065955 genotype and allele distribution in 64 EBV-associated gastric carcinoma samples, and found genotype CC may contribute more to the risk of developing EBVaGC (Kuang et al., 2016). Nevertheless, no significant difference in *FLG* expression was detected by immunohistochemical analysis in those studies. It would be speculative to guess that a single-nucleotide variation in intron may have little impact on the gene expression, or the intragenic CNV may also complicate the process

TABLE 1 The baseline patient data of STAD in TCGA database.

Variables	All patients	*FLG*-WT	*FLG*-MT	*p* Value
	(*n* = 343)	(*n* = 278)	(*n* = 65)	
Age				0.014*
<65	144 (42.0%)	126 (45.3%)	18 (27.7%)	
≥65	199 (58.0%)	152 (54.7%)	47 (72.3%)	
Gender				0.761
Female	119 (34.7%)	98 (35.3%)	21 (32.3%)	
Male	224 (65.3%)	180 (64.7%)	44 (67.7%)	
Grade				0.01*
G1 & G2	134 (39.1%)	99 (35.6%)	35 (53.8%)	
G3 & GX	209 (60.9%)	179 (64.4%)	30 (46.2%)	
T				1
T1 & T2	88 (25.7%)	71 (25.5%)	17 (26.2%)	
T3 & T4 & TX	255 (74.3%)	207 (74.5%)	48 (73.8%)	
M				0.079
M0	306 (89.2%)	244 (87.8%)	62 (95.4%)	
M1 & MX	37 (10.8%)	34 (12.2%)	3 (4.6%)	
N				0.09
N0 & N1	192 (56.0%)	149 (53.6%)	43 (66.2%)	
N2 & N3 & NX	151 (44.0%)	129 (46.4%)	22 (33.8%)	

of cancer pathogenesis. Interpretation of the results of prior studies may have been hampered by limited sample sizes and heterogeneous study populations. So, it is necessary to integrate the complexity of cancer genome data and multiplex optimization in a single method as bioinformatics to facilitate data integration and processing.

We found missense mutation is the predominant form of *FLG* genetic variation through the analysis of the amino acid changes in the *FLG* mutation group. Next, we demonstrated the existence of CNV alternations in the *FLG* mutation group, which would actually alter the *FLG* gene structure and function of tumor cells, so further analysis of these CNV regions should be treated as priority in the future. We finally identified 100 upregulated DEGs and 414 downregulated DEGs. The heatmap of all DEGs showed obvious difference between the *FLG*-mutant and non-mutant groups. To better understand the interactions among DEGs, we further performed functional enrichment analysis. The GO term enrichment analysis shows that DEGs were mainly involved in biological processes such as cornification, keratinocyte differentiation, skin development, endoplasmic reticulum lumen, and cornified envelope. As in previous studies, the role of filaggrin in skin physiology and disease is well-established; we thus surmise that *FLG* gene mutation may also have a role in epidermal differentiation, morphogenesis, and homeostasis of gastric cancer cells. These signatures may help to untangle remaining questions about the biological processes in STAD progression.

TMB and MSI can serve as the predicting factors for selecting patients that likely to benefit from immune checkpoint inhibition therapy. Notably, the mutant *FLG* expression was correlated with high TMB and MSI values. In addition, the patients in the *FLG*-MT group had a negative TIDE score, indicating lack of tumor immune evasion phenotypes. Although we did not find the differences in the immune score and stromal score in the *FLG* mutation group, but at different mutation levels, we observed significant changes in the immune cell infiltration. Filaggrin is a key protein involved in many inflammation dysregulated diseases, and lack of filaggrin protein can induce inflammation and T-cell infiltration (Brown and McLean, 2012). Our analysis also found that cell cycle and DNA repair–related pathways showed significant enrichment in the *FLG* mutation group. We have validated the performance of *FLG* mutation using the GDSC heterogeneous dataset containing molecular descriptors of drugs with transcriptomic expressions of STAD cell lines. We screened 138 drugs and compared the response to common anticancer drugs between the *FLG* mutation and non-mutation groups and identified significantly different responses to three of these drugs. The application of the Bruton's tyrosine kinase (BTK) inhibitor LFM-A13 in solid cancer has been discovered recently. It promotes apoptosis, has an antiproliferative effect, and increases the sensitivity of cancer cells to chemotherapy drugs (Uckun et al., 2011). AZD8055 is a small-molecule inhibitor of mammalian target of rapamycin (mTOR) kinase activity; mTOR plays an important, albeit complex, role in tissue

TABLE 2 Univariate and multivariate analyses with the Cox proportional hazards regression model based on *FLG* gene mutation for predicting OS in TCGA database.

	Univariate cox analysis				Multivariate cox analysis			
	HR	HR.95L	HR.95H	*p*-value	HR	HR.95L	HR.95H	*p*-value
Age (≥65 vs. < 65)	1.55	1.09	2.20	0.013733	1.81	1.27	2.59	0.001054
Gender (male vs. female)	1.21	0.84	1.73	0.300951	1.25	0.87	1.79	0.229855
Grade (G3 & GX vs. G1 & G2)	1.45	1.02	2.05	0.03755	1.34	0.92	1.94	0.122587
T-stage (T3 & T4 & TX vs. T1&T2)	1.74	1.14	2.65	0.010536	1.53	0.99	2.37	0.054883
M-stage (M1 & MX vs. M0)	1.77	1.08	2.90	0.024188	1.86	1.13	3.07	0.014555
N-stage (N2 & N3 & NX vs. N0 & N1)	1.66	1.19	2.31	0.002856	1.47	1.04	2.08	0.029323
FLG mutation (MT vs. WT)	0.66	0.42	1.04	0.075563	0.70	0.44	1.13	0.144729

TABLE 3 Univariate and multivariate analyses with the Cox proportional hazards regression model based on *FLG* gene mutation for predicting DFS in TCGA database.

	Univariate cox analysis				Multivariate cox analysis			
	HR	HR.95L	HR.95H	*p*-value	HR	HR.95L	HR.95H	*p*-value
Age (≥65 vs. <65)	1.09	0.71	1.68	0.690387	1.31	0.84	2.03	0.231999
Gender (male vs. female)	1.53	0.94	2.48	0.088376	1.50	0.92	2.45	0.101933
Grade (G3 & GX vs. G1 & G2)	1.51	0.96	2.37	0.073327	1.23	0.77	1.99	0.387691
T-stage (T3 & T4 & TX vs. T1&T2)	2.12	1.19	3.77	0.010885	1.73	0.96	3.13	0.069976
M-stage (M1 & MX vs. M0)	1.67	0.86	3.23	0.12934	1.68	0.87	3.27	0.124968
N-stage (N2 & N3 & NX vs. N0 & N1)	2.09	1.35	3.22	0.000851	1.81	1.16	2.84	0.009531
FLG mutation (MT vs. WT)	0.52	0.28	0.98	0.044748	0.59	0.31	1.14	0.117505

homeostasis and tumorigenesis (Marshall et al., 2011). The positive expression of p-mTOR was more frequent in advanced gastric cancers (Murayama et al., 2009).

KEGG pathway enrichment analysis reveals that DEGs were mainly involved in several pathways such as neuroactive ligand–receptor interaction, protein digestion and absorption, and chemical carcinogenesis. The neuroactive ligand–receptor interaction pathway was suggested to play a key role in the effect of DNA methylation on GC prognosis (Dai et al., 2021). The protein digestion and absorption pathway was found to be significantly enriched in GC tissues (Liu et al., 2020). In particular, the loss of *FLG* in neck squamous cell carcinoma tissue was found to result in a dramatic resistance to targeted therapies (Bai et al., 2021). GSEA analysis further revealed significant enrichment of ribosome, focal adhesion, dilated cardiomyopathy, regulation of actin cytoskeleton, and ECM receptor interaction in mutant *FLG* tissues. Actin cytoskeletal remodeling was proven to affect epithelial–mesenchymal transition in gastric cancer cells (Yang et al., 2017). Subsequently, in order to find the co-existence pattern among all the DEGs, we constructed the PPI network. From the network, several hub genes with high degrees were found. For example, the G-protein–coupled receptor *GPRC6A* is located in multiple tissues, including gastrointestinal epithelia; studies have shown that *GPRC6A* is involved in regulating glucose and fat metabolism in certain cancers such as prostate cancer progression (Pi et al., 2021). *GALNT9* (an initiator of O-glycosylation) is a member of a sub family that differs significantly in the sequence from other GALNAC-T members, and has been shown to be dysregulated in cancer by promoter methylation (Pangeni et al., 2015).

This was the first time we evaluated the relationship between *FLG*, clinicopathological features, and prognosis in STAD. We explored the prognostic value of *FLG* mutation in the TCGA-STAD database. Compared to the non-mutant group, the results showed that patients with mutant *FLG* had better overall survival and disease-free survival, but had no effect on progression-free survival. Combining *FLG*-mutant tumor types with clinicopathological features showed that *FLG* mutation was a protective factor but not an independent protective factor. Our current nomogram also showed that the *FLG* mutation had a significant influence on 1-, 3-, and 5-year prognosis, future studies are needed to externally validate

TABLE 4 Gene Ontology analyses using differentially expressed genes (DEGs).

Ontology	ID	Description	Count	*p*-value	Gene
BP	GO: 0070268	Cornification	24	1.96E-16	TGM1/KRT4/KLK5/KRT6C/KRT14/KRT13/SPINK5/KRT78/KRT5/KRT6A/KRT24/IVL/LCE3D/SPRR2F/SPRR3/KLK13/KRT6B/SPRR1A/SPRR2E/SPRR2B/SPRR2G/SPRR2D/KRT82/KRT38
BP	GO: 0030216	Keratinocyte differentiation	34	1.05E-13	TGM1/KRT4/KLK5/KRT6C/KRT14/KRT13/SPINK5/KRT78/KRT5/KRT6A/S100A7/KRT24/SCEL/IVL/LCE3D/CERS3/SPRR2F/SPRR3/KLK13/KRT6B/SPRR1A/SERPINB13/SPRR2E/SPRR2B/LCE3E/LELP1/SPRR2G/KRTAP13-1/KRTAP2-3/KRTAP4-6/KRTAP11-1/SPRR2D/KRT82/KRT38
BP	GO: 0043588	Skin development	40	1.17E-13	TGM1/KRT4/KLK5/KRT6C/KRT14/KRT13/LRP4/SPINK5/KRT78/DKK4/KRT5/COMP/KRT6A/S100A7/KRT24/SCEL/IVL/LCE3D/CERS3/SPRR2F/SPRR3/KLK13/DKK1/KRT6B/SPRR1A/SERPINB13/SPRR2E/SPRR2B/DACT2/FOXE1/LCE3E/LELP1/SPRR2G/KRTAP13-1/KRTAP2-3/KRTAP4-6/KRTAP11-1/SPRR2D/KRT82/KRT38
CC	GO: 0005788	Endoplasmic reticulum lumen	30	7.19E-11	ALB/AFP/FGG/APOB/APOA1/VTN/GHRL/ITIH2/FGA/CHGB/CASQ2/BPIFB2/APOA4/SERPINA10/CES1/VGF/AHSG/COL2A1/CASQ1/PENK/COL26A1/F2/NOTUM/AMELX/COL9A3/APOA2/MTTP/PDIA2/GCG/SPP2
CC	GO: 0001533	Cornified envelope	13	3.43E-09	TGM1/SCEL/IVL/LCE3D/SPRR2F/SPRR3/SPRR1A/SPRR2E/SPRR2B/LCE3E/LELP1/SPRR2G/SPRR2D
CC	GO: 0062023	Collagen-containing extracellular matrix	31	1.23E-08	FGG/APOA1/VTN/ITIH2/ANXA8/ADIPOQ/FGA/APOC3/APOA4/CBLN1/NCAM1/AHSG/COL2A1/ORM1/SERPINA3/FGB/PTPRZ1/L1CAM/COMP/COL26A1/ORM2/F2/THBS4/S100A7/AMELX/COL9A3/PRTN3/MMP8/S100A8/PRSS1/SPP2
MF	GO: 0048018	Receptor–ligand activity	41	2.88E-12	SCGB3A1/APOA1/GHRL/IL37/SCT/ANGPTL3/ADIPOQ/CHGB/FGF20/IGF2/SST/VIP/NTS/VGF/CARTPT/IL36A/DEFB4A/PENK/FGF19/F2/NTF4/THBS4/AMELX/PPBP/TTR/IFNW1/EPHA7/FGF3/INHA/EPGN/CSH2/NPPB/DKK1/GCG/LEFTY1/IFNL2/MLN/IFNL3/IL3/SLURP1/INSL5
MF	GO: 0030546	Signaling receptor activator activity	41	4.01E-12	SCGB3A1/APOA1/GHRL/IL37/SCT/ANGPTL3/ADIPOQ/CHGB/FGF20/IGF2/SST/VIP/NTS/VGF/CARTPT/IL36A/DEFB4A/PENK/FGF19/F2/NTF4/THBS4/AMELX/PPBP/TTR/IFNW1/EPHA7/FGF3/INHA/EPGN/CSH2/NPPB/DKK1/GCG/LEFTY1/IFNL2/MLN/IFNL3/IL3/SLURP1/INSL5
MF	GO: 0005179	Hormone activity	18	8.70E-10	GHRL/SCT/ADIPOQ/CHGB/IGF2/SST/VIP/NTS/VGF/CARTPT/PENK/TTR/INHA/CSH2/NPPB/GCG/MLN/INSL5

TABLE 5 Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis using differentially expressed genes.

ID	Description	Count	*p*-value	Gene
hsa04080	Neuroactive ligand–receptor interaction	29	1.92E-08	51738/6343/2862/6750/7432/4922/4886/155/5179/2147/6865/1136/2904/1443/5644/1129/886/2641/6863/2740/11255/2834/9248/4295/134864/3350/3361/57152/10022
hsa04974	Protein digestion and absorption	12	1.56E-05	1280/5222/4311/136227/1299/477/5644/10136/1358/643834/1357/643847
hsa05204	Chemical carcinogenesis	10	6.05E-05	1576/1543/1646/79799/127/54578/7367/6822/131/221357
hsa04972	Pancreatic secretion	11	7.31E-05	6343/1811/22802/3778/477/5644/886/10136/1358/9635/1357
hsa04976	Bile secretion	10	0.000121	6343/1576/79799/9971/477/570/54578/7367/6822/6555
hsa00980	Metabolism of xenobiotics by cytochrome P450	9	0.000198	1576/1543/79799/127/54578/7367/6822/131/221357
hsa04979	Cholesterol metabolism	7	0.000309	338/335/27329/345/337/336/8435
hsa04977	Vitamin digestion and absorption	5	0.000346	338/335/337/5948/9227
hsa00830	Retinol metabolism	8	0.000393	1576/1543/79799/127/9227/54578/7367/131

the proposed nomograms to establish their value in predicting the long-term prognosis of STAD patients.

A few limitations of the current study are worth mentioning. Although microarray-based bioinformatic analysis is a powerful analytic tool for a deeper understanding of molecular mechanisms and for identifying potential biomarkers of STAD, further efforts are needed. First, though the *FLG* gene has been demonstrated to have a high diagnostic and prognostic value in STAD patients, current

TABLE 6 Results of gene set enrichment analysis (GSEA).

Name	Size	Enrichment score	NES	*p*-value	Leading edge
KEGG_RIBOSOME	87	0.956547	1.656318	1.00E-10	Tags = 85%, list = 3%, signal = 83%
KEGG_FOCAL_ADHESION	199	0.830965	1.46027	2.21E-10	Tags = 39%, list = 8%, signal = 36%
KEGG_DILATED_CARDIOMYOPATHY	90	0.87904	1.521094	4.21E-07	Tags = 17%, list = 3%, signal = 16%
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	212	0.78167	1.373872	4.26E-07	Tags = 36%, list = 13%, signal = 32%
KEGG_ECM_RECEPTOR_INTERACTION	83	0.883822	1.530104	5.20E-07	Tags = 42%, list = 6%, signal = 40%
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	74	0.894349	1.542548	9.25E-07	Tags = 23%, list = 3%, signal = 22%
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM	83	0.874796	1.514478	2.17E-06	Tags = 17%, list = 3%, signal = 16%
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	80	0.887199	1.532384	2.31E-06	Tags = 34%, list = 6%, signal = 32%

understanding of detailed mechanisms is limited. For deeper understanding of mechanisms underlying the *FLG* gene functions, the mRNA expression levels of the *FLG* gene need to be performed by RT-PCR, and the protein expression levels of the *FLG* gene need to be performed by Western blot and the immunochemistry method in separate cohorts. Moreover, a more detailed examination by utilizing a combination of *in vitro* and *in vivo* techniques may further elucidate the diagnostic and therapeutic effects of the *FLG* gene in STAD patients. Second, as a retrospective study, potential selection bias and recall bias were inevitable, and more studies in settings with better statistics are necessary. Third, due to the incomplete clinical information on STAD from the TCGA database and the limited sample size, we need more long-term follow-up data and the clinical benefit of early detection for further validation and study.

Conclusion

Our study first comprehensively demonstrated the expression and function, and prognostic value of the *FLG* gene in STAD. We found that the *FLG*-MT group showed a higher mutation load and immunogenicity in STAD patients. We further identified DEGs between the *FLG*-WT and *FLG*-MT groups and performed GO analysis, pathway enrichment analysis, PPI network construction, and prognostic analysis to understand its role in STAD at the molecular level. Further analyses showed that the DFS and OS were significantly different by mutation status and *FLG* gene mutation might be a protective factor. We also demonstrated that *FLG* mutation patients showed comparably high mutation counts than *FLG* intact patients in DNA damage repairment-related pathways. Then we found that *FLG* mutations resulted in increased gastric cancer sensitivity to 24 chemotherapeutic drugs and small-molecule anticancer drugs, especially in LFM. A13, AZD8055, and X17. AAG. This study not only suggests the potential value of the *FLG* gene as a novel biomarker, but also offers new diagnostic and/or therapeutic avenues for STAD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

FY and LX collected the manuscript and analyzed data, analyzed the conclusion, and drafted the manuscript. YT reviewed the data and conclusion. LH presented the idea of this manuscript, supported the funding, analyzed the conclusion, drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was financially supported by the Key Clinical Program of Peking University Third Hospital (Grant No. BYSY2017006).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ajani, J. A., Lee, J., Sano, T., Janjigian, Y. Y., Fan, D., Song, S., et al. (2017). Gastric adenocarcinoma. *Nat. Rev. Dis. Prim.* 3, 17036. doi:10.1038/nrdp.2017.36

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. doi:10.1038/nature12477

Alsina, M., Miquel, J. M., Diez, M., Castro, S., and Tabernero, J. (2019). How I treat gastric adenocarcinoma. *ESMO Open* 4, e000521. doi:10.1136/esmoopen-2019-000521

Bai, Y., Zhao, Z., Boath, J., van Denderen, B. J., and Darido, C. (2021). The functional GRHL3-filaggrin axis maintains a tumor differentiation potential and influences drug sensitivity. *Mol. Ther.* 29, 2571–2582. doi:10.1016/j.ymthe.2021.03.016

Brown, S. J., and McLean, W. H. (2012). One remarkable molecule: Filaggrin. *J. Invest. Dermatol.* 132, 751–762. doi:10.1038/jid.2011.393

Dai, J., Nishi, A., Li, Z. X., Zhang, Y., Zhou, T., You, W. C., et al. (2021). DNA methylation signatures associated with prognosis of gastric cancer. *BMC Cancer* 21, 610. doi:10.1186/s12885-021-08389-0

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hecht, J. R., Bang, Y. J., Qin, S. K., Chung, H. C., Xu, J. M., Park, J. O., et al. (2016). Lapatinib in combination with capecitabine plus oxaliplatin in human epidermal growth factor receptor 2-positive advanced or metastatic gastric, esophageal, or gastroesophageal adenocarcinoma: TRIO-013/LOGiC--A randomized phase III trial. *J. Clin. Oncol.* 34, 443–451. doi:10.1200/JCO.2015.62.6598

Huang, Z., He, A., Wang, J., Lu, H., Zhang, R., Wu, L., et al. (2021). The circadian clock is associated with prognosis and immune infiltration in stomach adenocarcinoma. *Aging (Albany NY)* 13, 16637–16655. doi:10.18632/aging.203184

Jia, F., Teer, J. K., Knepper, T. C., Lee, J. K., Zhou, H. H., He, Y. J., et al. (2017). Discordance of somatic mutations between asian and caucasian patient populations with gastric cancer. *Mol. Diagn. Ther.* 21, 179–185. doi:10.1007/s40291-016-0250-z

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Kim, J., Copeland, C. E., Seki, K., Vogeli, B., and Kwon, Y. C. (2020). Tuning the cell-free protein synthesis system for biomanufacturing of monomeric human filaggrin. *Front. Bioeng. Biotechnol.* 8, 590341. doi:10.3389/fbioe.2020.590341

Krishnamurthy, K., Urioste, S. N., Cusnir, M., Schwartz, M., Alghamdi, S., Sriganeshan, V., et al. (2022). The mutational landscape of upper gastrointestinal adenocarcinomas-A study of similarities and differences. *Pathol. Res. Pract.* 232, 153830. doi:10.1016/j.prp.2022.153830

Kuang, X., Sun, L., Liu, S., Zhao, Z., Zhao, D., Liu, S., et al. (2016). Association of single nucleotide polymorphism rs2065955 of the filaggrin gene with susceptibility to Epstein-Barr virus-associated gastric carcinoma and EBV-negative gastric carcinoma. *Virol. Sin.* 31, 306–313. doi:10.1007/s12250-016-3721-9

Li, L., Zhu, Z., Zhao, Y., Zhang, Q., Wu, X., Miao, B., et al. (2019). FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci. Rep.* 9, 7827. doi:10.1038/s41598-019-43924-x

Li, J., Zhou, W., Wei, J., Xiao, X., An, T., Wu, W., et al. (2021). Prognostic value and biological functions of RNA binding proteins in stomach adenocarcinoma. *Onco. Targets. Ther.* 14, 1689–1705. doi:10.2147/OTT.S297973

Liguigli, W., Tomasello, G., Toppo, L., Ratti, M., and Passalacqua, R. (2014). Ramucirumab for metastatic gastric or gastroesophageal junction cancer: results and implications of the REGARD trial. *Future Oncol.* 10, 1549–1557. doi:10.2217/fon.14.106

Liu, J., Ma, L., Chen, Z., Song, Y., Gu, T., Liu, X., et al. (2020). Identification of critical genes in gastric cancer to predict prognosis using bioinformatics analysis methods. *Ann. Transl. Med.* 8, 884. doi:10.21037/atm-20-4427

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Marshall, G., Howard, Z., Dry, J., Fenton, S., Heathcote, D., Gray, N., et al. (2011). Benefits of mTOR kinase targeting in oncology: pre-clinical evidence with AZD8055. *Biochem. Soc. Trans.* 39, 456–459. doi:10.1042/bst0390456

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Mizrak, K. D., Harada, K., Shimodaira, Y., Amlashi, F. G., Lin, Q., Ajani, J. A., et al. (2017). Advanced gastric adenocarcinoma: optimizing therapy options. *Expert Rev. Clin. Pharmacol.* 10, 263–271. doi:10.1080/17512433.2017.1279969

Murayama, T., Inokuchi, M., Takagi, Y., Yamada, H., Kojima, K., Kumagai, J., et al. (2009). Relation between outcomes and localisation of p-mTOR expression in gastric cancer. *Br. J. Cancer* 100, 782–788. doi:10.1038/sj.bjc.6604915

Pangeni, R. P., Channathodiyil, P., Huen, D. S., Eagles, L. W., Johal, B. K., Pasha, D., et al. (2015). The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. *Clin. Epigenet.* 7, 57. doi:10.1186/s13148-015-0089-x

Pi, M., Nishimoto, S. K., and Darryl, Q. L. (2021). Explaining divergent observations regarding osteocalcin/GPRC6A endocrine signaling. *Endocrinology* 162, bqab011. doi:10.1210/endocr/bqab011

Pilehchian, L. M., Nikbakhsh, N., Samadani, A. A., Fattahi, S., Taheri, H., Shafaei, S., et al. (2017). FAT4 hypermethylation and grade dependent downregulation in gastric adenocarcinoma. *J. Cell Commun. Signal.* 11, 69–75. doi:10.1007/s12079-016-0355-5

Qiu, J., Sun, M., Wang, Y., and Chen, B. (2020). Identification of hub genes and pathways in gastric adenocarcinoma based on bioinformatics analysis. *Med. Sci. Monit.* 26, e920261. doi:10.12659/msm.920261

Qu, Y., Gao, N., and Wu, T. (2021). Expression and clinical significance of SYNE1 and MAGI2 gene promoter methylation in gastric cancer. *Med. Baltim.* 100, e23788. doi:10.1097/MD.0000000000023788

Ramezankhani, R., Solhi, R., Es, H. A., Vosough, M., and Hassan, M. (2021). Novel molecular targets in gastric adenocarcinoma. *Pharmacol. Ther.* 220, 107714. doi:10.1016/j.pharmthera.2020.107714

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., Mesirov, J. P., et al. (2006). GenePattern 2.0. *Nat. Genet.* 38, 500–501. doi:10.1038/ng0506-500

Russi, S., Verma, H. K., Laurino, S., Mazzone, P., Storto, G., Nardelli, A., et al. (2019). Adapting and surviving: Intra and extra-cellular remodeling in drug-resistant gastric cancer cells. *Int. J. Mol. Sci.* 20, E3736. doi:10.3390/ijms20153736

Salama, R. H., Rasheed, Z., Ahmed, A. A., Bin Saif, G. A., Elkholy, M. M., Abd El-Moniem, A. E., et al. (2021). Missense, silent, non-sense and frame-shift mutations in exon 3 of the filaggrin gene in patients with bronchial asthma, atopic dermatitis, allergic rhinitis and mixed atopy. *Nucleosides Nucleotides Nucleic Acids* 40, 357–367. doi:10.1080/15257770.2021.1880009

Sandilands, A., Sutherland, C., Irvine, A. D., and McLean, W. H. (2009). Filaggrin in the frontline: role in skin barrier function and disease. *J. Cell Sci.* 122, 1285–1294. doi:10.1242/jcs.033969

Satoh, T., Xu, R.-H., Chung, H. C., Sun, G.-P., Doi, T., Xu, J.-M., et al. (2014). Lapatinib plus paclitaxel versus paclitaxel alone in the second-line treatment ofHER2 -amplified advanced gastric cancer in asian populations: TyTAN—a randomized, phase III study. *J. Clin. Oncol.* 32, 2039–2049. doi:10.1200/JCO.2013.53.6136

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., and Lordick, F. (2020). Gastric cancer. *Lancet* 396, 635–648. doi:10.1016/S0140-6736(20)31288-5

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Uckun, F., Dibirdik, I., Sarkissian, A., and Qazi, S. (2011). *In vitro* and *in vivo* chemosensitizing activity of LFM-A13, a dual-function inhibitor of Bruton's tyrosine kinase and polo-like kinases, against human leukemic B-cell precursors. *Arzneimittelforschung.* 61, 252–259. doi:10.1055/s-0031-1296196

Vanden, H. J., Maddox, E., Maalouf, S. W., Iorns, E., and Tsui, R. (2018). Replication Study: Systematic identification of genomic markers of drug sensitivity in cancer cells. *Elife* 7, e29747. doi:10.7554/eLife.29747

Wang, Q., Liu, G., and Hu, C. (2019). Molecular classification of gastric adenocarcinoma. *Gastroenterol. Res.* 12, 275–282. doi:10.14740/gr1187

Wang, H., Shen, L., Li, Y., and Lv, J. (2020). Integrated characterisation of cancer genes identifies key molecular biomarkers in stomach adenocarcinoma. *J. Clin. Pathol.* 73, 579–586. doi:10.1136/jclinpath-2019-206400

Yang, Y., Liu, W., Zhao, Z., Zhang, Y., Xiao, H., Luo, B., et al. (2017). Filaggrin gene polymorphism associated with Epstein-Barr virus-associated tumors in China. *Virus Genes* 53, 532–537. doi:10.1007/s11262-017-1463-x

Yang, Y., Zhang, J., Chen, Y., Xu, R., Zhao, Q., Guo, W., et al. (2020). MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. *Clin. Transl. Med.* 10, e155. doi:10.1002/ctm2.155

Ye, Z., Zheng, M., Zeng, Y., Wei, S., Wang, Y., Lin, Z., et al. (2020). Bioinformatics analysis reveals an association between cancer cell stemness, gene mutations, and the immune microenvironment in stomach adenocarcinoma. *Front. Genet.* 11, 595477. doi:10.3389/fgene.2020.595477

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for comparing biological themes among gene clusters. *OMICS A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zhu, Z., Lee, P. H., Chaffin, M. D., Chung, W., Loh, P. R., Lu, Q., et al. (2018). A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat. Genet.* 50, 857–864. doi:10.1038/s41588-018-0121-0

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 August 2022
doi: 10.3389/fgene.2022.862977

Integrated Analysis Reveals the Potential Significance of HDAC Family Genes in Lung Adenocarcinoma

*Congkuan Song†, Weichen Lin†, Heng Meng, Ning Li and Qing Geng**

Department of Thoracic Surgery, Renmin Hospital of Wuhan University, Wuhan, China

Histone deacetylases comprise a family of 18 genes, and classical HDACs are a promising class of novel anticancer drug targets. However, to date, no systematic study has been comprehensive to reveal the potential significance of these 18 genes in lung adenocarcinoma (LUAD). Here, we used a systematic bioinformatics approach to comprehensively describe the biological characteristics of the HDACs in LUAD. Unsupervised consensus clustering was performed to identify LUAD molecular subtypes. The ssGSEA, CIBERSORT, MCP counter, and ESTIMATE algorithms were used to depict the tumor microenvironment (TME) landscape. The Cox proportional hazards model and LASSO regression analyses were used to construct the HDAC scoring system for evaluating the prognosis of individual tumors. In this study, three distinct HDAC-mediated molecular subtypes were determined, which were also related to different clinical outcomes and biological pathways. HDACsCluster-C subtype had lowest PD-L1/PD-1/CTLA4 expression and immune score. The constructed HDAC scoring system (HDACsScore) could be used as an independent predictor to assess patient prognosis and effectively identify patients with different prognosis. High- and low-HDACsScore groups presented distinct genetic features, immune infiltration, and biological processes. The high-HDACsScore group was more likely to benefit from immunotherapy, as well as from the application of common chemotherapeutic agents (cyclopamine, docetaxel, doxorubicin, gemcitabine, paclitaxel, and pyrimethamine). Overall, HDAC family genes play important roles in LUAD, and the three LUAD subtypes and the HDAC scoring system identified in this study would help enhance our perception of LUAD prognostic differences and provide important insights into the efficacy of immunotherapy and chemotherapy.

Keywords: histone deacetylases (HDAC), lung adenocarcinoma, prognosis, immune, drug

OPEN ACCESS

Edited by:
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

Reviewed by:
Xiaoju Hu,
The State University of New Jersey, United States
Tong Lu,
The Second Affiliated Hospital of Harbin Medical University, China

****Correspondence:***
Qing Geng
gengqingwhu@whu.edu.cn

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

Received: *26 January 2022*
Accepted: *22 June 2022*
Published: *22 August 2022*

Citation:
Song C, Lin W, Meng H, Li N and Geng Q (2022) Integrated Analysis Reveals the Potential Significance of HDAC Family Genes in Lung Adenocarcinoma. Front. Genet. 13:862977. doi: 10.3389/fgene.2022.862977

INTRODUCTION

Histone acetylation is accomplished by histone acetyltransferases (HATs) and histone deacetylases (HDACs), controlling the transcription level of genes and playing a key role in structural modification of chromosomes and the regulation of gene expression (Kim et al., 2020; Weinberger et al., 2012; Xu et al., 2012). Of them, HDACs, as gene-silencing complexes, can inhibit gene expression through transcription factors such as E2F1 and can also eliminate the acetylation of nonhistones (Fang et al., 2020; Kim et al., 2019). At present, it has been found that there are four categories of HDACs, including 18 subtypes. Class I includes HDAC1, HDAC2, HDAC3,

and HDAC8; class II includes HDAC4, HDAC5, HDAC6, HDAC7, HDAC9, and HDAC10; class III includes sirtuin (SIRT) 1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, and SIRT7; and class IV includes HDAC11. HDAC overexpression promotes histone deacetylation, enhancing the interaction between histones and DNA, and thus inhibiting the transcriptional processes of related genes (Kim et al., 2016; Sanchez et al., 2020). HDACs are often overexpressed in tumor cells (Yang et al., 2014). Many studies (Li et al., 2020; Maiti et al., 2019; Torres-Adorno et al., 2017) have confirmed the involvement of HDACs in different stages of cancer. HDACs have become a popular target for antitumor drug development (Dokmanovic and Marks., 2005; Bolden et al., 2006; Li and Zhu, 2014).

Lung adenocarcinoma (LUAD), as one of the most common histological subtypes in non–small cell lung cancer (NSCLC), has complex heterogeneity (Raparia et al., 2013; Travis et al., 2005), posing great challenges for clinical treatment. Although the diagnosis and treatment of LUAD have made great progress in recent years, there is still room for great improvement in the long-term prognosis of patients. Previous studies (Li and Seto., 2016; Mithraprabhu et al., 2014) have shown that an increased expression of HDAC was observed in multiple solid tumors and was related to the poor prognosis of patients. Compared with normal lung cells, lung cancer cells have abnormal histone modification patterns (Li and Seto., 2016), which play a crucial role in lung carcinogenesis. The mechanisms by which HDAC regulates tumorigenesis and progression are complex and diverse, and it can regulate oncogenic cell signaling pathways both by inhibiting the expression levels of tumor suppressor genes and by modifying key molecules (Li and Seto., 2016). A full understanding of the important role of HDAC in lung cancer is of great significance for analyzing the mechanism of tumorigenesis, improving the possibility of clinical application of HDAC inhibitors and individualized treatment strategies for lung cancer. HDAC family genes are still poorly studied in LUAD. To date, no systematic study has been comprehensive to reveal the potential significance of these 18 HDACs in LUAD. Therefore, this study used a systematic bioinformatics approach to comprehensively describe the biological characteristics of the HDACs in the LUAD, performing unsupervised clustering based on HDACs to identify HDAC-mediated molecular subtypes, while constructing the HDAC scoring system to explore their regulatory relationship with the tumor.

MATERIALS AND METHODS

Data Collection and Preprocessing

First, we searched for appropriate gene microarray containing transcriptomic data and clinical information for LUAD from the GEO database as well as publicly published literature, excluding datasets with sample sizes of less than 30 cases and lacking HDAC family genes (HDAC 1-11 and SIRT 1-7), and we finally included a six-gene microarray [GSE29013 (Xie et al., 2011), GSE30219 (Rousseaux et al., 2013), GSE31210 (Okayama et al., 2012; Yamauchi et al., 2012), GSE37745 (Botling et al., 2013; Goldmann et al., 2021; Jabs et al., 2017; Lohr et al., 2015), GSE50081 (Der et al., 2014), and GSE72094 (Schabath et al., 2016)] in this study. In addition, we also downloaded data on LUAD transcriptomes, somatic mutations, and clinical information from TCGA GDC (https://portal.gdc.cancer.gov/). In **Supplementary Table S1**, we summarize the number of samples and platform names of patients with LUAD from the abovementioned datasets. The CNV data were obtained from the UCSC Xena (http://xena.ucsc.edu/). For genes with multiple probe sets of signals, we averaged them to generate single-expression values. For different gene microarrays, we used the ""ComBat" algorithm of the "sva" package in R to further integrate into a metacohort to reduce batch effects resulting from nonbiotechnological bias (Irizarry et al., 2003; Johnson et al., 2007). Expression data before and after the removal of batches were analyzed by PCA with the base function "prcomp" of R.

Unsupervised Consensus Clustering

By integrating into a metacohort from the abovementioned six-gene microarray, based on the expression of related genes, we applied unsupervised consensus clustering analysis to identify different molecular clusters and classify patients for further analysis. The "ConsensusClusterPlus" package in R (Wilkerson and Hayes., 2010) was used to perform the abovementioned steps and also perform 1,000 replicates to ensure stability of the classification.

Functional and Pathway Enrichment Analysis

To investigate the differences in biological processes between different subgroups, GSVA enrichment analysis was performed using the "GSVA" R package (Hanzelmann et al., 2013). The gene set "c2. cp.kegg.v7.2. symbols" was derived from the MSigDB (http://www.gsea-msigdb.org/gsea/msigdb). The adjusted p-value of less than 0.05 was considered statistically significant. GO and KEGG enrichment analyses were also used to explore the activation of biological processes between groups. The associated genes were functionally annotated using the "clusterProfiler" R package (Yu et al., 2012), with a cutoff value of FDR <0.05.

Characterization of the Tumor Microenvironment and Immune Landscape

To better characterize the tumor microenvironment (TME) and immune landscape in different subtypes, we used multiple algorithms, including ssGSEA (Barbie et al., 2009), CIBERSORT (Newman et al., 2015), and MCP counter (Becht et al., 2016), to quantify the abundance of each immune infiltrating cell. Meanwhile, the ESTIMATE algorithm (Becht et al., 2016) was used to impute the tumor purity, stromal score, and immune score.

Establishment of an Evaluation System Associated With the HDAC Phenotypes

Univariate Cox analysis was performed on HDACsCluster phenotype-related genes, and then the genes with $p <$

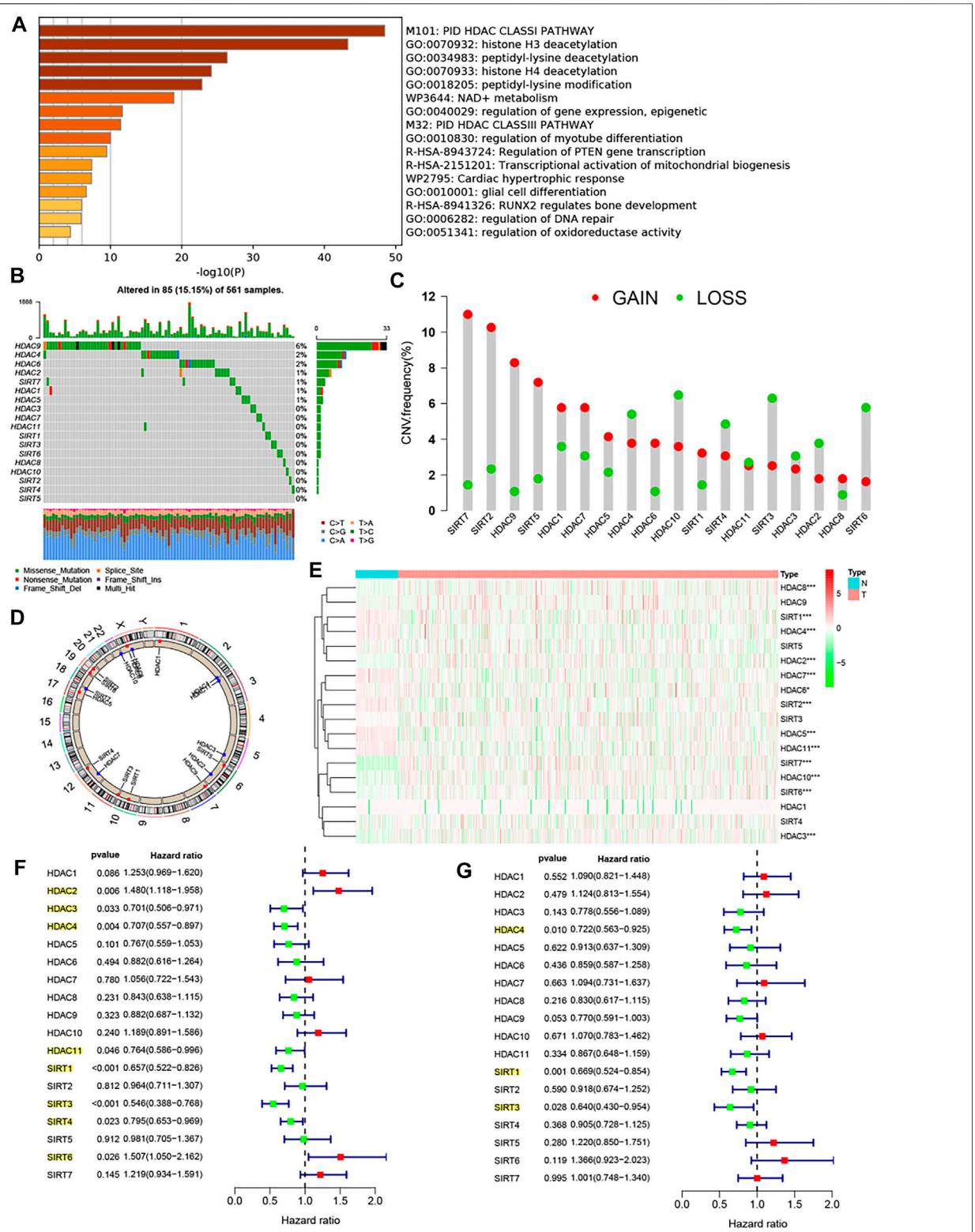


FIGURE 1 | Biological characterization of the HDAC family genes. **(A)** Metascape analysis revealed the biological function of the 18 HDAC family genes. **(B)** Mutation frequency of 18 HDAC family genes in 561 LUAD patients from the TCGA–LUAD cohort. Each column represented individual patients. The upper barplot showed TMB, and the number on the right indicated the mutation frequency in each regulator. The right barplot showed the proportion of each variant type. The stacked

(*Continued*)

FIGURE 1 | barplot showed fraction of conversions in each sample. **(C)** CNV variation frequency of 18 HDAC family genes in the TCGA–LUAD cohort. The height of the column represented the alteration frequency. The deletion frequency, green dot; the amplification frequency, red dot. **(D)** Location of CNV alteration of 18 HDAC family genes on 23 chromosomes using the TCGA–LUAD cohort. **(E)** Expression of 18 HDAC family genes between lung normal tissues and LUAD tissues. The asterisks represented the statistical *p*-value (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(F)** Univariate Cox analysis of 18 HDAC family genes in the TCGA–LUAD cohort. **(G)** Multivariate Cox analysis of 18 HDAC family genes in the TCGA–LUAD cohort.

0.05 were included in Lasso regression analysis for further dimensionality reduction. The filtered genes were subjected to multivariate Cox analysis (stepwise regression) to construct a scoring system for assessing patient outcomes. The calculation formula for HDACsScore was as follows: HDACsScore = Coef***G1****Expression***G1*** + Coef***G2****Expression***G2*** + ... + Coef***Gn****Expression***Gn***, where "Expression***Gn***" is the expression of "gene***n***", and "Coef***Gn***" presented the coefficient of "gene***n***".

Statistical Analysis

The "limma" package in R was used for gene differential expression analysis. The statistical difference between two groups was calculated using the Wilcoxon rank sum test. For comparisons of more than two groups, the Kruskal–Wallis test was used. Survival comparison between two or multiple groups was performed using the Kaplan–Meier method. We used the "pRRophetic" package in R (Geeleher et al., 2014) to impute the semi-inhibitory concentration (IC_{50}) of the drugs to assess the sensitivity of patients to a given drug. A website named Tumor Immune Dysfunction and Exclusion (TIDE) (http://tide.dfci.harvard.edu) was applied to compute TIDE scores and assess patient sensitivity to immunotherapy (Jiang et al., 2018). The $p < 0.05$ was considered statistically significant. All data processing was carried out using R3.6.2 software.

RESULTS

Biological Characterization of the HDAC Family Genes

By consulting the relevant reports, we obtained a total of 18 HDAC family genes, and they are HDAC1, HDAC2, HDAC3–HDAC11, SIRT1, SIRT2, and SIRT3–SIRT7. Subsequently, we revealed the biological function of the family genes using metascape analysis, as shown in **Figure 1A**, which were primarily involved in PID HDAC Class-I pathway, histone H3 and H4 deacetylation, peptidyl-lysine deacetylation, peptidyl-lysine modification, etc. We summarized the incidence of CNVs and somatic mutations of the 18 HDAC family genes in LUAD. Among 561 samples, 85 experienced mutations of HDAC family genes, with a frequency of 15.15%. Of these genes, HDAC9 exhibited the highest mutation frequency (6%) followed by HDAC4 and HDAC6 (2%), while HDAC3, HDAC7, HDAC10, HDAC11, as well as SIRT1-6 did not show any mutations in LUAD samples (**Figure 1B**). Considering the relatively higher mutation frequency of HDAC9, we further investigated whether genetic variations in HDAC9 could affect the expression of other HDAC genes. The results of the abovementioned analysis showed that this HDAC9 low-frequency mutation would promote HDAC2 expression and did not appear to affect the expression of other HDACs (**Supplementary Figure S1**). Further analysis of the 18 HDAC family genes revealed that CNV mutations were prevalent. SIRT7, SIRT2, HDAC9, SIRT5, HDAC1, HDAC7, HDAC5, HDAC6, SIRT1, and HDAC8 showed widespread CNV amplification. In contrast, HDAC4, HDAC10, SIRT4, HDAC11, SIRT3, HADC3, HDAC2, and SIRT6 had prevalent CNV deletions (**Figure 1C**). The locations of CNV alterations of the 18 HDAC family genes on chromosomes are shown in **Figure 1D**. To ascertain whether the abovementioned copy number variants affected the expression of HDAC family genes in LUAD patients, we performed correlation exploration using the Kruskal–Wallis test (**Supplementary Figure S2**) and investigated the transcriptomic levels of HDAC family genes between lung normal and LUAD samples in the TCGA–LUAD cohort (**Figure 1E**). The results revealed that the alterations of CNV could be the prominent factors resulting in perturbations on the HDAC family gene expression. Copy number amplification of these genes, such as HDAC1, HDAC4, HDAC5, HDAC6, HDAC7, HDAC10, SIRT2, SIRT4, SIRT5, SIRT6, and SIRT7, upregulated the expression of the corresponding genes, while their copy number deletion downregulated gene expression. Copy number amplification of these genes, such as HDAC8, HDAC9, and SIRT1, upregulated the expression of the corresponding genes, while their copy number deletion did not affect the gene expression. HDAC2 and HDAC3 expressions did not appear to be affected by copy number variation. Moreover, differential expression analysis of the abovementioned genes uncovered that the expressions of HDAC8, HDAC2, SIRT7, SIRT6, HDAC10, and HDAC3 were upregulated in LUAD tissues and the expressions of SIRT1, HDAC4, HDAC7, HDAC6, SIRT2, HDAC5, and HDAC11 were upregulated in lung normal tissues. From these, we observed that the copy number alterations of these genes did not cause differences in their expression between tumor and normal tissues.

Increasing evidence has reported a close link between HDAC family genes and cancer; however, there is not still a more systematic study to investigate the impact of these 18 family genes in the prognosis of LUAD. Univariate Cox regression analysis and Kaplan–Meier survival analysis were used to ascertain the relationship between these genes and the prognosis of LUAD patients. The results showed that whether in the univariate Cox regression analysis (**Figure 1F**) or in the Kaplan–Meier survival analysis (**Supplementary Figure S3**), HDAC3, HDAC4, HDAC11, SIRT1, SIRT3, and SIRT4 could be considered protective factors and were significantly associated with prolonged overall survival, while HDAC2 and SIRT6 were

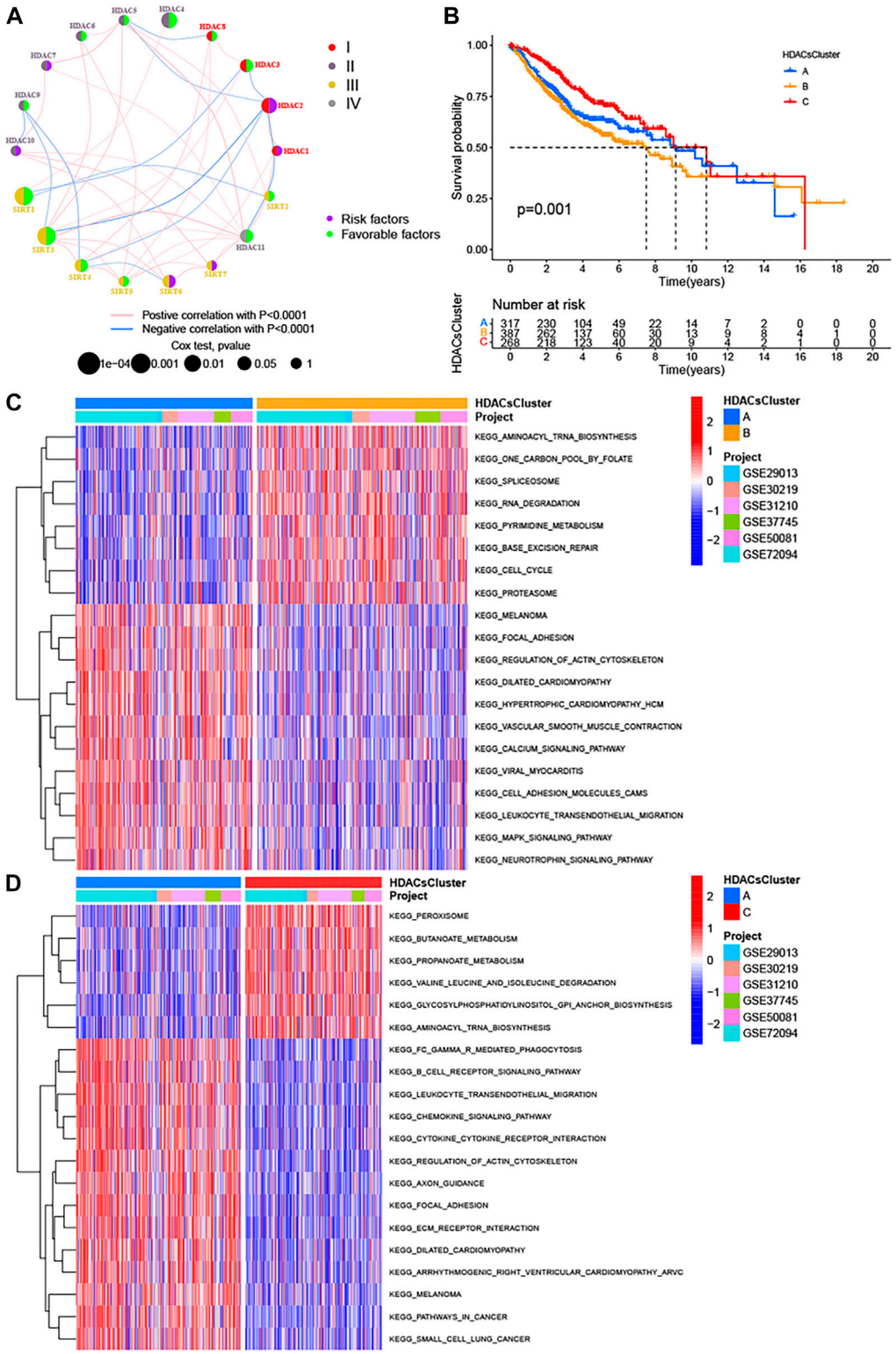


FIGURE 2 | Three distinct LUAD subtypes and biological characteristics of each subtype. **(A)** Interaction between 18 HDAC family genes in LUAD. The circle size represented the effect of each regulator on the prognosis, and the range of values calculated by the log-rank test was $p < 0.0001$, $p < 0.001$, $p < 0.01$, $p < 0.05$, and $p < 1$, respectively. Left half of the circle: purple represents prognostic risk factors, and green represents prognostic favorable factors. Right half of the circle: the types of 18 HDAC family genes. The lines linking regulators showed their interactions, and thickness showed the correlation strength between regulators. Negative correlation was marked with blue and positive correlation with pink. **(B)** Kaplan–Meier survival analyses for the three distinct LUAD subtypes based on 972 patients from

(*Continued*)

FIGURE 2 | six GEO cohorts (GSE29013, GSE30219, GSE31210, GSE37745, GSE50081, and GSE72094) including 317 cases in HDACsCluster-A, 387 cases in HDACsCluster-B, and 268 cases in HDACsCluster-C. **(C–D)** GSVA enrichment analysis showing the activation states of biological pathways in distinct LUAD subtypes. The heat map was used to visualize these biological processes, and red represented activated pathways and blue represented inhibited pathways. The LUAD cohorts were used as sample annotations. **(C)** HDACsCluster-A vs HDACsCluster-B; **(D)** HDACsCluster-C vs HDACsCluster-A.

recognized as risk factors. Subsequent multivariate Cox analysis revealed that HDAC4, SIRT1, SIRT3, and SIRT4 were independent good prognostic factors for LUAD patients (**Figure 1G**). The correlation between the expression of HDAC family genes and immune cell infiltration remains poorly explored, and our study showed that the vast majority of HDAC genes were negatively correlated with the vast majority of immune cells, whereas HDAC3, HDAC9, and SIRT2 showed the opposite trend (**Supplementary Figure S4**). In addition, the vast majority of HDAC genes also showed slight positive correlations with each other (**Supplementary Figure S4** and **Figure 2A**).

The abovementioned analyses characterized the biological characteristics of HDAC family genes in LUAD from biological function, transcriptome expression pattern, genomic alteration, prognosis, and immune cell infiltration, and these results indicated the important roles of HDAC family genes in LUAD.

Three Distinct LUAD Subtypes Were Identified Based on the Unsupervised Consensus Clustering of 18 HDAC Family Genes

Many studies (Raparia et al., 2013; Travis et al., 2005) have confirmed there is significant heterogeneity in LUAD, which also brings great challenges to the prognostic judgment and treatment decisions in LUAD patients. It is necessary to further determine the tumor subtypes with different characteristics to adopt precise strategies. Given the important roles of HDAC family genes in LUAD, we believed that they probably played critical roles in the formation of different tumor subtypes and were implicated in cancer pathogenesis and progression. Based on these hypotheses, we used the "ConsensusClusterPlus" package in R to classify patients with qualitatively different HDAC modification patterns based on the expression of 18 HDAC family genes, and three distinct tumor subtypes were eventually identified using unsupervised clustering (**Supplementary Figure S5A**), including 317 cases in subtype A, 387 cases in subtype B, and 268 cases in subtype C. We termed these subtypes as HDACsCluster A–C, among which HDACsCluster-C exhibited a prominent survival advantage, whereas HDACsCluster-B had the worst prognosis (**Figure 2B**). Additionally, we also noticed significant differences in the expression of HDAC family genes between distinct tumor subtypes. HDAC1, HDAC2, SIRT6, and SIRT7 were significantly elevated in the HDACsCluster-B subtype. HDAC9 and SIRT2 were markedly increased in the HDACsCluster-A subtype. Also, HDAC3, HDAC5, HDAC6, HDAC8, HDAC10, HDAC11, SIRT1, SIRT3, SIRT4, and SIRT5 were evidently increased in the HDACsCluster-C subtype (**Supplementary Figure S5B**). We could conclude that a vast number of HDAC family genes were upregulated in the HDACsCluster-C subtype and that most of these genes were prognostic-friendly genes (**Supplementary Figure S3**), which could explain why HDACsCluster-C subtype presented a best prognosis.

The Biologic Pathways and TME Landscape in Distinct Tumor Subtypes

To explore the biological behaviors across these different subtypes, we performed GSVA enrichment analyses in the meta-GEO cohort. **Supplementary Figure S6** reflects the PCA plots before and after the batch effect removal. As shown in **Figures 2C,D**; **Supplementary Figure S5C**, HDACsCluster-A presented the activation of important signals and pathways including leukocyte transendothelial migration, B cell receptor signaling pathway, and glycosphingolipid biosynthesis ganglio series. HDACsCluster-B was markedly enriched in aminoacyl tRNA biosynthesis, DNA replication, mismatch repair, homologous recombination, base excision repair etc, while HDACsCluster-C was prominently related to body material metabolism including butanoate metabolism; propanoate metabolism; valine, leucine, and isoleucine degradation; and aminoacyl tRNA biosynthesis. These results suggested significant differences in pathway activation across the three tumor subtypes. Subsequently, in the analysis of tumor immune cell infiltration, to our surprise, there were 19 immune cells (both immunosuppressive cells and immune-activated cells) with the highest infiltration levels in HDACsCluster-A (**Supplementary Figure S5D**). Consistent with this finding, ESTIMATE analysis also revealed HDACsCluster-A presented a highest immune score and lowest tumor purity (**Figures 3A,B**). Given that the expression of the immune checkpoints could somewhat reflect the therapeutic efficacy of patients against PD-1/L1 treatment, we also compared the expression levels of PD-L1, PD-1, and CTLA4 in three different clusters. We observed that compared with HDACsCluster-A and -B, HDACsCluster-C had lowest immune checkpoint expression (**Figures 3C–E**). This might somehow imply a poor response of HDACsCluster-C subtype to immunotherapy.

HDACsCluster Phenotype-Related DEGs and Characteristics in LUAD

Although the consensus clustering algorithm based on HDAC family gene expression classified LUAD patients into three HDACsCluster phenotypes, the underlying molecular features and expression perturbations within these phenotypes were not well known. Thus, we further examined the potential HDAC-related transcriptional expression change across three subtypes in

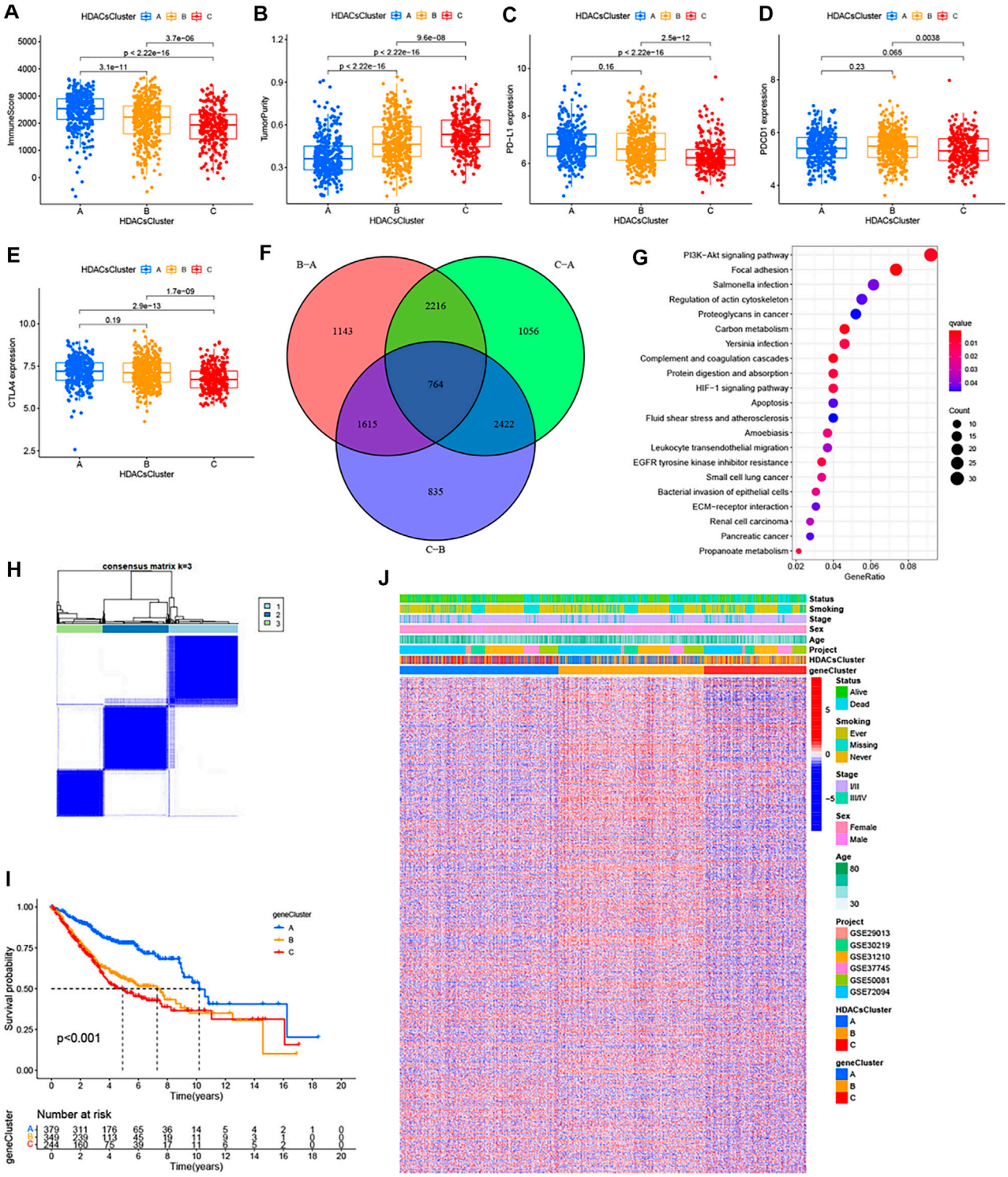


FIGURE 3 | Immune landscapes among the three distinct molecular subtypes and HDAC phenotype–related reclustering. **(A)** Immune score, **(B)** tumor purity, **(C)** PD-L1 expression, **(D)** PDCD1 expression, and **(E)** CTLA-4 expression differences in three distinct molecular subtypes. **(F)** 764 HDAC phenotype–related differentially expressed genes (DEGs) between three distinct molecular subtypes were shown in the Venn diagram. **(G)** Functional annotation for phenotype-related DEGs associated with the prognosis using KEGG enrichment analysis. **(H)** Unsupervised clustering of the HDAC phenotype–related genes and consensus matrices for k = 3. **(I)** Survival curves of the HDAC phenotype–related gene signatures were estimated by the Kaplan–Meier plotter. **(J)** Unsupervised clustering of HDAC phenotype–related DEGs. The geneCluster, HDACsCluster, project, age, sex, stage, smoking status, and survival status were used as patient annotations. Red represented high expression of regulators and green represented low expression.

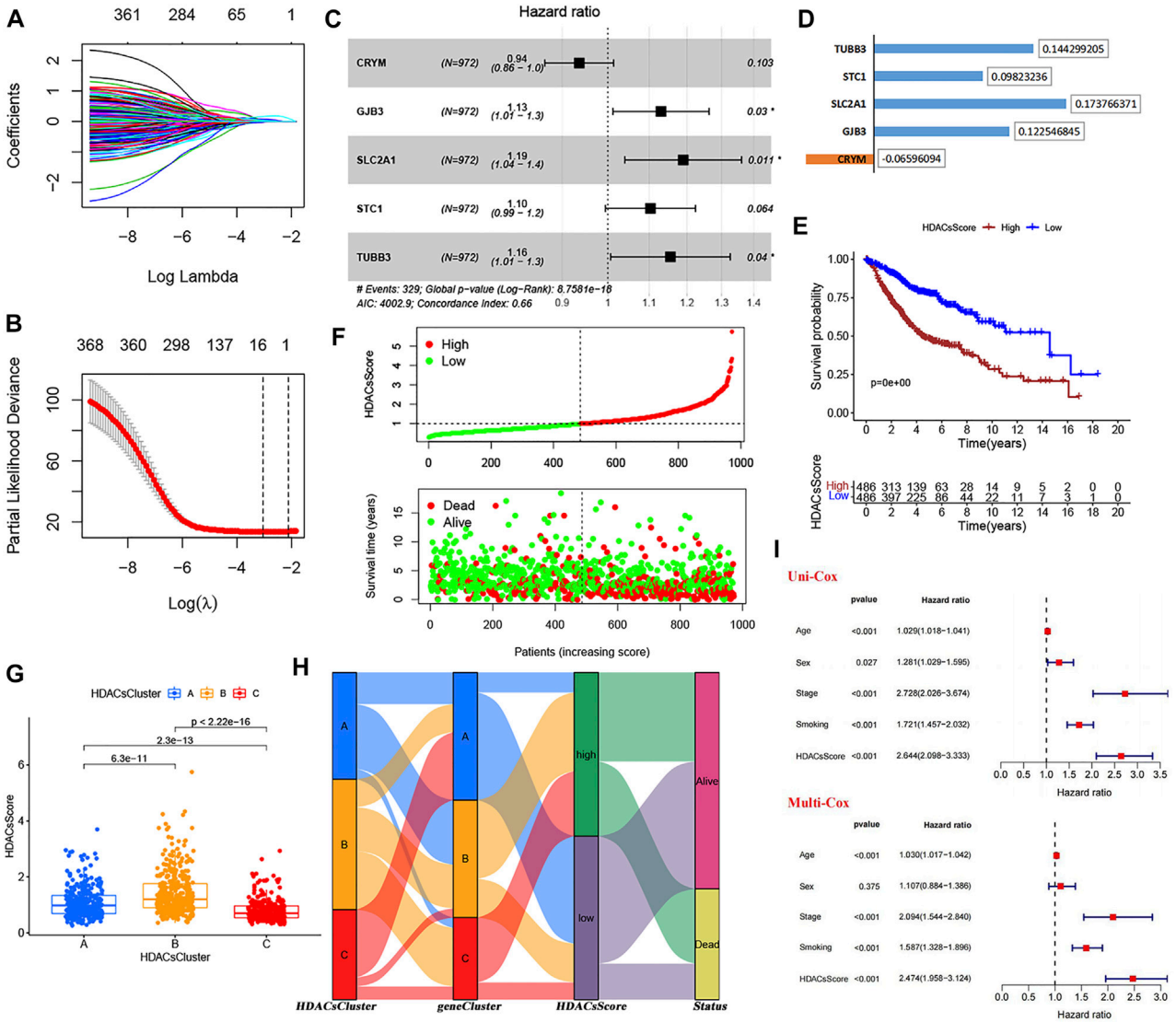


FIGURE 4 | Construction and validation of the HDAC scoring system. **(A–B)** LASSO regression analysis of 371 prognosis-related DEGs to avoid the model overfitting. **(C)** Multivariate Cox analysis (stepwise regression) of the filtered genes in the meta-GEO cohort. **(D)** Correlation coefficient of the model genes. **(E)** Survival analyses for low- (486 cases) and high-(486 cases) HDACsScore groups using Kaplan–Meier curves. **(F)** HDACsScore and OS status distribution of the model in the meta-GEO cohort. **(G)** Correlations between HDACsScore and HDAC-mediated molecular subtypes (HDACsCluster). **(H)** Alluvial diagram showing the changes of HDACsClusters, geneCluster, HDACsScore, and OS status. **(I)** Univariate and multivariate Cox analysis of clinical parameters and HDACsScore.

LUAD. The empirical Bayesian approach was applied to determine overlapping differentially expressed genes (DEGs) among the three subtypes. We determined 764 HDACsCluster phenotype-related DEGs using the limma package in R (**Figure 3F**). GO and KEGG enrichment analyses of these DEGs revealed that these genes were mainly involved in some vital biological processes and pathways, such as PI3K/Akt signaling pathway, focal adhesion, regulation of actin cytoskeleton, extracellular matrix organization, and collagen–containing extracellular matrix (**Figure 3G**; **Supplementary Figure S7A**). Subsequently, based on these genes, we once again performed the unsupervised clustering analysis and obtained three stable transcriptomic phenotypes (**Figure 3H**). These stratifications divided patients into three distinct subgroups (geneCluster-A, geneCluster-B, and geneCluster-C). Kaplan–Meier survival analysis revealed the patients in geneCluster-A showed the longest overall survival compared to the others (**Figure 3I**). The relationship between these three new clusters, HDACsCluster, clinical parameters, and the gene expression distribution was visualized as a heat map in **Figure 3J**, where we found the vast majority of the samples in HDACsCluster-C were corresponding to the geneCluster-A, the vast majority of the samples in HDACsCluster-B were corresponding to the geneCluster-C, and the vast majority of the samples in HDACsCluster-A were corresponding to the geneCluster-B. This relationship also corresponded one by one

to their respective prognostic differences (**Figure 2B**, **Figure 3I**). The heat map in **Figure 3J** revealed that the vast majority of 764 HDACsClusters phenotype-related DEGs were also highly expressed in the geneCluster-B. Moreover, HDAC8, HDAC11, SIRT1, SIRT3, and SIRT4 were obviously highly expressed in geneCluster-A than in geneCluster-C/B, while HDAC1, HDAC2, HDAC10, SIRT6, and SIRT7 had the highest expression value in the geneCluster-C (**Supplementary Figure S7B**). While our results (**Supplementary Figure S3**) indicate that HDAC8, HDAC11, SIRT1, SIRT3, and SIRT4 were good prognostic factors and HDAC1, HDAC2, SIRT6, and SIRT7 were unfavorable prognostic factors, this could also explain why geneCluster-A had the best prognosis, while geneCluster-C presented a poor prognosis (**Figure 3I**), suggesting that HDAC family genes did play an important role in LUAD prognosis and that the three subtypes identified in this study had good risk stratification performance.

Construction of the HDACsScore System

From the analysis mentioned above, we have identified 764 genes differentially expressed in three HDACsClusters. Based on these, univariate Cox analysis was performed, and 371 genes affecting prognosis were identified (**Supplementary Table S2**). The Lasso regression analysis was subsequently performed to avoid overfitting (**Figures 4A,B**). Filtered genes were included in the multivariate Cox regression analysis (stepwise regression) (**Figure 4C**), and we finally constructed a prognostic scoring system (we called HDACsScore), where a total of five genes (CRYM, GJB3, SLC2A1, STC1, and TUBB3) were included according to their risk coefficients (**Figure 4D**). Each patient was scored based on the following formula: HDACsScore = Expression*CRYM* * (−0.06596) + Expression*GJB3* * (0.122546) + Expression*SLC2A1* * (0.17376) + Expression*STC1* * (0.09823) + Expression*TUBB3* * (0.14429). Based on the median HDACsScore, we divided patients into high- and low-HDACsScore groups. Survival analysis suggested a worse prognosis in patients in the high-HDACsScore group (**Figure 4E**; **Supplementary Figure S7C**). **Figure 4F** and **Supplementary Figures S7D–E** also visually showed high HDACsScore linking a poor prognosis. Subsequently, we explored the correlation between HDACsCluster and HDACsScore, and this result showed that HDACsCluster-B had the highest HDACsScore, followed by HDACsCluster-A (**Figure 4G**). From **Figure 2B**, we identified HDACsCluster-B had the worst prognosis, followed by HDACsCluster-A. This further supported the good performance of HDACsScore in patient prognostic stratification. We also applied an alluvial diagram including HDACsCluster, geneCluster, HDACsScore, and status to visualize the attribute changes of individual patients (**Figure 4H**). Further testing was performed to determine whether HDACsScore could independently predict patient outcome, and the results also confirmed that HDACsScore could be used as an independent predictor to assess patient prognosis (**Figure 4I**). Overall, this HDACsScore system could effectively identify patients with different prognoses and was expected to be extended to clinical practice.

Genetic Features, Immune Infiltration, and Biological Processes of High- and Low-HDACsScore Groups

The abovementioned analyses suggested significant differences in prognosis between high- and low-HDACsScore groups. In order to probe deeper into the potential genomic alterations, we then compared the somatic mutation landscapes in the two groups based on TCGA data. The high-HDACsScore group presented more extensive genetic mutation than the low-HDACsScore group (**Figures 5A,B**). Of the 20 genes with the highest mutation frequency, these genes all had higher mutation frequencies in the high-HDACsScore group of tumors than the low-HDACsScore group. In the high-HDACsScore group, TP53 presented a mutation frequency of up to 54% and tops the list. In the low-HDACsScore group, TIN had the highest frequency of mutations at 35%. **Figure 5C** depicted the immune infiltration landscapes with clear differences in high- and low-HDACsScore groups. To explore the biological behaviors between these two groups, we performed GSVA enrichment analysis. As shown in **Figure 5D**, the low-HDACsScore group was markedly enriched in body material metabolism processes, including fatty acid metabolism; valine, leucine, and isoleucine degradation; limonene and pinene degradation; propanoate metabolism; and butanoate metabolism, while the high-HDACsScore group highly enriched in homologous recombination, cell cycle, DNA replication etc. To sum up, these data enabled us to depict the impact of HDACsScore classification (high and low) on genomic variation, immune landscape, and biological pathways more comprehensively, where we found significant differences between high- and low-HDACsScore groups, which might be the intrinsic mechanism for the significantly different clinical outcomes between the two groups. Also, further investigation was still necessary.

The Potential of HDACsScore in Predicting Immunotherapeutic and Chemotherapy Benefits

Accumulated evidence (Chan et al., 2019; Cristescu et al., 2018; Gong et al., 2018; Patel and Kurzrock, 2015) demonstrated patients with high-TMB status and high-immune checkpoint expression presented a durable clinical response to anti–PD-1/PD-L1 immunotherapy. In this study, the TMB quantification analyses confirmed that the high-HDACsScore group was markedly correlated with a higher TMB (**Figure 5E**), which was also able to be conjectured from **Figures 5A,B**. The Wilcoxon rank sum test revealed significant differences on expression of the immune checkpoints (PD-L1, PDCD1, and CTLA4) between low- and high-HDACsScore groups, and the high-HDACsScore group showed the higher immune checkpoint expression (**Figures 5F–H**). Based on the abovementioned findings, our preliminary inference was that the high-HDACsScore group might benefit more from immunotherapy. To further confirm this conjecture, we also compared the TIDE scores in the two groups. A lower TIDE score represented a higher response rate against both PD-1 and anti-

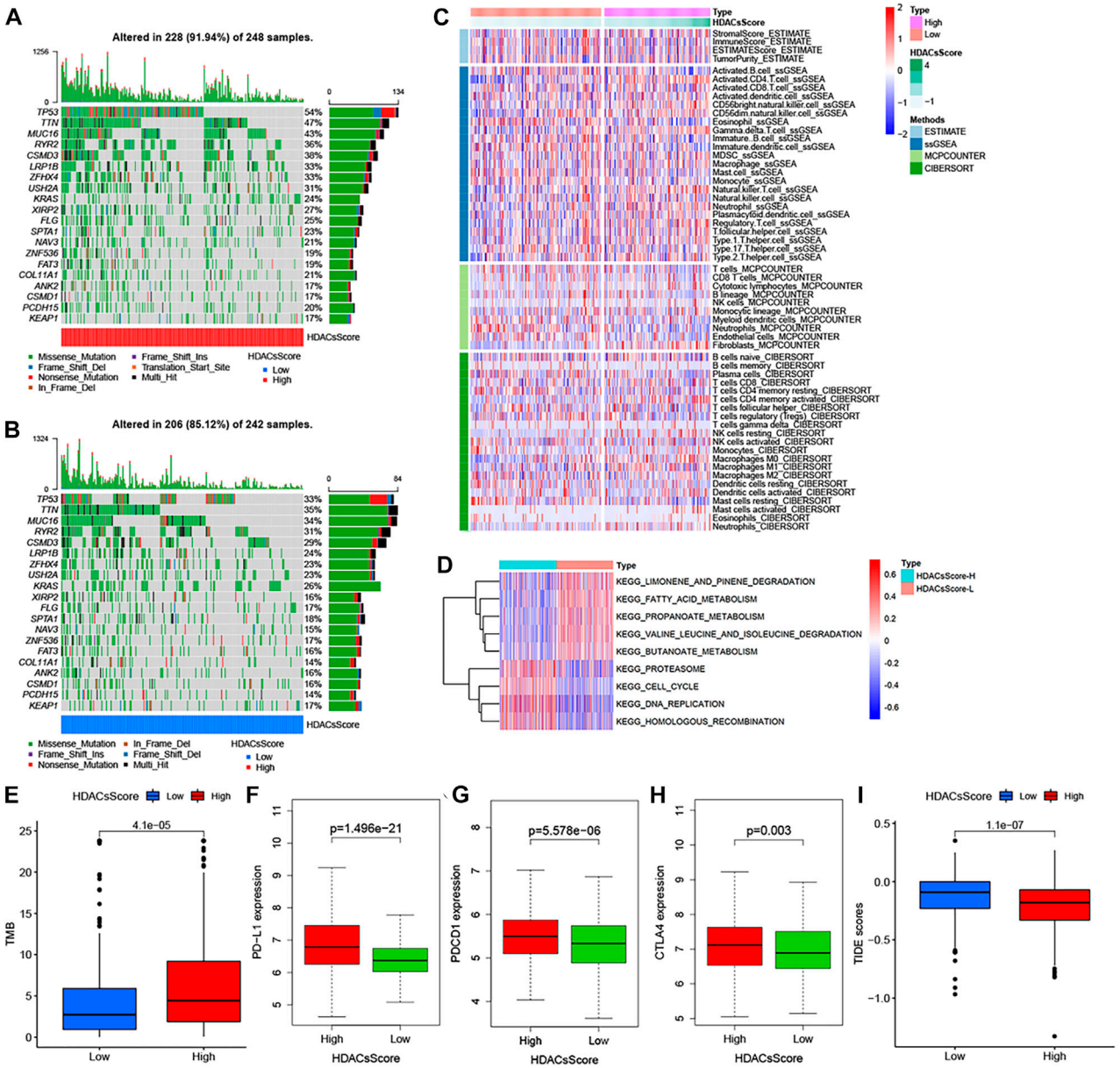


FIGURE 5 | Genetic alteration landscape, immune infiltration, and biological characteristics in different HDACsScore subgroups. **(A–B)** Waterfall plot of tumor somatic mutation established in high- **(A)** and low- **(B)** HDACsScore groups. Each column represented individual patients. The upper barplot showed TMB, and the number on the right indicated the mutation frequency in each gene. The right barplot showed the proportion of each variant type. **(C)** Comparison of tumor immune infiltrating cells based on ssGSEA, CIBERSORT, and MCP counter algorithms in the high- and low-HDACsScore groups. Red indicated the high infiltrating levels of immune cells and blue indicated the low infiltrating levels. ESTIMATE score, stromal score, immune score, tumor purity, and subtype are also shown below the annotation. **(D)** GSVA enrichment analysis showing the activation states of biological pathways in high- and low-HDACsScore groups. The Wilcoxon rank sum test revealed significant difference on TMB **(E)**, expression of PD-L1 **(F)**, PDCD1 **(G)**, CTLA4 **(H)**, and TIDE score **(I)** between high- and low-HDACsScore groups.

CTLA-4 drugs. This result revealed that the TIDE score was remarkably decreased in the high-HDACsScore group (**Figure 5I**). This further suggested a higher sensitivity to immunotherapy in the high-HDACsScore group.

Although many patients benefit from the rise of immunotherapy, there are still some patients who do not benefit from this advanced treatment. They had to return to traditional chemotherapy to prolong life. Effective identification of populations that may be sensitive to some type of chemotherapeutic agent is still being a matter of great significance. Therefore, we also used "pRRophetic" algorithms to compute the IC_{50} of some drugs in different patients to assess their sensitivity to a given drug. From the meta-GEO cohort (**Figure 6A**) and TCGA cohort data (**Figure 6B**), we found that the low-HDACsScore group had higher IC_{50} in six drugs (cyclopamine, docetaxel, doxorubicin, gemcitabine, paclitaxel, and pyrimethamine), indicating the patients in the high-HDACsScore

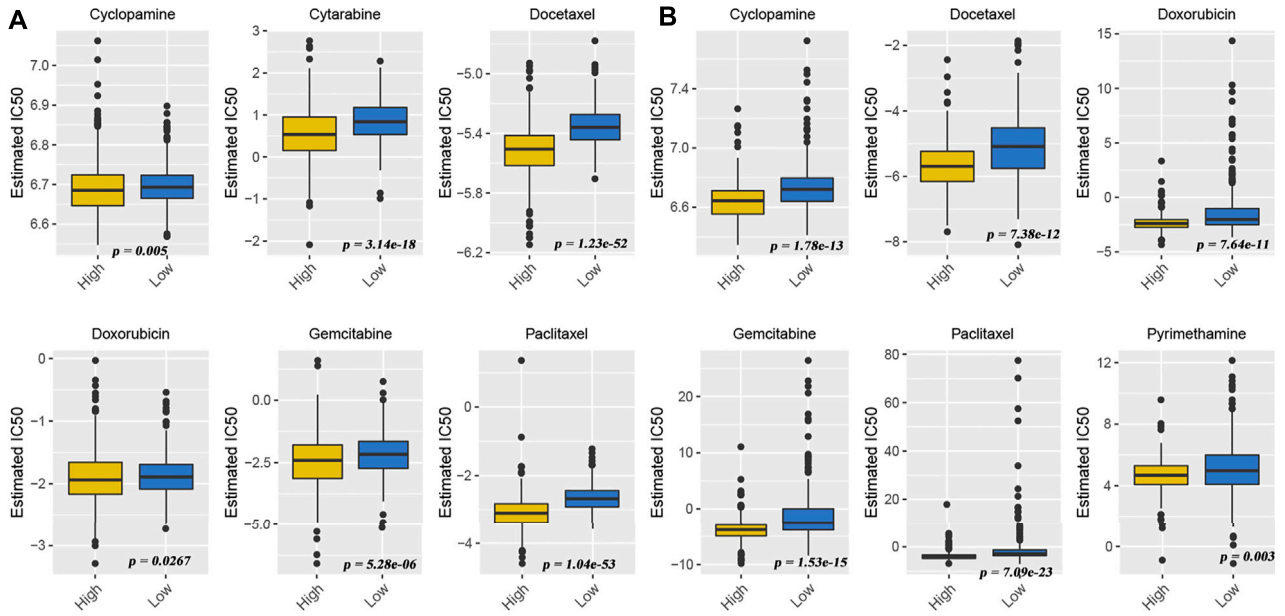


FIGURE 6 | Drug sensitivity comparison between high- and low-HDACsScore groups. Distribution of the estimated IC_{50} of cyclopamine, docetaxel, doxorubicin, gemcitabine, paclitaxel, and pyrimethamine between high- and low-HDACsScore groups in the meta-GEO cohort **(A)** and TCGA cohort **(B)**.

group were more sensitive to these drugs. The abovementioned results initially illustrated the importance of HDACsScore in predicting the efficacy of immunotherapy and chemotherapy in LUAD. These results could provide more clues in determining the personalized treatment strategies for LUAD patients.

DISCUSSION

To our knowledge, this study is the only comprehensive analysis of HDAC family genes in LUAD. Herein, a systematic bioinformatics approach was used to comprehensively characterize the biological function, dysregulated expression, genomic mutations, immune infiltration, and prognostic relationships of HDAC genes in LUAD. Based on unsupervised clustering of HDACs, this study identified three different molecular subtypes mediated by HDACs, which presented different immune infiltration, prognoses, and transcriptome expression. The HDAC scoring system we constructed can effectively identify patients with different prognoses and make preliminary judgments on their response to chemotherapy and immunotherapy. High- and low-score groups represented different genomic landscapes, immune features, and signaling pathway activation states. This is of great significance for further exploring the prognostic heterogeneity as well as the potential targets of action in LUAD patients.

In this study, we observed mutations in the HDAC genes occurred in 15.15% of the LUAD samples, in which HDAC9 showed the highest mutation frequency (6%). Further investigation showed that this low-frequency mutation would promote HDAC2 expression and did not appear to affect the expression of other HDAC genes. HDAC9 was not differentially expressed in LUAD tumor tissues versus normal tissues; however, Kaplan–Meier analysis suggested that its overexpression was associated with better prognosis. A similar phenomenon was also observed in clear cell renal cell carcinoma (Fu et al., 2020) and retinoblastoma (Zhang et al., 2016). In other tumors, high expression of HDAC9 was usually associated with poor prognosis in patients, such as pancreatic cancer (Li et al., 2020), breast cancer (Huang et al., 2018), and oral squamous cancer (Rastogi et al., 2016). The abovementioned results indicated that HDAC9 had different prognostic implications in different tumors, and its influence on the tumor biological behavior may be influenced by the cancer context. Additionally, the correlation between the expression of HDAC family genes and immune cell infiltration remains poorly explored, and our study showed that most HDAC genes were positively correlated with each other, but negatively correlated with most immune cell infiltration. Among them, HDAC9 was negatively associated with all the other HDAC genes (such as HDAC10, HDAC11, SIRT2, SIRT3, SIRT4, SIRT5, and SIRT6), but positively correlated with most immune cell infiltration (activated B cell, activated CD4 T cell, activated CD8 T cell, activated dendritic cell, and NK cell). In the study of Ning et al. (2020), HDAC9 expression could modulate the tumor microenvironment, which in turn affected tumor biology behaviors. These results suggested that HDAC9 was closely related to tumor immunity, which could provide new horizons for in-depth relevant studies.

Significant heterogeneity brought great challenges to the prognostic judgment and treatment decisions in LUAD patients. Further identification of tumor subtypes with different characteristics facilitated the clinical adoption of precise strategies. The abovementioned analysis suggested the important roles of HDAC family genes in LUAD; hence, we believed that they probably played critical roles in the formation of different tumor subtypes and were implicated in cancer pathogenesis and progression. The three molecular subtypes identified in this study presented significantly different prognoses, and the

expression of these 18 HDAC genes, biological pathway activation, and TME landscape across them also varied significantly. This may explain the internal mechanism of their different prognoses in the three molecular subtypes. We observed that compared with HDACsCluster-A and -B, HDACsCluster-C had the best survival but lowest immune checkpoint expression. This might somehow imply a poor response of HDACsCluster-C subtype to immunotherapy. To further explore the underlying molecular features and expression perturbations within these HDAC-mediated phenotypes, we performed a second unsupervised clustering and further determined that HDAC family genes indeed play an important role in LUAD prognosis, and the three subtypes identified in this study had good risk stratification performance. The construction of the HDAC prognostic scoring system aimed to eliminate prediction misjudgments caused by individual heterogeneity. Similar to previous prognostic systems (Li et al., 2021; Song et al., 2021; Zhang et al., 2020,2021), this HDAC prognostic scoring system accurately calculated the HDACsScore for each patient and then accurately predicts patient survival. We found that high HDACsScore was associated with poor prognosis, and HDACsScore could be used as an independent predictor to assess patient prognosis. In addition, we also observed that the high- and low-score groups were significantly different in somatic mutation characteristics, immune infiltration landscape, and biological pathway activation. Of the 20 genes with the highest mutation frequency, these genes changed more frequently in the high-HDACsScore group. In the high-HDACsScore group, TP53 mutation stated as high as 54%, topping the list. In the low-HDACsScore group, the TIN had the highest frequency of mutations at 35%. The material metabolic process was more active in the low-HDACsScore group, while the high-HDACsScore group was highly enriched in homologous recombination, cell cycle, and DNA replication. This seemed to explain why the high HDACsScore group presented a broader range of gene mutations.

Additionally, this study also investigated the potential of HDACsScore in predicting immunotherapeutic and chemotherapy benefits. It was shown that TMB status as well as immune checkpoint expression could suggest the tumor response to immunotherapy (Chan et al., 2019; Cristescu et al., 2018; Gong et al., 2018; Patel and Kurzrock., 2015). In our study, TMB quantification analyses confirmed that the high-HDACsScore group was markedly correlated with higher TMB. Also, the high-HDACsScore group also showed higher immune checkpoint expression. Furthermore, TIDE analysis also yielded consistent results. The abovementioned findings suggested a higher sensitivity to immunotherapy in the high-HDACsScore group. In addition, the IC_{50} of a given drug could reflect its sensitivity to the drug. This study found that the low-HDACsScore group had higher IC_{50} in cyclopamine, docetaxel, doxorubicin, gemcitabine, paclitaxel, and pyrimethamine, indicating patients in the high-HDACsScore group were more sensitive to these drugs. Overall, the abovementioned results initially illustrated the importance of HDACsScore in predicting the efficacy of immunotherapy and chemotherapy in LUAD. These results could provide more clues in determining the personalized treatment strategies for LUAD patients.

This study, based on a large LUAD cohort including seven independent datasets, revealed the potential significance of HDAC family genes in LUAD. However, there were still some limitations in this study. Although we used the "ComBat" algorithm of the "sva" package in R to further integrate into a metacohort, batch effects resulting from nonbiotechnological bias could not be completely eliminated. It was still difficult to avoid the bias imposed by the nature of the retrospective study. In addition, this study lacked extensive experimental validation with results derived from public database. Overall, the present study comprehensively evaluated the potential significance of HDAC family genes in over 1400 LUAD samples and described the multidimensional characterization of HDAC family genes in LUAD. More broadly, based on the HDAC genes, this study identified three LUAD subtypes with different genomic, transcriptome, immune infiltration, and metabolic pathway characteristics and constructed an HDAC scoring system for risk stratification and efficacy prediction, which would help enhance our perception of LUAD prognostic differences and provide important insights into the efficacy of immunotherapy and chemotherapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CS, WL, NL, and QG designed the study. WL and CS analyzed the data. QG prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81770095, 81700093, and 8210082163), the Fundamental Research Funds for the Central Universities (No. 2042021kf0081), and Science Fund for Creative Research Groups of the Natural Science Foundation of Hubei Province (No. 2020CFA027).

ACKNOWLEDGMENTS

This work benefited from open databases. We are grateful for the efforts made by the resources and staff to expand and improve the databases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.862977/full#supplementary-material

REFERENCES

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., et al. (2009). Systematic RNA Interference Reveals that Oncogenic KRAS-Driven Cancers Require TBK1. *Nature* 462 (7269), 108–112. doi:10.1038/nature08460

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17 (1), 218. doi:10.1186/s13059-016-1070-5

Bolden, J. E., Peart, M. J., and Johnstone, R. W. (2006). Anticancer Activities of Histone Deacetylase Inhibitors. *Nat. Rev. Drug Discov.* 5 (9), 769–784. doi:10.1038/nrd2133

Botling, J., Edlund, K., Lohr, M., Hellwig, B., Holmberg, L., Lambe, M., et al. (2013). Biomarker Discovery in Non-small Cell Lung Cancer: Integrating Gene Expression Profiling, Meta-Analysis, and Tissue Microarray Validation. *Clin. Cancer Res.* 19 (1), 194–204. doi:10.1158/1078-0432.CCR-12-1139

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of Tumor Mutation Burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., et al. (2018). Pan-tumor Genomic Biomarkers for PD-1 Checkpoint Blockade-Based Immunotherapy. *Science* 362 (6411), eaar3593. doi:10.1126/science.aar3593

Der, S. D., Sykes, J., Pintilie, M., Zhu, C.-Q., Strumpf, D., Liu, N., et al. (2014). Validation of a Histology-independent Prognostic Gene Signature for Early-Stage, Non-small-cell Lung Cancer Including Stage IA Patients. *J. Thorac. Oncol.* 9 (1), 59–64. doi:10.1097/JTO.0000000000000042

Dokmanovic, M., and Marks, P. A. (2005). Prospects: Histone Deacetylase Inhibitors. *J. Cell. Biochem.* 96 (2), 293–304. doi:10.1002/jcb.20532

Fang, Z., Lin, M., Li, C., Liu, H., and Gong, C. (2020). A Comprehensive Review of the Roles of E2F1 in Colon Cancer. *Am. J. Cancer Res.* 10 (3), 757–768.

Fu, Y., Piao, C., Zhang, Z., Zhu, Y., Sun, S., Bi, J., et al. (2020). Decreased Expression and Hypomethylation of HDAC9 Lead to Poor Prognosis and Inhibit Immune Cell Infiltration in Clear Cell Renal Cell Carcinoma. *Urologic Oncol. Seminars Orig. Investigations* 38 (9), e1–740. doi:10.1016/j.urolonc.2020.03.006

Geeleher, P., Cox, N., and Huang, R. S. (2014). PRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Goldmann, T., Marwitz, S., Nitschkowski, D., Krupar, R., Backman, M., Elfving, H., et al. (2021). PD-L1 Amplification Is Associated with an Immune Cell Rich Phenotype in Squamous Cell Cancer of the Lung. *Cancer Immunol. Immunother.* 70 (9), 2577–2587. doi:10.1007/s00262-020-02825-z

Gong, J., Chehrazi-Raffle, A., Reddi, S., and Salgia, R. (2018). Development of PD-1 and PD-L1 Inhibitors as a Form of Cancer Immunotherapy: A Comprehensive Review of Registration Trials and Future Considerations. *J. Immunother. cancer* 6 (1), 8. doi:10.1186/s40425-018-0316-z

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Huang, Y., Jian, W., Zhao, J., and Wang, G. (2018). Overexpression of HDAC9 Is Associated with Poor Prognosis and Tumor Progression of Breast Cancer in Chinese Females. *Ott* Vol. 11, 2177–2184. doi:10.2147/OTT.S164583

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* 4 (2), 249–264. doi:10.1093/biostatistics/4.2.249

Jabs, V., Edlund, K., König, H., Grinberg, M., Madjar, K., Rahnenführer, J., et al. (2017). Integrative Analysis of Genome-wide Gene Copy Number Changes and Gene Expression in Non-small Cell Lung Cancer. *PLoS One* 12 (11), e0187246. doi:10.1371/journal.pone.0187246

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* 8 (1), 118–127. doi:10.1093/biostatistics/kxj037

Kim, J. H., Yoon, C. Y., Jun, Y., Lee, B. B., Lee, J. E., Ha, S. D., et al. (2020). NuA3 HAT Antagonizes the Rpd3S and Rpd3L HDACs to Optimize mRNA and lncRNA Expression Dynamics. *Nucleic Acids Res.* 48 (19), 10753–10767. doi:10.1093/nar/gkaa781

Kim, S.-J., Baek, K. S., Park, H.-J., Jung, Y. H., and Lee, S.-M. (2016). Compound 9a, a Novel Synthetic Histone Deacetylase Inhibitor, Protects against Septic Injury in Mice by Suppressing MAPK Signalling. *Br. J. Pharmacol.* 173 (6), 1045–1057. doi:10.1111/bph.13414

Kim, T. H., Yoo, J.-Y., Choi, K.-C., Shin, J.-H., Leach, R. E., Fazleabas, A. T., et al. (2019). Loss of HDAC3 Results in Nonreceptive Endometrium and Female Infertility. *Sci. Transl. Med.* 11 (474), eaaf7533. doi:10.1126/scitranslmed.aaf7533

Li, H., Li, X., Lin, H., and Gong, J. (2020a). High HDAC9 Is Associated with Poor Prognosis and Promotes Malignant Progression in Pancreatic Ductal Adenocarcinoma. *Mol. Med. Rep.* 21 (2), 822–832. doi:10.3892/mmr.2019.10869

Li, Y., Gu, J., Xu, F., Zhu, Q., Chen, Y., Ge, D., et al. (2021). Molecular Characterization, Biological Function, Tumor Microenvironment Association and Clinical Significance of m6A Regulators in Lung Adenocarcinoma. *Brief. Bioinform.* 22 (4), bbaa225. doi:10.1093/bib/bbaa225

Li, Y., and Seto, E. (2016). HDACs and HDAC Inhibitors in Cancer Development and Therapy. *Cold Spring Harb. Perspect. Med.* 6 (10), a026831. doi:10.1101/cshperspect.a026831

Li, Y., Zhang, X., Zhu, S., Dejene, E. A., Peng, W., Sepulveda, A., et al. (2020b). HDAC10 Regulates Cancer Stem-like Cell Properties in KRAS-Driven Lung Adenocarcinoma. *Cancer Res.* 80 (16), 3265–3278. doi:10.1158/0008-5472.CAN-19-3613

Li, Z., and Zhu, W.-G. (2014). Targeting Histone Deacetylases for Cancer Therapy: From Molecular Mechanisms to Clinical Implications. *Int. J. Biol. Sci.* 10 (7), 757–770. doi:10.7150/ijbs.9067

Lohr, M., Hellwig, B., Edlund, K., Mattsson, J. S. M., Botling, J., Schmidt, M., et al. (2015). Identification of Sample Annotation Errors in Gene Expression Datasets. *Arch. Toxicol.* 89 (12), 2265–2272. doi:10.1007/s00204-015-1632-4

Maiti, A., Qi, Q., Peng, X., Yan, L., Takabe, K., and Hait, N. (2019). Class I Histone Deacetylase Inhibitor Suppresses Vasculogenic Mimicry by Enhancing the Expression of Tumor Suppressor and Anti-angiogenesis Genes in Aggressive Human TNBC Cells. *Int. J. Oncol.* 55 (1), 116–130. doi:10.3892/ijo.2019.4796

Mithraprabhu, S., Kalff, A., Chow, A., Khong, T., and Spencer, A. (2014). Dysregulated Class I Histone Deacetylases Are Indicators of Poor Prognosis in Multiple Myeloma. *Epigenetics* 9 (11), 1511–1520. doi:10.4161/15592294.2014.983367

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods.* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ning, Y., Ding, J., Sun, X., Xie, Y., Su, M., Ma, C., et al. (2020). HDAC9 Deficiency Promotes Tumor Progression by Decreasing the CD8+ Dendritic Cell Infiltration of the Tumor Microenvironment. *J. Immunother. Cancer* 8 (1), e000529. doi:10.1136/jitc-2020-000529

Okayama, H., Kohno, T., Ishii, Y., Shimada, Y., Shiraishi, K., Iwakawa, R., et al. (2012). Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-negative Lung Adenocarcinomas. *Cancer Res.* 72 (1), 100–111. doi:10.1158/0008-5472.CAN-11-1403

Patel, S. P., and Kurzrock, R. (2015). PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol. Cancer Ther.* 14 (4), 847–856. doi:10.1158/1535-7163.MCT-14-0983

Raparia, K., Villa, C., DeCamp, M. M., Patel, J. D., and Mehta, M. P. (2013). Molecular Profiling in Non-small Cell Lung Cancer: A Step toward Personalized Medicine. *Arch. Pathol. Lab. Med.* 137 (4), 481–491. doi:10.5858/arpa.2012-0287-RA

Rastogi, B., Raut, S. K., Panda, N. K., Rattan, V., Radotra, B. D., and Khullar, M. (2016). Overexpression of HDAC9 Promotes Oral Squamous Cell Carcinoma Growth, Regulates Cell Cycle Progression, and Inhibits Apoptosis. *Mol. Cell. Biochem.* 415 (1-2), 183–196. doi:10.1007/s11010-016-2690-5

Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., et al. (2013). Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci. Transl. Med.* 5 (186), 166r–186r. doi:10.1126/scitranslmed.3005723

Sanchez, H. N., Moroney, J. B., Gan, H., Shen, T., Im, J. L., Li, T., et al. (2020). B Cell-Intrinsic Epigenetic Modulation of Antibody Responses by Dietary Fiber-Derived Short-Chain Fatty Acids. *Nat. Commun.* 11 (1), 60. doi:10.1038/s41467-019-13603-6

Schabath, M. B., Welsh, E. A., Fulp, W. J., Chen, L., Teer, J. K., Thompson, Z. J., et al. (2016). Differential Association of STK11 and TP53 with KRAS Mutation-Associated Gene Expression, Proliferation and Immune Surveillance in Lung Adenocarcinoma. *Oncogene* 35 (24), 3209–3216. doi:10.1038/onc.2015.375

Song, C., Wu, Z., Wang, Q., Wang, Y., Guo, Z., Li, S., et al. (2021). A Combined Two-mRNA Signature Associated with PD-L1 and Tumor Mutational Burden for Prognosis of Lung Adenocarcinoma. *Front. Cell Dev. Biol.* 9, 634697. doi:10.3389/fcell.2021.634697

Torres-Adorno, A. M., Lee, J., Kogawa, T., Ordentlich, P., Tripathy, D., Lim, B., et al. (2017). Histone Deacetylase Inhibitor Enhances the Efficacy of MEK Inhibitor through NOXA-Mediated MCL1 Degradation in Triple-Negative and Inflammatory Breast Cancer. *Clin. Cancer Res.* 23 (16), 4780–4792. doi:10.1158/1078-0432.CCR-16-2622

Travis, W. D., Garg, K., Franklin, W. A., Wistuba, I. I., Sabloff, B., Noguchi, M., et al. (2005). Evolving Concepts in the Pathology and Computed Tomography Imaging of Lung Adenocarcinoma and Bronchioloalveolar Carcinoma. *Jco* 23 (14), 3279–3287. doi:10.1200/JCO.2005.15.776

Weinberger, L., Voichek, Y., Tirosh, I., Hornung, G., Amit, I., and Barkai, N. (2012). Expression Noise and Acetylation Profiles Distinguish HDAC Functions. *Mol. Cell* 47 (2), 193–202. doi:10.1016/j.molcel.2012.05.008

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Xie, Y., Xiao, G., Coombes, K. R., Behrens, C., Solis, L. M., Raso, G., et al. (2011). Robust Gene Expression Signature from Formalin-Fixed Paraffin-Embedded Samples Predicts Prognosis of Non-small-cell Lung Cancer Patients. *Clin. Cancer Res.* 17 (17), 5705–5714. doi:10.1158/1078-0432.CCR-11-0196

Xu, Y., Liu, X., Zhang, X., Zhang, G., Zhang, R., Liu, T., et al. (2012). Histone Acetylation of the Htr3a Gene in the Prefrontal Cortex of Wistar Rats Regulates Ethanol-Seeking Behavior. *Neural Regen. Res.* 7 (13), 1021–1028. doi:10.3969/j.issn.1673-5374.2012.13.009

Yamauchi, M., Yamaguchi, R., Nakata, A., Kohno, T., Nagasaki, M., Shimamura, T., et al. (2012). Epidermal Growth Factor Receptor Tyrosine Kinase Defines Critical Prognostic Genes of Stage I Lung Adenocarcinoma. *PLoS One* 7 (9), e43923. doi:10.1371/journal.pone.0043923

Yang, H., Salz, T., Zajac-Kaye, M., Liao, D., Huang, S., and Qiu, Y. (2014). Overexpression of Histone Deacetylases in Cancer Cells Is Controlled by Interplay of Transcription Factors and Epigenetic Modulators. *FASEB J.* 28 (10), 4265–4279. doi:10.1096/fj.14-250654

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m6A Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Gastric Cancer. *Mol. Cancer.* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, N., Zhang, H., Wang, Z., Dai, Z., Zhang, X., Cheng, Q., et al. (2021). Immune Infiltrating Cells-Derived Risk Signature Based on Large-Scale Analysis Defines Immune Landscape and Predicts Immunotherapy Responses in Glioma Tumor Microenvironment. *Front. Immunol.* 12, 691811. doi:10.3389/fimmu.2021.691811

Zhang, Y., Wu, D., Xia, F., Xian, H., Zhu, X., Cui, H., et al. (2016). Downregulation of HDAC9 Inhibits Cell Proliferation and Tumor Formation by Inducing Cell Cycle Arrest in Retinoblastoma. *Biochem. Biophysical Res. Commun.* 473 (2), 600–606. doi:10.1016/j.bbrc.2016.03.129

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Song, Lin, Meng, Li and Geng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/fgene.2022.897083

Check for updates

OPEN ACCESS

EDITED BY
Anton A. Buzdin,
European Organisation for Research and Treatment of Cancer, Belgium

REVIEWED BY
Hanyu Shen,
Wuxi Huishan District People's Hospital, China
Wubin Ding,
Children's Hospital of Philadelphia, United States

*CORRESPONDENCE
Bin Lu,
lubin03162@126.com
Zijian Zhou,
zhouzijiandoctor@163.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 15 March 2022
ACCEPTED 26 July 2022
PUBLISHED 24 August 2022

CITATION
Zhou Z, Wei J, Kuang L, Zhang K, Liu Y, He Z, Li L and Lu B (2022), Characterization of aging cancer-associated fibroblasts draws implications in prognosis and immunotherapy response in low-grade gliomas.
Front. Genet. 13:897083.
doi: 10.3389/fgene.2022.897083

COPYRIGHT
© 2022 Zhou, Wei, Kuang, Zhang, Liu, He, Li and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Characterization of aging cancer-associated fibroblasts draws implications in prognosis and immunotherapy response in low-grade gliomas

Zijian Zhou[1]*, Jinhong Wei[2], Lijun Kuang[2], Ke Zhang[3], Yini Liu[2], Zhongming He[2], Luo Li[1] and Bin Lu[1]*

[1]Department of Neurosurgery, Qingdao Municipal Hospital, Qingdao University, Qingdao, China, [2]School of Basic Medical Sciences, Southwest Medical University, Luzhou, China, [3]Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Sichuan, China

Background: Due to the highly variable prognosis of low-grade gliomas (LGGs), it is important to find robust biomarkers for predicting clinical outcomes. Aging cancer-associated fibroblasts (CAFs) within the senescent stroma of a tumor microenvironment (TME) have been recently reported to play a key role in tumor development. However, there are few studies focusing on this topic in gliomas.

Methods and Results: Based on the transcriptome data from TCGA and CGGA databases, we identified aging CAF-related genes (ACAFRGs) in LGGs by the weighted gene co-expression network analysis (WGCNA) method, followed by which LGG samples were classified into two aging CAF-related gene clusters with distinct prognosis and characteristics of the TME. Machine learning algorithms were used to screen out eight featured ACAFRGs to characterize two aging CAF-related gene clusters, and a nomogram model was constructed to predict the probability of gene cluster A for each LGG sample. Then, a powerful aging CAF scoring system was developed to predict the prognosis and response to immune checkpoint blockage therapy. Finally, the ACAFRGs were verified in two glioma-related external datasets. The performance of the aging CAF score in predicting the immunotherapy response was further validated in two independent cohorts. We also confirmed the expression of ACAFRGs at the protein level in glioma tissues through the Human Protein Atlas website and Western blotting analysis.

Conclusion: We developed a robust aging CAF scoring system to predict the prognosis and immunotherapy response in LGGs. Our findings may provide new targets for therapeutics and contribute to the exploration focusing on aging CAFs.

KEYWORDS

low-grade glioma, aging cancer-associated fibroblasts, tumor microenvironment, prognosis, immunotherapy response

Introduction

Low-grade gliomas (LGGs) encompassing grade II and III gliomas represent a group of primary tumors originating from the central nervous system and are very common in young adults compared to high-grade gliomas (grade IV, glioblastoma multiforme, GBM) (Perry and Wesseling, 2016). In 2016, the World Health Organization (WHO) updated the classification method for gliomas by combing histological diagnosis with molecular variations such as *IDH* mutation status and codeletion of the short arm of chromosome 1 and the long arm of chromosome 19 (1p/19q codeletion) (Komori, 2017). Previous studies revealed that glioma patients with mutant *IDH* exhibited a more favorable response to current therapy including radiation and chemotherapy, implying the correlation between molecular alterations and prognosis (Cairncross et al., 2014). Due to high heterogeneity, glioma patients had various clinical outcomes even with the same diagnosis. While LGG patients tend to get a longer survival time with the median overall survival ranging from 5.6–13.3 years, the prognosis of LGG patients can be highly variable (Hottinger et al., 2016; Ostrom et al., 2018). Exploration focusing on biomarkers for predicting prognosis is becoming a hot spot in cancer research.

