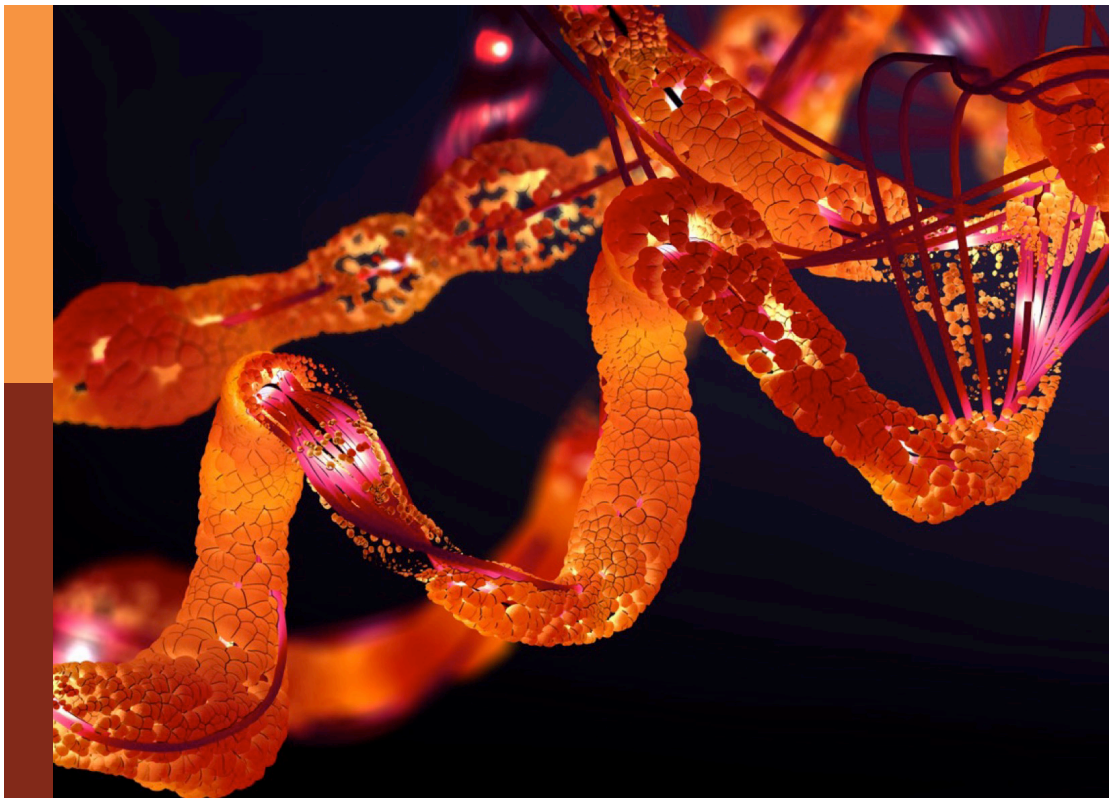
frontiers | Research Topics

Molecular biomarkers for the diagnosis, prognosis, and risk prediction of cancer

Edited by
Hossain Shekhar, Xin Wang, Ismail Hosen, Florence Le Calvez-Kelm, Tianhui Chen and William C. Cho

Published in
Frontiers in Molecular Biosciences



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 978-2-8325-2512-8
DOI 10.3389/978-2-8325-2512-8

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

Molecular biomarkers for the diagnosis, prognosis, and risk prediction of cancer

Topic editors

Hossain Shekhar — University of Dhaka, Bangladesh
Xin Wang — The Chinese University of Hong Kong, China
Ismail Hosen — University of Dhaka, Bangladesh
Florence Le Calvez-Kelm — International Agency For Research On Cancer (IARC), France
Tianhui Chen — Zhejiang Cancer Hospital, China
William C. Cho — QEH, Hong Kong, SAR China

Citation

Shekhar, H., Wang, X., Hosen, I., Calvez-Kelm, F. L., Chen, T., Cho, W. C., eds. (2023). *Molecular biomarkers for the diagnosis, prognosis, and risk prediction of cancer*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2512-8

Table of contents

08 m6A-Related lncRNAs Are Potential Biomarkers for the Prognosis of Metastatic Skin Cutaneous Melanoma
Suyang Huang, Shanshan Lyu, Zhe Gao, Weifeng Zha, Ping Wang, Yunyun Shan, Jianzhong He and Yang Li

22 Establishment of Novel Prostate Cancer Risk Subtypes and A Twelve-Gene Prognostic Model
Enchong Zhang, Fujisawa Shiori, Mo Zhang, Peng Wang, Jieqian He, Yuntian Ge, Yongsheng Song and Liping Shan