Most of the previous investigations have been focusing on tumor cells themselves while pioneering studies have claimed the significant importance of the crosstalk between tumor cells and the surrounding microenvironment in the course of tumor development (Radin and Tsirka, 2020). As a complex environment with dynamic alterations, the tumor microenvironment (TME) represents the non-tumoral components around tumor cells, including the extracellular matrix (ECM) and various cell populations such as immune cells, fibroblasts, and endothelial cells (Anderson and Simon, 2020). Fibroblasts, which constitute a major proportion of the TME, refer to a heterogeneous cell population derived from mesenchymal lineage cells and are collectively defined as cancer-associated fibroblasts (CAFs) (Piersma et al., 2020b). The past few years have witnessed significant strikes in the explorations of CAFs. Considering its well-established roles in epithelial–mesenchymal transition (EMT) (Erin et al., 2020) and maintenance of cancer stemness (Su et al., 2018), which is important for tumorigenesis and progression, CAFs are known to be closely related to prognosis in a variety of cancers (Miyai et al., 2020). Recently, T-cell-targeted immunotherapy is emerging as a robust treatment option for intractable cancers. As a novel type of immunotherapy, immune checkpoint blockage treatment such as CTLA4 and PD-1/PD-L1 antibodies has demonstrated pronounced success by activating T cells (Topalian et al., 2015). However, only a minority of patients get a favorable response to immunotherapy (Pitt et al., 2016). The T cell's capacity to kill tumor cells is significantly affected by the tumor stromal microenvironment, in which CAFs are understood to be a key player in immunosuppressive activity and reduce the efficacy of immune checkpoint blockage treatment (Baker et al., 2021; Miyai et al., 2022). Over the past decade, cancer has been usually recognized as a disease of aging and CAFs appear to be easily influenced by this age-related effect (Fane and Weeraratna, 2020). Despite the fact that senescence can be caused by tumor-independent manners, the senescence of CAFs is generally induced by signaling from tumor cells (Sahai et al., 2020). Senescent CAFs in the TME and their secretory profile (senescence-associated secretory phenotype, SASP) are known to influence all aspects of tumor development, including tumor initiation, progression, anti-tumor immunity, and chemoresistance (Yasuda et al., 2021b; Ruhland and Alspach, 2021). Overall, comprehensive analysis of aging CAFs is meaningful to determine biomarkers for predicting prognosis and immunotherapy response in gliomas. However, the explanation for the complicated roles of aging CAFs in the TME is hindered due to the lack of specific biomarkers to identify both the aging status and the cell type of CAFs *in vivo (Sahai et al., 2020)*. Moreover, there are no studies focusing on aging CAF-related genes or prediction models in gliomas at this time.

In this research, we identified aging CAF-related genes (ACAFRGs) by comprehensive analysis of the transcriptome data from TCGA and CGGA databases, based on which two distinct aging CAF-related gene clusters were determined. Subsequently, we constructed an aging CAF scoring system to predict the prognosis and immunotherapy response for LGG patients. Finally, we confirmed the expression of the aging CAF-related genes at the protein level. Our study may shed light on the exploration of aging CAFs and contribute to the development of aging CAF-targeted therapy for glioma patients in the future.

Materials and methods

Data acquisition

A dataset containing 508 LGG samples with the corresponding RNA sequencing (RNA-seq) data was downloaded from TCGA database (The Cancer Genome Atlas, http://cancergenome.nih.gov/). The annotation file, Genome Reference Consortium Human Build 38 (GRCh38), which was acquired from the Ensembl website (http://asia.ensembl.org/), was employed to annotate the RNA-seq data. The transcriptome data (dataset ID: mRNA-array_301) composed of 159 LGG samples were obtained from the CGGA database (Chinese Glioma Genome Atlas, http://cgga.org.cn/index.jsp) (Fang et al., 2017; Wang et al., 2017). The corresponding clinical information for LGG patients involved in the two datasets was also downloaded from the aforementioned websites. R software (version 4.1.1) was utilized for the bioinformatic analysis and visualization of the data.

Determination of ACAFRGs

In our study, the transcriptome data from TCGA database were first transformed to transcripts per million (TPM) values for further combination with the transcriptome data from the CGGA database. Robust multi-array average normalization would be performed for the transcriptome data by the normalizeBetweenArrays function in the limma R package (Smyth et al., 2005) when the distribution of gene expression values in the transcriptome data from different databases was not uniform, followed which quantile normalization and log2 transformation were carried out. Combat function in the sva R package (Leek et al., 2012) was used to remove the batch effect caused by non-biotechnological bias when merging the transcriptome data. In addition, two-dimensional principal component analysis (PCA) cluster plots were utilized to show the sample distribution before and after batch effect correction.

The stromal score which indicated the stromal components of the TME for each LGG sample was calculated by the ESTIMATE algorithm (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) through the estimate R package (Yoshihara et al., 2013) based on the gene expression values in the merged transcriptome data. Furthermore, LGG samples in the merged data were separated into high- and low-stromal score groups according to the optimal cut-off value which was determined through survminer and survival R packages. The differentially expressed genes (DEGs) between the high- and low-stromal score groups were screened out by $|\log_2 \text{FC (fold change)}| > 0.5$ and adjusted p-values (FDR, false discovery rate) < 0.05 through the limma R package. The robust DEGs were considered as stromal cell-related genes. The relative abundance of fibroblasts in the TME was quantified *via* the MCP counter (Becht et al., 2016). Based on the expression profiles of stromal cell-related genes, weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath, 2005) was utilized to determine the ACAFRGs by using the WGCNA R package. To construct the network, we first calculated the robust correlations between all the stromal cell-related genes across all LGG samples in the data. The optimal power parameter was set to amplify the strong connections between genes in the same gene modules and to penalize the weak connections between genes in different modules. In this study, the optimal power value was determined when the scale independence R^2 was higher than 0.90 and the mean connectivity degree of the co-expression network was relatively higher. A total of four phenotypes, namely, survival time, age, fibroblasts, and stromal score were involved in WGCNA. The Spearman method was used to analyze the correlation between MEs and phenotypes.

Aging CAF-related gene clusters

First, ACAFRGs with prognostic values were screened out *via* univariate Cox regression analysis by using the survival R package, in which $p < 0.05$ was considered statistically significant. Then, distinct aging CAF-related gene clusters were determined by a consensus clustering method using the ConsensusClusterPlus R package based on the expression profiles of prognostic ACAFRGs in the merged data (Wilkerson and Hayes, 2010). Our clustering analysis was based on the Partitioning Around Medoid (PAM) algorithm which was derived from the k-means machine learning algorithm. A total of 50 repetitions were conducted in the consensus clustering process for the stability of our classification and 80% of the LGG samples were involved in each iteration. The optimal number for the subgroup assignment was determined based on the consensus matrix heatmap and the relative change values of the area under the cumulative distribution function (CDF) curves.

Identification of the featured ACAFRGs for discriminating aging CAF-related gene clusters

Based on the expression profiles of ACAFRGs, two machine learning algorithms were adopted to select the key genes for discriminating aging CAF-related gene clusters, namely, least absolute shrinkage and selection operator (LASSO) logistic regression (Tibshirani, 1996) and support vector machine-recursive feature elimination (SVM-RFE) (Suykens and Vandewalle, 1999). The LASSO algorithm serves as a special instance of the penalized least squares regression with the L1-penalty function. LASSO logistic regression was carried out by using the glmnet R package, in which the optimal number of featured genes was determined when the lambda value was minimal. The SVM-RFE machine learning algorithm was performed with five-fold cross-validation by using the e1071 R package, in which the optimal number of featured genes was determined when the root mean square error (RMSE, cross-validation) was minimal. Afterward, the overlapping featured genes were selected for further analysis. Furthermore, the random forest (RF) machine learning algorithm was used to further screen out featured genes *via* the randomForest R package (Breiman, 2001), in which ntrees and mtry were set at 500 and 3, respectively. First, the optimal number of the random forest trees was determined when the cross-validation error presented minimal. Then, the random forest with the optimal number of trees was constructed. To obtain featured genes with high importance, the importance of each gene was calculated. The genes with importance >10 were selected as the featured genes for aging CAF-related gene clusters.

Based on the expression profiles of the featured ACAFRGs, we constructed a nomogram model to predict aging CAF-related gene cluster A. Calibration curves, decision curve analysis (DCA), and clinical impact curve were used to evaluate the performance of the model to predict aging CAF-related gene cluster A.

Aging CAF scoring system

Univariate cox regression analysis was implemented to determine whether the ACAFRGs were positively or negatively associated with prognosis, according to which the ACAFRGs were divided into favorable and unfavorable genes. Gene set variation analysis (GSVA) can quantify the enrichment with respect to specific functions or characteristics for individuals based on specific gene sets and transcriptome data (Hänzelmann et al., 2013). GSVA and single-sample gene set enrichment analysis (ssGSEA) were used to produce GSVA scores regarding the unfavorable and favorable gene sets for LGG samples by using the GSVA R package (Hänzelmann et al., 2013). First, the gene expression values in the transcriptome data were sequenced to obtain their rank. Then, the genes in the unfavorable and favorable gene sets were extracted, followed by which the expression levels of unfavorable and favorable genes were summed. Finally, we obtained the enrichment scores of unfavorable and favorable gene sets for each LGG sample. The aging CAF score for each LGG sample was calculated by the following formula: $agingCAFscore$ = GSVAscoreA − GSVAscoreB, in which GSVAscoreA represents the enrichment regarding unfavorable ACAFRGs and GSVAscoreB represents the enrichment regarding favorable ACAFRGs. Subsequently, LGG samples were classified into high- and low-aging CAF score groups according to the optimal cut-off value of aging CAF scores which was determined by survminer and survival R packages. Moreover, the aging CAF score and other common clinicopathological characteristics were taken into consideration for the construction of a nomogram model to predict the prognosis of LGG patients by using the rms R package.

Differential function enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and molecular functions (GO, Gene Ontology) between subgroups were analyzed by using GSVA, GSEABase, and limma R packages, in which the reference gene sets including "c5.go.mf.v7.4.symbols.gmt" and "c2.cp.kegg.v7.4.symbols.gmt" were downloaded from the GSEA database. First, the GSVA score for each function term was calculated for each sample based on the reference gene sets and the gene expression profiles. The function terms with $|\log_2 FC| > 0.1$ and adjusted p-values (FDR) < 0.05 between the two groups were then screened out and were considered differentially enriched. In the heatmaps, the values of the GSVA score were centered and scaled in the row direction, in which the rows are scaled to have mean zero and standard deviation one (z score). The top 20 differentially enriched function terms were shown in the heatmaps.

Exploration of the TME

The immune and stromal components of the TME were quantified *via* the ESTIMATE algorithm (Yoshihara et al., 2013). The relative abundance of essential immune and stromal cells in the TME was quantified *via* the MCP counter (Becht et al., 2016). CIBERSORT, a deconvolution algorithm based on linear support vector regression, was employed to further calculate the abundance of infiltrating immune cells in the TME based on the gene expression profiles of LGG samples (Newman et al., 2015). The SsGSEA method was also employed to quantify the immune cells based on the input immune cell-related gene set. In addition, we downloaded the results of the abundance of critical cells in the TME for all LGG samples from TCGA on the TIMER2.0 website (http://timer.cistrome.org/), including XCELL, TIMER, QUANTISEQ, MCP counter, EPIC, CIBERSORT, and CIBERSORT ABS. Tumor Immune Dysfunction and Exclusion (TIDE)-related scores for LGG samples were calculated to explore the immunotherapeutic response (http://tide.dfci.harvard.edu/).

Genetic mutation analysis

The genetic mutation data for LGG samples were retrieved from TCGA database. The maftools R package was used for the analysis of somatic variants. The cumulative nonsynonymous mutations per million bases in coding regions were defined as tumor mutation burdens (TMBs). LGG samples with exome nonsynonymous mutations were taken into account.

Validation in external datasets

Two independent datasets from the CGGA database (dataset ID: mRNAseq_325 and mRNAseq_693) were chosen as validation cohorts (Bao et al., 2014; Wang et al., 2015; Zhao et al., 2017; Liu et al., 2018; Zhao et al., 2021) to verify the aging CAF-related gene cluster and aging CAF score. In addition, the aging CAF score was further validated in two external datasets (GSE78220 and IMvigor210 cohort) to explore the performance in predicting the immunotherapeutic response (Hugo et al., 2016; Balar et al., 2017; Mariathasan et al., 2018).

Validation of the ACAFRGs at the protein level

A total of four ACAFRGs, namely, RBP1, PDPN, FKBP9, and MSN were randomly selected from the featured ACAFRGs. The differential expression patterns of the aforementioned genes

between normal and glioma tissues were explored on the Human Protein Atlas website (https://www.proteinatlas.org/).

Western blotting was implemented to further verify the differential expression levels of the featured ACAFRGs between normal and glioma tissues. Normal brain tissues were acquired from patients with epilepsy who received temporal lobe resection. Glioma tissues which were histologically diagnosed as grades II (G2) and III (G3) were obtained from patients who received tumor resection. In this study, three normal samples, six G2 glioma samples, and nine G3 glioma samples were involved. One normal sample, two G2 glioma samples, and three G3 glioma samples were taken together for Western blotting analysis every time. The optical density of the bands in Western blotting was analyzed by using ImageJ software (Software Version: 1.53q, Wayne Rasband and contributors, National Institutes of Health, United States). The differential analysis between different samples was conducted by the limma R package.

The collected tissues were separately homogenized and lysed in RIPA lysis buffer containing protease and phosphatase inhibitors at 0–4°C. The homogenized protein samples were centrifuged at 1,000 g for 15 min at 4°C to obtain the protein in the cytoplasm. A Bio-Rad protein assay kit was used to correct the protein content to an equal level. The protein samples were homogenized with the prepared loading buffer and then boiled for 5 min at 100°C. The same amounts of protein samples were added to SDS-PAGE and electrophoresed at 80 V for 1 h. Afterward, the protein was transferred to polyvinylidene difluoride (PVDF) membranes at 50 V for 1 h. The primary antibodies used in this study were as follows: RBP1, Podoplanin (PDPN), FKBP9, TIMP1, CHI3L1, moesin (MSN), and β-actin. After incubation with the primary antibodies for 12 h, the membranes were then incubated with the secondary anti-rabbit or anti-mouse horseradish peroxidase (HRP) antibodies. Finally, the membranes were visualized by enhanced chemiluminescence (ECL) solution.

Statistical analysis

The prognosis between different subgroups was compared using Kaplan–Meier survival analysis by survminer and survival R packages, in which the log-rank test was utilized for statistical analysis. Comparisons between the two groups were carried out by using Wilcoxon rank-sum tests. Comparisons of categorical variables between two groups were presented by chi-square tests. Comparisons of continuous variables between two groups were conducted *via* an independent Student's t-test. Two-tailed $p < 0.05$ was considered statistically significant.

Results

Determination of ACAFRGs

The schematic diagram of the workflow of this study is shown in Supplementary Figure S1. The corresponding clinicopathological information for LGG samples in the merged data is demonstrated in Supplementary Tables S1, S2. The inter-batch difference was corrected when merging the transcriptome data from TCGA and CGGA databases. As illustrated in the two-dimensional PCA cluster diagram (Supplementary Figures S2A,B), the inter-batch difference was removed. The Kaplan–Meier survival analysis indicated that the high-stromal score group tended to get a worse prognosis compared to the low score group (Figure 1A). As shown in the volcano plot (Figure 1B), the robust DEGs between the high and low-stromal score groups were screened out for further analysis. Given that CAFs represent a major component of the TME and accumulate in the tumor stroma across multiple cancers (Kalluri, 2016; Kobayashi et al., 2019; Piersma et al., 2020a; Miyai et al., 2020), the stroma-related DEGs were selected for the identification of CAF-related genes. Based on the expression profiles of stroma-related genes, a co-expression network was constructed by the WGCNA method. As shown in Figure 1C, the scale independence R^2 increased and the mean connection decreased when the soft threshold (power) value increased. The optimal power parameter was set at 10 to amplify the strong connections between genes in the same gene modules and to penalize the weak connections between different modules (scale independence $R^2 = 0.89$, mean connectivity = 7.02). The genes within the same module presented significant within-module connectivity and were defined as the hub genes for the specific modules (Figure 1D, Supplementary Figure S2C). The correlation between module eigengenes (MEs) and phenotypes including survival time, age, fibroblasts, and stromal score were subsequently analyzed (Figure 1E). A total of five gene modules were identified, in which the green module was positively associated with age ($r = 0.27$, p = 1e-12) and fibroblasts ($r = 0.81$, p = 3e-158) and the gray module was negatively associated with age ($r = -0.13$, p = 6e-04) and fibroblasts ($r = -0.41$, p = 2e-28). A total of 463 genes in the two modules were determined as ACAFRGs (Supplementary Table S3).

Aging CAF-related gene clusters

A total of 400 ACAFRGs with prognostic values were screened out *via* univariate Cox regression analysis. LGG samples in the merge data were classified into two gene clusters based on the expression profiles of prognostic ACAFRGs (Supplementary Figure S3A). As shown in the heatmap of the consensus matrix, samples were reasonably classified into two gene clusters ($k = 2$), in which samples with high consensus scores between them were more likely to be grouped into the same cluster. Moreover, no apparent increase was found in the area under the CDF curve when $k = 2$

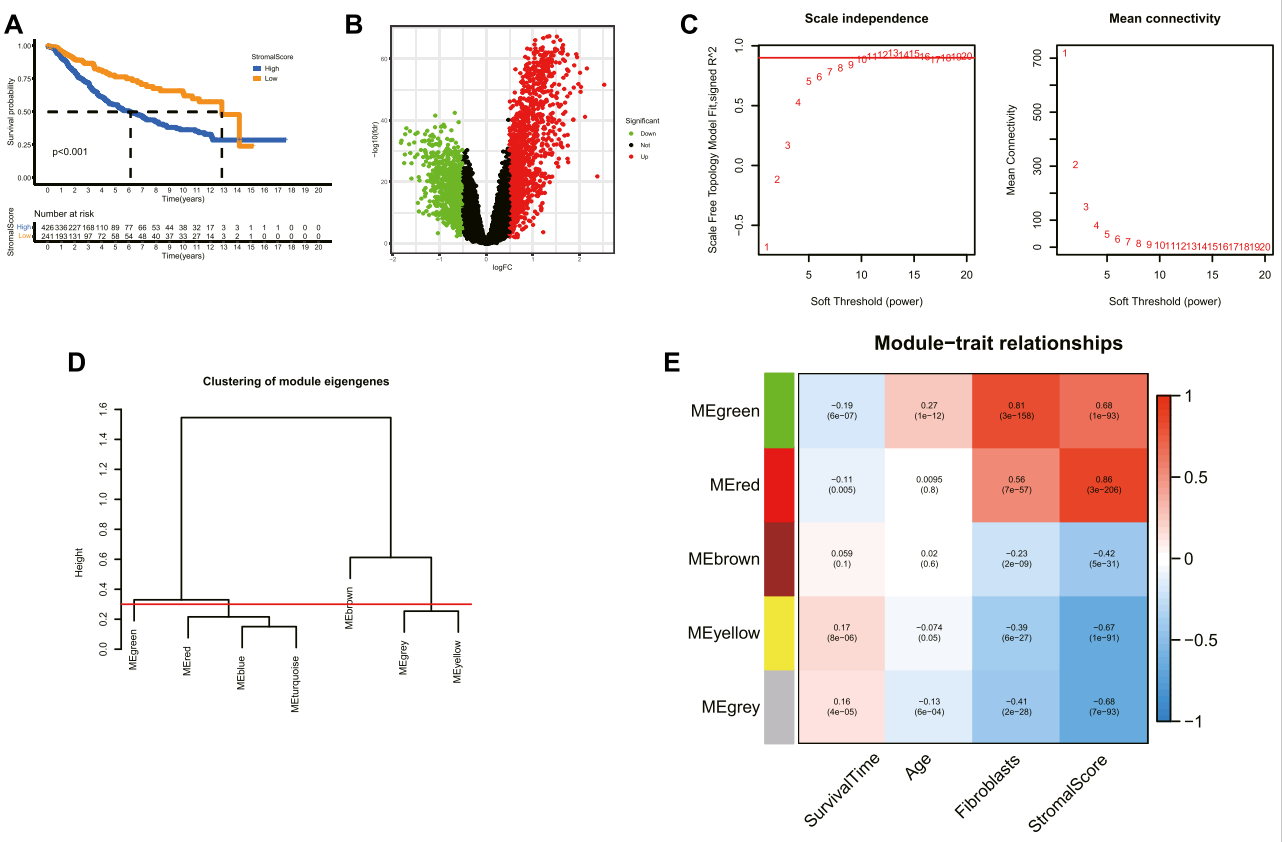


FIGURE 1
Determination of ACAFRGs. **(A)** Kaplan–Meier survival analysis for LGG patients assigned to high and low stromal scores ($p < 0.001$). **(B)** Volcano plot of the DEGs between high- and low-stromal score groups. Genes with $|\log_2 FC \text{ (fold change)}| > 0.5$ and adjusted p-values (FDR, false discovery rate) < 0.05 were considered significant. Green dots represent downregulated genes in the high-stromal score group and red dots represent upregulated genes in the high-stromal score group. **(C)** Scale independence index and mean connectivity values when soft threshold (power) ranges from 1–20. The red line was set at 0.90. **(D)** Clustering of the module eigengenes. The cut height was set at 0.30 as depicted with the red line. **(E)** Heatmap demonstrating the key gene modules associated with survival time, age, fibroblasts, and stromal scores. Pearson correlation coefficients and p-values were shown in cells. ACAFRGs, aging cancer-associated fibroblast related genes; PCA, principal component analysis; LGG, low-grade glioma; DEGs, differentially expressed genes.

(relative change <0.4). PCA confirmed the results of the subgroup assignment (Figure 2A). Kaplan–Meier analysis revealed that gene cluster A had shorter overall survival and shorter progression-free survival than gene cluster B, indicating that LGG samples in gene cluster A tended to get a worse prognosis (Figures 2B,C). The prognostic ACAFRGs were separated into gene types A and B, in which ACAFRGs in gene type A were downregulated in gene cluster A and upregulated in gene cluster B while ACAFRGs in gene type B were upregulated in gene cluster A and downregulated in gene cluster B (Figure 2D). In addition, the stromal and immune scores of gene cluster A were significantly higher than those of gene cluster B (Supplementary Figures S3B,C). The abundance of most types of the cells in the TME was higher in gene cluster A, including T cells, CD8 T cells, epithelial cells, fibroblasts, and macrophages (Figure 2E, Supplementary Figures S3B,C).

Identification of the featured ACAFRGs for discriminating aging CAF-related gene clusters

First, the LASSO logistic regression machine learning method was utilized to identify the featured ACAFRGs for discriminating two aging CAF-related gene clusters, in which 53 featured ACAFRGs were determined when the lambda value was minimal (Figure 3A). Subsequently, the SVM-RFE machine learning algorithm was performed to further determine the featured ACAFRGs, in which 31 featured ACAFRGs were identified when RMSE was minimal (Figure 3B). We obtained 15 overlapped genes *via* the two aforementioned methods (Figure 3C). Moreover, the random forest model was used to further screen out featured ACAFRGs based on the expression profiles of the aforementioned 15 featured ACAFRGs, in which 50 trees were determined when the cross-validation error

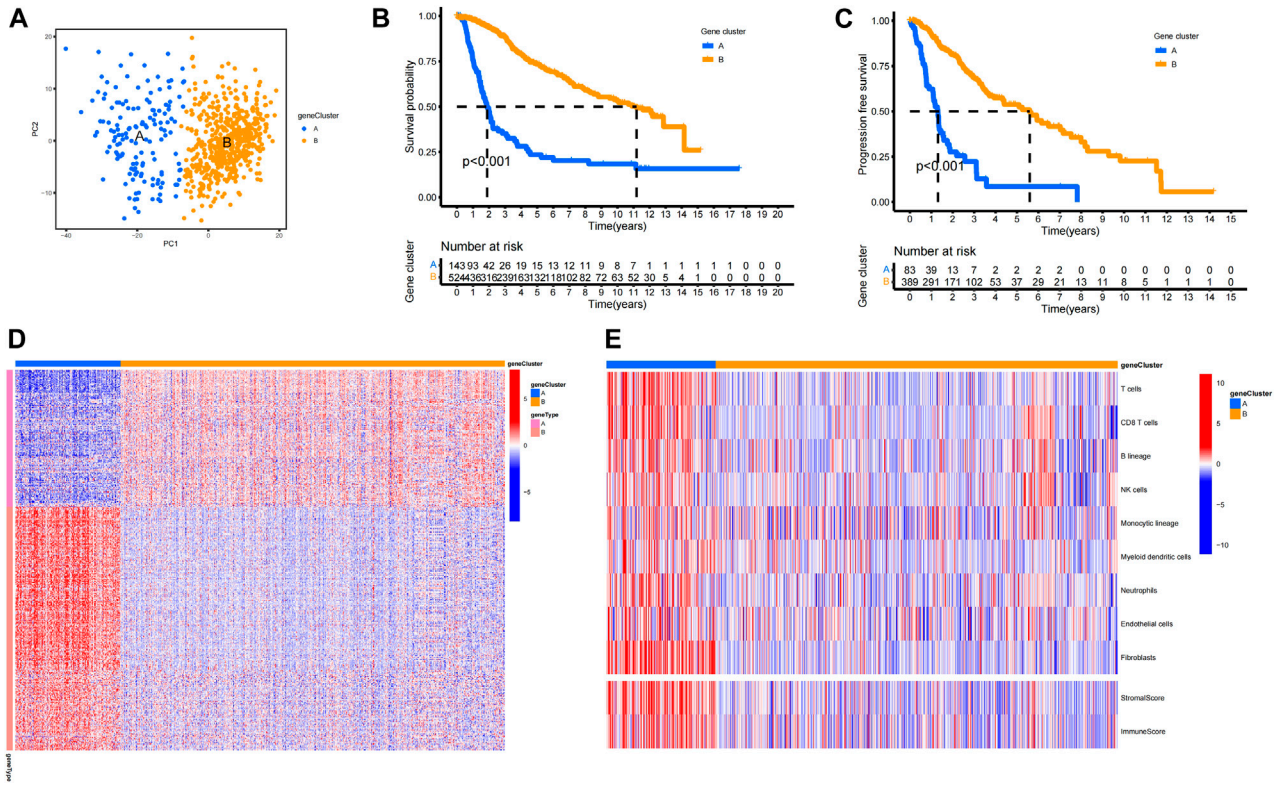


FIGURE 2
Aging CAF-related gene clusters. **(A)** PCA of LGG samples based on the expression profiles of ACAFRGs. Blue dots represent samples of gene cluster A and yellow dots represent samples of gene cluster B. **(B)** Kaplan–Meier analysis indicated that gene cluster A had a shorter overall survival than gene cluster B ($p < 0.001$). **(C)** Kaplan–Meier analysis indicated that gene cluster A had shorter progression-free survival than gene cluster B ($p < 0.001$). **(D)** Heatmap displaying the distinct expression patterns of ACAFRGs between two gene clusters, in which gene type A was upregulated in gene cluster B while gene type B was upregulated in gene cluster A. **(E)** Heatmap displaying the distinct characteristics of the TME between two gene clusters. ACAFRGs, aging cancer-associated fibroblast related genes; LGG, low-grade glioma; PCA, principal component analysis.

presented minimal (Figure 3D). Based on the determination of the optimal number of forest trees, the importance of each gene was calculated, followed by which eight ACAFRGs with importance higher than 10 were selected as the optimal featured ACAFRGs for discriminating aging CAF-related gene clusters (Figure 3E). ROC curves revealed the efficacy of each featured gene for discriminating aging CAF-related gene clusters (Figure 3F, Supplementary Figure S4), in which all the AUC values were higher than 0.930. A nomogram model combing the eight featured ACAFRGs was constructed to predict aging CAF-related gene cluster A (Figure 3G). The calibration curves indicated a good performance of the nomogram model to predict gene cluster A (Figure 3H). The red line in the DCA remained above the gray and black lines from 0 to 1, suggesting that the decisions based on the nomogram model were accurate (Figure 3I). The clinical impact curve confirmed the robust performance of the nomogram model (Figure 3J). Unsupervised clustering for LGG samples was conducted based on the expression of the eight featured ACAFRGs. We found that samples in the same gene cluster tended to be aggregated together, indicating that LGG samples could be well distinguished through the expression of the eight featured ACAFRGs (Figure 3K). Additionally, the differential expression patterns of the eight featured genes between gliomas and normal samples were analyzed through the GEPIA online tools (Li et al., 2021) (GEPIA, Gene Expression Profiling Interactive Analysis, http://gepia.cancer-pku.cn/). As shown in Supplementary Figure S5, most of the featured ACAFRGs were upregulated in gliomas compared to normal brain samples, except for FAM110B with no significant difference.

Comparison of the prognosis between low- and high-aging CAF score groups

Based on the expression profiles of the prognostic ACAFRGs, the aging CAF score was calculated for each LGG sample in the merged data through the GSVA method. LGG samples were then

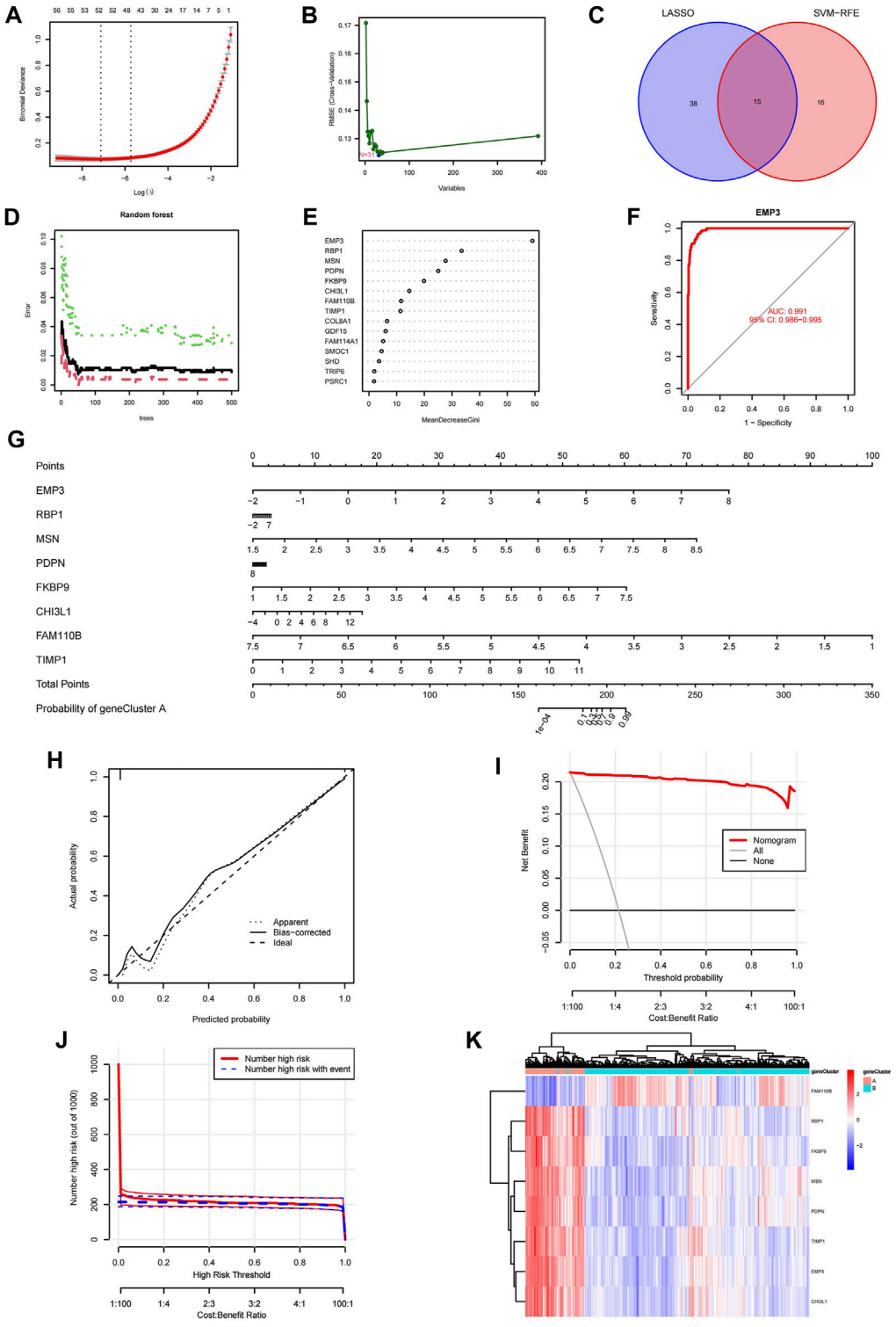


FIGURE 3
Identification of the featured ACAFRGs for discriminating aging CAF-related gene clusters. **(A)** Determination of the optimal number of featured ACAFRGs by LASSO logistic regression, in which 53 featured ACAFRGs were determined when the lambda value was minimal. **(B)** Determination of the optimal number of featured ACAFRGs by using the SVM-RFE algorithm (*N* = 31). **(C)** Venn plot showing the 15 overlapped featured ACAFRGs obtained by the aforementioned methods. **(D)** Optimal number of the random forest trees was determined when the cross-validation error presented minimal. The red dots represent the samples in gene cluster A, the green dots represent the samples in gene cluster B, and the black dots represent all the samples. **(E)** Importance of the featured genes in which eight ACAFRGs with importance higher than 10 were selected as the optimal

(*Continued*)

FIGURE 3
featured ACAFRGs. **(F)** ROC curve demonstrating the accuracy of the featured genes for discriminating two gene clusters (take EMP3 for example, AUC value = 0.991). **(G)** Construction of the nomogram model based on the eight featured ACAFRGs to calculate the probability of gene cluster A for each sample. **(H)** Calibration curve revealed the accuracy of the nomogram model. **(I)** DCA of the nomogram model. **(J)** Clinical impact curves of the nomogram model. **(K)** Unsupervised clustering of samples based on the expression of the eight featured ACAFRGs. ACAFRGs, aging cancer-associated fibroblast related genes; CAF, cancer-associated fibroblast; LASSO, least absolute shrinkage and selection operator; SVM-RFE, support vector machine-recursive feature elimination; ROC, receiver operating characteristic; AUC, area under curve; DCA, decision curve analysis.

divided into low- and high-aging CAF score groups. Kaplan–Meier survival analysis suggested that the high-aging CAF score group exhibited a worse prognosis than the low-score group in TCGA cohort (Figure 4A). ROC curves revealed that the accuracy of the aging CAF score was 0.874, 0.843, and 0.816 when predicting the 1, 2, and 3-year overall survival of LGG samples, respectively, in TCGA cohort (Figure 4B). Univariate Cox regression analysis indicated that the aging CAF score was significantly correlated with the prognosis of LGG samples and multivariate Cox regression analysis demonstrated that the aging CAF score served as an independent prognostic factor in TCGA cohort (Figure 4C). Similar results were acquired in the CGGA cohort (Supplementary Figures S6A–C). The relationship between the aging CAF score and multiple clinicopathological characteristics was explored in our research. We found that the proportion of glioma patients of grade III (G3) was significantly higher in the high-aging CAF score group than those in the low-score group (Figure 4D). IDH1 mutation was more frequent in the low-aging CAF score group (Figure 4E). Gliomas were more likely to recur or progress in the high-aging CAF score group than in the low-score group (Figure 4F). With respect to the therapeutic response to conventional treatment, more patients got complete or partial remission in the low-aging CAF score group compared to the high-score group (Figure 4G). Based on clinical features such as age, gender, and grade, LGG samples were stratified into different subgroups. We compared the aging CAF scores between subgroups with different clinical features. The aging CAF scores of old individuals were significantly higher than those of young individuals and the scores of samples with G3 were higher compared to G2. There was no correlation between the aging CAF score and gender (Supplementary Figure S7A). The overall survival of the high-aging CAF score group was substantially shorter than those of the low-score group even though samples were separated into subgroups with different clinical features (Supplementary Figure S7B). All these findings suggested that patients in the high-aging CAF score group tended to get a poor prognosis.

Consistent with the aforementioned results, we found that most samples of gene cluster A belonged to the high-aging CAF score group and all samples of gene cluster B were classified into the low-aging CAF score group (Supplementary Figure S6D). In addition, the aging CAF scores of gene cluster A were significantly higher than those of gene cluster B (Supplementary Figure S6E).

We further probed into the correlation between the expression patterns of ACAFRGs and aging CAF scores. As shown in Figures 4H,I, the expression levels of unfavorable ACAFRGs were upregulated with the increase in the aging CAF score while the expression levels of favorable ACAFRGs were downregulated with the increase in the aging CAF score. Differentially enriched functions between the two aging CAF score groups were analyzed to explore the underlying molecular mechanisms. We found that stroma-related functions were active in the high-aging CAF score group, including fibronectin binding, collagen binding, ECM receptor interaction, focal adhesion, and cell adhesion molecule (CAM)-related pathways, which may contribute to tumorigenesis and progression (Walker et al., 2018; Winkler et al., 2020; Bhargav et al., 2022) (Figures 4J,K).

Construction of a nomogram model

A nomogram model based on the aging CAF score and multiple clinicopathological factors was constructed to improve the predictive ability for prognosis (Figure 5A). As shown in Figure 5B, the values of the C-index for the aging CAF score, aging CAF score group, and nomogram model were 0.827, 0.942, and 0.849, respectively, indicating a good performance (the C-index of 0.5 represents a random chance and 1.0 represents ideal ability to predict the prognosis). The calibration curves of the nomogram model presented a good agreement between the prediction and the actual observation (Figure 5C). The AUC values of ROC curves for the nomogram model were 0.862 and 0.829 when predicting the 2 and 3-year overall survival, respectively, indicating the high accuracy of the model (Figures 5D,E). The results of DCA demonstrated that the nomogram model got great net benefits across a large range of risk thresholds (Figure 5F). All these findings revealed the powerful performance of the nomogram model.

Exploration of the correlation between the aging CAF score and TME

The abundance of critical compositions in the TME calculated through multiple methods was involved in our study to extensively explore the correlation between the TME

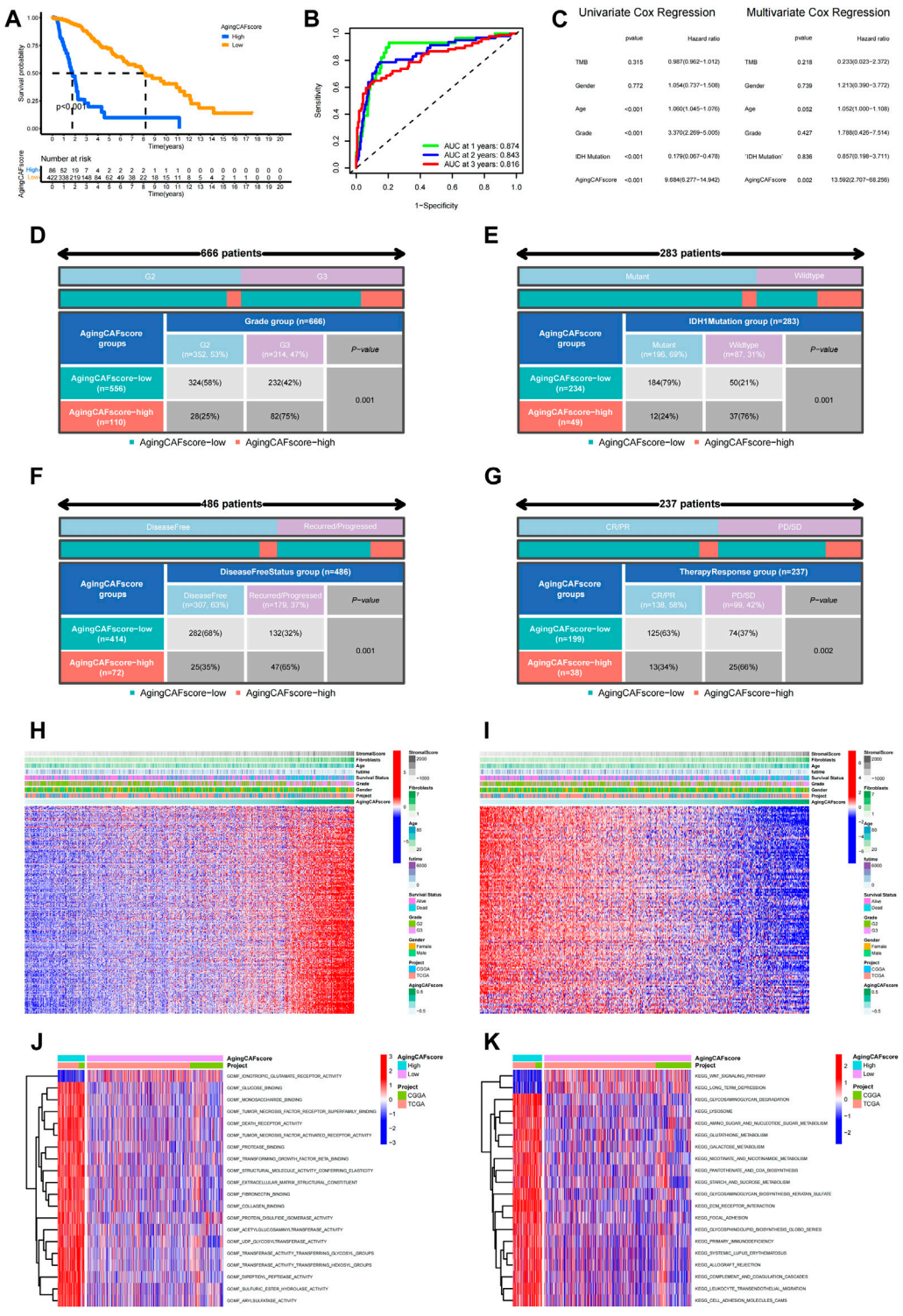


FIGURE 4
Comparison of the prognosis between low- and high-aging CAF score groups. **(A)** Kaplan–Meier survival analysis demonstrated that the high-aging CAF score group had a worse prognosis than the low-aging CAF score group in TCGA cohort ($p < 0.001$). **(B)** Time-dependent ROC curves of the aging CAF score in TCGA cohort. **(C)** Univariate/multivariate Cox regression analysis of the aging CAF score in TCGA cohort. **(D–G)** Correlation analysis between the aging CAF score group and grade **(D)**, IDH1 mutation status **(E)**, disease-free status **(F)**, and conventional therapy response **(G)**. All *p*-values less than 0.05. **(H)** Heatmap showing the expression patterns of unfavorable ACAFRGs with increasing aging CAF scores. **(I)** Heatmap showing the expression patterns of favorable ACAFRGs with increasing aging CAF scores. **(J)** Differentially enriched molecular functions between

(*Continued*)

FIGURE 4
low- and high-aging CAF score groups. **(K)** Differentially enriched KEGG pathways between low- and high-aging CAF score groups. The function terms with $|\log_2 FC| > 0.1$ and adjusted p-values (FDR) < 0.05 between two groups were considered differentially enriched. The values of the GSVA score for function terms were centered and scaled in the row direction. The top 20 differentially enriched function terms were shown in the heatmaps. CAF, cancer-associated fibroblast; ROC, receiver operating characteristic; AUC, area under curve; ACAFRGs, aging cancer-associated fibroblast-related genes; CR/PR, complete remission/partial remission; PD/SD, progressed disease/stable disease, GSVA, gene set variation analysis.

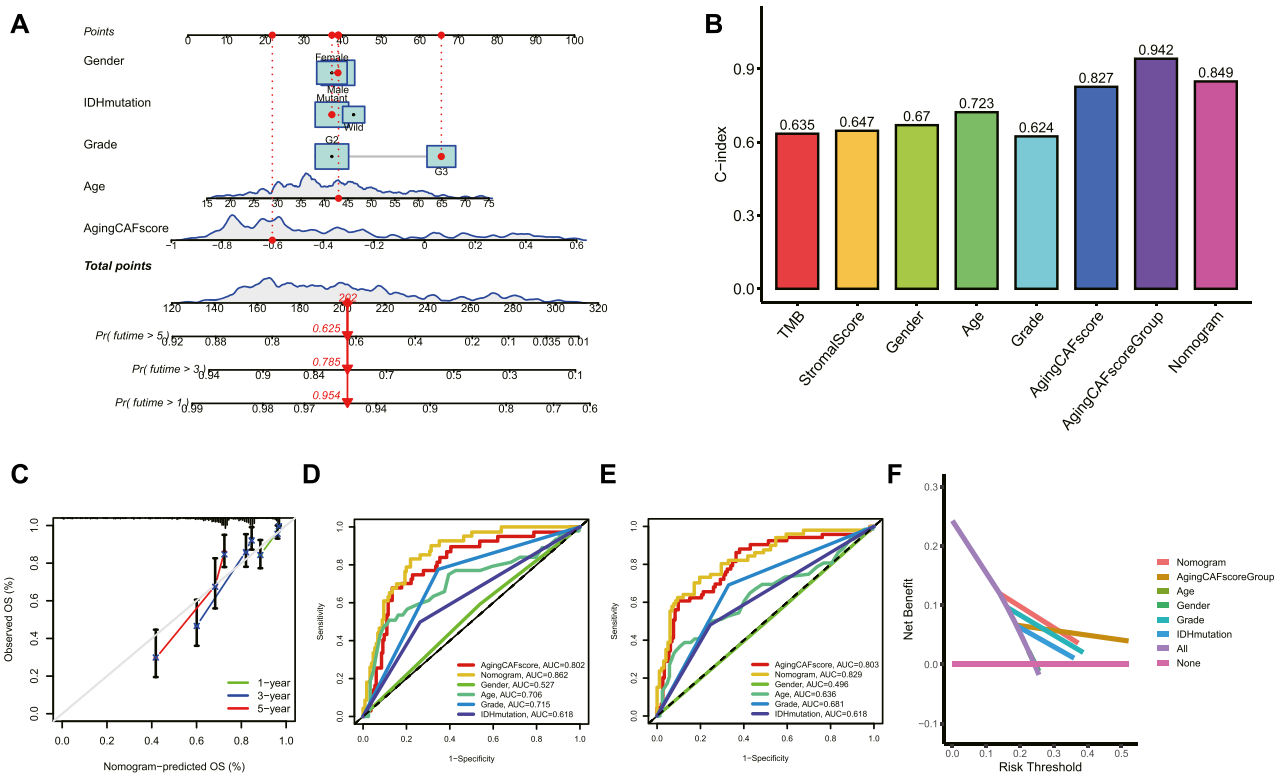


FIGURE 5
Construction of a nomogram model. **(A)** Nomogram model based on the aging CAF score and multiple clinicopathological factors was constructed. **(B)** Values of the C-index for the nomogram model and aging CAF score group were higher than those of other clinical factors. **(C)** Calibration curves for the nomogram model. **(D,E)** ROC curves of the nomogram model for predicting the 2 **(D)** and 3-year **(E)** overall survival. **(F)** DCA of the nomogram model for predicting the prognosis at 3 years. CAF, cancer-associated fibroblast; C-index, consistency index; ROC, receiver operating characteristic; AUC, area under curve; DCA, decision curve analysis.

and aging CAF score (Figure 6A, $p < 0.05$). We found that immune score, stromal score, microenvironment score, and the abundance of stromal cells such as CAFs and epithelial cells were positively correlated with the aging CAF score. The abundance of macrophages, especially M2 macrophages which served as an anti-inflammatory and tumor-promoting phenotype (Yunna et al., 2020), was significantly associated with the aging CAF score. $CD4^+$ T cells and $CD8^+$ T cells were also positively correlated with the aging CAF score. Consistently, the abundance of M2 macrophages and CAFs significantly increased in the high-aging CAF score group (Figure 6B). Similar results were obtained by quantifying the TME components for all samples in the merged data *via* MCP counter, CIBERSORT algorithm, and ssGSEA method, in which CAFs, M2 macrophages, regulatory T cells (Treg), and myeloid-derived suppressor cells (MDSCs) increased in the high-aging CAF score group (Supplementary Figures S8A–C). We also found that most of the immune checkpoints were highly expressed in the high-aging CAF score group compared to the low-score group (Figure 6C). The expression levels of most of the genes involved in the negative regulation of the cancer-immunity cycle increased in the high-aging CAF score group (Supplementary Figures 8D,E), the related gene list was downloaded from the Tracking Tumor Immunophenotype

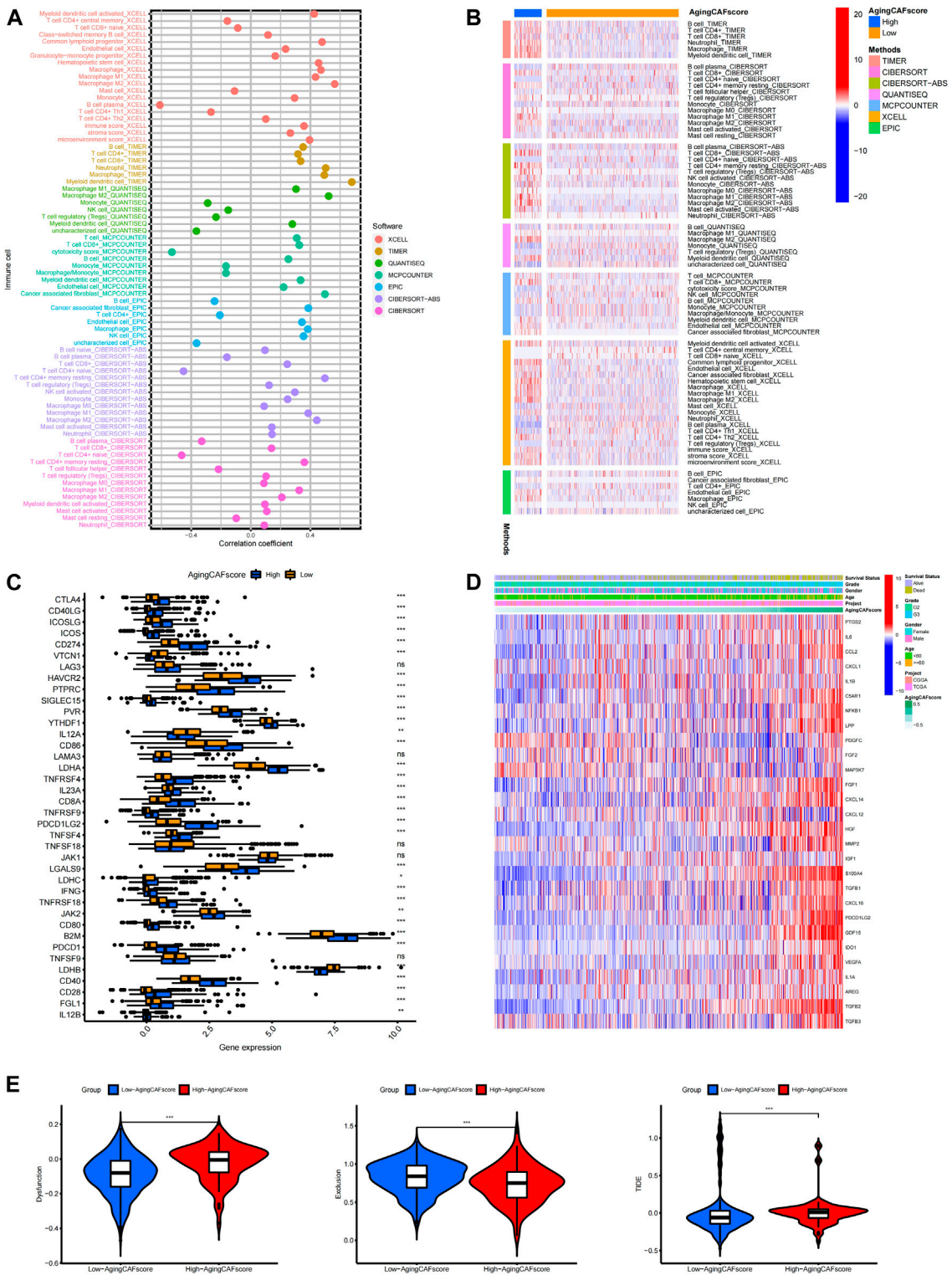


FIGURE 6
Exploration of the correlation between the aging CAF score and TME. **(A)** Correlation between the abundance of essential cells in the TME and aging CAF score. The cell types with *p*-values less than 0.05 were presented. **(B)** Heatmap showing the comparisons of the abundance of essential cells in the TME between low- and high-aging CAF score groups. The cell types with *p*-values less than 0.05 were presented. **(C)** Comparisons of the expression levels of immune checkpoints between the low- and high-aging CAF score groups. **(D)** Expression patterns of cytokines secreted by CAFs with increasing aging CAF scores. **(E)** Comparisons of TIDE-related scores between the two groups. * means $p < 0.05$, ** means $p < 0.01$, and ***means $p < 0.001$. CAF, cancer-associated fibroblast; TME, tumor microenvironment; TIDE, Tumor Immune Dysfunction and Exclusion.

website (http://biocc.hrbmu.edu.cn/). All these results suggested that the high-aging CAF score group tended to exhibit an immune-suppressive phenotype. Moreover, the expression levels of most of the cytokines secreted by CAFs upregulated with the increase in the aging CAF score (Figure 6D, Supplementary Figure S8F). Intriguingly, we found that some cytokines secreted by CAFs were also involved in senescence-associated secretory phenotype (SASP) factors which can be upregulated upon senescence, influence immune cell functions, and play tumor-promoting roles in the TME (Ruhland and Alspach, 2021), such as *IL6, AREG, CXCL12, TGFβ, VEGF,* and *CCL2*. These findings indicated that the aging CAF score developed by our study was genetically correlated with aging CAFs. The TIDE algorithm can be used to predict the potential response to immune checkpoint blockage treatment based on a comprehensive analysis of tumor immune dysfunction and exclusion mechanisms (Jiang et al., 2018). We detected that the dysfunction scores of the low-aging CAF score group were lower than those of the high-score group while the exclusion scores of the low-aging CAF score group were higher than those of the high-score group. Finally, we found that the low-aging CAF score group exhibited significantly lower TIDE scores than the high-score group, implying that patients in the low-aging CAF score group tended to benefit from immune checkpoint blockage treatment such as PD-1/PD-L1 blockage immunotherapy (all p values <0.001, Figure 6E).

In addition, the enrichment levels of the immune gene sets were quantified by single-sample gene set enrichment analysis (ssGSEA) based on the gene expression profiles of LGGs. Then, based on the ssGSEA scores of the immune cells, consensus clustering was performed to classify the LGG patients into different clusters, which were termed by immune subtypes (He et al., 2018). As shown in Supplementary Figure S9A, LGG samples were reasonable to be classified into three immune subtypes. We found that immune subtype B represented a subtype with high infiltration of immune cells, immune subtype C represented a subtype with low infiltration of immune cells, and immune subtype A represented a subtype with medium infiltration of immune cells (Supplementary Figure S9B). Furthermore, we explored the links between immune subtypes and the previously established model in our study. As shown in Supplementary Figures S9C,D, immune subtype B with high immunity showed the highest proportion of gene cluster A and the high-aging CAF score group while immune subtype C with low immunity showed the lowest proportion of gene cluster A and the high-aging CAF score group ($p < 0.001$).

Exploration of genetic mutations for samples with low- and high-aging CAF scores

Previous studies demonstrated the underlying correlation between genetic alterations and the tumor immune microenvironment (Rooney et al., 2015). Thus, we further explored the features of genetic mutations for samples with low- and high-aging CAF scores. We found that TMBs were significantly lower in the low-aging CAF score group than in the high-score group ($p = 1.7e-12$) and the TMB was positively correlated with the aging CAF score ($R = 0.22$, $p = 9.2e-7$, Figure 7A). Moreover, the LGG patients with high TMBs and high-aging CAF scores received the shortest overall survival while the patients with low TMBs and low-aging CAF scores had the longest overall survival. The prognosis of patients with low TMBs and high-aging CAF scores was worse than those of patients with high TMBs and low-aging CAF scores, indicating that the aging CAF score served as an independent prognostic factor independent of the TMB (Figure 7B). Finally, we identified the top 20 genes with the highest mutation frequencies in the low- and high-aging CAF score groups (Figures 7C,D). *IDH1*, *TP53,* and *ATRX* represented the top three frequently mutated genes in the low-aging CAF score group while *EGFR*, *PTEN,* and *NF1* presented the highest mutation frequencies in the high-aging CAF score group.

Validation of aging CAF-related genes in external datasets

Two independent cohorts were employed to verify the aging CAF-related genes. First, two distinct aging CAF-related gene clusters were identified in the validation cohort (dataset ID: mRNAseq_693) based on the expression profiles of aging CAF-related genes by the consensus clustering method (Supplementary Figure S10A). Consistent with the aforementioned results, the prognosis of gene cluster A was worse than that of gene cluster B (Figure 8A). Moreover, we found that the expression levels of the eight featured genes for discriminating the two gene clusters substantially differed between gene clusters A and B in the validation cohort (Supplementary Figure S10B). The ROC curves demonstrated the high accuracy of the eight featured genes for discriminating the two gene clusters (Supplementary Figure S10C). Unsupervised clustering for glioma samples in the validation cohort demonstrated that samples can be easily discriminated against based on the expression of the eight featured ACAFRGs (Figure 8B). A nomogram model was also built to predict the probability of gene cluster A based on the expression of these featured genes (Supplementary Figure S10D). The calibration curve, DCA curve, and clinical impact curve confirmed the robust performance of the nomogram model (Supplementary Figures S6E–G). Finally, samples in the validation cohort were assigned with specific aging CAF scores based on a similar method, followed by which samples were separated into low- and high-aging CAF score groups. The prognosis of the high-aging CAF score group was worse than those of the low-score group (Figure 8C). The ROC curves for predicting 1, 2, and 3-year overall survival further verified the accuracy of the aging CAF score (Figure 8D). In addition, similar results were obtained

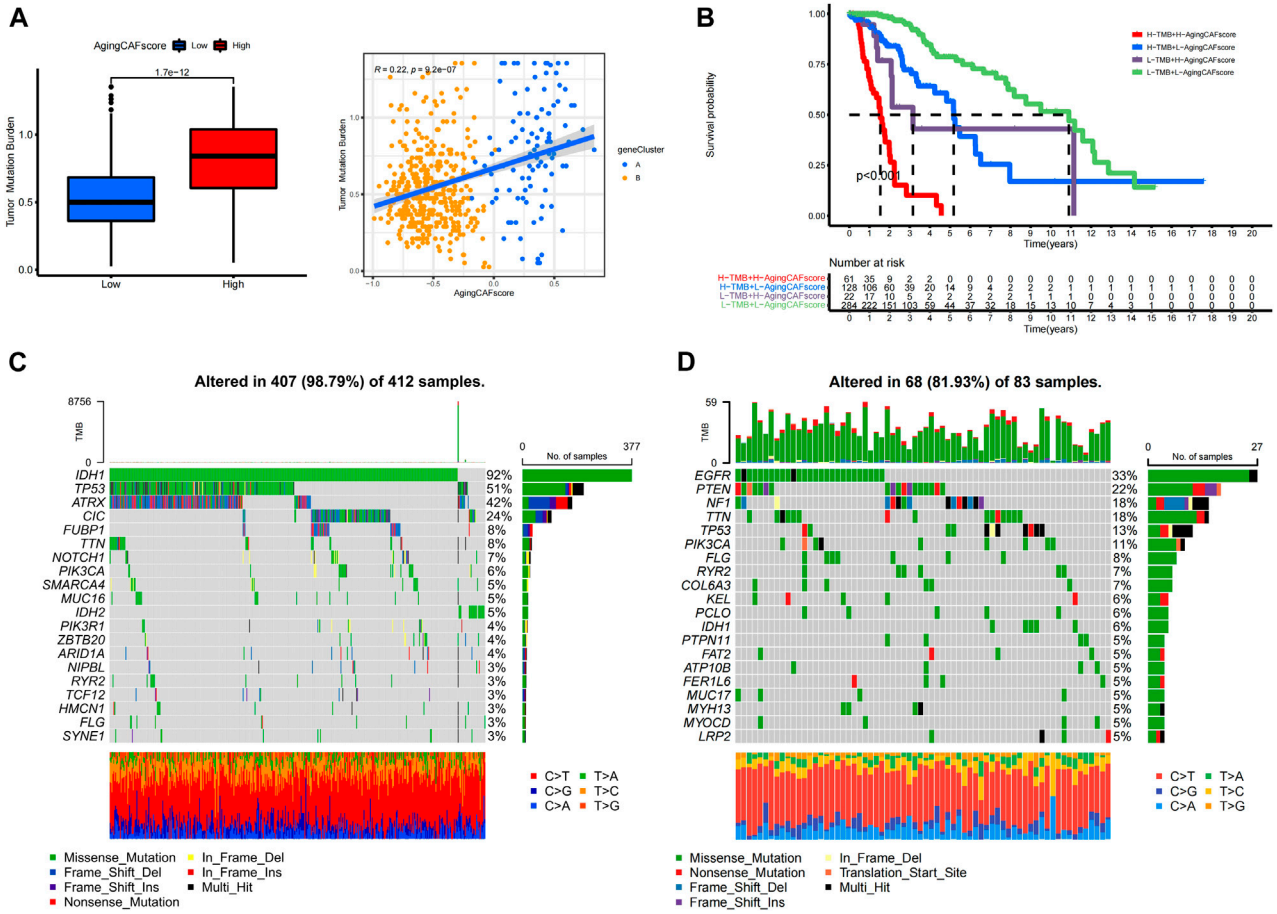


FIGURE 7
Exploration of genetic mutations for samples with low- and high-aging CAF scores. **(A)** Correlation between the TMB and aging CAF score. **(B)** Kaplan–Meier analysis of patients with different TMBs and aging CAF scores. **(C,D)** Top 20 genes with the highest mutation frequencies in the low- **(C)** and high- **(D)** aging CAF score groups. CAF, cancer-associated fibroblast; TMB, tumor mutation burden.

in another validation cohort (dataset ID: mRNAseq_325, Figures 8E–H, Supplementary Figure S11).

We also verified the performance of the aging CAF score in predicting the response to immune checkpoint blockage treatment in two independent cohorts. More patients were found to get favorable responses in the low-aging CAF score group compared to the high-score group ($p = 0.033$ in the GSE78220 cohort and $p = 0.03$ in the IMvigor210 cohort, Figures 8I,J).

Validation of the featured ACAFRGs at the protein level

A total of four ACAFRGs, namely, *RBP1, PDPN, FKBP9,* and *MSN* were randomly selected from the featured ACAFRGs. We found differential expression patterns of the aforementioned genes between normal and glioma tissues in immunohistochemistry staining on the Human Protein Atlas website (Figures 9A–D). Western blotting confirmed the high expression levels of six ACAFRGs in glioma tissues at the protein level (Figure 9E). As shown in Supplementary Figure S12, for the six molecules, we found that the optical density of G3 glioma samples was significantly higher than those of the normal brain samples. Although we detected that the six molecules were upregulated in the G2 glioma samples compared to the normal brain samples (Figure 9E), the statistical analysis of the optical density of the bands demonstrated no significant difference between them except PDPN and RBP1.

Discussions

As the most common stromal component in the TME, CAFs have been drawing increasing attention in cancer research for

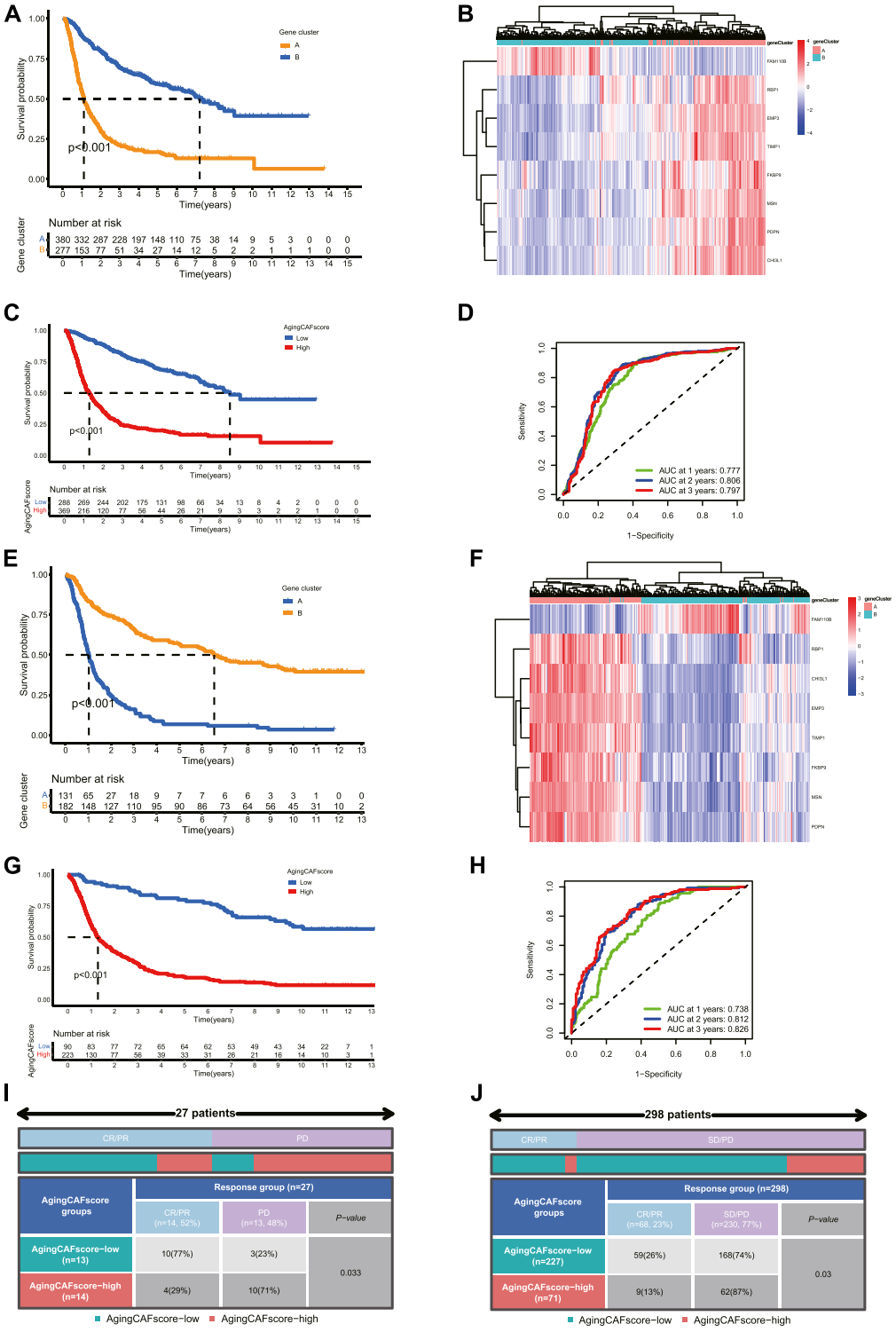


FIGURE 8
Validation of aging CAF-related genes in external datasets. **(A)** Comparison of the overall survival between gene clusters A and B in the validation cohort (dataset ID: mRNAseq_693). **(B)** Unsupervised clustering for samples based on the expression of eight featured ACAFRGs in the validation cohort (dataset ID: mRNAseq_693). **(C)** Comparison of the overall survival between the low- and high-aging CAF score groups in the validation cohort (dataset ID: mRNAseq_693). **(D)** Time-dependent ROC curves of the aging CAF score in the validation cohort (dataset ID: mRNAseq_693). **(E–H)** Similar results were obtained in another validation cohort (dataset ID: mRNAseq_325). **(I)** Comparison of the response to PD-1 immune checkpoint blockage treatment between the low- and high-aging CAF score groups in the GSE78220 cohort. **(J)** Comparison of the

(*Continued*)

FIGURE 8
response to PD-L1 immune checkpoint blockage treatment between the low- and high-aging CAF score groups in the IMvigor210 cohort. CAF, cancer-associated fibroblast; ACAFRGs, aging cancer-associated fibroblast related genes; ROC, receiver operating characteristic; AUC, area under curve; CR/PR, complete remission/partial remission; PD/SD, progressed disease/stable disease.

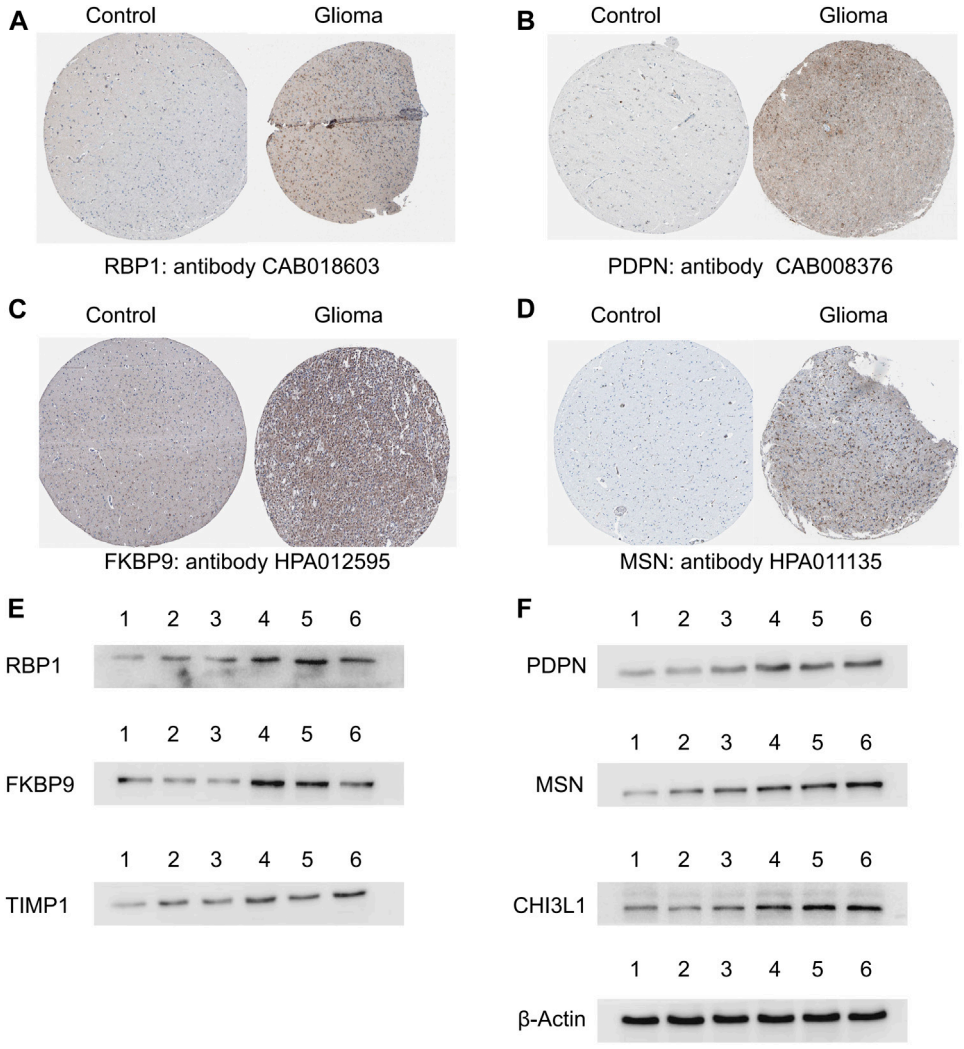


FIGURE 9
Validation of the featured ACAFRGs at the protein level. **(A–D)** Differential expression patterns of the featured genes between normal brain tissues and glioma tissues which were identified in immunohistochemistry staining on the Human Protein Atlas website. **(E)** Identification of the featured genes by Western blotting, in which lane 1 represents normal brain tissues, lanes 2 and 3 represent grade II glioma tissues, and lanes 4, 5, and 6 represent grade III glioma tissues. ACAFRGs, aging cancer-associated fibroblast related genes; Control: normal brain tissue.

their indispensable roles in tumor initiation and progression. Previous studies have revealed that CAFs have context-dependent functions, harboring both tumor-promoting and tumor-suppressive roles (Chen et al., 2021). The significant impact of CAFs on the regulation of anti-tumor immunity makes it possible to predict the immunotherapy response based on CAF-related biomarkers (Miyai et al., 2022). Moreover, considering that cancer has been previously

regarded as a disease of aging, CAFs are demonstrated to be particularly susceptible to aging-related impact in the context of tumor development (Yasuda et al., 2021a). Therefore, studies focusing on aging CAFs may provide a new direction for exploring biomarkers by drawing implications in predicting prognosis and immunotherapy in gliomas. In our study, we identified ACAFRGs in LGGs by the WGCNA method, based on which LGG samples were classified into two aging CAF-related gene clusters with distinct prognosis and characteristics of the TME. Machine learning algorithms were used to screen out the eight featured ACAFRGs to characterize two aging CAF-related gene clusters, and a nomogram model was constructed to predict the probability of gene cluster A for each LGG sample. Then, a powerful aging CAF scoring system was developed to predict the prognosis and response to immune checkpoint blockage therapy in the current research. Finally, the ACAFRGs were verified in two glioma-related external datasets. The performance of the aging CAF score in predicting the immunotherapy response was further validated in two independent cohorts with the information on immune checkpoint blockage treatment. We also confirmed the expression of ACAFRGs at the protein level in glioma tissues.

CAFs are a critical component in the stroma of the TME with a variety of functions, including generating and remodeling extracellular matrix components and complex interactions with tumor cells and other cell types in the TME (Sahai et al., 2020). Epigenetic alterations of CAFs enable the production and release of multiple cytokines, chemokines, exosomes, and metabolites, which impacts cancer progression, regulation of anti-tumor immunity, and metabolism (Chen et al., 2021). Based on the distinct transcriptome, single-cell RNA sequencing analysis has identified several subpopulations of CAFs with different functions, which has increased our understanding of the high heterogeneity of CAFs (Kieffer et al., 2020). Similar to the opposing effects of CAFs on tumor cells (tumor-permissive and tumor-suppressive effects), aging CAFs have also been demonstrated to exhibit both pro- and anti-tumorigenic activity (Ruhland and Alspach, 2021). Nevertheless, the substantial roles of aging CAFs in tumor promotion have always been underscored in recent years. Researchers have reported the dynamic evolving epigenetic changes of CAFs in the course of aging. Aging CAFs facilitate tumor progression mainly by the secretion of SASP factors which can cause chronic inflammation, promote angiogenesis, and enhance immunosuppressive activity (Yasuda et al., 2021a). However, it is important to note that current studies focusing on aging CAFs are facing the challenge of identifying both the aging status and the cell type of CAFs. Fibroblasts can be determined only by the absence of markers which define epithelial cells, endothelial cells, and immune cells (Sahai et al., 2020). In our study, we determined a total of 463 ACAFRGs which were significantly correlated with aging CAFs in LGG samples, such as *CDKN2B, CCL4, CCL19,* and *ISLR*. The *CDKN2B* (cyclin-dependent kinase inhibitor 2B) gene locates at exon 1 of *CDKN2B-AS1,* and the encoded protein serves as a regulator in cell cycle G1 progression by interacting with CDK kinases (Sibin et al., 2016). *CDKN2B* is found to be highly differentially expressed in aged individuals (Sebastiani et al., 2021). Considering that cellular senescence has previously been defined as a state of permanent cell cycle arrest, as a cell cycle arrest gene, *CDKN2B* has been reported to contribute to extracellular matrix deposition and cellular senescence (Rathi et al., 2020). CCL family members are involved in the chemokines and cytokines which can be strongly expressed by CAFs to enhance the pro-tumorigenic activity of myeloid cells (Monteran and Erez, 2019). Meflin (*ISLR*) is defined as a new cell surface marker for cancer-restraining CAFs in pancreatic and colon cancers. Tumor-suppressive roles of meflin-positive CAFs have been proposed, which are mediated by the regulation of collagen structures and bone morphogenetic protein (BMP) signaling in the TME (Takahashi et al., 2021). Meflin has been also reported to correlate with favorable prognosis and therapeutic response to immune checkpoint blockage treatment in patients with non-small cell lung cancer (NSCLC) (Miyai et al., 2022). Alternatively, meflin is determined as an unfavorable gene with a predictive value in patients with colon adenocarcinoma (Wang et al., 2021). Consistently, univariate Cox regression analysis revealed that meflin (*ISLR*) served as an unfavorable gene in LGGs in our research.

A total of 400 ACAFRGs with prognostic values were used to segregate LGG samples into two aging CAF-related gene clusters. We found that worse survival was associated with gene cluster A which was characterized by more infiltrating immune cells and fibroblasts in the TME, compared to gene cluster B. A total of eight featured ACAFRGs were determined to discriminate the two gene clusters, namely, *FAM110B, RBP1, FKBP9, MSN, PDPN, TIMP1, EMP3, and CHI3L1*, based on which a nomogram with robust performance was constructed to predict the probability to be grouped into gene cluster A for each glioma patient. In addition, each gene involved in the eight featured ACAFRGs exhibited high accuracy to characterize the gene clusters. For example, as a member of the FAM110 family (family with sequence similarity 110), *FAM110B* with an AUC value of 0.934 was downregulated in gene cluster A and the high-aging CAF score group (Supplementary Figure S4, Figure 3K, Figure 4I). Previous studies demonstrated that *FAM110B* participated in the regulation of the cell cycle and predicted favorable prognosis in NSCLC (Xie et al., 2020). In agreement with this study, *FAM110B* was found to be involved in the favorable gene set which was used to calculate the aging CAF score. As an unfavorable gene involved in ACAFRGs, *CHI3L1* was highly expressed in gene cluster A and the high-aging CAF score group (Figure 3K, Figure 4H). Consistent with our results, *CHI3L1* has been reported to be associated with poor prognosis in hepatocellular carcinoma (Wang et al., 2022). A recent study showed that *CHI3L1* was significantly correlated with severe state and adverse prognosis for COVID-19 patients (Kimura et al., 2021). Intriguingly, *CHI3L1* has been shown to positively regulate the PD-1/PD-L1 axis and other immune checkpoint molecules, potentially implying its impact on immunotherapy response (Ma et al., 2021). It is important to emphasize that the high-aging CAF score group has been found to receive less response to immune

checkpoint blockage therapy in the current research, which might be attributed to the aforementioned mechanism.

Based on the expression profiles of unfavorable and favorable ACAFRGs, we constructed a novel aging CAF scoring system. We detected that the high-aging CAF score predicted poor prognosis and a less favorable response to immune checkpoint blockage therapy in LGGs. The powerful performance of the aging CAF score was further verified in external cohorts. In this study, our analysis culminated in several important points: 1) as shown in Supplementary Figure S13, the aging CAF score was positively correlated with age (R = 0.17, p = 1.7e-5) and the age values of the high-aging CAF score group were significantly higher than those of the low-score group (p = 1.9e-12); 2) the aging CAF score was shown to be positively associated with the abundance of CAFs within the TME, as shown in Figure 6A; 3) the SASP factors and cytokines secreted by CAFs were robustly upregulated with the increasing aging CAF score. All these findings suggested that, except for the predictive performance, the aging CAF score developed in our study may serve as an indicator to quantify the abundance of aging CAFs in the TME. To some extent, the high aging CAF score predicted a poor prognosis and indicated more abundance of aging CAFs in the TME, which was consistent with the widely accepted concept that aging CAFs contributed to the proliferation and invasion of the surrounding cancer cells (Yasuda et al., 2021a). In addition, previous studies have revealed that aging CAFs promote the recruitment of M2 macrophages to enhance the immunosuppressive phenotype (Ruhland et al., 2016). We found that well-defined immune cells negatively regulating immune responses such as regulatory T cells (Tregs), M2 macrophages, and MDSCs presented high abundance in the high-aging CAF score group accompanied by the high expression of genes negatively regulating anti-tumor immunity (Figure 6B, Supplementary Figure S8B–E). This indicated that the high-aging CAF score group had an immunosuppressive phenotype, which was in accordance with the current point that aging CAFs acted as a major driver of immunosuppression within the TME (Ruhland and Alspach, 2021). To confirm the outstanding performance of the aging CAF score, we compared the risk score from multivariate Cox regression analysis based on ACAFRGs with it. First, we calculated the risk score for each LGG sample based on the expression profiles of prognostic ACAFRGs by using multivariate Cox regression analysis method. The critical genes involved in the construction of the risk score and the corresponding coefficients are listed in Supplementary Table S4. Similarly, LGG samples were divided into high- and low-risk score groups. As shown in Supplementary Figures S14A,B, LGG patients in the high-risk score group had a worse prognosis than those in the low-risk score group (p < 0.001) and the univariate/multivariate Cox regression analysis indicated that the risk score was significantly correlated with prognosis and served as an independent prognostic factor (all p values < 0.001). However, the ROC curves revealed that the accuracy of the aging CAF score was higher than those of the risk score for predicting the prognosis (Supplementary Figures S14C–E). For example, the AUC value of the aging CAF score for predicting the 2-year overall survival was 0.802 while the AUC value of the risk score was 0.776. Similarly, the results of DCA showed that the aging CAF score was a more powerful predictor than the risk score. Moreover, as shown in Supplementary Figure S14F, the C-index value of the aging CAF score group (0.942) was higher than those of the risk score group (0.913). Similar to the aging CAF score, the risk score can also predict the characteristics of the TME (Supplementary Figure S14G), immunotherapy response (Supplementary Figure S14H), aging CAF-related gene clusters (Supplementary Figure S14I), and tumor mutation burden (Supplementary Figure S14J). However, the risk score cannot predict the immunotherapy response in the validation cohorts (p = 0.853 in GSE78220, p = 0.077 in the IMvigor210 cohort, Supplementary Figures S14K,L). The robust capacity of the aging CAF score for predicting the immunotherapy response has been verified in the aforementioned external datasets in our study. Taking all these results into consideration, we believe that the aging CAF score in our study has a more powerful potential to predict prognosis and immunotherapy response than the risk score from multivariate Cox regression analysis.

Moreover, our results indicated that the high-aging CAF score group bore more TMBs than the low-score group (Figure 7A). It is to be noted that the top three genes with the highest mutation frequencies in the high-aging CAF score group were *EGFR*, *PTEN*, and *NF1*. The mutation frequency of *EGFR* in GBMs has been shown to be higher than those in LGGs. *EGFR* mutation has been reported to be an independent predictor of the prognosis in all grades of gliomas (Saadeh et al., 2017). *PTEN* mutation has also been found to be significantly associated with reduced survival in gliomas (Zhang et al., 2021). Wang et al. demonstrated that *NF1* was more frequently mutated in GBMs compared to LGGs and has been used to define the mesenchymal subtype of GBMs (Verhaak et al., 2010). In contrast, *IDH1* exhibited the highest mutation frequency in the low-aging CAF score group, which was consistent with the well-identified concept that *IDH* mutation was associated with better prognosis in gliomas (Turkalp et al., 2014). It seemed that LGGs in the high-aging CAF score group may represent a subgroup similar to GBMs from the point of genetic variation.

Overall, in our study, we determined a series of ACAFRGs in LGGs, based on which a robust aging CAF scoring system was developed to predict the prognosis and immunotherapy response. Our findings may provide new targets for therapeutics and contribute to future exploration focusing on aging CAFs.

Data availability statement

Publicly available datasets were analyzed in this study. We declare that the datasets supporting the findings of this study are available in TCGA database (https://portal.gdc.cancer.gov/),

CGGA database(http://cgga.org.cn/index.jsp), and GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Qingdao municipal hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

BL and ZZ contributed to the study conception and design. Material preparation and analysis were performed by ZZ, JW, LK, KZ, YL, and ZH. Data collection and visualization were conducted by ZZ, YL, LL, ZH, and BL. The first draft of the manuscript was written by ZZ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (82002642).

Acknowledgments

The authors would like to thank TCGA, GEO, and CGGA databases and the Human Protein Atlas website for data availability.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.897083/full#supplementary-material

References

Anderson, N. M., and Simon, M. C. (2020). The tumor microenvironment. *Curr. Biol.* 30 (16), R921–R925. doi:10.1016/j.cub.2020.06.081

Baker, A. T., Abuwarwar, M. H., Poly, L., Wilkins, S., and Fletcher, A. L. (2021). Cancer-associated fibroblasts and T cells: From mechanisms to outcomes. *J. I.* 206 (2), 310–320. doi:10.4049/jimmunol.2001203

Balar, A. V., Galsky, M. D., Rosenberg, J. E., Powles, T., Petrylak, D. P., Bellmunt, J., et al.fnm (2017). Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: A single-arm, multicentre, phase 2 trial. *Lancet* 389(10064), 67–76. doi:10.1016/S0140-6736(16)32455-2

Bao, Z.-S., Chen, H.-M., Yang, M.-Y., Zhang, C.-B., Yu, K., Ye, W.-L., et al. (2014). RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res.* 24 (11), 1765–1773. doi:10.1101/gr.165126.113

Becht, E., Giraldo, N., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218. doi:10.1186/s13059-016-1070-5

Bhargav, A. G., Domino, J. S., Chamoun, R., and Thomas, S. M. (2022). Mechanical properties in the glioma microenvironment: Emerging insights and theranostic opportunities. *Front. Oncol.* 11, 805628. doi:10.3389/fonc.2021.805628

Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324

Cairncross, J. G., Wang, M., Jenkins, R. B., Shaw, E. G., Giannini, C., Brachman, D. G., et al. (2014). Benefit from procarbazine, lomustine, and vincristine in oligodendroglial tumors is associated with mutation of IDH. *J. Clin. Oncol. official J. Am. Soc. Clin. Oncol.* 32 (8), 783–790. doi:10.1200/JCO.2013.49.3726

Chen, Y., McAndrews, K. M., and Kalluri, R. (2021). Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat. Rev. Clin. Oncol.* 18 (12), 792–804. doi:10.1038/s41571-021-00546-5

Erin, N., Grahovac, J., Brozovic, A., and Efferth, T. (2020). Tumor microenvironment and epithelial mesenchymal transition as targets to overcome tumor multidrug resistance. *Drug Resist. Updat.* 53, 100715. doi:10.1016/j.drup.2020.100715

Fane, M., and Weeraratna, A. T. (2020). How the ageing microenvironment influences tumour progression. *Nat. Rev. Cancer* 20 (2), 89–106. doi:10.1038/s41568-019-0222-9

Fang, S., Liang, J., Qian, T., Wang, Y., Liu, X., Fan, X., et al. (2017). Anatomic location of tumor predicts the accuracy of motor function localization in diffuse lower-grade gliomas involving the hand knob area. *Am. J. Neuroradiol.* 38 (10), 1990. doi:10.3174/ajnr.A5342

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14 (1), 7. doi:10.1186/1471-2105-14-7

He, Y., Jiang, Z., Chen, C., and Wang, X. (2018). Classification of triple-negative breast cancers based on Immunogenomic profiling. *J. Exp. Clin. cancer Res. CR* 37 (1), 327. doi:10.1186/s13046-018-1002-1

Hottinger, A. F., Hegi, M. E., and Baumert, B. G. (2016). Current management of low-grade gliomas. *Curr. Opin. Neurology* 29 (6). doi:10.1097/WCO.0000000000000390

Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165 (1), 35–44. doi:10.1016/j.cell.2016.02.065

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* 16 (9), 582–598. doi:10.1038/nrc.2016.73

Kieffer, Y., Hocine, H. R., Gentric, G., Pelon, F., Bernard, C., Bourachot, B., et al. (2020). Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov.* 10 (9), 1330–1351. doi:10.1158/2159-8290.CD-19-1384

Kimura, Y., Nakai, Y., Shin, J., Hara, M., Takeda, Y., Kubo, S., et al. (2021). Identification of serum prognostic biomarkers of severe COVID-19 using a quantitative proteomic approach. *Sci. Rep.* 11 (1), 20638. doi:10.1038/s41598-021-98253-9

Kobayashi, H., Enomoto, A., Woods, S. L., Burt, A. D., Takahashi, M., and Worthley, D. L. (2019). Cancer-associated fibroblasts in gastrointestinal cancer. *Nat. Rev. Gastroenterology Hepatology* 16 (5), 282–295. doi:10.1038/s41575-019-0115-0

Komori, T. (2017). The 2016 WHO classification of tumours of the central nervous system: The major points of revision. *Neurol. medico-chirurgica* 57 (7), 301–311. doi:10.2176/nmc.ra.2017-0010

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034

Li, C., Tang, Z., Zhang, W., Ye, Z., and Liu, F. (2021). GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. *Nucleic acids Res.* 49 (1), W242–W246. doi:10.1093/nar/gkab418

Liu, X., Li, Y., Qian, Z., Sun, Z., Xu, K., Wang, K., et al. (2018). A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. *NeuroImage Clin.* 20, 1070–1077. doi:10.1016/j.nicl.2018.10.014

Ma, B., Akosman, B., Kamle, S., Lee, C.-M., He, C. H., Koo, J. S., et al. (2021). CHI3L1 regulates PD-L1 and anti-CHI3L1-PD-1 antibody elicits synergistic antitumor responses. *J. Clin. investigation* 131 (21), e137750. doi:10.1172/JCI137750

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Miyai, Y., Esaki, N., Takahashi, M., and Enomoto, A. (2020). Cancer-associated fibroblasts that restrain cancer progression: Hypotheses and perspectives. *Cancer Sci.* 111 (4), 1047–1057. doi:10.1111/cas.14346

Miyai, Y., Sugiyama, D., Hase, T., Asai, N., Taki, T., Nishida, K., et al. (2022). Meflin-positive cancer-associated fibroblasts enhance tumor response to immune checkpoint blockade. *Life Sci. Alliance* 5 (6), e202101230. doi:10.26508/lsa.202101230

Monteran, L., and Erez, N. (2019). The dark side of fibroblasts: Cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front. Immunol.* 10, 1835. doi:10.3389/fimmu.2019.01835

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ostrom, Q. T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J. S. (2018). CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. *Neuro-Oncology* 20 (4), iv1–iv86. doi:10.1093/neuonc/noy131

Perry, A., and Wesseling, P. (2016). "Chapter 5 - histologic classification of gliomas," in *Handbook of clinical neurology*. Editors M. S. Berger and M. Weller (Elsevier), 71–95. doi:10.1016/b978-0-12-802997-8.00005-0

Piersma, B., Hayward, M. K., and Weaver, V. M. (2020a). Fibrosis and cancer: A strained relationship. *Biochimica biophysica acta. Rev. cancer* 1873 (2), 188356. doi:10.1016/j.bbcan.2020.188356

Piersma, B., Hayward, M. K., and Weaver, V. M. (2020b). Fibrosis and cancer: A strained relationship. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1873 (2), 188356. doi:10.1016/j.bbcan.2020.188356

Pitt, Jonathan M., Vétizou, M., Daillère, R., Roberti, María P., Yamazaki, T., Routy, B., et al. (2016). Resistance mechanisms to immune-checkpoint blockade in cancer: Tumor-intrinsic and -extrinsic factors. *Immunity* 44 (6), 1255–1269. doi:10.1016/j.immuni.2016.06.001

Radin, D. P., and Tsirka, S. E. (2020). Interactions between tumor cells, neurons, and microglia in the glioma microenvironment. *Int. J. Mol. Sci.* 21 (22), 8476. doi:10.3390/ijms21228476

Rathi, S., Danford, I., Gudiseva, H. V., Verkuil, L., Pistilli, M., Vishwakarma, S., et al. (2020). Molecular genetics and functional analysis implicate cdkn2bas1-CDKN2B involvement in POAG pathogenesis. *Cells* 9 (9), 1934. doi:10.3390/cells9091934

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Ruhland, M. K., and Alspach, E. (2021). Senescence and immunoregulation in the tumor microenvironment. *Front. Cell Dev. Biol.* 9 (2789), 754069. doi:10.3389/fcell.2021.754069

Ruhland, M. K., Loza, A. J., Capietto, A.-H., Luo, X., Knolhoff, B. L., Flanagan, K. C., et al. (2016). Stromal senescence establishes an immunosuppressive microenvironment that drives tumorigenesis. *Nat. Commun.* 7, 11762. doi:10.1038/ncomms11762

Saadeh, F. S., Mahfouz, R., and Assi, H. I. (2017). EGFR as a clinical marker in glioblastomas and other gliomas. *Int. J. Biol. Markers* 33 (1), 22–32. doi:10.5301/ijbm.5000301

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., et al. (2020). A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev.* 20 (3), 174–186. doi:10.1038/s41568-019-0238-1

Sebastiani, P., Federico, A., Morris, M., Gurinovich, A., Tanaka, T., Chandler, K. B., et al. (2021). Protein signatures of centenarians and their offspring suggest centenarians age slower than other humans. *Aging cell* 20 (2), e13290. doi:10.1111/acel.13290

Sibin, M. K., Dhananjaya, I. B., Narasingarao, K. V. L., Harshitha, S. M., Jeru-Manoj, M., and Chetan, G. K. (2016). Two gene polymorphisms (rs4977756 and rs11515) in CDKN2A/B and glioma risk in South Indian population. *Meta gene* 9, 215–218. doi:10.1016/j.mgene.2016.07.010

Smyth, G. K., Michaud, J., and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21 (9), 2067–2075. doi:10.1093/bioinformatics/bti270

Su, S., Chen, J., Yao, H., Liu, J., Yu, S., Lao, L., et al. (2018). $CD10^{+}GPR77^{+}$ cancer-associated fibroblasts promote cancer formation and chemoresistance by sustaining cancer stemness. *Cell* 172 (4), 841–856. e816. doi:10.1016/j.cell.2018.01.009

Suykens, J. A. K., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9 (3), 293–300. doi:10.1023/A:1018628609742

Takahashi, M., Kobayashi, H., Mizutani, Y., Hara, A., Iida, T., Miyai, Y., et al. (2021). Roles of the mesenchymal stromal/stem cell marker meflin/islr in cancer fibrosis. *Front. cell Dev. Biol.* 9, 749924. doi:10.3389/fcell.2021.749924

Tibshirani, R. (1996). *Regres. Shrinkage Sel.* Via *Lasso* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2015). Immune checkpoint blockade: A common denominator approach to cancer therapy. *Cancer cell* 27 (4), 450–461. doi:10.1016/j.ccell.2015.03.001

Turkalp, Z., Karamchandani, J., and Das, S. (2014). IDH mutation in glioma: New insights and promises for the future. *JAMA Neurol.* 71 (10), 1319–1325. doi:10.1001/jamaneurol.2014.1205

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* 17 (1), 98–110. doi:10.1016/j.ccr.2009.12.020

Walker, C., Mojares, E., and Del Río Hernández, A. (2018). Role of extracellular matrix in development and cancer progression. *Int. J. Mol. Sci.* 19 (10), 3028. doi:10.3390/ijms19103028

Wang, S., Chen, S., Jin, M., Hu, M., Huang, W., Jiang, Z., et al. (2022). Diagnostic and prognostic value of serum Chitinase 3-like protein 1 in hepatocellular carcinoma. *J. Clin. laboratory analysis* 36 (2), e24234. doi:10.1002/jcla.24234

Wang, W., Xu, C., Ren, Y., Wang, S., Liao, C., Fu, X., et al. (2021). A novel cancer stemness-related signature for predicting prognosis in patients with colon adenocarcinoma. *Stem cells Int.* 2021, 7036059. doi:10.1155/2021/7036059

Wang, Y., Qian, T., You, G., Peng, X., Chen, C., You, Y., et al. (2015). Localizing seizure-susceptible brain regions associated with low-grade gliomas using voxel-based lesion-symptom mapping. *Neuro-Oncology* 17 (2), 282–288. doi:10.1093/neuonc/nou130

Wang, Y., Wang, Y., Fan, X., Li, S., Liu, X., Wang, J., et al. (2017). Putamen involvement and survival outcomes in patients with insular low-grade gliomas. *J. Neurosurg.* 126 (6), 1788–1794. doi:10.3171/2016.5.Jns1685

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* 11 (1), 5120. doi:10.1038/s41467-020-18794-x

Xie, M., Cai, L., Li, J., Zhao, J., Guo, Y., Hou, Z., et al. (2020). FAM110B inhibits non-small cell lung cancer cell proliferation and invasion through inactivating wnt/β-catenin signaling. *OncoTargets Ther.* 13, 4373–4384. doi:10.2147/OTT.S247491

Yasuda, T., Baba, H., and Ishimoto, T. (2021a). Cellular senescence in the tumor microenvironment and context-specific cancer treatment strategies. *FEBS J.* 1, 1. doi:10.1111/febs.16231

Yasuda, T., Koiwa, M., Yonemura, A., Miyake, K., Kariya, R., Kubota, S., et al. (2021b). Inflammation-driven senescence-associated secretory phenotype in cancer-associated fibroblasts enhances peritoneal dissemination. *Cell Rep.* 34 (8), 108779. doi:10.1016/j.celrep.2021.108779

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4 (1), 2612. doi:10.1038/ncomms3612

Yunna, C., Mengru, H., Lei, W., and Weidong, C. (2020). Macrophage M1/M2 polarization. *Eur. J. Pharmacol.* 877, 173090. doi:10.1016/j.ejphar.2020.173090

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene Co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4 (1). doi:10.2202/1544-6115.1128

Zhang, P., Meng, X., Liu, L., Li, S., Li, Y., Ali, S., et al. (2021). Identification of the prognostic signatures of glioma with different PTEN status. *Front. Oncol.* 11, 633357. doi:10.3389/fonc.2021.633357

Zhao, Z., Meng, F., Wang, W., Wang, Z., Zhang, C., and Jiang, T. (2017). Comprehensive RNA-seq transcriptomic profiling in the malignant progression of gliomas. *Sci. Data* 4 (1), 170024. doi:10.1038/sdata.2017.24

Zhao, Z., Zhang, K.-N., Wang, Q., Li, G., Zeng, F., Zhang, Y., et al. (2021). Chinese glioma Genome Atlas (CGGA): A comprehensive resource with functional genomic data from Chinese glioma patients. *Genomics, Proteomics Bioinforma.* 19 (1), 1–12. doi:10.1016/j.gpb.2020.10.005

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fgene.2022.888672

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Aiyan Xing,
Shandong University, China
Lin Hu,
Soochow University, China

*CORRESPONDENCE
Qian Ren,
ldyy_renqian@lzu.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 03 March 2022
ACCEPTED 26 July 2022
PUBLISHED 25 August 2022

CITATION
Peng K, Ren X and Ren Q (2022), NcRNA-mediated upregulation of CAMK2N1 is associated with poor prognosis and tumor immune infiltration of gastric cancer. *Front. Genet.* 13:888672. doi: 10.3389/fgene.2022.888672

COPYRIGHT
© 2022 Peng, Ren and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

NcRNA-mediated upregulation of CAMK2N1 is associated with poor prognosis and tumor immune infiltration of gastric cancer

Kaipeng Peng[1,2,3], Xiangqing Ren[1,2,3] and Qian Ren[1,2,3]*

[1]The First Clinical Medical College, Lanzhou University, Lanzhou, China, [2]Department of Gastroenterology, The First Hospital of Lanzhou University, Lanzhou, China, [3]Key Laboratory for Gastrointestinal Diseases of Gansu Province, The First Hospital of Lanzhou University, Lanzhou, China

Gastric cancer (GC) is still notorious for its poor prognosis and aggressive characteristics. Though great developments have been made in diagnosis and therapy for GC, the prognosis of patient is still perishing. In this study, differentially expressed genes (DEGs) in GC were first screened using three Gene Expression Omnibus (GEO) datasets (GSE13911, GSE29998, and GSE26899). Second, The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) data were used to validate expression of these DEGs and perform survival analysis. We selected seven candidate genes (CAMK2N1, OLFML2B, AKR7A3, CYP4X1, FMO5, MT1H, and MT1X) to carry out the next analysis. To construct the ceRNA network, we screened the most potential upstream ncRNAs of the candidate genes. A series of bioinformatics analyses, including expression analysis, correlation analysis, and survival analysis, revealed that the SNHG10–hsa-miR-378a-3p might be the most potential regulatory axis in GC. Then, the expression of CAMK2N1, miR-378a-3p, and SNHG10 was verified in GC cell lines (GES-1, MGC-803, BGC-823, HGC-27, MKN-45, and AGS) by qRT-PCR and Western blotting. We found that SNHG10 and CAMK2N1 were highly expressed in gastric cancer lines, and the miR-378a-3p was lowly expressed in BGC-823, HGC-27, and MKN-45. Furthermore, CAMK2N1 levels were significantly negatively associated with tumor immune cell infiltration, biomarkers of immune cells, and immune checkpoint expression. In summary, our results suggest that the ncRNA-mediated high expression of CAMK2N1 is associated with poor prognosis and tumor immune infiltration of GC.

KEYWORDS

CAMK2N1, gastric cancer, tumor immune infiltration, prognosis, noncoding RNA

Introduction

Gastric cancer (GC) is one of the most common cancers in the world (Fu, 2015). The incidence of GC, especially in the Asian region, has observably increased over the past decades (Wang et al., 2015). In 2018, GC was the fifth incidence and third mortality among all cancers worldwide (Komatsu and Otsuji, 2019). This disease is a highly heterogeneous malignancy with different histological and molecular subtypes. Despite advances in surgery and neoadjuvant chemotherapy, the prognosis of patients with advanced gastric cancer remains poor, and the median overall survival (OS) was about 11 months, and the five-year survival rate was about 30% (Khorana et al., 2013; Digklia and Wagner, 2016; Smyth et al., 2016). Therefore, it is necessary to find effective therapeutic targets or seek valuable prognostic biomarkers in GC .

Non-coding RNAs (ncRNAs), which include miRNA, circRNA, and lncRNA, play a key role in human physiological and pathological activities (Anastasiadou et al., 2018; Kanwal and Lu, 2019). In recent years, ncRNAs have been the hotspot of human disease research (Esteller, 2011). NcRNAs have essential implications for human life activities; moreover, dysregulation of ncRNAs causes numerous disorders of human life activities, including cancer (Eddy, 2001). A competing endogenous RNA (ceRNA) hypothesis was proposed in 2011, which is a regulatory mechanism between mRNAs and ncRNAs (Salmena et al., 2011). The CeRNA mechanism assumes that competitively binding between lncRNAs, pseudogenes, circRNAs, and mRNAs to share miRNAs can regulate a series of biological functions (Karreth and Pandolfi, 2013). Increasing ncRNAs have been reported to influence tumor pathological processes (Wang et al., 2014a; Wang et al., 2017; Ding et al., 2019). In addition, ncRNAs have been found to have the value of potential diagnostic and prognostic biomarkers for cancer in blood (Qi et al., 2016; Zhang et al., 2017; Wei et al., 2019).

In this study, we constructed a ceRNA network linked to GC by a series of analytical processes. First, we identified differentially expressed genes (DEGs) associated with GC using three Gene Expression Omnibus (GEO) datasets (GSE13911, GSE29998, and GSE26899). Second, TCGA GC cohort was used to validate the expression of these DEGs and perform survival analysis. In addition, we excluded DEGs, which have been reported as prognostic biomarkers in GC. Subsequently, upstream miRNAs of the remaining DEGs were predicted using the starBase database. By expression correlation analysis and survival analysis, we identified the most potential miRNAs bound to CAMK2N1 and FOM5. Then, we identified the most potential lncRNA in the same way. Finally, we determined the association of CAMK2N1 expression with immune cell infiltration, biomarkers of immune cells, and immune checkpoints in GC. To confirm the reliability of the analysis results, we validated the expression of CAMK2N1 and ncRNAs in the human gastric cancer cell lines. To sum up, we found that ncRNA-related upregulation of CAMK2N1 correlates with poor prognosis and tumor immune infiltration of patients in GC.

TABLE 1 Table 1 Expression change of 10 candidate genes in gastric cancer.

Gene name	Log2FC[a]	Log2FC[b]	Log2FC[c]
CAMK2N1	1.11	1.30	1.89
OLFM2B	1.04	2.25	1.20
TNFAIP6	1.49	1.50	1.05
AKR7A3	−1.76	−2.17	−1.70
CYP4X1	−1.63	−1.57	−1.33
FMO5	−1.71	−1.78	−1.10
MT1H	−1.45	−2.50	−2.12
MT1M	−2.75	−3.07	−1.80
MT1X	−1.15	−1.51	−1.01
SIDT2	−1.51	−1.29	−1.09

FC[a] determined by GES13911.
FC[b] determined by GES29998.
FC[c] determined by GES26899.

Materials and methods

Microarray

The mRNA microarrays of GC were downloaded from the Xiantao database (https://www.xiantao.love/). We searched the GEO database with keywords such as “gastric cancer” and “mRNA.” GSE13911 (containing 37 tumor tissues and 32 normal tissues), GSE29998 (49 tumor tissues and 50 normal tissues), and GES26899 (95 tumor tissues and 13 normal tissues) were selected for the follow-up analysis of this study.

Differential expression analysis of mRNAs

The Xiantao database (https://www.xiantao.love/) is an online analytical tool, which was used to perform differential expression analysis based on the GEO database. DEGs were screened out using adjusted p-value <0.05, $|\log2FC| > 1$.

GEPIA database analysis

GEPIA (http://gepia.cancer-pku.cn/index.html) is a novel tool for gene-expression profiling and interactive analyses in different types of cancer which is based on TCGA and Genotype-

Tissue Expression (GTEx) data. We used this tool to determine DEG expression in GC. A p-value <0.05 was considered statistically significant.

Kaplan–Meier plotter analysis

The Kaplan–Meier plotter (http://kmplot.com/analysis/) is an online database which can access the prognostic values of genes or microRNAs in more than 20 cancer types including GC. A log rank p-value <0.05 was considered statistically significant.

starBase database analysis

starBase (http://starbase.sysu.edu.cn/) is a database for miRNA-related studies. starBase was introduced to perform expression correlation analysis for miRNA–DEGs, lncRNA–miRNA, and lncRNA–DEGs in GC. In addition, starBase was used to predict candidate miRNAs that could potentially bind to DEGs and lncRNAs that could bind to miRNAs. A p-value <0.05 was considered statistically significant.

miRNet database analysis

The miRNet database (https://www.mirnet.ca/) is a web server for integrated data from several miRNA-linked databases and was used to predict the potential lncRNAs binding to miRNAs. Then, these lncRNAs were intersected with the lncRNAs obtained from the starBase database to acquire the most potential upstream lncRNAs of hsa-miR-378a-3p and hsa-miR-140-5p.

TIMER database analysis

TIMER (https://cistrome.shinyapps.io/timer/) is an online tool for comprehensive relation analysis of tumor-infiltrating immune. TIMER was used to analyze the correlation of the CAMK2N1 expression level with the immune cell infiltration level or immune checkpoint expression level in GC. A p-value <0.05 was considered statistically significant.

LinkedOmics database analysis

The LinkedOmics database (http://www.linkedomics.org/login.php) is a website for analyzing multi-omics data based on TCGA datasets. We obtained the differentially expressed genes related to CAMK2N1 in GC using the LinkFinder module, which were analyzed by the Pearson correlation coefficient and visualized by volcano plot and heat maps. To get descriptive information, the differentially expressed genes related to CAMK2N1 were annotated using Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.

Cell lines

Normal human gastric mucosa epithelial cell line GES-1 and gastric cancer cell lines AGS, MGC-803, BGC-823, HGC-27, and MKN-45 were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). The cells were cultured in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, Waltham, United States) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, United States), 100 U/ml penicillin, and 100 μg/ml streptomycin in a humidified atmosphere at 37 °C with 5% CO_2.

RNA extraction and qRT-PCR analysis

Total RNA from cells was extracted using TRIzol solution (Invitrogen, CA, United States). RNA was reverse-transcribed into complementary DNA (cDNA) by One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China). qRT-PCR was conducted with Top Green qPCR SuperMix (TransGen Biotech, Beijing, China) at 94°C for 30 s, followed by 40 cycles at 94°C for 5 s and at 60°C for 30 s. Gene-specific qRT-PCR primers were purchased from Invitrogen (Shanghai, China). GAPDH or U6 was used as an internal control. The relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

Western blotting

Gastric cancer cells were lysed in RIPA buffer (Solarbio, Beijing, China) containing protease inhibitor for 30 min at 4 °C. The protein concentration was determined by the BCA protein assay kit (Thermo Fisher Scientific, MA, United States) according to the manufacturer's instructions. The protein extracts were separated by 12% SDS-PAGE electrophoresis (30 μg/lane) and transferred to polyvinylidenedifluoride (PVDF) membranes (Millipore, MA, United States). The membranes were blocked with TBST containing 5% non-fat milk for 1 h and incubated overnight at 4 °C with primary antibodies against CAMK2N1 (1:1000, Invitrogen, MA, United States) and β-actin (1:1000, PTM BIO, Shanghai, China). The horseradish peroxidase (HRP)-conjugated secondary antibody (1:10,000, Invitrogen, MA, United States) was incubated at room temperature for 1 h. A chemiluminescence detection reagent (Thermo Fisher Scientific, MA, United States) was used to detect the signals.

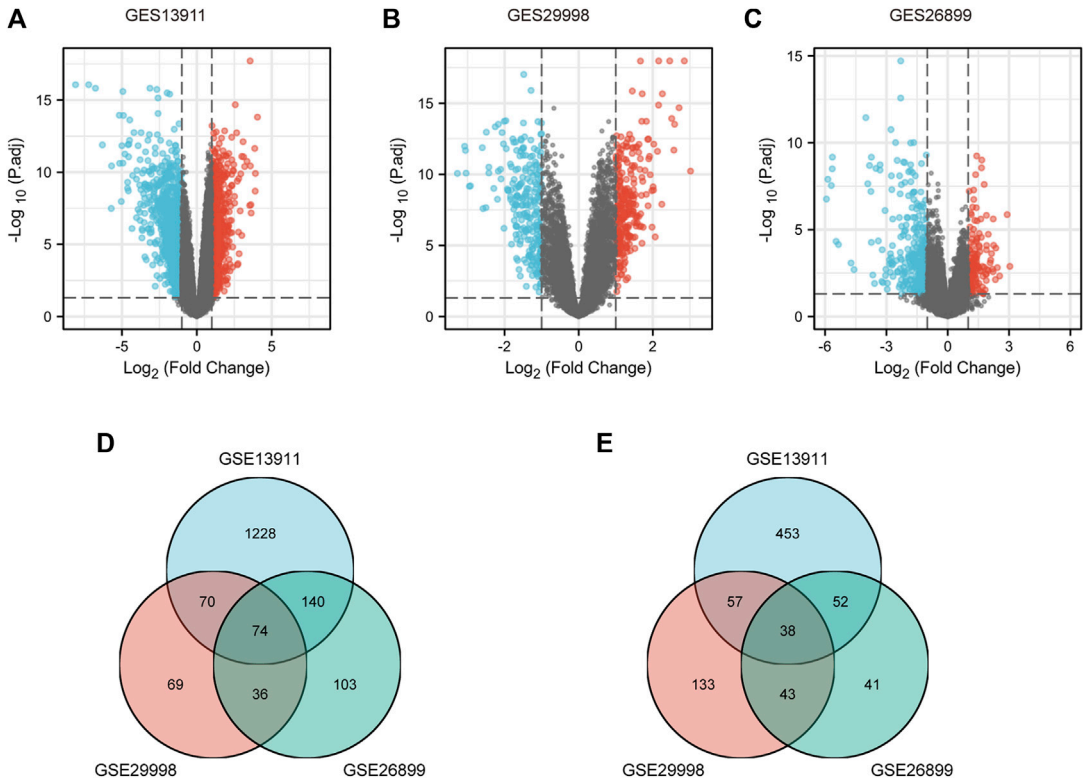


FIGURE 1
Identification of DEGs between gastric cancer and normal controls. **(A)** Volcano plots of gene expression profile data in GSE13911, containing 37 tumors tissues and 32 normal tissues. **(B)** Volcano plots of gene expression profile data in GSE29998, containing 49 tumors tissues and 50 normal tissues. **(C)** Volcano plots of gene expression profile data in GSE26899, containing 95 tumors tissues and 13 normal tissues. **(D)** Intersection of downregulated DEGs of GSE13911, GSE29998, and GSE26899. **(E)** Intersection of upregulated DEGs of GES19828, GSE13911, GSE29998, and GSE26899. DEG, differentially expressed gene.

The primer sequence

Please refer to Supplementary Table S1 for the details.

Statistical analysis

Statistical analysis was performed through GraphPad Prism (version 8, San Diego, CA). Student's t-tests were utilized for the comparison of two groups. Differences were considered statistically significant when $p < 0.05$.

Results

Screening of candidate genes related to GC

To identify the key genes associated with GC, GES13911, GES29998, and GES26899 datasets were selected to perform differential expression analysis using the Xiantao database. As described in Materials and methods, the cases of GSE13911, GSE29998, and GES26899 are, respectively, divided into two groups: "normal" group and "tumor" group. Differentially expressed genes (DEGs) of three GEO datasets were discovered as shown in Figure 1A, Figure 1B, and Figure 1C, respectively. This study aims to find the most potential genes associated with GC. Therefore, we obtained the upregulated and downregulated DEGs that appeared in three datasets. As shown in Figures 1D,E, 38 upregulated DEGs and 74 downregulated DEGs were finally identified. Finally, the 112 DEGs were defined as candidate genes and processed for the subsequent analysis.

Expression validation and survival analysis of candidate genes in GC

To verify the reliability of the analysis method, TCGA data and GTEx normal samples were used to validate the expression levels of 112 candidate genes by the GEPIA database. DEGs which had been reported as prognostic biomarkers in GC and

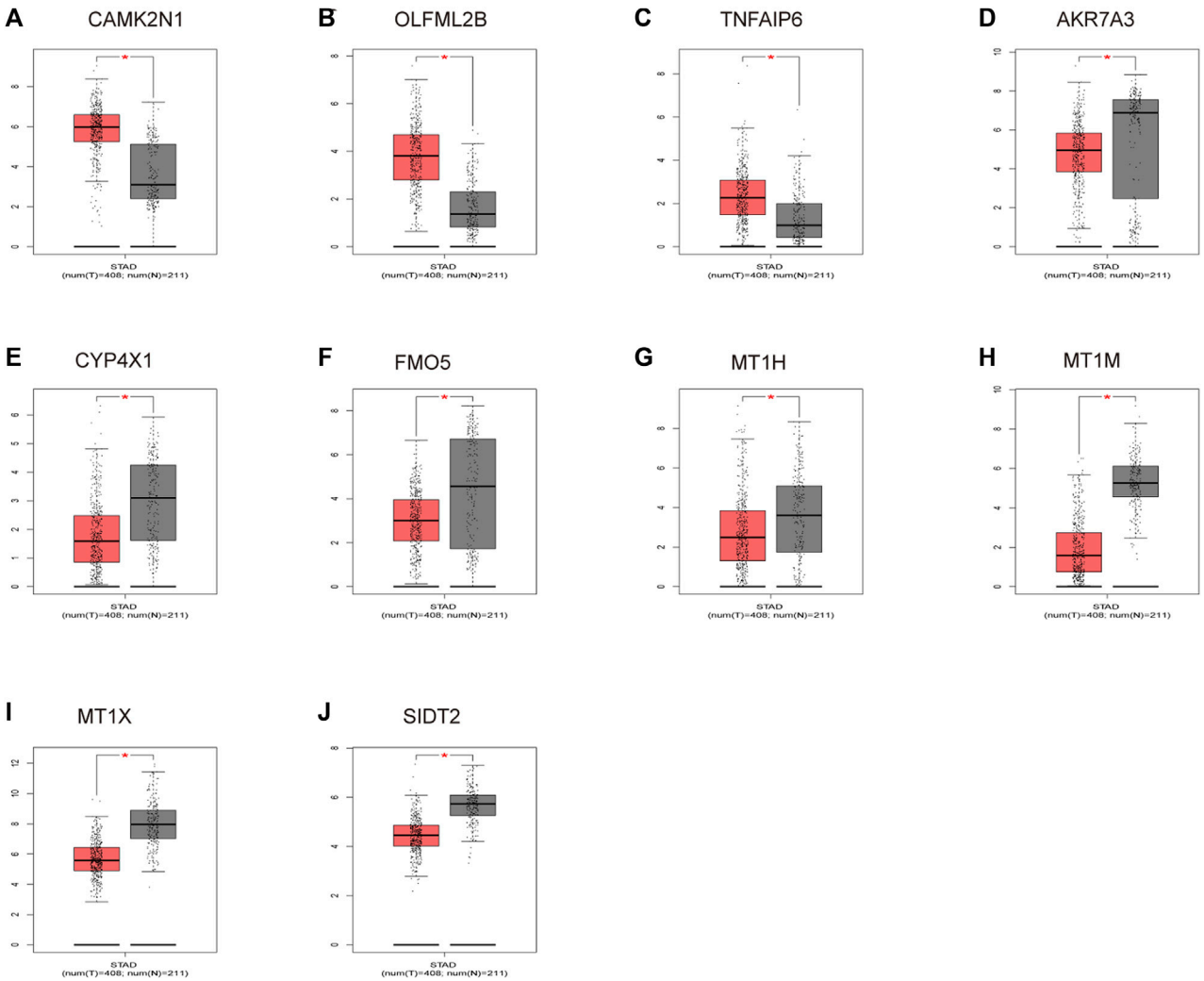


FIGURE 2
Expression levels of 10 key genes. **(A)** CAMK2N1. **(B)** OLFML2B. **(C)** TNFAIP6. **(D)** AKR7A3. **(E)** CYP4X1. **(F)** FMO5. **(G)** MT1H. **(H)** MT1M. **(I)** MT1X. **(J)** SIDT2. These genes were determined by the GEPIA database in gastric cancer. "*" represents a *p*-value < 0.05.

were not in accordance with the GEO expression levels were excluded. As shown in Figure 2, among the 112 candidate genes, the expression of 10 genes (CAMK2N1, OLFML2B, TNFAIP6, AKR7A3, CYP4X1, FMO5, MT1H, MT1M, MT1X, and SIDT2) was selected. In the following survival analysis, we focused on the 10 key genes. First, TCGA GC cohort was used to assess the prognostic values (including OS and RFS) of the 10 candidate genes by the Kaplan–Meier plotter database. As shown in Figure 3, high expression of CAMK2N1, OLFML2B, and SIDT2 indicated a poor OS, and high expression of TNFAIP6, AKR7A3, CYP4X1, FMO5, MT1H, MT1M, and MT1X forecasted a better OS. But, the OS of MT1M was not statistically significant, while the OS of TNFAIP6 and SIDT2 were not in accordance with expression levels in tumor tissues. For RFS, high expression of OLFML2B, MT1M, and SIDT2 indicated a poor RFS of patients with GC, whereas gastric cancer patients with higher expression of CYP4X1 possessed better RFS from TCGA cohort (Figure 4). Therefore, we conducted the next analysis on CAMK2N1, OLFML2B, AKR7A3, CYP4X1, FMO5, MT1H, and MT1X (Table 1).

Prediction and analysis of upstream miRNAs of seven key genes

The regulation of gene expression by ncRNAs has been widely recognized. To ascertain whether key genes were regulated by some ncRNAs, we predicted upstream miRNAs that could potentially bind to candidate genes and verified the correlation between miRNAs and candidate genes. To improve

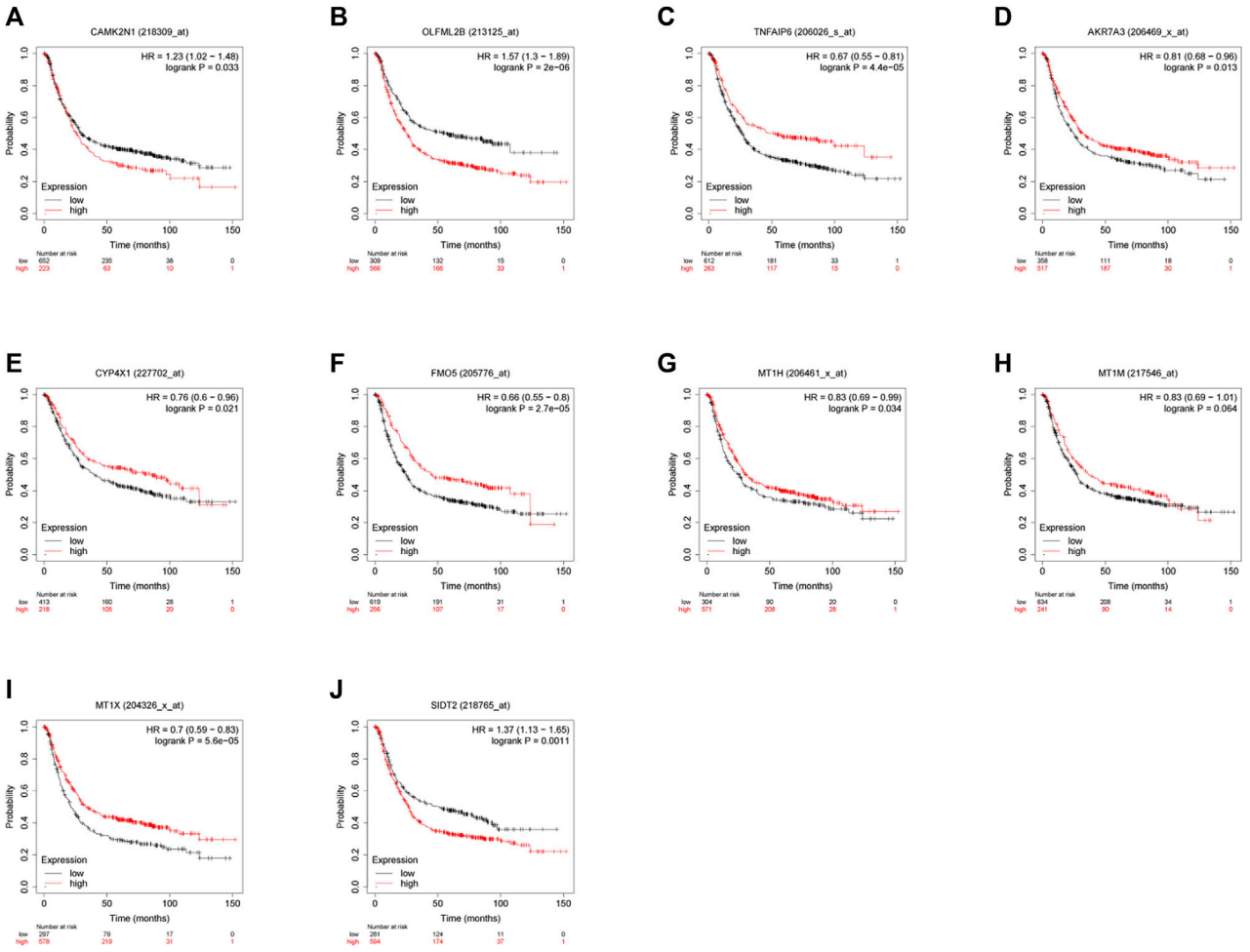
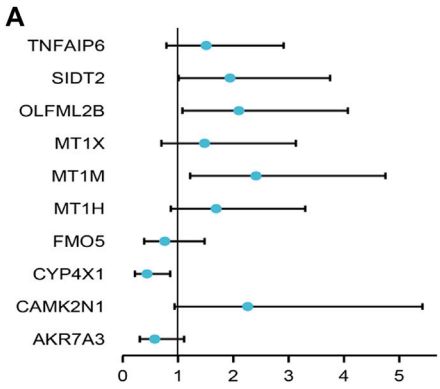


FIGURE 3
Prognostic values of 10 candidate genes in gastric cancer assessed using the Kaplan–Meier plotter database. **(A)** Prognostic value of CAMK2N1 in gastric cancer. **(B)** Prognostic value of OLFML2B in gastric cancer. **(C)** Prognostic value of TNFAIP6 in gastric cancer. **(D)** Prognostic value of AKR7A3 in gastric cancer. **(E)** Prognostic value of CYP4X1 in gastric cancer. **(F)** Prognostic value of FMO5 in gastric cancer. **(G)** Prognostic value of MT1H in gastric cancer. **(H)** Prognostic value of MT1M in gastric cancer. **(I)** Prognostic value of MT1X in gastric cancer. **(J)** Prognostic value of SIDT2 in gastric cancer. Log rank $p < 0.05$ was considered statistically significant.

visualization, a miRNA–mRNA regulatory network was established using Cytoscape software (Figures 5A,B). According to the mechanism of miRNA regulating the expression of target genes, there should be a negative correlation between miRNA and mRNA. As listed in Table 2, CAMK2N1 was significantly negatively correlated with hsa-miR-140-5p, hsa-miR-22-3p, hsa-miR-30e-5p, hsa-miR-138-5p, hsa-miR-18b-5p, and hsa-miR-378a-3p, and FMO5 was negatively correlated with hsa-miR-34c-5p and hsa-miR-526b-5p in GC. There were no statistical expression relationships or negative correlation between other mRNAs and the predicted miRNAs. Finally, the expression and prognostic value of candidate miRNAs in GC were determined by TCGA cohort. As presented in Figures 5C,D, hsa-miR-140-5p and hsa-miR-378a-3p were markedly downregulated in GC, and their upregulation was positively linked to patients' prognosis. hsa-miR-34c-5p was upregulated in GC, and its downregulation was positively linked to patients' prognosis. All these results suggest that hsa-miR-378a-3p–CAMK2N1 and hsa-miR-34c-5p–FMO5 could be the potential regulatory pathways.

Prediction and analysis of upstream lncRNAs of candidate miRNAs

The starBase database and miRNet database were utilized to predict the upstream potential lncRNAs that could potentially bind to the candidate miRNAs. As shown in Figure 6A,



B

Gene	HR	95%CI	logrank P
AKR7A3	0.58	0.31-1.11	0.098
CAMK2N1	2.26	0.94-5.42	0.061
CYP4X1	0.44	0.22-0.86	0.013
FMO5	0.76	0.39-1.48	0.42
MT1H	1.69	0.87-3.3	0.12
MT1M	2.41	1.22-4.75	0.0087
MT1X	1.48	0.7-3.13	0.31
OLFML2B	2.1	1.08-4.07	0.026
SIDT2	1.94	1.01-3.75	0.044
TNFAIP6	1.51	0.79-2.91	0.21

FIGURE 4
Prognostic values of 10 candidate genes in gastric cancer assessed using the Kaplan–Meier plotter database. **(A)** Prognostic values (relapse-free survival, RFS) of the 10 key genes in TCGA gastric cancer cohort. **(B)** Detailed information of **(A)**. Log rank $p < 0.05$ was considered statistically significant.

16 lncRNAs were obtained as the ncRNAs that bind to hsa-miR-378a-3p. According to the competing endogenous RNA (ceRNA) hypothesis, lncRNA could increase mRNA expression by competitively binding to shared miRNAs. Therefore, lncRNA should be positively correlated with mRNA (Figure 7). The expression correlation between the four lncRNAs and miRNAs or mRNAs in GC was also determined by the starBase database. Then, we determined the expression levels and prognostic values of these lncRNAs in TCGA GC cohort (Figures 6B,C). Of all the predicted lncRNAs, only SNHG10 can meet all the conditions. Taking all expression analysis, survival analysis, and correlation analysis into consideration, SNHG10/hsa-miR-378a-3p/CAMK2N1 might be the most potential regulatory axis in GC.

CAMK2N1 co-expression network in GC

The results of the co-expression pattern of CAMK2N1 are shown in Supplementary Figure S1A. Heat maps displayed the top 50 genes positively and the top 50 genes negatively associated with CAMK2N1 (Supplementary Figures S1B,C). GO term annotation showed that co-expressed genes of CAMK2N1 participate mainly in embryonic skeletal system development, epidermis development, isoprenoid metabolic process, organic cation transport, mesenchymal cell proliferation, modified amino acid transport, and columnar/cuboidal epithelial cell differentiation, etc. (Supplementary Figure S1D). The KEGG pathway analysis indicated enrichment in steroid hormone biosynthesis, glycerolipid metabolism, protein digestion and absorption, synaptic vesicle cycle, tight junction, and linoleic acid metabolism (Supplementary Figure S1E).

Validation of the expression level of CAMK2N1, SNHG10, and hsa-miR-378a-3p in GC cell lines

To further validate the role of CAMK2N1, SNHG10, and hsa-miR-378a-3p in GC, we assessed the expression of mRNA and protein in gastric cancer cell lines. As shown in Figure 8, there revealed significantly over-expression of CAMK2N1 in four gastric cancer cell lines (MGC-803, BGC-823, HGC-27, and MKN-45) as compared to the normal gastric mucosal cell line GES-1 (Figures 8A,B). We found that SNHG10 was highly expressed in gastric cancer lines, and the miR-378a-3p was lowly expressed in BGC-823, HGC-27, and MKN-45.

CAMK2N1 negatively correlates with immune cell infiltration in GC

Calcium/calmodulin-dependent protein kinase II inhibitor alpha (CAMK2N1) is an endogenous cellular inhibitor of CAMK2. Recent studies revealed that the expression level of CAMK2N1 is also involved in the tumorigenesis of human cancers. As shown in Figure 9A, a significant change in immune cell infiltration level under various copy numbers of CAMK2N1 in GC was observed. Correlation analysis could provide key clues for studying the function and mechanism of CAMK2N1. Thus, the correlation of the CAMK2N1 expression level with the immune cell infiltration level was assessed. As shown in Figure 9B, CAMK2N1 expression was significantly negatively associated with all analyzed immune cells, including B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells, in GC. This suggests that the overexpression

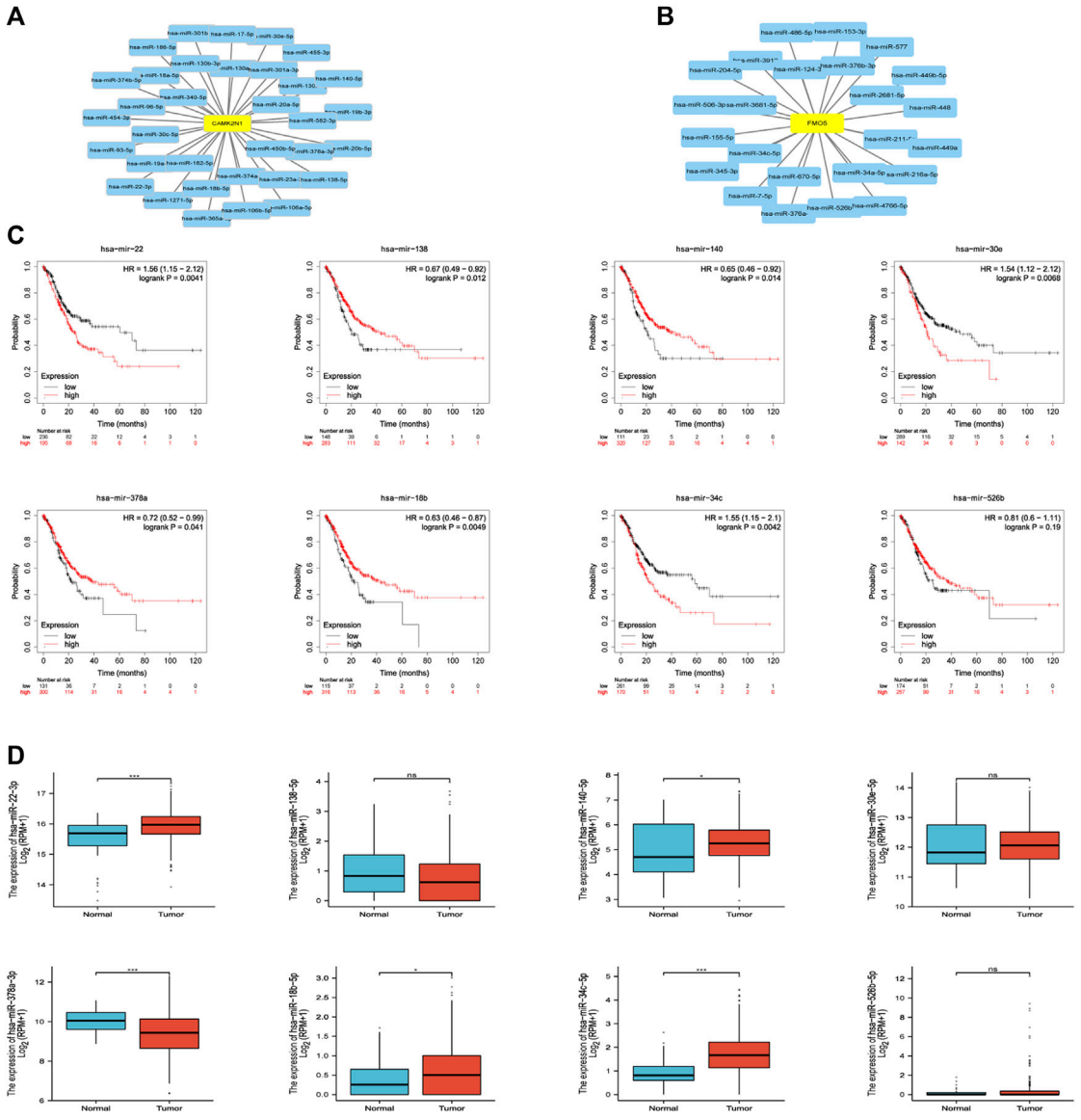


FIGURE 5
(A) miRNA–CAMK2N1 network established by Cytoscape software. **(B)** miRNA–FMO5 network established by Cytoscape software. **(C)** Prognostic values (overall survival, OS) of potential upstream miRNAs in TCGA gastric cancer cohort. **(D)** Expression of potential upstream miRNAs in TCGA gastric cancer cohort. "*" represents a p-value < 0.05; "***" represents a p-value < 0.01.

of CAMK2N1 may be accompanied by the decline of immune cell infiltration in GC.

Expression correlation of CAMK2N1 and biomarkers of immune cells in GC

To further explore the role of CAMK2N1 in tumor immune, we evaluated the expression correlation of CAMK2N1 with biomarkers of immune cells in GC using the GEPIA database. As listed in Table 3, CAMK2N1 was significantly negatively correlated with B-cell's biomarkers (CD19 and CD79A), $CD4^+$-T cell's biomarkers (CD4), $CD8^+$ T-cell's biomarkers (CD8A and CD8B), M2 macrophage's biomarkers (CD163, VSIG4, and MS4A4A), neutrophil's biomarkers (ITGAM and CCR7), and dendritic cell biomarkers (HLA-DPB1, HLA-DRA, HLA-DPA1, CD1C, and ITGAX) in GC. These findings partly support the hypothesis that CAMK2N1 is negatively linked to immune cell infiltration.

TABLE 2 Prediction of miRNAs binding to CAMK2N1 and FMO5.

Gene name	miRNA name	Predicting program	Number	R	P
CAMK2N1	hsa-miR-17-5p	PITA, miRmap, miRanda, PicTar, and TargetScan	5	0.012	8.15E-01
CAMK2N1	hsa-miR-18a-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	−0.011	8.38E-01
CAMK2N1	hsa-miR-19a-3p	PITA, microT, and miRanda	3	−0.031	5.54E-01
CAMK2N1	hsa-miR-19b-3p	PITA, microT, and miRanda	3	0.02	7.03E-01
CAMK2N1	hsa-miR-20a-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	−0.009	8.64E-01
CAMK2N1	hsa-miR-22-3p	PITA, miRanda, PicTar, and TargetScan	4	0.146	4.74E-03
CAMK2N1	hsa-miR-23a-3p	PITA, miRanda, and TargetScan	3	0.029	5.74E-01
CAMK2N1	hsa-miR-93-5p	PITA, miRmap, miRanda, PicTar, and TargetScan	5	0.016	7.62E-01
CAMK2N1	hsa-miR-96-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	0.026	6.16E-01
CAMK2N1	hsa-miR-106a-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	−0.092	7.48E-02
CAMK2N1	hsa-miR-30c-5p	PITA, microT, miRanda, PicTar, and TargetScan	5	0.028	5.92E-01
CAMK2N1	hsa-miR-182-5p	PITA, microT, miRanda, PicTar, and TargetScan	5	0.013	8.04E-01
CAMK2N1	hsa-miR-138-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	−0.169	1.03E-03
CAMK2N1	hsa-miR-140-5p	PITA, miRmap, miRanda, PicTar, and TargetScan	5	−0.145	5.12E-03
CAMK2N1	hsa-miR-106b-5p	PITA, miRmap, microT, miRanda, PicTar, and TargetScan	6	0.04	4.38E-01
CAMK2N1	hsa-miR-30e-5p	PITA, microT, miRanda, and TargetScan	4	−0.185	3.42E-04
CAMK2N1	hsa-miR-378a-3p	PITA and miRanda	2	−0.122	1.83E-02
CAMK2N1	hsa-miR-18b-5p	PITA, miRmap, microT, miRanda, and TargetScan	5	−0.107	3.95E-02
CAMK2N1	hsa-miR-20b-5p	PITA, miRmap, miRanda, PicTar, and TargetScan	5	−0.05	3.38E-01
CAMK2N1	hsa-miR-455-3p	PITA, miRmap, microT, and TargetScan	4	−0.061	2.44E-01
CAMK2N1	hsa-miR-450b-5p	PITA, miRmap, and microT	3	−0.098	6.00E-02
CAMK2N1	hsa-miR-1271-5p	PITA, miRmap, microT, miRanda, and TargetScan	5	−0.092	7.68E-02
FMO5	hsa-miR-34a-5p	miRmap and miRanda	2	0.041	4.26E-01
FMO5	hsa-miR-153-3p	miRmap and miRanda	2	0.284	2.53E-08
FMO5	hsa-miR-34c-5p	miRmap and miRanda	2	−0.103	4.61E-02
FMO5	hsa-miR-448	miRmap and miRanda	2	−0.003	9.55E-01
FMO5	hsa-miR-449a	miRmap and miRanda	2	0.037	4.77E-01
FMO5	hsa-miR-526b-5p	miRmap and microT	2	−-0.148	4.24E-03
FMO5	hsa-miR-449b-5p	miRmap and miRanda	2	0.058	2.68E-01

Correlation between CAMK2N1 and immune checkpoints in GC

PD1/PD-L1 and CTLA-4 are important immune checkpoints that are responsible for tumor immune escape. In order to explore the potential tumorigenic role of CAMK2N1 in GC, we assessed the relationship of CAMK2N1 with PD1, PD-L1, or CTLA-4. As shown in Figures 10A–C, CAMK2N1 expression was significantly negatively correlated with PD1, PD-L1, and CTLA-4 in GC, which was adjusted by purity. To further validate these results, we used the GEPIA database to examine the relationship between CAMK2N1 and immune checkpoints. We also found that there was a significant negative correlation of CAMK2N1 with PD1, PD-L1, or CTLA-4 in GC (Figures 10D–F). These results indicate that tumor immune escape might be involved in the CAMK2N1-mediated carcinogenesis of GC.

Discussion

Until now, GC is still notorious for its poor prognosis and aggressive characteristics. Though great developments have been made in diagnosis and therapy for GC, the prognosis of patients is still perishing. Exploring novel molecular mechanisms of GC carcinogenesis may provide assistance to the development of effective therapeutic targets or seek valuable and helpful prognostic biomarkers. Increasing evidence has demonstrated that CAMK2N1 plays a key role in the proliferation and progression of multiple human cancers (Wang et al., 2014b; Li et al., 2018; Xu et al., 2019; Peng et al., 2021). However, the role of CAMK2N1 in GC is still unclear, and further studies are needed.

In this study, we first conducted comprehensive expression analysis and survival analysis based on TCGA cohort and GEO microarray. Then, seven key genes (CAMK2N1, OLFML2B, AKR7A3, CYP4X1, FMO5, MT1H, and MT1X) were identified which may be associated with the prognosis of GC.

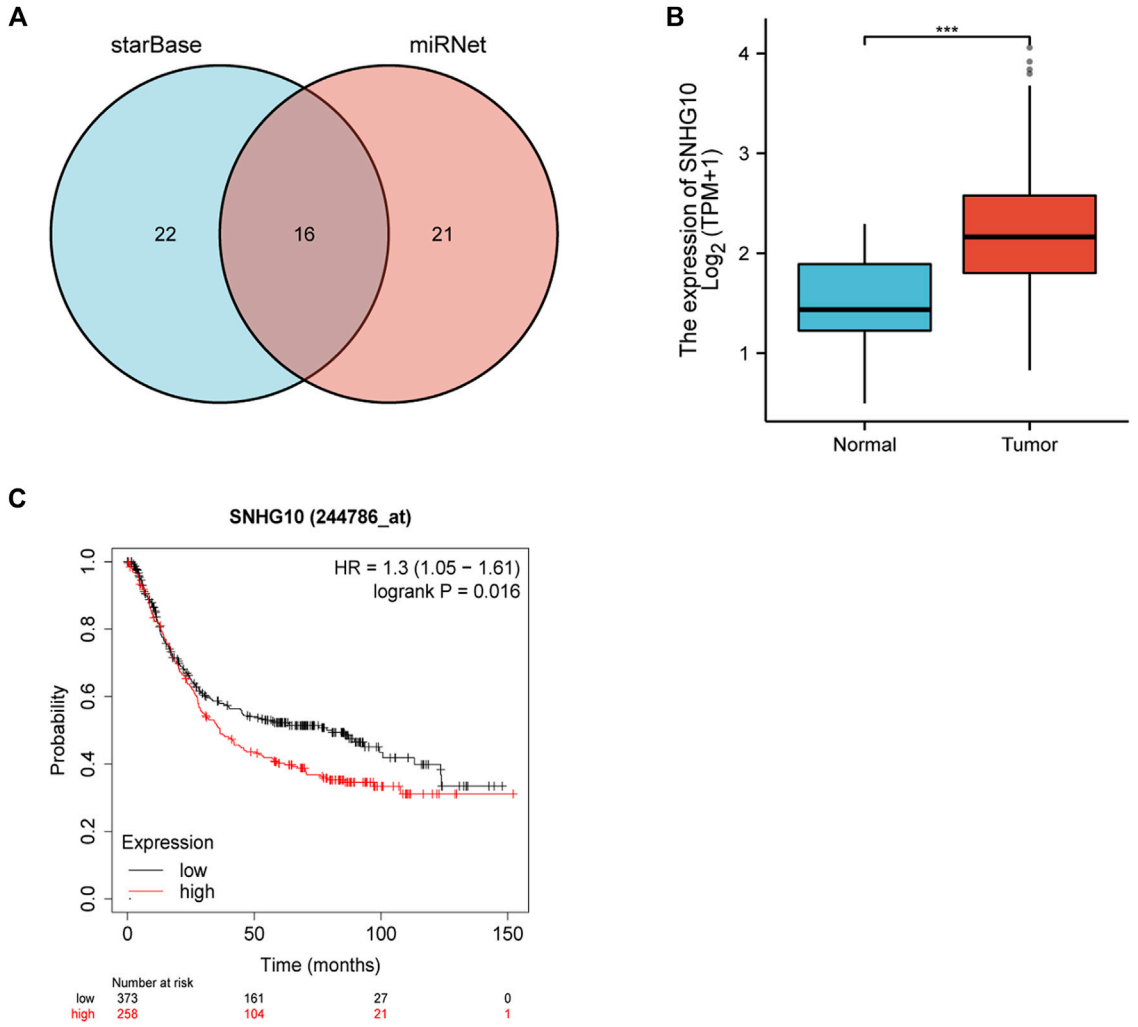


FIGURE 6
(A) Intersection of potential lncRNAs of hsa-miR-378-3p from starBase and miRNet databases. **(B)** Expression of SNHG10 in TCGA gastric cancer cohort. **(C)** Prognostic values (overall survival, OS) of SNHG10 in TCGA gastric cancer cohort.

Furthermore, we established a ceRNA network based on seven key genes.

It has been widely reported that ncRNAs (miRNAs, lncRNAs, and circular RNAs) can play important roles in regulating biological behaviors by ceRNA mechanisms (Gao et al., 2020; Ghafouri-Fard et al., 2020; Razavi et al., 2021). We used the starBase database by seven prediction programs, involving PITA, RNA22, miRmap, microT, miRanda, PicTar, and TargetScan, to predict possible upstream regulatory miRNAs that could potentially bind to key genes. MiRNAs play a role in suppressing target genes by binding to the mRNA of target genes. Therefore, the miRNAs should be negatively related to mRNA, and eight miRNAs were finally obtained, including six miRNAs for CAMK2N1 and two miRNAs for FMO5. In addition, we take into consideration the prognostic significance and expression of predicted upstream regulatory miRNAs. According to the mechanism of miRNA, these miRNAs which bind to CAMK2N1 should be tumor-suppressive miRNAs, and these miRNAs which bind to FMO5 are tumor-promoting. We conducted this analysis in TCGA cohort using the Xiantao database and Kaplan–Meier plotter database. At the end, two miRNA–mRNA pairs (hsa-miR-378a-3p–CAMK2N1 and hsa-miR-34c-5p–FMO5) owned the most potential functions in GC and were selected for subsequent analysis.

According to the ceRNA hypothesis, the potential lncRNAs competitively binding to common miRNAs were found. The potential lncRNAs of hsa-miR-378a-3p–CAMK2N1 and hsa-miR-34c-5p-–FMO5 were first

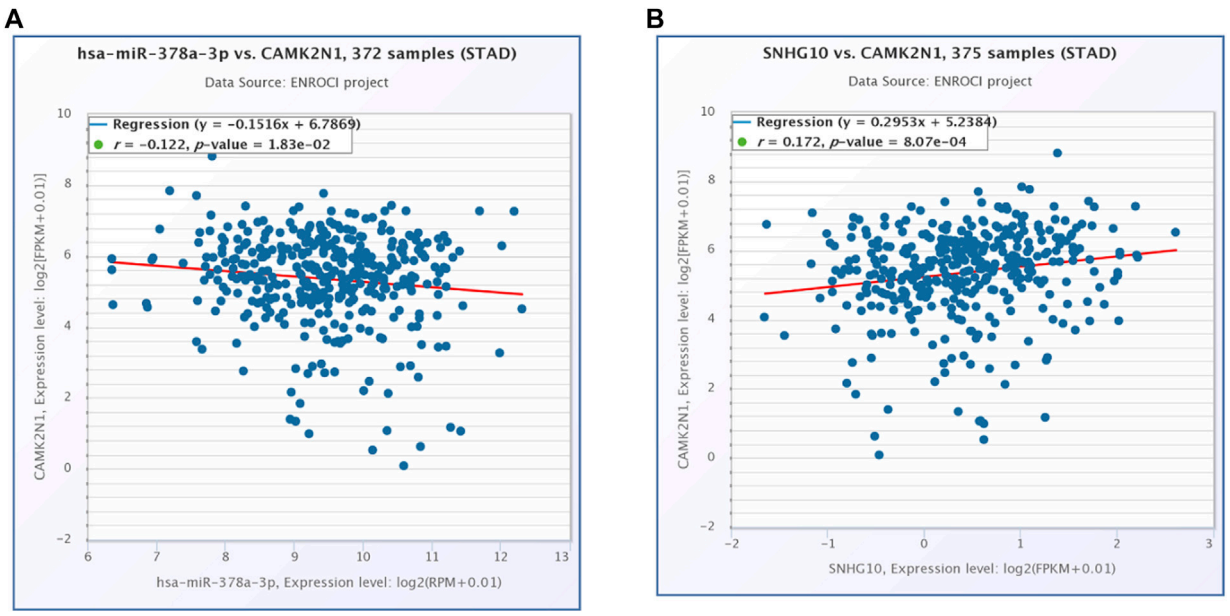


FIGURE 7
(A) Expression correlation of hsa-miR-378a-3p and CAMK2N1 in gastric cancer. **(B)** Expression correlation of CAMK2N1 and SNHG10 in gastric cancer.

obtained by the starBase database. By conducting expression analysis, survival analysis, and correlation analysis, SNHG10 was found to be the most potential upregulated lncRNA for CAMK2N1. Also, we did not find suitable lncRNAs to construct a ceRNA network with FOM5. SNHG10 has been reported to function as an oncogene in multiple cancers, including GC. For instance, SNHG10 (Yuan et al., 2021; Zhang et al., 2021) enhanced cell proliferation and migration in GC. Taking together, SNHG10/hsa-miR-378a-3p/CAMK2N1 might be the most potential regulatory axis in GC.

CAMK2N1 is an endogenous cellular inhibitor of CAMK2 which is located on chromosomal 1 p1. Recent studies revealed that the expression level of CAMK2N1 is also involved in the tumorigenesis of human cancers. CAMK2N1 is commonly used as a tumor suppressor in some cancers, such as prostate cancer and hepatocellular carcinoma. It can inhibit E2F1 to regulate the cell cycle to affect the proliferation of hepatocellular carcinoma and regulate the proliferation of prostate cancer by suppressing ErbB2 (Wang et al., 2014b; Peng et al., 2021). But we found that CAMK2N1 may play a carcinogenic role in GC using qRT-PCR and Western blotting.

In order to verify the accuracy of the aforementioned bioinformatics analysis results, the expression of CAMK2N1, hsa-miR-378a-3p, and SNHG10 were verified in five GC cell lines (AGS, MGC-803, BGC-823, HGC-27, and MKN-45) by RT-qPCR and Western blotting. We found that the expression of CAMK2N1 and SNHG10 in GC cell lines were higher than that in the normal gastric mucosal cell line GES-1, and the expression of hsa-miR-378a-3p in GC cell lines was lower than that in GES-1.

A large number of studies (Li et al., 2016; Liu et al., 2017; Zhang et al., 2018; Dieci et al., 2021) have confirmed that tumor immune cell infiltration can affect the efficacy and prognosis of chemotherapy, radiotherapy, or immunotherapy in tumor patients. Our work indicated that CAMK2N1 was observably negatively correlated with various immune cells, including B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells in GC. Moreover, CAMK2N1 was also notably negatively associated with biomarkers of these infiltrated immune cells. Our findings suggest that tumor immune infiltration may be partly responsible for the carcinogenic effect of CAMK2N1-mediated GC.

An immune checkpoint (Sharma and Allison, 2015; Wei et al., 2018; Kalbasi and Ribas, 2020) is a set of molecules expressed on immune cells that regulate the level of immune activation and play an important role in preventing autoimmunity (an abnormal immune function that attacks normal cells) from occurring. Abnormal expression and function of immune checkpoint molecules is one of the

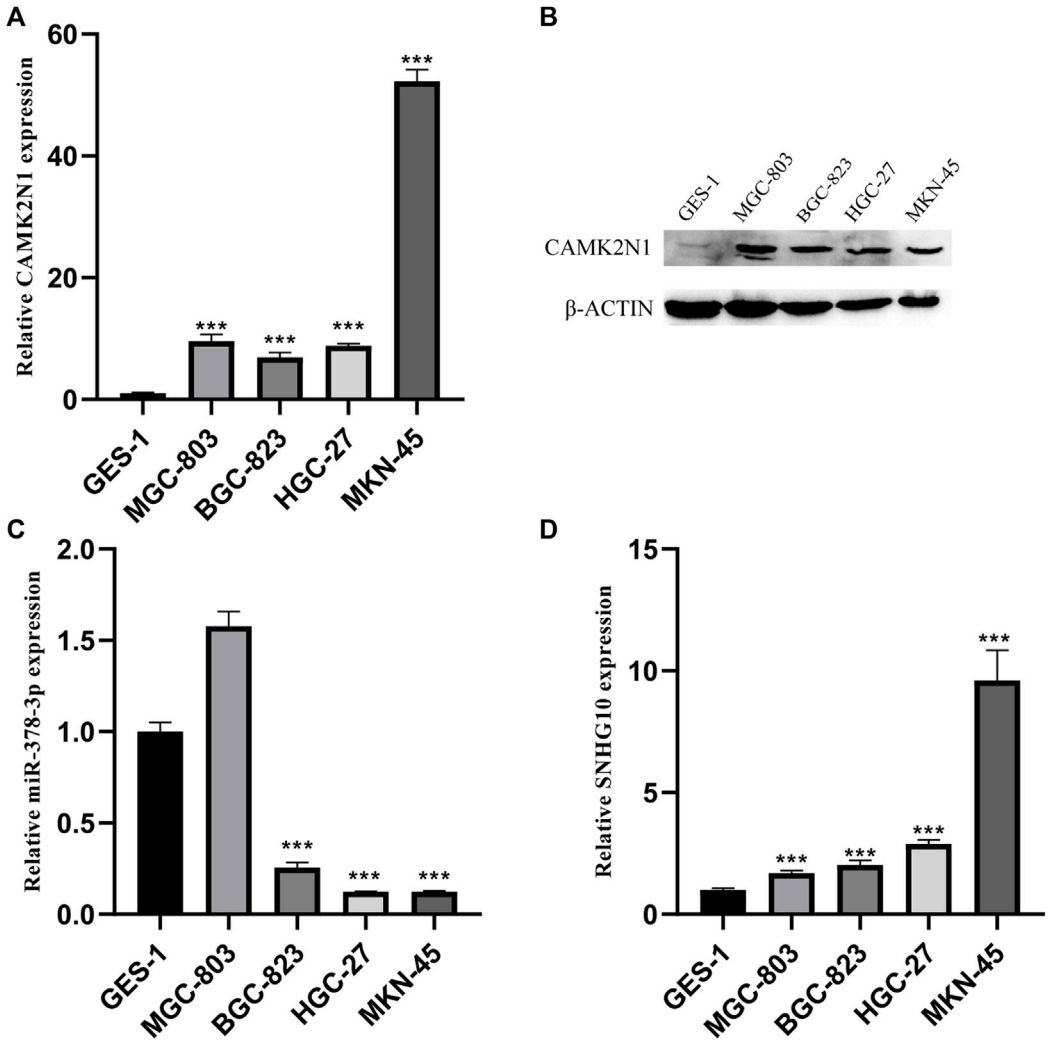


FIGURE 8
Expression levels of CAMK2N1, hsa-miR-378-3p, and SNHG10 in GC cell lines. **(A)** mRNA expression of CAMK2N1 in GC cell lines. **(B)** Protein expression of CAMK2N1 in GC cell lines. **(C)** Expression of hsa-miR-378a-3p in GC cell lines. **(D)** Expression of SNHG10 in GC cell lines.

important reasons for the occurrence of many diseases, such as overexpression or over-function of immune checkpoint molecules, in which immune function is suppressed, the body's immunity is low, and people are prone to cancer and other diseases. Conversely, if the immunosuppressive function of immune checkpoint molecules is poor, the immune function of the body will be abnormal. Efficacy of immunotherapy also depends on the sufficient expression of immune checkpoints in addition to adequate immune cell infiltration (de Miguel and Calvo, 2020; Morad et al., 2021). Therefore, we assessed the relationship between CAMK2N1 and immune checkpoints. The results demonstrated that high expression of CAMK2N1 was strongly linked to PD1, PD-L1, and CTLA-4 in GC, suggesting that targeting CAMK2N1 might increase the efficacy of immunotherapy in GC.

Conclusion

In summary, we elucidated that CAMK2N1 was highly expressed in GC and positively correlated with unfavorable prognosis in GC. We predicted an upstream regulatory mechanism of CAMK2N1 in GC, which was the SNHG10/hsa-miR-378a-3p/CAMK2N1 axis (Figure 11). In addition, our current findings also indicated that CAMK2N1 might play its oncogenic roles through decreasing tumor immune cell infiltration and immune checkpoint expression. However,

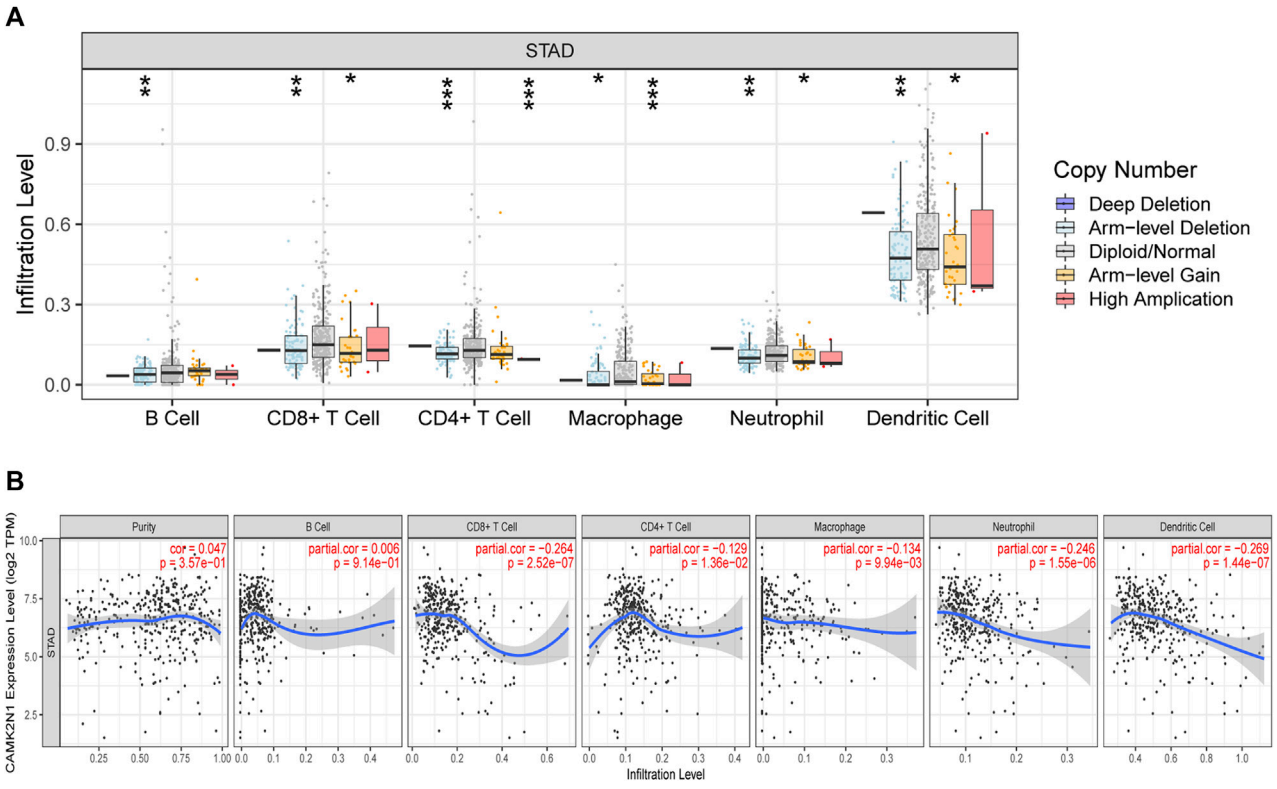


FIGURE 9
Relationship of immune cell infiltration with CAMK2N1 in gastric cancer. **(A)** Infiltration level of various immune cells under different copy numbers of CAMK2N1 in gastric cancer. **(B)** Correlation of the CAMK2N1 expression level with B cell, $CD8^+$ T cell, $CD4^+$ T cell, macrophage, neutrophil, or dendritic cell infiltration level in gastric cancer.

TABLE 3 Correlation analysis between CAMK2N1 and biomarkers of immune cells in gastric cancer determined using the GEPIA database.

Immune cell	Biomarker	P	R
B cell	CD19	2.20E-04	−0.18
	CD79A	9.4E−08	−0.26
$CD8^+$T cell	CD8A	1.5E−08	−0.28
	CD8B	1.5E−05	−0.21
$CD4^+$T cell	CD4	2.5E−06	−0.23
M1 macrophage	NOS2	2.70E-01	0.055
	IRF5	4.60E-01	0.036
	PTGS2	3.10E-01	0.051
M2 macrophage	CD163	6.10E-04	−0.17
	VSIG4	1.20E-03	−0.16
	MS4A4A	2.6E−06	−0.23
Neutrophil	ITGAM	1.00E-03	−0.16
	CCR7	6.5E−09	−0.28
Dendritic cell	HLA-DPB1	6.6E−08	−0.26
	HLA-DQB1	5.90E-02	−0.094
	HLA-DRA	5.4E−07	−0.25
	CD1C	2.4E−06	−0.23
	NRP1	2.50E-01	−0.057
	ITGAX	6.90E-03	−0.13

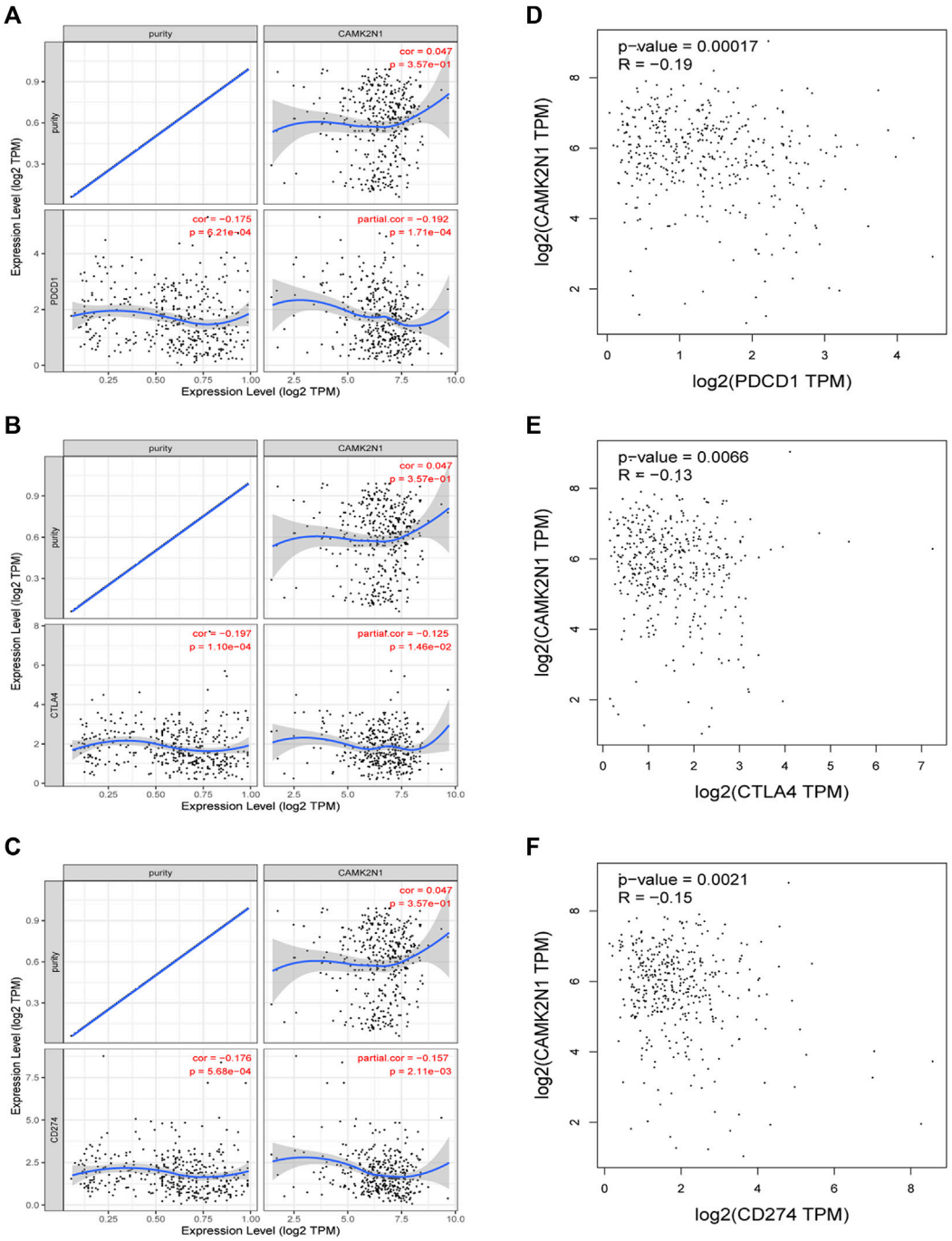


FIGURE 10
Correlation of CAMK2N1 with PDCD1, CTLA-4, and CD274 expression in gastric cancer. **(A)** Spearman correlation of CAMK2N1 with the expression of PDCD1 in gastric cancer adjusted by the TIMER database. **(B)** Spearman correlation of CAMK2N1 with the expression of CTLA-4 in gastric cancer adjusted by the TIMER database. **(C)** Spearman correlation of CAMK2N1 with the expression of CD274 in gastric cancer adjusted by the TIMER database. **(D)** Expression correlation of CAMK2N1 with PDCD1 in gastric cancer determined by the GEPIA database. **(E)** Expression correlation of CAMK2N1 with CTLA-4 in gastric cancer determined by the GEPIA database. **(F)** Expression correlation of CAMK2N1 with CD274 in gastric cancer determined by the GEPIA database.

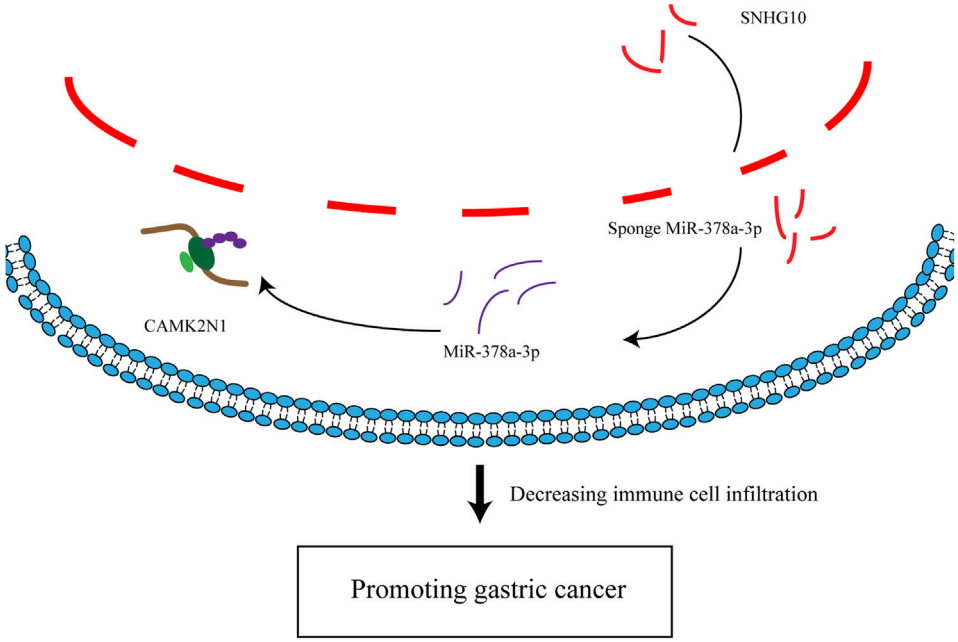


FIGURE 11
Model of the mechanism axis in carcinogenesis of GC.

these results should be validated by many more basic experiments and large clinical trials in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization, KP and QR; methodology, KP; software, KP and XR; formal analysis, KP and XR; investigation, KP; writing—original draft preparation, KP; writing—review and editing, KP; visualization, KP and XR; supervision, QR; and funding acquisition, QR. All authors have read and agreed to the published version of the manuscript.

Funding

This study was supported by the Health Industry Research Project of Gansu Province (GSWSKY 2020-07), the Fundamental Research Funds for the Central Universities (lzujbky-2021-ct17), and the Natural Science Foundation of Gansu Province (21JR7RA381).

Acknowledgments

The authors gratefully acknowledge the support of TCGA and the GEO database.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.888672/full#supplementary-material

References

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18 (1), 5–18. doi:10.1038/nrc.2017.99

de Miguel, M., and Calvo, E. (2020). Clinical challenges of immune checkpoint inhibitors. *Cancer Cell* 38 (3), 326–333. doi:10.1016/j.ccell.2020.07.004

Dieci, M. V., Miglietta, F., and Guarneri, V. (2021). Immune infiltrates in breast cancer: recent updates and clinical implications. *Cells* 10 (2), 223. doi:10.3390/cells10020223

Digklia, A., and Wagner, A. D. (2016). Advanced gastric cancer: current treatment landscape and future perspectives. *World J. Gastroenterol.* 22 (8), 2403–2414. doi:10.3748/wjg.v22.i8.2403

Ding, X., Jia, X., Wang, C., Xu, J., Gao, S. J., and Lu, C. (2019). A DHX9-lncRNA-MDM2 interaction regulates cell invasion and angiogenesis of cervical cancer. *Cell Death Differ.* 26 (9), 1750–1765. doi:10.1038/s41418-018-0242-0

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2 (12), 919–929. doi:10.1038/35103511

Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12 (12), 861–874. doi:10.1038/nrg3074

Fu, D. G. (2015). Epigenetic alterations in gastric cancer (Review). *Mol. Med. Rep.* 12 (3), 3223–3230. doi:10.3892/mmr.2015.3816

Gao, S., Ding, B., and Lou, W. (2020). microRNA-dependent modulation of genes contributes to ESR1's effect on ERα positive breast cancer. *Front. Oncol.* 10, 753. doi:10.3389/fonc.2020.00753

Ghafouri-Fard, S., Shoorei, H., Anamag, F. T., and Taheri, M. (2020). The role of non-coding RNAs in controlling cell cycle related proteins in cancer cells. *Front. Oncol.* 10, 608975. doi:10.3389/fonc.2020.608975

Kalbasi, A., and Ribas, A. (2020). Tumour-intrinsic resistance to immune checkpoint blockade. *Nat. Rev. Immunol.* 20 (1), 25–39. doi:10.1038/s41577-019-0218-4

Kanwal, F., and Lu, C. (2019). A review on native and denaturing purification methods for non-coding RNA (ncRNA). *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1120, 71–79. doi:10.1016/j.jchromb.2019.04.034

Karreth, F. A., and Pandolfi, P. P. (2013). ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* 3 (10), 1113–1121. doi:10.1158/2159-8290.CD-13-0202

Khorana, A. A., Dalal, M., Lin, J., and Connolly, G. C. (2013). Incidence and predictors of venous thromboembolism (VTE) among ambulatory high-risk cancer patients undergoing chemotherapy in the United States. *Cancer* 119 (3), 648–655. doi:10.1002/cncr.27772

Komatsu, S., and Otsuji, E. (2019). Essential updates 2017/2018: recent topics in the treatment and research of gastric cancer in Japan. *Ann. Gastroenterol. Surg.* 3 (6), 581–591. doi:10.1002/ags3.12284

Li, B., Severson, E., Pignon, J. C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 17 (1), 174. doi:10.1186/s13059-016-1028-7

Li, N., Nan, C. C., Zhong, X. Y., Weng, J. Q., Fan, H. D., Sun, H. P., et al. (2018). miR-182-5p promotes growth in oral squamous cell carcinoma by inhibiting CAMK2N1. *Cell. Physiol. biochem.* 49 (4), 1329–1341. doi:10.1159/000493411

Liu, X., Wu, S., Yang, Y., Zhao, M., Zhu, G., and Hou, Z. (2017). The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. *Biomed. Pharmacother.* 95, 55–61. doi:10.1016/j.biopha.2017.08.003

Morad, G., Helmink, B. A., Sharma, P., and Wargo, J. A. (2021). Hallmarks of response, resistance, and toxicity to immune checkpoint blockade. *Cell* 184 (21), 5309–5337. doi:10.1016/j.cell.2021.09.020

Peng, J. M., Tseng, R. H., Shih, T. C., and Hsieh, S. Y. (2021). CAMK2N1 suppresses hepatoma growth through inhibiting E2F1-mediated cell-cycle signaling. *Cancer Lett.* 497, 66–76. doi:10.1016/j.canlet.2020.10.017

Qi, P., Zhou, X. Y., and Du, X. (2016). Circulating long non-coding RNAs in cancer: current status and future perspectives. *Mol. Cancer* 15 (1), 39. doi:10.1186/s12943-016-0524-4

Razavi, Z. S., Tajiknia, V., Majidi, S., Ghandali, M., Mirzaei, H. R., Rahimian, N., et al. (2021). Gynecologic cancers and non-coding RNAs: epigenetic regulators with emerging roles. *Crit. Rev. Oncol. Hematol.* 157, 103192. doi:10.1016/j.critrevonc.2020.103192

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell* 146 (3), 353–358. doi:10.1016/j.cell.2011.07.014

Sharma, P., and Allison, J. P. (2015). The future of immune checkpoint therapy. *Science* 348 (6230), 56–61. doi:10.1126/science.aaa8172

Smyth, E. C., Verheij, M., Allum, W., Cunningham, D., Cervantes, A., Arnold, D., et al. (2016). Gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 27 (5), v38–v49. doi:10.1093/annonc/mdw350

Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., et al. (2014). The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* 344 (6181), 310–313. doi:10.1126/science.1251456

Wang, Q. X., Zhu, Y. Q., Zhang, H., and Xiao, J. (2015). Altered MiRNA expression in gastric cancer: a systematic review and meta-analysis. *Cell. Physiol. biochem.* 35 (3), 933–944. doi:10.1159/000369750

Wang, S., Liang, K., Hu, Q., Li, P., Song, J., Yang, Y., et al. (2017). JAK2-binding long noncoding RNA promotes breast cancer brain metastasis. *J. Clin. Investig.* 127 (12), 4498–4515. doi:10.1172/JCI91553

Wang, T., Guo, S., Liu, Z., Wu, L., Li, M., Yang, J., et al. (2014). CAMK2N1 inhibits prostate cancer progression through androgen receptor-dependent signaling. *Oncotarget* 5 (21), 10293–10306. doi:10.18632/oncotarget.2511

Wei, H., Pu, K., Liu, X. G., Li, B. X., Zhang, H. S., Wang, H., et al. (2019). The diagnostic value of circulating microRNAs as a biomarker for gastric cancer: a meta-analysis. *Oncol. Rep.* 41 (1), 87–102. doi:10.3892/or.2018.6782

Wei, S. C., Duffy, C. R., and Allison, J. P. (2018). Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discov.* 8 (9), 1069–1086. doi:10.1158/2159-8290.CD-18-0367

Xu, K., Hu, X., Sun, L., Liang, Q., Ouyang, G., Zhang, Y., et al. (2019). MicroRNA-532 exerts oncogenic functions in t(4;14) multiple myeloma by targeting CAMK2N1. *Hum. Cell* 32 (4), 529–539. doi:10.1007/s13577-019-00276-y

Yuan, X., Yang, T., Xu, Y., Ou, S., Shi, P., Cao, M., et al. (2021). SNHG10 promotes cell proliferation and migration in gastric cancer by targeting miR-495-3p/CTNNB1 Axis. *Dig. Dis. Sci.* 66 (8), 2627–2636. doi:10.1007/s10620-020-06576-w

Zhang, H., Liu, H., Shen, Z., Lin, C., Wang, X., Qin, J., et al. (2018). Tumor-infiltrating neutrophils is prognostic and predictive for postoperative adjuvant chemotherapy benefit in patients with gastric cancer. *Ann. Surg.* 267 (2), 311–318. doi:10.1097/SLA.0000000000002058

Zhang, K., Shi, H., Xi, H., Wu, X., Cui, J., Gao, Y., et al. (2017). Genome-wide lncRNA microarray profiling identifies novel circulating lncRNAs for detection of gastric cancer. *Theranostics* 7 (1), 213–227. doi:10.7150/thno.16044

Zhang, Y., Guo, H., and Zhang, H. (2021). SNHG10/DDX54/PBX3 feedback loop contributes to gastric cancer cell growth. *Dig. Dis. Sci.* 66 (6), 1875–1884. doi:10.1007/s10620-020-06488-9

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fgene.2022.865111

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Juan Manuel Mejia-Arangure, Universidad Nacional Autonoma de Mexico, Mexico
Eduardo Andres-Leon, Institute of Parasitology and Biomedicine López-Neyra (CSIC), Spain

*CORRESPONDENCE
Xuchun Che,
chexuchun@tmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 29 January 2022
ACCEPTED 08 August 2022
PUBLISHED 06 September 2022

CITATION
Wang Q, Yue C, Liu Q and Che X (2022), Exploration of differentially expressed mRNAs and miRNAs for pediatric acute myeloid leukemia.
Front. Genet. 13:865111.
doi: 10.3389/fgene.2022.865111

COPYRIGHT
© 2022 Wang, Yue, Liu and Che. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Exploration of differentially expressed mRNAs and miRNAs for pediatric acute myeloid leukemia

Qing Wang[1†], Chao Yue[2†], Qin Liu[3†] and Xuchun Che[4*]

[1]Department of Clinical Laboratory, The General Hospital of Tianjin Medical University, Tianjin, China, [2]Department of Blood Transfusion, The Fifth Central Hospital of Tianjin, Tianjin, China, [3]Department of Clinical Laboratory, The Fifth Central Hospital of Tianjin, Tianjin, China, [4]Department of Immunology, Tianjin Key Laboratory of Cellular and Molecular Immunology, Key Laboratory of Educational Ministry of China, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

Background: To establish a comprehensive differential gene profile for pediatric acute myeloid leukemia patients (pAML) based on two independent databases and verify the differentially expressed genes using *in vitro* and *in vivo* analyses.

Methods: The mRNA and miRNA sequencing information of GSE2191 and GSE35320, clinically recruited pAML individuals, and human AML cell line (NB4 cells) were utilized in the study.

Results: Compared with the control sample, pAML patients demonstrated a total of 778 differentially expressed genes, including 565 upregulated genes and 213 downregulated genes. The genes including *ZC3H15*, *BCLAF1*, *PPIG*, *DNTTIP2*, *SRSF11*, *KTN1*, *UBE3A*, *PRPF40A*, *TMED5*, and *GNL2* were the top 10 potential hub genes. At the same time, 12 miRNAs demonstrated remarkable differential expressions in pAML individuals compared with control individuals, as five upregulated and seven downregulated miRNAs. The *hsa-miR-133*, *hsa-miR-181*, and *hsa-miR-195* were significantly downregulated. Building a miRNA–mRNA regulatory network, *hsa-miR-133* regulated *ZC3H15*, *BCLAF1*, *SRSF11*, *KTN1*, *PRPF40A*, and *GNL2*. Using the NB4 cell model, *hsa-miR-133* treatment inhibited cell proliferation capacity, which could be attenuated by a single mRNA transfection or a combination of *ZC3H15* and *BCLAF1*. At the same time, *hsa-miR-133* mimic treatment could significantly accelerate cell apoptosis in NB4 cells, which was also *ZC3H15*- and *BCLAF1*-dependent. The concentrations of ZC3H15 and BCLAF1 were investigated in peripheral blood using the ELISA method for the clinical control and pAML samples. In pAML samples, the expression levels of ZC3H15 and BCLAF1 were significantly enhanced ($p < 0.01$), regardless of the classification.

Conclusion: Collectively, this study hypothesized several promising candidates for pAML formation.

KEYWORDS

pediatric acute myeloid leukemia, differential gene analysis, hsa-miR-133, ZC3H15, BCLAF1

Introduction

Acute myeloid leukemia (AML) is the most common malignancy of acute leukemia in humans, comprising approximately 80 percent of the cases (Short et al., 2018). Compared with adults, the incidence rate in children is relatively rare, but with disproportionate mortality (Elgarten and Aplenc, 2020). As recently reported, the pediatric acute myeloid leukemia (pAML) accounts for 25% of pediatric leukemia with a low overall survival ratio of 70% (Taga et al., 2016). Due to the decades of study, the treatments for pAML has been greatly improved by advances in the hematopoietic stem cell transplantation (HSCT) method, chemotherapy, supportive care, and optimal risk stratification (Tarlock and Meshinchi, 2015; Paulraj et al., 2019). Based on the various pAML subtypes, the clinical pAML procedures are conducted separately. For instance, children with *de novo* AML are administrated with a standard treatment, which includes four or five cycles of myelosuppressive chemotherapy with cytarabine and anthracyclines followed by a HSCT for a subset of patients. At the same time, children with acute promyelocytic leukemia (APL) are recommended with an all-trans retinoic acid (ATRA)–combined regimen. In contrast, myeloid leukemia children with Down syndrome (ML-DS) are generally converged with a less intensive regimen (Ribeiro, 2014).

The worldwide incidence of pAML is heterogeneous because of the variable prevalence of the risk factors. As reported, host factors such as age, race, and germline predisposition impact outcomes contribute to a large amount of pAML (Radhi et al., 2010; Shimada, 2017). In addition to the host factors, the elevated level of peripheral white blood cells (WBCs) has been also suggested as an unfavorable aspect for pAML (Woods et al., 1996). For adult AML, growing evidence based on retrospective assessments supports the association between the disease and the abnormal proliferation and differentiation of a clonal population of myeloid stem cells (De Kouchkovsky and Abdul-Hay, 2016). Except for the chromosomal rearrangements, the particular genetic mutations play a key role in the formation of AML, causing more than 97% of cases (Patel et al., 2012). Compared with adult AML, the pathological mechanism underlying pAML is still poorly understood and the molecular landscape behind pAML remains quite distinct (Bolouri et al., 2018). The treatments for pAML have evolved with years of work. If detected at an early stage, the pAML could be administrated with various choices of procedures. Unfortunately, for the majority of patients, pAML is diagnosed at a later stage so that the long-term prognosis remains unsatisfactory.

MicroRNAs (miRNAs) are emerging as a promising candidate for the molecular mechanism behind pAML, which attracts attention of many researchers. Previously, a study by Obulkasim et al. (2017)suggested that the biological subgroups of pAML were reflected by a common miRNA expression pattern, while the separate subtypes of pAML have distinct miRNA expression patterns. The miRNAs belong to a class of small endogenous RNAs that regulate gene expression post-transcriptionally and play a role in gene silencing and translation inhibition by binding to target genes. The miRNAs represent a highly conserved class of tissue-specific genes that have been found in all eukaryotic cells preserved across species since their discovery in 1993 (Vishnoi and Rani, 2017; Lu and Rothenberg, 2018). Generally speaking, they are short RNA molecules with 19–25 nucleotides in size. A single miRNA can target hundreds of mRNAs and influence the expression of many genes. The appropriate maintenance of miRNA expression is required for a balanced physiological environment.

Since the underlying molecular mechanism of pAML remains unclear, the disorder has brought great difficulty to clinical treatment. To address these issues, in this study, we used a combination of bioinformatics analysis and external clinical specimen experiments to establish differentially expressed gene profiling. Meanwhile, the potential signaling axis for pAML formation had also been explored. All of these promising outcomes enriched the precise early diagnosis of the disease, which provided tremendous help for the pAML study.

Materials and methods

Data source and differentially expressed gene analysis

The information on mRNA sequencing, miRNA sequencing as well as clinical data of children with AML were downloaded from the GEO database (GEO, https://www.ncbi.nlm.nih.gov/geo/).

The array chip numbering GSE2191 was established for pAML differentially expressed mRNA analysis. The database contains bone marrow samples from 54 pAML to 4 control individuals. All the specimens were used for analysis in the study. At the same time, the array chip numbering GSE35320 was utilized for pAML differentially expressed miRNA analysis. The database included the miRNA expression profiles of different AML subtypes from 102 pediatric patients in comparison to CD34+ cells from healthy donors and adult AML patients. In order to identify differentially expressed miRNAs for pAML, miRNA expression profiling of 110 pediatric AML patients was conducted for differentially expressed miRNA examination compared with 5 NB4, isotype control-–detected miRNAs.

The mRNA and miRNA expression analyses were performed by the genetic analysis group of HPS. Co (Tianjin, China). The information of selected samples was tested using the Affymetrix Human Genome U95 Version 2 Array platform.

For the original data measured by the chip, the RMA method was used to normalize first, and then the log2 logarithm of the normalized value was taken to generate the standardized data for downstream analysis. The limma package in R language (version:3.5.2) was developed to analyze the differentially expressed mRNAs as well miRNAs between different groups, taking the absolute value of the log-transformed differential expression multiple (Log2FC) > 1 and $p < 0.05$ as a standard for analysis.

Moreover, the final p-value was corrected using the Bonferroni method.

Functional enrichment analysis

For the obtained differentially expressed genes, we used the "clusterProfiler" function package in R language for enrichment analysis of GO (including biological process, molecular function, and cellular component) and KEGG pathway. When p-value < 0.05, we considered the corresponding entries to be significantly enriched (Ritchie et al., 2015).

Prediction of microRNA target genes

The target genes of miRNAs were predicted through the miRDB (http://mirdb.org/index.html, version 6.0) database (Yu et al., 2012). Meanwhile, the Cytoscape (https://cytoscape.org/, version 3.7.2) was performed to visualize the miRNA–mRNA regulatory network.

Cell culture

The human AML cell line (NB4 cell line) was purchased from Procell Co., China (Cat. NO: CL-0676) and maintained in a RPMI-1640 medium with L-glutamine/10% fetal bovine serum/1X PSF.

The *hsa-miR-133* mimics and corresponding negative control (NC siRNA) were constructed by Genewiz Corporation Co. (Tianjin, China). Lipofectamine™ 3000 was used for plasmid transfection.

Cell proliferation and apoptosis measurement

The cells were collected and measured for their proliferation capacity using the CCK-8 kit (Fisher, China). The absorbance was evaluated at 450 nm using the plate reader purchased from Thermo Fisher Scientific Co. At the same time, the cell apoptosis capacity was examined by flow cytometry after Annexin-V FITC/PI double staining. All experiments were conducted three times independently.

Sample collection

This study was a retrospective study using clinical recruited pAML patients from General Hospital of Tianjin Medical University. The 110 pAML patients were selected based on the following inclusion criteria: 1) age from 0 to 18 years; 2) abnormal expression of leukemia, a significant increase in lymphocyte, varying degrees of anemia based on blood routine tests; 3) increased expression level of proliferating leukemia cells for bone marrow puncture; 4) positive for other auxiliary diagnostic schemes. The exclusion criteria: 1) presence of other malignancy or 2) other circulatory system disorders. A total of 28 control children were recruited in the control group, which were excluded from pAML, other malignancy or circulatory system disorders. Based on the standard of FAB classification (French–American–British classification), the pAML individuals were further sub-grouped into M0–M7 classifications (M0: 13; M1: 9; M2: 25; M3: 11; M4: 8; M5: 37; M6: 7 without M7).

This study was in line with the medical ethics standards, and approved by the hospital ethics committee. All treatment and testing were performed with informed consent of patients or their families. For the pAML group, there were 28 males and 20 females, and the average age was 9.92 ± 1.13 years. For the control group, there were 16 males and 12 females, and the average age was 9.64 ± 1.48 years. There was no significant difference between two groups ($p > 0.05$).

ELISA analysis of hub genes

The protein concentrations of ZC3H15 and BCLAF1 were determined using the ELISA double antibody sandwich method from recruited individuals' peripheral blood. The specific operation was carried out in strict accordance with the instructions of the kit (Abcam company, United States). The experimental results were repeated three times independently and were tested by statistical methods.

Statistics

Excel 2022 was established for data analysis. The continuous variables were tested for normal distribution and continuous variables are presented by the mean ± standard deviation (x ± s).

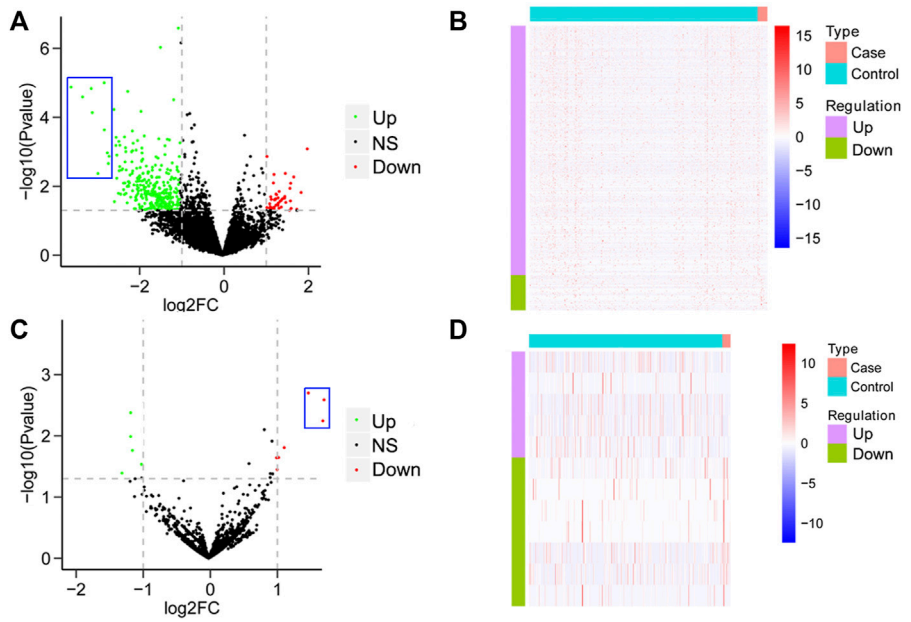


FIGURE 1
Analysis of differential mRNAs and miRNAs comparing pAML and control groups. **(A)** Volcano map of differentially expressed mRNAs between two groups. The horizontal axis represents the multiple differential expression (Log2FC), the vertical axis represents −log10 (FDR). The green dots indicate upregulated genes, and the red dots indicate downregulated genes, respectively. **(B)** Heat map of differentially expressed mRNAs. The horizontal axis represents the sample, vertical axis represents different genes. **(C)** Volcano map of differentially expressed miRNAs between two groups. **(D)** Heat map of differentially expressed miRNAs.

The student *t*-test was used for data comparison between two groups, with the $p < 0.05$ considered as a significant difference.

Results

Differentially expressed mRNA and miRNA analysis results

To establish a differential gene profiling for pAML, we first analyzed the mRNA information for pAML patients' specimens on GSE2191. Compared with the control sample, pAML patients' samples exhibited a total of 778 differentially expressed genes, including 565 upregulated genes and 213 downregulated genes (Figures 1A,B, Supplementary Table S1). The genes such as *ZC3H15*, *BCLAF1*, *PPIG*, *DNTTIP2*, *SRSF11*, *KTN1*, *UBE3A*, *PRPF40A*, *TMED5*, and *GNL2* were the top 10 potential hub genes (most significant difference among the two groups). It is interesting that all the primary hub genes were upregulated in pAML (indicated as blue rectangle in Figure 1A). At the same time, based on the investigation of array chip numbering GSE35320, 12 miRNAs demonstrated remarkable differential expressions in pAML individuals compared with control individuals, as five upregulated and seven downregulated miRNAs (Figures 1C,D; Supplementary Table S2). The *hsa-miR-133*, *hsa-miR-181*, and *hsa-miR-195* were significantly downregulated(indicated as blue rectangle in Figure 1C).

GO and KEGG enrichment analysis results

By performing GO and KEGG enrichment analyses on these 778 differentially expressed genes, we found these differentially expressed genes were in GO terms related to various biological processes such as immunoglobulin complex and blood microparticles. (Figure 2A). In addition, the NF-κB pathway, cell proliferation, and cell adhesion–related signaling pathways were significantly enriched in KEGG pathway analysis (Figure 2B).

miRNA–mRNA regulatory network

The three selected miRNAs (*hsa-miR-133*, *hsa-miR-181*, and *hsa-miR-195*) as well as 10 potential mRNAs (*ZC3H15*, *BCLAF1*, *PPIG*, *DNTTIP2*, *SRSF11*, *KTN1*, *UBE3A*, *PRPF40A*, *TMED5*, and *GNL2*) were further investigated for a regulatory network construction visualized using Cytoscape software. As shown in Figure 3, the three miRNAs were functionally

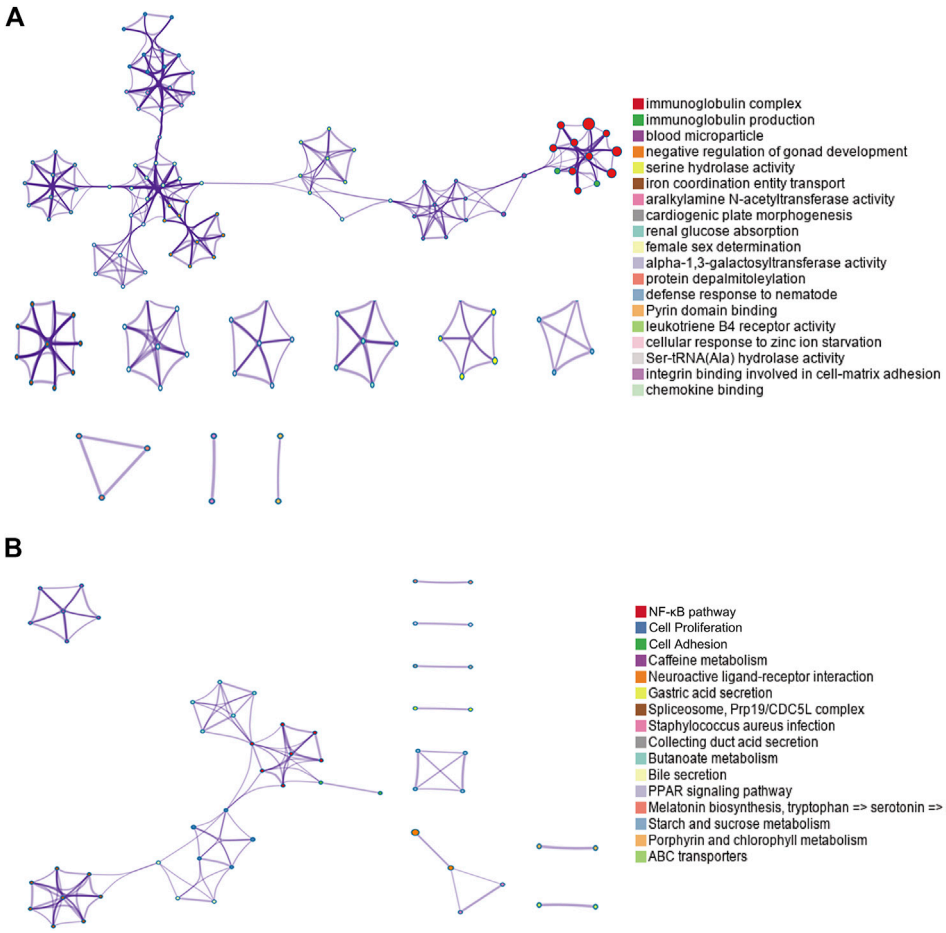


FIGURE 2
GO and KEGG enrichment results for differentially expressed genes. **(A)** Top GO term enrichment results with the largest number of genes. In the figure, the horizontal axis represents the number of enriched genes, and the vertical axis represents the name of each GO term, respectively. **(B)** Enrichment results of the KEGG pathways with the largest number of genes. The horizontal axis in the figure indicates the number of genes enriched, and the vertical axis indicates the name of each KEGG pathway, respectively.

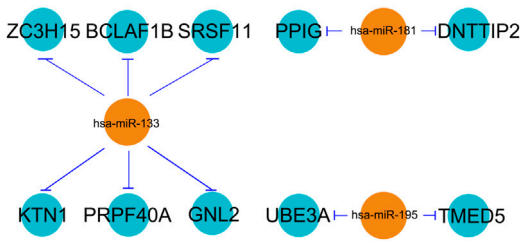


FIGURE 3
miRNA–mRNA regulatory network. The line type represents the interaction between miRNA and mRNA.

associated with multiple hub genes. Among which, *hsa-miR-133* regulated the largest number of target genes (6), which were *ZC3H15*, *BCLAF1*, *SRSF11*, *KTN1*, *PRPF40A*, and *GNL2*. At the same time, *hsa-miR-181* interacted with two hub genes, which were *PPIG* and *DNTTIP2*. Meanwhile, the functions of *UBE3A* and *TMED5* were manipulated by *hsa-miR-195*.

Verification of the differentially expressed genes using the *in vitro* cell model

ZC3H15 and *BCLAF1* were shown to be most significantly enhanced for pAML patients by differential gene analysis (Figure 1A). At the same time, both of the genes were demonstrated to be direct targets of *hsa-miR-133* by the miRNA–mRNA regulatory network results. Next, we sought to verify this using the *in vitro* cell model, which was the human AML cell line (NB4 cell line). The introduction of *hsa-miR-133* in NB4 cells obviously inhibited cell proliferation capacity, which could be

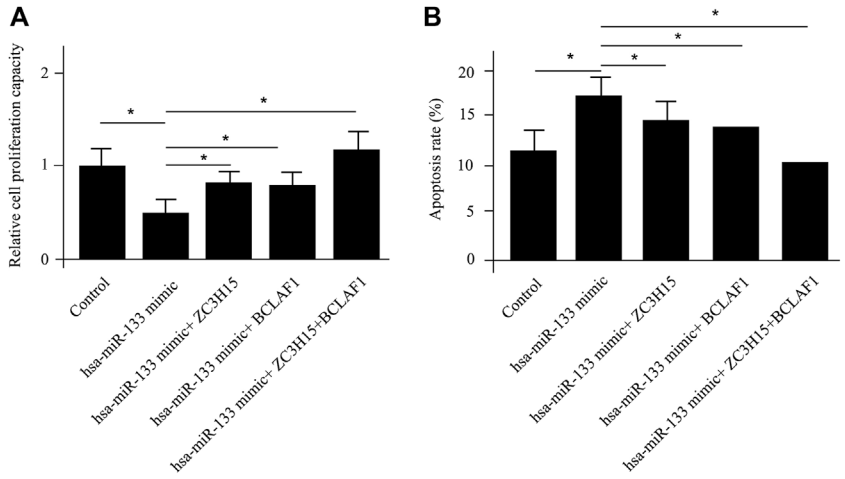


FIGURE 4
In vitro verification of differentially expressed mRNAs and miRNAs. The cell proliferation capacity **(A)** and cell apoptosis analysis **(B)** for NB4 cells with different treatments.

attenuated by a single mRNA transfection or a combination of *ZC3H15* and *BCLAF1* (Figure 4A). Moreover, *hsa-miR-133* mimic treatment could significantly accelerate cell apoptosis in NB4 cells, which was also *ZC3H15*- and *BCLAF1*-dependent (Figure 4B). These outcomes suggested that the increased expressions of *ZC3H15* and *BCLAF1* and the decreased expression of *hsa-miR-133* were hallmarks for pAML. Meanwhile, *ZC3H15* and *BCLAF1* were direct targets for *hsa-miR-133* to modulate AML cellular functions.

ZC3H15 and BCLAF1 were significantly differentially expressed for pAML *in vivo*

So far, *ZC3H15* and *BCLAF1* were suggested to be key regulatory factors for pAML progression, which were closely associated with cell proliferation and apoptosis. However, the pAML patients were sub-grouped into different classifications based on FAB criteria. The mRNA information of the array chip did not provide the inter-group special expression levels for the two key genes. In order to approve that the differential expression of mRNA was independent of FAB classification, we included pAML patients with different FAB classifications. To further verify the functions of hub genes in pAML patients, the protein expression levels of top two hub genes (*ZC3H15* and *BCLAF1*) were compared between clinical recruited control individuals and pAML individuals. The concentrations of ZC3H15 and BCLAF1 were investigated in peripheral blood using the ELISA method for the two groups.

For control samples, the protein concentration of ZC3H15 was 28.33 ± 5.51 ng/ml, ranging 21.3–35.4 ng/ml; while the protein concentration of BCLAF1 was 12.33 ± 3.39 ng/ml, ranging 9.2–19.4 ng/ml. On the other hand, in pAML samples (including all the FAB classifications), the expression levels of ZC3H15 and BCLAF1 were significantly enhanced ($p < 0.01$). The protein concentration of ZC3H15 was 45.68 ± 6.46 ng/ml, ranging 27.6–55.6 ng/ml; while the protein concentration of BCLAF1 was 22.17 ± 5.26 ng/ml, ranging 14.6–33.8 ng/ml. Furthermore, there were existing remarkable differences between control individuals and pAML patients of all the FAB groups (except for M7, which was not recruited in our study for the low sample size, as shown in Table 1). These outcomes were in line with the differential expression analysis results.

TABLE 1 Protein concentration of potential hub genes between control and different pAML groups.

Group	N	ZC3H15 (ng/ml, x ± s)	BCLAF1 (ng/ml, x ± s)
Control	28	28.33 ± 5.51*	12.33 ± 3.39*
pAML (total)	110	45.68 ± 6.46*	22.17 ± 5.26*
pAML (M0)	13	52.18 ± 6.35*	26.17 ± 5.08*
pAML (M1)	9	40.22 ± 5.73*	20.15 ± 4.29*
pAML (M2)	25	39.68 ± 5.33*	18.39 ± 4.21*
pAML (M3)	11	38.59 ± 6.12*	18.53 ± 3.88*
pAML (M4)	8	40.17 ± 6.31*	22.43 ± 5.40*
pAML (M5)	37	49.38 ± 7.02*	26.84 ± 6.25*
pAML (M6)	7	48.11 ± 7.11*	24.21 ± 6.32*

*as an indication of $p < 0.05$ compared with the control group.

Discussion

Even with years of hard work for the breakthroughs in pAML, the prevention, early screening, diagnosis, treatments, and overall outcomes are still inferior, highlighting the need for improved targeted therapies. In this study, we comprehensively compared differentially expressed mRNAs as well as miRNAs between pAML and control groups based on two independent databases. In this integrated study, the 10 mRNAs and three miRNAs were hypothesized as key regulators for pAML formation. Previously, more and more attention had been paid to the interaction between the signal transducer and activator of transcription 3 (STAT3) and AML in adults. Increases in cytokine ligands, such as IL-6, trigger intracellular tyrosine phosphorylation of STAT3, which is seen in up to 50% of AML cases and signifies a worse prognosis (Schuringa et al., 2000). In addition to STAT3, NPM1 and CEBPA confer to notable class II mutations while DNA-methylation–related genes DNMT3A, TET2, and IDH-1 and IDH-2 contribute to the class II mutation group for AML, which are found in about 27% and more than 40% of AML cases, respectively (Takahashi, 2011; Klein et al., 2018). However, the majority of these mutations were not significantly expressed in the pAML group specimens from our analysis, which might reflect a tremendous difference between pediatrics and adults for AML initiation as well as development. The DNA methylation events in somatic mutations are suggested to be highly prevalent in adults. Inversely, the structural alterations in methyltransferase genes are more prevalent in young children, but rarer or even absent in adults. These may explain the differences for the AML between children and adults considering differentially expressed gene profiles.

Previously, the connection between miRNA and leukemogenesis was built on the cell proliferation as well as cell adhesion regulation by particular miRNA families (Mott et al., 2007; Volinia et al., 2010). These are in dire need of better understanding of the functions of miRNA in pAML since most of the work focuses on the adult AML (Danen-van Oorschot et al., 2012). A study by Zhang et al. (2009) demonstrated the miR-expression differences between FAB-M1, FAB-M2, and FAB-M3 groups in pAML. Another study suggested the specific expression pattern of miR-99a, miR-125b, and let-7c in pAML and the upregulations of these factors stimulated leukemogenesis by switching the balance between TGF-ß and Wnt signaling pathways (Bonifant et al., 2018). In a large cohort, Obulkasim and his colleagues analyzed 665 miRNAs on 165 pediatric AML samples, which provided 14 key clusters for the differentially expressed miRNAs (Lu and Rothenberg, 2018). In this study, they claimed pAML samples with MLL rearrangements were classified with 89% accuracy using 37 miRNA expression signatures; 14 miRNAs were highly expressed and 23 were lowly expressed compared to the samples without MLL rearrangements. In consistent with our findings here, *hsa-miR-133*, *hsa-miR-181*, and *hsa-miR-195* were strikingly repressed in the pAML samples. Moreover, they supposed that *hsa-miR-181* and *hsa-miR-195* were functional as tumor suppressors while *hsa-miR-133* was only used to classify MLL-rearranged samples in pAML.

Except for miRNA analysis, we also hypothesized the key mRNAs for pAML, including *ZC3H15*, *BCLAF1*, *PPIG*, *DNTTIP2*, *SRSF11*, *KTN1*, *UBE3A*, *PRPF40A*, *TMED5*, and *GNL2*. ZC3H15 represents zinc finger CCCH-type containing 15, which is an immediate early erythropoietin response gene. Using the co-immunoprecipitation method, it can be found that ZC3H15 is functional via the signaling adapter protein tumor necrosis factor receptor–associated factor 2 (TRAF-2) with the NF-κB pathway for AML formation (Capalbo et al., 2013). Here, in this study, we also observed the key roles of ZC3H15 for pAML. Meanwhile, the NF-κB pathway was suggested as the primary signaling pathway by KEGG analysis (as shown in Figure 2). Taken these together, ZC3H15 may be a central factor for the pAML process through the NF-κB signaling pathway, raising an interesting direction for future pathological mechanism studies underlying pAML. BCLAF1 was originally initiated as a key of apoptosis and repressor of transcription, which is associated with antiapoptotic members of the Bcl2 family for various developmental processes such as T-cell activation and so on (Sarras et al., 2010). SRSF11 was previously demonstrated as a novel TERC-binding protein, which localizes to nuclear speckles and associates with active telomerase enzyme for cell cycle manipulation (Lee et al., 2015). Based on its function as a nuclear speckle–targeting factor essential for telomerase association with telomeres, SRSF11 has been deeply explored in cancer research. UBE3A is a dual-function protein, which consists of ubiquitin ligase as well as transcriptional co-activator function and has been shown to play a fundamental role in the modulation of synaptic function and plasticity (Vatsa and Jana, 2018). Except for these, the association between selected hub genes from this study and pAML has not been fully explored, which calls for a great point for future studies. Since *ZC3H15* and *BCLAF1* were suggested as two primary hub genes for pAML. We further investigated the expression level of them between control and pAML groups. Initially, we analyzed the expression in the cell line and observed the distinct expression pattern for both of them (data not shown). We preferred to utilize peripheral blood samples since pAML refers to a blood disease. Using clinical specimens, both ZC3H15 and BCLAF1 displayed a remarkable increased expression level in the pAML compared with the control group (as shown in Table 1), which confirmed the functions of these genes in clinical verification. It is worth noting that all the top hub genes showed increased expressions in the pAML group, which could be explained by the downregulations of top three miRNAs (as shown in Figures 1, 3). All of these deserve further investigation.

To conclude, in the light of unclear pathological mechanisms as well as missing potential biomarkers for pAML, we systematically generated a differential expression profile of mRNAs and miRNAs and provided several potential miRNA-dependent signaling axes for pAML patients. All the work here offered novel insights for the future pAML research.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Tianjin Medical University General Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

QW collected the data and analyzed the data. CY conducted the cellular experiments. QL wrote the manuscript. XC supervised the data and obtained the funding.

Funding

This research was funded by the Natural Science Foundation of Tianjin (18JCTPJC60300 and 18JCYBJC93000).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.865111/full#supplementary-material

References

Bolouri, H., Farrar, J. E., Triche, T., Ries, R. E., Lim, E. L., Alonzo, T. A., et al. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* 24 (1), 103–112. doi:10.1038/nm.4439

Bonifant, C. L., Velasquez, M. P., and Gottschalk, S. (2018). Advances in immunotherapy for pediatric acute myeloid leukemia. *Expert Opin. Biol. Ther.* 18 (1), 51–63. doi:10.1080/14712598.2018.1384463

Capalbo, G., Mueller-Kuller, T., Koschmieder, S., Klein, H. U., Ottmann, O. G., Hoelzer, D., et al. (2013). Characterization of ZC3H15 as a potential TRAF-2-interacting protein implicated in the NFκB pathway and overexpressed in AML. *Int. J. Oncol.* 43 (1), 246–254. doi:10.3892/ijo.2013.1924

Danen-van Oorschot, A. A., Kuipers, J. E., Arentsen-Peters, S., Schotte, D., de Haas, V., Trka, J., et al. (2012). Differentially expressed miRNAs in cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. *Pediatr. Blood Cancer* 58 (5), 715–721. doi:10.1002/pbc.23279

De Kouchkovsky, I., and Abdul-Hay, M. (2016). Acute myeloid leukemia: A comprehensive review and 2016 update. *Blood Cancer J.* 6 (7), e441. doi:10.1038/bcj.2016.50

Elgarten, C. W., and Aplenc, R. (2020). Pediatric acute myeloid leukemia: Updates on biology, risk stratification, and therapy. *Curr. Opin. Pediatr.* 32 (1), 57–66. doi:10.1097/MOP.0000000000000855

Klein, K., de Haas, V., and Kaspers, G. J. L. (2018). Clinical challenges in de novo pediatric acute myeloid leukemia. *Expert Rev. Anticancer Ther.* 18 (3), 277–293. doi:10.1080/14737140.2018.1428091

Lee, J. H., Jeong, S. A., Khadka, P., Hong, J., and Chung, I. K. (2015). Involvement of SRSF11 in cell cycle-specific recruitment of telomerase to telomeres at nuclear speckles. *Nucleic Acids Res.* 43 (17), 8435–8451. doi:10.1093/nar/gkv844

Lu, T. X., and Rothenberg, M. E. (2018). MicroRNA. *J. Allergy Clin. Immunol.* 141 (4), 1202–1207. doi:10.1016/j.jaci.2017.08.034

Mott, J. L., Kobayashi, S., Bronk, S. F., and Gores, G. J. (2007). mir-29 regulates Mcl-1 protein expression and apoptosis. *Oncogene* 26 (42), 6133–6140. doi:10.1038/sj.onc.1210436

Obulkasim, A., Katsman-Kuipers, J. E., Verboon, L., Sanders, M., Touw, I., Jongen-Lavrencic, M., et al. (2017). Classification of pediatric acute myeloid leukemia based on miRNA expression profiles. *Oncotarget* 8 (20), 33078–33085. doi:10.18632/oncotarget.16525

Patel, J. P., Gonen, M., Figueroa, M. E., Fernandez, H., Sun, Z., Racevskis, J., et al. (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* 366 (12), 1079–1089. doi:10.1056/NEJMoa1112304

Paulraj, P., Diamond, S., Razzaqi, F., Ozeran, J. D., Longhurst, M., Andersen, E. F., et al. (2019). Pediatric acute myeloid leukemia with t(7;21)(p22;q22). *Genes Chromosom. Cancer* 58 (8), 551–557. doi:10.1002/gcc.22740

Radhi, M., Meshinchi, S., and Gamis, A. (2010). Prognostic factors in pediatric acute myeloid leukemia. *Curr. Hematol. Malig. Rep.* 5 (4), 200–206. doi:10.1007/s11899-010-0060-z

Ribeiro, R. C. (2014). Advances in treatment of de-novo pediatric acute myeloid leukemia. *Curr. Opin. Oncol.* 26 (6), 656–662. doi:10.1097/CCO.0000000000000136

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Sarras, H., Alizadeh Azami, S., and McPherson, J. P. (2010). In search of a function for BCLAF1. *ScientificWorldJournal.* 10, 1450–1461. doi:10.1100/tsw.2010.132

Schuringa, J. J., Wierenga, A. T., Kruijer, W., and Vellenga, E. (2000). Constitutive Stat3, Tyr705, and Ser727 phosphorylation in acute myeloid leukemia cells caused by the autocrine secretion of interleukin-6. *Blood* 95 (12), 3765–3770. doi:10.1182/blood.v95.12.3765.012k50_3765_3770

Shimada, A. (2017). Pediatric acute myeloid leukemia with genetic alterations. *Rinsho. Ketsueki.* 58 (8), 983–990. doi:10.11406/rinketsu.58.983

Short, N. J., Rytting, M. E., and Cortes, J. E. (2018). Acute myeloid leukaemia. *Lancet* 392 (10147), 593–606. doi:10.1016/S0140-6736(18)31041-9

Taga, T., Tomizawa, D., Takahashi, H., and Adachi, S. (2016). Acute myeloid leukemia in children: Current status and future directions. *Pediatr. Int.* 58 (2), 71–80. doi:10.1111/ped.12865

Takahashi, S. (2011). Current findings for recurring mutations in acute myeloid leukemia. *J. Hematol. Oncol.* 4, 36. doi:10.1186/1756-8722-4-36

Tarlock, K., and Meshinchi, S. (2015). Pediatric acute myeloid leukemia: Biology and therapeutic implications of genomic variants. *Pediatr. Clin. North Am.* 62 (1), 75–93. doi:10.1016/j.pcl.2014.09.007

Vatsa, N., and Jana, N. R. (2018). UBE3A and its link with autism. *Front. Mol. Neurosci.* 11, 448. doi:10.3389/fnmol.2018.00448

Vishnoi, A., and Rani, S. (2017). MiRNA biogenesis and regulation of diseases: An overview. *Methods Mol. Biol.* 1509, 1–10. doi:10.1007/978-1-4939-6524-3_1

Volinia, S., Galasso, M., Costinean, S., Tagliavini, L., Gamberoni, G., Drusco, A., et al. (2010). Reprogramming of miRNA networks in cancer and leukemia. *Genome Res.* 20 (5), 589–599. doi:10.1101/gr.098046.109

Woods, W. G., Kobrinsky, N., Buckley, J. D., Lee, J. W., Sanders, J., Neudorf, S., et al. (1996). Timed-sequential induction therapy improves postremission outcome in acute myeloid leukemia: A report from the children's cancer group. *Blood* 87 (12), 4979–4989. doi:10.1182/blood.v87.12.4979.bloodjournal87124979

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, H., Luo, X. Q., Zhang, P., Huang, L. B., Zheng, Y. S., Wu, J., et al. (2009). MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia. *PLoS One* 4 (11), e7826. doi:10.1371/journal.pone.0007826

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fgene.2022.888601

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Yucong Zhang,
Huazhong University of Science and Technology, China
Hailong Wang,
Shanxi Medical University, China

*CORRESPONDENCE
Ming-You Dong,
mydong@ymcn.edu.cn
Run-Lei Du,
runleidu@whu.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 03 March 2022
ACCEPTED 22 August 2022
PUBLISHED 12 September 2022

CITATION
Liang X-Y, Zhang Y, He Y-N, Liu X-Y, Ding Z-H, Zhang X-D, Dong M-Y and Du R-L (2022), A cancer stem cell associated gene signature for predicting overall survival of hepatocellular carcinoma. *Front. Genet.* 13:888601. doi: 10.3389/fgene.2022.888601

COPYRIGHT
© 2022 Liang, Zhang, He, Liu, Ding, Zhang, Dong and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A cancer stem cell associated gene signature for predicting overall survival of hepatocellular carcinoma

Xin-Yi Liang[1], Yue Zhang[1], Ya-Nan He[1], Xue-Yi Liu[1], Zhi-Hao Ding[1], Xiao-Dong Zhang[1], Ming-You Dong[2]* and Run-Lei Du[1]*

[1]Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China, [2]The Key Laboratory of Molecular Pathology (For Hepatobiliary Diseases) of Guangxi, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China

Hepatocellular carcinoma (HCC) is the most prevalent type of primary liver cancer characterized by high mortality and morbidity rate. The lack of effective treatments and the high frequency of recurrence lead to poor prognosis of patients with HCC. Therefore, it is important to develop robust prediction tools for predicting the prognosis of HCC. Recent studies have shown that cancer stem cells (CSC) participate in HCC progression. The aim of this study was to explore the prognostic value of CSC-related genes and establish a prediction model based on data from The Cancer Genome Atlas (TCGA) database. In this study, 475 CSC-related genes were obtained from the Molecular Signature Database and 160 differentially expressed CSC-related genes in HCC patients were identified using the limma R package in the TCGA database. A total of 79 CSC-related genes were found to be associated with overall survival (OS). Using the least absolute shrinkage and selection operator (LASSO) and multivariate Cox regressions, a 3-gene signature (*RAB10*, *TCOF1*, and *PSMD14*) was constructed. Receiver operating characteristic (ROC) curves and Kaplan-Meier survival curves were constructed to test the prediction performance of the signature. Performance of the signature was validated using the International Cancer Genome Consortium (ICGC) dataset. In addition, immune feature and functional enrichment analyses were carried out to explore the underlying mechanisms. Moreover, a co-expression network was constructed using the weighted gene correlation network analysis (WGCNA) method to select genes significantly associated with risk scores in HCC in the TCGA dataset. The *SGO2* gene was found to be significantly associated with risk scores of HCC. *In vitro* experiments revealed that it can promote HCC cell proliferation. Therefore, *SGO2* may be a potential therapeutic target for HCC treatment. The constructed nomogram can help clinicians make decisions about HCC treatment.