39 Analyzing and Validating the Prognostic Value of a TNF-Related Signature in Kidney Renal Clear Cell Carcinoma
Wenhao Zhang, Changjiu Li, Fanding Wu, Ning Li, Yuwei Wang, Yixuan Hu, Tiantian Fang, Hui Yuan and Huadong He

48 A Comprehensive Analysis of the Downregulation of *miRNA-1827* and Its Prognostic Significance by Targeting *SPTBN2* and *BCL2L1* in Ovarian Cancer
Penghui Feng, Zhitong Ge, Zaixin Guo, Lin Lin and Qi Yu

65 *BRCA1* and *BRCA2* Variation in Taiwanese General Population and the Cancer Cohort
Jiasheng Chian, Siddharth Sinha, Zixin Qin and San Ming Wang

74 Characterization of Hypoxia-Related Molecular Subtypes in Clear Cell Renal Cell Carcinoma to Aid Immunotherapy and Targeted Therapy *via* Multi-Omics Analysis
Weimin Zhong, Hongbin Zhong, Fengling Zhang, Chaoqun Huang, Yao Lin and Jiyi Huang

88 Analysis of the Prognostic Value and Gene Expression Mechanism of *SHOX2* in Lung Adenocarcinoma
Nanhong Li, Yu Zeng, Min Tai, Biyun Lin, Di Zhu, Yi Luo, Xinle Ren, Xiaoying Zhu, Lanlan Li, Hongrong Wu and Jian Huang

101 ANLN Enhances Triple-Negative Breast Cancer Stemness Through TWIST1 and BMP2 and Promotes its Spheroid Growth
Alishba Maryam and Y. Rebecca Chin

111 Hub Long Noncoding RNAs with m6A Modification for Signatures and Prognostic Values in Kidney Renal Clear Cell Carcinoma
Gaoteng Lin, Huadong Wang, Yuqi Wu, Keruo Wang and Gang Li

124 Comprehensive Molecular Analyses of a Novel Mutational Signature Classification System with Regard to Prognosis, Genomic Alterations, and Immune Landscape in Glioma
Zaoqu Liu, Taoyuan Lu, Libo Wang, Long Liu, Lifeng Li and Xinwei Han

139 Plasma Levels of Heat Shock Protein 90 Alpha Associated With Colorectal Cancer Development
Wene Wei, Jiahui Zhou, Lipeng Chen, Haizhou Liu, Fuyong Zhang, Jilin Li, Shufang Ning, Shirong Li, Chen Wang, Yi Huang, Chang Zou and Litu Zhang

148 Phenotypic and Functional Analyses of B7S1 in Ovarian Cancer
Dongli Cai, Fang Wang, Changgang Wang and Liping Jin

158 Identification of Arp2/3 Complex Subunits as Prognostic Biomarkers for Hepatocellular Carcinoma
Shenglan Huang, Dan Li, LingLing Zhuang, Liying Sun and Jianbing Wu

175 Implications of Lifestyle Factors and Polygenic Risk Score for Absolute Risk Prediction of Colorectal Neoplasm and Risk-Adapted Screening
Hongda Chen, Li Liu, Ming Lu, Yuhan Zhang, Bin Lu, Ying Zhu, Jianbo Tian, Xinying Li, Shaofa Nie, Xiaoping Miao and Min Dai

184 Annexin A3, a Calcium-Dependent Phospholipid-Binding Protein: Implication in Cancer
Liu Yang, Pingan Lu, Xiaohui Yang, Kaiguo Li and Song Qu

198 Identification of a Prognostic Signature Associated With the Homeobox Gene Family for Bladder Cancer
Bingqi Dong, Jiaming Liang, Ding Li, Wenping Song, Jinbo Song, Mingkai Zhu, Shiming Zhao, Yongkang Ma and Tiejun Yang

211 Molecular Characterization of Advanced Colorectal Cancer Using Serum Proteomics and Metabolomics
Jun Rao, Xianghui Wan, Fangfang Tou, Qinsi He, Aihua Xiong, Xinyi Chen, Wenhao Cui and Zhi Zheng

222 Overview on the Role of E-Cadherin in Gastric Cancer: Dysregulation and Clinical Implications
Huichen Zhao, Huihui Hu, Beibei Chen, Weifeng Xu, Jing Zhao, Chen Huang, Yishu Xing, Huifang Lv, Caiyun Nie, Jianzheng Wang, Yunduan He, Sai-Qi Wang and Xiao-Bing Chen

233 Prognostic Prediction, Immune Microenvironment, and Drug Resistance Value of Collagen Type I Alpha 1 Chain: From Gastrointestinal Cancers to Pan-Cancer Analysis
Yi Liu, Jinmin Xue, Maoxi Zhong, Zhi Wang, Jie Li and Yuxi Zhu

249 Genomic Analysis Reveals Heterogeneity Between Lesions in Synchronous Primary Right-Sided and Left-Sided Colon Cancer
Hanqing Hu, Qian Zhang, Rui Huang, Zhifeng Gao, Ziming Yuan, Qingchao Tang, Feng Gao, Meng Wang, Weiyuan Zhang, Tianyi Ma, Tianyu Qiao, Yinghu Jin and Guiyu Wang

260 CUX1 Facilitates the Development of Oncogenic Properties *Via* Activating Wnt/β-Catenin Signaling Pathway in Glioma
Fan Feng, Zongqing Zhao, Yunfei Zhou, Yanhao Cheng, Xiujie Wu and Xueyuan Heng

276 Meta-Analysis of the Diagnostic Value of Cell-free DNA for Renal Cancer
Yipeng Xu, Yingjun Jiang, Mingke Yu, Jianmin Lou, Mei Song, Han Xu, Yingying Cui, Xiaowei Zeng, Qibo Wang, Hanyun Ma, Zongping Wang, Shaoxing Zhu, Guorong Li and An Zhao

284 The Oncogenic Roles of JC Virus T Antigen in Breast Carcinogenesis
Hua-chuan Zheng, Ying E, Zheng-guo Cui, Shuang Zhao and Yong Zhang

292 Identification of Potential Genomic Alterations and the circRNA-miRNA-mRNA Regulatory Network in Primary and Recurrent Synovial Sarcomas
Qing Yao, Yong-Lai He, Ning Wang, Shuang-Shuang Dong, Mei Er Tu He Ta Mi Shi, Xiao Feng, Hao Chen, Li-Juan Pang, Hong Zou, Wen-Hu Zhou, Feng Li and Yan Qi

303 Pan-Cancer Analysis Identified *C1ORF112* as a Potential Biomarker for Multiple Tumor Types
Jiaxuan Chen, Haoming Mai, Haitao Chen, Bin Zhou, Jinlin Hou and De-Ke Jiang

314 SMPDL3B Predicts Poor Prognosis and Contributes to Development of Acute Myeloid Leukemia
Huiqing Qu and Ye Zhu

325 Mass Spectrometry-Based Targeted Serum Monomethylated Ribonucleosides Profiling for Early Detection of Breast Cancer
Zhihao Fang, Yiqiu Hu, Jiani Chen, Kailun Xu, Kailai Wang, Shu Zheng and Cheng Guo

333 CDK13-Mediated Cell Cycle Disorder Promotes Tumorigenesis of High HMGA2 Expression Gastric Cancer
Zhouying Wu, Min Wang, Feng Li, Feng Wang, Jianchao Jia, Zongqi Feng, Xue Huo, Jie Yang, Wen Jin, Rina Sa, Wenming Gao and Lan Yu

346 CD74 Correlated With Malignancies and Immune Microenvironment in Gliomas
Shengchao Xu, Xizhe Li, Lu Tang, Zhixiong Liu, Kui Yang and Quan Cheng

361 Comprehensive Analysis of HDAC Family Identifies *HDAC1* as a Prognostic and Immune Infiltration Indicator and *HDAC1*-Related Signature for Prognosis in Glioma
Yuxiang Fan, Xinyu Peng, Yubo Wang, Baoqin Li and Gang Zhao

378 Upregulation of Translationally Controlled Tumor Protein Is Associated With Cervical Cancer Progression
Xiaoyu Zhu, Ji Ren, Dianqin Xu, Di Cheng, Wei Wang, Jie Ren, Ziwen Xiao, Hongmei Jiang, Yan Ding and Yujie Tan

389 SNX20AR/MiRNA-301a-3p/SNX20 Axis Associated With Cell Proliferation and Immune Infiltration in Lung Adenocarcinoma
Yixiao Yuan, Xiulin Jiang, Lin Tang, Juan Wang, Qianqian Liu, Xiaolan Zou and Lincan Duan

412 Characterization of METTL7B to Evaluate TME and Predict Prognosis by Integrative Analysis of Multi-Omics Data in Glioma
Xiaochuan Chen, Chao Li, Ying Li, Shihong Wu, Wei Liu, Ting Lin, Miaomiao Li, Youliang Weng, Wanzun Lin and Sufang Qiu