KEYWORDS

hepatocellular carcinoma, cancer stem cells, gene signature, immune, SGO2

Introduction

Liver cancer, which is ranked sixth and third in terms of morbidity and mortality among all malignant tumors, has become a major public health issue worldwide (Sung et al., 2021). Of all subtypes of primary liver cancer, hepatocellular carcinoma (HCC) accounts for 80% of all liver cancer patients globally (Rumgay et al., 2022). Recent etiologic studies have shown that infection with HBV or HCV is the leading cause of HCC, and it mostly occurs in countries and regions with low economic development (Ghouri et al., 2017). In addition, alcohol addiction, metabolism-related liver disease, and dietary toxins, including aflatoxins and aristolochic acid, are common risk factors of HCC in some developed countries (El-Serag, 2012; Yang et al., 2019a). To date, there is no elaborate treatment strategy for HCC, which calls for studies to find a cure for this disease. Several molecularly targeted drugs have been approved for the treatment of HCC, but they only have moderate effects and are not effective in all patients (Huang et al., 2020). Although surgery, chemotherapy, and radiotherapy are widely used for the HCC treatment, the prognosis of patients is still poor, with a 5-years survival rate of only about 18% (Jemal et al., 2017). Prediction of HCC prognosis is hampered by the complex tumorigenesis mechanism, high degree of heterogeneity, and frequent recurrence.

Recent studies have demonstrated that liver cancer stem cells may be responsible for these malignant properties of HCC (Yamashita and Wang, 2013). Cancer stem cells (CSCs), also known as tumor-initiating cells, are a subpopulation of cells within the tumor that possess stem cell-like capacity. Lee et al. (2022) revealed that they are self-renewable and can differentiate into aggressive tumor cells to promote tumor growth. These cells may be the cause of tumor relapse and high heterogeneity. With the advancement in experimental technology, several biological markers of liver CSCs have been identified in HCC and were recently reviewed by (Tsui et al., 2020). Some markers, including CK19, ABCG2, CD44, and CD133, were found to be correlated with highly invasive features, and patients with elevated expression levels of these markers had worse prognosis and shorter survival time (Yang et al., 2010). Therefore, the CSCs-related gene signatures have the potential to become powerful predictors of prognosis for HCC patients.

In this study, the mRNA expression profiles and corresponding clinical data of HCC patients were retrieved from public databases and used to identify the key genes associated with liver CSCs. Consequently, a 3-gene signature model was constructed using The Cancer Genome Atlas (TCGA) cohort, and validated in the International Cancer Genome Consortium (ICGC) cohort. We confirmed that this gene signature was an independent predictor of overall survival (OS) and the underlying mechanisms were explored, with the overarching goal of providing a new strategy for accurately predicting the prognosis of HCC patients.

Materials and methods

Data collection

Gene expression profiles of RNA-sequencing data and corresponding clinical information of HCC patients were extracted from TCGA database (https://portal.gdc.cancer.gov/repository). Another dataset consisting of 231 samples was downloaded from the ICGC database. The TCGA-LIHC dataset was used as the training cohort and ICGC dataset as the validation cohort. As the data from both TCGA and ICGC are publicly available. Therefore, this study was exempted from approval by the local ethics committee. Patients with no follow up data from both TCGA and ICGC were excluded from the analysis. All read count values were normalized.

A total of 475 CSC genes from 109 CSC-related gene sets were obtained from a published Articles (Liang et al., 2020). The 109 CSC-related gene sets are provided in Supplementary Table S1.

Construction and validation of a prognostic CSC-related gene signature

The "limma" R package was used to identify the differentially expressed genes (DEGs) between tumor and adjacent normal samples in the TCGA cohort, with false discovery rate (FDR) < 0.05 set as the cut-offs. Next, univariate Cox regression analysis of overall survival (OS) was performed to determine the prognostic value of CSC-related genes. 79 prognostic genes was obtained after examining the intersection between the two gene groups. Furthermore, to minimize overfitting, prognostic CSC-related genes were assessed by least absolute shrinkage and selection operator (LASSO) Cox proportional hazards regression using the "glmnet" R package. The value of penalty parameter (λ) corresponding to the lowest partial likelihood deviance was used to select the best model by 10-fold cross-validation. We got a list of genes with non-zero beta coefficients. Finally, a stepwise multivariate Cox regression analysis were used to establish a score system for calculating the survival risk for each HCC patient based on the expression level of each prognostic gene and its related regression coefficient. The following formula was used: risk score = sum (gene expression level * regression coefficient). Moreover, all patients were classified into high-risk or low-risk group according to the median value of risk score, which was used as the cut-off value. To observe the clustering conditions of the gene signature, principal component analysis (PCA), and t-SNE were performed using the "prcomp" function of the "stats" R package and the "Rtsne" R package, respectively. Time-dependent receiver operating characteristic (ROC) curves were generated using the "survivalROC" R package for evaluating the predictive capacity of the novel gene signature. The survival curves for

different groups were analyzed with the Kaplan-Meier method with log-rank test. Finally, univariate and multivariate Cox regression analyses were performed to determine whether the CSC-related gene signature possessed the independent prognostic value. To verify the effectiveness of the model, the β value derived from the TCGA set was applied to the ICGC set, all patients come from ICGC were also classified into high-risk or low-risk group according to the same median value of risk score in TCGA set.

Independence of the 3-CSC-related genes signature from clinical features of other TCGA-LIHC patients

Based on other clinical features (grade, age, TNM stage, and T stage) of TCGA-LIHC patients, univariate and multivariate Cox regression analyses were conducted to explore whether the prognostic model was an independent variable. To confirm the prognostic significance of the predictive model, TCGA-LIHC patients were divided into two groups according to different clinical characteristics. Patients were separately classified into the following subgroups: grade I/II, grade III/IV, stage I/II, stage III/IV, age <65, age ≥65, T1-T2, and T3-T4 subgroups. Survival outcome analysis was then performed to verify the independent prognostic significance of the gene signature in specific subgroups. The ideal cut-off value of the risk score was established using the surv_cutpoint function of "survminer" in R package.

Immune infiltration analysis

Single-sample gene set enrichment analysis (ssGSEA) was performed using "ssGSEA" R package to calculate the infiltrating score of 16 immune cells and the activity of 13 immune-related pathways of patients in the high-risk and low-risk groups. Supplementary Table S2 shows the representative gene sets of immune cells and related pathways.

Gene set variation analysis (GSVA)

Gene set variation analysis (GSVA) is a non-parametric and unsupervised GSE method which can calculate enrichment scores of predefined gene sets representing various biological processes in each sample (Hanzelmann et al., 2013). The GSVA was performed using the "clusterProfiler" R package to convert the gene expression profiles of patients in high-risk and low risk groups from TCGA and ICGC cohorts into the enrichment scores in biological signaling pathways or functions. The predefined pathway gene sets were obtained from the Kyoto Encyclopedia for Genes and Genomes (KEGG) database.

Weighted gene correlation network analysis (WGCNA)

To identify the risk score-related hub genes, WGCNA was applied to HCC samples from the TCGA database using the "WGCNA" R package. This analysis method has two parts. One part classifies genes into different modules according to their expression patterns and the other identifies modules that are highly correlated with traits (Langfelder and Horvath, 2008). The WGCNA analysis was carried out on the identified DEGs in HCC tumor samples. First, a gene expression similarity matrix was constructed through calculating the Pearson correlation coefficient between any two genes. Next, a soft threshold of $\beta = 8$ was used to judge the similarity of two genes, followed by converting the similarity matrix into a weighted adjacency matrix. Finally, a topological overlap matrix (TOM) was applied to further measure the connectivity of genes in the co-expression network. Genes were clustered according to the value of 1-TOM. In addition, different modules were divided from the identified DEGs by building the dynamic pruning tree, with each module containing at least 30 genes. Similar modules were merged at the cut-off value of 0.25.

Co-expression network was constructed and genes were divided into modules, after which the relationships among modules and traits were determined, which included risk score and risk groups (low-risk = 0, high-risk = 1). Highly correlated modules were selected for further analyses. Gene ontology (GO) and KEGG analyses were then performed using the "clusterProfiler" R package to evaluate the biological functions of genes in the modules.

Establishment and evaluation of a predictive nomogram

A nomogram was constructed based on gender, stage, grade, age, and risk score (Iasonos et al., 2008). The area under the ROC curve (AUC), 1-, 3-, and 5-years calibration curves, and decision curve analyses (DCA) (Vickers and Elkin, 2006) were then used to assess the nomogram's prediction accuracy and discriminatory capacity.

Cell culture

Human HCC cell lines Huh-7 and human kidney cells lines HEK293T were purchased from American Type Culture Collection. All cells were cultured in Dulbecco's modified Eagle medium (HyClone, USA) supplemented with 10% fetal bovine serum (Gibco, USA) and 100U penicillin/streptomycin (HyClone, USA) at 37°C in a humidified thermal incubator with 5% CO_2.

CRISPR-Cas9-mediated SGO2 knockdown

Two optimal guide RNAs (CCAGTCTATTGGCCGCAGAT and AATAGTTCAGATGTCGATAT) targeting come from the different sites of the human SGO2 gene exon 6 were designed on the CRISPOR website (http://crispor.tefor.net/). Next, the gRNAs were cloned into pLentiCRISPRv2 (Addgene plasmid #52961). HEK293T cells ($2x10^5$) were first seeded in 6-well plates and then transfected with pLentiCRISPRv2-gRNA plasmid or empty pLentiCRISPRv2 (2 μg) together with two lentiviral packaging plasmid psPAX2 (1 μg, Addgene plasmid #12260) and pMD2.G (1 μg, Addgene plasmid #12259) to generate lentivirus. After transfection for 48 h, the upper media containing lentivirus was collected and filtered using 0.45 μm membrane filters. On the other hand, Huh-7 cells were seeded in 6-well plates at a density of $2x10^5$ cells per well. After incubation for 24 h, the media were replaced with a mixture of two lentiviruses and fresh media (1:1:2) containing polybrene (10 μg/ml). The infected cells were selected by treatment with 1 μg/ml puromycine for 4 days until the death of control Huh-7 cells. Finally, the expression of *SGO2* was determined using quantitative real-time polymerase chain reaction (qRT-PCR).

Cell proliferation and colony formation

Cell Counting Kit-8 (CCK-8, Dojindo Laboratories, Kumamoto, Japan) was used to explore the effect of *SGO2* knockdown on cell proliferation of HCC cells. Briefly, cells were first seeded into 96-well plates at a density of $1x10^3$ cells per well. Next, the CCK-8 solution (10 μl in 90 μl DMEM) was added to each well and incubated at 37°C for 1 h. Finally, the optical density of the medium was measured at a wavelength of 450 nm.

Colony formation experiment: Cells were placed in 6-well plates (1,000 cells per well). After 14 days of incubation, the colonies were fixed in 4% paraformaldehyde (Cat. 15,700, Electron Microscopy Sciences, USA) for 30 min, stained with 0.1% crystal violet (Beyotime, Beijing, China) for 15 min, and then washed it. Finally, colonies were photographed and counted using ImageJ software.

Quantitative real-time PCR

Total RNA of wildtype and *SGO2* knockdown Huh-7 cells were extracted using TRIzol reagent (TAKARA, Japan). qRT-PCR was then performed with MonAmp™ ChemoHS qPCR Mix (Monad) in accordance with the manufacturer's instructions. Relative expression of *SGO2* mRNA was normalized to *β-actin* mRNA (internal control) and calculated using the $2^{-\Delta\Delta Ct}$ method. The sequences of primer used were as follows: *SGO2*: 5'- GCC CAGTCTATTGGCCGCAG -3' (forward) and 5'- TTCAATCTT TTCCCCAATAT -3' (reverse); *β-actin*: 5'- CACCATTGGCAA TGAGCGGTTC -3' (forward) and 5'- AGGTCTTTGCGGATG TCCACGT-3' (reverse).

TABLE 1 Clinical characteristics of the HCC patients used in this study.

	TCGA cohort	ICGC cohort
No. of patients	365	231
Age (median, range)	61(16-90)	69(31-89)
Gender (%)		
Female	119(32.6%)	61(26.4%)
Male	246(67.4%	170(72.6%)
Grade (%)		
Grade 1	55(15.1%)	NA
Grade 2	175(47.9%)	NA
Grade 3	118(32.3%)	NA
Grade 4	12(3.3%)	NA
Unknown	5(1.4%)	NA
Stage (%)		
I	170(46.6%)	36(15.6%)
II	84(23.0%)	105(45.5%)
III	83(22.7%)	71(30.7%)
IV	4(1.1%)	19(8.2%)
unknown	24(6.6%)	0(0.0%)
Survival status OS days (median)	556	780
Survival status		
Dead	130	189
Alive	235	42

Statistical analyses

The students *t* test was used to evaluate gene expressions differences between normal and tumor tissues, while the chi-square test was used to compare proportional differences between normal and tumor tissues. The Kaplan Meier and log-rank tests were used for comparisons of OS for various clinical subgroup. Univariate and multivariate Cox regression analysis were used to identify independent predictors of OS. For data analysis, the R program (version 4.1.0) was utilized. Unless otherwise stated, $p < 0.05$ denotes significance.

Results

Identification of prognostic cancer stem cell (CSC)-related DEGs in TCGA cohort

Supplementary Figure S1 shows the flow chart of the study. The detailed clinical characteristics of these patients come

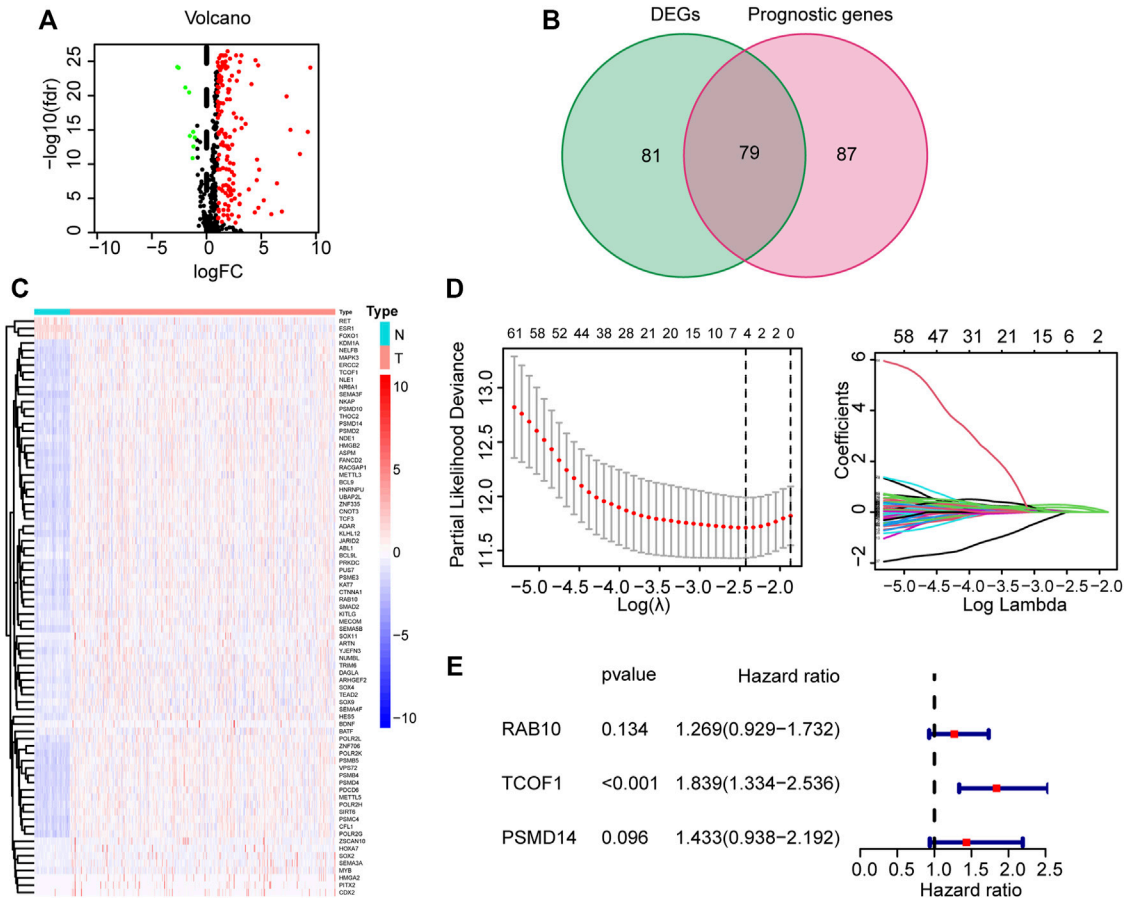


FIGURE 1
Identification and construction of CSC-related gene signature from TCGA corhot. **(A)** Volcano map of CSC-related DEGs in tumor and normal tissues from the TCGA dataset. **(B)** Venn plot used for selection of DEGs with prognostic value. **(C)** Heatmap illustrating differences in expression of 79 overlapping genes. **(D)** LASSO Cox regression analysis **(E)** Forest plot showing the relationship between the 3 genes and OS. CSC: cancer stem cell; DEG: Differentially expressed genes; LASSO: least absolute shrinkage and selection operator; OS: overall survival.

from TCGA are summarized in Table 1. A total of 160 CSC-related DEGs were identified between HCC cancer samples and adjacent normal samples (Supplementary Figures S2A,B). Among them, 79 DEGs were found to be associated with OS in the univariate Cox regression analysis (Figures 1A,B). Analysis of the heatmap plot showed that most of these prognostic DEGs were upregulated in the cancer samples (Figure 1C). GO analysis revealed that these CSC-related genes were mainly enriched in response to stem cell differentiation, mesenchyme development, mesenchymal cell differentiation, maintenance of stem cell population, and stem cell development (Supplementary Figure S2C). Moreover, KEGG pathway analysis revealed that these genes were correlated with the hippo signaling pathway, Wnt signaling pathway, TGF-beta pathway, and notch signaling pathway (Supplementary Figure S2D).

Construction and evaluation of a CSC-related prognostic signature in the TCGA cohort

To further filter the genes and reduce the risk of overfitting, LASSO Cox regression was employed. According to the multivariate Cox regression of OS, A 3-gene signature (*RAB10, TCOF1*, and *PSMD14*) was eventually identified based on the minimum value of λ (Figures 1D,E). Next, the risk score of each HCC patient in the TCGA dataset was calculated using the following formula: 0.238*expression level of *RAB10* + 0.609*expression level of *TCOF1*+0.360*expression level of *PSMD14*. Patients were stratified into two groups based on the risk score; patients with risk scores higher than the median value were classified in high-risk group (n = 182), whereas patients with risk scores lower than the median value were classified in low-risk group

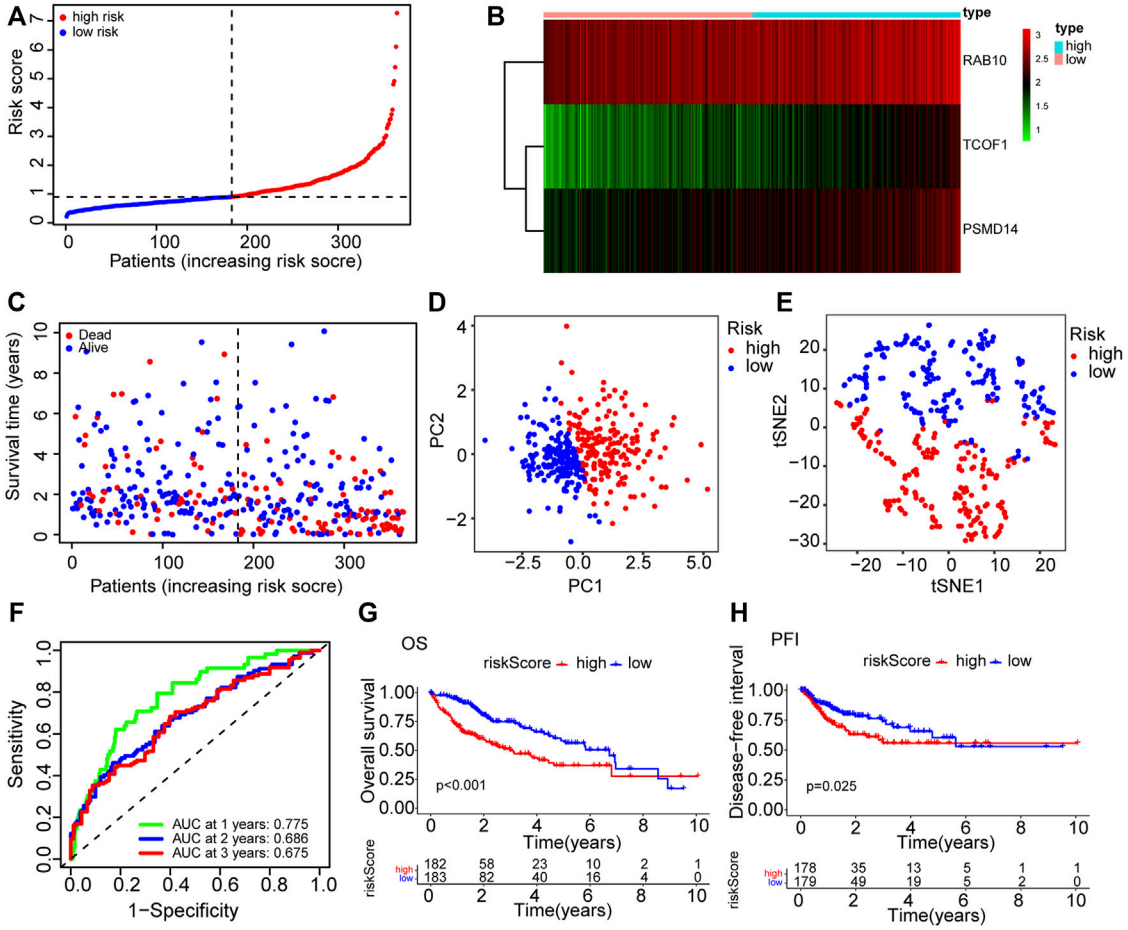


FIGURE 2
The prognosis prediction value of CSC-related gene signature in TCGA cohort. **(A)** The distribution of patients in high risk and low risk groups in the TCGA dataset. **(B)** Heatmap showing differences in expression levels of the 3 genes between the two risk groups. **(C)** The distribution of OS status and OS among various risk score groups. **(D)** PCA analysis of the TCGA dataset. **(E)** t-SNE analysis of the TCGA dataset. **(F)** AUC of time-dependent ROC curves for the risk score. **(G)** Kaplan-Meier survival curves showing the OS of patients in the two risk groups. **(H)** Kaplan-Meier survival curves showing the disease-free interval of patients in the two risk groups. AUC: area under the ROC curve; DFI: disease-free interval; OS: overall survival; PCA: Principal Component Analysis; ROC: Receiver Operating Characteristic.

(n = 183) (Figure 2A). Results showed that the three genes were significantly upregulated in the high-risk group (Figure 2B). PCA and t-SNE analysis revealed that patients in the two groups were well distributed in different directions (Figures 2D,E). Further analysis showed that patients in the high risk score group had shorter the survival time and mortality risk (Figure 2C). Time-dependent ROC curves were generated to evaluate the prediction power of the risk score. Results demonstrated that the area under the curve (AUC) for 1-, 2-, and 3-years OS was 0.775, 0.686, and 0.675, respectively, indicating that the gene signature could predict the prognosis of HCC (Figure 2F). Analysis of Kaplan-Meier survival curves showed that patients in the high-risk group had poor OS and platinum-free interval (PFI) ($p < 0.001$; p = 0.025, Figures 2G,H).

Validation of the prognostic model in the ICGC cohort

The dataset downloaded from the ICGC was used to verify the generality of the novel CSC-related prognostic model. The risk score was used to divide patients from the ICGC dataset into high-risk groups and low-risk groups (Figure 3A). Similarly, the three prognostic risk genes were upregulated in the high-risk group (Figure 3B). PCA and t-SNE analyses indicated that patients in the different risk groups were distributed in two discrete sections (Figures 3C,D). Results of the scatterplot and Kaplan-Meier survival curves showed that patients with high-risk scores had short survival time and high mortality rate (Figures 3E,F), consistent with findings in the TCGA cohort. The AUC of time-dependent ROC ranged from 0.705 to 0.703 for 3 years

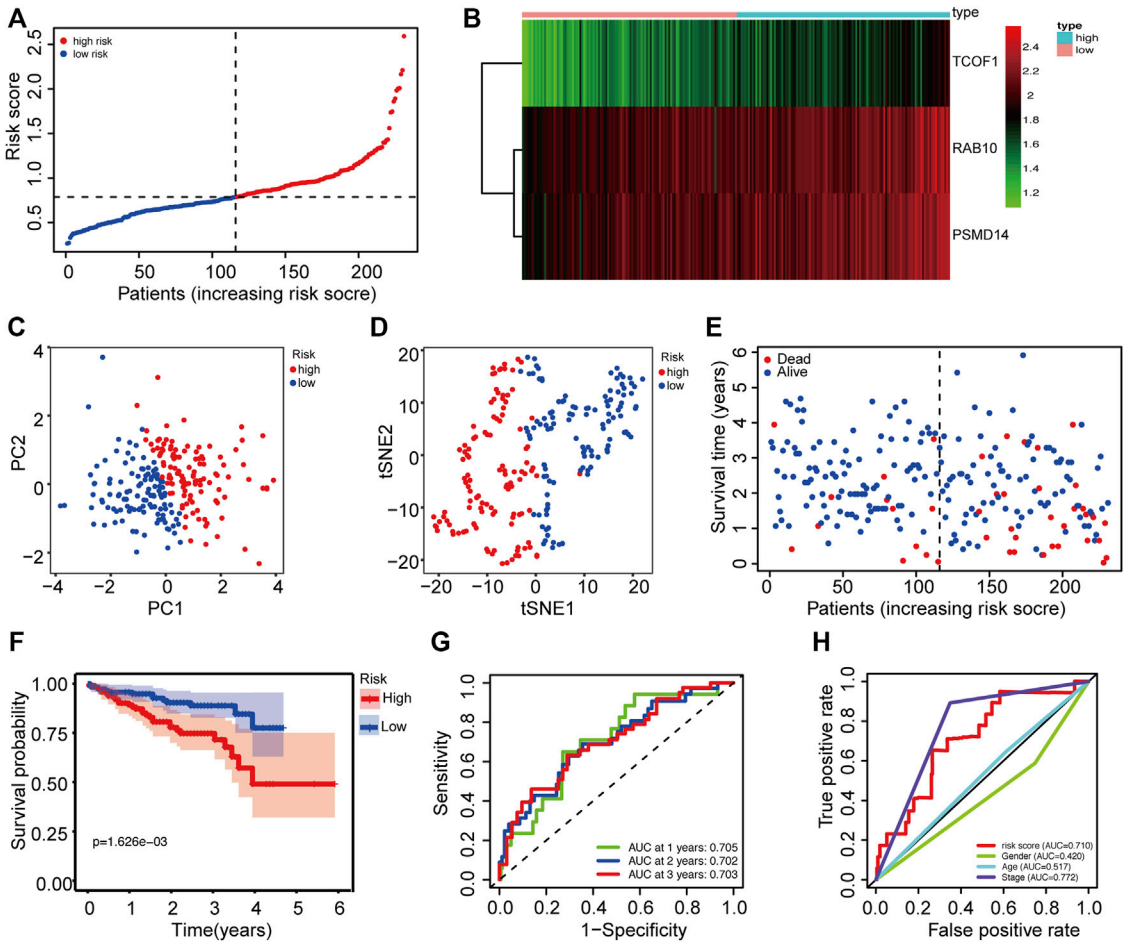


FIGURE 3
Validation of the prognostic prediction accuracy of the signature in the ICGC cohort. **(A)** The distribution of patients between high risk and low risk groups in the ICGC dataset. **(B)** Heatmap showing differences in expression levels of 3 genes between the two risk groups. **(C)** PCA analysis for the ICGC dataset. **(D)** t-SNE analysis for the ICGC dataset. **(E)** The distribution of OS status and OS among various risk score groups. **(F)** Kaplan-Meier survival curves for the OS of patients in the two risk groups. **(G)** AUC of time-dependent ROC curves for the risk scores. **(H)** AUC of ROC curves for the risk score and various clinical characteristics. AUC: area under the ROC curve; DFI: disease-free interval; OS: overall survival; PCA: Principal Component Analysis; ROC: Receiver Operating Characteristic.

(Figure 3G), indicating that the gene signature could predict the prognosis of patients in the ICGC cohort. Moreover, the risk score showed better prediction than most the clinical characteristics such as gender and age (Figure 3H).

Independent prognostic value of the 3-CSC-related genes signature

To further evaluate the prognostic value of the risk score, univariate and multivariate Cox regression analyses were applied among risk score and some clinicopathological characteristics in both TCGA and ICGC cohorts. The univariate Cox regression analysis results showed that the risk score ($p < 0.001$, HR = 1.623, 95% CI = 1.422-1.854) and tumor stage ($p < 0.001$, HR = 2.500, 95% CI = 1.721-3.632) were significantly associated with OS in the TCGA cohort (Figure 4A). In the multivariate Cox regression analysis of the two variables, the risk score ($p < 0.001$, HR = 1.635, 95% CI = 1.418-1.884) and tumor stage ($p < 0.001$, HR = 2.361, 95% CI = 1.624-3.432) were independent prognostic factor in the TCGA dataset (Figure 4C). The risk score was also an independent prognostic factor in the ICGC cohort ($p < 0.001$, HR = 3.534, 95% CI = 1.862-6.709) (Figure 4B, Figure 4D). Collectively, these results suggest that the constructed risk score was an independent prognostic predictor of OS.

To assess the prognostic value of the model in various clinic-pathological subgroups, clinical variables and samples were randomized into two subgroups in term of TNM stage, age, grade, and T stage. The obtained results indicated that the 3-CSC-related genes signature was significantly correlated with the

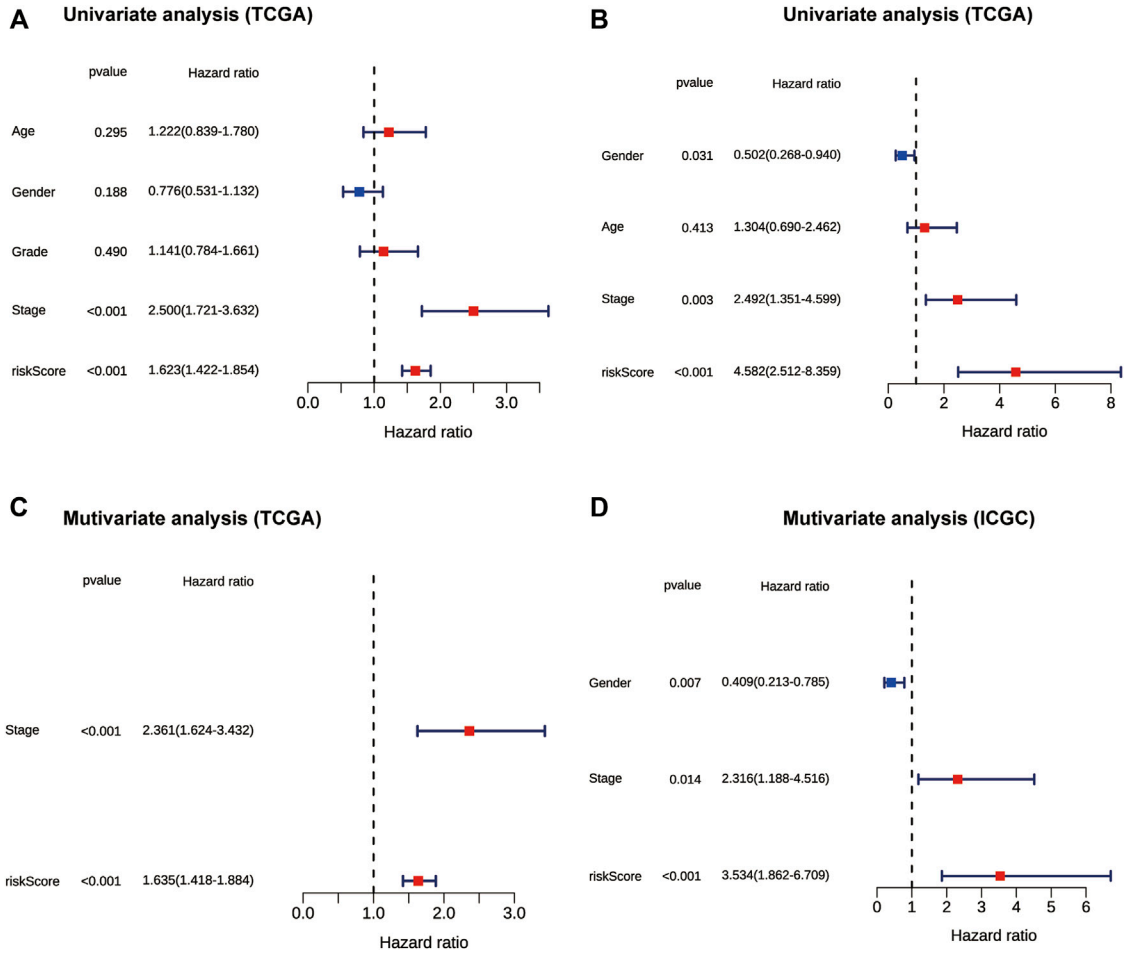


FIGURE 4
The forest plot showing results of univariate and multivariate Cox regression analyses of OS in the TCGA **(A,C)** and ICGC cohort **(B,D)**. OS: overall survival.

survival of patients in different clinic-pathological subgroups in the TCGA-LIHC cohort (Supplementary Figure S3).

The correlation between immune status and risk score

HCC patients were classified into high-risk and low-risk groups based on the risk score. It was evident that patients with high-risk scores were more likely to have poor prognosis, which has been confirmed above. However, the correlation between poor prognosis and high-risk score was not clear. We suspected that there might be a difference in immune status between the two risk groups. To this effect, the ssGSEA method was applied to all HCC samples in the high-risk and low-risk groups to evaluate immune cells infiltration and the activities of immune-related pathways. Figure 5A and Figure 5B show the enrichment scores of 16 types of infiltrating immune cells. In TCGA and ICGC cohorts, the scores of aDCs, macrophages, Th2 cells, and Tregs were significantly upregulated in the high-risk score group, whereas the scores of neutrophils and NK cells were lower than those in the low-risk group ($p < 0.05$). Among the screened immune-related functions or pathways, the expression of MHC class I was increased in the high-risk score group. In contrast, the activities of type I IFN response and type II IFN response were inhibited in the high-risk score group ($p < 0.05$, Figures 5C,D).

Identification of biological functions associated with the risk score

To further explore the differences in the activities of various biological signaling pathways between high-risk and low-risk

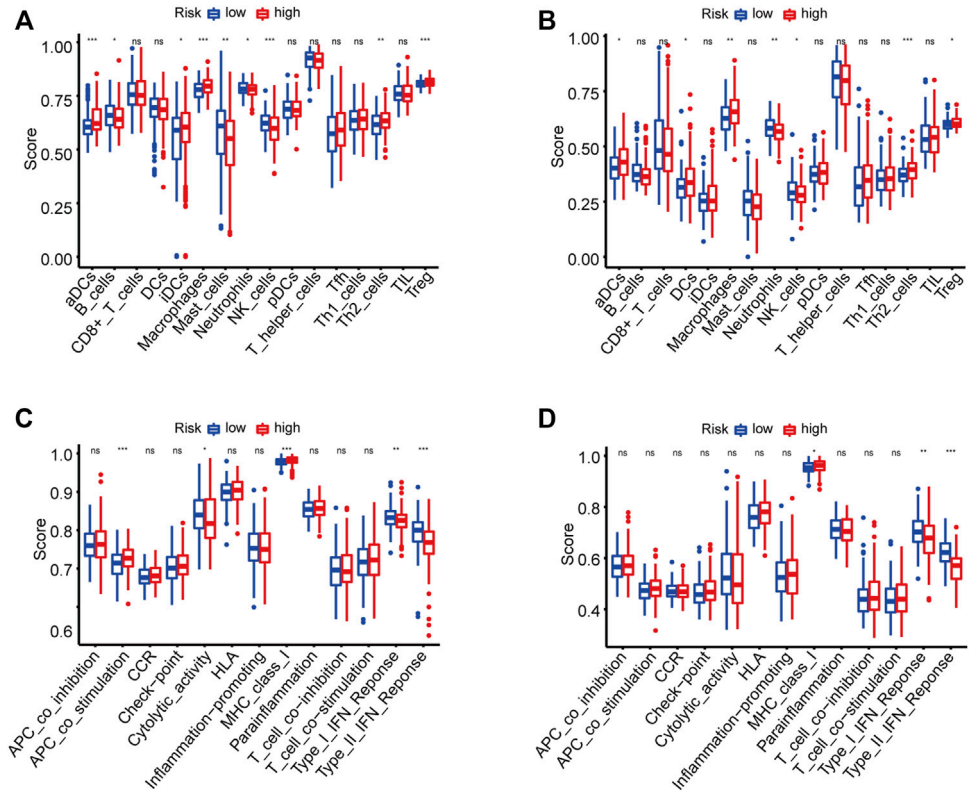


FIGURE 5
The enrichment scores of 16 types of immune infiltrating cells and 13 immune-related functions in the TCGA **(A,C)** and ICGC **(B,D)** cohorts. *p < 0.05, **p < 0.01, ***p < 0.001.

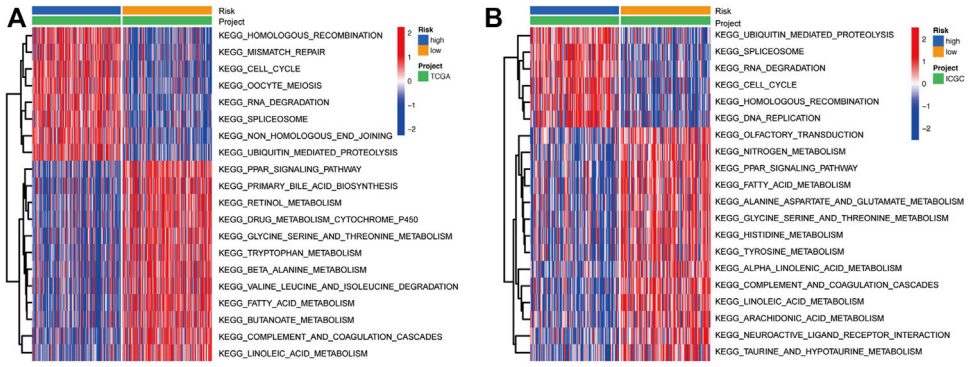


FIGURE 6
Heatmap showing the KEGG pathways enriched in high- and low-risk groups in the TCGA **(A)** and ICGC **(B)** cohorts. KEGG: Kyoto Encyclopedia of Genes and Genomes.

groups, GSVA enrichment analysis was performed for each sample from the TCGA and ICGC cohorts. Results showed that the common pathways and functions enriched in samples from high-risk groups in both TCGA and ICGC cohorts were homologous recombination, cell cycle, RNA degradation, splicesome, and ubiquitin mediated proteolysis (Figures 6A,B). On the other hand, the activities of various metabolism-related pathways in the low-risk groups were substantially higher than those

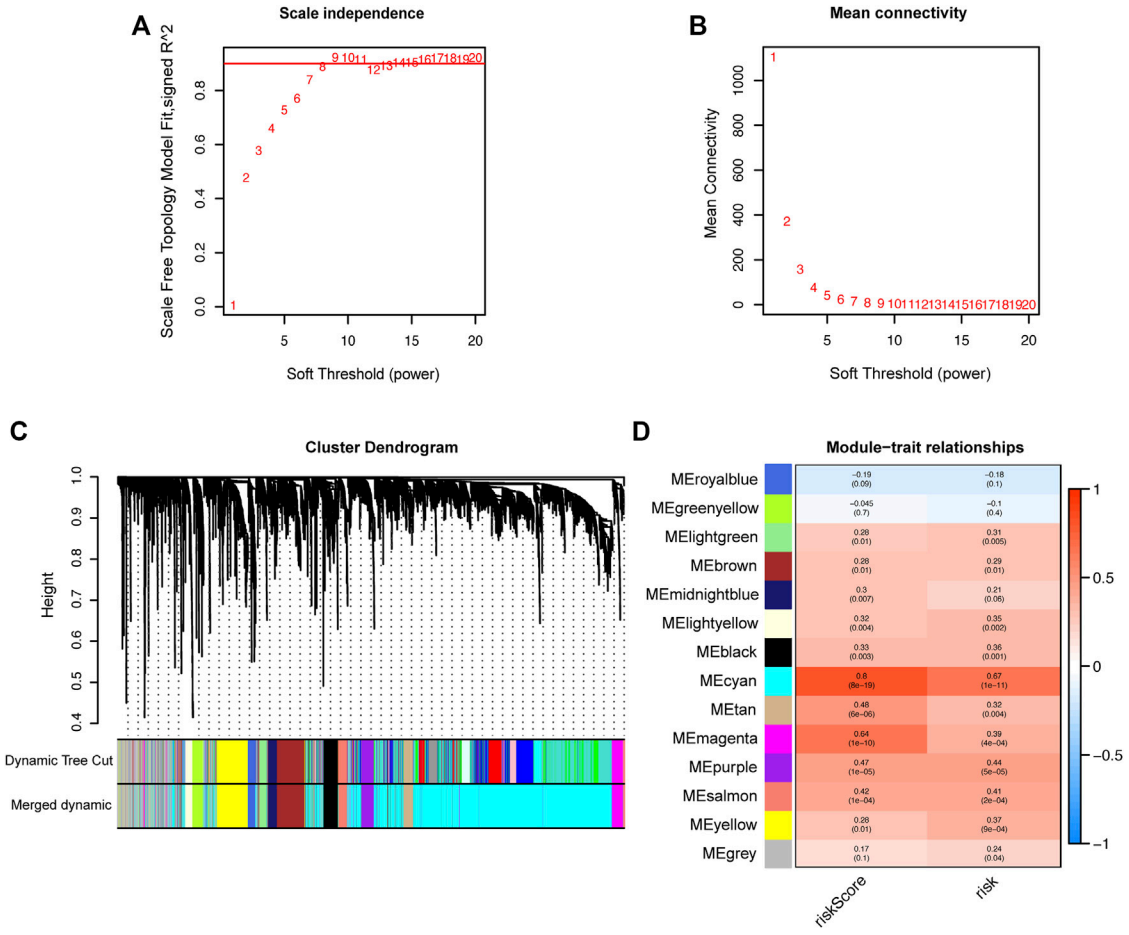


FIGURE 7
Identification of hub genes related to the risk score by WGCNA. **(A)** Analysis of the optimal soft-thresholding power based on scale-free network. **(B)** Mean connectivity analysis of soft-thresholding powers. **(C)** Cluster dendrogram of co-expressed genes. Each branch represents a module and is given a unique color. **(D)** Heatmap showing the correlation coefficients (upper row) and significance (*p* values showed in the brackets) between modules and risk score or risk groups. WGCNA: Weighted Gene Co-expression Network analysis.

in the high-risk groups, including glycine-serine and threonine metabolism, fatty acid metabolism, linoleic acid metabolism, and the PPAR signaling pathway. In addition, one immune-related pathway, complement and coagulation cascades, was enriched in the low-risk groups (Figures 6A,B).

Hub genes associated with high-risk in HCC patients

To screen out the hub genes responsible for the high-risk of poor prognosis in HCC, WGCNA was performed on the DEGs between tumor and normal tissues identified from TCGA database. A co-expression network of 5,301 genes was built and 14 modules were divided from it through setting the optimal soft-thresholding power at 8 (scale free R^2 = 0.85, Figures 7A–C). The correlation coefficients between each module and risk score or risk groups (high-risk = 1; low-risk = 0) were calculated and presented in a heatmap. The results suggested that the cyan module was the most closely and positively associated with risk score and the high-risk group, with correlation coefficients of 0.8 and 0.67, respectively ($p < 0.001$, Figure 7D). Therefore, the cyan module was selected for further analysis. GO and KEGG enrichment analyses were conducted on the 2,815 genes in the cyan module to reveal the functional link between key genes. The top 10 GO terms and KEGG pathways enriched from the genes of cyan module are shown in Supplementary Figure S4.

Genes in the module were ranked by the scores and *p* values of gene significance (GS) and module membership (MM). Notably, the top 10 genes are listed in Table 2. Given that the mechanism of the number 1 gene (*BUB1B*) in HCC has already been elucidated (Qiu et al., 2020), we chose the number 2 key gene (*SGO2*) for subsequent analyses. Notably, *in vitro* experiments were conducted to further investigate the

TABLE 2 Genes in the cyan module ranked by scores and *p* values of MM and GS.

Gene	MM	Correlation coefficients and *p*-values	
		Risk score	Risk group
BUB1B	0.8407 (3.29E-22)	0.8814 (8.60E-27)	0.5789 (2.28E-08)
SGO2	0.9126 (1.24E-31)	0.8558 (9.66E-24)	0.6264 (6.65E-10)
STIP1	0.7749 (5.29E-17)	0.8464 (9.16E-23)	0.7098 (2.43E-13)
NCAPG	0.7764 (4.21E-17)	0.8447 (1.34E-22)	0.5237 (7.33E-07)
PRC1	0.7758 (4.61E-17)	0.8339 (1.44E-21)	0.5418 (2.51E-07)
HJURP	0.8376 (6.54E-22)	0.8315 (2.40E-21)	0.5658 (5.53E-08)
SPC25	0.7419 (5.22E-15)	0.8312 (2.52E-21)	0.5789 (2.28E-08)
KPNB1	0.8374 (6.82E-22)	0.8294 (3.69E-21)	0.5746 (3.08E-08)
SPATS2	0.8872 (1.39E-27)	0.8278 (5.16E-21)	0.6553 (5.63E-11)
KIF23	0.8164 (4.77E-20)	0.8263 (6.89E-21)	0.5478 (1.74E-07)

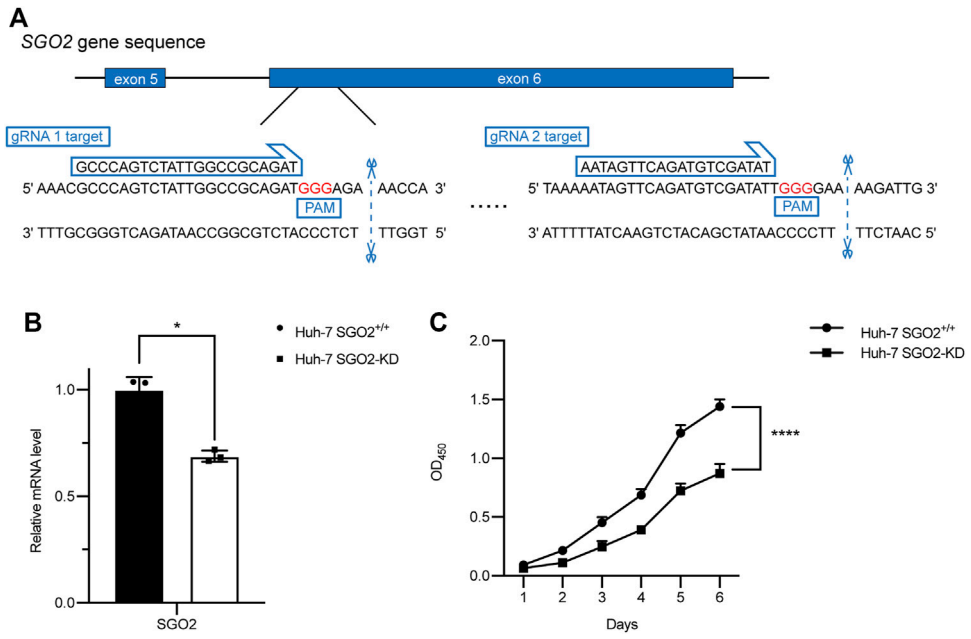


FIGURE 8
CRISPR-Cas9 mediated SGO2 knockdown inhibits the proliferation of HCC cells. **(A)** Two specific gRNAs targeting the gene sequence of *SGO2*. **(B)** Confirmation of expression level of SGO2 *mRNA* in Huh-7 SGO2 knockdown cells and control cells by Quantitative real-time PCR assay. **(C)** The cell proliferation status, evaluated daily for 6 days using the CCK-8 assay (n = 6). CCK-8: Cell Counting Kit-8; HCC: Hepatic cell carcinoma; PCR: Polymerase Chain Reaction.

biological functions of *SGO2* in HCC. Specifically, the expression level of *SGO2* was knocked down in Huh-7 cells by CRISPR/Cas 9 system as illustrated in Figure 8A. The success of *SGO2* knockdown was confirmed using qRT-PCR analysis (Figure 8B). The CCK-8 assay results showed that decreased expression of *SGO2* had a significant inhibitory effect on the proliferation of Huh-7 cells (Figure 8C).Clone formation experiments showed, Huh-7 cells with knockdown of *SGO2* showed worse ability to form colonies.The difference was statistically significant ($p < 0.05$), as shown in Supplementary Figure S5. The results showed that *SGO2* knockdown effectively inhibited the clone formation ability of HCC cells.

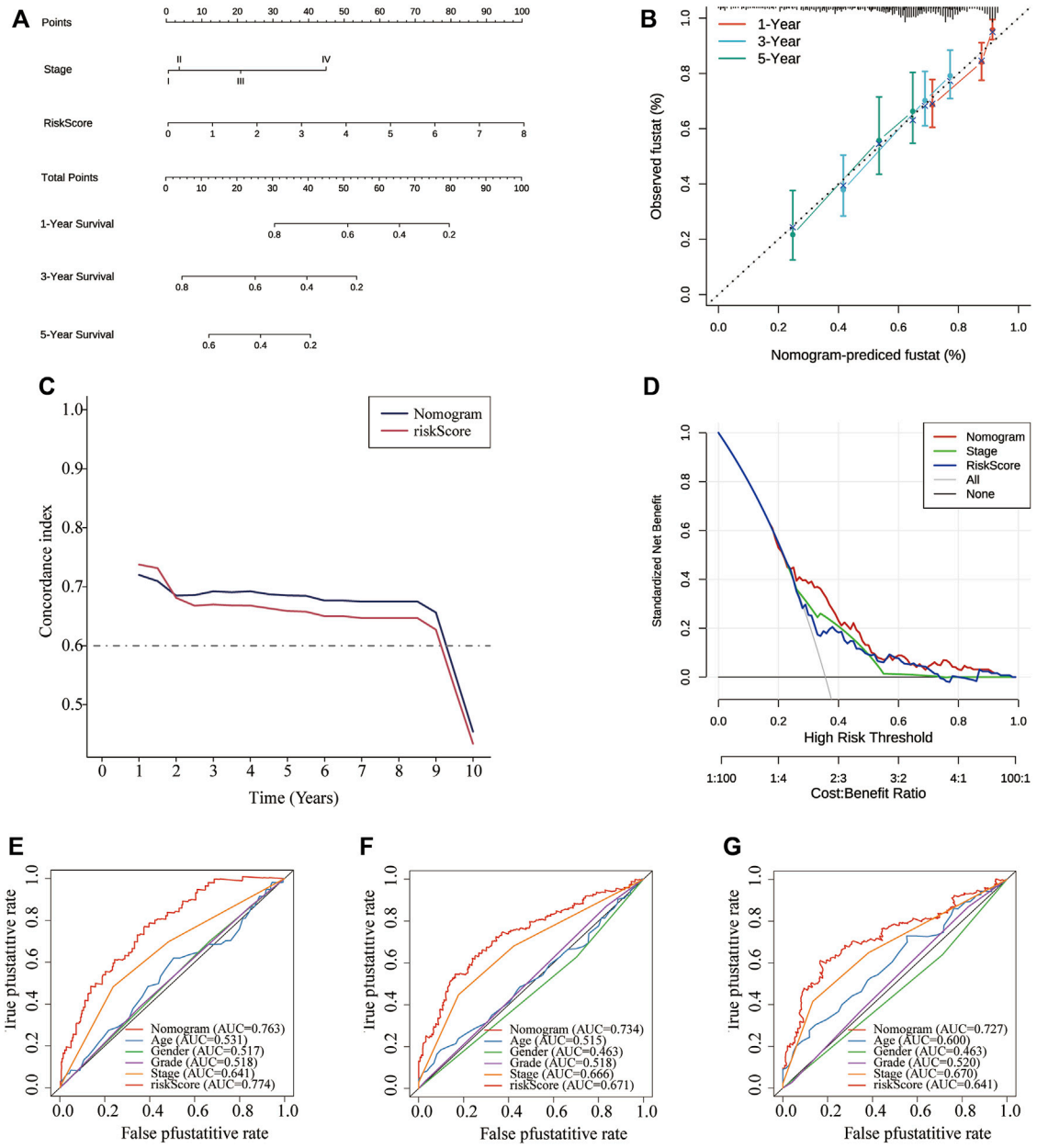


FIGURE 9
Construction of a nomogram based on independent prognostic factors for predicting OS in TCGA cohort. **(A)** The prediction performance of the nomogram for the OS at 1-, 3- and 5-years. **(B)** Calibration curves for the predicted 1-, 3-, and 5-years OS. **(D)** DCA curve. **(C)** Concordance index demonstrating the concordance measure of the predictor with patient survival. **(E–G)** ROC curves for the predicted 1-, 3-, and 5-years OS. DCA: decision curve analyses; LIHC: Liver hepatocellular carcinoma; OS: overall survival; ROC: Receiver Operating Characteristic.

Establishment and validation of the predictive nomogram

To make the prognostic signature more convenient for clinical application, we constructed a nomogram and tested its capacity to predict OS on the TCGA-LIHC cohort at 1-, 3-, and 5-years based on risk scores, and other signifcant independent risk factors (Figure 9A). The 1-, 3-, and 5-years OS calibration curves of the TCGA-LIHC data revealed that the nomogram had good predictive discrimination and accuracy (Figure 9B). Moreover,the nomogram had a higher consistency index compared to other clinical markers and the risk score (Figure 9C). In addition, Comparison of the net benefits of various models, such as none, risk score, all, nomogram, and

clinical indicators revealed that the nomogram had a higher net income and a wider threshold probability (Figure 9D). The analysis results revealed that the nomogram had a better prognostic ROC value compared to the 3-CSC-related-genes signature and other four clinical indicators, and it could predict OS for 1-, 3-, and 5-years (Figures 9E–G).Finally, the predictive value of nomogram was further validated in the ICGC dataset (Supplementary Figure S6). Altogether, results of ROC, DCA, calibration curve, and C index analyses indicated that the nomogram had better clinical benefits than risk scores based on the four 3-genes signatures alone.

Discussion

Herein, 456 cancer stem cell-related genes were identified in HCC tumor samples. The correlation of these genes with clinical outcomes was explored. Among them, three genes (*RAB10*, *TCOF1,* and *PSMD14*) were found to be significantly upregulated in tumor samples, and univariate Cox regression showed that they were also significantly associated with overall survival. Consequently, the three genes were used to construct a novel prognosis prediction model. The effectiveness of the model as a prognosis predictor was then confirmed using internal and external validation cohorts, with results showing that it was an independent variable of HCC prognosis.

Previous studies reported the biological functions of the three genes in cancers. Consistent with our results, the expression level of these genes was upregulated in human HCC (He et al., 2002; Lv et al., 2020; Wu et al., 2022). *RAB10* is a member of the RAS superfamily. It encodes a protein that functions as small GTPase and plays a crucial role in intercellular vesicular trafficking (Pereira-Leal and Seabra, 2000; Chua and Tang, 2018). Some studies have revealed that *RAB10* knockdown significantly suppressed proliferation of HCC cells both *in vitro* and in nude mice xenografts, suggesting that *RAB10* was involved in tumorigenesis (Wang et al., 2017). Furthermore, *RAB10* silencing could induce cell cycle arrest and apoptosis in HCC cells, thereby affecting a number of cellular signaling pathways (Wang et al., 2017). *TCOF1* encodes the protein named treacle, a nucleolar factor that can regulate the transcription of not only ribosomal DNA but even of DNA elsewhere in the genome (Valdez et al., 2004). It has been found to activate the transcription of oncogenes and increase ribosomal production in HCC cells, thereby promoting tumor growth (Wu et al., 2022). Wu *et al.* (2022) reported that expression of *TCOF1* can alter the immune microenvironment of HCC and induce the antitumor immune cell infiltration, which indirectly facilitates HCC progression. Furthermore, although not reported in HCC, the tumor-initiating capacity and self-renewal ability of triple-negative breast cancer cells were affected by *TCOF1* depletion, which indicates its significance in maintaining CSC stemness (Hu et al., 2022). *PSMD14* is a subunit of the 19S regulatory cap of the 26S proteasome that mediates substrate deubiquitination, a deubiquitinase belonging to the JAMM domain metalloprotease family. It has recently been identified as an oncogene, and has been shown to be associated with multiple solid tumors and poor prognosis (Luo et al., 2017; Song et al., 2017). Moreover, *PSMD14* has been found to interact with various signaling pathways in HCC, such as stabilizing E2F1 to hyperactivate its downstream pro-survival signaling thereby promoting cell proliferation (Wang et al., 2015); deubiquitinating TGF-β receptors (TGFBR1 and TGFBR2) and CAV1 to facilitate tumor metastasis (Wang et al., 2019), and protecting GRB2 from degradation to regulate HCC progression (Lv et al., 2020). In summary, these three genes modulate HCC tumorigenesis, and influence cancer progression and prognosis through complex mechanisms. Therefore, they can be use as risk factors for evaluating clinical outcomes in HCC patients.

In this study, patients were classified into two groups based on the 3-gene signature: high-risk or low-risk group. Experiments were conducted based to uncover the potential relationships between CSC and poor prognosis in HCC from a holistic point of view. Considering that most HCC cases were driven by chronic liver inflammation, the patients' immune status was of concern. GSVA results showed that that homologous recombination, cell cycle, RNA degradation, splicesome, and ubiquitin mediated proteolysis were positively enriched in TCGA and the ICGC cohort, which suggested that the dysregulation of these pathways was closely related to HCC development. The ssGSEA analysis found that high-risk patients tended to have higher infiltration levels of macrophages and Treg cells in TCGA and the ICGC cohort. Previous studies have demonstrated that increased tumor-associated macrophages (Zhang et al., 2019; Zhou et al., 2016) or Treg cells (Fu et al., 2007; Zhou et al., 2016) are related to poor prognosis in HCC patients due to their role in immune invasion. Of note, these immune cells have previously been identified as immunosuppressive cells in the tumor microenvironment that can inactivate anti-tumor immunity and help tumor cells escape immune attack, thereby promoting tumor growth. They have also been reported to be associated with poor prognosis of HCC patients (Dong et al., 2016; Fu et al., 2007; Lu et al., 2019). Moreover, some anti-tumor immune responses were impaired in the high-risk group, including decreased neutrophils and NK cells as well as the activity of type I and type II IFN response. In line with our results, a recent study by Dai et al. (2021) found that CSCs were involved in immune evasion, suggesting that they can be used as immunotherapeutic targets for HCC. Therefore, one reasonable explanation for the poor prognosis of patients in the high-risk group is the immunosuppressive

microenvironment created by cancer stem cells. Furthermore, GSVA results suggested that homologous recombination, cell cycle, RNA degradation, splicesome, and ubiquitin mediated proteolysis were positively correlated with risk score. In addition, the activity of many metabolism pathways, including glycine-serine and threonine metabolism, fatty acid metabolism, linoleic acid metabolism, and the PPAR signaling pathway, was inhibited in the high-risk group. These results suggest that dysfunction of these signaling pathways may contribute to cancer progression, which is worthwhile for future studies and provides new insights into the molecular mechanisms of tumorigenesis and the search for a HCC cure.

Finally, to determine key genes significantly associated with HCC prognosis, we performed WGCNA using DEGs identified from the TCGA dataset, with risk scores and risk groups as the traits. A gene module was identified from the constructed co-expression network. *BUB1B*, the first gene in the module, has been reported to promote HCC progression by activating the mTORC1 signaling pathway (Qiu et al., 2020). It has been found to have highly expressed in tumor tissues and HCC cell lines, and *BUB1B* knockdown significantly inhibited the proliferation, migration, and invasion of HCC cells (Qiu et al., 2020). Moreover, several bioinformatics analyses have considered it as a prognostic marker and potential therapeutic target for HCC (Fu et al., 2021; Yang et al., 2019b; Zhou et al., 2019; Zhuang et al., 2018). Therefore, we hypothesized that the second gene in the same module, *SGO2*, may play a similar oncogenic role as *BUB1*, although its biological function in HCC remains poorly understood. Results obtained after conducting the *in vitro* experiment suggested that *SGO2* plays a critical role in HCC cell proliferation, and knockdown of *SGO2* largely suppressed cell growth. However, further mechanistic studies should be performed to investigate the oncogenic role of *SGO2*, and explore its value as a new therapeutic target and prognostic marker of HCC.

In summary, this study identified three cancer stem cell related genes that were associated with poor prognosis in HCC patients. The three genes were used to construct a novel model for accurate and independent prediction of clinical outcome of HCC patients. Therefore, this model can be widely used to predict the prognosis of HCC patients, and provides valuable insights for improving patient outcomes. Moreover, this study demonstrates for the first time the tumor-promoting role of *SGO2* in HCC using cellular experiments. Collectively, our results provide a fundamental contribution to elucidating the pathogenesis of HCC as well as the search for new therapeutic targets in the future.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

X-YL and Z-HD conceived and designed the study. X-YP developed method. X-YL performed analyses and experiments. X-DZ revised the manuscript. M-YD and R-LD drafted the manuscript. All the authors read and agreed to the manuscript.

Funding

This study was supported by grants from project of Baise scientific research and technology development plan in 2021 (No. 20212347), and the Basic Ability Improvement Project for Young and Middle-aged Teachers in Colleges and Universities of Guangxi (Grant No. 2022KY0532).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.888601/full#supplementary-material

References

Chua, C., and Tang, B. L. (2018). Rab 10-a traffic controller in multiple cellular pathways and locations. *J. Cell. Physiol.* 233 (9), 6483–6494. [Journal Article; Research Support, Non-U.S. Gov't; Review]. doi:10.1002/jcp.26503

Dai, X., Guo, Y., Hu, Y., Bao, X., Zhu, X., Fu, Q., et al. (2021). Immunotherapy for targeting cancer stem cells in hepatocellular carcinoma. *Theranostics* 11 (7), 3489–3501. [Journal Article; Research Support, Non-U.S. Gov't; Review]. doi:10.7150/thno.54648

Dong, P., Ma, L., Liu, L., Zhao, G., Zhang, S., Dong, L., et al. (2016). CD86$^+$/CD206$^+$, diametrically polarized tumor-associated macrophages, predict hepatocellular carcinoma patient prognosis. *Int. J. Mol. Sci.* 17 (3), 320. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.3390/ijms17030320

El-Serag, H. B. (2012). Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 142 (6), 1264–1273. N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review]. doi:10.1053/j.gastro.2011.12.061

Fu, J., Xu, D., Liu, Z., Shi, M., Zhao, P., Fu, B., et al. (2007). Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients. *Gastroenterology* 132 (7), 2328–2339. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1053/j.gastro.2007.03.102

Fu, J., Zhang, X., Yan, L., Shao, Y., Liu, X., Chu, Y., et al. (2021). Identification of the hub gene BUB1B in hepatocellular carcinoma via bioinformatic analysis and *in vitro* experiments. *PeerJ* 9, e10943. doi:10.7717/peerj.10943

Ghouri, Y. A., Mian, I., and Rowe, J. H. (2017). Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog.* 16, 1

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1186/1471-2105-14-7

He, H., Dai, F., Yu, L., She, X., Zhao, Y., Jiang, J., et al. (2002). Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues. *Gene Expr.* 10 (5-6), 231–242. Gene Expr.; Research Support, Non-U.S. Gov't]. doi:10.3727/000000002783992406

Hu, J., Lai, Y., Huang, H., Ramakrishnan, S., Pan, Y., Ma, V., et al. (2022). TCOF1 upregulation in triple-negative breast cancer promotes stemness and tumour growth and correlates with poor prognosis. *Br. J. Cancer* 126 (1), 57–71. Br. J. Cancer; Research Support, Non-U.S. Gov't]. doi:10.1038/s41416-021-01596-3

Huang, A., Yang, X. R., Chung, W. Y., Dennison, A. R., and Zhou, J. (2020). Targeted therapy for hepatocellular carcinoma. *Signal Transduct. Target Ther.* 5 (1), 146. [Journal Article; Research Support, Non-U.S. Gov't; Review]. doi:10.1038/s41392-020-00264-x

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26 (8), 1364–1370. doi:10.1200/JCO.2007.12.9791

Jemal, A., Ward, E. M., Johnson, C. J., Cronin, K. A., Ma, J., Ryerson, B., et al. (2017). Annual report to the nation on the status of cancer, 1975-2014, featuring survival. *J. Natl. Cancer Inst.* 109 (9). doi:10.1093/jnci/djx030

Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Lee, T. K., Guan, X. Y., and Ma, S. (2022). Cancer stem cells in hepatocellular carcinoma - from origin to clinical implications. [Journal Article; Research Support, Non-U.S. Gov't; Review]. *Nat Rev Gastroenterol Hepatol.* 19 (1), 26–44.

Liang, Y., Su, Q., and Wu, X. (2020). Identification and validation of a novel six-gene prognostic signature of stem cell characteristic in colon cancer. *Front. Oncol.* 10, 571655. doi:10.3389/fonc.2020.571655

Lu, C., Rong, D., Zhang, B., Zheng, W., Wang, X., Chen, Z., et al. (2019). Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: Challenges and opportunities. *Mol. Cancer.* 18 (1), 130. Mol. Cancer; Research Support, Non-U.S. Gov't; Review]. doi:10.1186/s12943-019-1047-6

Luo, G., Hu, N., Xia, X., Zhou, J., and Ye, C. (2017). RPN11 deubiquitinase promotes proliferation and migration of breast cancer cells. *Mol. Med. Rep.* 16 (1), 331–338. doi:10.3892/mmr.2017.6587

Lv, J., Zhang, S., Wu, H., Lu, J., Lu, Y., Wang, F., et al. (2020). Deubiquitinase PSMD14 enhances hepatocellular carcinoma growth and metastasis by stabilizing GRB2. *Cancer Lett.* 469, 22–34. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1016/j.canlet.2019.10.025

Pereira-Leal, J. B., and Seabra, M. C. (2000). The mammalian rab family of small GTPases: Definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the ras superfamily. *J. Mol. Biol.* 301 (4), 1077–1087. doi:10.1006/jmbi.2000.4010

Qiu, J., Zhang, S., Wang, P., Wang, H., Sha, B., Peng, H., et al. (2020). BUB1B promotes hepatocellular carcinoma progression via activation of the mTORC1 signaling pathway. *Cancer Med.* 9 (21), 8159–8172. Cancer Med.; Research Support, Non-U.S. Gov't]. doi:10.1002/cam4.3411

Rumgay, H., Ferlay, J., de Martel, C., Georges, D., Ibrahim, A. S., Zheng, R., et al. (2022). Global, regional and national burden of primary liver cancer by subtype. *Eur. J. Cancer* 161, 108–118. doi:10.1016/j.ejca.2021.11.023

Song, Y., Li, S., Ray, A., Das, D. S., Qi, J., Samur, M. K., et al. (2017). Blockade of deubiquitylating enzyme Rpn11 triggers apoptosis in multiple myeloma cells and overcomes bortezomib resistance. *Oncogene* 36 (40), 5631–5638. [Journal Article; Research Support, N.I.H., Extramural]. doi:10.1038/onc.2017.172

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tsui, Y. M., Chan, L. K., and Ng, I. O. (2020). Cancer stemness in hepatocellular carcinoma: Mechanisms and translational potential. *Br. J. Cancer* 122 (10), 1428–1440. Br. J. Cancer; Research Support, Non-U.S. Gov't; Review]. doi:10.1038/s41416-020-0823-9

Valdez, B. C., Henning, D., So, R. B., Dixon, J., and Dixon, M. J. (2004). The Treacher Collins syndrome (TCOF1) gene product is involved in ribosomal DNA gene transcription by interacting with upstream binding factor. *Proc. Natl. Acad. Sci. U. S. A.* 101 (29), 10709–10714. [Journal Article; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, P.H.S.]. doi:10.1073/pnas.0402492101

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* 26 (6), 565–574. Med. Decis. Making; Research Support, N.I.H., Extramural]. doi:10.1177/0272989X06295361

Wang, B., Ma, A., Zhang, L., Jin, W. L., Qian, Y., Xu, G., et al. (2015). POH1 deubiquitylates and stabilizes E2F1 to promote tumour formation. *Nat. Commun.* 6, 8704. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1038/ncomms9704

Wang, W., Jia, W. D., Hu, B., and Pan, Y. Y. (2017). RAB10 overexpression promotes tumor growth and indicates poor prognosis of hepatocellular carcinoma. *Oncotarget* 8 (16), 26434–26447. doi:10.18632/oncotarget.15507

Wang, B., Xu, X., Yang, Z., Zhang, L., Liu, Y., Ma, A., et al. (2019). POH1 contributes to hyperactivation of TGF-β signaling and facilitates hepatocellular carcinoma metastasis through deubiquitinating TGF-β receptors and caveolin-1. *EBioMedicine* 41, 320–332. doi:10.1016/j.ebiom.2019.01.058

Wu, C., Xia, D., Wang, D., Wang, S., Sun, Z., Xu, B., et al. (2022). TCOF1 coordinates oncogenic activation and rRNA production and promotes tumorigenesis in HCC. *Cancer Sci.* 113 (2), 553–564. doi:10.1111/cas.15242

Yamashita, T., and Wang, X. W. (2013). Cancer stem cells in the development of liver cancer. *J. Clin. Invest.* 123 (5), 1911–1918. doi:10.1172/JCI66024

Yang, X. R., Xu, Y., Yu, B., Zhou, J., Qiu, S. J., Shi, G. M., et al. (2010). High expression levels of putative hepatic stem/progenitor cell biomarkers related to tumour angiogenesis and poor prognosis of hepatocellular carcinoma. *Gut* 59 (7), 953–962. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1136/gut.2008.176271

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019a). A global view of hepatocellular carcinoma: Trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* 16 (10), 589–604. N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review]. doi:10.1038/s41575-019-0186-y

Yang, W. X., Pan, Y. Y., and You, C. G. (2019b). CDK1, CCNB1, CDC20, BUB1, MAD2L1, MCM3, BUB1B, MCM2, and RFC4 may Be potential therapeutic targets for hepatocellular carcinoma using integrated bioinformatic analysis. *Biomed. Res. Int.* 2019, 1245072. doi:10.1155/2019/1245072

Zhang, Q., He, Y., Luo, N., Patel, S. J., Han, Y., Gao, R., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell.* 179 (4), 829–845. doi:10.1016/j.cell.2019.10.003

Zhou, S. L., Zhou, Z. J., Hu, Z. Q., Huang, X. W., Wang, Z., Chen, E. B., et al. (2016). Tumor-associated neutrophils recruit macrophages and T-regulatory cells to promote progression of hepatocellular carcinoma and resistance to sorafenib. *Gastroenterology* 150 (7), 1646–1658. doi:10.1053/j.gastro.2016.02.040

Zhou, Z., Li, Y., Hao, H., Wang, Y., Zhou, Z., Wang, Z., et al. (2019). Screening hub genes as prognostic biomarkers of hepatocellular carcinoma by bioinformatics analysis. *Cell. Transpl.* 28, 76S–86S. doi:10.1177/0963689719893950

Zhuang, L., Yang, Z., and Meng, Z. (2018). Upregulation of BUB1B, CCNB1, CDC7, CDC20, and MCM3 in tumor tissues predicted worse overall survival and Disease-Free survival in hepatocellular carcinoma patients. *Biomed. Res. Int.* 2018, 7897346. doi:10.1155/2018/7897346

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 19 September 2022
DOI 10.3389/fgene.2022.860067

Check for updates

OPEN ACCESS

EDITED BY
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

REVIEWED BY
Harpreet Kaur,
National Institutes of Health (NIH), United States
Lu Zhang,
Henan University, China

*CORRESPONDENCE
Bing Tang,
tangbing@mail.sysu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 22 January 2022
ACCEPTED 30 August 2022
PUBLISHED 19 September 2022

CITATION
Dong Y, Zhao Z, Simayi M, Chen C, Xu Z, Lv D and Tang B (2022), Transcriptome profiles of fatty acid metabolism-related genes and immune infiltrates identify hot tumors for immunotherapy in cutaneous melanoma.
Front. Genet. 13:860067.
doi: 10.3389/fgene.2022.860067

COPYRIGHT
© 2022 Dong, Zhao, Simayi, Chen, Xu, Lv and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Transcriptome profiles of fatty acid metabolism-related genes and immune infiltrates identify hot tumors for immunotherapy in cutaneous melanoma

Yunxian Dong[1†], Zirui Zhao[2†], Maijimi Simayi[3], Chufen Chen[1], Zhongye Xu[1], Dongming Lv[1] and Bing Tang[1*]

[1]Department of Burn and Plastic Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China, [2]Department of Gastrointestinal Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China, [3]Department of General Surgery, The First People's Hospital of Kashgar, Kashgar, China

Background: Immunotherapy with checkpoint inhibitors usually has a low response rate in some cutaneous melanoma (CM) cases due to its cold nature. Hence, identification of hot tumors is important to improve the immunotherapeutic efficacy and prognoses of CMs.

Methods: Fatty acid (FA) metabolism-related genes were extracted from the Gene Set Enrichment Analysis and used in the non-negative matrix factorization (NMF), copy number variation frequency, tumor mutation burden (TMB), and immune-related analyses, such as immunophenoscore (IPS). We generate a risk model and a nomogram for predicting patient prognoses and predicted the potential drugs for therapies using the Connectivity Map. Moreover, the NMF and the risk model were validated in a cohort of cases in the GSE65904 and GSE54467. At last, immunohistochemistry (IHC) was used for further validation.

Results: Based on the NMF of 11 FA metabolism-related DEGs, CM cases were stratified into two clusters. Cluster 2 cases had the characteristics of a hot tumor with higher immune infiltration levels, higher immune checkpoint (IC) molecules expression levels, higher TMB, and more sensitivity to immunotherapy and more potential immunotherapeutic drugs and were identified as hot tumors for immunotherapy. The risk model and nomogram displayed excellent predictor values. In addition, there were more small potential molecule drugs for therapies of CM patients, such as ambroxol. In immunohistochemistry (IHC), we could find that expression of *PLA2G2D*, *ACOXL*, and *KMO* was upregulated in CM tissues, while the expression of *IL4I1*, *BBOX1*, and *CIDEA* was reversed or not detected.

Conclusion: The transcriptome profiles of FA metabolism-related genes were effective for distinguishing CM into hot–cold tumors. Our findings may be valuable for development of effective immunotherapy for CM patients and for proposing new therapy strategies.

KEYWORDS

cutaneous melanoma, hot tumors, immunotherapy, precision medicine, fatty acid metabolism

1 Introduction

Cutaneous melanoma (CM) is the most malignant of all skin tumors (Bray et al., 2018). Most of the patients with CM are diagnosed already at clinically detectable stage III with or without in-transit metastases, which is called high-risk resectable melanoma, and suffer from a high-risk relapse (up to 70%) when treated with surgery alone (Gershenwald et al., 2017; Schadendorf et al., 2018). Conventional surgical procedures and chemotherapies are difficult to effectively cure CM due to chemotherapy failure and severe adverse effects (Amaria et al., 2019). As a result, patients diagnosed at an advanced stage of CM have an extremely poor prognosis, with a 5-year survival rate of less than 10% (Schadendorf et al., 2018; Allais et al., 2021). Over the past decade, immunotherapy, such as immune checkpoint inhibitors (ICIs), has significantly prolonged CM patients' overall survival (Hamid et al., 2019; Eggermont et al., 2021). Moreover, those ICIs usually have lower toxicity, high life quality, and treatment compliance in most CM patients. Unfortunately, the therapeutic efficacy and response rate in CM patients remain low. On the other hand, not all CM patients benefit from ICIs, although immunotherapy had achieved a lot. For example, programmed cell death protein 1(PD-1) inhibitor therapy usually has a response rate of one-third in CM (Galon and Bruni, 2019; Salmon et al., 2019; Tang et al., 2020).

The effectiveness of immunotherapy depends on the levels of circulating immune components in the body, $CD8^+$ T cell infiltration, and proliferative ability in the tumors (Galon and Bruni, 2019; Salmon et al., 2019; Lopez de Rodas and Schalper, 2021). According to the immune infiltration level of tumors, we could classify tumors into two categories, "hot" and "cold". The terms "hot" and "cold" are used to refer to T cell-infiltrated, inflamed but non-infiltrated, and non-inflamed tumors (Galon and Bruni, 2019). The hot tumors with high immune infiltration, particularly for $CD8^+$ T cells, usually are sensitive to immunotherapy because the pre-stored immune cells can effectively attack tumor cells following immunotherapy, such as checkpoint inhibitors (Ochoa de Olza et al., 2020; Too et al., 2021). Therefore, it is necessary to apply different treatment strategies for cold and hot tumors. Actually, early transformation of cold tumors into hot tumors can improve the efficacy of immunotherapy and prognosis of patients (Ji et al., 2021; Li et al., 2021).

It is well known that because of hypoxia and aggressive growth, the TME usually has high oxidation of fatty acid (FA) metabolism. In CM, lipid metabolism is associated with the resistance to targeted therapeutic drugs by altered expression of the FA transporter FATP2 (Alicea et al., 2020). A metabolic reprogramming to FA oxidation (FAO) can regulate the adaptation of BRAF-mutated melanoma to MAPK inhibitors (Aloia et al., 2019). A strong FAO in dendritic cells can attenuate therapeutic responses to anti-PD-1 treatment in melanoma by modulating Wnt5a-β-catenin-PPAR-γ signaling (Zhao et al., 2018). Moreover, the fatty acid receptor GPR120 may be a new marker for human melanoma (Oh et al., 2010; Kleemann et al., 2018).

However, the relationship between FA metabolism-related genes and immunotherapy remains largely unclear in CM. Moreover, there is little information on reliable biomarkers to distinguish cold from hot tumors, including CM (Maleki Vareki, 2018; Long et al., 2022). In this study, we analyzed the FA metabolism-related genes and immune infiltrates in CM and after determining their prognostic values, we generated and validated a risk model and nomogram. In addition, we screened some small-molecule drugs. Our findings indicated the risk model and clusters were valuable for prognosis and predicting immunotherapeutic responses in CM patients.

2 Materials and methods

2.1 Date preparation

The RNA-seq profiles of 471 CM and one non-tumor samples of the cancer genome atlas (TCGA),as well as 555 non-tumor skin specimens of the Genotype-Tissue Expression Project (GTEx), were obtained from the University of California Santa Cruz (UCSC). Batch normalization was performed in the data set by the sva R package. The differentially expressed genes (DEGs) were analyzed using Counts format profiles and converted into TPM format profiles for further analyses using the limma R package (Stupnikov et al., 2021). Their clinical data, such as overall survival (OS), and copy number, were also downloaded from UCSC. Furthermore, RNA-seq profiles, survival status, and OS time of CM patients were obtained from the GSE65904 and GSE54467 datasets for external validation from Gene Expression Omnibus (GEO). To reduce statistical bias, CM samples with missing OS values or short OS values (< 30 days) were excluded from all cohorts. As a result, 447 patients in the TCGA cohort and 278 in the external validation cohort were used for analyses. In addition, the fatty acid (FA) metabolism-related gene sets, including M14568, M22474, M34207, M34208, M29237, M23047, M15179, M16838, M25445, M23048, M13605, M11966, M15385, M16969, M14401, M34091, M13290, M11936, M14177, M12558, M16551, M13480, M15568, M18199, M26370, M26153, M40674, M18978, M26866, M26251, M5935, M36310, M37191, M11673, M29570,

M27727, M27719, M14690, M22174, M6999, M13591, M10250, M15813, M25014, M16181, M12334, M17829, M25044, M15938, M23782, M22457, M40405, M40498, M699, M6995, M9927, M15531, M27854, M39440, and M39596, containing 745 fatty acid metabolism-related genes were extracted. Their intersection genes were identified in the TCGA, GTEx, and external validation cohorts using the VennDiagram R package. Moreover, the tumor mutation burden data of these cohorts were obtained from the TCGA database.

2.2 FA DEGs and NMF

FA metabolism-related DEGs between the CM and non-tumor samples were identified, based on the criteria of a false discovery rate (FDR) < 0.05 and an absolute value Log2 fold change >1, using the limma R package in the TCGA and GTEx synthetic Counts matrix (Deng et al., 2021). This synthetic matrix had been normalized for identifying DEGs. Subsequently, 11 FA metabolism-related DEGs were identified using univariate Cox proportional hazard regression, and they had significant prognostic values (all $p < 0.05$) using the limma and survival R packages in the TCGA cohort (Zheng et al., 2021). These 11 FA metabolism-related DEGs with potent prognostic values were subjected to non-negative matrix factorization (NMF) analyses for sample clustering by NMF and survival R packages (Zhuo et al., 2020). The Kaplan–Meier survival curves of OS, t-distributed stochastic neighbor embedding (t-SNE), and Principal Component Analyses (PCA) of the clusters were analyzed using the survminer, Rtsne, ggplot2, and scatterplot3d R packages. Similar analyses were performed in the external validation cohort.

2.3 Copy number alterations

The copy number variation (CNV) frequencies of these FA metabolism-related prognostic genes were calculated, and their locations in human chromosomes were identified using the RCircos R package. For each prognostic gene, we compared the immune infiltration levels of CM patients with different somatic copy number alterations, such as deep deletion, arm-level deletion, and diploid/normal, using the Tumor Immune Estimation Resource (TIMER) (Kang et al., 2020).

2.4 Functional analysis

The potential functions and pathways of each set of genes in the cluster were analyzed using the Gene Set Enrichment Analyses (GSEA) software (https://www.gsea-msigdb.org/gsea/login jsp) and Curated gene set (kegg. v7. 4. symbols. gmt). Multi-GSEA diagrams were drawn using the plyr, grid, and gridExtra R packages.

2.5 Evaluation of immune infiltration

The immune infiltration of individual patients was analyzed using the CIBERSORT (R scrip v 1.03), and their immune scores, stromal scores, ESTIMATE (microenvironment) scores, and tumor purity were compared using the estimate R package, followed by visualizing them as the pheatmap using the ggtext packages (Gui et al., 2021). Subsequently, the immune cell infiltration, immune functions, and expression of genes for checkpoints were compared using the GSVA, GSEABase, ggpubr, reshape2, and limma R packages.

2.6 Tumor mutation burden analysis

The tumor mutation burden (TMB) data in the "Masked Somatic Mutation" type were processed by VarScan2 and analyzed in clusters using the maftools package (Koboldt et al., 2012). Furthermore, their overall survival was estimated by K–M survival analysis, and a box plot was made using the survminer, survival, ggplot2, ggpubr, and ggExtra R packages.

2.7 Immunotherapy-related exploration

According to the results of comprehensive immunogenomic analyses in The Cancer Immunome Atlas (TCIA), the response of each patient to immunotherapy was predicted and compared using the ggpubr R package (Charoentong et al., 2017). Additionally, the potential immunotherapeutic function of some drugs and their half-maximal inhibitory concentration (IC_{50}) in CM patients were predicted based on the data from Genomics of Drug Sensitivity in Cancer (GDSC) using the pRRophetic R package (Gui et al., 2021).

2.8 Construction and assessment of the risk model and nomogram

CM patients were randomly divided into training and testing sets. A risk model was generated using the least absolute shrinkage and selection operator (LASSO) in the glmnet R package. The risk score formula (Zheng et al., 2021):

$$\text{Riskscore} = \sum_{k=1}^{n} \text{coef}\left(\text{gene}^{k}\right) * \text{expr}\left(\text{gene}^{k}\right)$$

The coef (gene) meant the coefficient of the gene in the risk model, and expr (gene) was the expression of the gene in the risk model. The status, survival time, heatmaps, Kaplan–Meier survival analyses, and receiver operating characteristic (ROC) curves of individual patients in the training, testing, and entire sets were also analyzed using the pheatmap, survival, and timeROC R packages. For validation, a risk score of each

patient in the external validation cohort was calculated, and their survival was estimated by the Kaplan–Meier survival analysis. The specificity and sensitivity of this risk model were evaluated by ROC analysis.

According to patient's clinical data and risk scores, the independent prognostic factors for worse survival were identified using univariate Cox and multivariate Cox regression analyses and used for the generation of a nomogram using the nomogramEx and nomogramFormula R packages. The concordance of the nomogram was analyzed by the ROC curves using the timeROC and rms R packages.

2.9 Connectivity Map

For predicting potential small-molecule drugs that might reverse high risk in CM, the whole overlapping genes, including the upregulated and downregulated genes, were submitted into the CMap database (https://portals.broadinstitute.org/cmap/). The drugs with enrichment scores between −1 and 0 were considered candidate drugs for CM (all $p < 0.05$) (Gui et al., 2021).

2.10 Validation of protein expressions of FA metabolism-related genes by the Human Protein Atlas

2.10.1 Database

The protein expression of the FA metabolism-related genes between CM and normal tissues was determined using immunohistochemistry (IHC) from the Human Protein Atlas database (HPA) as well as our own preserved patients' tissue paraffin slides (Thul and Lindskog, 2018).

2.11 Histopathology

The tumor tissues were fixed in 10% formalin overnight and paraffin-embedded. The tissue sections (5 μm) were regularly stained with immunohistochemistry (IHC) staining. The stained tissue sections were photo-imaged and observed under a light microscope. The primary antibodies were used at 1:200 for *CIDEA* (Abcam, ab191193), *ACOXL* (Proteintech, 23366-1-AP), *PLA2G2D* (Abcam, ab47118), and 1:400 for *KMO* (Abcam, ab233529).

2.12 Statistical analysis

We made analyses with R version 3.6.3 (http://www.R-project.org) and its appropriate packages. All involved packages were described in MATERIALS AND METHODS. Data were analyzed with standard statistical tests as appropriate, while multiple testing was adjusted by the FDR method by R (Gui et al., 2021). Statistical significance was observed when $p < 0.05$.

3 Result

3.1 FA metabolism-related differentially expressed genes and NMF clusters

A total of 745 FA metabolism-related genes were identified in CM from the GSEA (Supplementary Table S1). Similarly, 533 of them were identified in CM from the GTEx, TCGA, GSE65904, and GSE54467 (Figure 1A). There were 51 DEGs in the TCGA cohort. Of them, 17 were upregulated and the others were downregulated (Figure 1B). Further analyses indicated that 11 DEGs were associated with the prognosis of CM (all $p < 0.05$, Figure 1C). Of them, *ALOX12B*, *CYP4F3*, *ALDH3A1*, *CIDEA*, and *BBOX1* were upregulated in CM, while the others were downregulated.

Based on comprehensive correlation coefficients and all heatmaps, we found the optimal total cluster number was set to k = 2 (Figures 1D,E). The heatmap (k = 2) indicated a clear boundary (Figure 1F). Compared with cluster 2, patients in cluster 1 had a worse OS (Figure 1G). To further verify the cluster distribution, the t-SNE, 2D PCA, and 3D PCA analyses clearly separated these prognostic genes (Figures 1H–J).

3.2 External verification of NMF clusters and analyses of copy number in CM

In the external verification cohort, the optimal total cluster number was k = 2 in NMF. The difference between the two clusters and their NMF ranks are displayed (Figure 2A, Supplementary Figures S1A,B). In addition, the t-SNE, PCA, and Kaplan–Meier survival curves of OS analyses revealed their distribution, and patients in cluster B had a worse OS in this population (Figures 2B–E).

Next, we calculated the CNV frequency of 11 prognostic genes and located them in human chromosomes. The percentages of gain CNV of *KMO, KCNJ10, CYP4F3,* and *BBOX1* were higher than those of the loss, while the frequency of gain CNV of *PLA2G2D, GPR31, IL4I1, ACOXL, ALDH3A1, ALOX12B,* and *CIDE*A was lower than that of loss (Figure 2F). As a result, the *KMO, KCNJ10, CYP4F3,* and *BBOX1* genes were marked in red, and the *PLA2G2D, GPR31, IL4I1, ACOXL, ALDH3A1, ALOX12B,* and *CIDEA* were marked in blue on RCircos 2D track plots (Figure 2G). Furthermore, we explored the changes in immune cell infiltration with a copy number alteration of prognostic genes in CM. Alterations of these genes were associated with infiltration levels of $CD8^+$ T cells or other

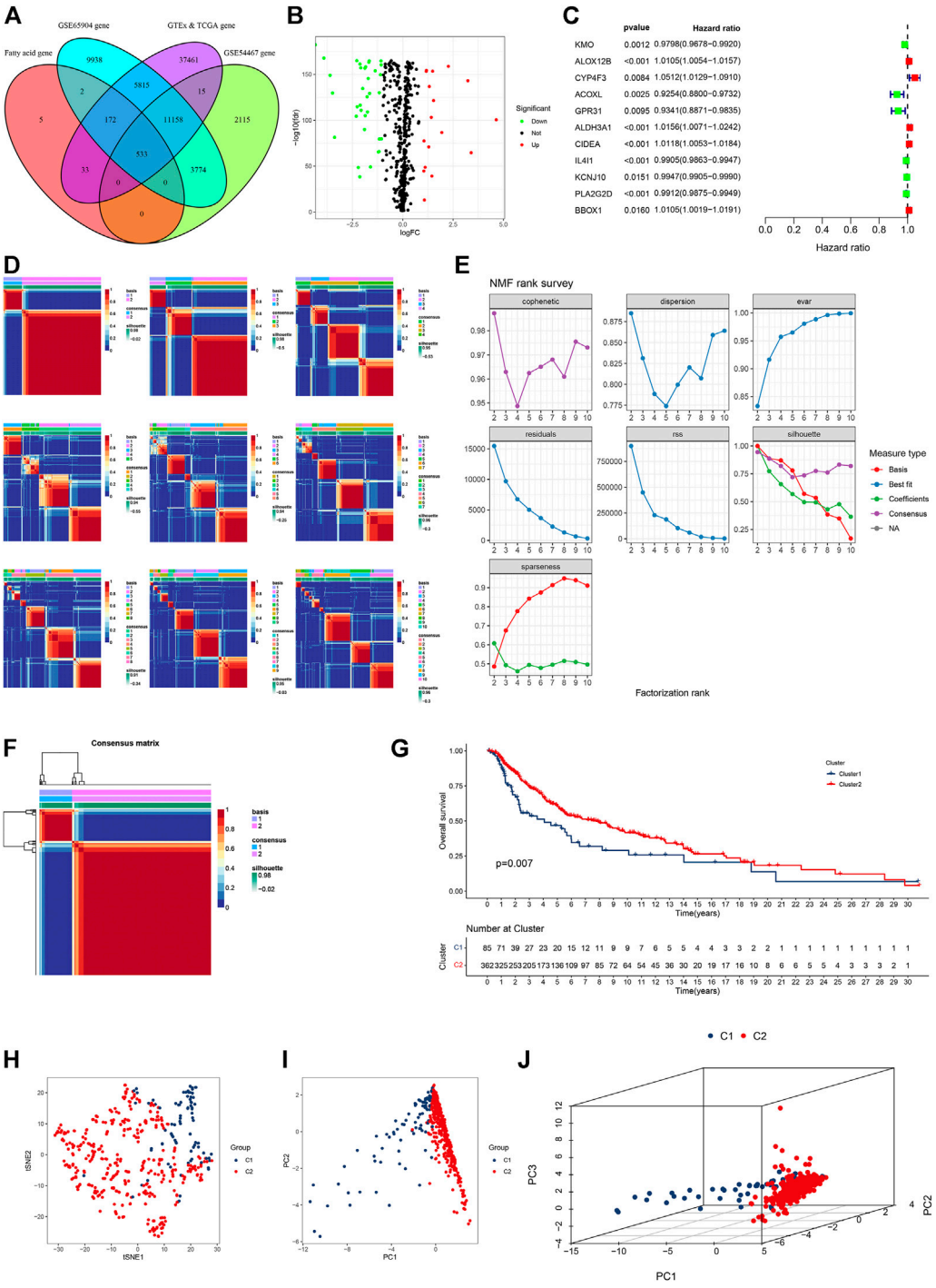


FIGURE 1
Identification of fatty acid metabolism-related genes in four gene subsets and clusters in the TCGA CM cohort. **(A)** Venn diagram displayed FA metabolism-related genes in CM cases from TCGA, GTEx, GSE65904, and GSE54467. **(B)** Volcano plot of 51 FA metabolism-related DEGs. **(C)** Forest plot of prognostic FA metabolism-related DEGs. **(D)** All the heatmaps of NMF consensus clustering. **(E)** NMF rank survey of cophenetic, dispersion, evar, residuals, rss, silhouette, and sparseness coefficients. **(F)** Heatmap of two clusters of CM. **(G)** Kaplan–Meier survival curves of OS in these clusters of CM cases. **(H–J)** t-SNE, PCA, and 3D PCA separated two clusters of CM. A *p* value of <0.05 was considered to indicate a statistically significant difference.

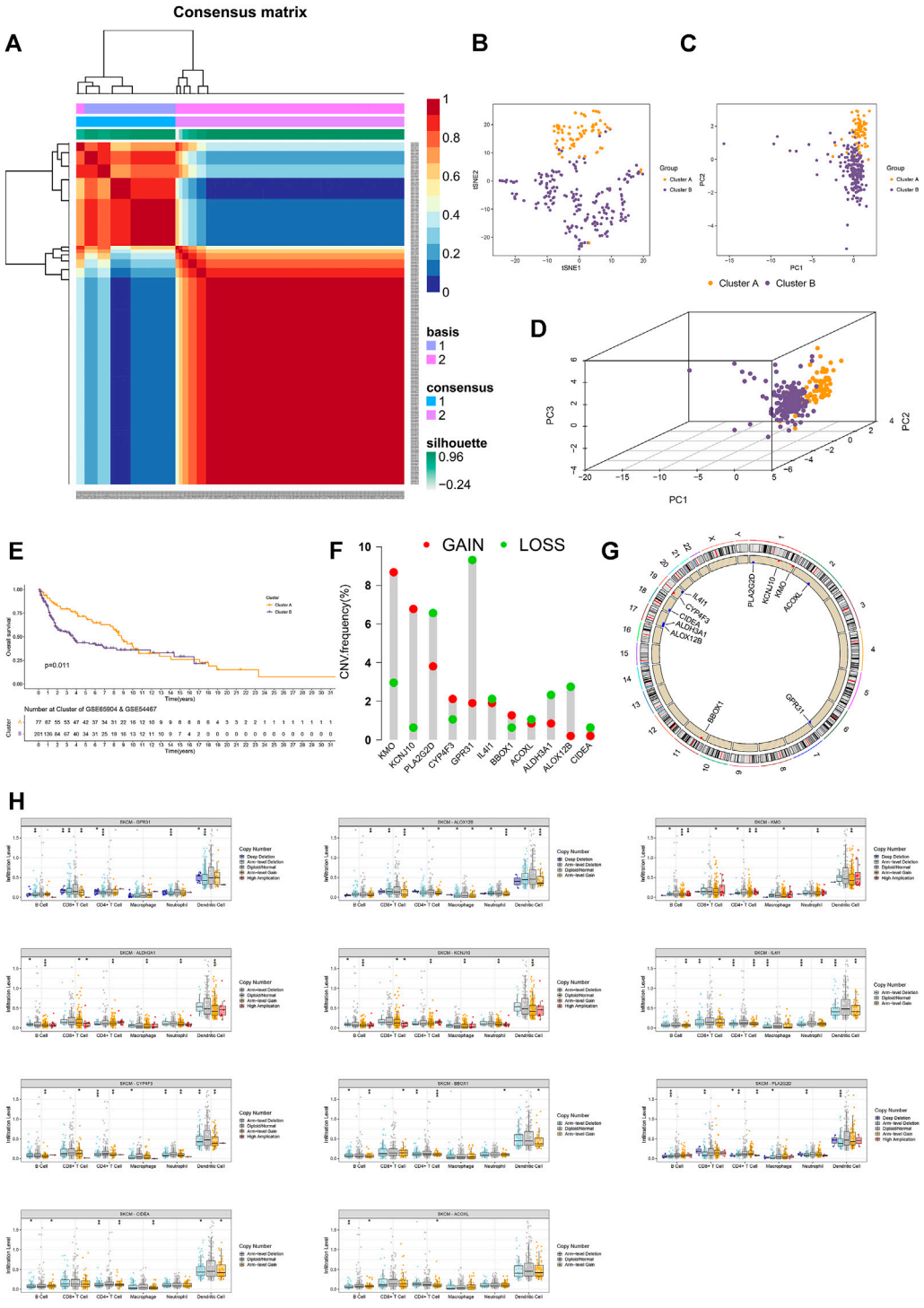


FIGURE 2
External validation and copy number analyses. **(A)** Heatmap of two clusters of cases in the external validation cohort. **(B–D)** t-SNE, PCA, and 3D PCA separated two clusters of CM cases in the external validation cohort. **(E)** Kaplan–Meier survival curves of OS in two clusters of CM cases in the external validation cohort. **(F)** CNV frequency of 11 FA metabolism-related genes. **(G)** CNV of prognostic genes on RCircos 2D track plot with the human genome. **(H)** Different types of immune infiltrates among samples with copy number of the indicated genes. Ns means no significant difference, *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. A p value of <0.05 was considered to indicate a statistically significant difference.

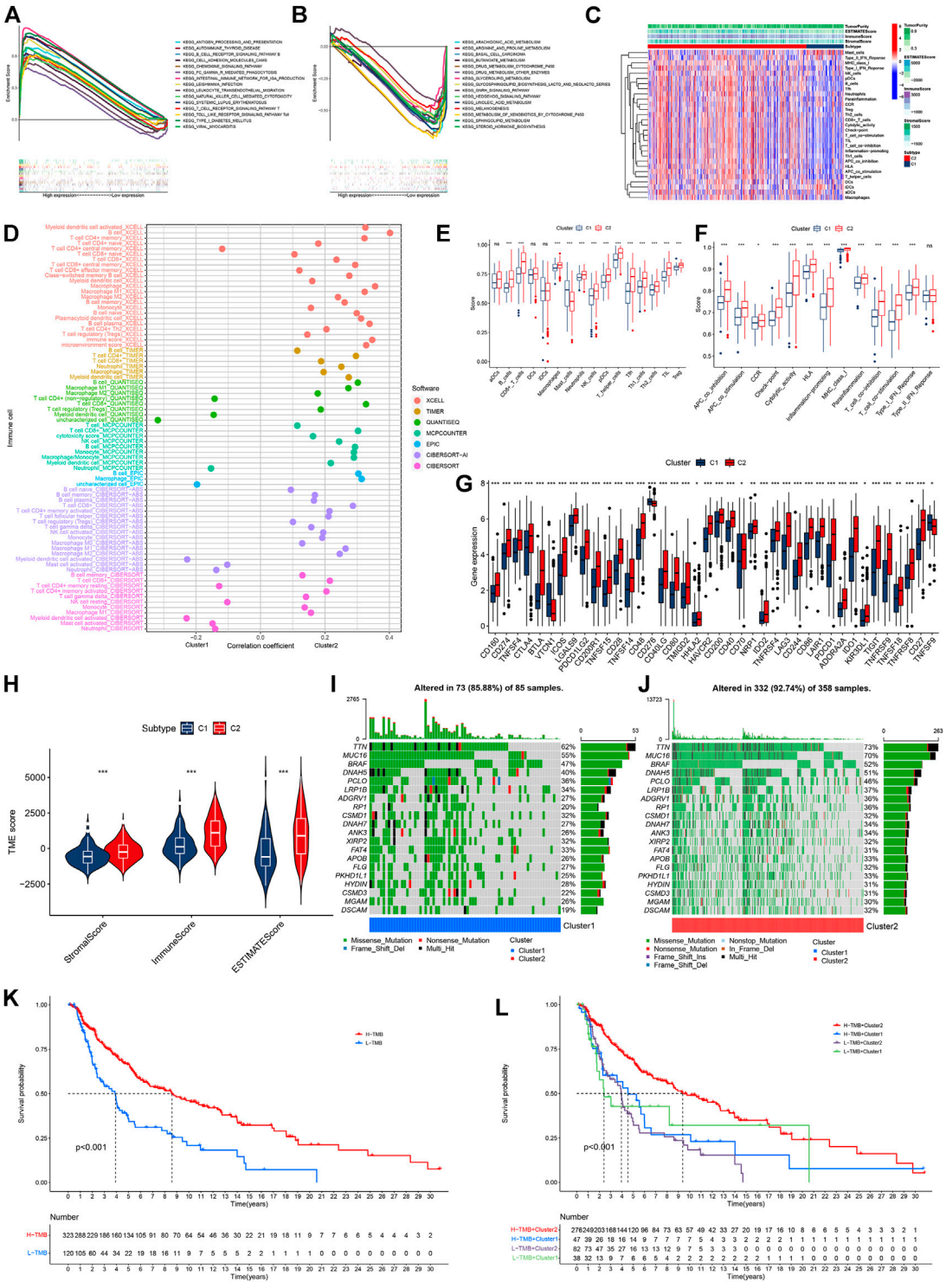


FIGURE 3
Analyses of tumor immune characteristics and TMB in these two clusters of CM cases. **(A)** Multi-GSEA analyses of cluster 2 CM cases. **(B)** multi-GSEA analyses of cluster 1 CM cases. **(C)** Heatmap of immune infiltrates between two clusters of CM cases. **(D)** Correlation coefficient of immune infiltrates in two clusters of CM cases. **(E,F)** Single-sample GSEA scores of immune cells and immune functions in two clusters of CM cases. **(G)** Comparisons of genes for expression of 40 checkpoints between two clusters of CM cases. **(H)** Comparisons of immune-related scores in two clusters of CM cases. **(I,J)** Waterfall plot of TMB of individual patients in cluster 1 and cluster 2. **(K)** Kaplan–Meier analysis of OS between the low- and high-TMB groups of CM cases. **(L)** Kaplan–Meier analysis of OS among four groups of CM cases. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. A p value of <0.05 was considered to indicate a statistically significant difference.

immune cells (Figure 2H). These indicated that clusters of CM with varying prognostic genes had different immune microenvironments, leading to different responses to immunotherapy (O'Sullivan et al., 2019).

3.3 Analyses of the immune microenvironment and TMB in CM

We performed GSEA to explore the biological functions of these clusters. Cluster 2 and cluster 1 of genes were involved in the top 15 pathways (all $p < 0.05$, FDR <0.05, |NES|>1.8, Supplementary Figures S1C,D). In cluster 2, almost all of the enriched pathways were associated with immunity, such as chemokine signaling, and natural killer cell-mediated cytotoxicity (Figure 3A) (Chow and Luster, 2014; Taniguchi and Karin, 2018). On the contrary, six out of the top 15 pathways enriched by the genes in cluster 1 were associated with FA metabolism, such as arachidonic acid metabolism, four pathways were related to tumor growth, and two were related to drug metabolism (Figure 3B). According to the heatmap of immune cell infiltration, we felt that the cluster 2 had a higher immune cell infiltration level, lower tumor purity, and more active immunity (Figure 3C). In the immune cell bubble chart, cluster 2 of CM had more immune cell infiltrates (Figure 3D, Supplementary Data Sheet S1). Based on the single-sample GSEA scores for immune cells and immune functions, 13 types of immune cells, such as $CD8^+$ T cells, and 12 immune functions, such as inflammation-promoting and check-point, had a higher score in cluster 2 of CM (ns means no significant difference, $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$, Figures 3E,F). Of course, almost all immune checkpoint genes, such as *CD274 (PD-L1), CTLA4, HAVCR2 (TIME3),*and *LAG3*, displayed a higher activation in cluster 2 (Figure 3G) (Galon and Bruni, 2019). Cluster 2 had a high immune score, stromal score, and ESTIMATE score (microenvironment score, Figure 3H). Therefore, cluster 2 of CM was considered the hot tumor to respond to immunotherapy because cluster 2 of CM displayed the characteristics of hot immune tumors, such as a high degree of $CD8^+$ T cells, high immune score, more active immune functions, and higher expression of *CTLA4*, *TIM3,* and *LAG3* (Galon and Bruni, 2019; Zheng et al., 2021). Previous studies have shown that a high TMB is significantly associated with improved OS and benefits from immunotherapy, such as *CTLA-4* blockade, and the TMB has been considered a potential immunotherapy parameter (Snyder et al., 2014; Mahoney et al., 2015; Chan et al., 2019). Hence, we analyzed the TMB of CM in the TCGA cohort to explore the responses of these clusters of CM to immunotherapy. The TMB frequency in cluster 2 of CM (92.74%) was higher than that of cluster 1 (85.88%) (Figures 3I,J). Consistently, the high-TMB group (H-TMB) of CM patients had better OS than those in the low-TMB group (L-TMB) (Figure 3K). Stratification analysis revealed that CM patients with high TMB displayed a better OS than those with low-TMB in both clusters (Figure 3L).

3.4 Investigation in immunotherapy

Compared with cluster 1, CM in cluster 2 possessed significantly higher TMB (Figure 4A). Accordingly, we investigated immune checkpoint-related scores in these clusters of CM. Compared with cluster 1 CM, cluster 2 CMs had significantly higher both *PD-1* and *CTLA4* immunophenoscore (IPS, Figure 4B). In addition, CM patients in cluster 2 were likely to be more sensitive to 12 potential immunotherapy-related drugs with lower IC_{50} than those in cluster 1 (Figure 4C) (Haikala et al., 2019; Song et al., 2019).

3.5 Risk model and external verification

With the LASSO regression analyses, a risk model of seven genes was established after control of the first-rank value of Log(λ) at the minimum likelihood of deviance (Figures 5A,B). The risk score formula was Risk score = KMO × (−0.0032) + CYP4F3 × (−0.0054) + ACOXL × (−0.0220) + CIDEA × 0.0135 + IL4I1 × (−0.0039) + PLA2G2D × (−0.0033) + BBOX1 × 0.0048. In addition, analyses of the survival status, survival time, expression of seven genes, and OS of patients clearly separated between low- and high-risk groups of CM patients in the training, testing, and entire sets of cases (Figures 5C–N). Hence, CM patients in the low-risk group displayed a better prognosis. The AUC for 1-, 2- and 3-year OS in the training set of cases was 0.671, 0.703, and 0.701, that of the testing set of cases was 0.664, 0.665, and 0.654, and that of the entire sets of cases was 0.663, 0.682, 0.676, respectively (Figures 5O–Q). After analyzing the clinical characteristic of the patients, we found our risk model was also available in age, gender, tumor stage, T stage. and N stage (Supplementary Figure S2A). Furthermore, stratification of CM patients was carried out into low- and high-risk groups, and the t-SNE, 2D PCA, and 3D PCA clearly separated them (Supplementary Figures S3A). Further validation revealed that CM patients in the low-risk group exhibited significantly better OS than those in the high-risk group ($p = 0.002$, Figure 5R). The AUC for 1-, 2-, and 3-year OS in the external validation cohort were 0.668, 0.684, and 0.690, respectively (Figure 5S).

3.6 Construction and assessment of a nomogram

The univariate Cox (uni-Cox) and multivariate Cox (multi-Cox) regression analyses indicated that patient's age, tumor T stage, tumor N stage, and risk score were independent risk factors

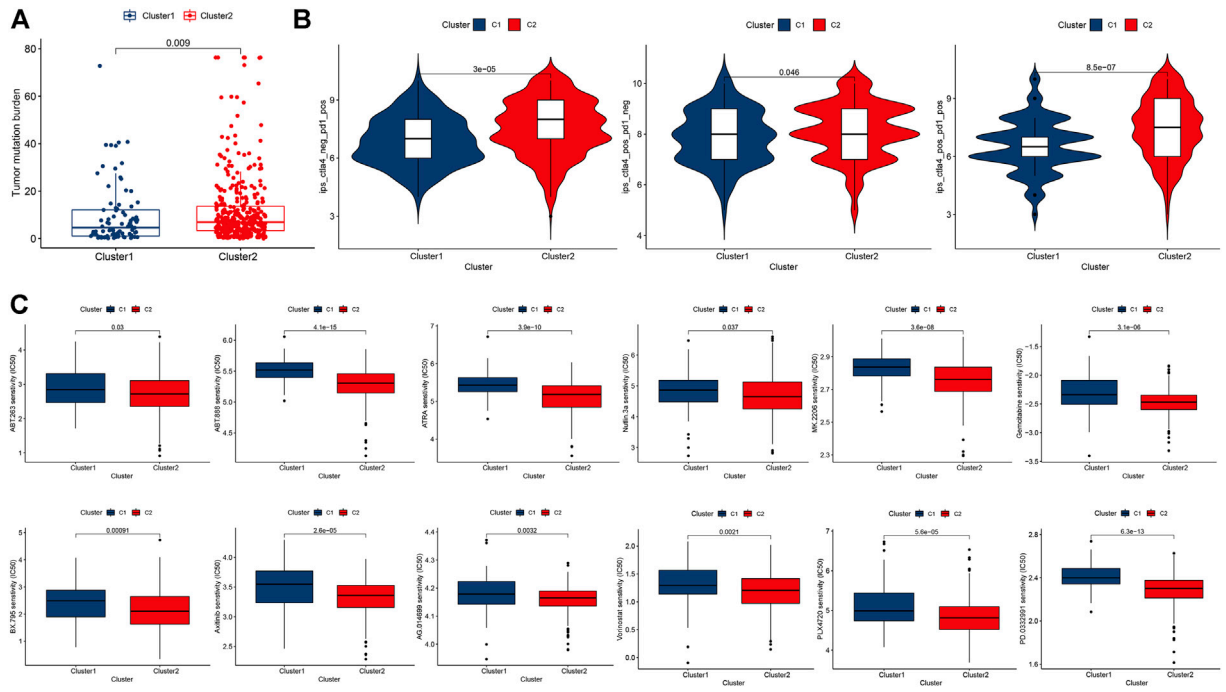


FIGURE 4
Comparisons of TMB, immune checkpoint gene expression, and drug sensitivity between two clusters of CM cases. **(A)** Levels of TMB in two clusters of CM cases. **(B)** *PD-1* and/or *CTLA4* in two clusters of CM cases. **(C)** Prediction of potential therapeutic drugs IC_{50} in two clusters of cases. A p value of <0.05 was considered to indicate a statistically significant difference.

for prognosis of CM patients. Their independent risk factors were associated with a worse prognosis for CM patients with hazard ratios and a 95% confidence interval (CI) (Figures 6A,B). With these four independent prognostic indexes, a nomogram was generated (Figure 6C). The AUC of a nomogram for 1-, 2-, and 3-year OS of CM patients was 0.766, 0.816, and 0.816, respectively (Figure 6D). The calibration plots of the nomogram for 1-, 2-, and 3-year OS of CM patients had a good concordance with the prediction (Figure 6E).

3.7 The potential value of the risk model in clinical application

Among the top 15 signaling pathways, 11 signaling pathways were associated with immunity, such as the T cell receptor signaling (all $p < 0.05$, FDR <0.05, |NES|>2.3) (Supplementary Figure S3B) (Courtney et al., 2018). The immune heatmap exhibited that CM in the low-risk group had more types of immune cells (Supplementary Figure S3C, Supplementary Data Sheet S2). The levels of $CD8^+$ T, B cell infiltrates, and other types of immune cells were also correlated with lower risk scores (Supplementary Figures S3D). Furthermore, CM in the low-risk group had higher immune scores (Supplementary Figures S3E). Thus, we tried to explore potential immunotherapy-related drugs for these risk groups of CM. As a result, we found that four drugs, such as metformin, had significantly different IC_{50} between both risk groups of CM (Figure 7A) (O'Sullivan et al., 2019; Cha et al., 2018). A total of five most related small-molecule drugs, ambroxol, tiletamine, mimosine, esculetin, and pizotifen, were identified (all $p < 0.05$), based on upregulated and downregulated gene expression between the low- and high-risk groups of CM (Supplementary Data Sheet S3) (Gui et al., 2021). Their 2D and 3D structure tomography are is in Figures 7B,C.

3.8 Verification of the protein expression of FA metabolism-related genes in normal skin and CM tissue

Finally, we further validated the expression of these key genes in normal skin and CM tissues. Melanoma arises from melanocytes in the epidermal layer of the skin, so we focused on the expression of these proteins in the epidermis. First, we searched the immunohistochemical slide information through the public database on the HPA website and found that the expression of *IL4I1* in normal skin epidermis was higher than that in CM tissue, while *BBOX1* protein was not detected in normal skin (Figures 8A,B). Next, our own

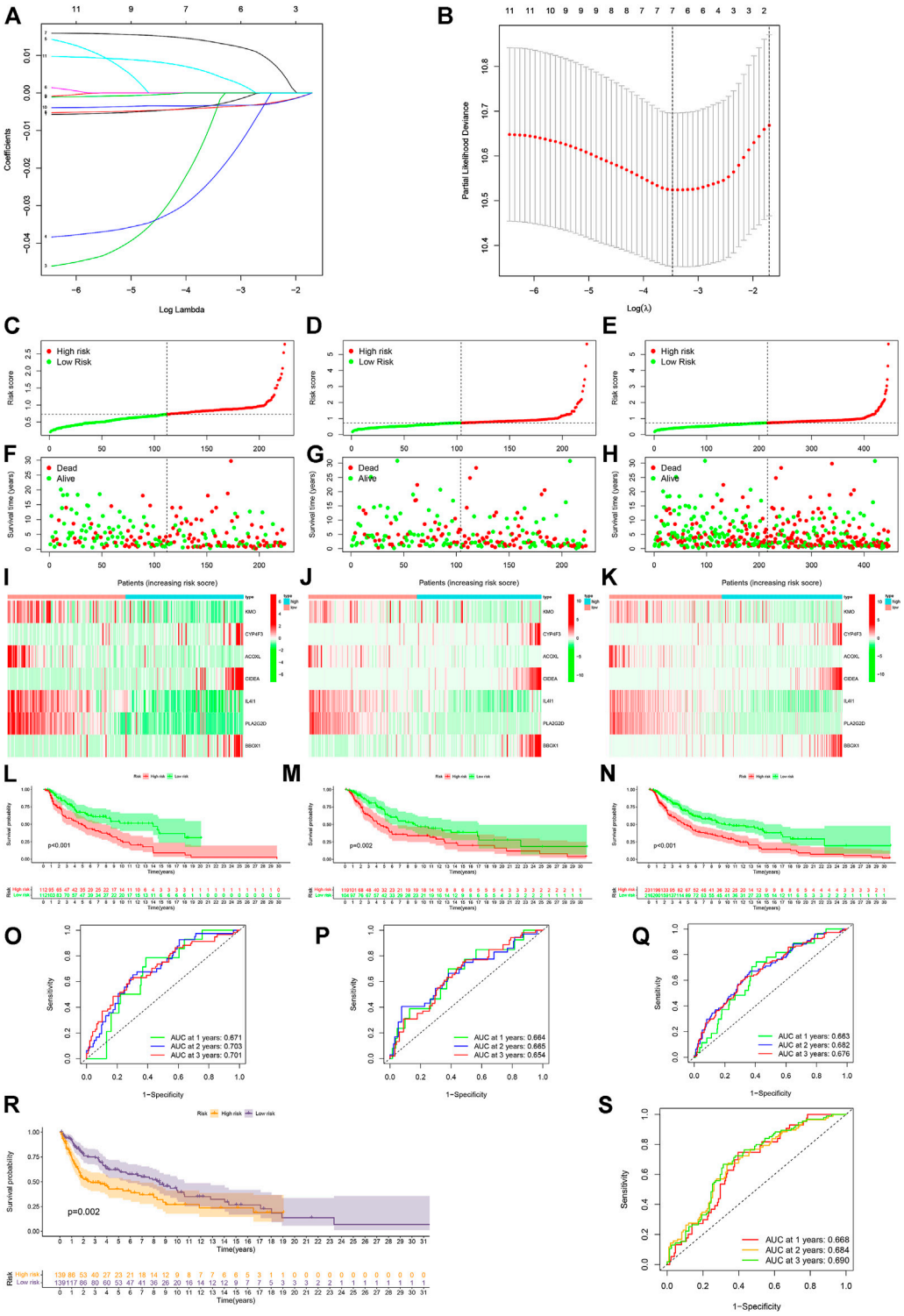


FIGURE 5
Construction and validation of the risk model. **(A,B)** Constructing a risk model of seven genes by LASSO regression. **(C–E)** Risk scores of the training, testing, and entire sets, respectively. **(F–H)** Survival status of individual cases between the low- and high-risk groups in the training, testing, and entire sets. **(I–K)** Heatmap of seven gene expression in the training, testing, and entire sets of CM cases. **(L–N)** Kaplan–Meier analysis of OS between the low- and high-risk groups of CM cases in the training, testing, and entire sets. **(O–Q)** ROC curves for 1-, 2- and 3-year OS of CM cases in the training, testing, and entire sets. **(R)** External validation of the risk model. **(S)** ROC curves for 1-, 2- and 3-year OS of CM case in the external validation cohort. A *p* value of <0.05 was considered to indicate a statistically significant difference.

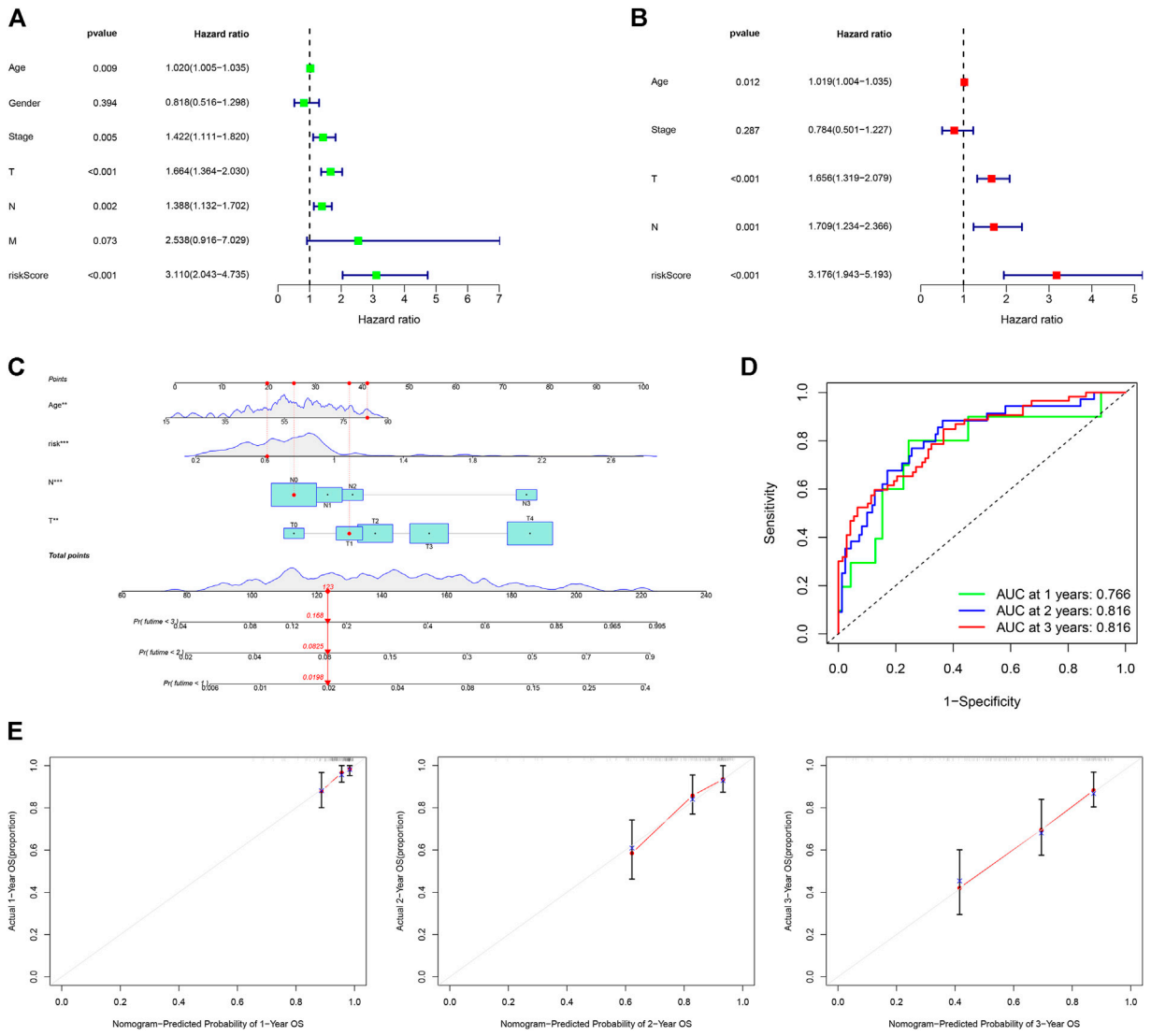


FIGURE 6
Construction and assessment of a nomogram. **(A,B)** Univariate and multivariate Cox analyses of clinical characteristics to identify independent risk factors for worse OS. **(C)** A nomogram based on age, T, N stage, and risk scores. **(D)** ROC curves for 1-, 2- and 3-year OS of CM cases using the nomogram. **(E)** Calibration ROS curves for 1-, 2- and 3-year OS.

immunohistochemical results showed that the expression of *PLA2G2D*, *ACOXL*, and *KMO* was higher in normal epidermal tissue than in CM tissue; however, the expression of *CIDEA* was reversed (Figures 8C–F).

4 Discussion

Currently, immunotherapy has been widely used in CM and achieved a good therapeutic effect in some CM patients. However, the therapeutic response rate of immunotherapies, particularly checkpoint inhibitors for CM patients remains low. It has been demonstrated that hot tumors usually respond to immunotherapies because they contain a lot of $CD8^+$ T cell infiltrates with a higher immune score, high expression of immune checkpoints (e.g., *PD-1* and *CTLA4*), and active inflammatory response. On the other hand, cold tumors generally do not respond well to immunotherapy although switching cold tumors into hot tumors by promoting immune cell infiltration into the tumor environment is feasible (Herrera et al., 2021). Therefore, the discovery of biomarkers to distinguish between cold and hot tumors is particularly important for immunotherapy.

A recent study has reported that tumor cell metabolism is crucial for shaping the tumor microenvironment and its

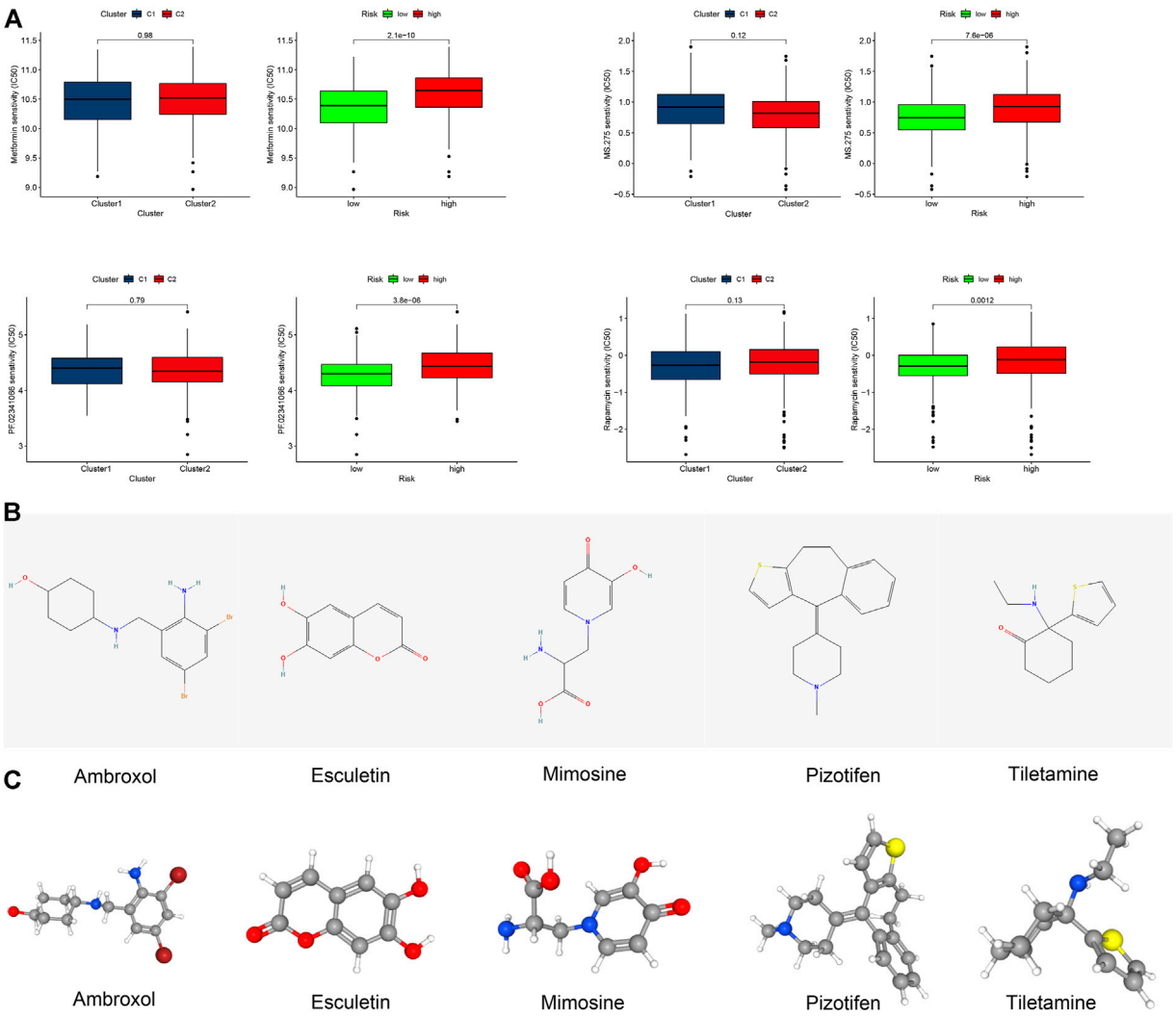


FIGURE 7
Clinical prediction of drug sensitivity in CM cases between two clusters or two risk groups. **(A)** Four potential drugs solely had significant IC_{50} differences between these two risk groups. **(B,C)** 2D structure illustrations and the 3D structure tomography of five candidate small-molecule drugs for CM. A *p* value of <0.05 was considered to indicate a statistically significant difference.

dysregulation is not only associated with the growth of tumors, but also with the therapeutic responses to immunotherapies (O'Sullivan et al., 2019). In addition, cancer cells can undergo metabolic reprogramming to support their survival when carcinogenic signals are blocked (DeBerardinis, 2020). FA, as a kind of important lipid molecule and energy source, is important for the growth of tumors and their therapeutic responses. The accumulation of FAs in the tumor microenvironment can affect the function and phenotype of immune cell infiltrates and FAO is crucial for CM metastasis and immune evasion (Li et al., 2018; Pascual et al., 2021). The accumulated FAs can limit anti-*CTLA-4* activity and inhibit tumor-specific and memory T cell infiltration into the tumors (Coutzac et al., 2020). In this study, we screened the FA metabolism-related DEGs in CM and found several DEGs had prognostic values in CM patients. Based on unique DEGs, we stratified CM patients into two clusters. In addition, we found that the frequency of CNV was critical for immune infiltrates in tumors and associated with the immunotherapeutic responses in CM patients, consistent with a previous report (Yang et al., 2021). Further analyses revealed that CM patients in Cluster 2 not only had more immune cell infiltrates (e.g., $CD8^+$ T and B), higher immune infiltration score, and more active immune function (e.g., Inflammatory promotion) but also displayed a higher immune checkpoint activity, such as *CTLA4*, *CD274* (*PD-*

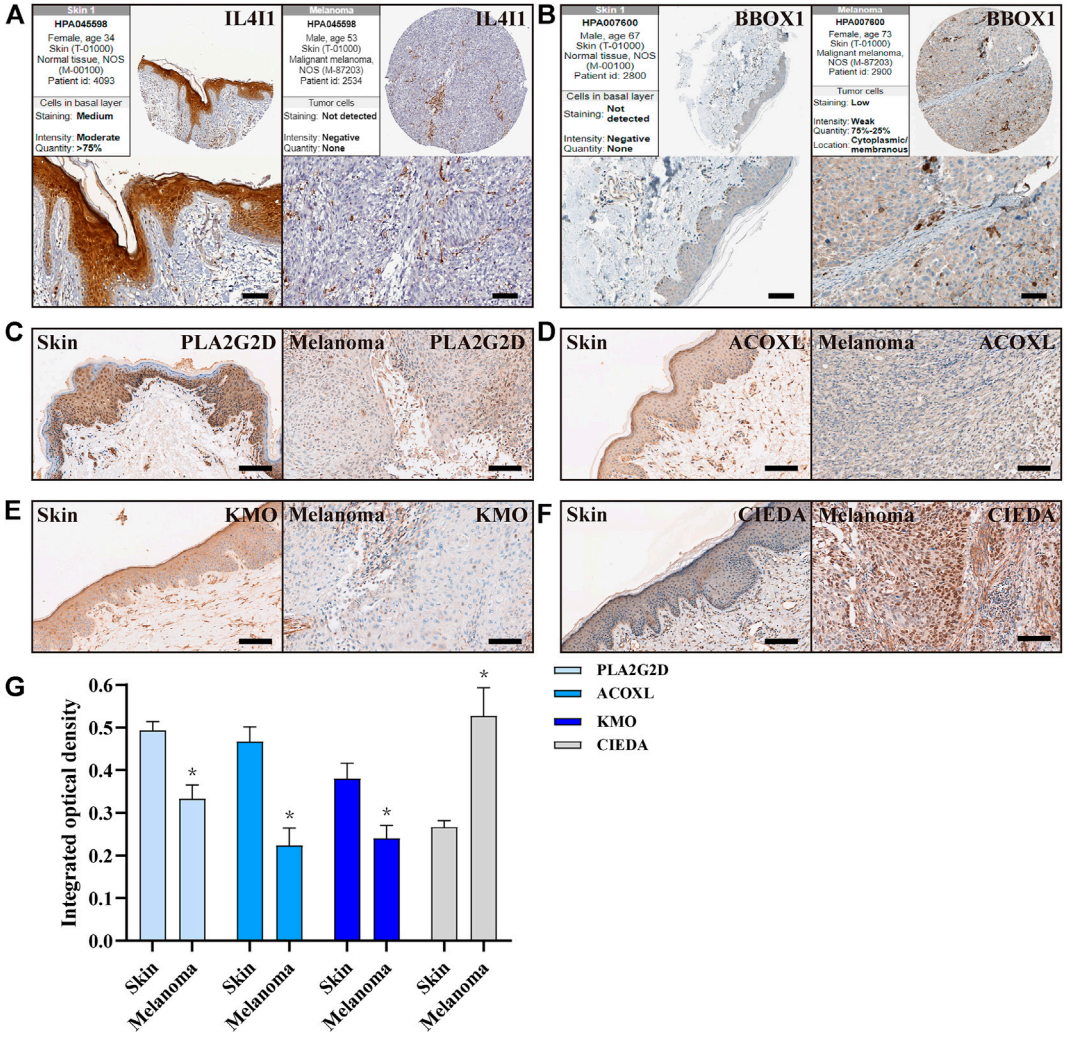


FIGURE 8
Expression of FA metabolism-related genes in normal skin and CM tissues. **(A,B)** IHC of the *IL4I1* and *BBOX1* in CM and normal skin tissues from HPA. **(C–F)** IHC of the *PLA2G2D*, *ACOXL*, *KMO,* and *CIDEA* in normal skin and CM tissues (n = 3), Scale bar = 100 μm. **(G)** Quantitative studies of *PLA2G2D*, *ACOXL*, *KMO,* and *CIDEA* were analyzed by counting the integrated optical density, respectively.

L1), *HAVCR2* (*TIME3*), and *LAG3*. Hence, the CM patients in cluster 2 had most of the characteristics of a hot tumor and might respond better to immunotherapy.

It is well known that a high TMB is associated with better immunotherapeutic responses in tumor patients. Tumor with a high TMB usually has a higher level of neoantigens, which can promote immune cell infiltration, enhancing the effect of immunotherapy (Maleki Vareki, 2018). We found that the TMB in the cluster 2 CM was significantly higher than that of the cluster 1 and was expected to have better survival. Similarly, CM in cluster 2 displayed higher levels of *PD-1* and *CTLA4* expression and a lower IC_{50} for many potential immunotherapeutic agents. Therefore, cluster analysis of the FA metabolism-related genes in CM effectively stratified patients for rational immunotherapy.

To predict patients' prognoses and explore the clinical application of FA metabolism-related genes in CM, we constructed a risk model using several FA metabolism-related DEGs, and validation revealed that this risk model had excellent sensitivity and specificity in separating CM patients for prognosis of OS in CM patients and their potential systemic therapy. Subsequently, we generated a nomogram using several independent risk factors, such as risk scores, age, tumor N stage, and T stage, and found that this nomogram had good validity and credibility for prognosis of CM patients.

Furthermore, five small-molecule drugs with potential therapeutic value were screened out through Cmap analysis, and they included ambroxol, tiletamine, mimosine, esculetin, and pizotifen. Previous studies have shown that tiletamine has potent cytotoxicity against melanoma cells by promoting ROS production and inducing cell cycle arrest, leading to melanoma cell apoptosis (Kyriakou et al., 2020). A combination of paclitaxel and ambroxol can synergistically kill lung cancer cells (He et al., 2020). Esculetin can inhibit the proliferation of pancreatic cancer cells to modulate their apoptosis by enhancing KEAP1 activity (Arora et al., 2016). Pizotifen has antitumor activity and is commonly used in gastrointestinal cancers. These small-molecule drugs may also have a potential therapeutic effect on CM (Jiang et al., 2020). Therefore, the FA metabolism can not only be used as a liquid biopsy method to quickly and effectively separate the cold from hot tumors to assist in rational immunotherapy but also predict patients' prognosis and potential treatment drugs.

The DEGs for separating clusters in our study are associated with development and progression of several types of malignancies. *ALDH3A1* over-expression can enhance the secretion of *PD-L1* in melanoma cells *in vitro*, and the levels of *ALDH3A1* expression are consistently correlated with those of *PD-L1* and *COX-2* in clinical melanoma and lung cancer samples (Terzuoli et al., 2019). *ALOX12B*, an immunosuppressive factor, can inhibit immunity and promote tumor progression (Uderhardt et al., 2012; Rooney et al., 2015). *BBOX1* inhibitors can restrain the progression of triple-negative breast cancer (Liao et al., 2020). *CYP4F3* has been identified as a cancer promoter of lung cancer (Yin et al., 2017). *PLA2G2D* has been found to have potential as a potential biomarker of adaptive resistance to immune checkpoint inhibitors (Cindy Yang et al., 2021). However, *PLA2G2D* expression is associated with delayed tumor growth by enhancing anti-tumor immunity in a mouse model of skin cancer (Miki et al., 2013). The *IL-1β* expression is significantly upregulated in glioblastoma and negatively correlated with the expression levels of *KCNJ10*, suggesting that *KCNJ10* may promote inflammation in the TME (Brandalise et al., 2020). Except for *ALDH3A1*, the precise roles of other factors we studied in CM remain to be determined. Our findings may provide new insights into the mechanisms underlying the immune regulation of CM.

We recognized that our studies had limitations. First, we only obtained data from TCGA and other public databases, but we did not explore real clinical samples to validate our findings. Second, although several independent data sets were used for validation of our findings, the retrospective study in nature might have potential bias. Therefore, the reliability and values of this model need to be further validated by well-designed prospective, multi-center, large-scale studies. However, our findings may provide a reliable reference for the specific interpretation of metabolic reprogramming in CM. It is worth mentioning that the risk model and nomogram have good clinical prognostic value, and the cluster analysis may help improve the current predicament in predicting immunotherapeutic responses of CM patients. Furthermore, our cluster analyses and model may be valuable for development of new therapeutic strategies for precision medicine and personalized immunotherapy and may contribute to improving the prognoses of CM patients.

5 Conclusion

Our data indicated that the FA metabolism-related DEGs were effective for identification of hot tumors and improving immunotherapeutic responses and prognosis of CM. Exploring the mechanism underlying disordered FA metabolism may help not only for precision medicine but also for developing immunotherapy strategies for CM.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee for Clinical Research and Animal Trials of the First Affiliated Hospital of Sun Yat-sen University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization: YD, ZZ, and BT; methodology, analysis, visualization, and writing (original draft preparation): YD, ZZ; writing (review and editing): YD, ZZ, MS, CC, DL, ZX, ZZ, and BT; project administration and funding acquisition: BT. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (NO. 81871565 and NO.

82072181) and the Sun Yat-sen University Clinical Research 5010 Program (NO. 2018003).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.860067/full#supplementary-material

SUPPLEMENTARY FIGURE S1
NMF of the external validation cohort and GSEA of clusters in the TCGA cohort. (A,B) All the heatmaps and NMF rank surveys of the external validation cohort. (C) GSEA of cluster 2 CM in the TCGA cohort. (D) GSEA of cluster 1 CM in the TCGA cohort.

SUPPLEMENTARY FIGURE S2
Prognosis value of the risk model.

SUPPLEMENTARY FIGURE S3
Clinical analyses of risk group CM patients. (A) t-SNE, PCA, and 3D PCA of risk groups. (B) GSEA of the low-risk group. (C) Immune heatmap of risk groups. (D) Correlation between risk scores and immune infiltrates. (E) Comparisons of immune-related scores in risk groups. *$p<0.05$, **$p<0.01$, and ***$p<0.001$. A p value of < 0.05 was considered to indicate a statistically significant difference.

References

Alicea, G. M., Rebecca, V. W., Goldman, A. R., Fane, M. E., Douglass, S. M., Behera, R., et al. (2020). Changes in aged fibroblast lipid metabolism induce age-dependent melanoma cell resistance to targeted therapy via the fatty acid transporter FATP2. *Cancer Discov.* 10, 1282–1295. doi:10.1158/2159-8290.CD-20-0329

Allais, B. S., Beatson, M., Wang, H., Shahbazi, S., Bijelic, L., Jang, S., et al. (2021). Five-year survival in patients with nodular and superficial spreading melanomas in the US population. *J. Am. Acad. Dermatol.* 84, 1015–1022. doi:10.1016/j.jaad.2020.11.047

Aloia, A., Mullhaupt, D., Chabbert, C. D., Eberhart, T., Fluckiger-Mangual, S., Vukolic, A., et al. (2019). A fatty acid oxidation-dependent metabolic shift regulates the adaptation of BRAF-mutated melanoma to MAPK inhibitors. *Clin. Cancer Res.* 25, 6852–6867. doi:10.1158/1078-0432.CCR-19-0253

Amaria, R. N., Menzies, A. M., Burton, E. M., Scolyer, R. A., Tetzlaff, M. T., Antdbacka, R., et al. (2019). Neoadjuvant systemic therapy in melanoma: Recommendations of the international neoadjuvant melanoma consortium. *Lancet. Oncol.* 20, e378–e389. doi:10.1016/S1470-2045(19)30332-8

Arora, R., Sawney, S., Saini, V., Steffi, C., Tiwari, M., and Saluja, D. (2016). Esculetin induces antiproliferative and apoptotic response in pancreatic cancer cells by directly binding to KEAP1. *Mol. Cancer* 15, 64. doi:10.1186/s12943-016-0550-2

Brandalise, F., Ratto, D., Leone, R., Olivero, F., Roda, E., Locatelli, C. A., et al. (2020). Deeper and deeper on the role of BK and Kir4.1 channels in glioblastoma invasiveness: A novel summative mechanism? *Front. Neurosci.* 14, 595664. doi:10.3389/fnins.2020.595664

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 68, 394–424. doi:10.3322/caac.21492

Cha, J. H., Yang, W. H., Xia, W., Wei, Y., Chan, L. C., Lim, S. O., et al. (2018). Metformin promotes antitumor immunity via endoplasmic-reticulum-associated degradation of PD-L1. *Mol. Cell* 71, 606–620. e607. doi:10.1016/j.molcel.2018.07.030

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* 30, 44–56. doi:10.1093/annonc/mdy495

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18, 248–262. doi:10.1016/j.celrep.2016.12.019

Chow, M. T., and Luster, A. D. (2014). Chemokines in cancer. *Cancer Immunol. Res.* 2, 1125–1131. doi:10.1158/2326-6066.CIR-14-0160

Cindy Yang, S. Y., Lien, S. C., Wang, B. X., Clouthier, D. L., Hanna, Y., Cirlan, I., et al. (2021). Pan-cancer analysis of longitudinal metastatic tumors reveals genomic alterations and immune landscape dynamics associated with pembrolizumab sensitivity. *Nat. Commun.* 12, 5137. doi:10.1038/s41467-021-25432-7

Courtney, A. H., Lo, W. L., and Weiss, A. (2018). TCR signaling: Mechanisms of initiation and propagation. *Trends biochem. Sci.* 43, 108–123. doi:10.1016/j.tibs.2017.11.008

Coutzac, C., Jouniaux, J. M., Paci, A., Schmidt, J., Mallardo, D., Seck, A., et al. (2020). Systemic short chain fatty acids limit antitumor effect of CTLA-4 blockade in hosts with cancer. *Nat. Commun.* 11, 2168. doi:10.1038/s41467-020-16079-x

DeBerardinis, R. J. (2020). Tumor microenvironment, metabolism, and immunotherapy. *N. Engl. J. Med.* 382, 869–871. doi:10.1056/NEJMcibr1914890

Deng, Z. M., Hu, W., Dai, F. F., Yuan, M. Q., Hu, M., and Cheng, Y. X. (2021). Immune-related genes to construct a novel prognostic model of breast cancer: A chemosensitivity-based study. *Front. Immunol.* 12, 734745. doi:10.3389/fimmu.2021.734745

Eggermont, A. M. M., Blank, C. U., Mandala, M., Long, G. V., Atkinson, V. G., Dalle, S., et al. (2021). Adjuvant pembrolizumab versus placebo in resected stage III melanoma (EORTC 1325-MG/KEYNOTE-054): Distant metastasis-free survival results from a double-blind, randomised, controlled, phase 3 trial. *Lancet. Oncol.* 22, 643–654. doi:10.1016/S1470-2045(21)00065-6

Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* 18, 197–218. doi:10.1038/s41573-018-0007-y

Gershenwald, J. E., Scolyer, R. A., Hess, K. R., Sondak, V. K., Long, G. V., Ross, M. I., et al. (2017). Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *Ca. Cancer J. Clin.* 67, 472–492. doi:10.3322/caac.21409

Gui, C. P., Wei, J. H., Chen, Y. H., Fu, L. M., Tang, Y. M., Cao, J. Z., et al. (2021). A new thinking: Extended application of genomic selection to screen multiomics data for development of novel hypoxia-immune biomarkers and target therapy of clear cell renal cell carcinoma. *Brief. Bioinform.* 22, bbab173. doi:10.1093/bib/bbab173

Haikala, H. M., Anttila, J. M., Marques, E., Raatikainen, T., Ilander, M., Hakanen, H., et al. (2019). Pharmacological reactivation of MYC-dependent apoptosis induces susceptibility to anti-PD-1 immunotherapy. *Nat. Commun.* 10, 620. doi:10.1038/s41467-019-08541-2

Hamid, O., Robert, C., Daud, A., Hodi, F. S., Hwu, W. J., Kefford, R., et al. (2019). Five-year survival outcomes for patients with advanced melanoma treated with pembrolizumab in KEYNOTE-001. *Ann. Oncol.* 30, 582–588. doi:10.1093/annonc/mdz011

He, W., Xiao, W., Zhang, X., Sun, Y., Chen, Y., Chen, Q., et al. (2020). Pulmonary-affinity paclitaxel polymer micelles in response to biological functions of ambroxol enhance therapeutic effect on lung cancer. *Int. J. Nanomedicine* 15, 779–793. doi:10.2147/IJN.S229576

Herrera, F. G., Ronet, C., Ochoa de Olza, M., Barras, D., Crespo, I., Andreatta, M., et al. (2021). Low dose radiotherapy reverses tumor immune desertification and resistance to immunotherapy. *Cancer Discov.* 12, 108–133. doi:10.1158/2159-8290.CD-21-0003

Ji, S., Li, J., Duan, X., Zhang, J., Zhang, Y., Song, M., et al. (2021). Targeted enrichment of enzyme-instructed assemblies in cancer cell lysosomes turns immunologically cold tumors hot. *Angew. Chem. Int. Ed. Engl.* 60, 26994–27004. doi:10.1002/anie.202110512

Jiang, Y., Wang, W., Wu, X., and Shi, J. (2020). Pizotifen inhibits the proliferation and invasion of gastric cancer cells. *Exp. Ther. Med.* 19, 817–824. doi:10.3892/etm.2019.8308

Kang, K., Xie, F., Mao, J., Bai, Y., and Wang, X. (2020). Significance of tumor mutation burden in immune infiltration and prognosis in cutaneous melanoma. *Front. Oncol.* 10, 573141. doi:10.3389/fonc.2020.573141

Kleemann, J., Hrgovic, I., Ter-Nedden, J., Kleimann, P., Steinhorst, K., Harle, K., et al. (2018). Fatty acid receptor GPR120: A novel marker for human melanoma. *Melanoma Res.* 28, 271–276. doi:10.1097/CMR.0000000000000445

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi:10.1101/gr.129684.111

Kyriakou, S., Mitsiogianni, M., Mantso, T., Cheung, W., Todryk, S., Veuger, S., et al. (2020). Anticancer activity of a novel methylated analogue of L-mimosine against an *in vitro* model of human malignant melanoma. *Invest. New Drugs* 38, 621–633. doi:10.1007/s10637-019-00809-0

Li, J., Fang, Y., Zhang, Y., Wang, H., Yang, Z., and Ding, D. (2021). Supramolecular self-assembly-facilitated aggregation of tumor-specific transmembrane receptors for signaling activation and converting immunologically cold to hot tumors. *Adv. Mat.* 33, e2008518. doi:10.1002/adma.202008518

Li, X. X., Wang, Z. J., Zheng, Y., Guan, Y. F., Yang, P. B., Chen, X., et al. (2018). Nuclear receptor Nur77 facilitates melanoma cell survival under metabolic stress by protecting fatty acid oxidation. *Mol. Cell* 69, 480–492. doi:10.1016/j.molcel.2018.01.001

Liao, C., Zhang, Y., Fan, C., Herring, L. E., Liu, J., Locasale, J. W., et al. (2020). Identification of BBOX1 as a therapeutic target in triple-negative breast cancer. *Cancer Discov.* 10, 1706–1721. doi:10.1158/2159-8290.CD-20-0288

Long, J., Wang, D., Wang, A., Chen, P., Lin, Y., Bian, J., et al. (2022). A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. *Genome Med.* 14, 20. doi:10.1186/s13073-022-01024-y

Lopez de Rodas, M., and Schalper, K. A. (2021). Tumour antigen-induced T cell exhaustion - the archenemy of immune-hot malignancies. *Nat. Rev. Clin. Oncol.* 18, 749–750. doi:10.1038/s41571-021-00562-5

Mahoney, K. M., Rennert, P. D., and Freeman, G. J. (2015). Combination cancer immunotherapy and new immunomodulatory targets. *Nat. Rev. Drug Discov.* 14, 561–584. doi:10.1038/nrd4591

Maleki Vareki, S. (2018). High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J. Immunother. Cancer* 6, 157. doi:10.1186/s40425-018-0479-7

Miki, Y., Yamamoto, K., Taketomi, Y., Sato, H., Shimo, K., Kobayashi, T., et al. (2013). Lymphoid tissue phospholipase A2 group IID resolves contact hypersensitivity by driving antiinflammatory lipid mediators. *J. Exp. Med.* 210, 1217–1234. doi:10.1084/jem.20121887

O'Sullivan, D., Sanin, D. E., Pearce, E. J., and Pearce, E. L. (2019). Metabolic interventions in the immune response to cancer. *Nat. Rev. Immunol.* 19, 324–335. doi:10.1038/s41577-019-0140-9

Ochoa de Olza, M., Navarro Rodrigo, B., Zimmermann, S., and Coukos, G. (2020). Turning up the heat on non-immunoreactive tumours: Opportunities for clinical development. *Lancet. Oncol.* 21, e419–e430. doi:10.1016/S1470-2045(20)30234-5

Oh, D. Y., Talukdar, S., Bae, E. J., Imamura, T., Morinaga, H., Fan, W., et al. (2010). GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* 142, 687–698. doi:10.1016/j.cell.2010.07.041

Pascual, G., Dominguez, D., Elosua-Bayes, M., Beckedorff, F., Laudanna, C., Bigas, C., et al. (2021). Dietary palmitic acid promotes a prometastatic memory via Schwann cells. *Nature* 599, 485–490. doi:10.1038/s41586-021-04075-0

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61. doi:10.1016/j.cell.2014.12.033

Salmon, H., Remark, R., Gnjatic, S., and Merad, M. (2019). Host tissue determinants of tumour immunity. *Nat. Rev. Cancer* 19, 215–227. doi:10.1038/s41568-019-0125-9

Schadendorf, D., van Akkooi, A. C. J., Berking, C., Griewank, K. G., Gutzmer, R., Hauschild, A., et al. (2018). *Melanoma. Lancet* 392, 971–984. doi:10.1016/S0140-6736(18)31559-9

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199. doi:10.1056/NEJMoa1406498

Song, Y., Yin, W., Dan, Y., Sheng, J., Zeng, Y., and He, R. (2019). Chemerin partly mediates tumor-inhibitory effect of all-trans retinoic acid via CMKLR1-dependent natural killer cell recruitment. *Immunology* 157, 248–256. doi:10.1111/imm.13065

Stupnikov, A., McInerney, C. E., Savage, K. I., McIntosh, S. A., Emmert-Streib, F., Kennedy, R., et al. (2021). Robustness of differential gene expression analysis of RNA-seq. *Comput. Struct. Biotechnol. J.* 19, 3470–3481. doi:10.1016/j.csbj.2021.05.040

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, necroptosis, and pyroptosis in anticancer immunity. *J. Hematol. Oncol.* 13, 110. doi:10.1186/s13045-020-00946-7

Taniguchi, K., and Karin, M. (2018). NF-κB, inflammation, immunity and cancer: Coming of age. *Nat. Rev. Immunol.* 18, 309–324. doi:10.1038/nri.2017.142

Terzuoli, E., Bellan, C., Aversa, S., Ciccone, V., Morbidelli, L., Giachetti, A., et al. (2019). ALDH3A1 overexpression in melanoma and lung tumors drives cancer stem cell expansion, impairing immune surveillance through enhanced PD-L1 output. *Cancers (Basel)* 11, E1963. doi:10.3390/cancers11121963

Thul, P. J., and Lindskog, C. (2018). The human protein atlas: A spatial map of the human proteome. *Protein Sci.* 27, 233–244. doi:10.1002/pro.3307

Too, N. S. H., Ho, N. C. W., Adine, C., Iyer, N. G., and Fong, E. L. S. (2021). Hot or cold: Bioengineering immune contextures into *in vitro* patient-derived tumor models. *Adv. Drug Deliv. Rev.* 175, 113791. doi:10.1016/j.addr.2021.05.001

Uderhardt, S., Herrmann, M., Oskolkova, O. V., Aschermann, S., Bicker, W., Ipseiz, N., et al. (2012). 12/15-lipoxygenase orchestrates the clearance of apoptotic cells and maintains immunologic tolerance. *Immunity* 36, 834–846. doi:10.1016/j.immuni.2012.03.010

Yang, X., Hu, Y., Yang, K., Wang, D., Lin, J., Long, J., et al. (2021). Cell-free DNA copy number variations predict efficacy of immune checkpoint inhibitor-based therapy in hepatobiliary cancers. *J. Immunother. Cancer* 9, e001942. doi:10.1136/jitc-2020-001942

Yin, J., Liu, H., Liu, Z., Owzar, K., Han, Y., Su, L., et al. (2017). Pathway-analysis of published genome-wide association studies of lung cancer: A potential role for the CYP4F3 locus. *Mol. Carcinog.* 56, 1663–1672. doi:10.1002/mc.22622

Zhao, F., Xiao, C., Evans, K. S., Theivanthiran, T., DeVito, N., Holtzhausen, A., et al. (2018). Paracrine wnt5a-beta-catenin signaling triggers a metabolic Program that drives dendritic cell tolerization. *Immunity* 48, 147–160. e147. doi:10.1016/j.immuni.2017.12.004

Zheng, Y., Tian, H., Zhou, Z., Xiao, C., Liu, H., Liu, Y., et al. (2021). A novel immune-related prognostic model for response to immunotherapy and survival in patients with lung adenocarcinoma. *Front. Cell Dev. Biol.* 9, 651406. doi:10.3389/fcell.2021.651406

Zhuo, S., Chen, Z., Yang, Y., Zhang, J., Tang, J., and Yang, K. (2020). Clinical and biological significances of a ferroptosis-related gene signature in glioma. *Front. Oncol.* 10, 590861. doi:10.3389/fonc.2020.590861

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fgene.2022.834471

Check for updates

OPEN ACCESS

EDITED BY
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

REVIEWED BY
Taobo Hu,
Peking University People's Hospital, China
Yan Wei Cai,
China Academy of Chinese Medical Sciences, China
Jimei Wang,
Fudan University, China

*CORRESPONDENCE
Wang Zhen-Yu,
wzyxshp@163.com
Zhang Wei-Jun,
zwjxshp@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 13 December 2021
ACCEPTED 03 August 2022
PUBLISHED 28 September 2022

CITATION
Yu-Qing H, Peng-Ping L, Ke S, Ke-Xing Y, Wei-Jun Z and Zhen-Yu W (2022), Comprehensive analysis of liquid-liquid phase separation-related genes in prediction of breast cancer prognosis.
Front. Genet. 13:834471.
doi: 10.3389/fgene.2022.834471

COPYRIGHT
© 2022 Yu-Qing, Peng-Ping, Ke, Ke-Xing, Wei-Jun and Zhen-Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive analysis of liquid-liquid phase separation-related genes in prediction of breast cancer prognosis

Huang Yu-Qing[1,2†], Li Peng-Ping[1,2†], Sun Ke[1,2], Yin Ke-Xing[1,2], Zhang Wei-Jun[1,2*] and Wang Zhen-Yu[1,2*]

[1]The First People's Hospital of Xiaoshan, Hangzhou, China, [2]Xaioshan Hospital Affiliated to Wenzhou Medical University, Wenzhou, China

Objective: Liquid-liquid phase separation (LLPS) is a functional unit formed by specific molecules. It lacks a membrane and has been reported to play a crucial role in tumor drug resistance and growth by regulating gene expression and drug distribution. However, whether LLPS could be used to predict cancer prognosis was not clear. This study aimed to construct a prognostic model for breast cancer based on LLPS-correlated genes (LCGs).

Methods: LCGs were identified using the PhaSepDB, gene expression profile and clinical characteristics of breast cancer were obtained from TCGA and cBioportal. The PanCancer Atlas (TCGA) cohort was used as the training cohort to construct the prognostic model, while the Nature 2012 and Nat Commun 2016 (TCGA) cohort and GEO data were used as test cohort to perform external verification. Data analysis was performed with R4.2.0 and SPSS20.0.

Results: We identified 140 prognosis-related LCGs (pLCGs) ($p< 0.01$) in all cohorts, 240 pLCGs ($p< 0.01$) in the luminal cohort, and 28 pLCGs ($p< 0.05$) in the triple-negative breast cancer (TNBC) cohort. Twelve genes in all cohorts (training cohort: 5/10-year ROC values were 0.76 and 0.77; verification cohort: 5/10-year ROC values were 0.61 and 0.58), eight genes in the luminal cohort (training cohort: 5/10-year ROC values were 0.79 and 0.75; verification cohort: 5/10-year ROC values were 0.62 and 0.62), and four genes in the TNBC cohort (training cohort: 5/10-year ROC values were 0.73 and 0.79; verification cohort: 5/10-year ROC values were 0.55 and 0.54) were screened out to construct the prognostic prediction model. The 17-gene risk-score was constructed in all cohorts (1/3/5-year ROC values were 0.88, 0.83, and 0.81), and the 11-gene risk-score was constructed in the luminal cohort (1/3/5-year ROC values were 0.67, 0.85 and 0.84), and the six-gene risk-score was constructed in the TNBC cohort (1/3/5-year ROC value were 0.87, 0.88 and 0.81). Finally, the risk-score

Abbreviations: K-M curve, Kaplan Meier curve; FGA, fraction of genomic alteration; LCGs, LLPS-correlated genes; LLPS, liquid-liquid phase separation; pLCGs, prognosis-related LCGs; TCGA, The Cancer Genome Atlas; TNBC, triple negative breast cancer; TMB, tumor mutation burden.

and clinical factors were applied to construct nomograms in all cohorts (1/3/5-year ROC values were 0.89, 0.79 and 0.75, C-index = 0.784), in the luminal cohort (1/3/5-year ROC values were 0.84, 0.83 and 0.85, C-index = 0.803), and in the TNBC cohort (1/3/5-year ROC values were 0.95, 0.84 and 0.77, C-index = 0.847).

Discussion: This study explored the roles of LCGs in the prediction of breast cancer prognosis.

KEYWORDS

liquid-liquid phase separation, multi-gene risk-score, prediction model, breast cancer, TCGA

Introduction

During the past decades, medical science has made obvious progress in the treatment of breast cancer, especially for *HER2*-positive breast cancer based on the development of *HER2*-targeted drugs, such as trastuzumab. However, because of the lack of useful gene targets, paclitaxel-centered combination chemotherapy was still the first-line treatment strategy for patients with *HER2*-negative breast carcinoma, including luminal and triple-negative breast cancer (TNBC), the resistance of which made for a worse prognosis (Foulkes et al., 2010; Pandya-Jones et al., 2020).

Recently, it has been reported that RNA and proteins can interact with each other to form a droplet-like unit by multivalent weak interactions based on intrinsically disordered regions (IDR), folded proteins, DNA/RNA molecular scaffolds, and other structures, which was called a liquid-liquid phase separation (LLPS) (Alberti et al., 2019; Li et al., 2021a). Many biological processes, including transcription, chromatin organization, X chromosome inactivation (XCI), DNA damage response (DDR), autophagy, and even tumor growth and metastasis, have been proven to involve LLPS to achieve their specific functions (Du and Chen, 2018; Hahn, 2018; Ries et al., 2019). For example, the YAP protein formed a liquid aggregate in the nucleus, which induced the transcription of its target genes and promoted the growth of MBA-MD-231 breast cancer cells *in vivo* and in vitro (Li et al., 2021a).

LLPSs appear at different phases and spaces in cells to perform specific functions. The components involved in the formation of LLPS can be quite different. Recent studies have shown that long noncoding RNAs (lncRNAs) can regulate cellular functions by interacting with target proteins to form dynamic LLPS (Pandya-Jones et al., 2020). For example, the lncRNA Xist formed condensates in the inactive X(Xi) group by binding to multiple proteins, such as PTBP1, MATR3, TDP-43, and CELF1, with self-aggregation and heterotypic protein-protein interactions, which provided a new way for gene silencing (Pandya-Jones et al., 2020). LncRNA NEAT1 exhibited phase-separated condensate properties, and was able to bind to NONO/SFPQ with the formation of LLPS *in vitro*. In addition, lncRNAs interacted with oncogenes to form LLPS, which were involved in regulating tumor development (Yamazaki et al., 2018). For example, the lncRNA SNHG9 promoted LATS1 to experience LLPS, which further promoted the YAP signaling pathway-induced growth of breast cancer cells (Li et al., 2021a).

In previous research, immunological genes, autophagy-related genes, and some other genes were reported to be useful in tumor prognosis prediction (Shen et al., 2020; Li et al., 2021b; Jiang et al., 2021; Jiang et al., 2022), but few studies focused on the roles of LLPS-related genes (LCGs) in tumor prognosis prediction. For example, prognosis prediction models were based on previously constructed LCGs for ovarian cancer, lung squamous cell carcinoma, and glioma (Qiu et al., 2021; Zheng et al., 2022); the risk model based on LCGs identified a good/bad prognosis cluster. However, an LCG-based risk model has not been reported for breast cancer and its subtypes; so, in this study we constructed a nomogram based on LCGs.

Methods and materials

Data collection and collation

Gene expression profiles of the training cohort were collected from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov), and the data of the verification cohort was collected from cBioportal (METRABRIC, Nature 2012 & Nat Commun 2016; http://www.cbioportal.org) and GEO58812. The clinical characteristics of TCGA were obtained from cBioportal (http://www.cbioportal.org). LLPS-related genes were selected from PhaSepDB, an online database that records all LLPS-related genes (http://db.phasep.pro). A total of 1077 records (673 in the luminal cohort, 171 in the TNBC cohort, 0 excluded) were selected from the training cohort, and 1904 records were selected (1140 in the luminal cohort, 199 in the TNBC cohort, and 604 were excluded) from the verification cohort (Figure 1). Clinical factors included "Age (<45, 45 ~ 64, >64)", "clinical stage (I-II, III-IV)", "T stage

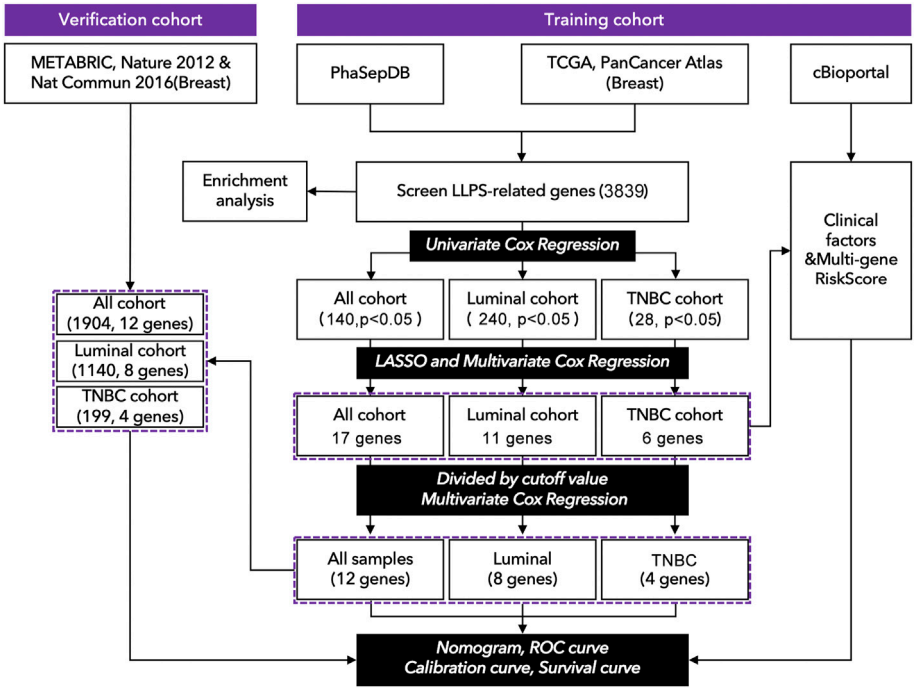


FIGURE 1
Technology roadmap of study. (1) Gene expression profiles were collected from TCGA: https://portal.gdc.cancer.gov. (2) Clinical characteristics were collected from cBioportal http://www.cbioportal.org. (3) The genomic list of LLPS-related genes was collected from PhaSepDB: http://db.phasep.pro. (4) Data analysis was performed in R4.0.1 and SPSS 20.0.

(T1-2, T3-4)", "N stage (N0, N1-3), M stage, recurrence status, and radiation therapy", genomic factors included "Tumor mutation burden (TMB)" and "risk-score (multi-gene risk-score)".

Identification of prognostic signature LCGs and construction of an LCG-based risk-score

The gene expression profile was collected from TCGA, and the LCGs were exported from PhaSepDB. Next, 3839 genes were identified by taking the intersection between the data from TCGA and PhaSepDB. Univariate Cox regression (*limma* package in R, $p< 0.01$ in all breast and luminal cohorts, $p< 0.05$ in the TNBC cohort, fold changes >1.5) was performed to identify prognosis-related LCGs. Least absolute shrinkage and selection operator (LASSO) Cox regression (*glmnet* package in R) was performed to narrow the array of candidate genes. Multivariate Cox regression was performed in R to select genes for constructing nomograms or multi-gene risk-scores, in which an LLPS-related risk-score was constructed according to the formula:

$$Riskscore = Exp\,(gene1) \times Coef\,(gene1) + Exp\,(gene2) \times Coef\,(gene2) + \ldots$$

The "*Exp*" mean expression value of a gene in multivariate Cox regression. "*Coef*" mean coefficient of the corresponding gene. Cohorts were divided into "High-risk group" and "Low-risk group" by risk-score.

Enrichment analysis

A total of 340 differentially-expressed LCGs were identified by KEGG and GO ("*clusterProfiler, org.Hs.eg.db, and enrichplot*" packages in R) analysis.

Identification of clinical factors and construction of prognosis prediction model

Clinical factors (clinical stage, N stage, T stage, age) and genomic factors (risk-score, TMB) were subjected to univariate Cox regression and multivariate Cox regression analysis in SPSS 20.0. Nomograms, calibration analysis, and Kaplan Meier (K-M) curves were constructed in R ("*survival, rms, regplot, and*

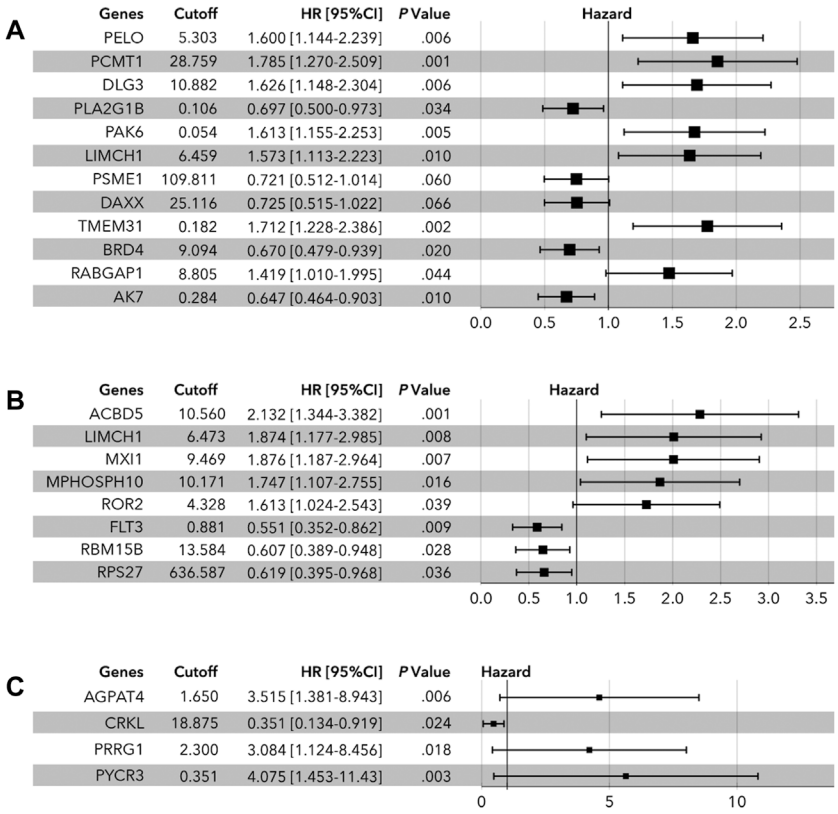


FIGURE 2
Multivariate Cox regression analysis of LLPS-related genes. **(A)** In all cohort, PELO, PCMT1, DLG3, PLA2G1B, PAK6, LIMCH1, PSME1, DAXX, TMEM31, BRD4, RABGAP1, and AK7 were identified to determine the risk-score. **(B)** In the luminal cohort, ACBD5, LIMCH1, MXI1, MPHOSPH10, ROR2, FLT3, RBM15B, and ROS27 were identified to determine the risk-score. **(C)** In the TNBC cohort, AGPAT4, CRKL, PRRG1, and PYCR3 were identified to calculate the risk-score.

survminer" packages), and the C-index was calculated in R ("*survcomp*").

Statistics analysis

All data analyses were performed using R4.0.1 and SPSS 20.0 and some data analyses were performed with online tools (http://www.sangerbox.com/tool). All core R codes related to this study were uploaded on ZENODO (https://zenodo.org) [DIO: 10.5281/zenodo.6497469; link: https://doi.org/10.5281/zenodo.6497469].

Results

Identification of prognosis-related LCGs in subtypes of breast cancer

As shown in Figure 1, 3839 LCGs were identified by searching the intersection between the data from TCGA and PhaSepDB, and those genes were subjected to univariate Cox regression analysis. The samples of breast cancer from TCGA were divided into all cohorts, the luminal cohort, and the TNBC cohort, in which 140 prognosis-related LCGs (pLCGs) were identified in all cohort ($p < 0.05$, Figure 1), 240 pLCGs were identified in the luminal cohort ($p < 0.05$, Figure 1), and 28 pLCGs were identified in the TNBC cohort ($p < 0.05$, Figure 1).

Construction of LGCs-based prognosis prediction model

The above candidate genes were subjected to LASSO analysis, and 17 genes in all cohorts were selected for multivariate Cox regression analysis-11 genes from the luminal cohort and 6 genes from the TNBC cohort (Figure 1 and Supplementary Figure S1). The samples were divided into high-expression and low-expression groups by median expression of the selected genes. Then, by multivariate

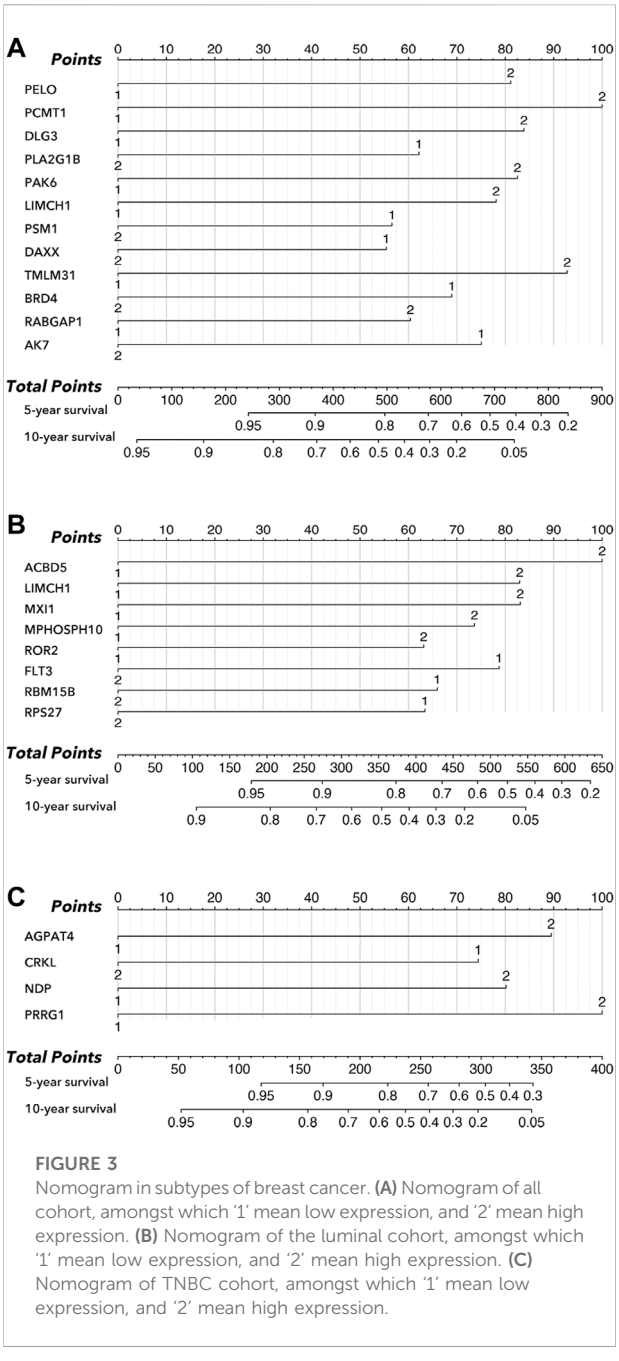


FIGURE 3
Nomogram in subtypes of breast cancer. **(A)** Nomogram of all cohort, amongst which '1' mean low expression, and '2' mean high expression. **(B)** Nomogram of the luminal cohort, amongst which '1' mean low expression, and '2' mean high expression. **(C)** Nomogram of TNBC cohort, amongst which '1' mean low expression, and '2' mean high expression.

Cox regression analysis, we identified **PELO** (1.600 [1.44–2.239], p = 0.006); **PCMT1** (1.785[1.270–2.509], p = 0.001); **DLG3** (1.626[1.148–2.304], p = 0.006); **PLA2G1B** (0.697[0.500–0.973], p = 0.034); **PAK6** (1.613[1.155–2.253], p = 0.005); **LIMCH1** (1.573[1.113–2.223], p = 0.010); **PSME1** (0.721[0.512–1.014], p = 0.060); **DAXX** (0.725[0.515–1.022], p = 0.066); **TMEM31** (1.712[1.228–2.386], p = 0.002); **BRD4** (0.670[0.479–0.939], p = 0.020); **RABGAP1** (1.419 [1.010–1.995], p = 0.044); and **AK7** (0.647[0.464–0.903], p = 0.010) as factors for constructing prognosis prediction model in all cohorts (Figure 2A). We also identified 11 genes: **ACBD5** (2.132[1.344–3.382], p = 0.001); **LIMCH1** (1.874 [1.177–2.985], p = 0.008); **MXI1** (1.876[1.187–2.964], p = 0.007); **MPHOSPH10** (1.747[1.107–2.755], p = 0.016); **ROR2** (1.613[1.024–2.543], p = 0.039); **FLT3** (0.551 [0.352–0.862], p = 0.009); **RBM15B** (0.607[0.389–0.948], p = 0.028); and **RPS27** (0.619[0.395–0.968], p = 0.036) as factors for constructing a prognosis prediction model in the luminal cohort (Figure 2B). We also identified the six genes: **AGPAT4** (3.515[1.381–8.943], p = 0.006); **CRKL** (0.351 [0.134–0.919], p = 0.024); **PRRG1** (3.084[1.124–8.456], p = 0.018); and **PYCR3** (4.075[1.453–11.430], p = 0.003) as factors for constructing a prognosis prediction model in the TNBC cohort (Figure 2C).

As shown in Figure 3, an LCG-based nomogram was constructed for the training cohort (all cohort, luminal cohort, and TNBC cohort), in which '1' mean low expression and '2' mean high expression. Samples were divided into low-risk and high-risk groups by predicted risk-score calculated by multivariate Cox regression analysis. As shown in Figure 4, the low-risk group had a better prognosis than the high-risk group in all cohorts (Figure 4A, $p < 0.0001$), the luminal cohort (Figure 4E, $p < 0.0001$), and the TNBC cohort (Figure 4I, $p = 0.002$). The ROC value of the nomograms in all cohorts was 0.76 (5-year survival) and 0.77 (10-year survival) (Figure 4B); in the luminal cohort it was 0.79 (5-year survival) and 0.75 (10-year survival) (Figure 4F); and in all cohorts it was 0.73 (5-year survival) and 0.79 (10-year survival) (Figure 4J). In addition, calibration analysis was performed to assess the predictive ability of the nomogram (Figures 4C,D,G,H,K,L).

To further verify the prediction ability of the above nomogram (prognosis prediction model), data from METABRIC (Nature 2012 & Nat Commun 2016) was used to construct the verification cohort (Figure 1). We used the above-identified genes to construct the prediction model in all cohorts, the luminal cohort, and the TNBC cohort. As shown in Figure 5, the high-risk group had a worse prognosis than the low-risk group in all cohorts (Figure 5A, $p< 0.0001$) and the luminal cohort (Figure 5E, $p< 0.0001$), while there was no difference in the TNBC cohort (Figure 5I, $p = 0.41$). In addition, the ROC values of the nomogram in the verification cohort were not good: 0.61 (5 years) and 0.58 (10 years) in all cohorts (Figure 5B), 0.62 (5 years, 10 years) in the luminal cohort (Figure 5F), and 0.56 (5 years) and 0.54 (10 years) in the TNBC cohort (Figure 5J). Calibration analysis was also performed (Figures 5C,D,G,H,K,L).

Enrichment analysis

A total of 340 differentially-expressed LCGs were subjected to KEGG and GO analysis to identify molecular signaling

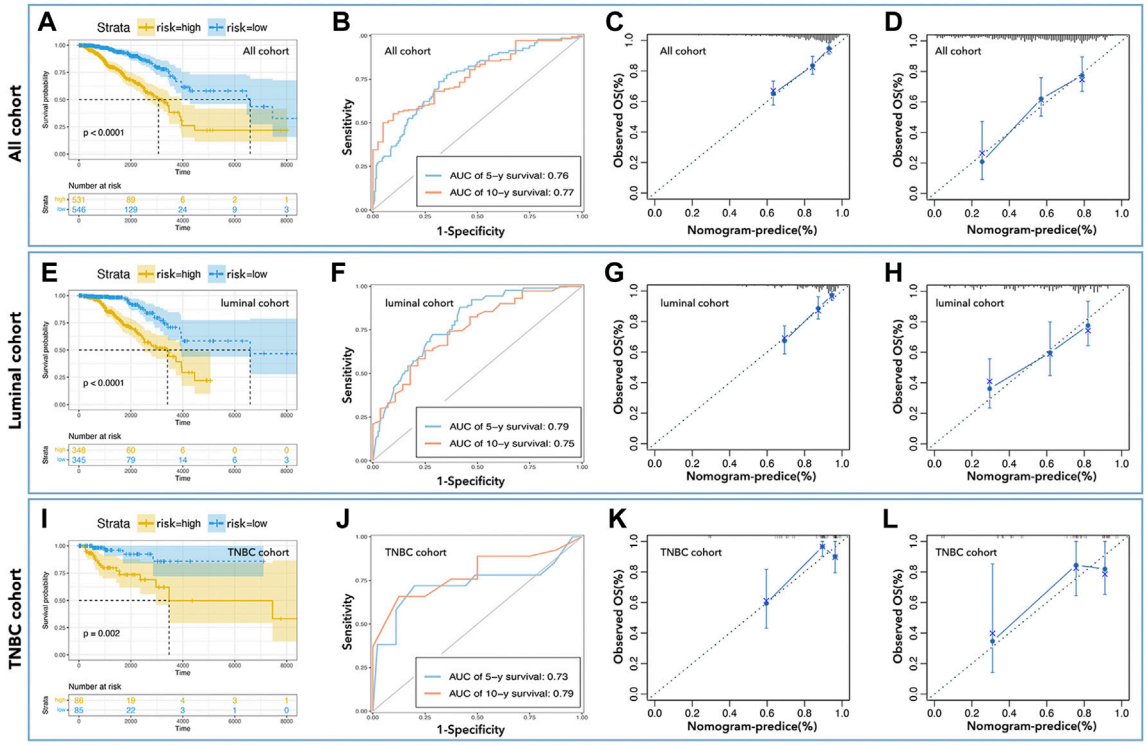


FIGURE 4
Inner verification of monogram. K-M curve showed low-risk with better prognosis in all cohorts **(A)**, the luminal cohort **(E)**, and the TNBC cohort **(I)**. The AUC values of nomograms in all cohorts **(B)** were 0.76 (5-year survival) and 0.77 (10-year survival), in the luminal cohort **(F)** were 0.79 (5-year survival) and 0.75 (10-year survival), and in all cohort **(J)** were 0.73 (5-year survival) and 0.79 (10-year survival). Calibration analysis was performed to assess the prediction accuracy of nomograms in all cohorta **(C,D)**, luminal cohort **(G,H)**, and TNBC cohort **(K,L)**.

pathways. As the GO analysis showed, LCGs were involved in RNA binding, protein targeting to ER, and translational initiation, etc. (Figure 6A). KEGG analysis showed that LCGs were involved in the NOD-like receptor signaling pathway, focal adhesion, tight junctions, and spliceosomes, etc. (Figure 6B).

Prognosis prediction model based on LCGs risk-score

To further explore the effects of LCGs on the prediction of prognosis, multi-gene risk-scores were calculated. As shown in Figure 7, 17 pLCGs were used to obtain multi-gene risk-scores in all cohorts, all of which had different expressions between tumor and non-tumor tissues (Figure 7A); 11 pLCGs were used to determine risk-scores in the luminal cohort, all of which had different expressions between tumor and non-tumor tissues (Figure 7D); 6 pLCGs were used to calculate risk-scores in the TNBC cohort, and *AGPAT4*, *CRKL*, *NDP*, *PRRG1*, and *PYCR3* all had different expression between tumor and non-tumor tissues (Figure 7G). The K-M curve showed that the group with the lowest risk-score had the best prognosis (**Fig. B, E, and H**). We determined the prognostic prediction ability of risk-scores in breast cancer and its subtypes, and found that the ROC value of the risk-score based on 17 LCGs was 0.88 (1 year), 0.83 (3 years), and 0.81 (5 years) in all cohorts (Figure 7C); the ROC value of the risk-score based on 11 LCGs was 0.67 (1 year), 0.85 (3 years), and 0.84 (5 years) in the luminal cohort (Figure 7F); the ROC value of the risk-score based on 6 LCGs was 0.87 (1 year), 0.88 (3years), and 0.81 (5 years) in the TNBC cohort (Figure 7I).

LCG-based risk-score in the regulation of genomic instability and tumor immunity

To explore the roles of LCGs in the regulation of genomic instability and tumor immunity, we calculated the differences in fraction of genomic alteration (FGA), microsatellite instability (MSI), gene mutation (mutation) and tumor mutation burden (TMB). As shown in Figure 8, the group with higher LCG-based risk-scores also had a higher level of FGA and MSI, while it was accompanied by

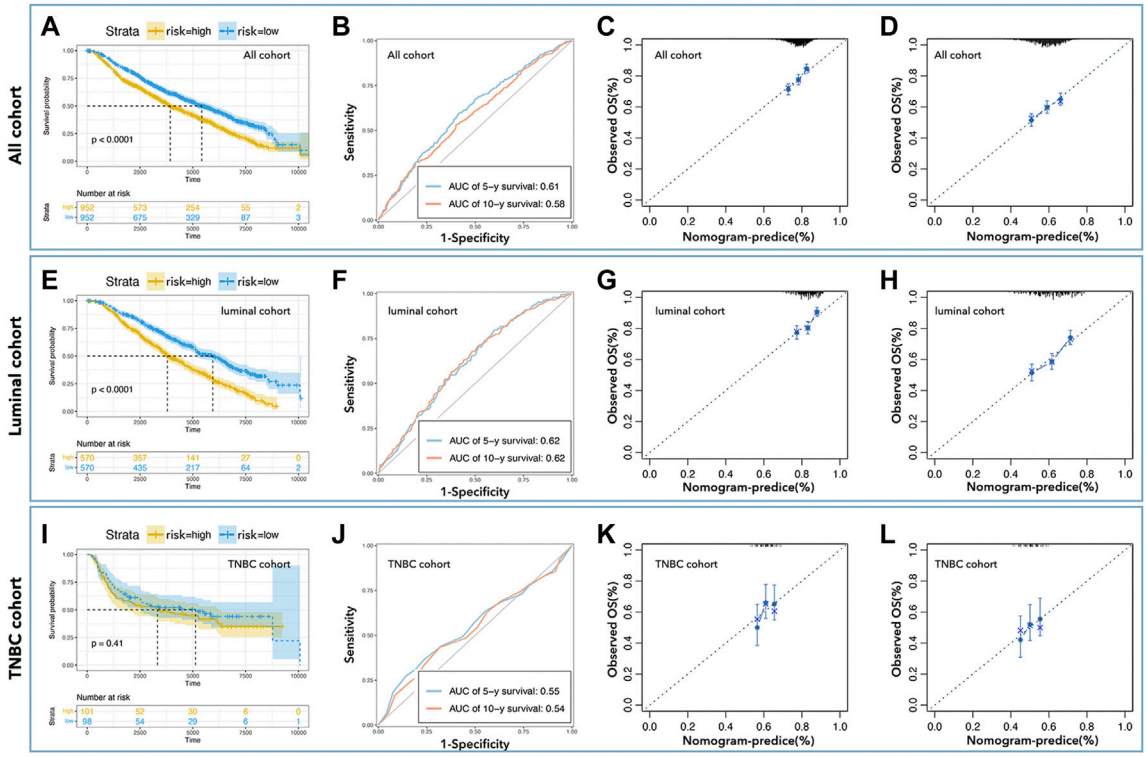


FIGURE 5
External verification of predicted models. K-M curve showed low-risk with better prognosis in all cohorts **(A)** and the luminal cohort **(E)**, but not in the TNBC cohort **(I)**. The AUC values of nomogram in all cohorts **(B)** were 0.61 (5-year survival) and 0.58 (10-year survival), in the luminal cohort **(F)** were 0.62 (5-year survival, 10-year survival), and in the TNBC cohort **(J)** were 0.55 (5-year survival) and 0.54 (10-year survival). Calibration analysis was performed to assess the prediction accuracy of nomograms in all cohorts **(C,D)**, the luminal cohort **(G,H)**, and the TNBC cohort **(K,L)**.

a lower level of mutation and TMB. In immunity analysis, we found that the group with the higher LCG-based risk-score was accompanied by a lower immunity score independent of the StromalScore, ImmuneScore, and EstimateScore (Figure 8B). In addition, we explored the differences in immune cell infiltration. As shown in Figure 8C, a higher LCG-based risk-score group was accompanied by a lower infiltration of memory B cells, plasma cells, $CD8^+$ T cells, resting memory $CD4^+$ T cells, γδ T cells, and resting NK cells, while it was accompanied by a higher infiltration of M0 macrophages and M2 macrophages (Figure 8C).

Prognosis prediction model based on LCG-based risk-score and clinical factors

For further development of the prognosis prediction model, we put clinical data into the construction of nomograms. Table 1 showed the clinical characteristics and genomic features of breast cancer (TCGA), and we excluded 'recurrence, M stage, radiation' whose proportion of subgroups was <10%. We identified 'age, clinical stage, N stage, and risk-score' as factors for constructing nomograms by univariate Cox regression and multivariate Cox regression in all cohorts (Table 2); we identified 'age, clinical stage, and risk-score' as factors for constructing nomograms in the luminal cohort and the TNBC cohort (Table 2). As shown in Figure 9, the total points provided a point-to-point survival prediction, such that a score of 44.9 corresponded to a 5-year death probability of 73.8%, a 3-year death probability of 49.2%, and a 1-year death probability of 8.42% in all cohort breast samples (Figure 9A). The same methods for prognosis prediction were used in the luminal and TNBC cohorts (Figures 9C,E). We divided the sample nomograms into a low-risk group and a high-risk group and the results showed that the low-risk group displayed a better prognosis than the high-risk group (Figures 9B,D,F, p <0.0001).

Inner verification was performed to assess the predictive ability of the above nomograms. ROC curves showed that the AUC values were 0.89 (1-year survival), 0.79 (3 years), and 0.75 (5 years) in all cohorts (Figure 10A); the AUC values were 0.84 (1-year survival), 0.83 (3 years), and 0.85 (5 years) in the luminal

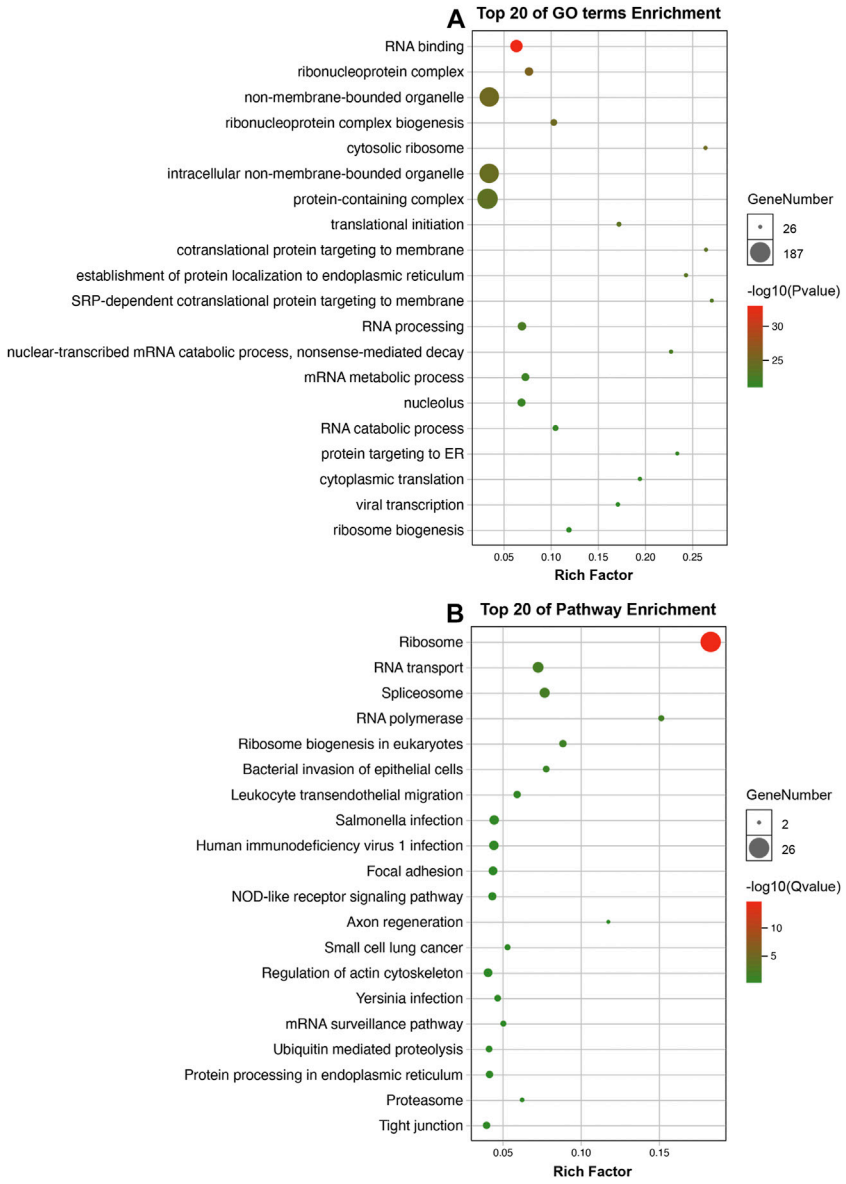


FIGURE 6
Enrichment analysis. **(A)** GO analysis of LLPS-related genes; **(B)** KEGG analysis of LLPS-related genes.

cohort (Figure 10E); and the AUC values were 0.95 (1-year survival), 0.84 (3 years), and 0.77 (5 years) in the TNBC cohort (Figure 10I). Calibration analysis was shown in Figure 10, which implied that nomograms were accurate in their prediction of breast cancer prognosis and its subtypes. The C-index was also calculated to assess the predictive ability of the nomograms, and Table 3 shows that the C-index was 0.784 [0.741–0.827] for nomograms in all cohorts, 0.803 [0.756–0.850] for nomograms in the luminal cohort, and 0.847 [0.759–0.934] for nomograms in the TNBC cohort.

Discussion

According to the National Cancer Report 2019, breast cancer has become the most common type of tumor in women, with more than 300,000 new breast cancers and more than 66,000 deaths every year (Siegel et al., 2020). Amongst them, HER2-negative breast cancer has to rely on paclitaxel-based combination chemotherapy because of the lack of effective molecular targeted therapy strategies. However, continuous low-sensitivity chemotherapy can easily cause drug resistance, reduce chemotherapy's

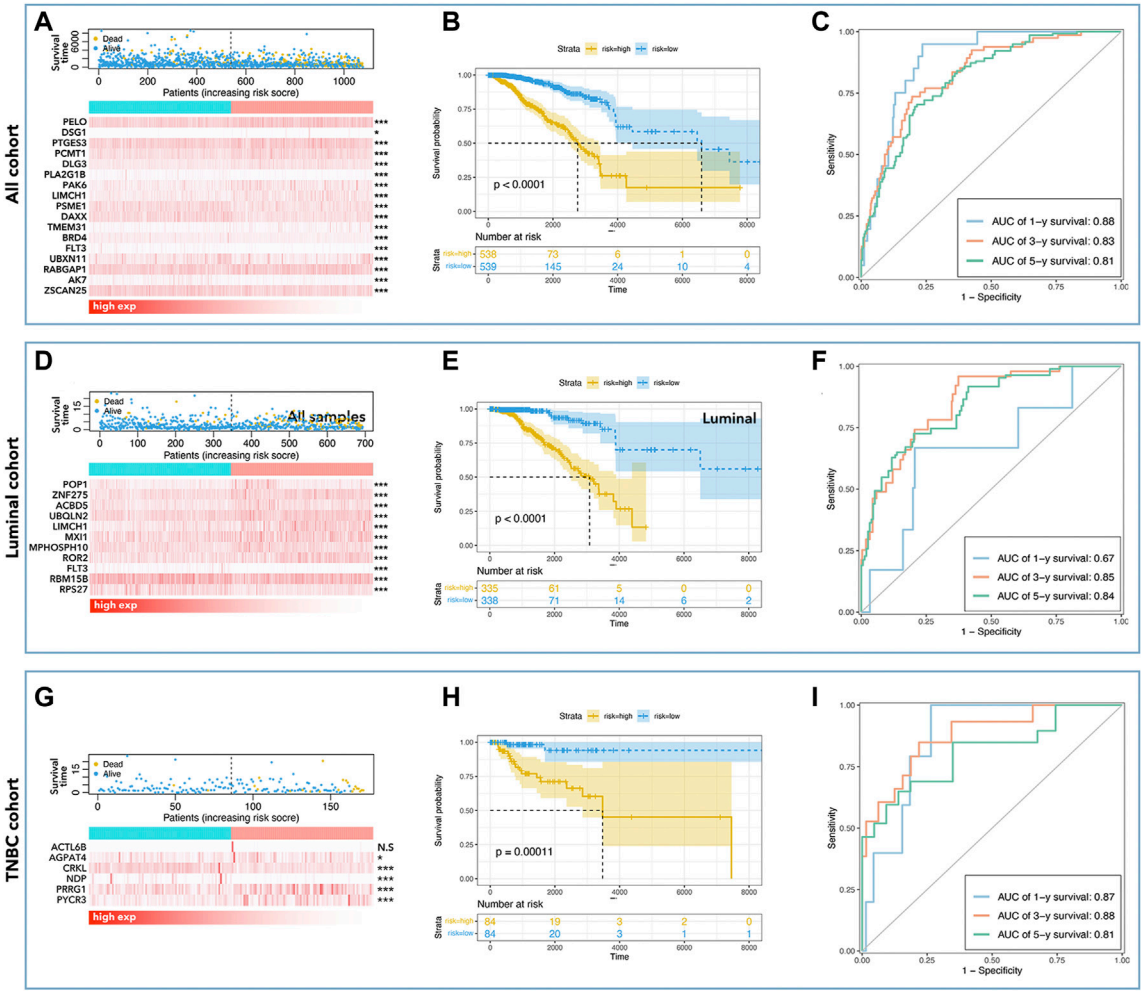


FIGURE 7
LLPS-related multi-gene risk-score. Risk-score-mortality relationship and selected LLPS-related gene expression differences are shown for all cohort **(A)**, luminal cohort **(D)**, and TNBC cohort **(G)**. The K-M curve showed that low-risk had better prognosis in all cohorts **(B)**, the luminal cohort **(E)**, and in the TNBC cohort **(H)**. The AUC values of risk-score in all cohorts **(C)** were 0.88 (1 year), 0.83 (3 years), and 0.81 (5 years), in the luminal cohort **(F)** were 0.67 (1 year), 0.85 (3 years), and 0.84 (5 years), and in the TNBC cohort **(I)** were 0.87 (1 year), 0.88 (3 years), and 0.81 (5 years).

clinical benefits, and increase the risk of recurrence and metastasis (Foulkes et al., 2010). Therefore, the development of sensitive chemotherapy is pivotal to the current clinical treatment of breast cancer, but is still a very difficult problem in scientific research.

In eukaryotic cells, there are many structures lacking membranes such as nucleoli, premature cell leukemia nuclei (PML NB), P bodies in *C. elegans*, etc., and they perform key functions in metabolic processes (Boeynaems et al., 2018). Previous studies have pointed out that membrane-free structures are formed from LLPS, which is also named membrane-free condensate or biological condensate. Recently, it was reported that LLPS was involved in neurological diseases and tumor processes. For example, tau species that formed LLPS under cellular conditions could serve as intermediates for tau aggregate formation (Wegmann et al., 2018); cAMP-dependent protein kinase (type I regulatory subunit) produced LLPS as part of their functional role in cAMP signaling to form biomolecular condensates enriched in cAMP and PKA activity, which was critical for effective cAMP compartmentation and played roles in atypical liver cancer (Zhang et al., 2020); YAP protein formed a liquid aggregate in the nucleus and promoted the growth of breast cancer cells by inducing the transcription of oncogenes (Li et al., 2021a). In addition, some studies were reported to apply LCGs to construct multi-gene risk-scores to

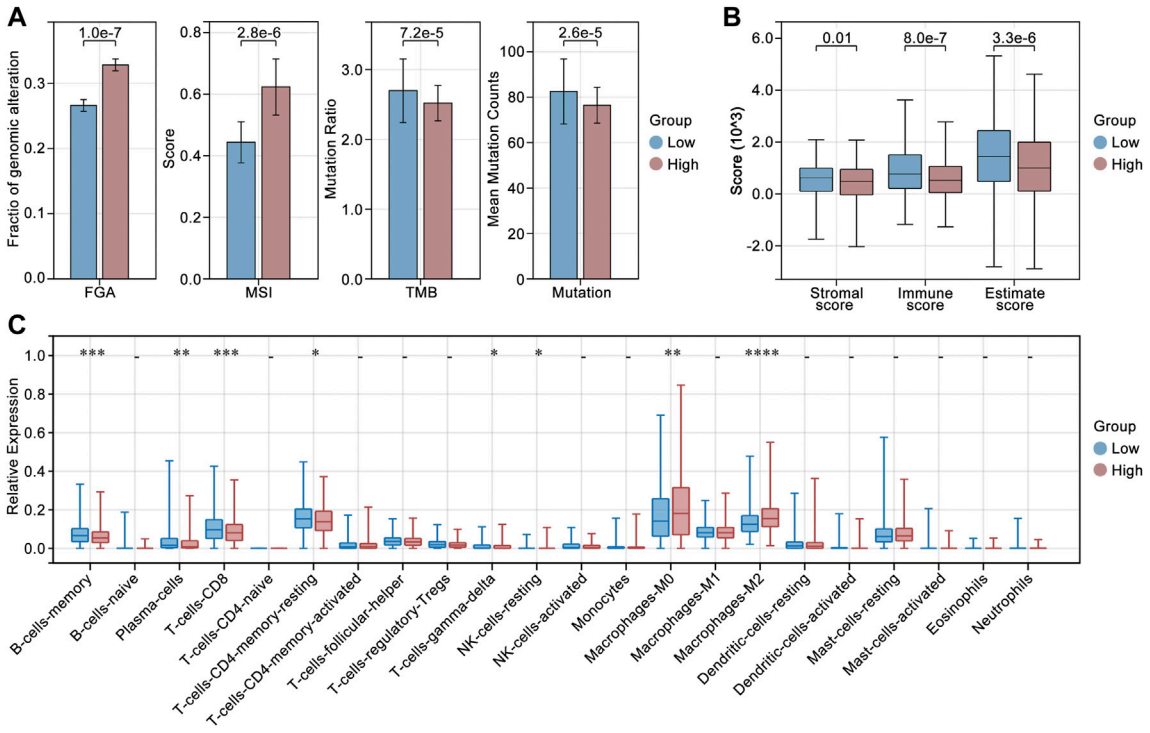


FIGURE 8
Genomic instability and tumor immunity analysis. **(A)** Differences in fraction of genomic alteration (FGA), microsatellite instability (MSI), gene mutation (mutation) and tumor mutation burden (TMB) in high risk-score and low risk-score groups. **(B)** Immune score in high risk-score and low risk-score groups. **(C)** Differences in immune cell infiltration in high risk-score and low risk-score groups.

assess the effects of LLPS in the prediction of tumor prognosis in ovarian epithelial cancer and lung cancer (Qiu et al., 2021; Zhuge et al., 2021). These two studies showed that LCGs were useful in distinguishing subgroups in which one group had a better prognosis, while the other had a worse prognosis. However, the visual prognosis prediction tool used was not given. So, we intended to use LCGs and clinical factors to construct visual prognosis prediction tools.

In this study, we identified differentially expressed pLCGs in breast cancer, and then selected pLCGs as factors to construct the prognosis prediction tool. Twelve genes were identified and samples were divided into low-expression and high-expression groups by the expression of those 12 genes in all cohorts. Although the 12-gene-based nomogram showed medium accuracy in the prediction of prognosis (AUC>0.7, Figure 4A), its performance in inner verification was not good enough (0.7 > AUC>0.5, Figure 5B) to apply it in the prediction of breast cancer prognosis. It was worse in the prediction of 1-year survival and 3-year survival in breast cancer (all cohorts) (data not shown). Meanwhile, we observed that the results in the luminal and TNBC cohorts were too weak to be useful. However, to our surprise, the LCG-based risk-score exactly divided breast cancer samples into better prognosis groups and worse prognosis in all cohorts, the luminal cohort, and the TNBC cohorts (Figures 7B,E,H). Afterwards, we applied LCG-based risk-scoring to construct a nomogram. Although the ROC curve displayed medium strength in the prediction of prognosis (AUC value > 0.8, Figures 7C,F,I; C-index>0.7, data not shown), calibration analysis showed that the results were not good enough (data not shown). So, we combined risk-score and clinical factors to construct a better nomogram for the prediction of prognosis in breast cancer. Inner verification showed that the LLPS-related-gene-based and clinical-factor-based nomograms gave good results for prediction of breast cancer prognosis, especially in the TNBC cohort, for which the AUC value (Figure 10I) of 1-year survival prediction was 0.95, 3-year survival prediction was 0.84, 5-year survival prediction was 0.77, C-index was 0.847 (Table 3), and the calibration analysis results were good (Figures 10J–L). Unfortunately, we did not find any available data that contained both gene expression profiles and clinical characteristics together to perform external verification.

TABLE 1 | Clinical and genomic characteristics of subtypes of breast cancers. (1) Data from TCGA, from which the total sample was 1077 records, the luminal cohort was 693 records, the TNBC cohort was 171 records, and the other cohort was 213 records. (2) The asterisk * means that factors whose proportion of subtypes was <10% were selected for further analysis.

Item		All samples (n = 1077)	Luminal (n = 693)	TNBC (n = 171)	Others (n = 213)
Clinical characteristics			**Number (%)**		
Age (years)[a]	*1 (<45)*	158 (14.7)	98 (14.1)	32 (18.7)	28 (13.1)
	2 (45~64)	584 (54.2)	360 (51.9)	97 (56.7)	127 (59.6)
	3 (>64)	335 (31.1)	235 (33.9)	42 (24.6)	58 (27.2)
Recurrence	*1 (Yes)*	100 (9.3)	58 (10.4)	16 (9.4)	26 (12.2)
	2 (No)	778 (72.2)	501 (72.3)	125 (73.1)	152 (71.4)
	3 (Na)	199 (18.5)	134 (19.3)	30 (17.5)	35 (16.4)
M stage	*1 (No)*	896 (83.2)	575 (83.0)	149 (87.1)	172 (80.8)
	2 (Yes)	21 (1.9)	12 (1.7)	3 (1.8)	6 (2.8)
	3 (Na)	160 (14.9)	106 (15.3)	19 (11.1)	35 (16.4)
N stage[a]	*1 (N0)*	510 (47.4)	314 (45.3)	107 (62.6)	89 (41.8)
	2 (N1-3)	547 (50.8)	365 (52.7)	64 (37.4)	118 (55.4)
	3 (Na)	20 (1.9)	14 (2.0)	0 (0.0)	6 (2.8)
T stage[a]	*1 (T1-2)*	899 (83.5)	584 (84.3)	149 (87.1)	166 (77.9)
	2 (T3-4)	174 (16.2)	106 (15.3)	21 (12.3)	47 (22.1)
	3 (Na)	4 (0.4)	3 (0.4)	1 (0.6)	0 (0.0)
Clinical stage[a]	*1 (I~II)*	793 (73.6)	505 (72.9)	145 (84.8)	143 (67.1)
	2 (III~IV)	266 (24.7)	176 (25.4)	23 (13.5)	67 (31.5)
	3 (Na)	18 (1.7)	12 (1.7)	3 (1.8)	3 (1.4)
Radiation	*1 (yes)*	977 (90.7)	634 (91.5)	154 (90.1)	189 (88.7)
	2 (No)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	3 (Na)	100 (9.3)	59 (8.5)	17 (9.9)	24 (11.3)
Genomic characteristics			Median ± Sd.		
TMB[a]		2.746 ± 9.317	2.546 ± 9.729	3.350 ± 4.815	2.924 ± 10.723

TMB: tumor mutation burden

[a]Factors are selected into further analysis whose proportion of subgroups ≥ 10%.

However, we still performed independent verification by GEO data. As shown in Supplementary Figure S2, the LCG-based risk-score results were not good enough. The ROC value of 1/3/5/7-year overall survival were 0.30, 0.58, 0.64 and 0.61 (Supplementary Figure S2A), while the ROC value of 1/3/5/7-year metastasis-free survival were 0.30, 0.58, 0.64 and 0.61 (Supplementary Figure S2B). Furthermore, the LCG-based risk-score and clinical factor (age) were applied simultaneously to construct a nomogram, and the results showed that the ROC values for 1/3/5/7-year overall survival were 0.90, 0.7, 0.77 and 0.75 (Supplementary Figure S2C), while the ROC value for 1/3/5/7-year metastasis-free survival were 0.94, 0.74, 0.76 and 0.74 (Supplementary Figure S2D).

In general, this study demonstrated that the prediction ability of nomograms based only on LLPS-related genes was not good enough to be applied in breast cancer therapy. However, the prognosis prediction tools, based on LCG-based risk-scores and clinical factors, had medium accuracy, which means that LCGs are useful for constructing a prognosis prediction model when combined with clinical factors. Next, we will collect our gene expression profiles and clinical data to make an external verification to further assess the prognosis prediction tools in the TNBC cohort.

The limitations of this work are obvious: 1. This work did not detect the real expression levels of those genes used to calculate risk-scores in fresh frozen breast cancer tissues by RNA-sequence assay. The low expression of those selected genes could result in large fluctuations in the assessment of prognosis; 2. Because of limitations in the collection of clinical data (breast cancer has good prognosis, so it is not easy to make a follow-up over 5 years or more in a limited time) this study did not perform external verification. 3. This study did not make a comparison between our prognostic model and another multi-gene-based model in the prediction of breast cancer prognosis.

TABLE 2 | Identifying clinical factors in the nomogram. (1) Age, clinical stage, N stage, and risk-score were selected in all cohorts; (2) Age, clinical stage, and risk-score were selected in the luminal cohort; (3) Age, clinical stage, and risk-score were selected in the TNBC cohort.

Item			Univariate cox Analysis			Multivariate cox Analysis		
			HR	95% Cl	*p* value	HR	95% Cl	*p* value
All	Age*	<45						
		45~64	1.078	0.650–1.788	0.771	1.359	0.813–2.269	0.242
		>64	2.354	1.406–3.943	0.001	2.884	1.702–4.886	<0.001
	TMB		0.967	0.898–1.042	0.380			
	Clinical stage*	I~II					–	
		III~IV	2.789	1.982–3.923	<0.001	2.275	1.489–3.476	<0.001
	T stage	T1~2						
		T3~4	1.871	1.278–2.741	0.001			
	N stage*	N0						
		N1~3	2.215	1.543–3.181	<0.001	1.679	1.067–2.643	0.025
	RiskScore*		1.051	1.039–1.063	<0.001	1.051	1.039–1.063	<0.001
TNBC	Age*	<45						
		45~64	0.831	0.273–2.529	0.745	3.460	0.657–18.226	0.143
		>64	1.930	0.590–6.320	0.277	6.664	1.272–34.922	0.025
	TMB		0.830	0.597–1.153	0.267			
	Clinical stage*	I~II						
		III~IV	6.284	2.555–15.454	<0.001	4.340	1.397–13.486	0.011
	T stage	T1~2						
		T3~4	2.429	0.802–7.356	0.116			
	N stage	N0						
		N1~3	4.123	1.596–10.649	0.003			
	RiskScore*		1.038	1.025–1.051	<0.001	1.040	1.021–1.059	<0.001
Luminal	Age*	<45						
		45~64	0.886	0.445–1.765	0.732	1.054	0.520–2.135	0.885
		>64	2.276	1.159–4.470	0.017	2.639	1.308–5.232	0.007
	TMB		1.006	0.959–1.055	0.804			
	Clinical stage*	I~II						
		III~IV	2.289	1.467–3.571	<0.001	2.340	1.471–3.720	<0.001
	T stage	T1~2						
		T3~4	1.643	0.995–2.714	0.053			
	N stage	N0						
		N1~3	1.881	1.183–2.991	0.008			
	RiskScore*		1.104	1.081–1.128	<0.001	1.096	1.071–1.121	<0.001

TABLE 3 | C-index of nomograms. The C-index value of nomograms was 0.784 [0.741–0.827] in all cohorts, 0.803 [0.756–0.850] in the luminal cohort, and 0.847 [0.759–0.934] in the TNBC cohort.

Item	C index	Lower value	Upper value
All cohort	0.784	0.741	0.827
Luminal cohort	0.803	0.756	0.850
TNBC cohort	0.847	0.759	0.934

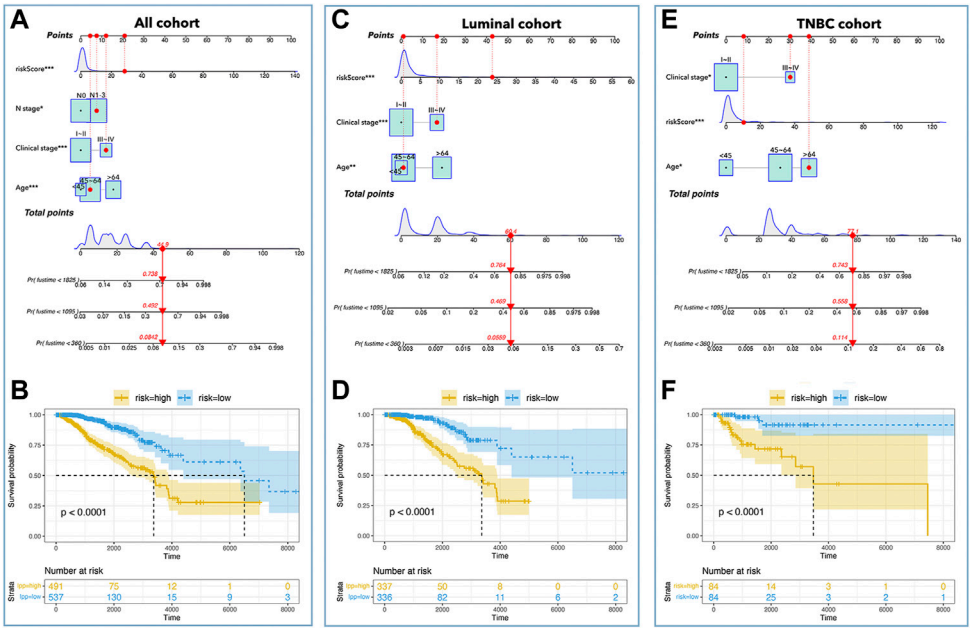


FIGURE 9
Nomograms in subtypes of breast cancer. Nomograms of all cohorts **(A)**, luminal cohort **(C)**, and TNBC cohort **(E)**. The K-M curve showed that a low-risk-score had better prognosis in all cohorts **(B)**, luminal cohort **(D)**, and TNBC cohort **(F)**.