429 Significance of Parkinson Family Genes in the Prognosis and Treatment Outcome Prediction for Lung Adenocarcinoma
Yanqi Li, Xiao Lu, Jiao Zhang, Quanxing Liu, Dong Zhou, Xufeng Deng, Yuan Qiu, Qian Chen, Manyuan Li, Guixue Yang, Hong Zheng and Jigang Dai

443 Immunohistochemical Analysis of CYP11B2, CYP11B1 and β-catenin Helps Subtyping and Relates With Clinical Characteristics of Unilateral Primary Aldosteronism
Luyan Sun, Yiran Jiang, Jing Xie, Hongyuan Zhu, Luming Wu, Xu Zhong, Weiwei Zhou, Tingwei Su and Weiqing Wang

454 Identification of Mast Cell-Based Molecular Subtypes and a Predictive Signature in Clear Cell Renal Cell Carcinoma
Hanxiang Liu and Yi Yang

469 Clinical Application Value of Circulating Cell-free DNA in Hepatocellular Carcinoma
Yuyuan Zhang, Zaoqu Liu, Kun Ji, Xin Li, Caihong Wang, Zhigang Ren, Yang Liu, Xinju Chen, Xinwei Han, Lingfang Meng, Lifeng Li and Zhen Li

479 Development and Verification of an Immune-Related Gene Pairs Prognostic Signature in Hepatocellular Carcinoma
Xiaofei Feng, Shanshan Mu, Yao Ma and Wenji Wang

490 Pan-Cancer Analysis of the Associations of TGFBI Expression With Prognosis and Immune Characteristics
Yun Chen, Han Zhao, Yao Feng, Qin Ye, Jing Hu, Yue Guo and Yunzhi Feng

507 Prognostic Value of N6-Methyladenosine-Related lncRNAs in Early-Stage Colorectal Cancer: Association With Immune Cell Infiltration and Chemotherapeutic Drug Sensitivity
Zhizhong Xiong, Xianzhe Li, Shi Yin, Minghao Xie, Chaobin Mao, Fengxiang Zhang, Huaxian Chen, Longyang Jin and Lei Lian

521 *SYDE1* Acts as an Oncogene in Glioma and has Diagnostic and Prognostic Values
Zhenyuan Han, Xiaodong Zhuang, Biao Yang, Lihui Jin, Pengjie Hong, Junqing Xue, Shunjie Chen and Zhen Tian

538 ARHGAP11A Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric Cancer
Biao Fan, Ke Ji, Zhaode Bu, Ji Zhang, Heli Yang, Jialin Li and Xiaojiang Wu

550 Aberrant Methylation and Immune Microenvironment Are Associated With Overexpressed Fibronectin 1: A Diagnostic and Prognostic Target in Head and Neck Squamous Cell Carcinoma
Surui Sheng, Bing Guo, Zhentao Wang, Zhihua Zhang, Jieyu Zhou and Zirong Huo

562 Molecular and Clinical Characteristics of Primary Pulmonary Lymphoepithelioma-Like Carcinoma
Ying Fan, Qianyun Shan, Jiali Gong, Jing Qin and Hongyang Lu

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 05 May 2021
doi: 10.3389/fmolb.2021.687760

m6A-Related lncRNAs Are Potential Biomarkers for the Prognosis of Metastatic Skin Cutaneous Melanoma

Suyang Huang[1†], Shanshan Lyu[2†], Zhe Gao[1], Weifeng Zha[1], Ping Wang[1], Yunyun Shan[1], Jianzhong He[3] and Yang Li[1]**

[1]Department of Dermatology, The Third People's Hospital of Hangzhou, Hangzhou, China, [2]Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [3]Department of Pathology, The Fifth Affiliated Hospital, Sun Yat-Sen University, Zhuhai, China

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong

Reviewed by:
Harpreet Kaur,
Independent researcher, New Orleans, LA, United States
Leping Li,
National Institute of Environmental Health Sciences (NIEHS), United States

***Correspondence:**
Jianzhong He
hejizh2010@163.com
Yang Li
13093786976@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 30 March 2021
Accepted: 21 April 2021
Published: 05 May 2021

Citation:
Huang S, Lyu S, Gao Z, Zha W, Wang P, Shan Y, He J and Li Y (2021) m6A-Related lncRNAs Are Potential Biomarkers for the Prognosis of Metastatic Skin Cutaneous Melanoma. Front. Mol. Biosci. 8:687760. doi: 10.3389/fmolb.2021.687760

Background: The incidence of skin cutaneous melanoma (SKCM) has risen more rapidly than any other solid tumor in the past few decades. The median