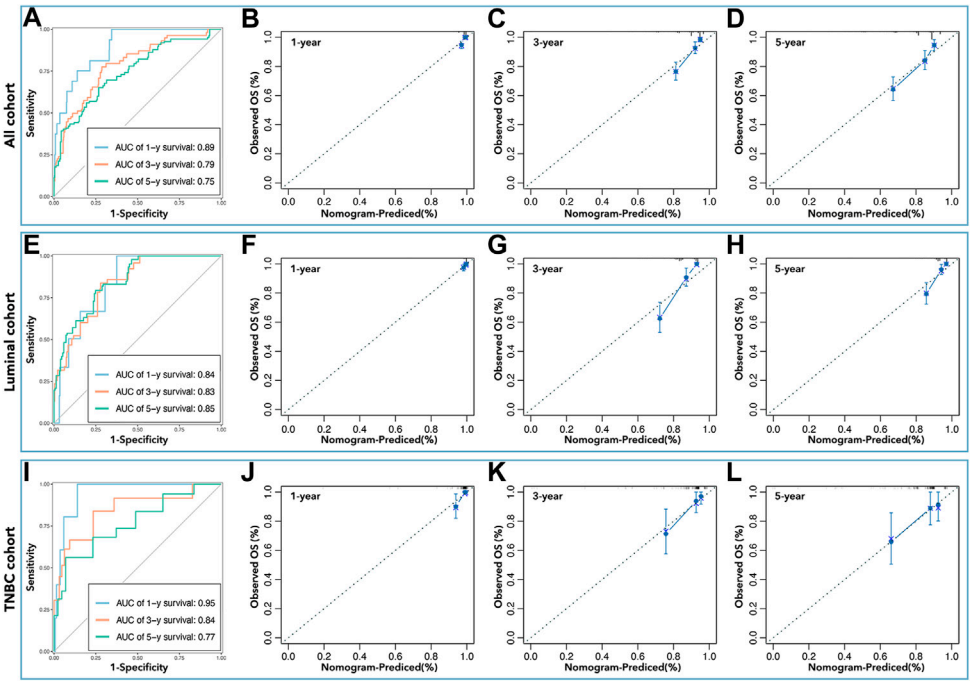


FIGURE 10
Inner verification of nomograms. The AUC values of nomograms in all cohorts **(A)** were 0.89 (1 year), 0.79 (3 years), and 0.75 (5 years), in the luminal cohort **(E)** were 0.84 (1 year), 0.83 (3 years), and 0.85 (5 years), and in the TNBC cohort **(I)** were 0.95 (1 year), 0.84 (3 years), and 0.77 (5 years). Calibration analysis was performed to assess the prediction accuracy of nomograms in all cohorts **(B–D)**, the luminal cohort **(F–H)**, and the TNBC cohort **(J–L)**.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

The original research design was made by WZ-Y, LP-P, and HY-Q, and the final paper revision was performed by WZ-Y. Data collection and collation were performed by HY-Q, data analysis was performed by LP-P, SK, and YK-X. The original paper was written by HY-Q, and revised by LP-P. ZW-J independently verified the results of this study.

Funding

All works of this paper are supported by "High-level Talent Research Fund of Xiaoshan District" from WZ-Y, ZW-J, and LP-P, and supported by Hangzhou Municipal Health Commission General Projects (No. B20220666).

Acknowledgments

We would like to extend our sincere gratitude to The First People's Hospital of Xiaoshan District and to Dean Yao Xiao-Cong for encouragement and support to young doctors in performing this research.

Conflict of interest

The authors declare that the research was performed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.834471/full#supplementary-material

SUPPLEMENTARY FIGURE S1
LASSO analysis to narrow the candidate genes. Binomial deviance and coefficients in all cohorts **(A,D)**, in the luminal cohort **(B,E)**, and the TNBC cohort **(C,F)**.

SUPPLEMENTARY FIGURE S2
External verification of LCG-based risk-score and age. **(A)** ROC value of 1/3/5/7-year overall survival in LCG-based risk-score. **(B)** ROC value of 1/3/5/7-year metastasis-free survival in LCG-based risk-score. **(C)** ROC value of 1/3/5/7-year overall survival in LCG-based risk-score and age. **(D)** ROC value of 1/3/5/7-year metastasis-free survival in LCG-based risk-score and age.

References

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* 176, 419–434. doi:10.1016/j.cell.2018.12.035

Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., et al. (2018). Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* 28, 420–435. doi:10.1016/j.tcb.2018.02.004

Du, M., and Chen, Z. J. (2018). DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* 361, 704–709. doi:10.1126/science.aat1022

Foulkes, W. D., Smith, I. E., and Reis-Filho, J. S. (2010). Triple-negative breast cancer. *N. Engl. J. Med.* 363, 1938–1948. doi:10.1056/NEJMra1001389

Hahn, S. (2018). Phase separation, protein disorder, and enhancer function. *Cell* 175, 1723–1725. doi:10.1016/j.cell.2018.11.034

Jiang, F., Wu, C., Wang, M., Wei, K., and Wang, J. (2022). An autophagy-related long non-coding RNA signature for breast cancer. *Comb. Chem. High. Throughput Screen.* 25, 1327–1335. doi:10.2174/1386207324666210603122718

Jiang, F., Wu, C., Wang, M., Wei, K., and Wang, J. (2021). Identification of novel cell glycolysis related gene signature predicting survival in patients with breast cancer. *Sci. Rep.* 11, 3986. doi:10.1038/s41598-021-83628-9

Li, R. H., Tian, T., Ge, Q. W., He, X. Y., Shi, C. Y., Li, J. H., et al. (2021). A phosphatidic acid-binding lncRNA SNHG9 facilitates LATS1 liquid-liquid phase separation to promote oncogenic YAP signaling. *Cell Res.* 31, 1088–1105. doi:10.1038/s41422-021-00530-9

Li, X., Chen, J., Yu, Q., Huang, H., Liu, Z., Wang, C., et al. (2021). A signature of autophagy-related long non-coding RNA to predict the prognosis of breast cancer. *Front. Genet.* 12, 569318. doi:10.3389/fgene.2021.569318

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W. R., Damianov, A., et al. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature* 587, 145–151. doi:10.1038/s41586-020-2703-0

Qiu, Y., Pan, M., and Chen, X. (2021). A liquid-liquid phase separation-related gene signature as prognostic biomarker for epithelial ovarian cancer. *Front. Oncol.* 11, 671892. doi:10.3389/fonc.2021.671892

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., et al. (2019). m(6 A enhances the phase separation potential of mRNA. *Nature* 571, 424–428. doi:10.1038/s41586-019-1374-1

Shen, Y., Peng, X., and Shen, C. (2020). Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics* 112, 2640–2646. doi:10.1016/j.ygeno.2020.02.015

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *Ca. A Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., et al. (2018). Tau protein liquid-liquid phase separation can initiate tau aggregation. *Embo J.* 37, e98049. doi:10.15252/embj.201798049

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y. S., Fox, A. H., et al. (2018). Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. *Mol. Cell* 70, 1038–1053. e1037. doi:10.1016/j.molcel.2018.05.019

Zhang, J. Z., Lu, T. W., Stolerman, L. M., Tenner, B., Yang, J. R., Zhang, J. F., et al. (2020). Phase separation of a PKA regulatory subunit controls cAMP compartmentation and oncogenic signaling. *Cell* 182, 1531–1544. e1515. doi:10.1016/j.cell.2020.07.043

Zheng, J., Wu, Z., Qiu, Y., Wang, X., and Jiang, X. (2022). An integrative multi-omics analysis based on liquid-liquid phase separation delineates distinct subtypes of lower-grade glioma and identifies a prognostic signature. *J. Transl. Med.* 20, 55. doi:10.1186/s12967-022-03266-1

Zhuge, L., Zhang, K., Zhang, Z., Guo, W., Li, Y., and Bao, Q. (2021). A novel model based on liquid-liquid phase separation-Related genes correlates immune microenvironment profiles and predicts prognosis of lung squamous cell carcinoma. *J. Clin. Lab. Anal.* 36, e24135. doi:10.1002/jcla.24135

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fgene.2022.867872

Check for updates

OPEN ACCESS

EDITED BY
Nguyen Minh Duc,
Pham Ngoc Thach University of Medicine, Vietnam

REVIEWED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China
Ian Kerr,
University of Nottingham, United Kingdom

*CORRESPONDENCE
Jing Wang,
wangj@sysucc.org.cn
Zhongping Chen,
chenzhp@sysucc.org.cn
Jian Wang,
wangjian2@sysucc.org.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 04 March 2022
ACCEPTED 08 July 2022
PUBLISHED 11 October 2022

CITATION
Hu Y, Zhu S, Xu R, Wang M, Chen F, Zhang Z, Feng B, Wang J, Chen Z and Wang J (2022), Delta-catenin attenuates medulloblastoma cell invasion by targeting EMT pathway. *Front. Genet.* 13:867872. doi: 10.3389/fgene.2022.867872

COPYRIGHT
© 2022 Hu, Zhu, Xu, Wang, Chen, Zhang, Feng, Wang, Chen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Delta-catenin attenuates medulloblastoma cell invasion by targeting EMT pathway

Yuanjun Hu[1,2,3†], Sihan Zhu[1,2†], Rizhen Xu[4], Manxia Wang[5], Furong Chen[1], Zeshun Zhang[4], Binghong Feng[5], Jian Wang[2*], Zhongping Chen[1,2*] and Jing Wang[1,2*]

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China, [2]Department of Neurosurgery/Neuro-Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China, [3]Department of Neurosurgery, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [4]Department of Surgery, Guangzhou University of Chinese Medicine, Guangzhou, China, [5]Department of Pharmacology, Guangdong Pharmaceutical University, Guangzhou, China

Background: Medulloblastoma is the most common pediatric malignant tumor in central nervous system. Although its prognosis has been improved enormously by the combination treatments with surgery, radiotherapy, and chemotherapy, it still could progress *via* invasion and distant dissemination. We aimed to investigate molecular mechanisms of medulloblastoma invasion in the current work.

Methods: The gene expression profile of medulloblastoma were analyzed based on the data deposited in Gene Expression Omnibus (GEO) and filtered according to brain specific proteins in the Uniprot. Delta-catenin was identified and further analyzed about its expression and roles in the prognosis of medulloblastoma patient. The function of delta-catenin on cell invasion and migration were investigated by transwell and wound healing assay. Whether delta-catenin participates in the epithelial-mesenchymal transition (EMT) regulated invasion was also studied.

Results: Delta-catenin expression was highly upregulated in tumor tissues compared to normal tissues from medulloblastoma patients in five independent, nonoverlapping cohorts. Furthermore, delta-catenin expression level was upregulated in WNT subgroup, and significantly correlated with better prognosis, and associated with metastasis through GEO database analysis. Functional assays indicated that delta-catenin inhibited medulloblastoma cell invasion and migration through regulating the key factors of EMT pathway, such as E-cadherin and vimentin.

Conclusion: Delta-catenin might be a positive predictor for prognosis of medulloblastoma patients, through attenuating medulloblastoma cell invasion by inhibiting EMT pathway.

KEYWORDS

medulloblastoma, prognosis, invasion, delta-catenin, EMT

Introduction

Medulloblastoma is the most common malignant pediatric brain tumor. In recent decades, therapeutic strategy and prognosis have been based on risk stratification of medulloblastoma patients by age at presentation, extent of tumor resection, and tumor metastases. Molecular profiling has revealed four molecular subgroups of medulloblastoma—wingless (WNT), sonic hedgehog (SHH), Group 3, and Group 4—with distinct molecular and clinical profiles (Northcott et al., 2011; Northcott et al., 2012; Coltin, 2021). The latest WHO classification of tumors of the central nervous system simplified medulloblastoma into three subgroups—WNT, SHH and non-WNT/non-SHH (Louis et al., 2021). Although this has improved the guidance for treatment, medulloblastoma invasion and dissemination occurs in all subgroups, and is still the major cause of medulloblastoma deaths (Fults et al., 2019). Leptomeningeal dissemination of medulloblastoma has long been considered to occur *via* cerebrospinal fluid, and hematogenous transfer has recently been described (Garzia et al., 2018). Metastases are found in 5%–45% of medulloblastoma cases at initial diagnosis, least often in WNT subgroup (5–10%) and most often in Group 3 (40–45%) (Juraschka and Taylor, 2019). Although nearly all medulloblastoma cases ultimately progress to metastasis *via* invasion and dissemination, this clinical challenge remains the least understood aspect of medulloblastoma pathogenesis and disease progression.

Delta-catenin, a catenin encoded by CTNND2, is a known neuroprotein that interacts with presenilin-1 (Zhou et al., 1997). Actually, abnormal function of delta-catenin is associated with several neurodevelopmental disorders, such as cri-du-chat syndrome, autism and schizophrenia (Medina et al., 2000; Vrijenhoek et al., 2008; Turner et al., 2015). The important role of delta-catenin in neurodevelopment may related with its function on dendrite development (Martinez et al., 2003; Arikkath et al., 2008). Recent work has revealed delta-catenin is overexpressed in a series of peripheral tissue neoplasms such as prostate and lung, suggesting its value as a cancer biomarker (Lu et al., 2014). Delta-catenin was mainly studied as a oncoprotein in several tumor types (Huang et al., 2018; Shen et al., 2021; Zeng et al., 2009; Zhang et al., 2015). However, the anti-tumor property of delta-catenin was also reported in a few studies (Westbrook et al., 2005; Frattini et al., 2013). Delta-catenin is a member of the p120-catenin family of catenin protein and it could bind to E-cadherin in competition with p120ctn, which belongs to the same family, indicating that it may participate in cell adhesion and EMT (Reynolds and Roczniak-Ferguson, 2004; Zhang et al., 2014). EMT, which involves changes in morphology and increased cell motility, is currently considered to be a major pathway in metastasis. Because medulloblastoma is a neuroepithelial-derived tumor, its invasion and dissemination are likely inseparable from EMT (Fults et al., 2019).

Although the role of delta-catenin has been examined in glioma (Shimizu et al., 2019; Wang et al., 2011), there are no prior reports of its role in medulloblastoma. We therefore aimed to elucidate the molecular mechanisms of medulloblastoma invasion, to improve the therapeutic response and prolong survival in medulloblastoma patients. We examined delta-catenin expression in medulloblastoma, and its effect on invasion and dissemination, using bioinformatic analysis and *in vitro* experiments. Delta-catenin expression was associated with improved survival. Delta-catenin alleviated medulloblastoma invasion *in vitro* by targeting EMT pathway.

Materials and methods

Bioinformatics analysis

The Differentially Expressed Genes (DEGs) were analyzed using the GEO database (GSE74195, GSE66354, and GSE86574) by "limma" R package. After Benjamini–Hochberg (BH) multiple test adjustment, DEGs with absolute $\log_2$ fold change (FC) > 1 and $p < 0.05$ were considered to be included for subsequent analysis. Then the DEGs were considered as delta-catenin related compared with the brain specific protein list in the Uniprot.

We further compared delta-catenin expression use GEO datasets between normal and medulloblastoma patient tissues. Then we used the clinical information from GEO datasets and the medulloblastoma patient cohort (74) from Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China) (Table 1) to analyze the relevance between delta-catenin expression and clinical characteristics such as prognosis, molecular subgroup and metastasis. The patients data from GSE85217 were divided into two independent nonoverlapping cohorts (7:3 ratio) for mutual verification (Chen et al., 2021; Li et al., 2020). All the GEO datasets used in the bioinformatics analysis were listed in Supplementary Table S1.

The bioinformatics analysis was conducted mainly in R v. 38.0. The "combat" function of the "sav" R package was used for batch-effect correction. The "Coxph" function, and the "survival" R package, were used to estimate the relationship between delta-catenin expression and survival. We used R package "survivalROC" to plot the receiver operating characteristic (ROC) curve of overall survival and calculate the area under the curve (AUC). 0.5 generally indicates some predictive ability. The higher the AUC is, the more accurate the prediction result is. BioGRID7 (https://thebiogrid.org/) is a biomedical interaction repository. The database can be used to retrieve publications on protein and genetic interactions, chemical interactions and post-translational modifications of important model organism species. Gene set enrichment analysis (GSEA) was performed using the GSEA software and its results were visualized using "Cluster Profiler" and "ggplot2" in R.

TABLE 1 Summary of clinical information of medulloblastoma cohorts.

	GSE85217 dataset 1	GSE85217 dataset 2	GSE30074	GSE21140	SYSUCC cohort
Case included (*n*)	534	229	30	103	74
Age (mean ± SD)	5.6 (±8.01)	8.6 (±7.39)	NA	9.4 (±8.39)	11.3 (±8.62)
Overall survival	N = 429	N = 183	N = 30	NA	N = 74
Year (mean ± SD)	4.77 (±3.70)	5.41 (±3.92)	4.25 (±2.29)	—	3.52 (±3.55)
Status (alive/dead)	309/120	144/39	18/12	—	44/30
Gender	N = 411	N = 183	N = 30	N = 103	N = 74
Male	288	132	19	63	52
Female	123	51	11	40	22
Subgroup	N = 763	NA	NA	N = 103	NA
WNT	70	—	—	8	—
SHH	223	—	—	33	—
Non-WNT/SHH	470	—	—	62	—
Metastasis	N = 400	N = 172	NA	NA	NA
M0	273	124	—	—	—
M1	127	48	—	—	—

Abbreviation: SHH, sonic hedgehog; WNT, wingless; M0, Non-Metastasis; M1, metastasis; SYSUCC, Sun Yat-sen University Cancer Center.

Cell culture and patient specimens

The medulloblastoma cell lines Daoy and ONS-76, and a tool cell line HEK293T were obtained from the Cell Bank of State Key Laboratory of Oncology in South China. The cells were cultured in DMEM with 10% fetal bovine serum (Gibco, Waltham, MA, United States) in a humidified incubator at 37°C with 5% CO_2. All the specimen from patients treated in our institute were residuals after surgery. The patients were all informed and provided signed consent regarding the use of their biological specimens and clinical information for research purposes. This study was approved by the ethics committee and institutional review board of Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China), in accordance with the Helsinki Declaration.

Immunohistochemistry and scoring

Paraffin-embedded tissue samples (n = 74) from medulloblastoma patients who underwent surgery at Sun Yat-sen University Cancer Center from 2011 to 2020 were used for immunohistochemical staining (IHC). Medulloblastoma diagnosis was based on the World Health Organization Classification of Central Nervous System Tumors (2016). Delta-catenin expression were detected by immunohistochemistry following standard protocol as per our previous paper (Wang et al., 2022). Briefly, the paraffin-embedded tissue slides were first heated at 65°C for 2 h, then sequentially deparaffinized in xylene, rehydrated *via* an ethanol gradient, antigen-retrieved using citric acid buffer (pH 6.0), and blocked with goat serum. Tissue samples were incubated overnight at 4°C with the antibody against delta-catenin (Cat# sc-81793; Santa Cruz Biotechnology, Inc.). HRP-labeled goat anti-rabbit/mouse antibody was then added to the slides for 1 h at 26°C. Finally, the tissues were stained with diaminobenzidine (DAB) and counterstained with hematoxylin. The slides were visualized and photographed using an automatic slide scanner (KF-PRO-020) at ×40 magnification. Quantitative image analyses were conducted using the HALO software (Indica labs, Corrales, MN, United States) using the multiplex IHC modules (Sun et al., 2019). The histochemistry score was used as the grouping criterion.

Lentiviral infection

To establish stable knockdown and overexpression (OE) lines of Daoy and ONS-76, the plasmids pSlenti-U6-shRNA (CTNND2)-CMV-EGFP-F2A-Puro-WPRE (OBiO Technology, Shanghai, China) and pEZ-Lv105-hCTNND2-Puro (GeneCopoeia, Guangzhou, China) were used. The plasmid was co-transfected with lentivirus packaging plasmids (PLP1, PLP2, and VSVG; Invitrogen, United States) into HEK293T with transfection reagent Lipofectamine 3000 (L3000008, Thermo Fisher Scientific, United States). The supernatant of HEK293T was collected twice, at 48 and 72 h after transfection. Cells (Daoy and ONS-76) were infected with the filtered supernatant containing virus. Puromycin (2 μg/ml) was added for 48 h after the infection and selected for 7 days. The target sequence for

delta-catenin was 5′-GCAACAACACTGCAAGCAA-3′ and 5′-GCTAAAGGCGAACACACTT-3′.

Real-time PCR

Total RNA from the cell lines was extracted following the protocol of RNA-Quick Purification Kit (ES Science, China). The reverse transcription were performed with HisScript III All-in-one RT SuperMix (Vazyme Biotech Co., Ltd.). Then real-time PCR was performed using ChamQ SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd. China), and triplicate samples were run on a Bio-Rad CFX96 qPCR system according to the manufacturer's protocol. The Ct values for delta-catenin was normalized to β-actin, and the $2^{-\Delta\Delta Ct}$ method was used for quantitative analysis. The primer sequences for delta-catenin were: 5′-AGGTCCCCGTCCATTGATAG-3′ and 5′-ACTGGTGCT GCAACATCTGAA-3′. The sequences for β-actin were: 5′- CTCCATCCTGGCCTCGCTGT-3′ and 5′- GCTGTC ACCTTCACCGTTCC-3′.

Western blotting

Cells were collected and washed twice in PBS, lysed in RIPA buffer (Biotime Biotechnology, Shanghai, China) with PMSF on ice for 30 min, then centrifuged at 12,000 × g for 15 min at 4°C. Protein concentration was determined with the BCA kit (Cat# 23227; Thermo Fisher Scientific). And the lysates were then subjected to SDS-PAGE (10% gel). The separated protein bands were transferred onto PVDF membrane (IPVH00010, Millipore, United States). After been blocking in milk (5% in TBST) for 1 h at 26°C, the membrane was probed with primary antibodies, such as for delta-catenin (ab184917; Abcam), E-cadherin (Cat# 3195; cell signaling technology), and vimentin (Cat# 10366-1-AP; Proteintech), beta-tubulin (Cat# 2128S, CST) and GAPDH (Cat# 2118S, CST) at 4°C for 12–16 h. After been washed in TBST for three times (15 min each time), the membranes were incubated with corresponding HRP-conjugated secondary antibodies (Cat# 7074S and Cat# 7076S, CST) for 1 h at 26°C. The bands were visualized using Ncm-ECL Ultra (New Cell & Molecular Biotech Co., Ltd., China), captured using an ChemiDoc Touch Imaging System (Bio-Rad, United States), and quantified using ImageJ (v1.8.0).

Immunofluorescence staining

Daoy (shNC and sh delta-catenin) cells were seeded in confocal dishs 24 h in advance. Then the dishs were fixed with 4% formaldehyde for 15 min at 26°C, rinsed three times in PBS (5 min each time), and blocked in blocking buffer (PBS with 5% v/v normal goat serum, 0.3% v/v Triton X-100) at 26°C for 60 min. The cells were then incubated overnight at 4°C in diluted primary E-cadherin antibody (1:100, Cat# 3195; CST). Then the slides were rinsed three times in PBS for 5 min each and incubated with fluorochrome-conjugated secondary antibody (diluted in blocking buffer at 1:500, Cat# 8889S, CST) for 1 h at 26°C in dark. The slides were rinsed three times again in 1× PBS for 5 min each, and DAPI was added to counter-stain the nucleus. Images were obtained using a Zeiss LSM 880 with fast Airyscan.

Cell migration and invasion assay

Cells were suspended in serum-free DMEM (2.5×10^5 cells/ml), 200 μl was inoculated into the upper chamber and 700 μl of complete medium was simultaneously added to the lower chamber. When placing the chamber, avoid air bubbles attached to the bottom of the chamber. After culturing for 15 h in incubator, the upper chamber was fixed in 4% paraformaldehyde for 30 min, then stained with 0.5% crystal violet for 7 min. The images were captured by the microscope (Nikon ECLIPSE Ni) when the chambers were dried and then counted the number of cells by ImageJ software (v1.8.0). For the invasion assay, 0.2% matrigel (Cat# 356234, Corning, NY, United States) was added into upper chamber before cells were seeded. For delta-catenin-overexpression (OE) cells, more cells (1×10^6 cells/ml) were suspended and culturing time extended to 24 h.

Wound healing assay

Cells were inoculated in 6-well plate (5×10^5 cells/well) and then the plates were placed in the incubator for pre-culture (at 37°C, 5% CO_2) overnight. Scratching on the cell surface with tip vertically next day and washing three times with PBS, then adding serum-free medium and placing in the incubator. Capturing images at the same point for each group in 0, 12, 24, and 48 h. The healing area between scratches were measured by ImageJ software (v1.8.0).

Statistical analysis

All the cell culture experiments were performed three independent replicates. Statistical analyses were performed using SPSS version 23.0 (IBM Corp, Armonk, NY) and GraphPad Prism version 8.0 (La Jolla, CA, United States). Survival analyses were performed using the Kaplan-Meier survival curves. All the data are shown as mean and standard deviation (mean ± SD). $p < 0.05$ was considered statistically significant for Student's t-tests or one-way ANOVA.

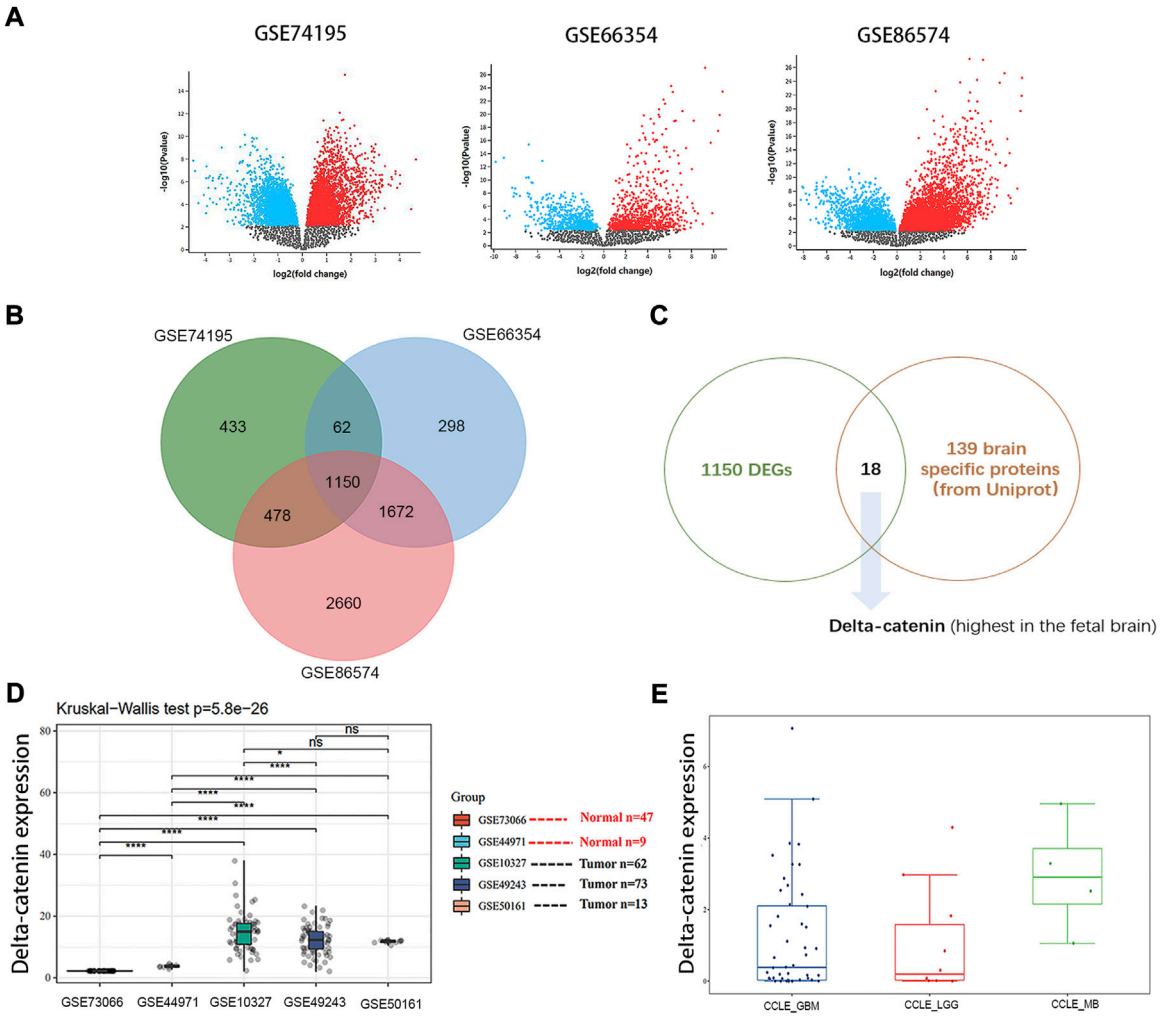


FIGURE 1
Delta-catenin is identified by a filtering strategy. **(A)** DEGs for medulloblastoma in three independent, nonoverlapping datasets were identified. **(B)** The overlap of the three datasets contains 1150 DEGs. **(C)** Flow diagram for screening out delta-catenin. **(D)** Delta-catenin expression is particularly high in medulloblastoma tissue compared to normal tissue controls ($p < 0.0001$) in five independent nonoverlapping cohorts. **(E)** Delta-catenin expression in various brain tumor cell lines (e.g., U87 and U251 for GBM, SF767 and SW1088 for LGG, Daoy and D283 for MB).

Results

Delta-catenin is identified by a filtering strategy

Several thousand differentially expressed genes were pulled out in the following datasets, 2123 in GSE74195, 3182 in GSE66354 and 5960 in GSE86574 (Figure 1A). The overlap of the three datasets contains 1150 DEGs (Figure 1B). Given that the cell origin of medulloblastoma is cerebellum and different subtypes arise within the distinct cell origins (Gibson et al., 2010), those brain specific proteins were more relevant to the development of medulloblastoma. Eighteen candidates were selected based on 139 brain specific proteins from Uniprot (Supplementary Table S2) and 1150 DEGs (Supplementary Table S3). Considering that medulloblastoma mainly occurs in children and the expression of delta-catenin is high in fetal brain (Turner et al., 2015), we then determined to further study the role of delta-catenin (Figure 1C).

Delta-catenin is verified highly expressed in medulloblastoma tissues and cells

Gene Expression Profiling Interactive Analysis (GEPIA) is an interactive web application for gene expression analysis based on 9736 tumors and 8587 normal samples from the TCGA and the GTEx databases using the output of a standard processing

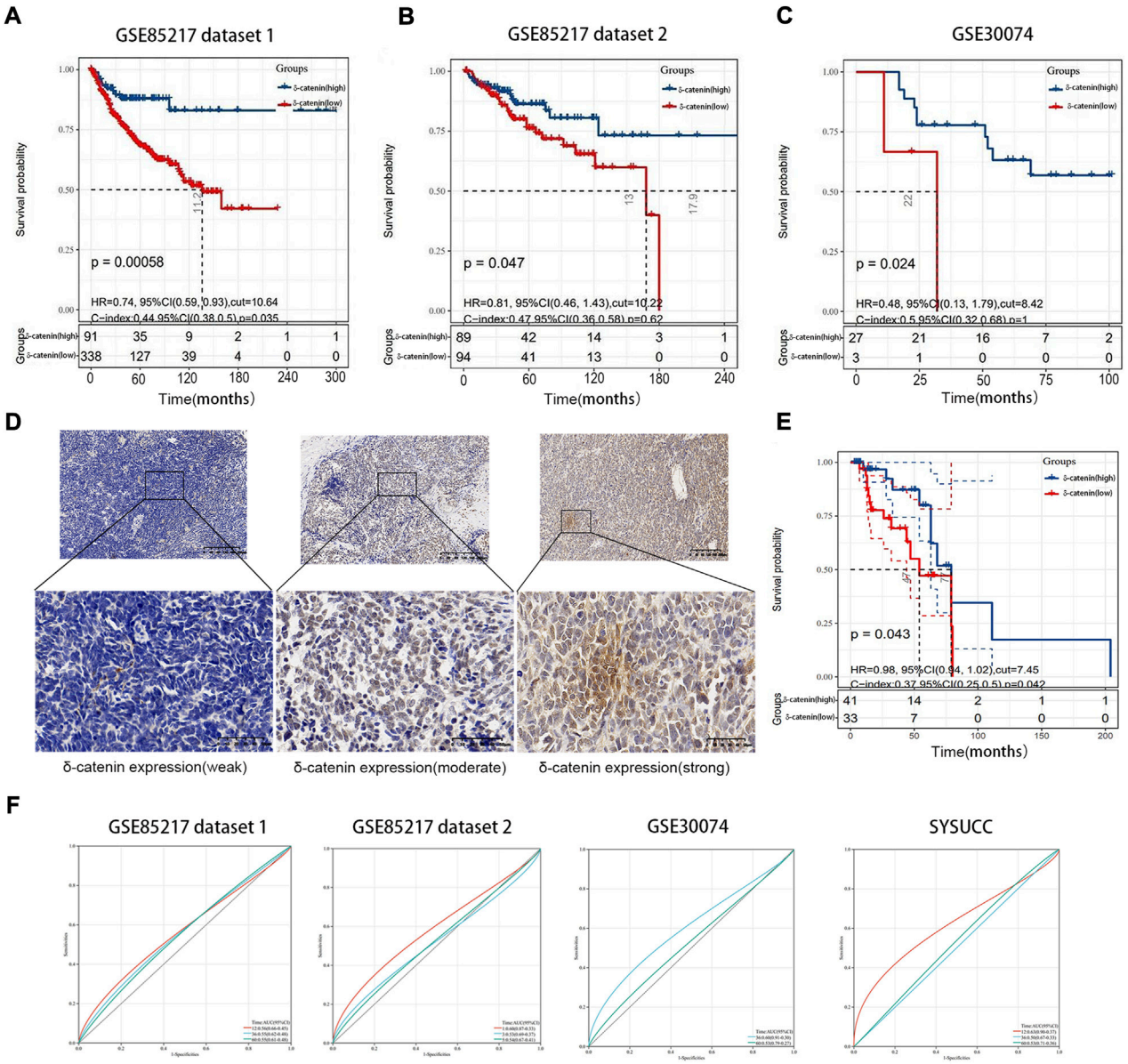


FIGURE 2
Delta-catenin expression and relevance for medulloblastoma prognosis. **(A-C)** Kaplan-Meier curves for medulloblastoma patients, based on delta-catenin expression. Data obtained from two nonoverlapping cohorts. **(D)** Immunohistochemistry reveals weak to strong delta-catenin expression in medulloblastoma tissues. **(E)** Kaplan-Meier curve for 74 medulloblastoma cases at the Sun Yat-sen University Cancer Center (SYSUCC), according to delta-catenin expression. **(F)** ROC curves of delta-catenin expression on the overall survival rate in four cohorts and AUC corresponding to each curve. The red, blue, and green lines represent the ROC curve of 1-year, 3-year, and 5-year overall survival rates.

pipeline for RNA sequencing data (Tang et al., 2017; Tang et al., 2019). Using online GEPIA, we analyzed expression of delta-catenin between tumor and normal tissue in various cancer types (Supplementary Figure S1). Delta-catenin expression was significantly higher in normal brain tissues and glioma patient tissues (both low-grade glioma and glioblastoma), than in other tumor types, supporting that delta-catenin may be a neural-specific protein. However, there are no medulloblastoma data in the GEPIA database which is based on TCGA and GTEx datasets. We then turned to GEO database, and examined the clinical significance of delta-catenin in medulloblastoma, 56 normal and 211 medulloblastoma samples (non-paired) were included. Its expression was significantly upregulated in medulloblastoma compared to the normal controls ($p < 0.0001$). Delta-catenin expression in medulloblastoma cell was comparable with that in glioblastoma and low-grade glioma cells (Figures 1D,E).

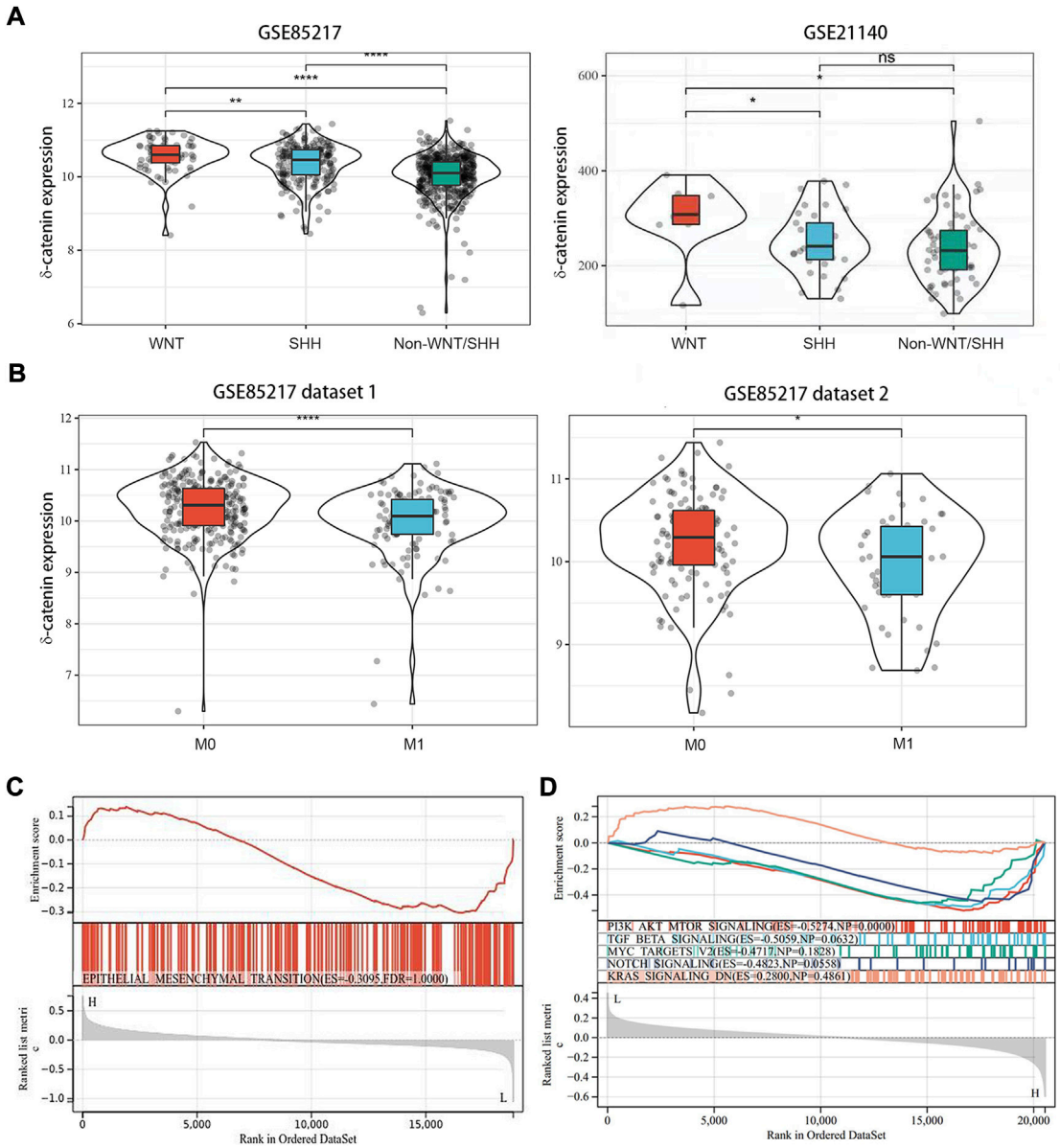


FIGURE 3
Delta-catenin expression in the medulloblastoma molecular subgroups, and its effects on metastasis. **(A)** Delta-catenin expression by medulloblastoma subgroup: delta-catenin was enriched in WNT subgroup. Data obtained from three nonoverlapping cohorts. SHH: sonic hedgehog; WNT: wingless. **(B)** Delta-catenin mRNA expression in tumor samples from non-metastatic and metastatic tumors. Delta-catenin expression is significantly higher in non-metastatic than in metastatic samples (Student's *t*-test). **(C)** GSEA of genes whose expression is downregulated by delta-catenin reveals the over-representation of EMT-signaling-related genes. **(D)** GSEA analysis showed that the metastasis-related signaling pathways such as PI3K/AKT, TGF-beta and so on were highly enriched in the low delta-catenin expression group.

High delta-catenin expression is a favorable prognostic factor in medulloblastoma

Using data from two independent and nonoverlapping medulloblastoma patient cohorts, we examined the association between delta-catenin expression level and medulloblastoma prognosis. We divided the Cavalli cohort dataset (GSE85217) (Cavalli et al., 2017) into two groups (7:3 ratio) for analysis. Overall survival analysis of medulloblastoma patients revealed that higher delta-catenin expression was directly associated with favorable overall survival (Figures 2A–C). To further confirm the finding, we detected delta-catenin expression in 74 medulloblastoma patients recruited in our institute, the typical staining pattern were

shown in Figure 2D. Similarly, high delta-catenin expression was associated with better prognosis (Figure 2E). These findings suggest that delta-catenin is a favorable predictor for medulloblastoma patients. ROC curve and AUC values reflect diagnostic values of markers. We drew 1-, 3-, and 5-year ROC curves of delta-catenin expression on the overall survival rate in four cohorts and calculated the AUC corresponding to each curve. The AUCs of the 1-, 3-, and 5-year GSE85217 dataset1 cohorts were 0.56, 0.55, and 0.55, respectively. The AUCs of the 1-, 3-, and 5-year GSE85217 dataset2 cohorts were 0.60, 0.53, and 0.54, respectively. The AUC values of the 3-, and 5- year GSE30074 cohorts were 0.60 and 0.53, respectively. The AUC values of the 1-, 3-, and 5-year SYSUCC cohorts were 0.63, 0.50, and 0.53 (Figure 2F).

Delta-catenin is enriched in WNT subgroup and is associated with dissemination

In our study, delta-catenin expression was significantly higher in WNT group of medulloblastoma than in non-WNT groups (SHH, Group 3, and Group 4) (Figure 3A) based on two datasets GSE85217 and GSE21140. Metastasis frequency was, in fact, lowest in WNT subgroup in medulloblastoma patients (Fults et al., 2019). Based on these two findings, we speculate that delta-catenin participates in regulating medulloblastoma invasion and dissemination. Interestingly, data of the cohorts from Cavalli (Cavalli et al., 2017), confirmed that high delta-catenin expression was associated with low tumor dissemination (Figure 3B). Based on GSEA analysis, the EMT molecules were highly enriched in the low-delta-catenin expression group (Figure 3C). In addition, we also enriched some signaling pathways that have been widely reported to affect the metastasis of medulloblastoma, such as PI3K/AKT, TGF-beta, MYC, Notch, KRAS signaling pathway (Li et al., 2021). The metastasis-related signaling pathways were also enriched in the low delta-catenin expression group (Figure 3D). These findings supports our hypothesis that delta-catenin suppresses both invasion and migration in medulloblastoma. In addition, the BioGRID database showed interactions between delta-catenin and a large number of proteins. The most significant correlations were EGFR, CLK1, ATN1, LNX1, TRIM47, NUDT12, TTR, ESR1, KAT28, MCM2, PTGER4, THUMPD3, PDZD2, LRRC7, ZMYND19, NR3C1 and ZBTB33 (Supplementary Figure S2).

Delta-catenin attenuates medulloblastoma cell invasion and migration

To evaluate the role of delta-catenin expression in medulloblastoma dissemination, we established stable knockdown and OE cells of Daoy and ONS-76. RT-qPCR and western blotting were used to confirm infection efficiency (Figures 4A,B). Transwell assay was applied to test the effect of delta-catenin on medulloblastoma invasion and migration. In both medulloblastoma cell lines, delta-catenin-knockdown significantly promoted medulloblastoma invasion and migration, whereas delta-catenin overexpression attenuated them (Figures 4C,D). Wound healing assay also showed that delta-catenin-knockdown accelerated medulloblastoma migration (Figure 4E). Actually, we also found that delta-catenin could inhibit the proliferation of medulloblastoma cells (Supplementary Figure S3). These are consistent with our bioinformatics findings, and supports that delta-catenin affects medulloblastoma invasion and dissemination *in vitro*.

Delta-catenin attenuates medulloblastoma cell invasion and migration by targeting epithelial-mesenchymal transition

Next, we examined how delta-catenin inhibits medulloblastoma invasion and migration, and hence metastasis, by observing how it affects EMT pathway, a major pathway of metastasis. We detected the expression of EMT markers (E-cadherin and vimentin) in delta-catenin-knockdown medulloblastoma cells and control cells. Delta-catenin knockdown significantly reduced the expression of E-cadherin and increased that of vimentin (Figure 5A). Moreover, *Via* immunofluorescence, we found that knockdown of delta-catenin broke down the continuous arrangement of E-cadherin in the adjacent areas of cell to discrete and disorganized form (Figure 5B). These data indicated that delta-catenin regulated both the quantity and structure of E-cadherin, and therefore attenuate medulloblastoma cell invasion and migration *via* EMT pathway.

Discussion

We aimed to elucidate the molecular mechanisms of medulloblastoma invasion, to improve the therapeutic response and prolong survival in medulloblastoma patients. Our bioinformatics analysis revealed that delta-catenin expression was positively associated with survival of medulloblastoma patients. It was enriched in WNT subgroup, which had the lowest metastasis rate among the four subgroups. Delta-catenin inhibited medulloblastoma cell invasion *in vitro*, and EMT pathway may be the underlying mechanism.

These findings are consistent with various studies showing that delta-catenin inhibits tumors. For example, CTNND2 loss-of-function mutation was found to be common in glioblastoma. Based on analysis of TCGA data, reduced CTNND2 expression was associated with poor prognosis of glioblastoma, especially for

10.3389/fgene.2022.867872

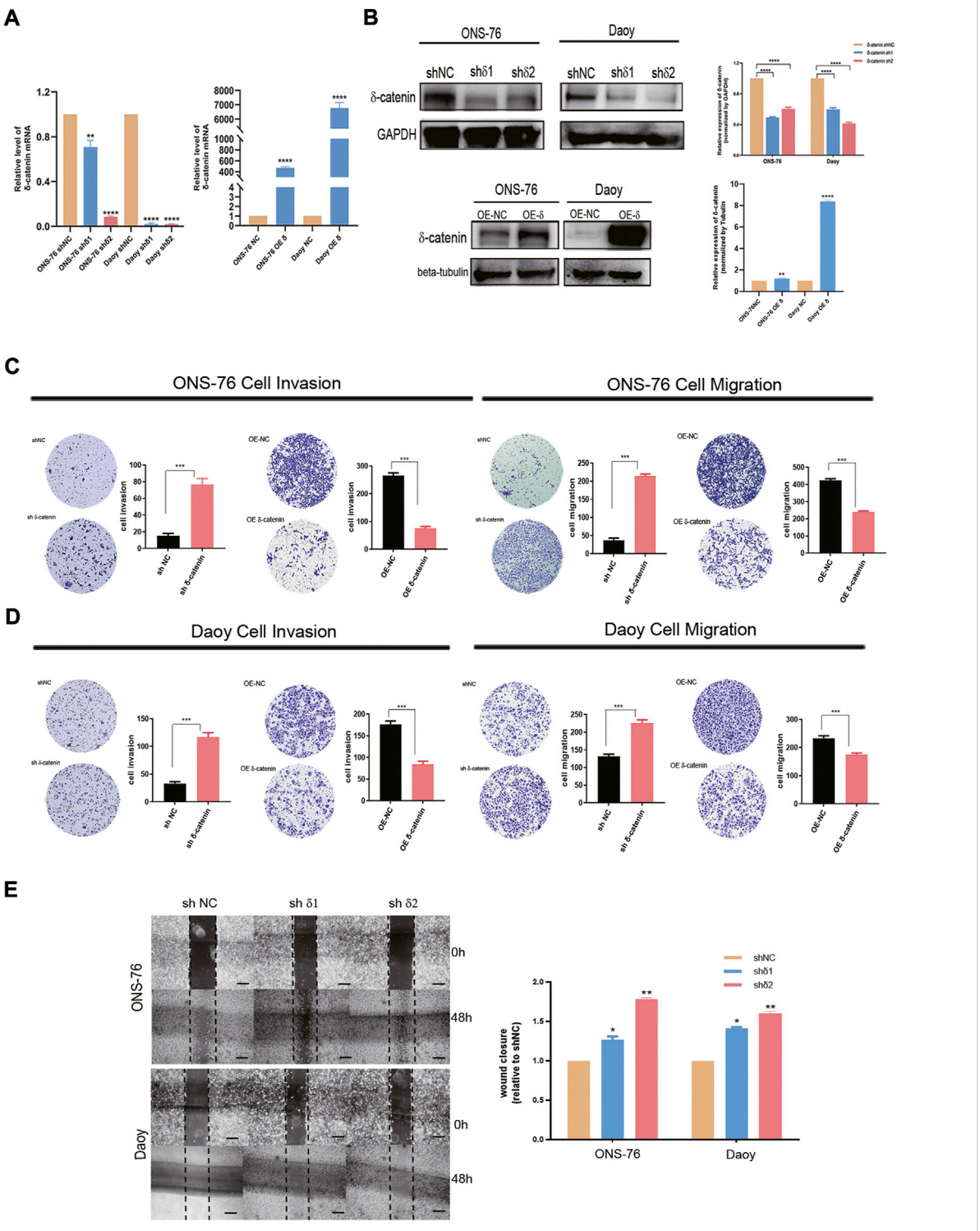


FIGURE 4
Effects of delta-catenin knockdown and overexpression in Daoy and ONS-76 cells migration and invasion. **(A)** Real-time quantitative PCR analysis of delta-catenin mRNA following lentiviral transfection. β-actin was used as a control. **(B)** Western blot analysis of delta-catenin expression following lentiviral transfection. GAPDH and Tubulin were used as controls. **(C)** Representative images (×10 magnification) of invasion and migration by delta-catenin-knockdown and overexpressing ONS-76 cells. **(D)** Representative images (×10 magnification) showing invasion and migration by delta-catenin-knockdown and overexpressing Daoy cells. **(E)** Cellular migration in both the sh#1 and sh#2 groups of ONS-76 and Daoy cells were promoted, as determined *via* a wound-healing assay. Scale bars, 100 μm.

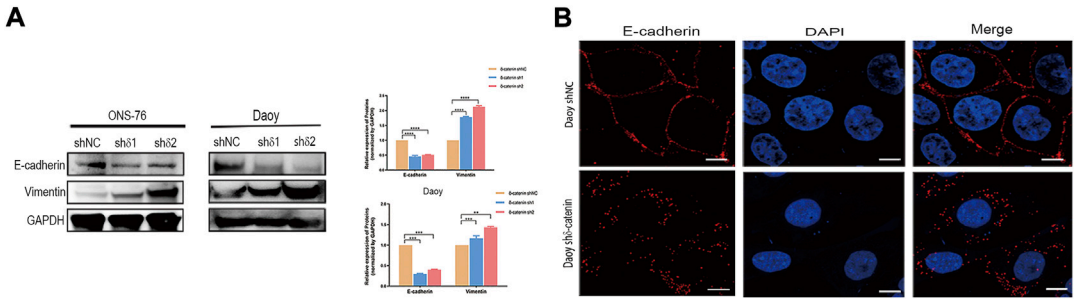


FIGURE 5
Delta-catenin attenuates medulloblastoma cell migration and invasion by targeting EMT pathway. **(A)** Western blot analysis of EMT pathway markers (E-cadherin and vimentin) following lentiviral transfection of ONS-76 and Daoy cells. GAPDH was used as a control. **(B)** Representative images (×100 magnification) of E-cadherin in Daoy sh delta-catenin and control cell lines, *via* immunofluorescence staining. Scale bars, 5 μm.

the mesenchymal type (Frattini et al., 2013). Delta-catenin inhibited glioma cell proliferation and self-renewal, followed by phenotypic transformation from the aggressive mesenchymal cell type to the neuronal cell type (Frattini et al., 2013). Among 25 identified genes that potentially suppress EMT, CTNND2 was the most likely to suppress EMT (Westbrook et al., 2005).

However, delta-catenin was also reported to serve as an oncoprotein (Huang et al., 2018; Shen et al., 2021; Zeng et al., 2009; Zhang et al., 2015). It could not only promote macrophage migration (Wu et al., 2020), but also promote cancer cell invasion, dissemination, and metastasis in various tumors and causes polygonal cells to develop irregular or elongated fibroblastic morphology (Lu et al., 1999; Dai et al., 2011; Huang et al., 2018). In brain tumors, it is associated with the malignant progression of astrocytoma and promotes astrocytoma cell invasion (Wang et al., 2011). Furthermore, delta-catenin might promote bevacizumab-induced glioma invasion (Shimizu et al., 2019).

Prior to now, there has been a poor understanding of the context-dependent roles of delta-catenin in cancer. Two explanations for this have been proposed. First, it may be primarily due to variation in CTNND2. In prostate cancer, wild-type and mutant delta-catenin both exhibited pro-oncogenic and tumor-suppressive functions; that study found a protein truncation, caused by a nonsense mutation of delta-catenin, mainly occurring in the cytoplasm, promoted carcinoma progression *via* various pathways (Li et al., 2020). CTNND2 has a vast number of known mutations; the Sanger COSMIC database has recorded 541 unique mutations of delta-catenin, in 38 different tissues, and some of these mutations alter delta-catenin structure and function in tumors (Lu et al., 2016). This may therefore explain the contrasting reported effects of delta-catenin.

The second possible reason might be related to epigenetic modification. Post-translational phosphorylation of delta-catenin, which alters its function substantially, has been increasingly described. Phosphorylation of delta-catenin plays an important role in determining delta-catenin's role in neuronal development and oncogenesis (Oh et al., 2009; Poore et al., 2010; Chen et al., 2021). And a classic dual-function model has been identified, involving a phospho-switch in delta-catenin that can trigger two opposite effects on dendrite development (Baumert et al., 2020). Thus, its phosphorylation status may cause these contrasting effects.

Medulloblastoma, which is differ from the other pediatric brain tumor such as ependymoma and brainstem glioma (Duc, 2020; Duc et al., 2020), is the most common malignant brain tumor in children. The challenge in medulloblastoma therapy is to address tumor cell invasion and dissemination, which cannot completely be prevented by treatment currently. Radiotherapy of the total central nervous system and chemotherapy are typically used to avoid tumor cell infiltration and metastasis, but they can cause intellectual and neurological disabilities in survivors (Mabbott et al., 2008). The fact that metastasis is rare in WNT subgroup suggests that unexamined molecular mechanisms in this group may help to prevent medulloblastoma invasion and dissemination. Delta-catenin is enriched in WNT subgroup and associated with lower tumor dissemination in our study. Delta-catenin has been reported promoting β-catenin nuclear localization and activating WNT pathway, and therefore accelerate tumor progression (Nopparat et al., 2015; Huang et al., 2018; Ju et al., 2020). However, WNT activation was found an unexpected tumor suppressor in medulloblastoma in several studies (Pöschl et al., 2014; Zinke et al., 2015; Manoranjan et al., 2020). The role of delta-catenin in WNT activation might be the reason that it is related with better prognosis in medulloblastoma patients according to this unexpected finding.

This study has some certain limitations. The cell line in our study are all belong to SHH subgroup (Ivanov et al., 2016). Given that delta-catenin is enriched in WNT Subgroup, the exploring of delta-catenin function in WNT subgroup cell model would reflect its role in cell invasion precisely.

However, there is no WNT cell line available except for primary cells derived from medulloblastoma tissue (Ivanov et al., 2016; Manoranjan et al., 2020), which we would establish and use in further research. Exploring the underlying molecular mechanisms which delta-catenin works in WNT subgroup would also be our future direction. In addition, the validation of delta-catenin's role in invasion is also lacking in other subgroup cell lines (Group3/4) in our study, which need further validation in the future.

Conclusion

In conclusion, we found that high expression of delta-catenin, which is enriched in WNT subgroup, is a favorable prognostic factor in medulloblastoma. The EMT pathway may participate in this reduced invasion and metastasis.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, and the raw data of this paper have been uploaded onto the Research Data Deposit (RDD) with an RDD number of RDDB2022275274.

Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Sun Yat-sen University Cancer Center (protocol code GZR 2017-007). According to our institutional standards, all the patients provided a written informed consent, either by themselves or by a legally authorized representative, to collect and analyze blood samples, pathology samples, and publish their results.

Author contributions

YH, RX, and MW performed the experiments. SZ conducted the bioinformatic analysis. RX and FC supervised and collected clinical specimens. YH, SZ, and JiW wrote and revised the article. JaW, ZZ, BF, and ZC supervised the study. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China to JiW (81772677 and 82073109).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.867872/full#supplementary-material

References

Arikkath, J., Israely, I., Tao, Y., Mei, L., Liu, X., and Reichardt, L. (2008). Erbin controls dendritic morphogenesis by regulating localization of delta-catenin. *J. Neurosci.* 28, 7047–7056. doi:10.1523/jneurosci.0451-08.2008

Baumert, R., Ji, H., Paulucci-Holthauzen, A., Wolfe, A., Sagum, C., Hodgson, L., et al. (2020). Novel phospho-switch function of delta-catenin in dendrite development. *J. Cell Biol.* 219, e201909166. doi:10.1083/jcb.201909166

Cavalli, F., Remke, M., Rampasek, L., Peacock, J., Shih, D., Luu, B., et al. (2017). Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell* 31, 737–754. e736. doi:10.1016/j.ccell.2017.05.005

Chen, G., An, N., Zhu, Y., Zhou, R., Noh, M., Kim, H., et al. (2021). bFGF-mediated phosphorylation of δ-catenin increases its protein stability and the ability to induce the nuclear redistribution of β-catenin. *Am. J. Cancer Res.* 11, 3877–3892.

Chen, H., Lin, F., Zhang, J., Lv, X., Zhou, J., Li, Z., et al. (2021). Deep learning radiomics to predict PTEN mutation status from magnetic resonance imaging in patients with Glioma. *Front Oncol.* 11, 734433. doi:10.3389/fonc.2021.734433

Coltin, H. R. V. (2021). Molecular and clinical correlates of medulloblastoma subgroups: A narrative review. *Glioma* 4, 92–99. doi:10.4103/glioma.glioma_18_21

Dai, S., Wang, Y., Zhang, J., Zhang, D., Zhang, P., Jiang, G., et al. (2011). Upregulation of δ-catenin is associated with poor prognosis and enhances transcriptional activity through Kaiso in non-small-cell lung cancer. *Cancer Sci.* 102, 95–103. doi:10.1111/j.1349-7006.2010.01766.x

Duc, N., Huy, H., Nadarajan, C., and Keserci, B. (2020). The role of predictive model based on quantitative basic magnetic resonance imaging in differentiating medulloblastoma from ependymoma. *Anticancer Res.* 40, 2975–2980. doi:10.21873/anticanres.14277

Duc, N. (2020). The role of diffusion tensor imaging metrics in the discrimination between cerebellar medulloblastoma and brainstem glioma. *Pediatr. Blood Cancer* 67, e28468. doi:10.1002/pbc.28468

Frattini, V., Trifonov, V., Chan, J., Castano, A., Lia, M., Abate, F., et al. (2013). The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* 45, 1141–1149. doi:10.1038/ng.2734

Fults, D., Taylor, M., and Garzia, L. (2019). Leptomeningeal dissemination: A sinister pattern of medulloblastoma growth. *J. Neurosurg. Pediatr.* 23, 613–621. doi:10.3171/2018.11.Peds18506

Garzia, L., Kijima, N., Morrissy, A., De Antonellis, P., Guerreiro-Stucklin, A., Holgado, B., et al. (2018). A hematogenous route for medulloblastoma leptomeningeal metastases. *Cell* 172, 1549–1062. e1014. doi:10.1016/j.cell.2018.05.033

Gibson, P., Tong, Y., Robinson, G., Thompson, M., Currle, D., Eden, C., et al. (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature* 468, 1095–1099. doi:10.1038/nature09587

Huang, F., Chen, J., Wang, Z., Lan, R., Fu, L., and Zhang, L. (2018). δ-Catenin promotes tumorigenesis and metastasis of lung adenocarcinoma. *Oncol. Rep.* 39, 809–817. doi:10.3892/or.2017.6140

Ivanov, D., Coyle, B., Walker, D., and Grabowska, A. (2016). *In vitro* models of medulloblastoma: Choosing the right tool for the job. *J. Biotechnol.* 236, 10–25. doi:10.1016/j.jbiotec.2016.07.028

Ju, L., Shan, L., Yin, B., and Song, Y. (2020). δ-Catenin regulates proliferation and apoptosis in renal cell carcinoma via promoting β-catenin nuclear localization and activating its downstream target genes. *Cancer Med.* 9, 2201–2212. doi:10.1002/cam4.2857

Juraschka, K., and Taylor, M. (2019). Medulloblastoma in the age of molecular subgroups: A review. *J. Neurosurg. Pediatr.* 24, 353–363. doi:10.3171/2019.5.Peds18381

Li, M., Deng, Y., and Zhang, W. (2021). Molecular determinants of medulloblastoma metastasis and leptomeningeal dissemination. *Mol. Cancer Res.* 19, 743–752. doi:10.1158/1541-7786.Mcr-20-1026

Li, M., Nopparat, J., Aguilar, B., Chen, Y., Zhang, J., Du, J., et al. (2020). Intratumor δ-catenin heterogeneity driven by genomic rearrangement dictates growth factor dependent prostate cancer progression. *Oncogene* 39, 4358–4374. doi:10.1038/s41388-020-1281-9

Li, W., Chen, Q., Huang, T., Wu, P., Shen, L., Huang, Z., et al. (2020). Identification and validation of a prognostic lncRNA signature for Hepatocellular Carcinoma. *Front Oncol.* 10, 780. doi:10.3389/fonc.2020.00780

Louis, D., Perry, A., Wesseling, P., Brat, D., Cree, I., Figarella-Branger, D., et al. (2021). The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro. Oncol.* 23, 1231–1251. doi:10.1093/neuonc/noab106

Lu, Q., Aguilar, B., Li, M., Jiang, Y., and Chen, Y. (2016). Genetic alterations of δ-catenin/NPRAP/neurojungin (CTNND2): Functional implications in complex human diseases. *Hum. Genet.* 135, 1107–1116. doi:10.1007/s00439-016-1705-3

Lu, Q., Lanford, G., Hong, H., and Chen, Y. (2014). δ-Catenin as a potential cancer biomarker. *Pathol. Int.* 64, 243–246. doi:10.1111/pin.12156

Lu, Q., Paredes, M., Medina, M., Zhou, J., Cavallo, R., Peifer, M., et al. (1999). delta-catenin, an adhesive junction-associated protein which promotes cell scattering. *J. Cell Biol.* 144, 519–532. doi:10.1083/jcb.144.3.519

Mabbott, D., Penkman, L., Witol, A., Strother, D., and Bouffet, E. (2008). Core neurocognitive functions in children treated for posterior fossa tumors. *Neuropsychology* 22, 159–168. doi:10.1037/0894-4105.22.2.159

Manoranjan, B., Venugopal, C., Bakhshinyan, D., Adile, A., Richards, L., Kameda-Smith, M., et al. (2020). Wnt activation as a therapeutic strategy in medulloblastoma. *Nat. Commun.* 11, 4323. doi:10.1038/s41467-020-17953-4

Martinez, M., Ochiishi, T., Majewski, M., and Kosik, K. (2003). Dual regulation of neuronal morphogenesis by a delta-catenin-cortactin complex and Rho. *J. Cell Biol.* 162, 99–111. doi:10.1083/jcb.200211025

Medina, M., Marinescu, R., Overhauser, J., and Kosik, K. (2000). Hemizygosity of delta-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics* 63, 157–164. doi:10.1006/geno.1999.6090

Nopparat, J., Zhang, J., Lu, J., Chen, Y., Zheng, D., Neufer, P., et al. (2015). δ-Catenin, a Wnt/β-catenin modulator, reveals inducible mutagenesis promoting cancer cell survival adaptation and metabolic reprogramming. *Oncogene* 34, 1542–1552. doi:10.1038/onc.2014.89

Northcott, P., Korshunov, A., Pfister, S., and Taylor, M. (2012). The clinical implications of medulloblastoma subgroups. *Nat. Rev. Neurol.* 8, 340–351. doi:10.1038/nrneurol.2012.78

Northcott, P., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C., Mack, S., et al. (2011). Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* 29, 1408–1414. doi:10.1200/jco.2009.27.4324

Oh, M., Kim, H., Yang, I., Park, J., Cong, W., Baek, M., et al. (2009). GSK-3 phosphorylates delta-catenin and negatively regulates its stability via ubiquitination/proteosome-mediated proteolysis. *J. Biol. Chem.* 284, 28579–28589. doi:10.1074/jbc.M109.002659

Poore, C., Sundaram, J., Pareek, T., Fu, A., Amin, N., Mohamed, N., et al. (2010). Cdk5-mediated phosphorylation of delta-catenin regulates its localization and GluR2-mediated synaptic activity. *J. Neurosci.* 30, 8457–8467. doi:10.1523/jneurosci.6062-09.2010

Pöschl, J., Bartels, M., Ohli, J., Bianchi, E., Kuteykin-Teplyakov, K., Grammel, D., et al. (2014). Wnt/β-catenin signaling inhibits the Shh pathway and impairs tumor growth in Shh-dependent medulloblastoma. *Acta Neuropathol.* 127, 605–607. doi:10.1007/s00401-014-1258-2

Reynolds, A., and Roczniak-Ferguson, A. (2004). Emerging roles for p120-catenin in cell adhesion and cancer. *Oncogene* 23, 7947–7956. doi:10.1038/sj.onc.1208161

Shen, Y., Lee, H., Zhou, R., Kim, H., Chen, G., Cho, Y., et al. (2021). δ-Catenin participates in EGF/AKT/p21^{Waf} signaling and induces prostate cancer cell proliferation and invasion. *Int. J. Mol. Sci.* 22, 5306. doi:10.3390/ijms22105306

Shimizu, T., Ishida, J., Kurozumi, K., Ichikawa, T., Otani, Y., Oka, T., et al. (2019). δ-Catenin promotes bevacizumab-induced glioma invasion. *Mol. Cancer Ther.* 18, 812–822. doi:10.1158/1535-7163.Mct-18-0138

Sun, R. C., Dukhande, V. V., Zhou, Z., Young, L. E. A., Emanuelle, S., Brainson, C. F., et al. (2019). Nuclear glycogenolysis modulates histone acetylation in human non-small cell lung cancers. *Cell Metab.* 30, 903–916. e907. doi:10.1016/j.cmet.2019.08.014

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 47, W556–W560. doi:10.1093/nar/gkz430

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). Gepia: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi:10.1093/nar/gkx247

Turner, T., Sharma, K., Oh, E., Liu, Y., Collins, R., Sosa, M., et al. (2015). Loss of δ-catenin function in severe autism. *Nature* 520, 51–56. doi:10.1038/nature14186

Vrijenhoek, T., Buizer-Voskamp, J., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., et al. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am. J. Hum. Genet.* 83, 504–510. doi:10.1016/j.ajhg.2008.09.011

Wang, J., Xi, S., Zhao, Q., Xia, Y., Yang, Q., Cai, H., et al. (2022). Driver mutations in ADGRL3 are involved in the evolution of ependymoma. *Lab. Invest.* 102, 702–710. doi:10.1038/s41374-021-00721-3

Wang, M., Dong, Q., Zhang, D., and Wang, Y. (2011). Expression of delta-catenin is associated with progression of human astrocytoma. *BMC cancer* 11, 514. doi:10.1186/1471-2407-11-514

Westbrook, T., Martin, E., Schlabach, M., Leng, Y., Liang, A., Feng, B., et al. (2005). A genetic screen for candidate tumor suppressors identifies REST. *Cell* 121, 837–848. doi:10.1016/j.cell.2005.03.033

Wu, B., Liu, J., Bian, E., Hu, W., Huang, C., Meng, X., et al. (2020). Blockage of Kv1.3 regulates macrophage migration in acute liver injury by targeting δ-catenin through RhoA signaling. *Int. J. Biol. Sci.* 16, 671–681. doi:10.7150/ijbs.38950

Zeng, Y., Abdallah, A., Lu, J., Wang, T., Chen, Y., Terrian, D., et al. (2009). delta-Catenin promotes prostate cancer cell growth and progression by altering cell cycle and survival gene profiles. *Mol. Cancer* 8, 19. doi:10.1186/1476-4598-8-19

Zhang, D., Zhang, J., and Wang, E. (2015). δ-catenin promotes the malignant phenotype in breast cancer. *Tumour Biol.* 36, 569–575. doi:10.1007/s13277-014-2680-8

Zhang, H., Dai, S., Zhang, D., Liu, D., Zhang, F., Zheng, T., et al. (2014). Delta-catenin promotes the proliferation and invasion of colorectal cancer cells by binding to E-cadherin in a competitive manner with p120 catenin. *Target. Oncol.* 9, 53–61. doi:10.1007/s11523-013-0269-6

Zhou, J., Liyanage, U., Medina, M., Ho, C., Simmons, A., Lovett, M., et al. (1997). Presenilin 1 interaction in the brain with a novel member of the Armadillo family. *Neuroreport* 8, 2085–2090. doi:10.1097/00001756-199705260-00054

Zinke, J., Schneider, F., Harter, P., Thom, S., Ziegler, N., Toftgård, R., et al. (2015). β-Catenin-Gli1 interaction regulates proliferation and tumor growth in medulloblastoma. *Mol. Cancer* 14, 17. doi:10.1186/s12943-015-0294-4

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 21 October 2022
DOI 10.3389/fgene.2022.890247

Check for updates

OPEN ACCESS

EDITED BY
Apeng Chen,
Lanzhou Veterinary Research Institute (CAAS), China

REVIEWED BY
Antonio Russo,
University of Palermo, Italy
Giuseppe Luigi Banna,
United Lincolnshire Hospitals NHS Trust, United Kingdom

*CORRESPONDENCE
Weizhang Shen,
shenweizhang2015@126.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 05 March 2022
ACCEPTED 06 October 2022
PUBLISHED 21 October 2022

CITATION
Gao L and Shen W (2022), Light at the end of the tunnel: Clinical features and therapeutic prospects of KRAS mutant subtypes in non-small-cell lung cancer. *Front. Genet.* 13:890247. doi: 10.3389/fgene.2022.890247

COPYRIGHT
© 2022 Gao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Light at the end of the tunnel: Clinical features and therapeutic prospects of KRAS mutant subtypes in non-small-cell lung cancer

Liyuan Gao and Weizhang Shen*

Department of Oncology and Hematology, The Second Hospital of Jilin University, Changchun, China

Lung cancer is one of the most common causes of cancer-related deaths, and non-small-cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases. Kirsten rat sarcoma virus (KRAS), one of the three subtypes of the RAS family, is the most common oncogene involved in human cancers and encodes the key signaling proteins in tumors. Oncogenic KRAS mutations are considered the initiating factors in 30% of NSCLC cases, accounting for the largest proportion of NSCLC cases associated with driver mutations. Because effective inhibition of the related functions of KRAS with traditional small-molecule inhibitors is difficult, the KRAS protein is called an "undruggable target." However, in recent years, the discovery of a common mutation in the KRAS gene, glycine 12 mutated to cysteine (G12C), has led to the design and synthesis of covalent inhibitors that offer novel strategies for effective targeting of KRAS. In this review, we have summarized the structure, function, and signal transduction pathways of KRAS and discussed the available treatment strategies and potential treatment prospects of KRAS mutation subtypes (especially G12C, G12V, and G12D) in NSCLC, thus providing a reference for selecting KRAS mutation subtypes for the treatment of NSCLC.

KEYWORDS

non-small-cell lung cancer, KRAS mutation, KRAS G12C, KRAS G12V, KRAS G12D

1 General structure of the KRAS protein and the KRAS signaling pathway

The activated RAS gene was first discovered in 1982, which is also the first identified proto-oncogene. The RAS gene family includes HRAS, KRAS, and NRAS, which are located on chromosomes 11, 12, and 1, respectively (Cox and Der 2010). KRAS is the most frequently mutated subtype. The RAS gene has three mutation sites, namely, G12, G13, and Q61. KRAS differs from NRAS and HRAS because it is the only RAS subtype that is predominantly mutated at 12 sites (Cox and Der 2010). KRAS encodes and manufactures the KRAS protein, which is a membrane GTPase. Structurally, the KRAS protein mainly includes a G-domain and C-terminal structural elements [also known as a hypervariable

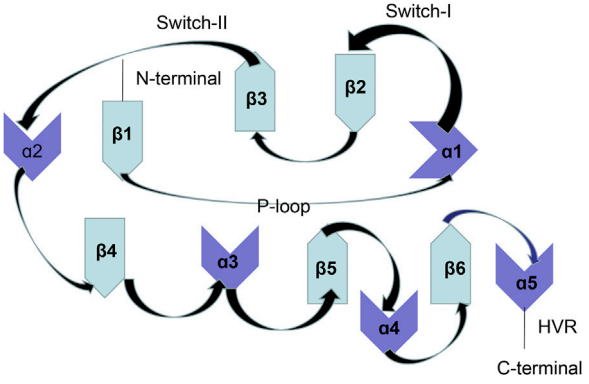


FIGURE 1
KRAS protein structure.

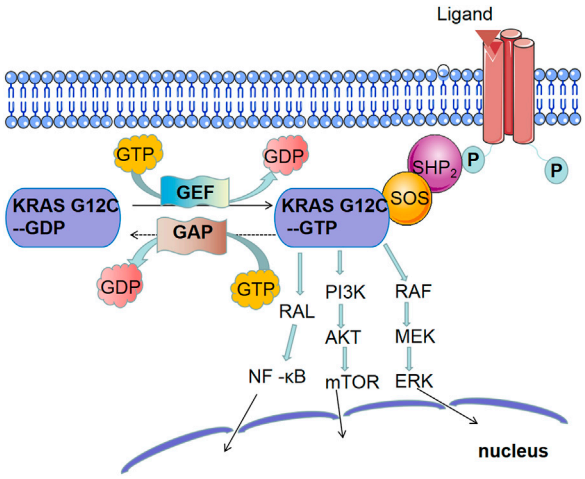


FIGURE 2
Main signaling pathway of KRAS mutations: ① MAPK signaling pathway: continuous activation of KRAS leads to the activation of RAF, the first kinase in the mitogen-activated protein kinase (MAPK) pathway, which phosphorylates MEK and then activates extracellular signal-regulated kinase (ERK). ERK both activates the cytoplasmic matrix and is transported to the nucleus to stimulate the expression of multiple genes involved in cell proliferation, survival, differentiation, and cell cycle regulation. Extensive studies have shown that MAPK signaling pathway plays an important role in RAS-mediated tumorigenesis. ② PI3K signaling pathway: phosphatidylinositol-4, 5-diphosphate trikinase (PI3K) also plays a key role in RAS-mediated tumorigenesis. When KRAS is activated, PI3K converts phosphatidylinositol 4, 5-phosphate (PIP2) to phosphatidylinositol 3, 4, 5-triphosphate (PIP3) through phosphorylation, which activates phosphoinositol-dependent kinase 1 (PDK1) and phosphorylates AKT, a serine/threonine-specific protein kinase. AKT activates mammalian target of rapamycin (mTOR). ③ NF-κB signaling pathway: At rest, NF-κB forms a complex with its inhibitory factors, which anchors NF-κB to the cytosol. When KRAS is activated, the RAL-NF-κB pathway is activated.

region (HVR)]. The G-domain includes a switch region (switch I and II) and a P-loop and forms the basis for biological functioning of the GTPase protein. HVR mainly anchors RAS to membranes (Hancock et al., 1990; Welman et al., 2000; Abraham et al., 2010). The switch region forms the binding interface for effector proteins and the regulatory factors of RAS such as GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). The structure of the KRAS protein is shown in Figure 1.

The KRAS protein can transmit extracellular signals to the nucleus and participate in the regulation of signaling pathways related to the growth, proliferation, differentiation, and apoptosis of tumor cells. Its activity is mainly mediated by guanosine triphosphate (GTP) or guanosine diphosphate (GDP) and is activated when bound to GTP and inactivated when bound to GDP. Because the GTPase activity of KRAS is not strong, GAPs are required to hydrolyze and inactivate KRAS-bound GTP (Vetter and Wittinghofer 2001). GAPs have an SH2 domain and can directly bind to the activated receptor. GDP release requires the participation of GEFs. Son of Sevenless (SOS1) is a GEF that promotes KRAS activation (Jeng et al., 2012). SHP2 is a protein tyrosine phosphatase that interacts with RTK and SOS1 to mediate KRAS activation (Mainardi et al., 2018). KRAS mutations promote the interaction between GAPs and KRAS, which promotes the gradual accumulation of active KRAS by inhibiting the hydrolysis of GTP, thereby breaking the regulatory cycle of active and inactive KRAS (Scheffzek et al., 1997). In addition, KRAS mutations can persistently activate the common KRAS-dependent downstream pathways, such as RAF/MEK/ERK, PI3K/AKT/mTOR, and RAL/NF-κB signaling pathways. The continuous activation of downstream molecules leads to tumorigenesis. The signal transduction pathway of the KRAS protein is shown in Figure 2.

Therapeutic targeting of KRAS mutations is difficult because of the structure and regulatory pathways of the KRAS protein (Canon et al., 2019; Fell et al., 2020; Goebel et al., 2020): ① The protein surface is excessively smooth, and only one GTP binding site is available. ② The lack of favorable small-molecule-binding pockets makes it difficult for drug molecules to compete with GTP. The binding of GTP to KRAS is very strong, reaching the level of Pmol, and the concentration of GTP in cells is very high, thus making it difficult for small-molecule inhibitors directly targeting GTP-binding pockets to compete with it. ③ KRAS is mainly regulated in the cell membrane. Owing to the aforementioned reasons, it is difficult for drug developers to find binding pockets for the small-molecule drug candidates on the surface of KRAS. Therefore, although KRAS is the earliest identified oncogene, the development of small-molecule inhibitors of mutated KRAS subtypes is difficult. For a long time, the KRAS protein was described as

an "unavailable target." However, the discovery of the KRAS G12C (glycine 12 mutated to cysteine)-targeted inhibitors AMG510 and MRTX849 offers novel strategies for effective targeting of KRAS.

2 KRAS mutation in non-small-cell lung cancer

KRAS-mutated tumors are the most common subtype of cancer with potentially targeted molecules in non-small-cell lung cancer (NSCLC) (Jordan et al., 2017). KRAS mutations are associated with smoking history, adenocarcinoma, and female sex among patients with lung cancer (Shepherd et al., 2013). Moreover, the incidence of KRAS mutations is different between the East and the West. KRAS mutations are found in approximately 20%–25% of lung cancer cases in the Western countries (Dogan et al., 2012; Shepherd et al., 2013; El Osta et al., 2019) and 10%–15% of lung cancer cases in Asian countries (Dearden et al., 2013; Yoshizawa et al., 2013).

KRAS mutations have been frequently discovered as driver mutations that are mutually exclusive with other driver genes, such as EGFR/ALK/BRAF/ROS1 (Renaud et al., 2016; Adachi et al., 2020; West et al., 2022). KRAS is co-mutated with other genes in 53.5% of cases. TP53, STK11, MET (amplification), ERBB2 (amplification), KEAP1, ATM, SMARCA4, ROS1, and PIK3CG are among the co-mutated genes. KRAS/TP53 co-mutation is the most common, accounting for approximately 39.4% of all cases (Scheffler et al., 2019). Additionally, studies have shown that patients with lung adenocarcinoma with TP53 and KRAS co-mutations have a higher programmed cell death-ligand 1 (PD-L1) expression than patients with either TP53 or KRAS mutation, suggesting that this population may benefit from anti-programmed cell death protein 1 (PD-1) immunotherapy (Dong et al., 2017).

A majority of KRAS mutations affect codons 12, 13, or 61. Mutations at codon 12 are the most common, occurring in > 90% of KRAS mutation-driven cancer cases. G12V (glycine 12 mutated to valine) and G12C are the two most common allelic mutations, accounting for > 50% of all KRAS mutation-driven cancer cases (Herdeis et al., 2021).

Among the subtypes of KRAS mutations, KRAS G12C and G12V mutations are epidemiologically associated with smoking history, whereas G12D mutations are more common among individuals who do not smoke. KRAS mutations can be divided into two types: transmutation (purine nucleotide substitution for a pyrimidine or *vice versa*) or transfer mutation (purine to purine or pyrimidine to pyrimidine substitution). Transmutation is more common among individuals who smoke, whereas metastatic mutations are more common among individuals who do not smoke (Tomasini et al., 2016).

3 KRAS G12C mutation

The KRAS G12C mutant subtype has a cysteine residue (glycine position 12 is mutated to cysteine) and is the most common KRAS mutation in NSCLC (Scheffler et al., 2019).

3.1 KRAS G12C-targeted drugs

Owing to the special characteristics of KRAS G12C mutations, specific irreversible inhibitors of KRAS G12C have been developed. Recently, various inhibitors that target KRAS G12C have been gradually introduced in clinical practice. At present, adagrasib and sotorasib are FDA-approved drugs for the treatment of patients with lung cancer with KRAS G12C mutation. To achieve therapeutic efficacy, the main mechanism of action is the binding of these drugs with KRAS G12C, which locks the oncoprotein in an inactive state and prevents it from sending signals that drive uncontrolled cell growth. However, the targeted drug does not affect the wild-type KRAS protein (Lito et al., 2016; Patricelli et al., 2016).

3.1.1 Sotorasib (AMG510)

A phase 1/2 clinical trial (NCT03600883, Table 2) showed promising results of sotorasib therapy among patients with advanced KRAS-G12C-positive solid tumors. According to the initial data shared in September 2020 (Hong et al., 2020), patients with NSCLC in the 960-mg cohort had an objective response rate (ORR) of 32.2% and a disease control rate (DCR) of 88.1%. Dose-limiting toxicity was not observed in safety studies, and treatment-related adverse events did not result in death in any case. The most common adverse reactions were diarrhea (29.5%), fatigue (23.3%), and nausea (20.9%). These results were recently validated by a phase II trial (Skoulidis et al., 2021), which included 122 patients with NSCLC. The ORR was 37.1%, with a median duration of response of 11 months and a DCR of 80.6%. The median progression-free survival (PFS) was 6.8 months, and the median overall survival (OS) was 12.5 months. In addition, a phase III clinical trial (NCT04303780, Table 2) is currently underway for comparing the efficacy of sotorasib and docetaxel in patients with NSCLC with KRAS G12 mutation before treatment. At present, sotorasib is undergoing phase II trials as a first-line treatment agent (NCT04933695, Table 2).

3.1.2 Adagrasib (MRTX849)

Mirati Therapeutics developed MRTX849 (adagrasib), an orally bio-available covalent KRAS G12C inhibitor (KRAS G12C Inhibitor, Mirati Therapeutics, 2021). A phase 1/2 study (NCT03785249, Table 2) demonstrated the potential of MRTX849 as an effective and potent KRAS G12C inhibitor. The preclinical studies on adagrasib (Hallin et al., 2020) have revealed the following: ① Strong action: in multiple cell models of KRAS G12C mutation, adagrasib showed good

efficacy at very low concentrations (nanomolar level). ② Long half-life: adagrasib is the only inhibitor of KRAS G12C mutant with a half-life of approximately 24 h. ③ High selectivity: the selectivity of KRAS G12C mutant was > 1000 times that of wild-type KRAS and other proteins. In a recent phase II single-arm trial, patients with advanced NSCLC who had previously received platinum-based dual-agent therapy and immune checkpoint inhibitor (ICI) therapy were enrolled. ORR was the primary endpoint as determined by an independent review board, whereas response duration, PFS, OS, and safety were secondary endpoints. As of 15 October 2021, 116 patients were enrolled. The median follow-up period was 12.9 months at the time of data analysis, and 98.3% of the patients had previously received chemotherapy and ICI therapy. There were 112 patients with evaluable lesions, with an ORR of 42.9%, a response duration of 8.5 months (95% CI, 6.2–13.8 months), and a median PFS of 6.5 months. As of 15 January 2022, the median follow-up period was 15.6 months, and the median OS was 12.6 months (95% CI, 9.2–19.2 months). The intracranial ORR was 33.3% in 33 previously treated patients with stable central nervous system (CNS) metastasis. Approximately 44.8% of treatment-related adverse events of ≥ grade 3 were observed, with the most common toxic effects being observed in the gastrointestinal (diarrhea, nausea, and vomiting) and hepatic (elevated liver enzyme levels) systems (Jänne et al., 2022).

3.2 Resistance mechanism and overcoming strategy of KRAS G12C-targeted drugs

Despite the availability of therapeutic agents for targeting KRAS G12C mutations and positive clinical outcomes, patients and physicians have to deal with the possibility of cancer cells developing resistance to sotorasib and adagrasib during clinical trials and actual treatment. Drug resistance mechanisms can be divided into two types: intrinsic and acquired drug resistance mechanisms.

3.2.1 Low dependence on KRAS signaling

Intrinsic resistance refers to the presence of drugs before treatment. Some preclinical studies have shown that intrinsic resistance may be one of the important reasons for the heterogeneous response of patients to clinically targeted KRAS G12C therapy, which may result from the low dependence on KRAS signaling (Canon et al., 2019; Jiao and Yang 2020). As mentioned above, the RAS protein functions through various pathways, mainly including the mitogen-activated protein kinase (MAPK)–ERK and PI3K–AKT–mTORC1 pathways. Previous studies have shown that all KRAS-mutant cells do not depend on KRAS activation to maintain survival and may have different levels of dependence on KRAS signaling. In a study, activation of the two major downstream effectors, namely, ERK and AKT, was not inhibited after KRAS knockdown (Singh et al., 2009). Another study showed that some KRAS-independent cell lines or those with a low level of dependence survived even after KRAS was completely inhibited. In addition, a majority of these cells showed PI3K-dependent MAPK pathway activation, indicating that inhibition of the MAPK pathway was more extensive and effective (Muzumdar et al., 2017). These findings suggest that MAPK pathway inhibitors are potential options for overcoming KRAS G12C-targeted drug resistance.

AMG510 and MRTX849 have shown good therapeutic results in clinical trials. However, rapid adaptive resistance and MAPK signal reactivation after inhibitor treatment have been reported (Ryan et al., 2020). Complementary therapeutic strategies may help to realize the full potential of targeting KRAS mutants for cancer treatment. Proteolysis-targeting chimera (PROTAC)-mediated degradation serves as a complementary strategy for regulating KRAS mutants. PROTACs are emerging drugs that degrade target proteins through cellular proteasome degradation mechanisms. In 2020, Bond et al. (Bond et al., 2020) developed the PROTAC molecule LC-2 that degrades KRAS G12C mutants. LC-2 covalently binds KRAS G12C to an MRTX849 warhead, inducing rapid and continuous degradation of KRAS G12C, leading to the inhibition of the MAPK signal in KRAS G12C-mutant cell lines. It reduces oncogenic KRAS levels and downstream signal transduction in cancer cells and is the first reported compound that can degrade endogenous KRAS G12C.

3.2.2 Synthesis of novel KRAS G12C protein

In a study (Xue, 2020), the main mechanism underlying the resistance of cancer cells to KRAS mutant inhibitors was investigated, and KRAS was found to function in both activated and inactivated states, whereas inhibitors were found to act only on inactivated KRAS. When inhibitors act on tumor cells, the function of KRAS is inhibited. However, some tumor cells can bypass this effect and resume proliferation. This rapid differential response occurs because some quiescent cells produce a new KRAS G12C protein in response to the inhibition of MAPK signaling. The newly synthesized KRAS G12C protein is maintained in an active, drug-insensitive state by epidermal growth factor receptor and Aurora kinase signaling. Cells that do not have these adaptations, or are inhibited by drugs, remain sensitive to drug treatment.

3.2.3 RTK-mediated activation of wild-type RAS

RTKs are the largest class of enzyme-associated receptors, which are cell surface receptors for many growth factors, cytokines, and hormones. Studies have shown that RTK-

mediated activation of wild-type RAS (NRAS or HRAS) can drive the inhibition of the KRAS pathway and drug resistance caused by KRAS G12C inhibitors (Jiao and Yang 2020; Ryan et al., 2020). Therefore, the use of RTK inhibitors may be a potentially effective strategy for overcoming the resistance to KRAS G12C inhibitors.

As discussed above, feedback reactivation mediated by multiple RTKs is a key mechanism for developing resistance to KRAS G12C-targeted agents. Targeting transduction molecules downstream of RTKs, such as SHP2 and SOS1, is one of the primary focus areas of current clinical research. SHP2 is a protein tyrosine phosphatase that activates the SOS-regulated RAS–GTP load to mediate cellular signaling *via* the RAS/MAP kinase signaling pathway (Ruess et al., 2018). Overactivation of SHP2 can stimulate the activation of several signaling pathways, including RAS–MAPK, Akt, and STAT5, thereby promoting the occurrence and development of tumors (Dong et al., 1996). Therefore, targeting SHP2 and SOS1 can be combined with KRAS G12C inhibitors. SHP2 is a phosphatase associated with the activation–inactivation mechanism of KRAS. In a study, a biopsy was performed on a patient with advanced lung cancer treated with AMG 510, and the potential drug resistance factors were identified. Because HER2 activation mediates resistance to targeted drugs in EGFR-mutated lung cancer, the study assessed its functional relevance in KRAS G12C-positive lung cancer. In controlled trials, sotorasib treatment combined with the SHP2 inhibitor TNO155 (NCT03114319, Table 2) has been found to synergistically inhibit cell proliferation and clonal growth in KRAS G12-positive models of HER2 overexpression (Ho et al., 2021).

3.2.4 Mechanisms of acquired drug resistance

Acquired resistance mechanisms of KRAS G12C inhibitors can be divided into the following three categories (Awad et al., 2021): 1) On-target resistance of KRAS G12C inhibitors: KRAS G12C mutation transforms to G12D/R/V/W and other subtypes or G12C amplification. 2) Histological transformation: lung adenocarcinoma with KRAS G12C mutation can be histologically transformed to squamous cell carcinoma. 3) Off-target resistance of KRAS G12C inhibitors: acquired bypass resistance mechanisms include MET amplification; activation mutations of NRAS, BRAF, MAP2K1, and RET; oncogenic fusing of ALK, RET, BRAF, RAF1, and FGFR3; and loss-of-function mutations in NF1 and PTEN. Recent studies (Adachi et al., 2020) have shown that activation of epithelial–mesenchymal transition (EMT) leads to primary and acquired resistance to KRAS G12C inhibitors. In cells with KRAS G12C inhibitor resistance, EMT is induced *via* activation of the PI3K pathway, which eventually leads to the occurrence of drug resistance. In addition, AMG 510 increases PD-1 expression on CD8+ T cells, which may lead to immunosuppression of the tumor immune microenvironment, leading to secondary drug resistance.

Immune checkpoint antibodies block the PD-1 pathway by targeting PD-L1 or PD-1 and have shown good clinical efficacy in treating various malignancies, including NSCLC, in which KRAS/RAF mutations are the common driving events (Brahmer et al., 2015). A meta-analysis conducted by Kim et al. (2017) showed that compared with docetaxel, ICIs improved the overall survival rate of patients with advanced NSCLC with KRAS mutation. In a preclinical study on AMG 510 monotherapy, AMG 510 degraded tumors but did not cure them in the absence of immune cells in mice models, indicating that immune ability can promote the therapeutic effects of AMG 510 in mice with tumors. Subsequently, Amgen proposed the combination of AMG510 and the pembrolizumab antibody, and in 9 of 10 mice treated with this combination therapy, tumors disappeared permanently, significantly improving survival. In addition, mice treated with the combination therapy developed the ability to reject KRAS G12D tumors, suggesting that the combination therapy drives an acquired immune response. The combination of AMG 510 and anti-PD-1 antibody enhances tumor-specific T-cell response and further enhances antitumor T-cell activity. Amgen is currently testing this combination in patients with NSCLC (Canon et al., 2019). In a study, compared with monotherapy, the combination of AMG 510 and an MAPK inhibitor significantly enhanced the antitumor activity both *in vitro* and *in vivo* (Canon et al., 2019).

Therefore, combination therapy is necessary to improve the treatment of KRAS G12C-positive NSCLC and reduce the impact of drug resistance on the therapeutic efficacy. To date, many preclinical studies have provided a theoretical basis for combination therapy to enter clinical trials.

3.3 Prognostic nature of KRAS G12C mutation in non-small-cell lung cancer treated with different strategies

At present, the treatment of NSCLC mainly includes surgery, chemotherapy, immunotherapy, and targeted therapy. To improve therapeutic guidance in clinical practice, we examined the prognostic characteristics of the KRAS mutation subtypes in NSCLC treated with different regimes based on previous studies.

Some studies have shown that KRAS mutation is an independent prognostic factor for the poor prognosis of patients with NSCLC, which indicates shortened OS and an increased risk of tumor recurrence (Renaud et al., 2015). Compared with the KRAS wild-type, the KRAS G12C mutation suggests a poor prognosis. PFS and OS

TABLE 1 Selected major studies about the prognostic relevance of KRAS G12C status in lung cancer.

Authors	Pts	Treatments	Comparison	Findings	Results
Liu. (2020)	(Stage I–IV) N = 130	Various	KRAS G12C vs. KRAS WT	1. OS: KRAS G12C vs. KRAS WT (18.3months vs. 26.7 months)	Poor prognosis
Finn et al. (2021)	(Stage I–III) N = 1014	Various	KRAS G12C vs. KRAS other	1. OS: KRAS G12C vs. KRAS other (HR = 1.39; 95% CI: 1.03–1.89, p = 0.031)	Poor prognosis
			KRAS G12C vs. KRAS WT	2. OS:KRAS G12C vs. KRAS WT (HR = 1.32; p = 0.028)	
Park et al. (2017)	(Stage IIB–IV) N = 334	Chemotherapy	KRAS G12C vs. KRAS WT	1. PFS:KRAS G12C vs. KRAS WT (HR = 1.96; p = 0.045)	Poor prognosis
Svaton et al. (2016)	(Advanced NSCLC) N = 127	Chemotherapy	KRAS G12C vs. KRAS other	1. OS: KRAS other vs. KRAS G12C mut (median 10.3 vs. 6.4 months, p = 0.011).	Poor prognosis
			KRAS mut vs. KRAS WT	2. OS: KRAS WT vs. KRAS mut (16.1 vs. 7.2 months, p = 0.008).	
Lei et al. (2020)	(Stage IIIB–IV) N = 75	Chemotherapy	KRAS G12C vs. KRAS other	1. PFS: KRAS G12C VS. KRAS other (median 4.7 vs. 2.5 months, $p < 0.05$).	Better prognosis
Nadal et al. (2015)	(Resected lung adenocarcinoma) N = 179	Surgical operation	KRAS G12C vs. KRAS other	1. DFS: KRAS WT vs. KRAS mut (p = 0.009).	Poor prognosis
			KRAS WT vs. KRAS mut	2. DFS:KRAS G12C vs. KRAS other ($p <$ 0.001).	
Izar et al. (2014)	(Resected stage I lung adenocarcinoma) N = 312	Surgical operation	KRAS mut vs. KRAS WT	1. DFS: KRAS mut vs. KRAS WT (HR = 3.62,$p < 0.0001$).	Better prognosis
			KRAS G12C/G12V mut vs. KRAS other	1. DFS: G12C/G12V mut vs. KRAS other (p = 0.0271)	
Wu et al. (2021)	N = 246	Immune checkpoint inhibitor (ICI)	Kras G12C vs. KRAS other	1. PFS:KRAS G12C vs. KRAS other (median 4.8 VS. 2.1 m)	Better prognosis
Fiala et al. (2013)	(Advanced NSCLC) N = 448	Targeted therapy	KRAS G12C vs. KRAS other	1. PFS: KRAS G12C vs. KRAS other (4.3 week vs. 9.0 week).	Poor prognosis
				2. OS: KRAS G12C vs. KRAS other (9.3 week vs. 12.1 week).	
Ihle et al. (2012)	(Patients with refractory non-small-cell lung cancer) N = 341	Targeted therapy	KRAS G12C vs. KRAS other	1. MST (median survival time) KRAS G12C/G12V vs. KRAS other 1.84 VS. 3.35 months.	Poor prognosis

KRAS WT, KRAS wild type; KRAS other, KRAS mutant subtypes that do not contain the KRAS G12C mutation.

rates are lower in patients with KRAS G12C mutation receiving different treatment schemes than in patients with other KRAS mutation subtypes (Izar et al., 2014; Svaton et al., 2016; Park et al., 2017; Liu et al., 2020; Finn et al., 2021). If the therapeutic efficacy of ICIs (Wu et al., 2021) and the resection of stage I lung adenocarcinoma are favorable, better DFS is observed (Nadal et al., 2015). In a study involving 75 patients with clinical stage II–IV KRAS-mutant NSCLC, KRAS G12C mutation was identified as a predictive biomarker for better survival benefits from first-line chemotherapy in patients with advanced NSCLC with KRAS mutations (Lei et al., 2020). In another study on patients with advanced NSCLC who received EGFR–TKI treatment, KRAS G12C mutation was identified as a strong negative predictor of therapeutic benefits (Fiala et al., 2013). In addition, KRAS G12C mutation predicts worse PFS, whereas other KRAS mutants predicted better PFS in patients with NSCLC receiving targeted therapy (Ihle et al., 2012) (Table 1).

4 KRAS G12V mutation in non-small-cell lung cancer

The KRAS G12V mutant has a valine residue (glycine at position 12 is mutated to valine). The KRAS G12V mutation occurs in approximately 21% of patients with NSCLC with KRAS mutations (Xie et al., 2021). Individuals who smoke are more predisposed to NSCLC with KRAS G12V mutation.

4.1 Prognostic and predictive role of KRAS G12V mutation in non-small-cell lung cancer

In a study by Jia et al. (2017) on Chinese patients with advanced NSCLC who received cisplatin- or carboplatin-based chemotherapy, PFS was significantly shorter in patients with advanced NSCLC with KRAS G12V mutation than in patients with wild-type KRAS or other KRAS mutations. Another study

TABLE 2 Ongoing clinical trials related to KRAS mutated NSCLC.

Drug	Mechanism of action	Clinical trials	Phase	Clinical staging associated with NSCLC
AMG510	KRAS G12C inhibitor	NCT03600883	I/II	Advanced with KRAS G12C mutation
AMG510	KRAS G12C inhibitor	NCT04303780	III	Advanced with KRAS G12C mutation
AMG510	KRAS G12C inhibitor	NCT04933695	II	Advanced with KRAS G12C mutation
MRTX849	KRAS G12C inhibitor	NCT03785249	I/II	Advanced with KRAS G12C mutation
TNO155	SHP2 inhibitor	NCT03114319	I	Advanced with KRAS G12C mutation
mRNA-5671	mRNA-derived KRAS targeted cancer vaccine	NCT03948763	I	Advanced with KRASG12C, G12D, G12V, or G13C mutation

showed that KRAS mutation subtypes can predict disease recurrence and metastasis after surgical treatment of NSCLC, and patients with NSCLC with KRAS G12V mutation are predisposed to pleura–pericardial metastasis after surgery (Renaud et al., 2016). In addition, patients with KRAS G12V-positive lung cancer have a poor OS and a high recurrence rate.

4.2 KRAS G12V mutation treatment strategies

KRAS mutations may be a potential biomarker for predicting the survival benefit of immunotherapy. Lan et al. (2018) reported that the expression of PD-L1 is significantly correlated with KRAS gene mutations. Pan et al. (2021) reported that the expression of PD-L1 was higher in patients with NSCLC with KRAS G12V mutation than in patients with wild-type KRAS. KRAS G12V mutation can induce PD-L1 expression through the transforming growth factor-β/EMT signaling pathway to promote the immune escape of KRAS-mutant NSCLC. Therefore, the beneficial effects of PD-1/PD-L1 immunotherapy may be better among patients with NSCLC with KRAS G12V mutation than among patients with other KRAS mutations.

In a study on the clinical efficacy of chemotherapeutic agents in patients with advanced NSCLC with KRAS mutations, platinum-based chemotherapy with taxane had the best ORR, especially when used in combination with bevacizumab. Compared with pemetrexed, taxanes significantly improve PFS. The KRAS mutation subtypes respond differently to chemotherapeutic regimens. In particular, patients with KRAS G12V mutation respond well to taxane treatment (Mellema et al., 2015). In a recent retrospective study, patients with KRAS G12V- or G12A-positive advanced NSCLC who received first-line platinum-based chemotherapy with taxane had longer PFS than those who received platinum-based chemotherapy with pemetrexed or gemcitabine. In addition, compared with wild-type KRAS, KRAS G12V mutation is associated with increased sensitivity to cisplatin in NSCLC cell lines (Garassino et al., 2011).

5 KRAS G12D mutation in non-small-cell lung cancer

The KRAS G12D mutation refers to the presence of an aspartic acid residue (glycine at position 12 is mutated to aspartic acid). KRAS G12D mutations occur in approximately 15% of patients with NSCLC with KRAS mutations (Xie et al., 2021). Epidemiologically, KRAS G12D mutation is more likely to occur in patients with NSCLC who do not smoke.

The use of PD-L1 inhibitor therapy in patients with KRAS mutations remains controversial. A study on the influence of KRAS mutations on immune biomarkers revealed a significantly decreased expression of PD-L1 protein and immune cell infiltration (activated CD4 memory T cells, helper T cells, M1 macrophages, and NK cells) in the KRAS G12D/TP53 mutant group. KRAS G12D/TP53 co-mutation drives immune suppression and may serve as a negative predictive biomarker (Gao et al., 2020).

6 Discussion and outlook

KRAS mutations occur in a significant proportion of patients with NSCLC; therefore, early detection and treatment are critical. The genetic characteristics, mutational mechanisms, and susceptibility of various KRAS mutant subtypes to treatment should be extensively investigated. In terms of diagnostic technology, liquid biopsy has been continuously improved and optimized since the discovery of circulating tumor cells in the blood and the introduction of precision oncology. Liquid biopsy can allow early detection of cancer and is more sensitive than traditional cancer screening methods such as radiology or imaging. In addition, it can decrease overall healthcare costs (Heitzer et al., 2019). However, the development of new detection methods is bound to encounter biological and technical challenges. Recently, a study from the Italian Scientific Society discussed the most pressing technical issues in liquid biopsy with the hope of improving its application in clinical practice (Russo et al., 2021).

In terms of treatment, research into KRAS G12C-targeted drugs has offered novel therapeutic strategies for patients with KRAS mutations. However, targeted resistance will inevitably occur in clinical settings. Therefore, the identification of patients who might benefit from targeted monotherapy or combination therapy is important, and methods to identify such patients should be developed. In addition, the selection and implementation of a combination therapy regimen are important to improve the therapeutic efficacy and reduce the incidence of side effects. Furthermore, the treatment of KRAS-mutant cancer should not be limited to the currently available treatment strategies. In recent years, mRNA-based immunotherapy has entered clinical trials and achieved success in the treatment of solid malignancies. Moderna introduced G12C, G12D, G12V, and G13C mutant antigen candidate mRNA products in 2017, which entered phase I clinical trials in 2019 (NCT03948763, Table 2) (Bear et al., 2021; Miao et al., 2021). Future studies should be focused on examining the relationship between the immunological microbial environment and KRAS-induced immune responses. Other cancer therapies, including cutting-edge techniques such as RNAi and CRISPR technologies, which can be used to knockdown genes, transfer payloads into tumors, and activate the immune system, may also help to eliminate KRAS-mutated tumors. In addition, novel therapeutic strategies should be developed for the future treatment of cancer.

Author contributions

LG contributed to the ideation and design of this article and wrote the first draft of the manuscript. WS contributed to the revision of the manuscript. All authors read and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abraham, S. J., Muhamed, I., Nolet, R., Yeung, F., and Gaponenko, V. (2010). Expression, purification, and characterization of soluble K-Ras4B for structural analysis. *Protein Expr. Purif.* 73, 125–131. doi:10.1016/j.pep.2010.05.015

Adachi, Y., Ito, K., Hayashi, Y., Kimura, R., Tan, T. Z., Yamaguchi, R., et al. (2020). Epithelial-to-Mesenchymal transition is a cause of both intrinsic and acquired resistance to KRAS G12C inhibitor in KRAS G12C-mutant non-small cell lung cancer. *Clin. Cancer Res.* 26, 5962–5973. doi:10.1158/1078-0432.ccr-20-2077

Awad, M. M., Liu, S., Rybkin, I. I., Arbour, K. C., Dilly, J., Zhu, V. W., et al. (2021). Acquired resistance to KRAS(G12C) inhibition in cancer. *N. Engl. J. Med.* 384, 2382–2393. doi:10.1056/NEJMoa2105281

Bear, A. S., Blanchard, T., Cesare, J., Ford, M. J., Richman, L. P., Xu, C., et al. (2021). Biochemical and functional characterization of mutant KRAS epitopes validates this oncoprotein for immunological targeting. *Nat. Commun.* 4365, 4365. doi:10.1038/s41467-021-24562-2

Bond, M. J., Chu, L., Nalawansha, D. A., Li, K., and Crews, C. M. (2020). Targeted degradation of oncogenic KRAS(G12C) by VHL-recruiting PROTACs. *ACS Cent. Sci.* 6, 1367–1375. doi:10.1021/acscentsci.0c00411

Brahmer, J., Reckamp, K. L., Baas, P., Crinò, L., Eberhardt, W. E., Poddubskaya, E., et al. (2015). Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* 373, 123–135. doi:10.1056/NEJMoa1504627

Canon, J., Rex, K., Saiki, A. Y., Mohr, C., Cooke, K., Bagal, D., et al. (2019). The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* 575, 217–223. doi:10.1038/s41586-019-1694-1

Cox, A. D., and Der, C. J. (2010). Ras history: The saga continues. *Small GTPases* 1, 2–27. doi:10.4161/sgtp.1.1.12178

Dearden, S., Stevens, J., Wu, Y. L., and Blowers, D. (2013). Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). *Ann. Oncol.* 24, 2371–2376. doi:10.1093/annonc/mdt205

Dogan, S., Shen, R., Ang, D. C., Johnson, M. L., D'Angelo, S. P., Paik, P. K., et al. (2012). Molecular epidemiology of EGFR and KRAS mutations in 3, 026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers. *Clin. Cancer Res.* 18, 6169–6177. doi:10.1158/1078-0432.ccr-11-3265

Dong, C., Waters, S. B., Holt, K. H., and Pessin, J. E. (1996). SOS phosphorylation and disassociation of the Grb2-SOS complex by the ERK and JNK signaling pathways. *J. Biol. Chem.* 271, 6328–6332. doi:10.1074/jbc.271.11.6328

Dong, Z. Y., Zhong, W. Z., Zhang, X. C., Su, J., Xie, Z., Liu, S. Y., et al. (2017). Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin. Cancer Res.* 23, 3012–3024. doi:10.1158/1078-0432.ccr-16-2554

El Osta, B., Behera, M., Kim, S., Berry, L. D., Sica, G., Pillai, R. N., et al. (2019). Characteristics and outcomes of patients with metastatic KRAS-mutant lung adenocarcinomas: The lung cancer mutation consortium experience. *J. Thorac. Oncol.* 14, 876–889. doi:10.1016/j.jtho.2019.01.020

Fell, J. B., Fischer, J. P., Baer, B. R., Blake, J. F., Bouhana, K., Briere, D. M., et al. (2020). Identification of the clinical development candidate MRTX849, a covalent KRAS(G12C) inhibitor for the treatment of cancer. *J. Med. Chem.* 63, 6679–6693. doi:10.1021/acs.jmedchem.9b02052

Fiala, O., Pesek, M., Finek, J., Benesova, L., Belsanova, B., and Minarik, M. (2013). The dominant role of G12C over other KRAS mutation types in the negative prediction of efficacy of epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer. *Cancer Genet.* 206, 26–31. doi:10.1016/j.cancergen.2012.12.003

Finn, S. P., Addeo, A., Dafni, U., Thunnissen, E., Bubendorf, L., Madsen, L. B., et al. (2021). Prognostic impact of KRAS G12C mutation in patients with NSCLC: Results from the European thoracic oncology platform lungscape project. *J. Thorac. Oncol.* 16, 990–1002. doi:10.1016/j.jtho.2021.02.016

Gao, G., Liao, W., Ma, Q., Zhang, B., Chen, Y., and Wang, Y. (2020). KRAS G12D mutation predicts lower TMB and drives immune suppression in lung adenocarcinoma. *Lung Cancer* 149, 41–45. doi:10.1016/j.lungcan.2020.09.004

Garassino, M. C., Marabese, M., Rusconi, P., Rulli, E., Martelli, O., Farina, G., et al. (2011). Different types of K-Ras mutations could affect drug sensitivity and tumour behaviour in non-small-cell lung cancer. *Ann. Oncol.* 22, 235–237. doi:10.1093/annonc/mdq680

Goebel, L., Müller, M. P., Goody, R. S., and Rauh, D. (2020). KRasG12C inhibitors in clinical trials: a short historical perspective. *RSC Med. Chem.* 11, 760–770. doi:10.1039/d0md00096e

Hallin, J., Engstrom, L. D., Hargis, L., Calinisan, A., Aranda, R., Briere, D. M., et al. (2020). The KRAS(G12C) inhibitor MRTX849 provides insight toward therapeutic susceptibility of KRAS-mutant cancers in mouse models and patients. *Cancer Discov.* 10, 54–71. doi:10.1158/2159-8290.cd-19-1167

Hancock, J. F., Paterson, H., and Marshall, C. J. (1990). A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* 63, 133–139. doi:10.1016/0092-8674(90)90294-o

Heitzer, E., Haque, I. S., Roberts, C. E. S., and Speicher, M. R. (2019). Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* 20, 71–88. doi:10.1038/s41576-018-0071-5

Herdeis, L., Gerlach, D., McConnell, D. B., and Kessler, D. (2021). Stopping the beating heart of cancer: KRAS reviewed. *Curr. Opin. Struct. Biol.* 71, 136–147. doi:10.1016/j.sbi.2021.06.013

Ho, C. S. L., Tüns, A. I., Schildhaus, H. U., Wiesweg, M., Grüner, B. M., Hegedus, B., et al. (2021). HER2 mediates clinical resistance to the KRAS(G12C) inhibitor sotorasib, which is overcome by co-targeting SHP2. *Eur. J. Cancer* 159, 16–23. doi:10.1016/j.ejca.2021.10.003

Hong, D. S., Fakih, M. G., Strickler, J. H., Desai, J., Durm, G. A., Shapiro, G. I., et al. (2020). KRAS(G12C) inhibition with sotorasib in advanced solid tumors. *N. Engl. J. Med.* 383, 1207–1217. doi:10.1056/NEJMoa1917239

Ihle, N. T., Byers, L. A., Kim, E. S., Saintigny, P., Lee, J. J., Blumenschein, G. R., et al. (2012). Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *J. Natl. Cancer Inst.* 104, 228–239. doi:10.1093/jnci/djr523

Izar, B., Zhou, H., Heist, R. S., Azzoli, C. G., Muzikansky, A., Scribner, E. E., et al. (2014). The prognostic impact of KRAS, its codon and amino acid specific mutations, on survival in resected stage I lung adenocarcinoma. *J. Thorac. Oncol.* 9, 1363–1369. doi:10.1097/jto.0000000000000266

Jänne, P. A., Riely, G. J., Gadgeel, S. M., Heist, R. S., Ou, S. I., Pacheco, J. M., et al. (2022). Adagrasib in non-small-cell lung cancer harboring a *KRAS*^{G12C} mutation. *N. Engl. J. Med.* 387, 120–131. doi:10.1056/NEJMoa2204619

Jeng, H. H., Taylor, L. J., and Bar-Sagi, D. (2012). Sos-mediated cross-activation of wild-type Ras by oncogenic Ras is essential for tumorigenesis. *Nat. Commun.* 1168, 1168. doi:10.1038/ncomms2173

Jia, Y., Jiang, T., Li, X., Zhao, C., Zhang, L., Zhao, S., et al. (2017). Characterization of distinct types of KRAS mutation and its impact on first-line platinum-based chemotherapy in Chinese patients with advanced non-small cell lung cancer. *Oncol. Lett.* 14, 6525–6532. doi:10.3892/ol.2017.7016

Jiao, D., and Yang, S. (2020). Overcoming resistance to drugs targeting *KRAS*^{G12C} mutation. *Innovation*. 1, 100035. doi:10.1016/j.xinn.2020.100035

Jordan, E. J., Kim, H. R., Arcila, M. E., Barron, D., Chakravarty, D., Gao, J., et al. (2017). Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discov.* 7, 596–609. doi:10.1158/2159-8290.cd-16-1337

Kim, J. H., Kim, H. S., and Kim, B. J. (2017). Prognostic value of KRAS mutation in advanced non-small-cell lung cancer treated with immune checkpoint inhibitors: A meta-analysis and review. *Oncotarget* 8, 48248–48252. doi:10.18632/oncotarget.17594

Lan, B., Ma, C., Zhang, C., Chai, S., Wang, P., Ding, L., et al. (2018). Association between PD-L1 expression and driver gene status in non-small-cell lung cancer: a meta-analysis. *Oncotarget* 9, 7684–7699. doi:10.18632/oncotarget.23969

Lei, L., Wang, W. X., Yu, Z. Y., Liang, X. B., Pan, W. W., Chen, H. F., et al. (2020). A real-world study in advanced non-small cell lung cancer with KRAS mutations. *Transl. Oncol.* 13, 329–335. doi:10.1016/j.tranon.2019.12.004

Lito, P., Solomon, M., Li, L. S., Hansen, R., and Rosen, N. (2016). Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. *Science* 351, 604–608. doi:10.1126/science.aad6204

Liu, S. Y., Sun, H., Zhou, J. Y., Jie, G. L., Xie, Z., Shao, Y., et al. (2020). Clinical characteristics and prognostic value of the KRAS G12C mutation in Chinese non-small cell lung cancer patients. *Biomark. Res.* 22, 22. doi:10.1186/s40364-020-00199-z

Mainardi, S., Mulero-Sánchez, A., Prahallad, A., Germano, G., Bosma, A., Krimpenfort, P., et al. (2018). SHP2 is required for growth of KRAS-mutant non-small-cell lung cancer *in vivo*. *Nat. Med.* 24, 961–967. doi:10.1038/s41591-018-0023-9

Mellema, W. W., Masen-Poos, L., Smit, E. F., Hendriks, L. E., Aerts, J. G., Termeer, A., et al. (2015). Comparison of clinical outcome after first-line platinum-based chemotherapy in different types of KRAS mutated advanced non-small-cell lung cancer. *Lung Cancer* 90, 249–254. doi:10.1016/j.lungcan.2015.09.012

Miao, L., Zhang, Y., and Huang, L. (2021). mRNA vaccine for cancer immunotherapy. *Mol. Cancer* 41, 41. doi:10.1186/s12943-021-01335-5

Muzumdar, M. D., Chen, P. Y., Dorans, K. J., Chung, K. M., Bhutkar, A., Hong, E., et al. (2017). Survival of pancreatic cancer cells lacking KRAS function. *Nat. Commun.* 1090, 1090. doi:10.1038/s41467-017-00942-5

Nadal, E., Beer, D. G., and Ramnath, N. (2015). KRAS-G12C mutation is associated with poor outcome in surgically resected lung adenocarcinoma. *J. Thorac. Oncol.* 10, e9–e10. doi:10.1097/jto.0000000000000438

Pan, L. N., Ma, Y. F., Li, Z., Hu, J. A., and Xu, Z. H. (2021). KRAS G12V mutation upregulates PD-L1 expression via TGF-β/EMT signaling pathway in human non-small-cell lung cancer. *Cell Biol. Int.* 45, 795–803. doi:10.1002/cbin.11524

Park, S., Kim, J. Y., Lee, S. H., Suh, B., Keam, B., Kim, T. M., et al. (2017). KRAS G12C mutation as a poor prognostic marker of pemetrexed treatment in non-small cell lung cancer. *Korean J. Intern. Med.* 32, 514–522. doi:10.3904/kjim.2015.299

Patricelli, M. P., Janes, M. R., Li, L. S., Hansen, R., Peters, U., Kessler, L. V., et al. (2016). Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. *Cancer Discov.* 6, 316–329. doi:10.1158/2159-8290.cd-15-1105

Renaud, S., Falcoz, P. E., Schaëffer, M., Guenot, D., Romain, B., Olland, A., et al. (2015). Prognostic value of the KRAS G12V mutation in 841 surgically resected Caucasian lung adenocarcinoma cases. *Br. J. Cancer* 113, 1206–1215. doi:10.1038/bjc.2015.327

Renaud, S., Seitlinger, J., Falcoz, P. E., Schaeffer, M., Voegeli, A. C., Legrain, M., et al. (2016). Specific KRAS amino acid substitutions and EGFR mutations predict site-specific recurrence and metastasis following non-small-cell lung cancer surgery. *Br. J. Cancer* 115, 346–353. doi:10.1038/bjc.2016.182

Ruess, D. A., Heynen, G. J., Ciecielski, K. J., Ai, J., Berninger, A., Kabacaoglu, D., et al. (2018). Mutant KRAS-driven cancers depend on PTPN11/SHP2 phosphatase. *Nat. Med.* 24, 954–960. doi:10.1038/s41591-018-0024-8

Russo, A., Incorvaia, L., Del Re, M., Malapelle, U., Capoluongo, E., Gristina, V., et al. (2021). The molecular profiling of solid tumors by liquid biopsy: a position paper of the AIOM-SIAPEC-IAP-SIBioC-SIC-SIF Italian scientific societies. *ESMO Open* 6, 100164. doi:10.1016/j.esmoop.2021.100164

Ryan, M. B., Fece de la Cruz, F., Phat, S., Myers, D. T., Wong, E., Shahzade, H. A., et al. (2020). Vertical pathway inhibition overcomes adaptive feedback resistance to KRAS(G12C) inhibition. *Clin. Cancer Res.* 26, 1633–1643. doi:10.1158/1078-0432.ccr-19-3523

Scheffler, M., Ihle, M. A., Hein, R., Merkelbach-Bruse, S., Scheel, A. H., Siemanowski, J., et al. (2019). K-ras mutation subtypes in NSCLC and associated Co-occuring mutations in other oncogenic pathways. *J. Thorac. Oncol.* 14, 606–616. doi:10.1016/j.jtho.2018.12.013

Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F., et al. (1997). The ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic ras mutants. *Science* 277, 333–338. doi:10.1126/science.277.5324.333

Shepherd, F. A., Domerg, C., Hainaut, P., Jänne, P. A., Pignon, J. P., Graziano, S., et al. (2013). Pooled analysis of the prognostic and predictive effects of KRAS mutation status and KRAS mutation subtype in early-stage resected non-small-cell lung cancer in four trials of adjuvant chemotherapy. *J. Clin. Oncol.* 31, 2173–2181. doi:10.1200/jco.2012.48.1390

Singh, A., Greninger, P., Rhodes, D., Koopman, L., Violette, S., Bardeesy, N., et al. (2009). A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival. *Cancer Cell* 15, 489–500. doi:10.1016/j.ccr.2009.03.022

Skoulidis, F., Li, B. T., Dy, G. K., Price, T. J., Falchook, G. S., Wolf, J., et al. (2021). Sotorasib for lung cancers with KRAS p.G12C mutation. *N. Engl. J. Med.* 384, 2371–2381. doi:10.1056/NEJMoa2103695

Svaton, M., Fiala, O., Pesek, M., Bortlicek, Z., Minarik, M., Benesova, L., et al. (2016). The prognostic role of KRAS mutation in patients with advanced NSCLC treated with second- or third-line chemotherapy. *Anticancer Res.* 36, 1077–1082.

Tomasini, P., Walia, P., Labbe, C., Jao, K., and Leighl, N. B. (2016). Targeting the KRAS pathway in non-small cell lung cancer. *Oncologist* 21, 1450–1460. doi:10.1634/theoncologist.2015-0084

Vetter, I. R., and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science* 294, 1299–1304. doi:10.1126/science.1062023

Welman, A., Burger, M. M., and Hagmann, J. (2000). Structure and function of the C-terminal hypervariable region of K-Ras4B in plasma membrane targetting and transformation. *Oncogene* 19, 4582–4591. doi:10.1038/sj.onc.1203818

West, H. J., McCleland, M., Cappuzzo, F., Reck, M., Mok, T. S., Jotte, R. M., et al. (2022). Clinical efficacy of atezolizumab plus bevacizumab and chemotherapy in KRAS-mutated non-small cell lung cancer with STK11, KEAP1, or TP53 comutations: subgroup results from the phase III IMpower150 trial. *J. Immunother. Cancer* 10, e003027. doi:10.1136/jitc-2021-003027

Wu, S. G., Liao, W. Y., Su, K. Y., Yu, S. L., Huang, Y. L., Yu, C. J., et al. (2021). Prognostic characteristics and immunotherapy response of patients with nonsquamous NSCLC with kras mutation in East asian populations: A single-center cohort study in taiwan. *JTO Clin. Res. Rep.* 2, 100140. doi:10.1016/j.jtocrr.2020.100140

Xie, M., Xu, X., and Fan, Y. (2021). KRAS-mutant non-small cell lung cancer: An emerging promisingly treatable subgroup. *Front. Oncol.* 11, 672612. doi:10.3389/fonc.2021.672612

Xue, J. Y., Zhao, Y., Aronowitz, J., Mai, T. T., Vides, A., Qeriqi, B., et al. (2020). Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature* 577, 421–425. doi:10.1038/s41586-019-1884-x

Yoshizawa, A., Sumiyoshi, S., Sonobe, M., Kobayashi, M., Fujimoto, M., Kawakami, F., et al. (2013). Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. *J. Thorac. Oncol.* 8, 52–61. doi:10.1097/JTO.0b013e3182769aa8

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fgene.2022.873218

Check for updates

OPEN ACCESS

EDITED BY
Jian-Guo Zhou,
University of Erlangen Nuremberg, Germany

REVIEWED BY
Ankur Sharma,
Harry Perkins Institute of Medical Research, Australia
Andrew Dellinger,
Elon University, United States

*CORRESPONDENCE
Wenzhe Gao,
gaowenzhexy@163.com
Jichun Sun,
sunjichun_1978@163.com
Hongwei Zhu,
zhw_0509@yeah.net

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 10 February 2022
ACCEPTED 12 August 2022
PUBLISHED 24 October 2022

CITATION
Li W, Liu J, Zhu W, Jin X, Yang Z, Gao W, Sun J and Zhu H (2022), Identification of biomarkers for hepatocellular carcinoma based on single cell sequencing and machine learning algorithms.
Front. Genet. 13:873218.
doi: 10.3389/fgene.2022.873218

COPYRIGHT
© 2022 Li, Liu, Zhu, Jin, Yang, Gao, Sun and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of biomarkers for hepatocellular carcinoma based on single cell sequencing and machine learning algorithms

Weimin Li[1,2†], Jixing Liu[1†], Wenjuan Zhu[3], Xiaoxin Jin[4], Zhi Yang[5], Wenzhe Gao[1*], Jichun Sun[1*] and Hongwei Zhu[1*]

[1]Department of Hepatobiliary and Pancreatic Surgery, The Third Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]School of Information, Hunan University of Humanities, Science and Technology, Loudi, China, [3]Division of Nephrology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong, China, [4]The Second Xiangya Hospital, Central South University, Changsha, Hunan, China, [5]Xiangya Hospital, Central South University, Changsha, Hunan, China

Hepatocellular carcinoma (HCC) remains one of the most lethal cancers around the world. Precision oncology will be crucial for further improving the prognosis of HCC patients. Compared with traditional bulk RNA-seq, single-cell RNA sequencing (scRNA-seq) enables the transcriptomes of a great deal of individual cells assayed in an unbiased manner, showing the potential to deeply reveal tumor heterogeneity. In this study, based on the scRNA-seq results of primary neoplastic cells and paired normal liver cells from eight HCC patients, a new strategy of machine learning algorithms was applied to screen core biomarkers that distinguished HCC tumor tissues from the adjacent normal liver. Expression profiles of HCC cells and normal liver cells were first analyzed by maximum relevance minimum redundancy (mRMR) to get a top 50 signature gene feature. For further analysis, the incremental feature selection (IFS) method and leave-one-out cross validation (LOOCV) were conducted to build an optimal classification model and to extract 21 potentially essential biomarkers for HCC cells. Our results provided new insights into HCC pathogenesis that might be valuable for HCC diagnosis and therapy.

KEYWORDS

hepatocellular carcinoma, ScRNA-seq, biomarkers, MRMR algorithm, support vector machine

Introduction

Hepatocellular carcinoma (HCC), with an annual incidence of 8.3 per 1,00,000 in population around the world, remains one of the most lethal malignancies in the digestive system. It is estimated that the 5-years survival rate for HCC patients is 18%, only a little bit higher than pancreatic cancer among all cancers, indicating that HCC is still one of the worst prognostic tumors worldwide (Siegel et al., 2020). Fortunately, with the development of modern cancer therapies, which integrate diverse neoadjuvant and adjuvant strategies with classic surgical resection, the survival rate of HCC has been

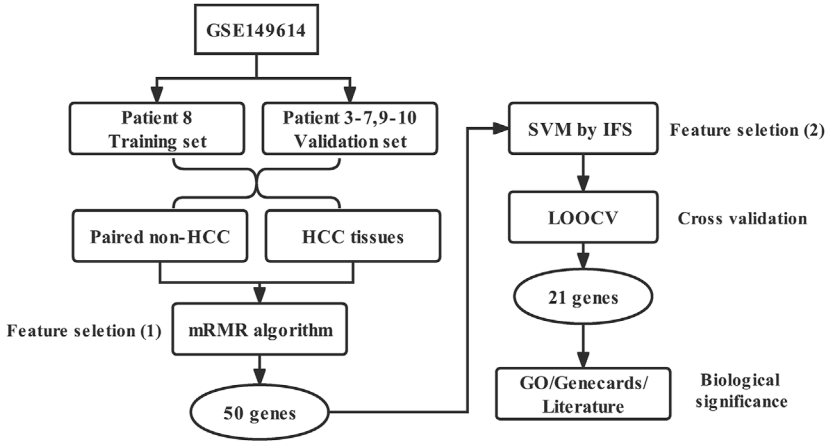


FIGURE 1
Workflow diagram of this study.

gradually improving in the past few decades (McGlynn et al., 2020). However, the highly heterogeneous nature of HCC determines that a large proportion of patients receiving standardized treatment will inevitably relapse (Petrowsky et al., 2020). Thus, precision oncology, including novel predictive and therapeutic oncogenetic markers, signals in tumor immune microenvironment and microbiome, etc. will be crucial for further improving the prognosis of HCC patients (Nault and Villanueva, 2020).

Tumor heterogeneity is the biggest obstacle to the development of precision treatment for HCC, which is decided by heterogeneous HCC cells, a changeful, complex microenvironment, and their involuted interaction. With the rapid development of profiling technology, bulk DNA and RNA sequencing have provided a lot of information about molecular phenotypes and evolutionary characteristics of HCC. In 2020, Amanda J. Craig and her colleagues reviewed the most important and common genetic alterations of HCC, including mutations in the TERT promoter, TP53, and CTNNB1, copy number variations in multiple genes, and aberrations in DNA methylation at the genome level (Craig et al., 2020). Rebouissou and Nault (2020) discovered signal pathways that are frequently altered in HCC patients included telomere maintenance, including Wnt/β-Catenin, P53/cell cycle regulation, oxidative stress, epigenetic modifiers, AKT/mTOR signaling, and MAPK pathway. These findings have generated classification schemas of HCC molecular subtypes. However, these results still encounter many difficulties in real-world clinical applications. On the one hand, bulk sequencing could only detect the average condition of gene alterations or expression status, unable to distinguish the gene expression signature of diverse cells in cancer samples, or get a categorized gene feature between tumors and normal tissues. On the other hand, limited by sample size and traditional methods for differential expression analysis (like Limma and EdgeR), results of bulk RNA-seq often lack stability and repeatability among different batches of experiments. These drawbacks reduce the potential practical value of bulk RNA-seq results.

The advent of single-cell RNA sequencing (scRNA-seq) is a revolutionary development in the field of profiling. Since proposed by Professor F Tang in 2009 (Tang et al., 2010), researchers have immediately realized the infinite value of this technology. It enables the transcriptomes profiling of a lot of individual cells assayed in an unbiased manner, allowing researchers to sort and study the specific characteristics of a single cell or a group of cells individually (Stegle et al., 2015). This technology perfectly fits the research on cancer that has innegligible internal heterogeneity. Thus, scRNA-seq for tumor research has sprung up in the past decade, it has been applied in multiple cancers including HCC. Zhang et al. (2019) detected HCC specific immune cells for five HCC patients using scRNA-seq and discovered that CD45 immune cells, LAMP3(+) dendritic cells, and tumor-associated macrophages were specific infiltrating immune cells in HCC and were associated with patients' poor prognosis. Ho et al. (2019) grouped HCC stem cells for two subgroups through scRNA-seq according to the expression of EPCAM; they also identified a CD24/CD44-enriched cell subpopulation within the EPCAM(+) cells which might indicate a novel stemness-related cell subclone of HCC. These studies illustrate the significance of scRNA-seq for deeply understanding the evolutionary differences among HCC patients, the heterogeneity between HCC tissues and normal livers, between HCC parenchymal cells and microenvironmental

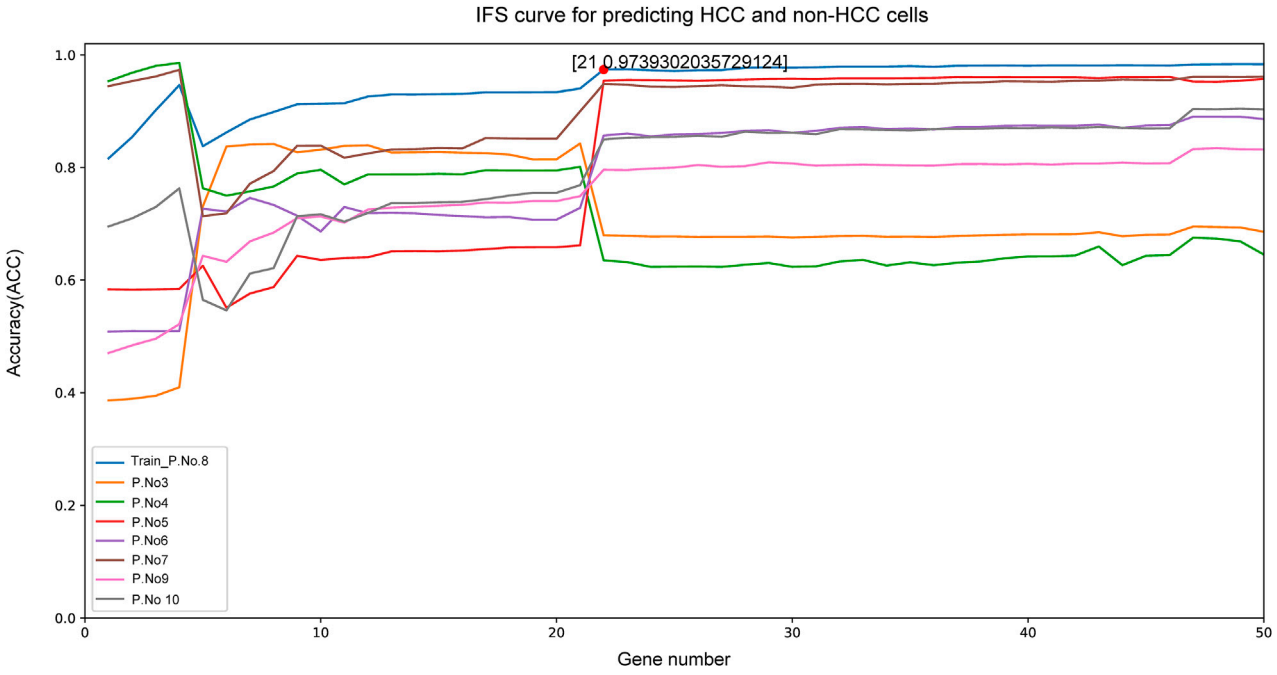


FIGURE 2
The IFS curve of the top 50 mRMR genes. The *x*-axis was the included number of top genes and the *y*-axis was the prediction performance. The blue plot was the training set from patient NO. 8. The peak MCC was 0.974 when top 21 genes included to IFS model. These 21 genes were chosen as the optimal HCC biomarkers.

TABLE 1 The 21 optimal HCC biomarker genes got from IFS method.

Rank	Gene	Rank	Gene	Rank	Gene
1	SPP1	8	GPR18	15	SLAMF6
2	FCN3	9	AKNA	16	TRGC1
3	FCRL6	10	FCMR	17	STAT4
4	S1PR5	11	AC092580.4	18	SCML4
5	CD8A	12	AIM1	19	HBB
6	SAA1	13	GZMM	20	PLAC8
7	CD160	14	IGHA1	21	APOA2

mesenchymal cells. Thus, the multi-dimensional interpretation of tumor heterogeneity by scRNA-seq will effectively solve the current clinical problem of chemoresistance and tumor recurrence and guide tumor immunotherapy and targeted therapy for HCC patients. However, limitations still exist in scRNA-seq research nowadays. First, it seems difficult for scRNA-seq to provide specific genetic markers to guide clinical diagnosis and treatment, which focus more on the alteration of organism and histology levels, making single-cell profiling too "microscopic." Second, even though scRNA-seq provides a temporal map of the tumor microenvironment and cell development and many new clusters of tumor progenitors and immune cells were identified, it is still hard to conduct molecular biological research on these new discoveries to elucidate the pathogenesis underlying the course of diseases. In a word, these limitations were mainly caused by insufficient mining of scRNA-seq big data. Innovative algorithmic strategies are demanded to provide new biological implications for scRNA-seq.

Max-Relevance and Min-Redundancy (mRMR) algorithm provides a highly robust feature selection scheme in machine learning and has been applied in multi-omics medical research in recent years. However, during the process of continuously adding features, mRMR only considers the local optimal solution. Thus, after obtaining the feature set with the importance ranking from high to low through this algorithm, a secondary feature selection is usually followed. For example, Morgan et al. (2021) applied mRMR along with an explainable boosting machine (EBM) classifier for CT radiomics to predict local failure following chemoradiation for head and neck cancer patients. Gao et al. (2020) chose an mRMR plus Random Forest model to find the lncRNA signature in bulk RNA-seq for immunophenotype prediction in Glioblastoma. In scRNA-seq, Cheng et al. (2020) applied an mRMR plus Support Vector Machines (SVM) to screen core biomarkers that distinguish the

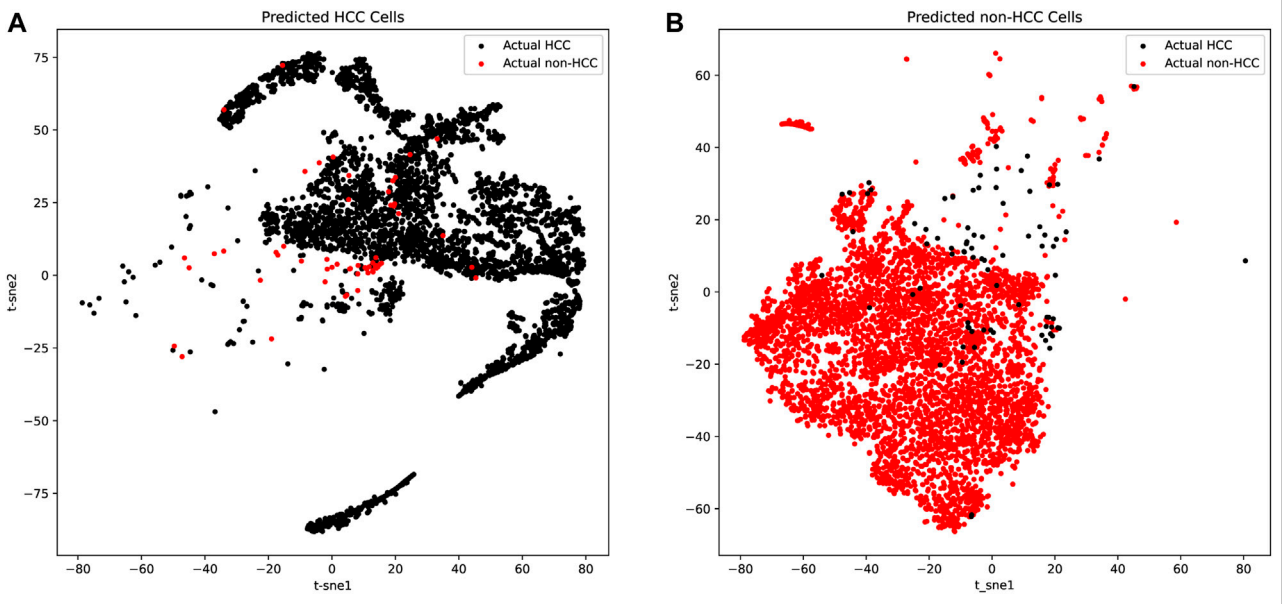


FIGURE 3
The t-SNE plots of predicted HCC cells and normal liver cells. **(A)** The t-SNE plots of predicted HCC cells. The true positive cells (black dots) account for the vast majority, while the false positive cells (red dots) occupied only a few and was mixed in the true positive cells, which was difficult to distinguish. **(B)** The t-SNE plots of predicted normal liver cells. The true negative cells (red dots) account for the vast majority, while the false negative cells (black dots) occupied only a few.

TABLE 2 Confused matrix of the 21 selected genes.

	Predicted HCC	Predicted non-HCC
Actual HCC	4,713	82
Actual non-HCC	105	4,728

discrepancy between GBM tumor and pericarcinomatous environment. Based on the above research, we believed that the algorithmic scheme centered on mRMR might be used to screen biomarkers between cancer and non-cancer in scRNA-seq as well. This might have biological significance in assisting tumor diagnosis and tumor tissue identification during biopsy as well as providing novel parenchyma and stromal biomarkers for a certain cancer type.

In this study, based on paired scRNA-seq results of HCC and adjacent normal liver cells from eight patients, we designed a new computational strategy, consisting of machine learning algorithms, to screen core biomarkers that could distinguish the discrepancy between HCC and normal liver tissue. Gene expression profiles of tumor cells and paired hepatocytes were analyzed by maximum relevance minimum redundancy (mRMR) to get a 50-hub-gene feature. For further screening and classification of the 50-gene-feature, a support vector machine (SVM) algorithm was adopted. Results yielded a gene set with 21 genes that might be essential biomarkers for HCC tumor patients.

Materials and methods

Single cell gene expression profiles of HCC tumors and normal liver tissues

Single-cell gene expression profile data of HCC was obtained and downloaded from Gene Expression Omnibus (GEO) database in NCBI, the accession number was GSE149614. In this dataset, >70,000 single-cell transcriptomes for 10 HCC patients were sequenced and further measured using Illumina NovaSeq 6000 platform (GPL24676). Here we extracted expression profiles of patient 8 as our training set and patient three to seven and 9-10 as our validation set. Patient No. 1 and 2 were excluded from our study for a lack of sequencing data of paired normal liver tissues. The number of expressed genes was counted in paired HCC and normal liver samples of each patient. We utilized this dataset to further establish our gene feature for the purpose of discriminating HCC cells from normal liver cells.

mRMR ranking of discriminative genes

To achieve the goal of best discriminating the two types of tissues using the least number of genes, the Max-Relevance and Min-Redundancy (mRMR) algorithm was applied (Peng et al., 2005). This algorithm aimed to find a gene set that had the biggest correlation between the selected genes and samples (Max-

TABLE 3 GO (BP) enrichment results of the 21 selected genes.

GO_Term	P	Adj.P	Genes
Lymphocyte mediated immunity	2.49E-05	0.014115	CD8A/CD160/GZMM/IGHA1/SLAMF6
Leukocyte mediated cytotoxicity	0.000165	0.030107	CD160/GZMM/SLAMF6
Positive regulation of natural killer cell mediated cytotoxicity	0.000262	0.030107	CD160/SLAMF6
Positive regulation of natural killer cell mediated immunity	0.000384	0.030107	CD160/SLAMF6
Alpha-beta T cell activation	0.000402	0.030107	CD160/GPR18/SLAMF6
Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0.000447	0.030107	CD8A/GZMM/IGHA1/SLAMF6
Cell killing	0.000648	0.030107	CD160/GZMM/SLAMF6

TABLE 4 Summary for markers related to the malignant phenotype or clinical prognosis of HCC.

GeneName	Location	Function summaries	Related pathways	Reported functions in HCC
SPP1	4q22.1	1. Forming an integral part of the mineralized matrix and is to cell-matrix interaction. 2. Acting as a cytokine enhancing IFN-γ and IL-12, reducing production of IL-10, essential in type I immunity.	1. Cytokine activity 2. Integrin binding 3. Protein binding 4. Extracellular matrix binding	1. Prognostic marker for HCC Zheng et al. (2018), Ouyang et al. (2020) 2. Enhancer of cell growth Wang et al. (2019)
SAA1	11p15.1	1. A major acute phase protein that is highly expressed in response to inflammation 2. Major biomarker for diverse tumors	1. Heparin binding and chemoattractant activity 2. Activated TLR4 signalling 3. Signaling by GPCR	Lowly expressed in HCC patients, indicating worse prognosis Zhang et al. (2020)
STAT4	2q32.2-q32.3	A member of the STAT family of transcription factors activated by cytokines	1. DNA-binding transcription factor activity 2. Sequence-specific DNA binding 3. JAK-STAT signaling pathway	Tumor suppressor in HCC that inhibit proliferation and promote apoptosis Li et al. (2016)
PLAC8	4q21.22	A highly conservative protein, physiology function unknown	1. Chromatin binding activity 2. Differentiation of white and brown adipocyte	Downregulated in HCC, indicating poor prognosis when lowly expressed by promoting cell proliferation Zou et al. (2016)

Relevance), but the least correlation between genes inside this gene set (Min-Redundancy). The redundancy between genes was minimized as genes with similar expression characteristics were removed, except for the most representative genes remained. This method was confirmed effective in finding core biomarkers in sequencing analysis, especially in scRNA-seq with large and spare expression data (Cheng et al., 2020). It effectively overcame the shortcomings of traditional differential expression analysis in bulk RNA-seq, helping us to get a smaller number of biomarkers with the highest representation.

The mathematical model of this algorithm was shown as followed. First, we defined all genes, selected genes, and to be selected genes as Ω, Ω_s, and Ω_t, respectively. The relevance (D) of gene g from Ω_t with cell type t can be measured with mutual information (I)

$$\boldsymbol{D = I(g,t)}$$

And the redundancy R of the gene g with the selected genes in Ω_s are

$$\boldsymbol{R = \frac{1}{m}\left(\sum_{g_i \epsilon \Omega s} I(g,g_i)\right)}$$

Now, our goal is to get the gene g_j from Ω_t so that D takes the maximum value ($D_{\max}$) and R takes the minimum value ($R_{\min}$), which can be expressed as the following function

TABLE 5 Summary for markers reported in other cancers.

GeneName	Location	Function summaries	Related pathways	Reported functions in other cancers
FCN3	1p36.11	1. Calcium-independent lectin activity, found in all human serum.	1. Carbohydrate binding activity	Highly expressed in ovarian cancer and leukaemia patients Szala et al. (2013), Sokołowska et al. (2020).
		2. Functioning in innate immunity	2. Antigen binding activity	
		3. Related to innate immune. system and creation of C4 and C2 activators.	3. Complement pathway in association with MASPs and sMAP	
APOA2	1q23.3	1. The second most abundant protein in high density lipoprotein (HDL) particles	1. Lipoprotein metabolism	A minimally invasive biomarker for detecting pancreatic cancer, bladder cancer, and metastatic renal cell cancer patients Vermaat et al. (2012), Chen et al. (2015), Sato et al. (2020)
		2. Stabilizing HDL structure by its association with lipids	2. Signaling by GPCR	
			3. Protein homodimerization	
			4. Activity of lipid binding	
AC092580.4	2p25.1	Affiliated with the lncRNA class	NA	1. Highly expressed in relapse AML patients Feng et al. (2018)
S1PR5	19p13.2	1. Receptor for a bioactive lysophospholipid S1P	1. G protein-coupled receptor activity	2. Highly expressed and promoting proliferation and invasion in clear cell carcinoma (ccRCC) and colon cancer patients Peng et al. (2020), Zhou et al. (2020)
		2. Both intracellular as a second messenger and extracellular as a receptor ligand	2. Sphingosine-1-phosphate receptor activity	
AKNA	9q32	1. Centrosomal protein that plays a key role in cell delamination	1. RNA polymerase II proximal promoter sequence-specific DNA binding activity	1. A tumor suppressor in gastric cancer by modulating EMT Wang et al. (2020)
		2. An epithelial-to-mesenchymal transition (EMT) regulator	2. Proximal promoter DNA-binding transcription activator activity	2. An immune activator in cervical cancer Manzo-Merino et al. (2018)
		3. A transcription factor that specifically activates the expression of the CD40		
AIM1	6q21	NA	NA	1. A classical tumor suppressor with high mutational frenquency in melanoma Ray et al. (1997)
				2. Suppressing tumor migration in prostate cancer Haffner et al. (2017)

$$\max_{g_j \in \Omega_s}\left[I(g_j, t) - \frac{1}{\mathrm{m}}\left(\sum_{g_i \in \Omega s} I(g_j, g_i)\right)\right] \quad (j = 1, 2 \ldots, \mathrm{n})$$

After n rounds of evaluation, all genes (Ω) will be ranked as a new gene list

$$S = \{g_1', g_2', \ldots, g_i', \ldots g_N'\}$$

The subscript i here reflects the trade-off between relevance with tissue type and redundancy with selected genes. The smaller index i is, the better discriminating power the gene has, and the higher of the corresponding gene g_i ranks.

Screening and optimization of single cell HCC biomarkers

We then constructed 50 support vector machines (SVM) classifiers and applied an incremental feature selection (IFS) method (Ye et al., 2017) using Top 50 mRMR genes to further screen optimized biomarker genes. The 50 gene sets are defined as

$$S_k = \{g_1', g_2', \ldots, g_k'\} \quad k = (1, 2, \ldots, 50)$$

Each candidate gene set includes the top k genes in the mRMR gene set S.

To prevent overfitting and evaluate the generalization ability of prediction performance for each SVM classifier, the leave-one-out cross validation (LOOCV) (Cheng et al., 2017) was then applied. Here we briefly described the procedure of LOOCV. Supposed that a dataset has N samples, in each round of LOOCV, there are (N–1) samples adopted for training and the remaining one sample for testing. This process keeps running until all the N samples have been tested for one time after N rounds.

Since the positive and negative sample sizes are imbalance, the Matthews correlation coefficient (MCC) (Matthews, 1975), which considered both sensitivity and specificity, seems idealized for our IFS optimizing process. The calculation formula of MCC is shown as followed:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP, TN, FP, and FN are the abbreviation of true positive, true negative, false positive, and false negative, respectively.

TABLE 6 Summary for immune cell markers.

GeneName	Location	Function summaries	Related pathways	Immune cell type	Reported in HCC
IGHA1	14q32.33	Constant region of immunoglobulin heavy chains	1. Cell surface interactions 2. Response to elevated platelet cytosolic Ca2+ 3. Antigen binding activity 4. Immunoglobulin receptor binding activity	B cells	Not Reported
FCRL6	1q23.2	MHC class II receptor	protein phosphatase binding	NK cells and CTLs	Not Reported
CD8A	2p11.2	1. A classic surface glycoprotein on most CTLs mediateing immune cell interactions 2. A coreceptor for MHC class I molecule: peptide complex	1. Protein homodimerization activity 2. Coreceptor activity	CTLs	Reported Sangro et al. (2020)
CD160	1q21.1	1. A transmembrane on immune cells, mainly NK cells and activated T cells 2. Upon persistent antigen stimulation, it may contribute to CTL exhaustion	1. Innate Lymphoid Cell Differentiation Pathways 2. Class I MHC mediated antigen processing and presentation 3. MHC class I receptor activity	NK cells and activated T cells	Reported Sun et al. (2018)
GPR18	13q32.3	1. A cannabinoid-activated orphan G protein-coupled receptor 2. Selected expressed on immune cells	1. Peptide ligand-binding receptors 2. G protein-coupled receptor activity	Tumor-infiltrating B lymphocytes (TIL-Bs) and $CD8^+$ T cells	Not Reported
GZMM	19p13.3	A member of granzymes, serine proteases acctivity	1. Serine-type endopeptidase activity 2. Endopeptidase activity 3. Creation of C4 and C2 activators	NK cells	Not Reported
SLAMF6	1q23.2-q23.3	A type I transmembrane protein, belonging to the CD2 subfamily of the immunoglobulin superfamily	1. Class I MHC mediated antigen processing and presentation 2. Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	Mainly in NK cells, also existing in T and B cells	Not Reported

After all the above procedures, an IFS curve was finally formed. The *x*-axis of this curve denoted the number of genes in the SVM classifier (1–100) and the *y*-axis indicated the MCCs of it. Based on the IFS curve, an inflecting point that represented the usage of relatively few genes to get a relatively higher prediction accuracy was marked. The *x*-coordinate value of this inflecting point indicated number of genes in the final biomarker gene set and the *y*-coordinate value represented the prediction performance. This point was regarded as the most suitable SVM model for the final HCC biomarkers.

Biological significance analysis

For the gene signature predicted by the above algorithms, we then analyzed their expressed cell types, chromosomal location, and functions by GO, GENECARDS and literature reviewing. For GO, biological process (BP) was enriched and the *p* value as well as false discovery rate (FDR) based on hypergeometric distribution were calculated, FDR <0.05 was considered significantly enriched. GENECARDS database was available in https://www.genecards.org/. Literature reviewing was applied using NCBI pubmed databases (www.ncbi.nlm.nih.gov/pubmed/) to search publications for every gene in recent 10 years.

The workflow diagram of this study was shown in Figure 1.

Results and discussion

Identifying the most discriminative feature by mRMR algorithm

After mRMR algorithm was applied, a feature, composed of top 50 most significant genes was established. This gene set was listed in Supplementary Table S1. Based on the principle of the mRMR algorithm, we believed that this feature was the most relevant one to distinguish HCC tumor cells from normal liver cells and had the least redundancy among the elements inside this gene set.

Further screening for the optimal HCC biomarker genes by IFS method

Given that our aim was to discriminate sample groups most significantly using biomarker genes as few as possible, the feature of 50 genes formed by mRMR algorithm was obviously too large to possess a practical value. Thus, we needed to choose an optimized group from these 50 genes as the final marker. To achieve this goal, IFS method was adopted. In the first round, only rank-first gene in mRMR was included as feature gene, then an SVM classifier was built to predict the group of each sample and validation was achieved by LOOCV and quantified by MCC value. In the second round, the rank-second gene was added into the previous 1-gene-feature and the above steps were repeated. This process kept repeated for 500 times until all 50 mRMR genes were included in the SVM model and an IFS curve was formed. As shown in Figure 2, the best peak MCC was 0.974 when 21 genes were included in SVM model in patient No.8 (train set), this peak MCC was also detected in other patients for validation (Figure 2). This peak was also validated effective in other patients except for patient No. 3 and No. 4, with MCC value no more than 0.7 but acceptable. Thus, these 21 genes were adopted as our final optimal HCC biomarkers (Table 1).

We further applied t-distributed stochastic neighbor embedding (t-SNE) for predicted HCC and non-HCC cells to detect both the tumor purity and the robustness of our classifier based on the 21 genes. As shown in Figures 3A,B and Table 2, there were only a few false positive (red dots in Figure 3A) and false negative dots (black dots in Figure 3B) mixed with true positive and true negative samples. However, the proportion of those false dots was extremely low with true dots and hard to classify. These t-SNE plots suggested that the HCC cells might contain non-HCC cells and vice versa, but most cells from the corresponding group were acceptable and the algorithms we applied could get the robust single cell biomarkers even when there were little tissue purity issues.

The biological functions of the 21 hub genes

The machine learning methods provided us with a new set of gene features for HCC to identify tumor from paired normal liver tissues. However, nothing was learned from the biological significance of this gene set. We first performed Gene Ontology (GO) enrichment analysis for biological process (BP) analysis of the selected 21 genes (Table 3). Results of GO enrichment showed that they were enriched in Natural Killer (NK) cells and T cells (Tc) related pathways, indicating that changes in immune microenvironment are the core difference between HCC tumors and normal liver tissues.

To explore the functions of these 21 genes in more depth from a biomedical perspective, we reviewed the location, basic functions, as well as related biological pathways and processes for each gene through *Genecards* database (https://www.genecards.org/). Literatures about the biological functions of these 21 genes in HCC and/or other malignant tumors were also thoroughly searched through Pubmed database. We finally divided these 21 genes into four categories (Siegel et al., 2020) Markers related to the malignant phenotype or clinical prognosis of HCC (Table 4); (McGlynn et al., 2020) Markers without reports in HCC but were related to the pathogenesis and/or prognosis of other malignant tumors (Table 5); (Petrowsky et al., 2020) Marker genes expressed by immune cells (Table 6); (Nault and Villanueva, 2020) Other genes that have not yet been studied in cancers, including *FCMR*, *TRGC1*, and *HBB*. Subsequent research was worth exploring the role of these markers in the pathogenesis of HCC and their applicating prospects in HCC diagnosis, monitoring and treatment. Furthermore, it was worth mentioning the immune-specific genes in Table 6. These genes were all markers for Natural Killer (NK) cells and T cells, indicating that changes in cytotoxic effects might play a vital role in the HCC immunity. These immune markers might be promising targets for enhancing the efficacy of HCC immunotherapy.

In summary, using an integrated machine learning strategy, mainly composed of mRMR and IFS, we analyzed scRNA-seq data from eight paired HCC tissues and normal liver tissues. A 21-gene-feature consisted of both cancer markers and immune cell markers was established. This feature was regarded as the core to distinguish HCC from normal liver tissue. Considering that the tissue obtained from clinical needle biopsy is often a mixture of tumor parenchyma, stroma and normal liver tissue, this 21-gene-feature might help both the clinical diagnosis of HCC and the identification of biopsy-obtained tissue types. Besides, given that these 21 genes, most of which had not been fully explored in HCC, were expressed in different parenchymal and mesenchymal cells, the following research might focus on their biological function and molecular mechanism in distinct HCC-related cell cluster.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

JS and HZ designed the concept of this research. WL and JL conducted all the studies in this paper. WZ and WG validated and proofread data. WG wrote this paper. XJ and ZY revised this paper.

Funding

This article was supported by the following fundings 1) National Natural Science Foundation of China (No. 81803085); 2) Natural Science Foundation of Hunan Province, China (Nos 2019JJ40414, 2018JJ2612, 2019JJ50285, and 2020JJ4919); 3) Science and Technology Project of Changsha, Hunan, China (No. kq2004146); 4) Scientifific Research Project for Distinguished Young Scientists of Provincial Education Department of Hunan, China (No. 18B460).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.873218/full#supplementary-material

References

Chen, H., Chen, Y., Tsai, R., Chen, M., Chen, S., Xiao, M., et al. (2015). A sensitive and selective magnetic graphene composite-modified polycrystalline-silicon nanowire field-effect transistor for bladder cancer diagnosis. *Biosens. Bioelectron.* 66, 198–207. doi:10.1016/j.bios.2014.11.019

Cheng, H., Garrick, D., and Fernando, R. (2017). Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction. *J. Anim. Sci. Biotechnol.* 8, 38. doi:10.1186/s40104-017-0164-6

Cheng, Q., Li, J., Fan, F., Cao, H., Dai, Z., Wang, Z., et al. (2020). Identification and analysis of glioblastoma biomarkers based on single cell sequencing. *Front. Bioeng. Biotechnol.* 8, 167. doi:10.3389/fbioe.2020.00167

Craig, A., von Felden, J., Garcia-Lezana, T., Sarcognato, S., and Villanueva, A. (2020). Tumour evolution in hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 17 (3), 139–152. doi:10.1038/s41575-019-0229-4

Feng, Y., Shen, Y., Chen, H., Wang, X., Zhang, R., Peng, Y., et al. (2018). Expression profile analysis of long non-coding RNA in acute myeloid leukemia by microarray and bioinformatics. *Cancer Sci.* 109 (2), 340–353. doi:10.1111/cas.13465

Gao, M., Wang, X., Han, D., Lu, E., Zhang, J., Zhang, C., et al. (2020). A six-lncRNA signature for immunophenotype prediction of glioblastoma multiforme. *Front. Genet.* 11, 604655. doi:10.3389/fgene.2020.604655

Haffner, M., Esopi, D., Chaux, A., Gürel, M., Ghosh, S., Vaghasia, A., et al. (2017). AIM1 is an actin-binding protein that suppresses cell migration and micrometastatic dissemination. *Nat. Commun.* 8 (1), 142. doi:10.1038/s41467-017-00084-8

Ho, D., Tsui, Y., Sze, K., Chan, L., Cheung, T., Lee, E., et al. (2019). Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. *Cancer Lett.* 459, 176–185. doi:10.1016/j.canlet.2019.06.002

Li, J., Liang, L., Liu, Y., Luo, Y., Liang, X., Luo, D., et al. (2016). Clinicopathological significance of STAT4 in hepatocellular carcinoma and its effect on cell growth and apoptosis. *Onco. Targets. Ther.* 9, 1721–1734. doi:10.2147/OTT.S100040

Manzo-Merino, J., Lagunas-Martínez, A., Contreras-Ochoa, C., Lizano, M., Castro-Muñoz, L., Calderón-Corona, C., et al. (2018). The human papillomavirus (HPV) E6 oncoprotein regulates CD40 expression via the AT-hook transcription factor AKNA. *Cancers* 10 (12), E521. doi:10.3390/cancers10120521

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* 405 (2), 442–451. doi:10.1016/0005-2795(75)90109-9

McGlynn, K., Petrick, J., and El-Serag, H. (2020). *Epidemiology of hepatocellular carcinoma*. Baltimore, Md: Hepatology.

Morgan, H., Wang, K., Dohopolski, M., Liang, X., Folkert, M., Sher, D., et al. (2021). Exploratory ensemble interpretable model for predicting local failure in head and neck cancer: The additive benefit of CT and intra-treatment cone-beam computed tomography features. *Quant. Imaging Med. Surg.* 11 (12), 4781–4796. doi:10.21037/qims-21-274

Nault, J., and Villanueva, A. (2020). *Biomarkers for hepatobiliary cancers*. Baltimore, Md: Hepatology.

Ouyang, G., Yi, B., Pan, G., and Chen, X. (2020). A robust twelve-gene signature for prognosis prediction of hepatocellular carcinoma. *Cancer Cell. Int.* 20, 207. doi:10.1186/s12935-020-01294-9

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (8), 1226–1238. doi:10.1109/TPAMI.2005.159

Peng, R., Wang, Y., Mao, L., Fang, F., and Guan, H. (2020). Identification of core genes involved in the metastasis of clear cell renal cell carcinoma. *Cancer Manag. Res.* 12, 13437–13449. doi:10.2147/CMAR.S276818

Petrowsky, H., Fritsch, R., Guckenberger, M., De Oliveira, M., Dutkowski, P., and Clavien, P. (2020). Modern therapeutic approaches for the treatment of malignant liver tumours. *Nat. Rev. Gastroenterol. Hepatol.* 17, 755–772. doi:10.1038/s41575-020-0314-8

Ray, M., Wistow, G., Su, Y., Meltzer, P., and Trent, J. (1997). AIM1, a novel non-lens member of the betagamma-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. *Proc. Natl. Acad. Sci. U. S. A.* 94 (7), 3229–3234. doi:10.1073/pnas.94.7.3229

Rebouissou, S., and Nault, J. (2020). Advances in molecular classification and precision oncology in hepatocellular carcinoma. *J. Hepatol.* 72 (2), 215–229. doi:10.1016/j.jhep.2019.08.017

Sangro, B., Melero, I., Wadhawan, S., Finn, R., Abou-Alfa, G., Cheng, A., et al. (2020). Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. *J. Hepatol.* 73 (6), 1460–1469. doi:10.1016/j.jhep.2020.07.026

Sato, Y., Kobayashi, T., Nishiumi, S., Okada, A., Fujita, T., Sanuki, T., et al. (2020). Prospective study using plasma apolipoprotein A2-isoforms to screen for high-risk status of pancreatic cancer. *Cancers* 12 (9), E2625. doi:10.3390/cancers12092625

Siegel, R., Miller, K., and Jemal, A. (2020). Cancer statistics, 2020. *Ca. Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Sokołowska, A., Świerzko, A., Gajek, G., Gołos, A., Michalski, M., Nowicki, M., et al. (2020). Associations of ficolins and mannose-binding lectin with acute myeloid leukaemia in adults. *Sci. Rep.* 10 (1), 10561. doi:10.1038/s41598-020-67516-2

Stegle, O., Teichmann, S., and Marioni, J. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16 (3), 133–145. doi:10.1038/nrg3833

Sun, H., Xu, J., Huang, Q., Huang, M., Li, K., Qu, K., et al. (2018). Reduced CD160 expression contributes to impaired NK-cell function and poor clinical outcomes in patients with HCC. *Cancer Res.* 78 (23), 6581–6593. doi:10.1158/0008-5472.CAN-18-1049

Szala, A., Sawicki, S., Swierzko, A., Szemraj, J., Sniadecki, M., Michalski, M., et al. (2013). Ficolin-2 and ficolin-3 in women with malignant and benign ovarian tumours. *Cancer Immunol. Immunother.* 62 (8), 1411–1419. doi:10.1007/s00262-013-1445-3

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V., et al. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* 5 (3), 516–535. doi:10.1038/nprot.2009.236

Vermaat, J., Gerritse, F., van der Veldt, A., Roessingh, W., Niers, T., Oosting, S., et al. (2012). Validation of serum amyloid α as an independent biomarker for progression-free and overall survival in metastatic renal cell cancer patients. *Eur. Urol.* 62 (4), 685–695. doi:10.1016/j.eururo.2012.01.020

Wang, G., Sun, D., Li, W., and Xin, Y. (2020). AKNA is a potential prognostic biomarker in gastric cancer and function as a tumor suppressor by modulating EMT-related pathways. *Biomed. Res. Int.* 2020, 6726759. doi:10.1155/2020/6726759

Wang, J., Hao, F., Fei, X., and Chen, Y. (2019). SPP1 functions as an enhancer of cell growth in hepatocellular carcinoma targeted by miR-181c. *Am. J. Transl. Res.* 11 (11), 6924

Ye, Y., Zhang, R., Zheng, W., Liu, S., and Zhou, F. (2017). Rifs: A randomly restarted incremental feature selection algorithm. *Sci. Rep.* 7 (1), 13013. doi:10.1038/s41598-017-13259-6

Zhang, Q., He, Y., Luo, N., Patel, S., Han, Y., Gao, R., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell.* 179 (4), 829–845. e20. doi:10.1016/j.cell.2019.10.003

Zhang, W., Kong, H., Gao, X., Dong, Z., Lu, Y., Huang, J., et al. (2020). Immune infiltration-associated serum amyloid A1 predicts favorable prognosis for hepatocellular carcinoma. *World J. Gastroenterol.* 26 (35), 5287–5301. doi:10.3748/wjg.v26.i35.5287

Zheng, Y., Huang, Q., Ding, Z., Liu, T., Xue, C., Sang, X., et al. (2018). Genome-wide DNA methylation analysis identifies candidate epigenetic markers and drivers of hepatocellular carcinoma. *Brief. Bioinform.* 19 (1), 101–108. doi:10.1093/bib/bbw094

Zhou, H., Yin, X., Bai, F., Liu, W., Jiang, S., and Zhao, J. (2020). The role and mechanism of S1PR5 in colon cancer. *Cancer Manag. Res.* 12, 4759–4775. doi:10.2147/CMAR.S239118

Zou, L., Chai, J., Gao, Y., Guan, J., Liu, Q., and Du, J. (2016). Down-regulated PLAC8 promotes hepatocellular carcinoma cell proliferation by enhancing PI3K/Akt/GSK3β/Wnt/β-catenin signaling./β-catenin signaling. *Biomed. Pharmacother. = Biomedecine Pharmacother.* 84, 139–146. doi:10.1016/j.biopha.2016.09.015

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 December 2022
DOI 10.3389/fgene.2022.882004

Check for updates

OPEN ACCESS

EDITED BY
Saurav Mallik,
Harvard University, United States

REVIEWED BY
Jichun Ma,
Jiangsu University Affiliated People's Hospital, China
Yashika Rustagi,
Dana–Farber Cancer Institute, United States
Sudeshna Rakshit,
SRM Institute of Science and Technology, India

*CORRESPONDENCE
Fanping Wang,
wangfanping118@126.com
Mingyong Wang,
wmy118@126.com

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 23 February 2022
ACCEPTED 28 November 2022
PUBLISHED 09 December 2022

CITATION
Ren Z, Huang X, Lv Q, Lei Y, Shi H, Wang F and Wang M (2022), High expression of B4GALT1 is associated with poor prognosis in acute myeloid leukemia.
Front. Genet. 13:882004.
doi: 10.3389/fgene.2022.882004

COPYRIGHT
© 2022 Ren, Huang, Lv, Lei, Shi, Wang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

High expression of B4GALT1 is associated with poor prognosis in acute myeloid leukemia

Zhihong Ren[1], Xiaoyu Huang[2], Qing Lv[3], Yiming Lei[1], Haiqiang Shi[4], Fanping Wang[1]* and Mingyong Wang[1,5]*

[1]School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, China, [2]Henan Key Laboratory of Immunology and Targeted Drug, School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, China, [3]School of Nursing and Health, Hennan University, Kaifeng, China, [4]School of Public Health, Xinxiang Medical University, Xinxiang, China, [5]Xinxiang Key Laboratory of Immunoregulation and Molecular Diagnostics, Xinxiang Medical University, Xinxiang, China

Acute myeloid leukemia is the most prevalent type of leukemia in adults and is prone to relapse and chemoresistance, with a low long-term survival rate. Therefore, the identification of quality biomarkers constitutes an urgent unmet need. High expression of beta-1,4-galactosyltransferase 1 (B4GALT1) has been observed in several cancer types; however, its function in acute myeloid leukemia has rarely been studied. Therefore, our study obtained gene expression data from The Cancer Genome Atlas (TCGA) database to analyze the relationship between B4GALT1 and LAML. We compared the expression of B4GALT1 in LAML and healthy samples using the Wilcoxon rank-sum test. Furthermore, the association between B4GALT1 and survival rates was investigated using Kaplan-Meier analysis and Cox regression. The nomogram obtained by Cox analysis predicts the effect of B4GALT1 on the prognosis. To assess B4GALT1-related genes' enrichment pathway and function and the correlation between B4GALT1 and immune features, GO/KEGG, protein-protein interaction network, and single sample gene set enrichment analysis were used. In addition, B4GALT1-specific siRNAs were used to verify the effect of B4GALT1 on apoptosis. The results showed that B4GALT1 is overexpressed in LAML and has some reference value in the diagnostic and prognostic assessment of LAML. Moreover, functional enrichment showed that B4GALT1 and its 63 associated genes were closely associated with the negative regulation of the apoptotic signaling pathway. Silencing B4GALT1 significantly promoted apoptosis. In addition, B4GALT1 expression was positively correlated with the infiltration levels of macrophages, regulatory T-cell (Tregs), and Th17 cells; in contrast, B4GALT1 expression was negatively correlated with the infiltration levels of T helper cells, Mast cells, and NK cells. In conclusion, our study shows that B4GALT1 may play a vital role in the occurrence of LAML.

Abbreviations: LAML, acute myeloid leukemia; B4GALT1, Beta-1,4-galactosyltransferase one; TCGA, the cancer genome atlas; GTEx, Genotype-Tissue Expression; DEGs, Differentially Expressed Genes; ssGSEA, single-sample Gene Set Enrichment Analysis; GO, Gene Ontology; CC, cellular component; MF, molecular function; BP, biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes; STRING, Search Tool for the Retrieval of Interacting Genes; PPI, Protein-protein interaction; ROC, receiver operating characteristic; AUC, area under the curve.

KEYWORDS
B4GALT1, acute myeloid leukemia, biomarker, immune infiltration, prognosis

Introduction

Acute myeloid leukemia (LAML) is a malignant clonal disease of hematopoietic stem cells, characterized by the proliferation of leukemic cells in the bone marrow and other hematopoietic tissues due to uncontrolled cell cycle, blocked apoptosis, impaired differentiation, and other mechanisms, which in turn infiltrate other tissues and organs (Dohner et al., 2015; Briot et al., 2018; Rudat et al., 2018). Many efforts have been aimed at improving the prognosis of LAML, but since 1970, the standard treatment for most types of LAML has not changed. The primary treatment drug, cytarabine, usually develops drug resistance; one of the reasons is that tumor immune cells evade the drug-induced apoptosis or autophagy. Therefore, most patients will eventually relapse. Recently, the development of molecularly targeted drugs has prolonged the survival of LAML patients and improved clinical outcomes, highlighting the need for new targeted therapies (Estey, 2016). Therefore, exploring more effective biomarkers may provide a new molecular therapeutic strategy for LAML.

Beta-1,4-galactosyltransferase 1 (B4GALT1) is a member of the β-1, 4-galactosyltransferase gene family that encodes a type II membrane-bound glycoprotein that transfers galactose to a similar receptor sugar in the β-1, 4-bonded form (Youakim et al., 1994). Because the promoter region upstream of the B4GALT1 start site contains the ubiquitous transcription factor Sp1, it has long been thought to be a housekeeping gene (Rajput et al., 1996). B4GALT1 plays a vital role in many biological processes, including cell growth, sperm-egg interaction, cell adhesion, migration, and brain development (Nakamura et al., 2001; Nixon et al., 2001; Qasba et al., 2008; Han et al., 2010). Recently, there has been increasing evidence that abnormal expression of B4GALT1 can lead to the development and malignant transformation of various tumors. Tang et al. (2013) found that β1,4-galactosyltransferase one affects the growth and apoptosis of hepatocellular carcinoma cells by regulating EGFR. Estrogen and arachidonic acid have been reported to promote proliferation and cell adhesion of human breast cancer cells by inducing B4GALT1 expression (Choi et al., 2012; Villegas-Comonfort et al., 2012). In glioblastoma, B4GALT1 regulates apoptosis and autophagy, enhancing tumor proliferation, migration, and invasion (Wang et al., 2020). However, the role of B4GALT1 in LAML and its prognostic value have rarely been reported.

Our study aims to use gene expression data obtained from the TCGA database to clarify the potential biological effects between B4GALT1 and LAML patients. Secondly, Kaplan-Meier and prognostic nomogram models were used to explore B4GALT1's potential diagnostic and prognostic value. Moreover, systematically assess the importance of B4GALT1 in LAML through bioinformatics and statistical methods. Our study shows that B4GALT1 plays a critical role in the carcinogenesis and progression of LAML, and it may be used as a biomarker to predict patient survival.

Materials and methods

Data source

Messenger RNA (mRNA) expression data were obtained from the TCGA website (https://cancergenome.nih.gov), which contains gene expression data (HTSeq-FPKM) and clinical information for 151 LAML patients. Next, the Level-3 HTSeq-FPKM data was converted to transcripts per million reads (TPM). Ethical approval and patient informed permission are unnecessary because our study follows the TCGA and GTEx guidelines.

Timer database analysis

TIMER (http://timer.cistrome.org/) (Li et al., 2017) is an online analysis tool with multiple functions. It can explore gene expression levels in multiple tumor para-cancer and tumor tissues in the TCGA dataset.

Prognostic model generation and prediction

A nomogram was built based on independent prognostic indicators derived in multivariate analysis to customize the estimated survival probabilities of 1, 3, and 5 years. The RMS package (version 6.2–0) was employed to generate nomograms that included significant clinical characteristics and calibration plots. The calibration curves were visually assessed by comparing the nomogram-predicted probabilities to the observed rates, with the 45° line representing the best predictive values. The discrimination of the nomogram was determined using the concordance index (C-index), which was generated using a bootstrap technique with 1000 resamples. Moreover, the predictive accuracies of the nomogram and various clinicopathological prognostic factors were compared using the C-index and ROC analyses.

GO and KEGG pathway enrichment analysis

To understand the biological processes and pathways in which B4GALT1-related genes may be involved, we performed Gene Ontology (GO) and KEGG pathway analyses on the top 600 genes associated with B4GALT1 and 63 genes associated with LAML prognosis in LAML using the ClusterProfiler (version 3.14.3) package in R (Yu et al., 2012). Biological process (BP), cellular component (CC), and molecular function (MF) categories were included in the GO analysis.

Protein–protein interaction network

The Search Tool for the Retrieval of Interacting Genes (STRING), a database of known and predicted protein-protein interactions (PPI), was used to construct a PPI network for the differentially expressed genes (DEGs) (Szklarczyk et al., 2019). A composite score threshold of 0.4 was set as the cutoff criterion. The network is imported into the Cytoscape 3.7.0 application for visualization.

Immune cell infiltration estimation with ssGSEA

Immune infiltration analysis of B4GALT1 was conducted by ssGSEA using the GSVA package in R (3.6.3) (Hanzelmann et al., 2013). Marker genes for the 24 immune cell types used in ssGSEA were obtained from Bindea and Gabriela et al. (Bindea et al., 2013). The association between B4GALT1 and the enrichment fraction of each immune cell was investigated using the Spearman correction. The enrichment scores of high- and low-B4GALT1 expression groups were analyzed with the Wilcoxon rank-sum test.

Cell culture and transfection

The Department of Hematology donated the human leukemia cell line KG1a, Zhujiang Hospital, Southern Medical University, and K562 cells were purchased from the Institute of Hematology & Blood Diseases Hospital Chinese Academy of Medical Sciences (Tianjin, China). All cell lines were cultured in RPMI 1640 (Gibco, United States) medium containing 10% fetal bovine serum (Gibco) in a 37°C, 5% CO2 incubator. We established two ADR sublines, KG1A/ADR and K562/ADR, from the corresponding parental cell lines by stepwise exposing cells to doxorubicin at concentrations increasing from 0.02 ug/ml to one ug/ml for 3 months. B4GALT1 small interfering RNA as well as the corresponding control RNA were transfected into logarithmic growth phase cells. Transfection was performed using Lipofectamine 3000 transfection reagent (Invitrogen, United States) according to the manufacturer's protocol. B4GALT1 small interfering RNA as well as the corresponding control RNA were designed by RiboBio.

Quantitative reverse transcription PCR (RT-qPCR)

The total RNA was extracted from cells with Trizol reagents. According to the manufacturer's instructions, cDNA was synthesized from 1 μg of mRNA with a high-capacity cDNA reverse transcription kit. Subsequently, cDNA was amplified by QPCR with the SYBR Premix Ex Taq kit according to the manufacturer's instructions using the ABI7300 Sequence Detection System. PCR conditions were as follows: one cycle at 95°C for 3 min, followed by 40 cycles at 95°C for 30 s and 64°C for 1 min. All assays were performed in triplicate and were calculated based on the $2^{-\Delta\Delta CT}$ method. For qRT-PCR, the following primers were used: human B4GALT1, 5′- CCAGGCGGGAGACACTATATT-3′ (Forward) and 5′-CACCTGTACGCATTATGGTCAT-3′ (Reverse); human β-actin, 5′-AGAGCTACGAGCTGCCTG AC-3′ (Forward) and 5′- AGCACTGTGTTGGCGTACAG-3′ (Reverse).

CCK-8

To assess proliferation, K562 and HL60 cells transfected with B4GALT1 siRNA were seeded in a 96-well plate and cultured in a 37°C incubator. At the same time point for the next 4 days, add 10 μl of Cell Counting Kit-8 to each well and incubate for 3 h. Absorbance (450 nm) was measured by a Microplate Reader (Thermo Scientific).

Western blotting

Total protein was extracted from the cells using ice-cold RIPA lysis buffer and the protein concentration was measured using the BCA Protein Assay Kit. 30 μg total proteins were applied on 10% SDS-PAGE and transferred onto 0.45 mm PVDF membranes (Millipore, United States). The membranes were blocked with 5% non-fat milk for 2 h at room temperature, and then incubated with primary antibodies overnight at 4°C. After washing with TBST buffer three times, the membranes were incubated with secondary HRP-conjugated antibodies for 1 h at room temperature. Capture specific bands using the ECL detection system. The primary antibodies were anti-B4GALT1 (Abcam, ab121326, 1:1000), anti-Bcl-2(Proteintech, 12789-1-AP, 1:1000), anti-Bax (Proteintech, 50,599-2-lg, 1:1000), anti-C-caspase (Proteintech, 19677-1-AP, 1:1000) and anti-β-actin (Proteintech, 20536-1-AP, 1:1000).

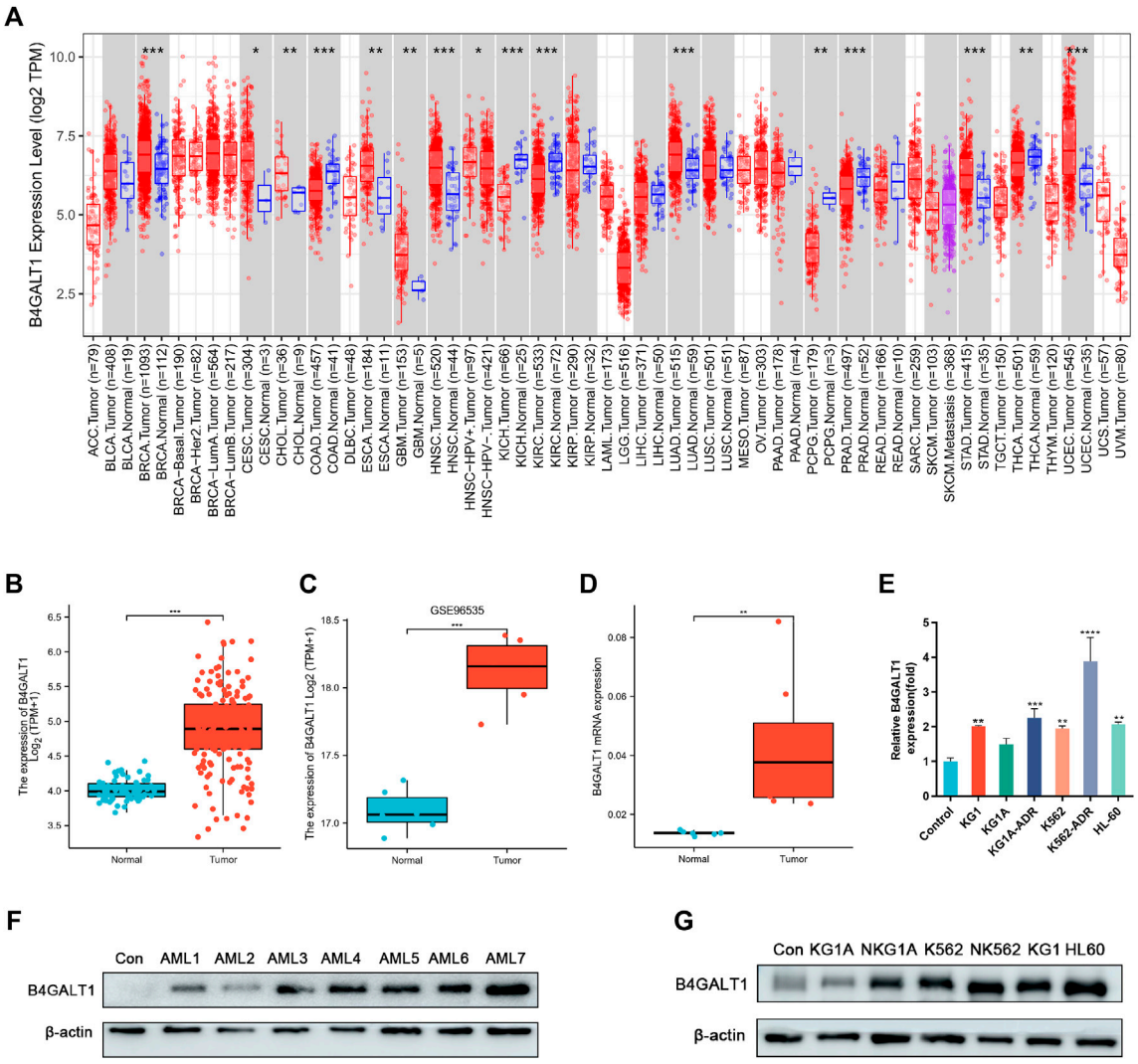


FIGURE 1
The expression levels of B4GALT1 in different human cancers. **(A)** TIMER was used to detect the expression levels of B4GALT1 in different tumors in The Cancer Genome Atlas (TCGA) database. **(B)** Expression level of B4GALT1 in paired normal and LAML samples. **(C)** Expression of B4GALT1 in tumor and normal samples from the GSE96535 dataset in the Gene Expression Omnibus (GEO) database. **(D)** Quantitative mRNA expression of B4GALT1 by qRT-PCR in LAML patient samples ($n = 8$) and controls ($n = 6$). **(E)** Transcriptional level of B4GALT1 in different LAML cell lines. **(F–G)** Western blot detecting the protein expression level of B4GALT1 in patients and different LAML cell lines. Analysis between two groups: Wilcoxon Rank sum test; NS:$p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Cell apoptosis assay

Cell apoptosis was performed with Annexin V-fluorescein isothiocyanate (FITC)/propidium iodide (PI) staining kit (Beyotime Biotechnology, Jiangsu, China). According to the manufacturer's instruction, the cells were seeded in 6-well plates and incubated for 24 h, followed by transfection for 48 h. Cells were collected, centrifuged, washed twice with PBS, resuspended in 400 μl binding buffer and then incubated with 5 μl Annexin V-FITC and 10 μl PI at room temper-ature for 15 min in the dark. Cells were analyzed for fluorescence with flow cytometer (BD Corp).

Statistical analysis

R (Version 3.6.3) was used to perform all statistical analyses and visualizations. The Wilcoxon rank sum test was used for the

TABLE 1 Association between B4GALT1 expression and clinicopathologic features in LAML samples from the TCGA database.

Characteristic	Levels	Low expression of B4GALT1	High expression of B4GALT1	*p*
N		75	76	
Gender, n (%)	Female	38 (25.2%)	30 (19.9%)	0.223
	Male	37 (24.5%)	46 (30.5%)	
Race, n (%)	Asian	0 (0%)	1 (0.7%)	1.000
	Black or African American	6 (4%)	7 (4.7%)	
	White	67 (45%)	68 (45.6%)	
Age, n (%)	<=60	49 (32.5%)	39 (25.8%)	0.114
	>60	26 (17.2%)	37 (24.5%)	
WBC count (x109/L), n (%)	<=20	35 (23.3%)	42 (28%)	0.417
	>20	39 (26%)	34 (22.7%)	
BM blasts (%), n (%)	<=20	24 (15.9%)	36 (23.8%)	0.078
	>20	51 (33.8%)	40 (26.5%)	
PB blasts (%), n (%)	<=70	37 (24.5%)	35 (23.2%)	0.810
	>70	38 (25.2%)	41 (27.2%)	
Cytogenetic risk, n (%)	Favorable	23 (15.4%)	8 (5.4%)	0.003
	Intermediate	38 (25.5%)	44 (29.5%)	
	Poor	12 (8.1%)	24 (16.1%)	
FAB classifications, n (%)	M0	9 (6%)	6 (4%)	<0.001
	M1	15 (10%)	20 (13.3%)	
	M2	27 (18%)	11 (7.3%)	
	M3	14 (9.3%)	1 (0.7%)	
	M4	6 (4%)	23 (15.3%)	
	M5	2 (1.3%)	13 (8.7%)	
	M6	0 (0%)	2 (1.3%)	
	M7	1 (0.7%)	0 (0%)	
Cytogenetics, n (%)	Normal	30 (22.2%)	39 (28.9%)	0.007
	+8	3 (2.2%)	5 (3.7%)	
	del (5)	1 (0.7%)	0 (0%)	
	del (7)	1 (0.7%)	5 (3.7%)	
	inv (16)	3 (2.2%)	5 (3.7%)	
	t (15; 17)	10 (7.4%)	1 (0.7%)	
	t (8; 21)	7 (5.2%)	0 (0%)	
	t (9; 11)	0 (0%)	1 (0.7%)	
	Complex	12 (8.9%)	12 (8.9%)	
FLT3 mutation, n (%)	Negative	47 (32%)	55 (37.4%)	0.259
	Positive	26 (17.7%)	19 (12.9%)	
IDH1 R132 mutation, n (%)	Negative	70 (47%)	66 (44.3%)	0.256
	Positive	4 (2.7%)	9 (6%)	
IDH1 R140 mutation, n (%)	Negative	70 (47%)	67 (45%)	0.152
	Positive	3 (2%)	9 (6%)	
IDH1 R172 mutation, n (%)	Negative	73 (49%)	74 (49.7%)	0.497
	Positive	0 (0%)	2 (1.3%)	
RAS mutation, n (%)	Negative	71 (47.3%)	71 (47.3%)	0.719
	Positive	3 (2%)	5 (3.3%)	
NPM1 mutation, n (%)	Negative	64 (42.7%)	53 (35.3%)	0.023
	Positive	10 (6.7%)	23 (15.3%)	
Age, meidan (IQR)		51 (36.5, 63)	60 (48, 68)	0.006

B4GALT1 expression in unpaired samples. The connection between clinical-pathologic characteristics and B4GALT1 expression was evaluated using the Kruskal-Wallis test and the Wilcoxon signed-rank test. Cox regression analysis and the Kaplan-Meier method evaluated the prognostic factors. The influence of B4GALT1 expression on survival and other clinical information was compared using multivariate Cox analysis. In addition, ROC analysis was performed on the pROC software package (Version 1.17.0.1) to evaluate the effectiveness of B4GALT1 expression in distinguishing LAML from healthy samples. The calculated AUC value ranges from 0.5 to one and implies a discriminative potential of 50–100%. p values <0.05 were considered statistically significant in all tests.

Results

The mRNA expression level of B4GALT1 in pan-cancers and LAML

To explore the expression levels of B4GALT1 in normal and tumor tissues, we analyzed the expression levels of B4GALT1 mRNA in different tumor and normal tissues using the TIMER database. We found that B4GALT1 is highly expressed in eight tumor types and low in six tumor types (Figure 1A). Furthermore, due to the lack of normal control samples of LAML in the TCGA database, we further integrated the healthy tissue data from the GTEx database and performed Wilcoxon rank sum test for the expression of B4GALT1 in LAML. The results showed that B4GALT1 was overexpressed in LAML (Figure 1B).

To further verify the relationship between B4GALT1 and LAML, we validated B4GALT1 expression in LAML and the prognosis by GSE96535 and GSE71014 in the GEO (Gene Expression Omnibus) database. The results were as predicted in the TCGA database, B4GALT1 was highly expressed in LAML, and high B4GALT1 expression was associated with poor prognosis in LAML patients (Figure 1C, Supplementary Figure S1). At the same time, we collected bone marrow specimens from six healthy individuals and eight LAML patients and similarly validated that B4GALT1 was highly expressed in LAML (Figure 1D, Figure 1F). In addition, the expression levels of B4GALT1 in LAML cell lines were detected by QPCR and Western blotting. The results showed that the expression levels of B4GALT1 were significantly higher in LAML cells and drug-resistant cells than in the control (Figure 1E, Figure 1G).

Association of B4GALT1 expression with clinicopathological status

The clinicopathological characteristics of all LAML patients are demonstrated, as shown in Table 1. The study included 68 women and 83 men. According to the median value of B4GALT1 expression (log2 (TPM+1)), B4GALT1 was low in 75 LAML patients and high in the remaining 76 patients. Correlation analysis suggested that B4GALT1 expression was significantly correlated with Cytogenetic risk ($p = 0.003$), FAB classification ($p < 0.001$), NPM1 mutation ($p = 0.023$), and Age ($p = 0.006$). In addition, the incidence of t (8; 21) and t (15; 17) was lower in the B4GALT1 high expression group ($p < 0.007$).

Logistic regression analysis was used to validate further the relationship between B4GALT1 expression as a categorical dependent variable and clinicopathological features of LAML. The results showed that overexpression of B4GALT1 was significantly and positively associated with unfavorable cytogenetic risk (dominance ratio [OR], 0.256; $p = 0.002$) and NPM1 mutation (OR, 2.777; $p = 0.015$) (Table 2). Besides, ROC curve analysis was used to assess the potential value of B4GALT1 in distinguishing LAML patients from healthy individuals, with an AUC of 0.915 indicating that B4GALT1 showed biomarker potential (Figure 2A). Furthermore, Wilcoxon Rank SUM test results showed that high expression of B4GALT1 was significantly related to FAB classification (non-M3 type; $p < 0.001$), cytogenetic risk (intermediate/poor; $p < 0.001$), and mutated nucleophosmin one gene (NPM1) mutation (positive; $p = 0.003$) (Figures 2B–D).

High expression of B4GALT1 is an independent prognostic factor in LAML patients

A Kaplan-Meier analysis was performed to determine whether B4GALT1 expression affects patient survival. The results show that, compared with the low B4GALT1 expression group, high B4GALT1 expression was more strongly related to the poor prognosis of LAML (HR = 1.68, $p = 0.019$) (Figure 3A). Subgroup analysis showed that a high B4GALT1 expression was significantly correlated with poor prognosis in LAML in the following cases: male (HR = 2.14, $p = 0.011$), WBC count ($\leq 20 \times 10^9$/L) (HR = 2.16, $p = 0.012$), BM blasts ≤20% (HR = 2.72, $p = 0.006$), PB blasts ≤70% (HR = 2.15, $p = 0.023$), FLT3 mutation-negative (HR = 2.06, $p = 0.006$), IDH1 R132 mutation-negative (HR = 1.92, $p = 0.004$), R140 mutation-negative (HR = 1.80, $p = 0.011$), R172 mutation-negative (HR = 1.64, $p = 0.025$), RAS mutation-negative (HR = 1.61, $p = 0.034$), and NPM1 mutation-negative (HR = 2.15, $p = 0.003$) (Figures 3B–K).

In addition, we used univariate Cox analysis to evaluate the link between B4GALT1 expression and clinical features of LAML patients to understand better the importance and mechanisms of B4GALT1 expression in LAML. The results show that B4GALT1 (high-vs. low-, $p = 0.019$), cytogenetic risk (favorable vs. poor and intermediate, $p < 0.001$) and age (>60 vs. ≤ 60, $p < 0.001$) (Table 3) are Predictors of poor overall survival (OS).

TABLE 2 The relationship between the clinicopathological factors of LAML and B4GALT1 expression was determined by the logistic analysis.

Characteristics	Total(N)	Odds ratio (OR)	*p*-Value
WBC count (x109/L) (>20 vs. <=20)	150	0.726 (0.381–1.379)	0.330
PB blasts (%) (>70 vs. <=70)	151	1.141 (0.602–2.166)	0.687
BM blasts (%) (>20 vs. <=20)	151	0.523 (0.267–1.009)	0.055
Cytogenetic risk (Favorable vs. Poor& Intermediate)	149	0.256 (0.100–0.597)	0.002
FLT3 mutation (Positive vs. Negative)	147	0.624 (0.305–1.263)	0.193
IDH1 R132 mutation (Positive vs. Negative)	149	2.386 (0.739–9.149)	0.164
IDH1 R140 mutation (Positive vs. Negative)	149	3.134 (0.892–14.580)	0.097
RAS mutation (Positive vs. Negative)	150	1.667 (0.394–8.372)	0.495
NPM1 mutation (Positive vs. Negative)	150	2.777 (1.243–6.590)	0.015

Construction and validation of a nomogram based on the B4GALT1

To predict the prognosis of LAML patients more accurately, the RMS R package is used to construct a nomogram based on the results of Cox regression analysis to predict the probability of patient survival at 1, 3, and 5 years (Figure 4A). All patients' observation findings were compatible with the nomogram calibration curve's prediction results (Figure 4B).

Differential genes correlated with B4GALT1 in LAML

To elucidate the role and significance of B4GALT1 expression in LAML, we examined the differentially expressed genes and gene co-expression pattern of B4GALT1 in LAML. The volcano plot showed 756 up-regulated genes and 553 down-regulated genes, which were statistically significant between the two groups (|log fold change (logFC)| > 1.5, *p*.adj <0.05) (Supplementary Figure S2). Heatmap shows the top 25 genes positively correlated with B4GALT1 and the bottom 25 negatively correlated genes (Figure 5A). We used the ClusterProfiler package to identify B4GALT1-related genes' biological processes and pathways (top 600). BP analysis showed that these related genes were mainly enriched in negative regulation of apoptotic signaling pathway, positive regulation of interleukin-6 production, and positive regulation of cell growth (Figure 5B). The results of KEGG analysis revealed that these related genes were mainly involved in the NOD-like receptor signaling pathway, Necroptosis, and Toll-like receptor signaling pathway (Figure 5C).

To identify genes with the same regulatory direction in high B4GALT1 and non-surviving patients, we intersected the 600 highest B4GALT1-related genes with 300 survival-related genes in LAML, detecting a total of 63 genes associated with B4GALT1 and LAML survival-related genes (Figure 6A). Further GO functional enrichment of these 63 genes showed that DEGs were significantly enriched in negative regulation of apoptotic signaling pathway, activation of MAPK activity, and regulation of I-kappaB kinase/NF-kappaB signaling (Figure 6B).

To further explore the interactions between these 63 proteins, we constructed a PPI network for DEG using the STRING database. The most important 10 central nodes are identified by the MCC algorithm using cytoHubba. We found more robust interaction networks between these proteins and that these genes were explicitly associated with negative regulation of the apoptotic signaling pathway (Figure 7A, Supplementary Figure S3). Gene co-expression analysis showed that most of the proteins in the network showed a strong positive correlation (Figure 7B).

B4GALT1 promoted proliferation, suppressed apoptosis of LAML cells

To further investigate the regulatory role of B4GALT1 on apoptosis, we used B4GALT1-targeted siRNA sequences to interfere with the expression of B4GALT1. The results showed that the expression of B4GALT1 was significantly decreased (Figures 8A–D). Then, CCK-8 was applied to determine the effect of B4GALT1 on cell viability. The results showed that knockdown of B4GALT1 significantly inhibited the viability of K562 and HL60 cells (Figures 8E,F). Subsequently, apoptosis was determined by flow cytometry and western blot. We found that knockdown of B4GALT1 in K562 and HL60 cells promoted their apoptosis (Figures 8G,H). In conclusion, these results suggest that B4GALT1 affects the proliferation of LAML cells by regulating apoptosis.

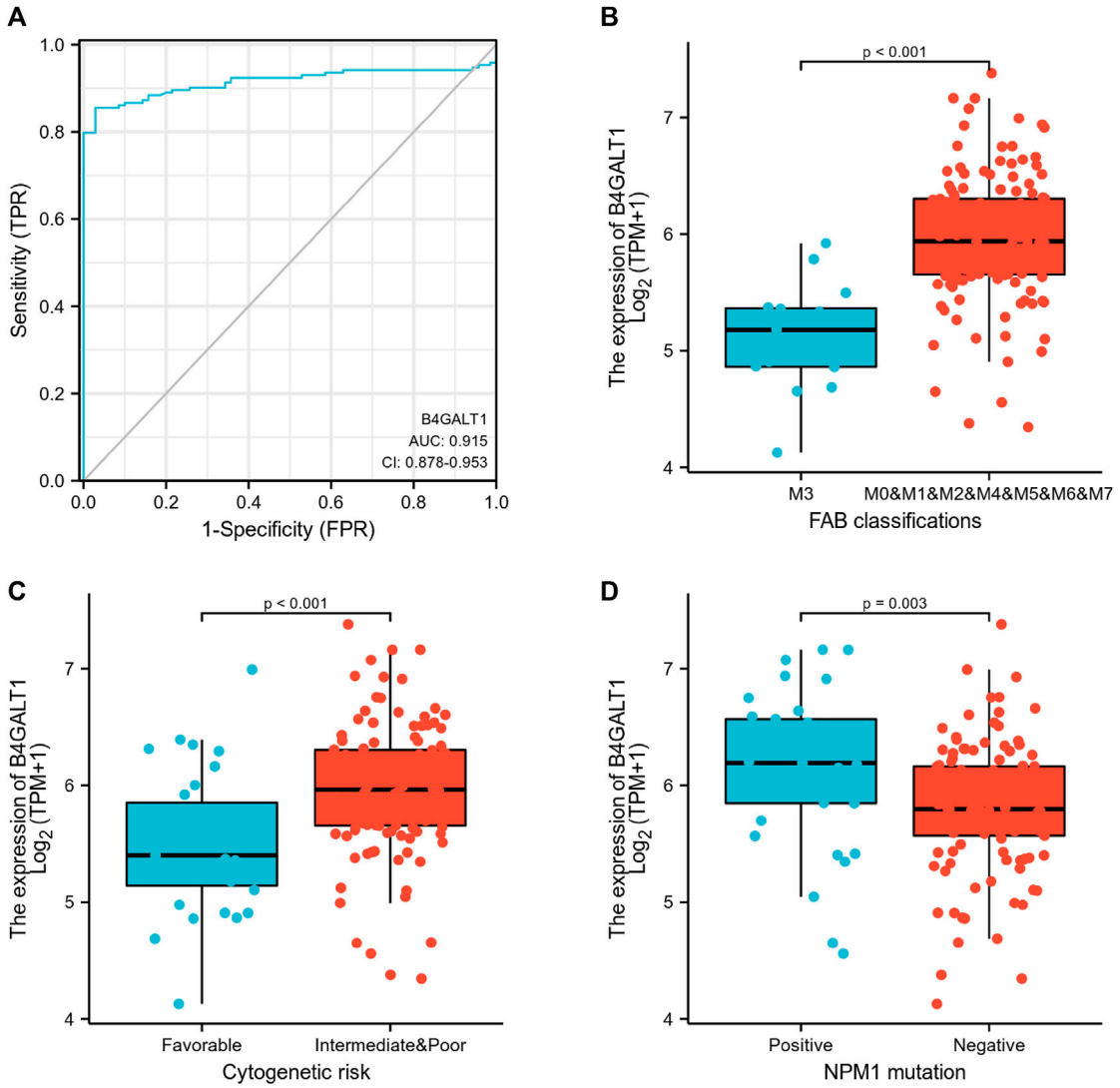


FIGURE 2
Association between B4GALT1 expression and clinical features. **(A)** ROC analysis of B4GALT1 shows promising discrimination power between tumor and normal tissues. **(B)** FAB classification; **(C)** cytogenetics risk; **(D)** NPM1 mutation.

Correlation between B4GALT1 expression and immune infiltration

Based on Spearman correlation analysis, the expression level of B4GALT1 (TPM) is correlated with the level of immune cell infiltration (generated by ssGSEA). The results showed that the infiltration levels of Dendritic cells (DC), iDC, Macrophages, Neutrophils, NK CD56dim cells, Tem, Tgd, Th1 cells, Th17 cells, and Treg were significantly increased in patients with high B4GALT1 expression. In contrast, Mast cells, NK cells, and T helper cells were significantly decreased in patients with high expression of B4GALT1. (Figures 9A,B).

We further analyzed the correlation between the expression level of B4GALT1 and immune infiltration in LAML, and the results showed that the expression level of B4GALT1 was positively correlated with the infiltrating levels of Macrophages (r = 0.557, $p < 0.001$) (Figure 9C), Tregs (r = 0.234, $p = 0.004$) (Figure 9D), and Th17 cells (r = 0.354, $p < 0.001$) (Figure 9E). In contrast, the expression level of B4GALT1 was negatively correlated with the infiltrating levels of T helper cells (r = -0.327, $p < 0.001$) (Figure 9F), Mast cells (r = -0.327, $p = 0.005$) (Figure 9G), and NK cells (r = -0.234, $p = 0.004$) (Figure 9H).

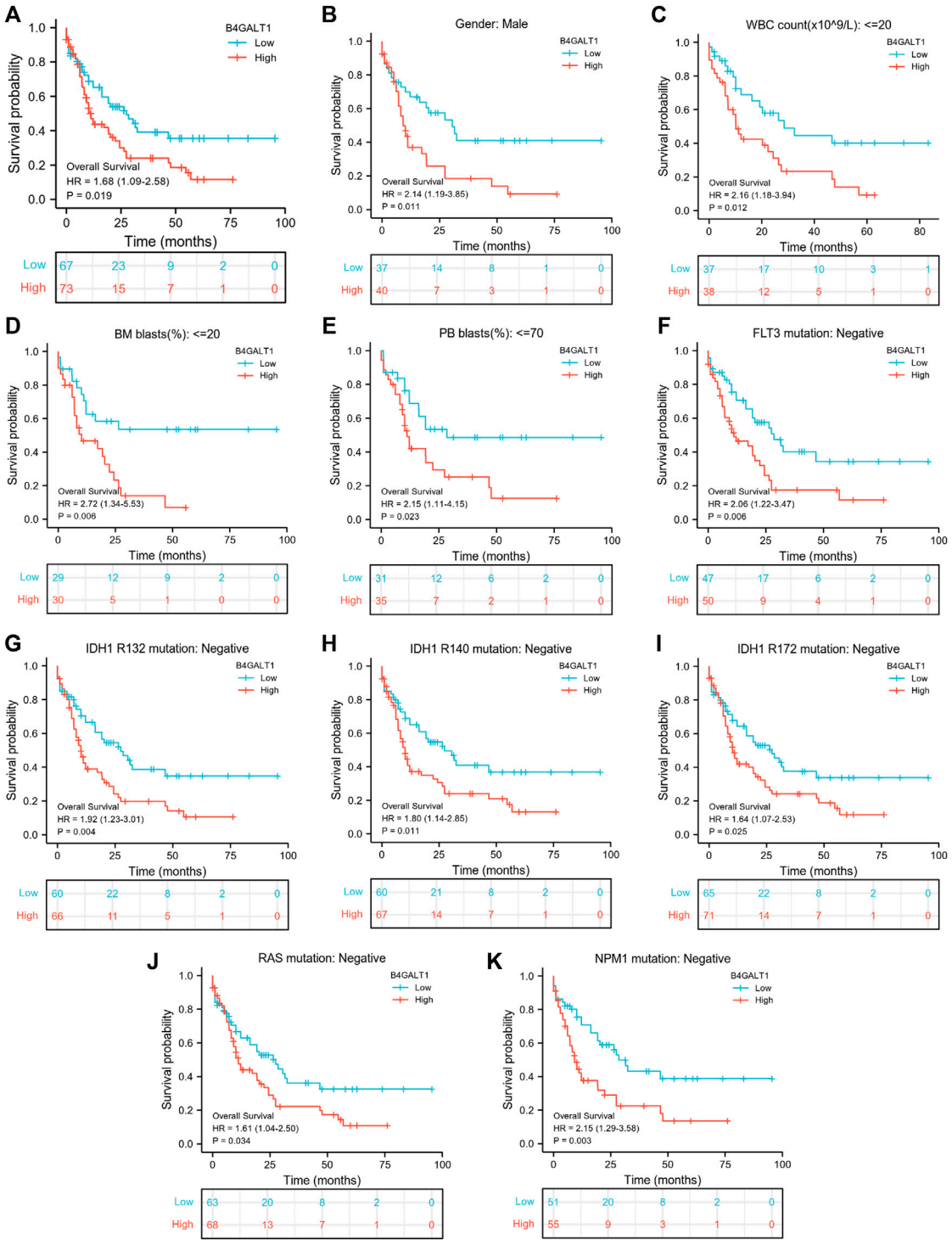


FIGURE 3
Kaplan-Meier survival curves compare B4GALT1's high and low expression in LAML patients. **(A)** Overall survival. **(B)** Kaplan-Meier curves in LAML patients with Male. **(C)** Kaplan-Meier curves in LAML patients with WBC counts <= 20 × 10^9/L. **(D)** Kaplan-Meier curves in LAML patients with BM blasts <20%. **(E)** Kaplan-Meier curves in LAML patients with PB blasts ≤70%. **(F–K)** Kaplan-Meier curves in subgroups with FLT3 mutation-negative, IDH1 R132 mutation-negative, IDH1 R140 mutation-negative, R172 mutation-negative, RAS mutation-negative, and NPM1 mutation-negative in LAML patients.

TABLE 3 Univariate and multivariate Cox's regression analysis of factors associated with OS in LAML.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p*-Value	Hazard ratio (95% CI)	*p*-Value
WBC count (x109/L)	139				
<=20	75	References			
>20	64	1.161 (0.760–1.772)	0.490		
BM blasts (%)	140				
<=20	59	References			
>20	81	1.165 (0.758–1.790)	0.486		
PB blasts (%)	140				
<=70	66	References			
>70	74	1.230 (0.806–1.878)	0.338		
Cytogenetic risk	138				
Poor&Intermediate	107	References			
Favorable	31	0.312 (0.160–0.606)	<0.001	0.434 (0.216–0.873)	0.019
Age	140				
<=60	79	References			
>60	61	3.333 (2.164–5.134)	<0.001	2.826 (1.800–4.439)	<0.001
Race	138				
Asian&Black or African American	11	References			
White	127	1.200 (0.485–2.966)	0.693		
Gender	140				
Female	63	References			
Male	77	1.030 (0.674–1.572)	0.892		
FLT3 mutation	136				
Negative	97	References			
Positive	39	1.271 (0.801–2.016)	0.309		
IDH1 R132 mutation	138				
Negative	126	References			
Positive	12	0.588 (0.238–1.452)	0.249		
IDH1 R140 mutation	138				
Negative	127	References			
Positive	11	1.131 (0.565–2.264)	0.727		
IDH1 R172 mutation	138				
Negative	136	References			
Positive	2	0.610 (0.085–4.385)	0.623		
RAS mutation	139				
Negative	131	References			
Positive	8	0.643 (0.235–1.760)	0.390		
NPM1 mutation	139				
Negative	106	References			
Positive	33	1.137 (0.706–1.832)	0.596		
B4GALT1	140				
Low	67	References			
High	73	1.677 (1.090–2.579)	0.019	1.267 (0.802–2.002)	0.311

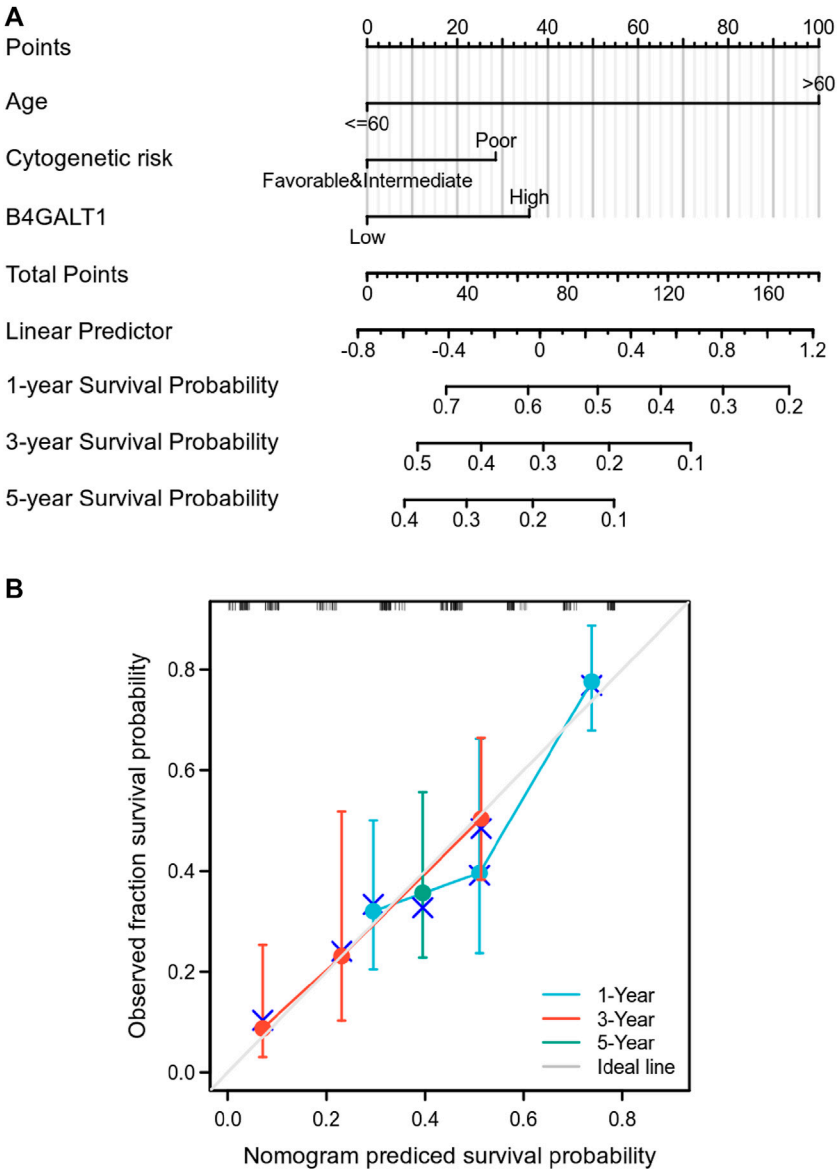


FIGURE 4
Relationship between B4GALT1 and other clinical factors with overall survival (OS). **(A)** Nomogram for predicting the probability of 1, 3, and 5-year OS for LAML patients. **(B)** Calibration plot of the nomogram for predicting the OS likelihood.

Discussion

As an essential component of cells, glycan chains are widely involved in cell adhesion, growth, differentiation, proliferation, apoptosis, and signal transduction through protein glycosylation modifications. Disorders of glycosylation can contribute to the development and progression of many tumors. It was shown that the expression and activity of Beta-1,4-galactosyltransferase one correlated with the malignancy of various tumors, including hepatocellular carcinoma, breast cancer, glioblastoma, and lung cancer (Zhu et al., 2005; Choi et al., 2012; Villegas-Comonfort et al., 2012; Tang et al., 2013; Wang et al., 2020). However, there are fewer studies on B4GALT1 in LAML. Our study used gene expression data from TCGA for bioinformatics analysis to evaluate B4GALT1's potential mechanism and role in LAML. We have observed that the high expression of B4GALT1 in LAML is associated with poorer OS.

Meanwhile, high B4GALT1 expression was associated with a low incidence of favorable genetic abnormalities t (8; 21) and t (15; 17). Moreover, B4GALT1 has some reference value distinguishing LAML patients from healthy individuals. Therefore, B4GALT1 may become a new biomarker for

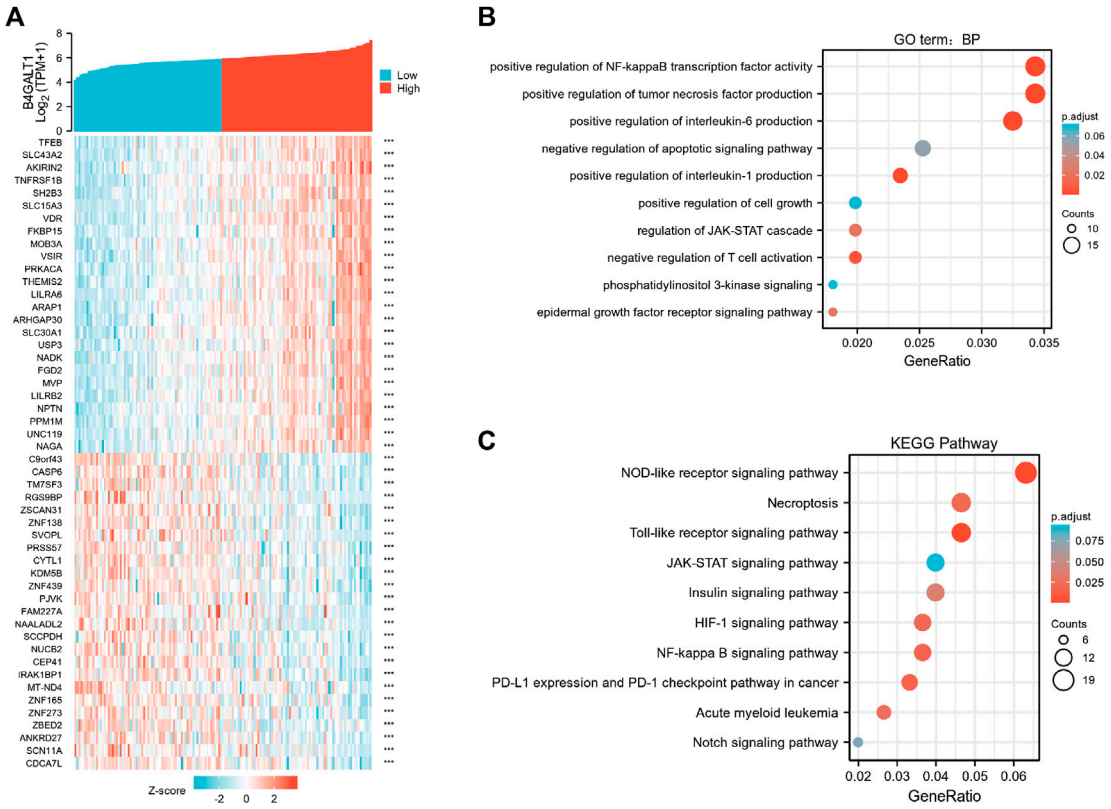


FIGURE 5
Functional clustering analysis of B4GALT1-related genes. **(A)** Heatmap showing the top 50 genes in acute myeloid leukemia (LAML) that were positively and negatively related to B4GALT1. Red represents positively related genes, and blue represents negatively related genes. **(B)** Enriched GO terms in the "biological process" category; **(C)** KEGG pathway annotations. The B4GALT1-axis represented the proportion of DEGs, and the *Y*-axis represented different categories. The different colors indicate different *p* values, and the different sizes represent gene numbers.

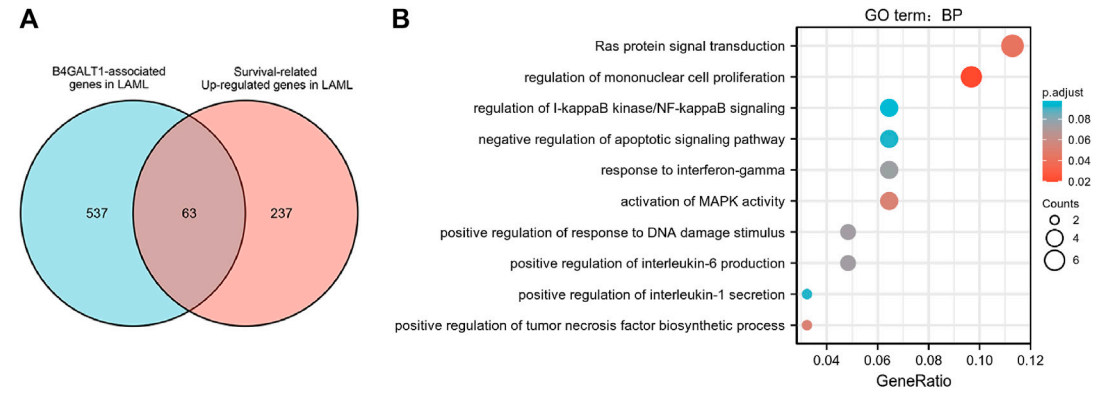


FIGURE 6
Functional clustering analysis of B4GALT1-related genes. **(A)** Venn diagram of B4GALT1-related genes and survival-related upregulated genes in LAML. **(B)** GO enrichment analysis of B4GALT1-related genes and LAML survival-related genes in LAML.

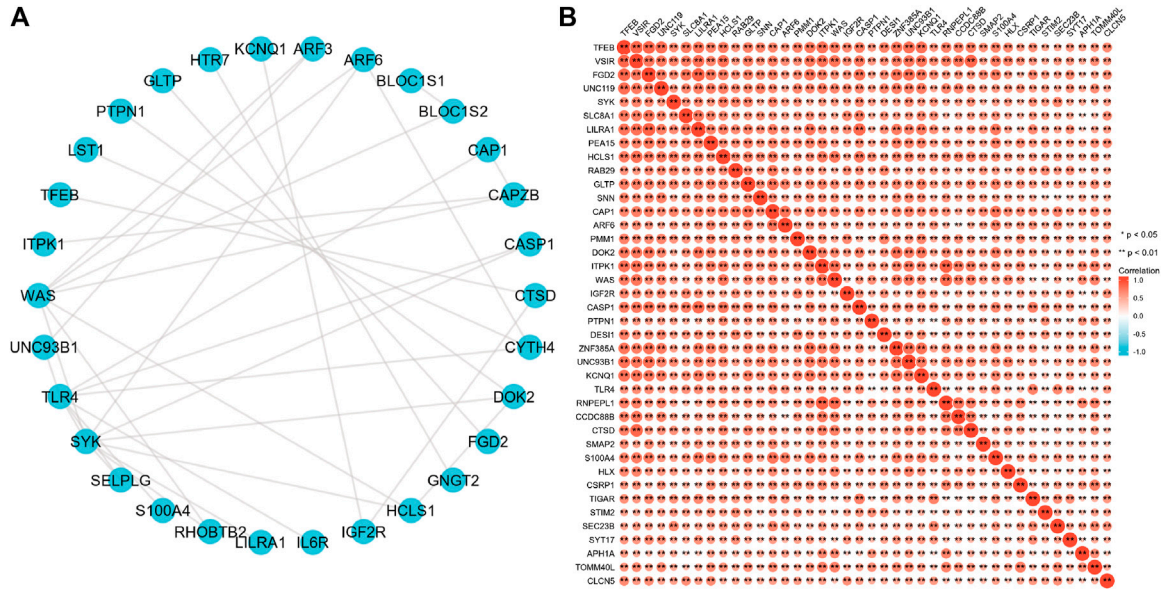


FIGURE 7
B4GALT1-associated gene interaction network **(A)** and gene co-expression matrix **(B)**.

LAML patients. Notably, to explain the potential molecular mechanisms by which B4GALT1 affects LAML prognosis, we performed GO and KEGG pathway analysis on 63 genes associated with B4GALT1. The results showed that B4GALT1 was significantly associated with the negative regulation of the apoptotic signaling pathway. These findings suggest that B4GALT1 is not only a new adverse prognostic factor but can also influence the pathways involved in LAML development and progression and may be a promising therapeutic target.

Chemotherapy is currently the preferred clinical treatment option for LAML. The widespread use of standard therapies can achieve complete remission (CR) in approximately 60% of patients at initial treatment. However, approximately 76% of patients eventually relapse or die, suggesting that treatment resistance has become a significant barrier to LAML prognosis (Tallman et al., 2005; Dohner et al., 2010; Dohner et al., 2015; Komanduri and Levine, 2016). Studies have shown that an important cause of tumor cell resistance to conventional anticancer drugs and some targeted therapies is an evasion of apoptosis (Aleksakhina et al., 2019). Previous studies have focused on BCL-2 family members in LAML (Sillar and Enjeti, 2019). However, resistance to drugs based on the BCL-2 family as targets has emerged during clinical use, which is partly attributed to the functional compensation of non-targeted anti-apoptotic molecules (Mazumder et al., 2012; Pan et al., 2015; Pan et al., 2017; Guieze et al., 2019; Nechiporuk et al., 2019). Discovering new candidates is critical to clinical treatment. Our study found that B4GALT1 expression was elevated in drug-resistant LAML cell lines. To further analyze the regulatory mechanisms of B4GALT1 affecting LAML progression and prognosis, we identified 63 overlapping genes as our hub genes and performed GO and KEGG functional enrichment analysis.

Interestingly, the negative regulation of the apoptotic signaling pathway was significantly enriched in the B4GALT1 high expression group. We further silenced B4GALT1 and showed significant apoptosis in LAML cell lines after B4GALT1 silencing. In addition, we found that B4GALT1 was significantly elevated in patients with non-M3 leukemia and cytogenetic risk (intermediate/poor). The Cox regression analysis suggests that B4GALT1 may have the ability to be an independent predictor of poor prognosis in LAML. Thus, our data suggest that B4GALT1 may act as a non-targeted anti-apoptotic molecule to regulate LAML progression and drug resistance by negatively regulating apoptosis, but further experimental validation is needed to demonstrate the effect of B4GALT1 on LAML development.

More and more essential and clinical studies have shown that the tumor microenvironment (TME) plays a crucial role in tumors' occurrence, development, and metastasis. The latest study report shows that B4GALT1 is a newly discovered PD-L1 glycosyltransferase, and in triple-negative breast cancer, RBMS1 can bind to the 3′-UTR of B4GALT1 to stabilize its mRNA. Subsequently, highly expressed B4GALT1 can promote PD-L1 glycosylation and increase PD-L1 protein stability, weakening the body's antitumor immune response and promoting tumor immune escape (Zhang et al., 2022).

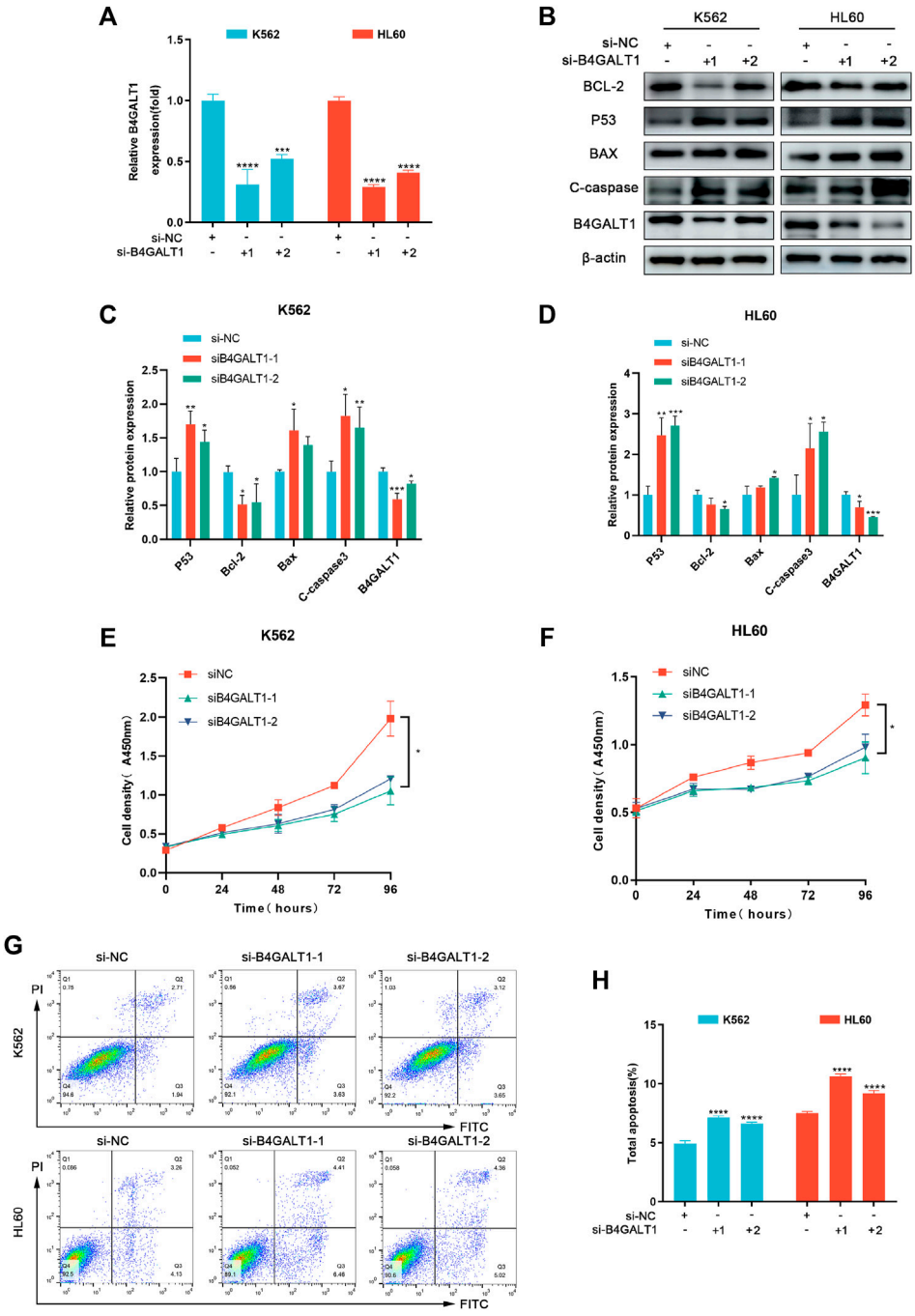


FIGURE 8
B4GALT1 suppressed apoptosis of LAML cells. **(A)** The expression of B4GALT1 was determined by qRT-PCR in B4GALT1 knockdown LAML cells. **(B–D)** The apoptosis of LAML cells was measured by western blot. **(E–F)** CCK-8 was used to detect the proliferation ability of K562 cells **(E)** and HL60 cells **(F)**. **(G–H)** The apoptosis of LAML cells was measured by staining with Annexin V/PI. NS:$p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

However, in LAML, the relationship between B4GALT1 and immune infiltration has not been reported. Therefore, we investigated the correlation between B4GALT1 and immune infiltration in LAML. We found that the abundance of Macrophages, Treg cells, and Th17 cells increased significantly in the high expression group of B4GALT1.

In contrast, the abundance of T helper cells, Mast cells, and NK cells decreased significantly. Tumor-associated macrophages

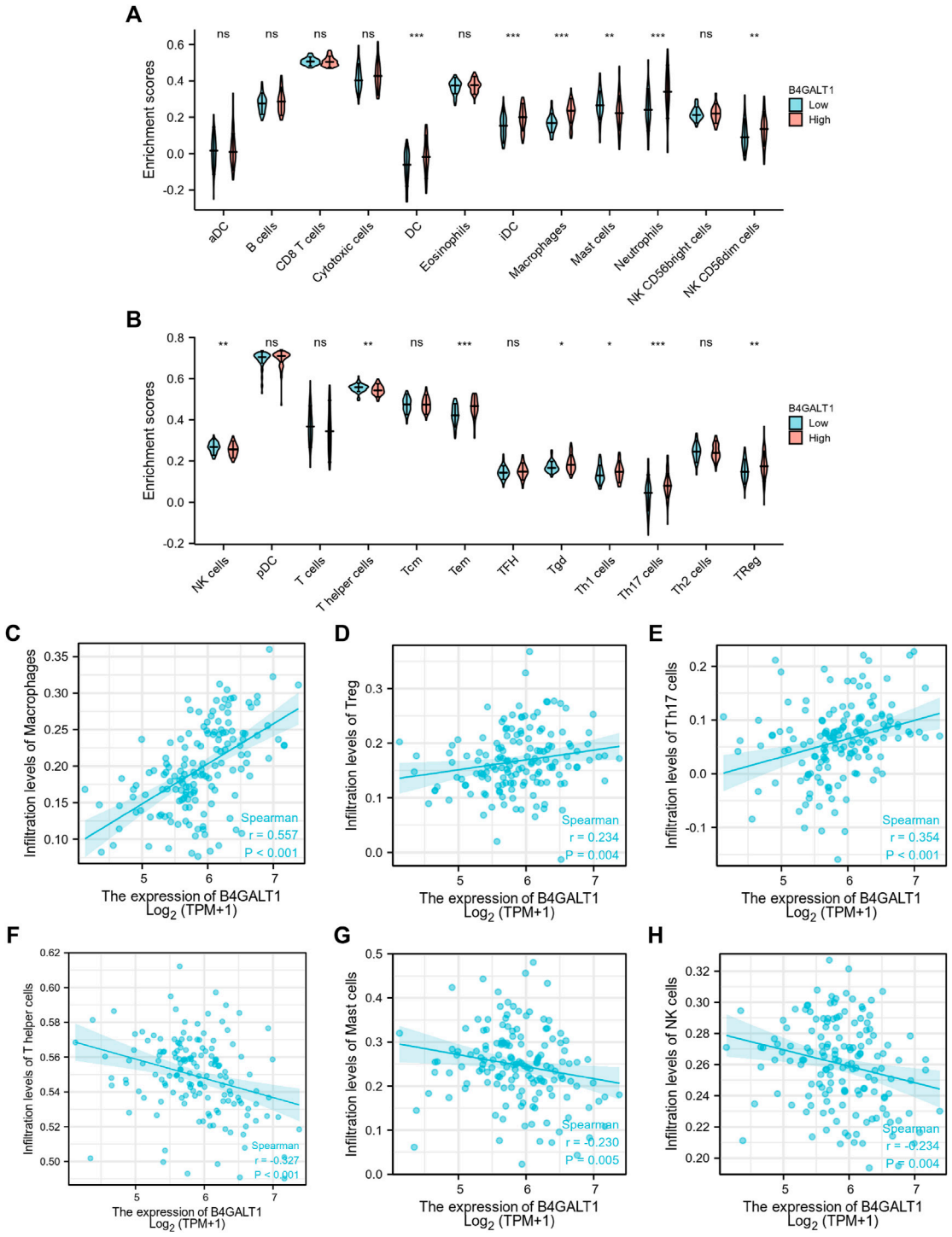


FIGURE 9
Correlation between the infiltration level of 24 immune cells and the expression level of B4GALT1. **(A-B)** Differential distribution of immune cells in patients with high B4GALT1 expression and low B4GALT1 expression. **(C-H)** Correlation between the expression level of B4GALT1 and immune infiltration in acute myeloid leukemia (LAML): **(C)** Macrophages, **(D)** regulatory T-cell, **(E)** Th17 cells, **(F)** T helper cells, **(G)** Mast cells, **(H)** NK cells. NS: $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

(TAMs), one of the most critical subpopulations of the many tumor-infiltrating immune cells, play a crucial role in the interaction between the immune system and tumor cells. In most human tumors, tumor-associated macrophage infiltration and upregulation of their associated gene expression severely affect tumor prognosis and treatment outcome (Hu et al., 2016). Studies have reported that TAMs promote tumor angiogenesis and hematogenous cell metastasis by secreting many pro-angiogenic factors, such as vascular endothelial growth factor VEGF (Lin et al., 2006; DeNardo et al., 2009). TAM can also produce immunosuppressive factors such as IL-10, TGFβ, and PGE2, which can promote tumorigenesis and development by suppressing the antitumor immune response, among which IL-10 can significantly reduce the effect of antitumor therapy by suppressing the antitumor immune response of chemotherapy drugs (Ruffell et al., 2014). Large amounts of TAMs secrete pleiotrophin (PTN) protein in glioblastoma tumors. Glioblastoma tumor stem cells (CSCs) have large amounts of its receptor, PTPRZ1, which activate a series of signaling pathways that produce CSCs and maintain their malignant behavior, promoting tumor growth and progression and leading to increased mortality in patients (Shi et al., 2017).

Moreover, Treg cells and Th17 cells also play an essential role in the immune escape of tumors. Treg cells are one of the most critical immunosuppressive components of the bone marrow environment (Ustun et al., 2011). Studies have shown that the proportion of Tregs in peripheral blood (PB) and bone marrow (BM) in patients with LAML increases, leading to a poor prognosis for patients (Szczepanski et al., 2009; Yang and Xu, 2013). Recent studies have shown that Tregs, an essential component of the acute myeloid leukemia (LAML) microenvironment, secretes IL-10 that interacts with IL-10 R on the surface of cancer cells and further activates the PI3K/AKT signaling pathway, promoting stemness of leukemic stem cells and the development of leukemia (B4GALT1u et al., 2021). In addition, it has been reported that in LAML, Th17 cells secrete TNF-alpha and activate $CD4^{+}CD25^{+}$ Treg cells through TNFR2 receptors, leading to increased expression of tumor necrosis factor-alpha and Treg cells, which in turn leads to an immunosuppressed state in patients (Wang et al., 2018).

Furthermore, natural killer cells (NK) are the first line of defense of the body's immune system. Unlike T and B-cell, they can rapidly kill tumor- or virus-infected cells without antigenic stimulation (Mandal and Viswanathan, 2015). Therefore, based on our findings and the above-reported studies, how B4GALT1 mediates immune escape in LAML remains further investigated.

Although our study can bring new insights into the connection between B4GALT1 and LAML, there are some limitations. First, we only analyzed the data in the database and did not collect a large sample ourselves to validate our conclusions, which may lead to biased results. Therefore, to improve the reliability of our results, we need to collect a wider sample of data to validate our conclusions. Second, the amount of our experimental data was too small, resulting in insufficient validation of our conclusions. We need further experiments to validate the biological functions of B4GALT1 *in vitro* and *in vivo*. In addition, although our data suggest that B4GALT1 may be involved in antitumor immunosuppression in LAML, further experiments are needed to verify the relationship between the two. Based on the limitations of the above studies, our laboratory has started to develop several programs.

Conclusion

Our study found a significant upregulation of B4GALT1 expression in LAML, which was also associated with a poor prognosis. High expression of B4GALT1 may mediate chemotherapy resistance in LAML. In addition, B4GALT1 may influence LAML progression by regulating apoptosis and immune infiltration. Therefore, B4GALT1 is expected to be a potential biomarker for LAML diagnosis and prognosis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

ZR, QL, and XH were responsible for the design of the study, data acquisition, and analysis, as well as drafting the manuscript. YL, FW, and MW participated in data acquisition, analysis, and interpretation. All authors read and approved the final manuscript.

Funding

This work was supported by the Henan University Science and Technology Innovation Team Support Program (20IRTSTHN030), Outstanding Youth Project of Henan Natural Science Foundation (212,300,410,013).

Acknowledgments

Special thanks are given to MW for his dedicated revision of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.882004/full#supplementary-material

References

Aleksakhina, S. N., Kashyap, A., and Imyanitov, E. N. (2019). Mechanisms of acquired tumor drug resistance. *Biochim. Biophys. Acta. Rev. Cancer* 1872 (2), 188310. doi:10.1016/j.bbcan.2019.188310

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Briot, T., Roger, E., Thepot, S., and Lagarce, F. (2018). Advances in treatment formulations for acute myeloid leukemia. *Drug Discov. Today* 23 (12), 1936–1949. doi:10.1016/j.drudis.2018.05.040

Choi, H. J., Chung, T. W., Kim, C. H., Jeong, H. S., Joo, M., Youn, B., et al. (2012). Estrogen induced beta-1, 4-galactosyltransferase 1 expression regulates proliferation of human breast cancer MCF-7 cells. *Biochem. Biophys. Res. Commun.* 426 (4), 620–625. doi:10.1016/j.bbrc.2012.08.140

DeNardo, D. G., Barreto, J. B., Andreu, P., Vasquez, L., Tawfik, D., Kolhatkar, N., et al. (2009). CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages. *Cancer Cell* 16 (2), 91–102. doi:10.1016/j.ccr.2009.06.018

Dohner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Buchner, T., Burnett, A. K., et al. (2010). Diagnosis and management of acute myeloid leukemia in adults: Recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 115 (3), 453–474. doi:10.1182/blood-2009-07-235358

Dohner, H., Weisdorf, D. J., and Bloomfield, C. D. (2015). Acute myeloid leukemia. *N. Engl. J. Med.* 373 (12), 1136–1152. doi:10.1056/NEJMra1406184

Estey, E. (2016). Acute myeloid leukemia - many diseases, many treatments. *N. Engl. J. Med.* 375 (21), 2094–2095. doi:10.1056/NEJMe1611424

Guieze, R., Liu, V. M., Rosebrock, D., Jourdain, A. A., Hernandez-Sanchez, M., Martinez Zurita, A., et al. (2019). Mitochondrial reprogramming underlies resistance to BCL-2 inhibition in lymphoid malignancies. *Cancer Cell* 36 (4), 369–384. doi:10.1016/j.ccell.2019.08.005

Han, Y., Zhou, X., Ji, Y., Shen, A., Sun, X., Hu, Y., et al. (2010). Expression of beta-1, 4-galactosyltransferase-I affects cellular adhesion in human peripheral blood CD4+ T cells. *Cell. Immunol.* 262 (1), 11–17. doi:10.1016/j.cellimm.2009.08.004

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hu, Y., He, M. Y., Zhu, L. F., Yang, C. C., Zhou, M. L., Wang, Q., et al. (2016). Tumor-associated macrophages correlate with the clinicopathological features and poor outcomes via inducing epithelial to mesenchymal transition in oral squamous cell carcinoma. *J. Exp. Clin. Cancer Res.* 35, 12. doi:10.1186/s13046-015-0281-z

Komanduri, K. V., and Levine, R. L. (2016). Diagnosis and therapy of acute myeloid leukemia in the era of molecular risk stratification. *Annu. Rev. Med.* 67, 59–72. doi:10.1146/annurev-med-051914-021329

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). Timer: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Lin, E. Y., Li, J. F., Gnatovskiy, L., Deng, Y., Zhu, L., Grzesik, D. A., et al. (2006). Macrophages regulate the angiogenic switch in a mouse model of breast cancer. *Cancer Res.* 66 (23), 11238–11246. doi:10.1158/0008-5472.CAN-06-1278

Mandal, A., and Viswanathan, C. (2015). Natural killer cells: In health and disease. *Hematol. Oncol. Stem Cell Ther.* 8 (2), 47–55. doi:10.1016/j.hemonc.2014.11.006

Mazumder, S., Choudhary, G. S., Al-Harbi, S., and Almasan, A. (2012). Mcl-1 Phosphorylation defines ABT-737 resistance that can be overcome by increased NOXA expression in leukemic B cells. *Cancer Res.* 72 (12), 3069–3079. doi:10.1158/0008-5472.CAN-11-4106

Nakamura, N., Yamakawa, N., Sato, T., Tojo, H., Tachi, C., and Furukawa, K. (2001). Differential gene expression of beta-1, 4-galactosyltransferases I, II and V during mouse brain development. *J. Neurochem.* 76 (1), 29–38. doi:10.1046/j.1471-4159.2001.00004.x

Nechiporuk, T., Kurtz, S. E., Nikolova, O., Liu, T., Jones, C. L., D'Alessandro, A., et al. (2019). The TP53 apoptotic network is a primary mediator of resistance to BCL2 inhibition in AML cells. *Cancer Discov.* 9 (7), 910–925. doi:10.1158/2159-8290.CD-19-0125

Nixon, B., Lu, Q., Wassler, M. J., Foote, C. I., Ensslin, M. A., and Shur, B. D. (2001). Galactosyltransferase function during mammalian fertilization. *Cells Tissues Organs* 168 (1-2), 46–57. doi:10.1159/000016805

Pan, R., Ruvolo, V., Mu, H., Leverson, J. D., Nichols, G., Reed, J. C., et al. (2017). Synthetic lethality of combined bcl-2 inhibition and p53 activation in AML: Mechanisms and superior antileukemic efficacy. *Cancer Cell* 32 (6), 748–760. doi:10.1016/j.ccell.2017.11.003

Pan, R., Ruvolo, V. R., Wei, J., Konopleva, M., Reed, J. C., Pellecchia, M., et al. (2015). Inhibition of Mcl-1 with the pan-Bcl-2 family inhibitor (-) BI97D6 overcomes ABT-737 resistance in acute myeloid leukemia. *Blood* 126 (3), 363–372. doi:10.1182/blood-2014-10-604975

Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. (2008). Structure and function of beta -1, 4-galactosyltransferase. *Curr. Drug Targets* 9 (4), 292–309. doi:10.2174/138945008783954943

Rajput, B., Shaper, N. L., and Shaper, J. H. (1996). Transcriptional regulation of murine beta1 4-galactosyltransferase in somatic cells Analysis of a gene that serves both a housekeeping and a mammary gland-specific function. *J. Biol. Chem.* 271 (9), 5131–5142. doi:10.1074/jbc.271.9.5131

Rudat, S., Pfaus, A., Cheng, Y. Y., Holtmann, J., Ellegast, J. M., Buhler, C., et al. (2018). RET-mediated autophagy suppression as targetable co-dependence in acute myeloid leukemia. *Leukemia* 32 (10), 2189–2202. doi:10.1038/s41375-018-0102-4

Ruffell, B., Chang-Strachan, D., Chan, V., Rosenbusch, A., Ho, C. M., Pryer, N., et al. (2014). Macrophage IL-10 blocks CD8+ T cell-dependent responses to chemotherapy by suppressing IL-12 expression in intratumoral dendritic cells. *Cancer Cell* 26 (5), 623–637. doi:10.1016/j.ccell.2014.09.006

Shi, Y., Ping, Y. F., Zhou, W., He, Z. C., Chen, C., Bian, B. S., et al. (2017). Tumour-associated macrophages secrete pleiotrophin to promote PTPRZ1 signalling in glioblastoma stem cells for tumour growth. *Nat. Commun.* 8, 15080. doi:10.1038/ncomms15080

Sillar, J. R., and Enjeti, A. K. (2019). Targeting apoptotic pathways in acute myeloid leukaemia. *Cancers (Basel)* 11 (11), 1660. doi:10.3390/cancers11111660

Szczepanski, M. J., Szajnik, M., Czystowska, M., Mandapathil, M., Strauss, L., Welsh, A., et al. (2009). Increased frequency and suppression by regulatory T cells in patients with acute myelogenous leukemia. *Clin. Cancer Res.* 15 (10), 3325–3332. doi:10.1158/1078-0432.CCR-08-3010

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi:10.1093/nar/gky1131

Tallman, M. S., Gilliland, D. G., and Rowe, J. M. (2005). Drug therapy for acute myeloid leukemia. *Blood* 106 (4), 1154–1163. doi:10.1182/blood-2005-01-0178

Tang, W., Weng, S., Zhang, S., Wu, W., Dong, L., Shen, X., et al. (2013). Direct interaction between surface β1, 4-galactosyltransferase 1 and epidermal growth factor receptor (EGFR) inhibits EGFR activation in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* 434 (3), 449–454. doi:10.1016/j.bbrc.2013.03.094

Ustun, C., Miller, J. S., Munn, D. H., Weisdorf, D. J., and Blazar, B. R. (2011). Regulatory T cells in acute myelogenous leukemia: Is it time for immunomodulation? *Blood* 118 (19), 5084–5095. doi:10.1182/blood-2011-07-365817

Villegas-Comonfort, S., Serna-Marquez, N., Galindo-Hernandez, O., Navarro-Tito, N., and Salazar, E. P. (2012). Arachidonic acid induces an increase of beta-1, 4-galactosyltransferase I expression in MDA-MB-231 breast cancer cells. *J. Cell. Biochem.* 113 (11), 3330–3341. doi:10.1002/jcb.24209

Wang, M., Zhang, C., Tian, T., Zhang, T., Wang, R., Han, F., et al. (2018). Increased regulatory T cells in peripheral blood of acute myeloid leukemia patients rely on tumor necrosis factor (TNF)-alpha-TNF receptor-2 pathway. *Front. Immunol.* 9, 1274. doi:10.3389/fimmu.2018.01274

Wang, P., Li, X., and Xie, Y. (2020). B4GalT1 regulates apoptosis and autophagy of glioblastoma *in vitro* and *in vivo*. *Technol. Cancer Res. Treat.* 19, 1533033820980104. doi:10.1177/1533033820980104

Yang, W., and Xu, Y. (2013). Clinical significance of Treg cell frequency in acute myeloid leukemia. *Int. J. Hematol.* 98 (5), 558–562. doi:10.1007/s12185-013-1436-3

Youakim, A., Dubois, D. H., and Shur, B. D. (1994). Localization of the long form of beta-1, 4-galactosyltransferase to the plasma membrane and Golgi complex of 3T3 and F9 cells by immunofluorescence confocal microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 91 (23), 10913–10917. doi:10.1073/pnas.91.23.10913

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, J., Zhang, G., Zhang, W., Bai, L., Wang, L., Li, T., et al. (2022). Loss of RBMS1 promotes anti-tumor immunity through enabling PD-L1 checkpoint blockade in triple-negative breast cancer. *Cell Death Differ.* 29, 2247–2261. doi:10.1038/s41418-022-01012-0

Zhu, X., Jiang, J., Shen, H., Wang, H., Zong, H., Li, Z., et al. (2005). Elevated beta1, 4-galactosyltransferase I in highly metastatic human lung cancer cells. Identification of E1AF as important transcription activator. *J. Biol. Chem.* 280 (13), 12503–12516. doi:10.1074/jbc.M413631200

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fgene.2023.885930

Check for updates

OPEN ACCESS

EDITED BY
Jitian Li,
Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), China

REVIEWED BY
Shuwen Han,
Huzhou Central Hospital, China
Zhixun Zhao,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Yinan Jiang,
University of Pittsburgh, United States
Jianxiang Shi,
Zhengzhou University, China

*CORRESPONDENCE
Kefeng Ding,
dingkefeng@zju.edu.cn
Xue Li,
xueli157@zju.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Genetics and Oncogenomics, a section of the journal Frontiers in Genetics

RECEIVED 28 February 2022
ACCEPTED 14 February 2023
PUBLISHED 03 March 2023

CITATION
Yu L, Wang L, Sun J, Zhou X, Hu Y, Hu L, He Y, Lin C, Chen J, Xu X, Dunlop MG, Theodoratou E, Ding K and Li X (2023), N6-methyladenosine related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer. *Front. Genet.* 14:885930. doi: 10.3389/fgene.2023.885930

COPYRIGHT
© 2023 Yu, Wang, Sun, Zhou, Hu, Hu, He, Lin, Chen, Xu, Dunlop, Theodoratou, Ding and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

N6-methyladenosine related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer

Lili Yu[1,2], Lijuan Wang[2], Jing Sun[2], Xuan Zhou[2], Yeting Hu[1], Lidan Hu[3], Yazhou He[4], Chunqing Lin[5], Jie Chen[6], Xiaolin Xu[2], Malcolm G. Dunlop[7], Evropi Theodoratou[7,8], Kefeng Ding[1,9]* and Xue Li[2]*

[1]Department of Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2]Analytics of The Second Affiliated Hospital and Department of Big Data in Health Science School of Public Health, Center of Clinical Big Data, Zhejiang University School of Medicine, Hangzhou, China, [3]The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [4]Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China, [5]National Cancer Center, National Clinical Research Center for Cancer, and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [6]Center for Global Health, Zhejiang University, Hangzhou, China, [7]Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom, [8]Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom, [9]Cancer Center, Zhejiang University, Hangzhou, China

N6-methyladenosine (m6A) modification has been demonstrated to exhibit a crucial prognostic effect on colorectal cancer (CRC). Nonetheless, potential mechanism of m6A in survival rate and immunotherapeutic response remains unknown. Here we investigated the genes associated with m6A regulators and developed a risk score for predicting the overall survival (OS) of CRC patients. RNA-seq transcriptomic profiling data of COAD/READ samples were obtained from The Cancer Genome Atlas (TCGA) database. Absolute Shrinkage and Selection Operator (LASSO)- Cox regression analysis was conducted to identify the m6A-related gene expression signatures and the selected genes were inputted into stepwise regression to develop a prognostic risk score in TCGA, and its predictive performance of CRC survival was further validated in Gene Expression Omnibus (GEO) datasets. According to our results, the risk score comprising 18 m6A-related mRNAs was significantly associated with CRC survival in both TCGA and GEO datasets. And the stratified analysis also confirmed that high-risk score acted as a poor factor in different age, sex, T stage, and tumour, node, metastasis (TNM) stages. The m6A-related prognostic score in combination with clinical characteristics yielded time-dependent area under the receiver operating characteristic curve (AUCs) of 0.85 (95%CI: 0.79–0.91), 0.84 (95%CI: 0.79–0.90) and 0.80 (95%CI: 0.71–0.88) for the prediction of the 1-, 3-, 5-year OS of CRC in TCGA cohort. Furthermore, mutation of oncogenes occurred more frequently in the high-risk group and the composition of immune cells in tumour microenvironment (TME) was significantly distinct between the low- and high-risk groups. The low-risk group had a lower microsatellite instability (MSI) score, T-cell exclusion score

and dysfunction score, implying that low-risk patients may have a better immunotherapy response than high-risk patients. In summary, a prognostic risk score derived from m6A-related gene expression signatures could serve as a potential prognostic predictor for CRC survival and indicator for predicting immunotherapy response in CRC patients.

KEYWORDS

colorectal cancer, prognostic risk score, gene expression, overall survival, immune responses

Introduction

Colorectal cancer (CRC) is the third most common cancer and a leading cause of cancer mortality worldwide (Bray et al., 2018). Although the survival time of CRC patients has been significantly extended by clinical treatment, the 5-year OS of CRC patients is still not ideal, with a rate of approximate 68% (Yuan et al., 2021). Presently, emerging evidence has shown that the discovery and application of molecular biomarkers may provide important clinical implications on the prognosis and treatment of CRC patients (Bramsen et al., 2017).

N6-methyladenosine (m6A) is one of the most prominent and abundant forms of internal RNA modification involved in stabilizing transcripts, affecting degradation process of mRNA and non-coding RNA, and promoting initiate translational efficiency (Wang et al., 2014; Liu et al., 2015; Ma et al., 2019; Sun et al., 2019). This modification regulated by methyltransferases, demethylases, and binding proteins, is a dynamic reversible process in mammalian cells, which are also known as "writers", "erasers", and "readers" (Yang et al., 2018). Of note, some of these effects are mediated by m6A "readers" proteins, which can selectively recognize m6A and exert a regulatory function on the m6A-marked mRNAs (He and He, 2021). Several recent systematic studies of the cross-link between m6A modification, substrate genes, and post-modification regulation to reveal the biological role of m6A in cancer development comprehensively (Luo et al., 2018; Chang et al., 2019). METTL3 enhances translation of oncogene BRD4 through forming an mRNA loop in lung adenocarcinoma, and promotes expression of SRY (sex-determining region Y)-box 2 (SOX2) through IGF2BP2-directed suppression of RNA degradation in CRC (Choe et al., 2018; Li et al., 2019). The m6A functions induced by m6A related modification enzymes can be influenced by environmental exposure (e.g., reactive oxygen species, inflammation, and cyclobutene pyrimidine dimers) and genomic signals (e.g., somatic mutation), thereby epigenetics provides a molecular basis for cancer development (Li D. et al., 2020; Rong et al., 2021). To date, accumulating evidence demonstrated that dysregulated m6A methylation modification is associated with multiple biological processes, including dysregulate cell proliferation and death, immunomodulatory abnormality and tumour malignant progression (Fu et al., 2014), thus could be closely related to a variety of human diseases, in particular cancer (Hong, 2018). For instance, it is shown that YTHDF2 may act as a tumour suppressor to restrain cell proliferation and growth *via* destabilizing the EGFR mRNA in hepatocellular carcinoma (Zhong et al., 2019). Previous study on the pathological role of m6A modification in CRC reported that METTL3, one of m6A regulators, directly induced m6A-glucose transporter 1 (GLUT1)-mammalian target of rapamycin complex 1 (mTORC1) axis to promote CRC initiation and progression (Chen et al., 2021). Likewise, another experimental study showed that METTL3 stabilizes HK2 and SLC2A1 (GLUT1) expression in CRC through an m6A-IGF2BP2/3- dependent mechanism, thereby pointing to the notion that m6A modification is a promising indicator of controlling human CRC aggressiveness (Shen et al., 2020). However, the specific role of m6A regulators in the dysregulation of mRNAs in CRC prognosis remains unclear.

The tumour microenvironment (TME), which is composed of various cancer cells, stromal cells, and distinct recruited cells (infiltrating immune cells, bone marrow-derived cells), plays a vital role in tumour progression and affects the clinical benefit from novel strategies of immunological checkpoint blockade (ICB) (Hanahan and Coussens, 2012; Topalian et al., 2012). ICB treatment, such as those programmed cell death protein 1 (PD1), programmed death-ligand 1 (PDL1) and cytotoxic T-lymphocyte antigen 4 (CTLA-4) is now the first class of immunotherapy to have a broad impact on survival for cancer patients, across a wide variety of tumour histologies and treatment settings (Littman, 2015; Lonberg and Korman, 2017; Ribas and Wolchok, 2018; Wieder et al., 2018). Emerging studies have made efforts to understand the heterogeneity and complexity of the TME by elaborate analysis of m6A modification, therefore improving immunotherapy strategies (Li N. et al., 2020). Predicting the immunotherapy response of CRC patients based on multiple m6A-related biomarkers has the potential to develop a personalised treatment strategy and therefore to increase the success of ICB (Fang and Declerck, 2013; Quail and Joyce, 2013; Binnewies et al., 2018).

In this study, we sought to elucidate the m6A related mRNAs signatures for predicting the overall survival (OS) and immune responses of CRC patients using transcriptome data from The Cancer Genome Atlas (TCGA) (2012) (Cancer Genome Atlas Network, 2012) and Gene Expression Omnibus (GEO) (Smith et al., 2010; Marisa et al., 2013) datasets. We focused on the m6A-related genes and developed a multivariate Cox prediction model for the OS of CRC patients and examined its prognostic ability in immunotherapy response. We additionally explored the candidate drugs targeting these m6A-related gene signatures using the publicly available Genomics of Drug Sensitivity in Cancer (GDSC) database for predicting drug sensitivity (Yang et al., 2013). Findings from this study are helpful to predict the prognosis of CRC and develop personalized CRC treatment strategies.

Materials and methods

Study population and datasets

A study sample of 644 CRC patients from the TCGA was used as a training dataset. RNA-seq [Fragments Per Kilobase of transcript per Million mapped reads (FPKM normalized)] were acquired from Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) using the R package "TCGAbiolinks", which was specifically developed for integrative analysis with Genetic Data Commons (GDC) data (Colaprico et al., 2016). Then FPKM values were transformed into transcripts per kilobase million (TPM) values. The corresponding clinicopathological information and somatic mutation data of CRC patients were obtained from the cBioPortal database (https://portal.gdc.cancer.gov/). Two study samples (GSE39582, N = 566; GSE17536, N = 177) from the GEO database were used as validation datasets, and their normalized microarray gene expression data and clinicopathological data were obtained online (https://www.ncbi.nlm.nih.gov/geo/). Those RNA probe sets were re-annotated using the Ensemble database (http://www.ensembl.org). CRC patients with missing survival data and OS values or OS < 30 days were excluded in order to reduce statistical bias in this analysis.

Identification of m6A-related prognostic genes

The expression matrices of 21 m6A regulators were retrieved from the TCGA, including the expression data of eight writers (*METTL3, METTL14, METTL16, RBMX2, RBM15B, WTAP, KIAA1429*, and *ZC3H13*), two erasers (*FTO* and *ALKBH5*), and eleven readers (*YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, IGF2BP1, EMR1, LRPPRC, HNRNPA2B1, HNRNPC*, and *ELAVL1*). Based on the RNA-seq data, Pearson's correlation analysis was firstly implemented to identify m6A-related genes, using the criteria of | Pearson R| >0.3 and $p < 0.001$. Univariable and multivariable Cox regression models (false discovery rate, FDR<0.05) and the least absolute shrinkage and selection operator (LASSO) Cox regression were conducted subsequently to select the m6A-related prognostic genes that were distinctly related to the OS of CRC patients. The proteins of m6A regulators and m6A-related prognostic genes in CRC and normal tissues were further examined by using immunohistochemistry data in the human Protein Atlas (HPA) (https://www.proteinatlas.org/) database, which provided expression levels of 24,000 protein in different tissues and cells (Uhlen et al., 2017).

Development and validation of the m6A-related prognostic risk score

A weighted prognostic risk score of m6A-related gene expression was constructed based on the following formula: Risk score = $\sum_{i=1}^{n} Coef\ (Genei) \times Expr\,(Genei)$, where Coef ($Gene_i$) was the coefficient of genes correlated with CRC survival, and Expr ($Gene_i$) was the expression of genes. The prognostic value of the risk score was evaluated by Kaplan-Meier survival curves with log-rank tests in both TCGA and GEO study samples. Multivariate Cox regression analysis was performed to evaluate the prediction performance of the m6A-related prognostic risk score. Patients with CRC were further stratified into low- and high-risk groups based on the median value of the prognostic risk score of m6A-related genes.

Analysis of the molecular characteristics in the low- and high-risk groups

To explore the biological function and alternative pathways of these m6A-related gene signatures, we performed a co-expression and pathway enrichment analysis based on the TCGA database, using the Kyoto Encyclopaedia of Genes and Genomes Pathway (KEGG pathway) as reference (Kanehisa and Goto, 2000). Linear regression was performed to detect co-expressed genes (FDR<0.05). In the gene mutation analysis, we obtained somatic mutation information from the cBioPortal database, and the quantity and quality of gene mutations were analysed in low- and high-risk groups by using the Maftools package in R.

Exploration of immunotherapeutic response between low- and high-risk groups

To depict immune characteristics of CRC patients, the entire expression data were imported into CIBERSORT (https://cibersort.stanford.edu/) and a deconvolution algorithm using support vector regression was used and iterated 1,000 times to determine the relative proportions of 22 immune cell types in tumours. The relative proportions of immune cell types and clinicopathologic factors were compared between the low- and high-risk groups. The tumour Immune Dysfunction and Exclusion (TIDE) score was calculated online (http://tide.dfci.harvard.edu/) to predict the likelihood of immunotherapeutic response between the low- and high-risk groups.

Prediction of potential compounds targeting therapeutic sensitivity in CRC patients

To obtain potential compounds with differential therapeutic sensitivity, we investigated the predictive capacity of the low- and high-risk groups in responding immunotherapy. The 50% inhibiting concentration half-maximal inhibitory concentration (IC_{50}) value of 138 anti-cancer drugs was inferred from the GDSC website based on the COAD/READ dataset of the TCGA project. The "pRRophetic" algorithm (Geeleher et al., 2014) was used to predict the IC_{50} of compounds in the low- and high-risk groups separately.

Statistical analysis

An independent *t*-test was performed to compare continuous variables between two groups. Categorical data were tested using the

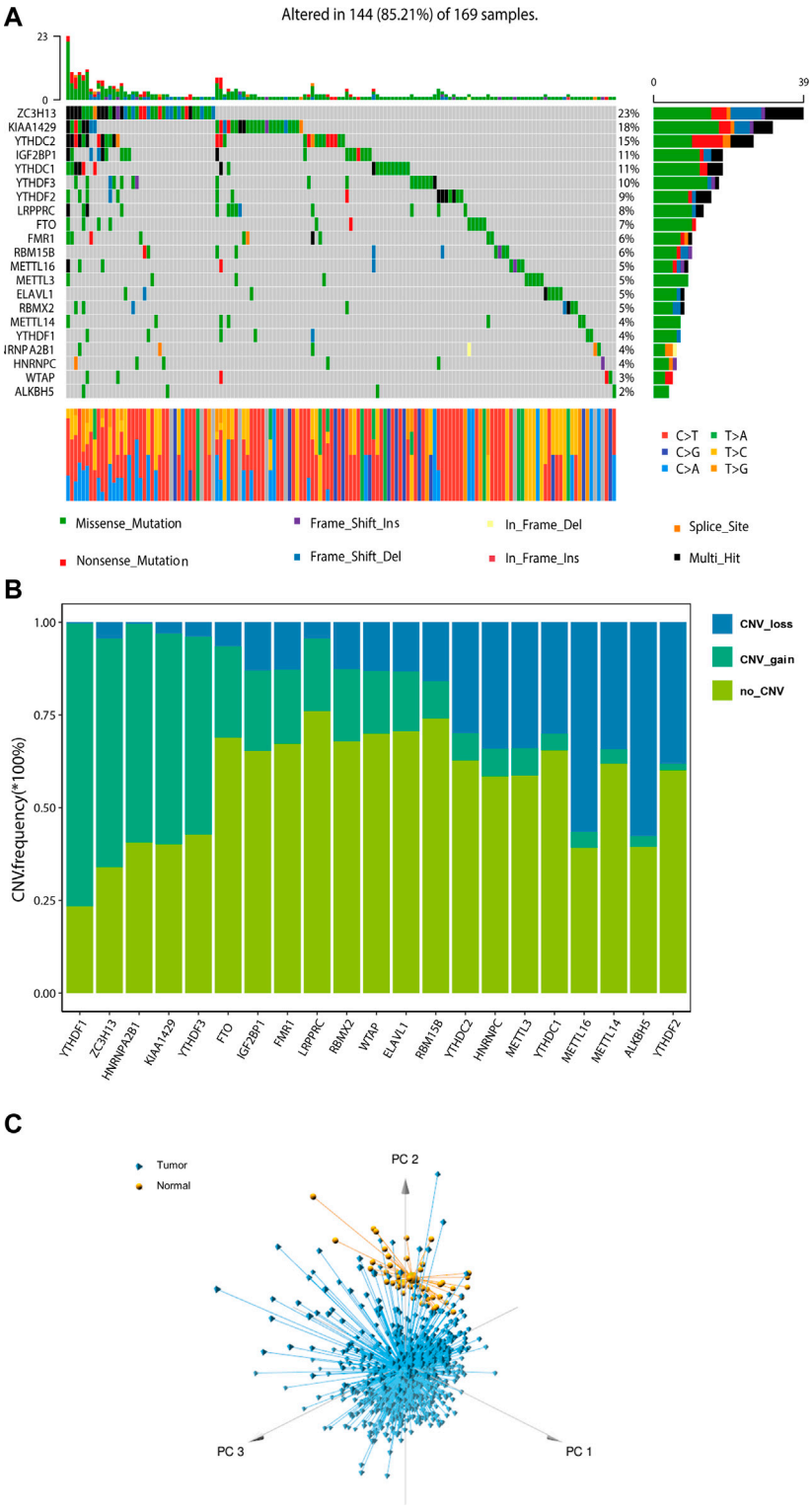


FIGURE 1
Landscape of genetic of m6A regulators in colorectal cancer. **(A)** The mutation frequency of 21 m6A regulators in 169 patients with CRC from TCGA cohort. **(B)** Bar graphs showing the frequency of CNV gain (green), loss (blue) and non CNV (yellow) of m6A regulators in TCGA-COAD/READ cohort. **(C)** Principal component analysis for the expression profiles of 21 m6A regulators to distinguish tumours from normal samples in TCGA cohort.

TABLE 1 Correlation between subgroups and clinicopathological factors in the TCGA cohort.

Variable	Overall (*n* = 551)	High-risk (*n* = 275)	Low-risk (*n* = 276)	*p*-value
Age, n (%)				0.730
<60 years	164 (29.8%)	80 (29.1%)	84 (30.4%)	
≥60 years	387 (70.2%)	195 (70.9%)	192 (69.6%)	
Gender, n (%)				0.581
Female	254 (46.1%)	130 (47.2%)	124 (44.9%)	
Male	297 (53.9%)	145 (52.7%)	152 (55.1%)	
Cancer type, n (%)				0.368
COAD	408 (74.0%)	199 (72.4%)	209 (75.7%)	
READ	143 (26.0%)	76 (27.6)	67 (24.3%)	
AJCC stage, n (%)				<0.001
I	96 (17.4%)	35 (12.7%)	61 (22.1%)	
II	201 (36.5%)	85 (30.9%)	116 (42.0%)	
III	163 (29.6%)	98 (35.6%)	65 (23.6%)	
IV	79 (14.3%)	50 (18.2%)	29 (10.5%)	
Not available	12 (2.2%)	7 (2.6%)	5 (1.8%)	
T stage, n (%)				0.004
T1-2	116 (21.1%)	44 (16.0%)	72 (26.1%)	
T3-4	435 (78.9%)	231 (84.0%)	204 (73.9%)	
M stage, n (%)				0.020
M0	411 (74.6%)	193 (70.2%)	218 (79.0%)	
M1-x	135 (24.5%)	79 (28.7%)	56 (20.3%)	
Not available	5 (0.9%)	3 (1.1%)	2 (0.7%)	
N stage, n (%)				<0.001
N0	314 (57.0%)	128 (46.5%)	186 (67.4%)	
N1-x	237 (43.0%)	147 (53.5%)	90 (32.6%)	
Radiotherapy, n (%)				0.301
No	457 (83.0%)	224 (81.5%)	233 (84.4%)	
Yes	27 (4.9%)	16 (5.8%)	11 (4.0%)	
Not available	67 (12.1%)	35 (12.7%)	32 (11.6%)	
Survival status, n (%)				<0.001
Alive	438 (79.5%)	189 (68.7%)	249 (90.2%)	
Dead	113 (20.5%)	86 (31.3%)	27 (9.8%)	

χ^2 test. Pearson correlation analysis was implemented to identify m6A-related genes (with the | Pearson r | >0.05 and $p < 0.001$). Univariate survival analysis was performed by K-M survival analysis with the log-rank test to calculate the significance of differences in the OS. Multivariate survival analysis was performed using the Cox regression model to estimate the hazard ratio (HR). The time-dependent area under the receiver operating characteristic curve (AUC) was estimated to evaluate the predictive power of the risk score and TNM stage to the OS. Stratification analysis was performed to investigate the survival difference in subgroups, including age, sex, T stage, N stage, M stage, American Joint Committee on Cancer (AJCC) TNM stage and radiation therapy history. A nomogram of the risk score and other predictors was set up accordingly for the prediction of the 1-, 3-, 5- year OS. The *p* values were two-sided and $p < 0.05$ was considered as statistically significant.

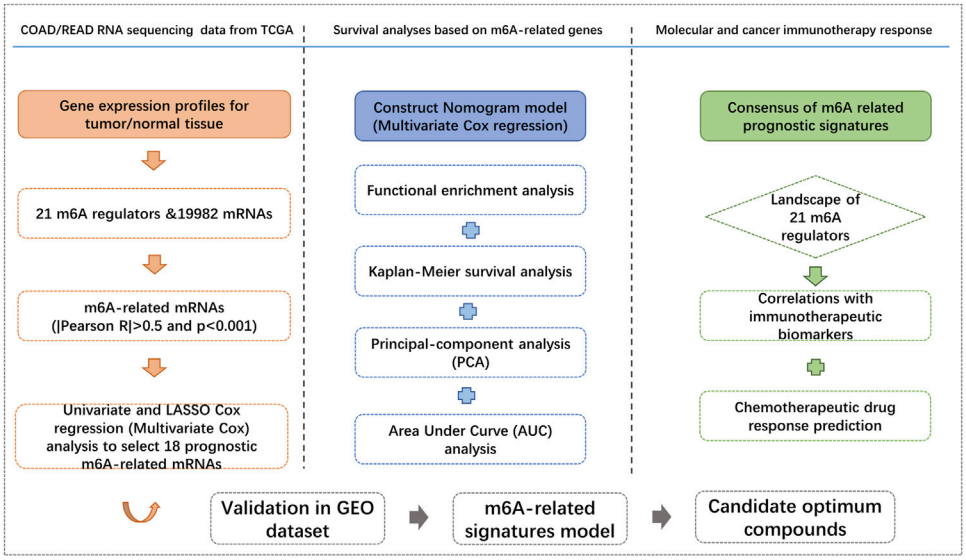


FIGURE 2
Flow chart of this study.

Results

Landscape of genetic variation of m6A regulators in CRC patients

A total of 21 m6A regulators, namely, 8 "writers", 2 "erasers", and 11 "readers", were included in this study. We firstly assessed the prevalence of somatic mutations and copy number variations (CNV) of these 21 m6A regulators. Among the 551 samples, 169 (30.67%) had mutations in any of the m6A modification regulators (Figure 1A; Table 1). *ZC3H13* exhibited the highest mutation frequency (23%) followed by *KIAA1429* (18%) and *YTHDC2* (15%), while demethylases *ALKBH5* (2%) and *WTAP* (3%) showed low number of mutations in CRC samples. Somatic copy number alterations of these m6A regulators were then examined, and we found that *METTL14* (34%), *METTL16* (56%), *ALKBH5* (58%) and *YTHDF2* (38%) had a widespread frequency of CNV deletions (Figures 1B, C). To ascertain whether the above genetic variations influenced the expression of m6A regulators in CRC patients, we investigated the mRNA alterations of the m6A regulators between paired normal and tumour samples of CRC patients. This showed that alterations of CNV were prominent factors, resulting in perturbations on the m6A regulators expression. Compared to the normal colon tissues, regulators with CNV gain demonstrated markedly higher expression in CRC tissues (e.g., *YTHDF1* and *KIAA1429*) (Figure 1B; Supplementary Figure S1). And *vice versa*, some regulators showed downregulated mRNA expression but with high frequency of CNV loss (e.g., *ALKBH5*). This analysis showed the high heterogeneity of genetic and expressional alteration landscape of m6A regulators between normal and tumour samples, demonstrating that the expression imbalance of m6A regulators may be important in the initiation and progression of CRC.

Identification of m6A-related genes in patients with CRC

A total of 551 COAD/READ patients from the TCGA database were included in our study to calculate the prognostic risk score of m6A-related genes. The detailed workflow for risk model construction and subsequent analyses is shown in Figure 2. We abstracted the matrix expression of 21 m6A regulators and 19,982 mRNAs from the TCGA database. Correlations between these 21 m6A regulators and 19,982 mRNAs were examined and we identified 4,274 mRNAs that were significantly correlated with m6A regulators base on the criteria of |Pearson R|>0.5 and $p < 0.001$. To identify m6A-related genes that correlated with the OS of CRC patients, we screened from 4,274 m6A-associated mRNAs in the TCGA training set using univariate Cox regression analysis. At FDR<0.05, fifty-seven m6A-related mRNAs correlated significantly with OS (Supplementary Table S1).

Construction of the prognostic risk score based on m6A-related gene expression signatures

To avoid overfitting, the LASSO-Cox regression was applied to optimise the selection of gene signatures in relation to the OS. Consequently, 18 m6A-related mRNAs (*PMM2*, *ERI1*, *NEK9*, *USP53*, *CNOT3*, *CDK5RAP2*, *ING5*, *HMGXB4*, *SH3D19*, *UBE2H*, *CLK1*, *SFPQ*, *UBP1*, *PDCD6IP*, *ZNF248*, *SCL25A53*, *CLCC1* and *GPR125*) were finally selected to construct a m6A-related prognostic risk score for CRC survival (Supplementary Figures S2A, S2B). The correlation between m6A regulators and m6A-related gene expression in the TCGA dataset is showed in Supplementary Figure S3. Twelve out of the 18 gene products and 17 out of the 21 m6A regulators were obtained from the HPA database, and the

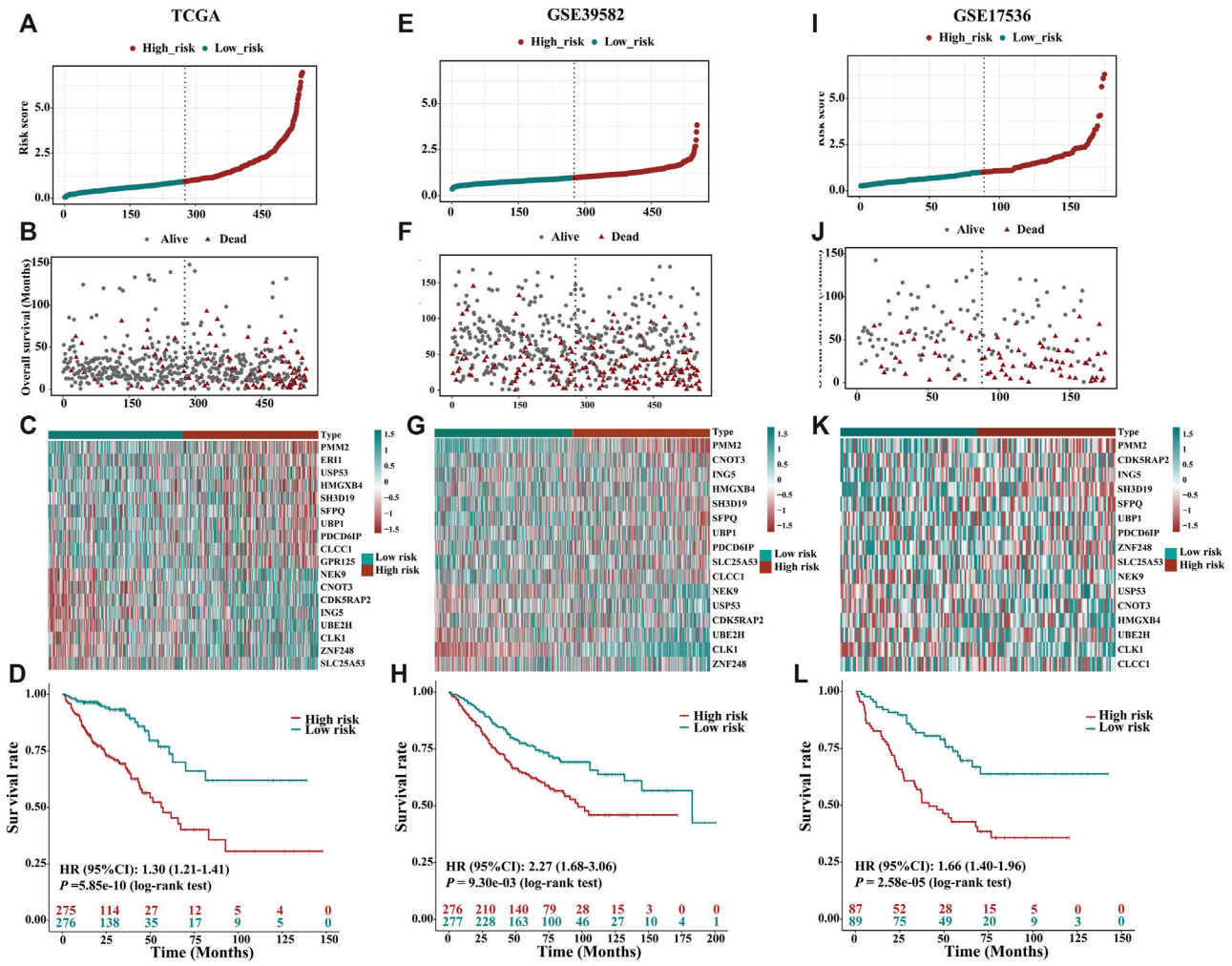


FIGURE 3
Prognostic value of the risk patterns of the 18 m6A-related gene signatures in the TCGA training dataset, GSE39582 and GSE17536 validation dataset. **(A)** Distribution of m6A-related gene expression model-based risk score for TCGA. **(B)** Different patterns of survival status and survival time between the high- and low-risk subgroups for TCGA. **(C)** Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient for TCGA. **(D)** Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups for TCGA. **(E)** Distribution of m6A-related gene expression model-based risk score for the GSE39582. **(F)** Different patterns of survival status and survival time between the high- and low-risk subgroups for the GSE39582. **(G)** Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient for the GSE39582. **(H)** Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups for the GSE39582. **(I)** Distribution of m6A-related gene expression model-based risk score for the GSE17536. **(J)** Different patterns of survival status and survival time between the high- and low-risk subgroups for the GSE17536. **(K)** Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient for the GSE17536. (**L**) Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups for the GSE17536.

other six genes and four m6A regulators were not available or in low reliability. The immunohistochemistry-stained proteins of the 17 m6A regulators and 12 genes in CRC and normal tissues were shown in Supplementary Figures S4A, S4B. A weighted prognostic risk score of m6A-related gene expression was constructed based on the gene expression levels of the 18 selected markers. CRC patients were separated into high- and low-risk groups based on the median value of the prognostic risk score constructed by the m6A-related gene expression signatures. The distribution of risk scores between the low- and high-risk groups is depicted in Figure 3A, and the survival status and survival time of CRC patients in the low- and high-risk groups are shown in Figure 3B. The expression levels of the 18 m6A-related genes in the low- and high-risk groups are shown in Figure 3C. Kaplan–Meier survival curves showed that CRC patients with higher risk scores had worse clinical outcomes (lower OS rates and a shorter OS time, HR = 1.30, 95%CI: 1.21–1.41; p = 5.85e-10, log-rank test) (Figure 3D). Based on the entire gene expression profiles, 21 m6A regulators and the expression profile of the 18 m6A-related genes, PCA analysis was further conducted to test the difference between the low- and high-risk groups (Supplementary Figures S5A–C). As showed Supplementary Figures S5A, B, the gene expression profiles of the low- and high-risk groups were differently distributed (Supplementary Figures S5C).

TABLE 2 Clinical information of CRC cohorts from GEO dataset.

Accession number	Platform	Tumor samples	Survival data	Stage	Gender	PMID
GSE39582	GPL570	566	553	T stage I-IV: 553	M:322; F:263	23700391
GSE17536	GPL570	177	176	Stage I-IV: 176	M:96; F:81	19914252

Validation of the prognostic risk score based on m6A-related gene expression signatures

Detailed clinicopathologic characteristics of CRC patients in TCGA and GEO datasets are shown in **Table 1** and **Table 2**. The expression of 18 m6A-related genes was closely correlated with the OS of CRC patients as determined by K-M analysis (Supplementary Figures S6). According to the subgroups classified by sex, age, AJCC TNM stage or tumour stage, the OS of the low-risk group continued to be superior to that of the high-risk group (Supplementary Figures S7A–H). To validate the prognostic capability, we calculated the risk scores for CRC patients in two GEO (GSE39582, N = 553; GSE17536, N = 176) datasets using the same formula. As showed in Figures 3E–G; Figures 3I–K, patients stratified into the high-risk group had a significantly worse prognosis than those in the low-risk group (Figure 3H, HR = 2.27, 95%CI: 1.68–3.06, p = 9.30e-03, log-rank test; Figure 3L, HR = 1.66, 95%CI: 1.40–1.96, p = 2.58e-05, log-rank test), which was consistent with the results of TCGA dataset.

Molecular characteristics of the low- and high-risk groups stratified by the prognostic risk score

To demonstrate the potential mechanisms and pathways involved in the molecular heterogeneity leading to the different outcomes between the low- and high-risk groups, we performed functional enrichment analysis with annotation of KEGG gene set. We found that m6A-related gene expression signatures were differentially enriched (FDR<0.05) in the pathways related to cancer, immune response, and neural signaling between the two groups (Supplementary Table S2), and pathways that more than half of the gene signatures enriched in were summarized in Supplementary Figure S8. When examining the somatic mutations, we found that the top 20 cancer driver genes mutated more frequently in the high-risk group than in the low-risk group, including four tumour suppressor genes (e.g., APC, TP53, LRP1B and ZFHX4) and the other sixteen genes (e.g., TTN, KARS, MUC16, SYNE1, PIK3CA, FAT4, RYR2 DNAH5, RYR1 and FBXW7, etc.) are oncogenes in disease development and progression (Figures 4A,B), and significant co-occurrences were also observed among mutations of these genes (as shown in Figure 4C).

Estimation of the tumour immune microenvironment and cancer immunotherapy response

To analyse the composition of immune cells in different risk groups, we used the Wilcoxon test to compare the distribution of immune cells. As shown in Figure 5A, we found that CD8 T cells, Tregs regulatory T cells, M0 macrophages, and resting natural killer (NK) cells were more abundant in the high-risk group, while plasma cells, resting memory CD4 T cells, activated memory CD4 T cells and M2 macrophages were more abundant in the low-risk group. Likewise, activated memory CD4 T cells and M2 macrophages are significantly distributed in different stages of CRC patients (Supplementary Figure S9). The correlations between the m6A-related signature model and immunotherapeutic biomarkers were then investigated. Compared with that in the low-risk group, PD1 and CTLA4 expression in the high-risk was significantly higher, suggesting that high-risk CRC patients have a potential response to anti-PD-1 immunotherapy (Figures 5B–D). Additionally, higher TIDE prediction score represented a higher potential for immune evasion, which suggested that the patients were less likely to benefit from (ICB) therapy. In our results, the low-risk group had a lower TIDE score than the high-risk group, implying that low-risk patients may have a better immunotherapy response than high-risk patients. Also, we found that the high-risk group had a higher microsatellite instability (MSI) score, T-cell exclusion score and dysfunction score (Figures 5E–H). To find the potency of m6A-related prognostic score as a biomarker for predicting the response of CRC patients to drugs, "pRRophetic" algorithm was used to infer the therapeutic response based on the IC50 value of the 138 anti-cancer drugs in TCGA-COAD/READ patients. We found 50 chemotherapeutic drugs displaying differential IC50 between these two groups (Supplementary Figure S10).

Construction of nomogram based on prognostic risk score and clinical characteristics

We next investigated the distribution of the risk score of patients with CRC using different conventional clinical information (including sex, T stage, N stage, M stage and AJCC TNM stage), and confirmed that CRC patients with higher T, N or TNM stage had a higher risk score (Figure 6A). Univariate Cox analysis showed that age, radiation history, T stage, N stage and the prognostic risk score were significantly associated with the prognosis of CRC (Figure 6B). Multivariate Cox analysis confirmed that the prognostic risk score based on m6A-related gene expression signatures was an independent predictor of CRC survival (Figure 6C). Multivariate Cox prediction models combing prognostic risk score and clinical characteristics yielded AUCs of 0.854 (95%CI: 0.795–0.913), 0.844 (95%CI: 0.790–0.898) and 0.796 (95%CI: 0.708–0.883) for the

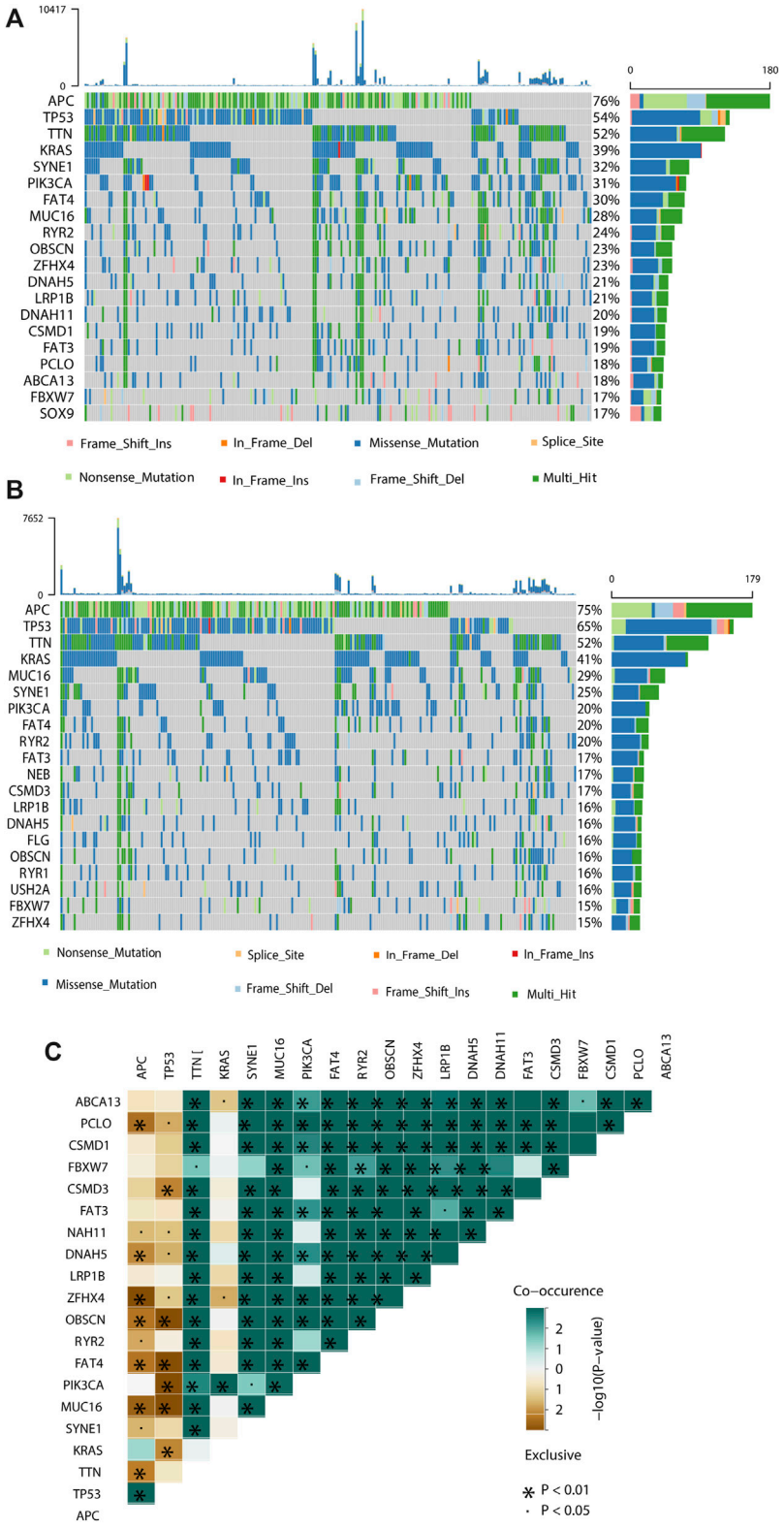


FIGURE 4
Molecular characteristics of different risk subgroups. **(A, B)** Waterfall plot displays tumour somatic mutation information of the genes with high mutation frequencies in the high-risk subgroup **(A)** and low-risk subgroup **(B)**. Mutated genes (rows, top 20) are ordered by mutation rate; samples (columns) are arranged to emphasize mutual exclusivity among mutations. The right shows the mutation percentage, and the top shows the overall number of mutations. The color coding indicates the mutation type. **(C)** The co-expression patterns of top 20 mutated genes in CRC patients.

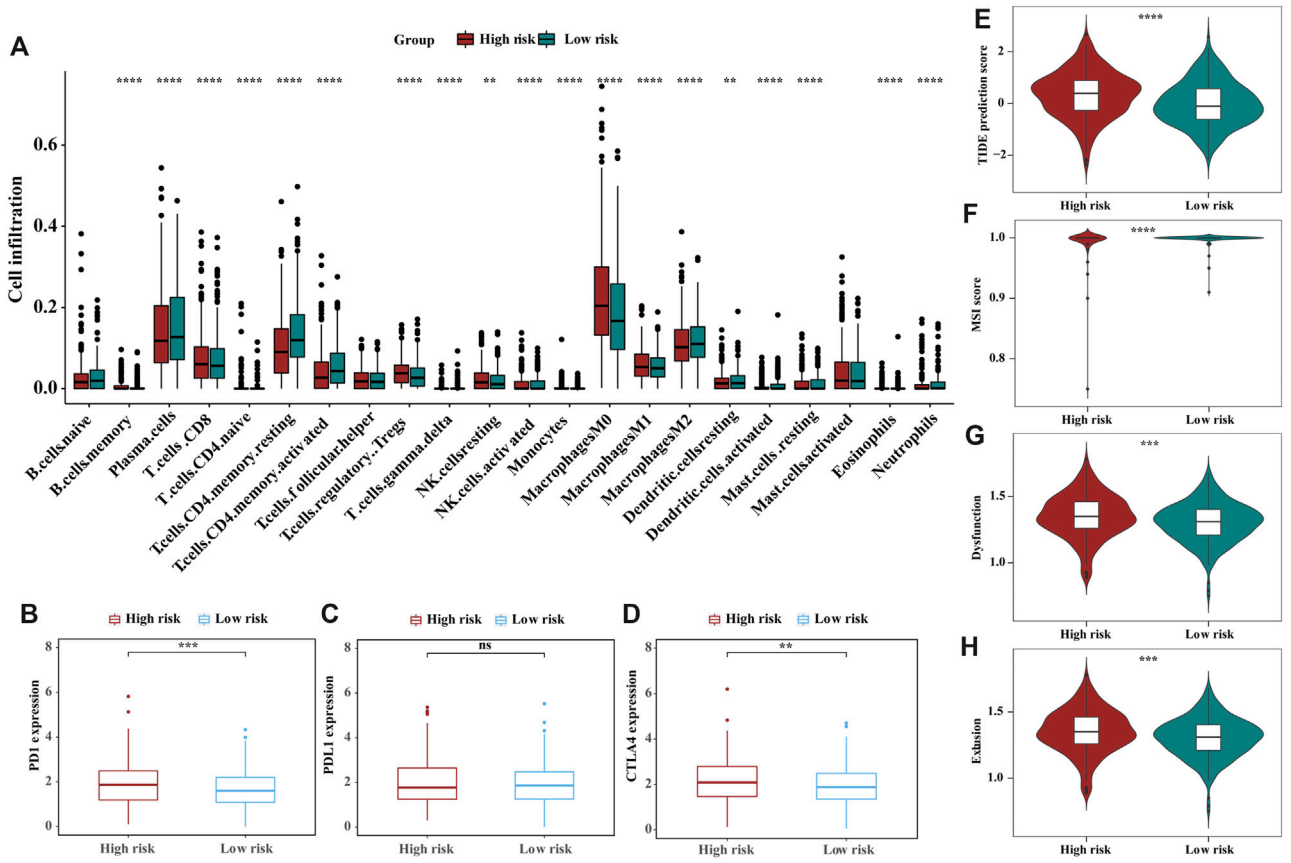


FIGURE 5
The landscape and estimation of the tumor immune microenvironment using the m6A-related gene signatures model. **(A)** The proportions of TME cells in different risk subgroups. Significant statistical differences between the two subgroups were assessed using the Wilcoxon test, the asterisks represented the statistical *p*-value (blank, not significant; **p* < 0.05; ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001). **(B–D)** Expression of the immune checkpoints PD1**(B)**, PDL1**(C)** and CTLA4 **(D)** between high- and low-risk groups. **(E–H)** TIDE **(E)**, MSI **(F)**, and T-cell exclusion **(G)** and dysfunction **(H)** score in the high- and low-risk patients. The scores between the two risk subgroups were compared through the Wilcoxon test (**p* < 0.05; ***p* < 0.01; ****p* < 0.001; ns, not significant).

prediction of the 1-, 3-, 5-year OS (Figures 7A–C), which displayed superior predictive performance over the model that only included clinical characteristics with AUCs of 0.808 (95%CI: 0.740–0.875), 0.793 (95%CI: 0.730–0.856) and 0.755 (95%CI: 0.665–0.845). Calibration plots showed that the observed vs predicted rates of 1-, 3-, 5-year OS had good concordance (Figures 7D–F). Accordingly, based on the risk score and clinical characteristics, a prognostic nomogram was established for the prediction of OS in CRC patients as shown in Figure 7G. The validation results (Supplementary Figure S11) were consistent with the findings in TCGA training set, which indicated that m6A-based model had a stable OS-predictive ability.

Discussion

Here, we developed a prognostic risk score based on m6A-related gene expression signatures and performed external validation to assess its prediction accuracy. Our study indicated that the m6A-based prognostic risk score was an independent predictor for CRC survival and had improved the prediction accuracy of CRC survival when combined with clinical characteristics. When stratified by this risk score, a worse survival rate, lower immunogenicity, and greater number of somatic mutations were shown in the high-risk group. The low-risk group had a lower TIDE score than the high-risk group for predicting immunotherapy response, implying that low-risk patients could benefit more from immunotherapy than high-risk patients.

Evidence from numerous studies have been discovered that RNA modifications regulate most steps of the gene expression, from DNA transcription to RNA translation (Helm and Motorin, 2017; Delaunay and Frye, 2019), through the effect of CNV and mutations to m6A regulators, including the alterations of RBM15, YTHDF2, YTHDC1, YTHDC2, and METTL14 (Zhang Q. et al., 2020). Besides, recent studies suggest that m6A modification plays important roles in RNA metabolism and cell proliferation, with significant implications on a variety of cell-physiological processes and cancer development (Zhang C. et al., 2020; Tian et al., 2020). In our study, functional enrichment analyses indicated that *CDK5R4P2*, *CLK1*, *CNOT3*, *GPR125*,

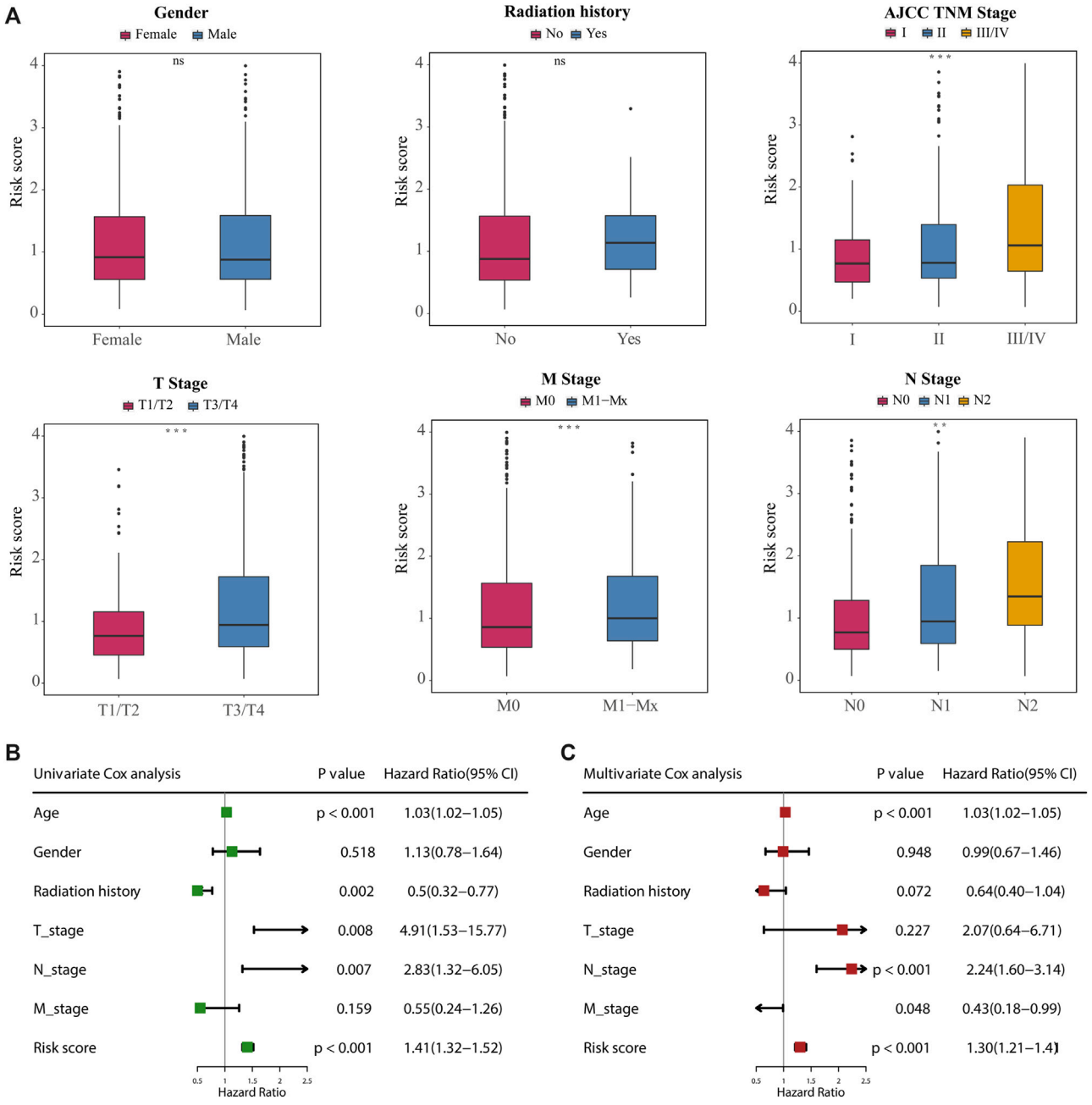


FIGURE 6
Correlation between the 18-gene expression signatures and clinical characteristics. **(A)** Difference analysis of the distribution of risk scores in different T, N, M, AJCC TNM stages, gender, and radiation history. Statistical difference of two groups was compared by the Wilcoxon test and three or more groups were compared by the Kruskal–Wallis test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns not significant). **(B, C)** Univariate **(B)** and multivariate **(C)** Cox regression analyses of correlations between the 18-gene expression signatures and clinical characteristics with OS, and revealed that the risk score based on the m6A-related gene expression signatures was an independent prognostic predictor in the TCGA dataset.

ING5, *SFPQ* and *UBE2H* are mainly involved in the neural, destabilization and metabolic processes of mRNA signatures, and influence the growth, differentiation and communications of multiple colon cell types. Interestingly, *GPR125* and *SFPQ* were enriched in a neural signaling pathway in relation to Spinocerebellar ataxia. Additionally, *CLK1,* a novel CLK kinases inhibitor, has been reported to impair the growth of CRC cell lines and organoids, and inhibit anchorage-independent colony formation and cell migration, thus promoting cytotoxicity (Sohail et al., 2021). *UBE2H* belongs to the ubiquitin-conjugating enzyme (*UBE2*) family, and there are several studies investing the role of UBE2 family in carcinogenesis, especially malignant breast cancer (Ayesha et al., 2016) and lung cancer (Jiang et al., 2017; Liu and Xu, 2018). Evidence from the existing research suggested that *MET-UBE2H* might be a novel prognostic biomarker or target in lung

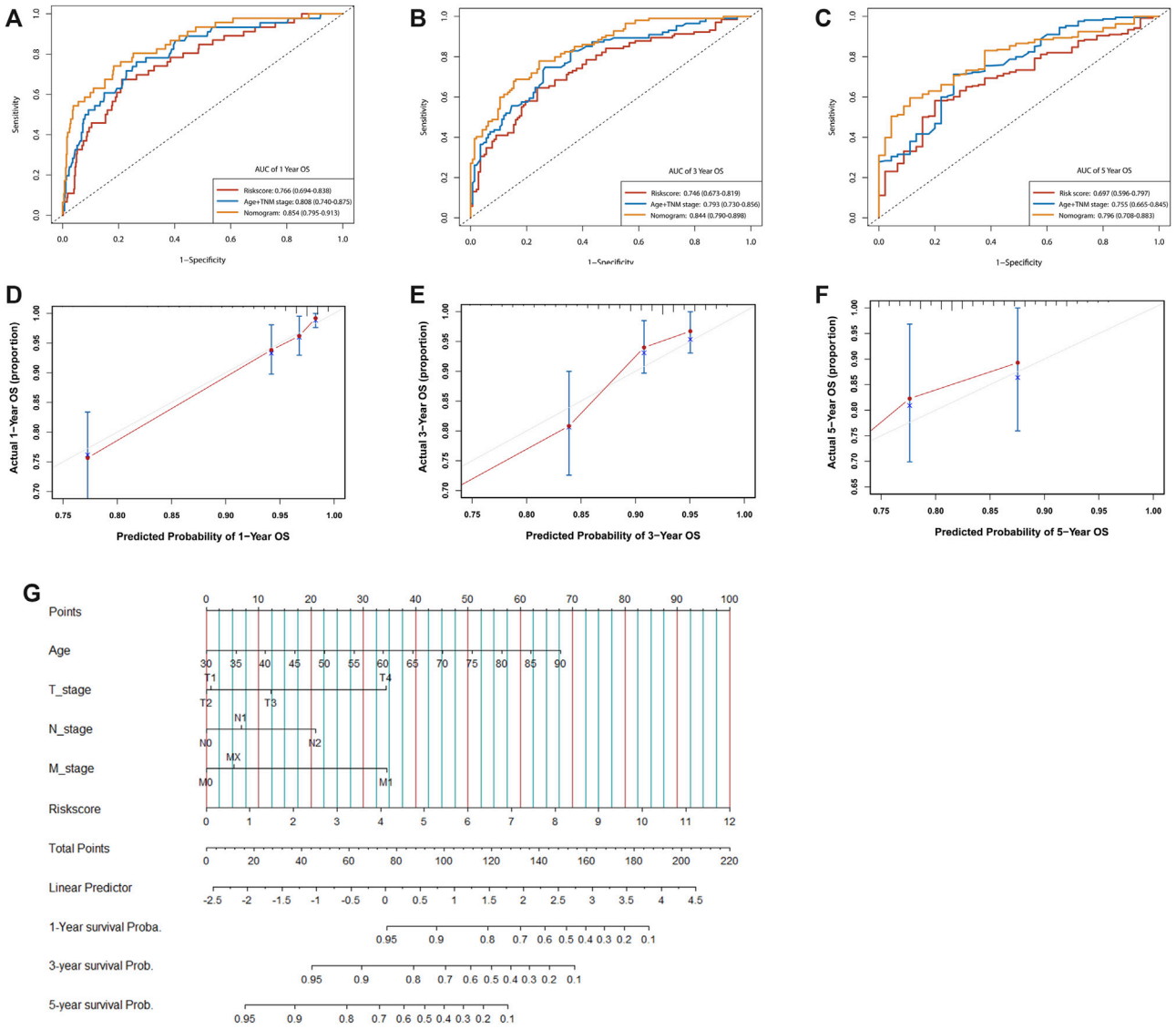


FIGURE 7
Assessment of the prognostic risk model of the m6A-related gene expression signatures and clinical features in CRC. **(A–C)** Time-dependent receiver operating characteristic (ROC) curves for the nomogram, risk score, and clinical characteristics in the TCGA dataset on predicting 1- **(A)**, 3- **(B)**, and 5-year **(C)** OS. **(D–F)** The calibration plot of the nomogram predicts the probability of the 1- **(D)**, 3- **(E)**, and 5- **(F)** year OS. **(G)** Nomogram for predicting the 1-, 3-, and 5-year OS of patients with CRC.

adenocarcinoma (Zhu et al., 2018). And *UBE2H* was also identified as an m6A-related hub gene closely related to the clinicopathology and prognosis of CRC using a prognostic signature model (Zhang and Zhang, 2021). In concordance with our findings, Cejas et al also found that *CNOT3* overexpression in colon tissues was associated with worse prognosis outcomes of CRC (Cejas et al., 2017).

Our study firstly developed a prognostic risk score based on 18 m6A-related gene expression signatures that could be used as an index to predict the OS of CRC patients, and further validated its predictive performance in two independent external datasets. The risk stratification analysis showed that the m6A-based prognostic risk score had a good prognostic accuracy in predicting the OS in both the TCGA and validation datasets. Time-dependent AUC also confirmed that combination of the m6A-related prognostic risk score with clinical predictors (TNM stage and age) displayed superior predictive performance over the model that only included clinical characteristics of OS for CRC patients. Stratified analysis also confirmed that the risk score could predict CRC survival with good performance in different clinical subgroups (age, T stage, AJCC TNM stage). Taken together, this m6A-based prognostic risk score could be used as an independent predictor for CRC survival and the application of risk score in combination with clinical characteristics could improve the prediction accuracy of CRC survival.

Using this m6A-related prognostic risk score as a classifier, CRC patients were stratified into low- and high-risk groups to

gain further biological insight into the gene mutations and immunologic nature of CRC patients in different risk groups. We found that m6A-related gene expression signatures were differentially enriched in the pathways related to cancer, immune response, and neural signaling between the two groups. When examining the somatic mutations, we found that the top 20 cancer driver genes mutated more frequently in the high-risk group than in the low-risk group, and significant co-occurrences were also observed among mutations of these genes. By examining the immunologic nature of CRC patients in different risk groups, we found high-risk group generally had higher monocytes and macrophages M1 infiltration and fractions of T cells CD8, and lower memory resting CD4 T cells than low-risk patients. Evidence from experimental studies observed that the infiltration levels of $CD8^+$T increased in YTHDF1-deficient mouse tumour, thereby enhancing an elevated antigen-specific $CD8^+$T cell antitumor response *in vivo* (Han et al., 2019). Additionally, it has been reported that CRC patients enriched with M1 phenotype and the high islet density of M1 macrophages would have poor prognosis (Zhang et al., 2012), as well as M0 macrophages (Zhang et al., 2021). And researches have revealed that the strategies converting M2 macrophages to M1 macrophages of tumour associated macrophages (TAMs) suppressed tumour growth (Dong et al., 2020). Of note, another study of human CRC specimens illustrated that those with high densities of $CD4^+$T were associated with a lower likelihood of tumour relapse and improved OS (Galon et al., 2006), which are consistent with the findings from our study. These indicate that the m6A-related gene expression signatures may modulate the TME phenotypes to influence the survival of CRC patients.

Emerging pieces of evidence showed that different TME phenotypes might have different degrees of benefit from immunotherapeutic treatment (Wang et al., 2019). It is reported that less immunogenic cancer cells are selected for during tumour development in immune-competent hosts to evade antitumor immune responses (Dunn et al., 2002), which may result in increased immunosuppressive cells (e.g., regulatory T cells and TAMs) and expression of immunosuppressive molecules (e.g., CTLA4 and PD1). As expected, we found that CTLA4 and PD1 expression levels was significantly higher in high-risk CRC patients. A Tumour Immune Dysfunction and Exclusion (TIDE) score has been increasingly used as an index for predicting immunotherapeutic response (Jiang et al., 2018). Consistently, using the TIDE algorithm, we estimated the immune response and found that patients in the low-risk group have a superior response to immunotherapy. Chemotherapy results indicated that the high-risk patients with CRC were more sensitive to 24 chemotherapies than low-risk patients. These results suggested that the poorer prognosis for high-risk patients could be due to higher immunosuppression in the TME, and that TME may influence the response of chemotherapy and immunotherapy. Based on these findings, this m6A-based risk score might also be used as an indicator for predicting immunotherapy response among CRC patients.

Our study also provides insight into the process and mechanism of m6A modification of gene expression signatures for future studies. However, we are also aware of several limitations in this study. Although the m6A-related gene signatures prognostic risk score showed superior performance on the prediction of CRC survival and the response to immunotherapy, it should be prospectively validated in real clinical settings and the clinicopathological factors should also be considered. Moreover, distribution of immune cells in this m6A-based classifier (e.g., T cells and macrophages infiltration) could be influenced by the difference in research datasets, sequencing method and sample size, and both the TIDE and MSI scores focused on the function and status of T cells, which could not fully reflect the complexity of the TME involved in the immunotherapeutic response. Thus, further observational and experimental studies should be performed to elucidate the accuracy of this prognostic risk score in the prediction of CRC survival, and to understand how targeted immunotherapy against m6A regulators could be applied in the clinic to achieve much improved cancer therapy in the future.

In conclusion, we developed a prognostic risk score based on the expression signature of 18 genes associated with m6A modification to predict the OS of CRC patients and their response to immunotherapy. This work highlights the clinical implications of this risk score in distinguishing immune and molecular characteristics and identifying response of target treatments. The derived m6A-related risk score showed the potential to be used as a prognostic and therapeutic indicator for the prediction of CRC prognosis and the development of individualized CRC treatment strategy.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

LY contributed to data acquisition analysis and interpretation, and drafted the paper; LW, JS, and XZ contributed to supervised statistical data analysis; YeH contributed to design, data acquisition and control; LiH contributed to design, data acquisition and control; YeH contributed to design, data acquisition and control, and critically revised the paper; CL and JC contributed to conception, design and critically revised the paper; XL and KD contributed to conception, design, supervision and critically revised the paper; ET and MD contributed to supervision and critically revised the paper. All authors gave final approval and agree to be accountable for all aspects of the work.

Funding

XL is supported by the Natural Science Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001) and the National Nature Science Foundation

of China (82204019). KD is supported by the Project of the regional diagnosis and treatment center of the Health Planning Committee (No. JBZX-201903). ET is supported by a CRUK Career Development Fellowship (C31250/A22804). This work is also funded by a grant to MGD as Project Leader with the MRC Human Genetics Unit Centre Grant (U127527198).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer ZZ declared a shared parent affiliation with the author(s) CL to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.885930/full#supplementary-material

References

Ayesha, A. K., Hyodo, T., Asano, E., Sato, N., Mansour, M. A., Ito, S., et al. (2016). UBE2S is associated with malignant characteristics of breast cancer cells. *Tumour Biol.* 37 (1), 763–772. doi:10.1007/s13277-015-3863-7

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24 (5), 541–550. doi:10.1038/s41591-018-0014-x

Bramsen, J. B., Rasmussen, M. H., Ongen, H., Mattesen, T. B., Ørntoft, M. W., Árnadóttir, S. S., et al. (2017). Molecular-subtype-specific biomarkers improve prediction of prognosis in colorectal cancer. *Cell Rep.* 19 (6), 1268–1280. doi:10.1016/j.celrep.2017.04.045

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487 (7407), 330–337. doi:10.1038/nature11252

Cejas, P., Cavazza, A., Yandava, C. N., Moreno, V., Horst, D., Moreno-Rubio, J., et al. (2017). Transcriptional regulator CNOT3 defines an aggressive colorectal cancer subtype. *Cancer Res.* 77 (3), 766–779. doi:10.1158/0008-5472.Can-16-1346

Chang, G., Leu, J. S., Ma, L., Xie, K., and Huang, S. (2019). Methylation of RNA N(6)-methyladenosine in modulation of cytokine responses and tumorigenesis. *Cytokine* 118, 35–41. doi:10.1016/j.cyto.2018.06.018

Chen, H., Gao, S., Liu, W., Wong, C. C., Wu, J., Wu, J., et al. (2021). RNA N(6)-methyladenosine methyltransferase METTL3 facilitates colorectal cancer by activating the m(6)A-GLUT1-mTORC1 Axis and is a therapeutic target. *Gastroenterology* 160 (4), 1284–1300.e16. doi:10.1053/j.gastro.2020.11.013

Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., et al. (2018). mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature* 561 (7724), 556–560. doi:10.1038/s41586-018-0538-8

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: An R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Delaunay, S., and Frye, M. (2019). RNA modifications regulating cell fate in cancer. *Nat. Cell Biol.* 21 (5), 552–559. doi:10.1038/s41556-019-0319-0

Dong, H., Yang, Y., Gao, C., Sun, H., Wang, H., Hong, C., et al. (2020). Lactoferrin-containing immunocomplex mediates antitumor effects by resetting tumor-associated macrophages to M1 phenotype. *J. Immunother. Cancer* 8 (1), e000339. doi:10.1136/jitc-2019-000339

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002). Cancer immunoediting: From immunosurveillance to tumor escape. *Nat. Immunol.* 3 (11), 991–998. doi:10.1038/ni1102-991

Fang, H., and Declerck, Y. A. (2013). Targeting the tumor microenvironment: From understanding pathways to effective clinical trials. *Cancer Res.* 73 (16), 4965–4977. doi:10.1158/0008-5472.Can-13-0661

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^{6}A RNA methylation. *Nat. Rev. Genet.* 15 (5), 293–306. doi:10.1038/nrg3724

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313 (5795), 1960–1964. doi:10.1126/science.1129139

Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468

Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour immunity controlled through mRNA m(6)A methylation and YTHDF1 in dendritic cells. *Nature* 566 (7743), 270–274. doi:10.1038/s41586-019-0916-x

Hanahan, D., and Coussens, L. M. (2012). Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21 (3), 309–322. doi:10.1016/j.ccr.2012.02.022

He, P. C., and He, C. (2021). m(6) A RNA methylation: from mechanisms to therapeutic potential. *Embo J.* 40 (3), e105977. doi:10.15252/embj.2020105977

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: Predict and validate. *Nat. Rev. Genet.* 18 (5), 275–291. doi:10.1038/nrg.2016.169

Hong, K. (2018). Emerging function of N6-methyladenosine in cancer. *Oncol. Lett.* 16 (5), 5519–5524. doi:10.3892/ol.2018.9395

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Jiang, Y., Huang, Y., Du, Y., Zhao, Y., Ren, J., Ma, S., et al. (2017). Identification of prognostic genes and pathways in lung adenocarcinoma using a bayesian approach. *Cancer Inf.* 16, 1176935116684825. doi:10.1177/1176935116684825

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Li, D., Zhu, X., Li, Y., and Zeng, X. (2020). Novel insights into the roles of RNA N(6)-methyladenosine modification in regulating gene expression during environmental exposures. *Chemosphere* 261, 127757. doi:10.1016/j.chemosphere.2020.127757

Li N., N., Kang, Y., Wang, L., Huff, S., Tang, R., Hui, H., et al. (2020). ALKBH5 regulates anti-PD-1 therapy response by modulating lactate and suppressive immune cell accumulation in tumor microenvironment. *Proc. Natl. Acad. Sci. U. S. A.* 117 (33), 20159–20170. doi:10.1073/pnas.1918986117

Li, T., Hu, P. S., Zuo, Z., Lin, J. F., Li, X., Wu, Q. N., et al. (2019). METTL3 facilitates tumor progression via an m(6)A-IGF2BP2-dependent mechanism in colorectal carcinoma. *Mol. Cancer* 18 (1), 112. doi:10.1186/s12943-019-1038-7

Littman, D. R. (2015). Releasing the brakes on cancer immunotherapy. *Cell* 162 (6), 1186–1190. doi:10.1016/j.cell.2015.08.038

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518 (7540), 560–564. doi:10.1038/nature14234

Liu, Z., and Xu, L. (2018). UBE2S promotes the proliferation and survival of human lung adenocarcinoma cells. *BMB Rep.* 51 (12), 642–647. doi:10.5483/BMBRep.2018.51.12.138

Lonberg, N., and Korman, A. J. (2017). Masterful antibodies: Checkpoint blockade. *Cancer Immunol. Res.* 5 (4), 275–281. doi:10.1158/2326-6066.Cir-17-0057

Luo, J., Liu, H., Luan, S., He, C., and Li, Z. (2018). Aberrant regulation of mRNA m^{6}A modification in cancer development. *Int. J. Mol. Sci.* 19 (9), 2515. doi:10.3390/ijms19092515

Ma, S., Chen, C., Ji, X., Liu, J., Zhou, Q., Wang, G., et al. (2019). The interplay between m6A RNA methylation and noncoding RNA in cancer. *J. Hematol. Oncol.* 12 (1), 121. doi:10.1186/s13045-019-0805-7

Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M. P., Vescovo, L., et al. (2013). Gene expression classification of colon cancer into molecular subtypes: Characterization, validation, and prognostic value. *PLoS Med.* 10 (5), e1001453. doi:10.1371/journal.pmed.1001453

Quail, D. F., and Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* 19 (11), 1423–1437. doi:10.1038/nm.3394

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359 (6382), 1350–1355. doi:10.1126/science.aar4060

Rong, D., Sun, G., Wu, F., Cheng, Y., Sun, G., Jiang, W., et al. (2021). Epigenetics: Roles and therapeutic implications of non-coding RNA modifications in human cancers. *Mol. Ther. Nucleic Acids* 25, 67–82. doi:10.1016/j.omtn.2021.04.021

Shen, C., Xuan, B., Yan, T., Ma, Y., Xu, P., Tian, X., et al. (2020). m(6)A-dependent glycolysis enhances colorectal cancer progression. *Mol. Cancer* 19 (1), 72. doi:10.1186/s12943-020-01190-w

Smith, J. J., Deane, N. G., Wu, F., Merchant, N. B., Zhang, B., Jiang, A., et al. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 138 (3), 958–968. doi:10.1053/j.gastro.2009.11.005

Sohail, M., Shkreta, L., Toutant, J., Rabea, S., Babeu, J. P., Huard, C., et al. (2021). A novel class of inhibitors that target SRSF10 and promote p53-mediated cytotoxicity on human colorectal cancer cells. *Nar. Cancer* 3 (2), zcab019. doi:10.1093/narcan/zcab019

Sun, T., Wu, R., and Ming, L. (2019). The role of m6A RNA methylation in cancer. *Biomed. Pharmacother.* 112, 108613. doi:10.1016/j.biopha.2019.108613

Tian, J., Ying, P., Ke, J., Zhu, Y., Yang, Y., Gong, Y., et al. (2020). ANKLE1 N(6)-Methyladenosine-related variant is associated with colorectal cancer risk by maintaining the genomic stability. *Int. J. Cancer* 146 (12), 3281–3293. doi:10.1002/ijc.32677

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* 366 (26), 2443–2454. doi:10.1056/NEJMoa1200690

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357 (6352), eaan2507. doi:10.1126/science.aan2507

Wang, S., He, Z., Wang, X., Li, H., and Liu, X. S. (2019). Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife* 8, e49020. doi:10.7554/eLife.49020

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505 (7481), 117–120. doi:10.1038/nature12730

Wieder, T., Eigentler, T., Brenner, E., and Röcken, M. (2018). Immune checkpoint blockade therapy. *J. Allergy Clin. Immunol.* 142 (5), 1403–1414. doi:10.1016/j.jaci.2018.02.042

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yang, Y., Hsu, P. J., Chen, Y. S., and Yang, Y. G. (2018). Dynamic transcriptomic m(6)A decoration: Writers, erasers, readers and functions in RNA metabolism. *Cell Res.* 28 (6), 616–624. doi:10.1038/s41422-018-0040-8

Yuan, C., Zhao, X., Wangmo, D., Alshareef, D., Gates, T. J., and Subramanian, S. (2021). Tumor models to assess immune response and tumor-microbiome interactions in colorectal cancer. *Pharmacol. Ther.* 231, 107981. doi:10.1016/j.pharmthera.2021.107981

Zhang, C., Chen, L., Peng, D., Jiang, A., He, Y., Zeng, Y., et al. (2020). METTL3 and N6-methyladenosine promote homologous recombination-mediated repair of DSBs by modulating DNA-RNA hybrid accumulation. *Mol. Cell* 79 (3), 425–442.e7. doi:10.1016/j.molcel.2020.06.017

Zhang, L., Wang, S., Wang, Y., Zhao, W., Zhang, Y., Zhang, N., et al. (2021). Effects of hypoxia in intestinal tumors on immune cell behavior in the tumor microenvironment. *Front. Immunol.* 12, 645320. doi:10.3389/fimmu.2021.645320

Zhang, Q., Cai, Y., Kurbatov, V., Khan, S. A., Lu, L., Zhang, Y., et al. (2020). Gene alterations of N6-methyladenosine (m(6)A) regulators in colorectal cancer: A TCGA database study. *Biomed. Res. Int.* 2020, 8826456. doi:10.1155/2020/8826456

Zhang, Q. W., Liu, L., Gong, C. Y., Shi, H. S., Zeng, Y. H., Wang, X. Z., et al. (2012). Prognostic significance of tumor-associated macrophages in solid tumor: A meta-analysis of the literature. *PLoS One* 7 (12), e50946. doi:10.1371/journal.pone.0050946

Zhang, Z., and Zhang, X. (2021). Identification of m6A-related biomarkers associated with prognosis of colorectal cancer. *Med. Sci. Monit.* 27, e932370. doi:10.12659/msm.932370

Zhong, L., Liao, D., Zhang, M., Zeng, C., Li, X., Zhang, R., et al. (2019). YTHDF2 suppresses cell proliferation and growth via destabilizing the EGFR mRNA in hepatocellular carcinoma. *Cancer Lett.* 442, 252–261. doi:10.1016/j.canlet.2018.11.006

Zhu, Y. C., Wang, W. X., Song, Z. B., Zhang, Q. X., Xu, C. W., Chen, G., et al. (2018). MET-UBE2H fusion as a novel mechanism of acquired EGFR resistance in lung adenocarcinoma. *J. Thorac. Oncol.* 13 (10), e202–e204. doi:10.1016/j.jtho.2018.05.009

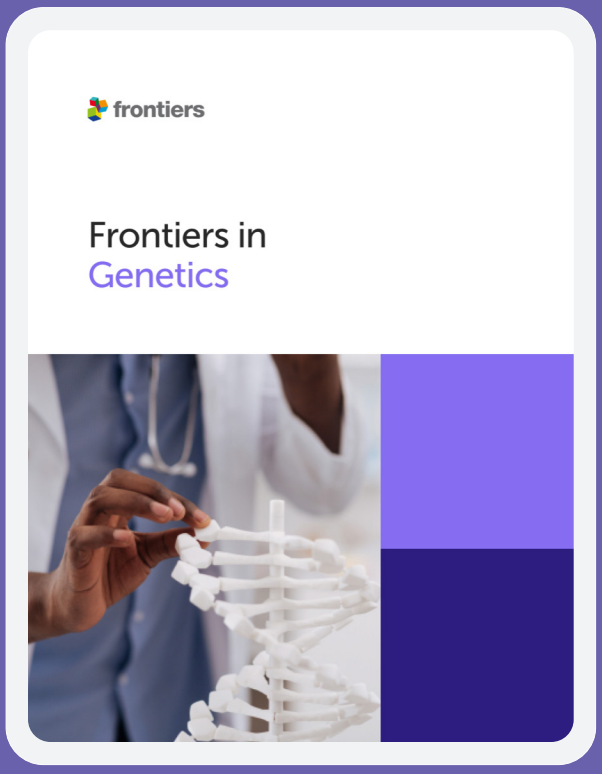
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact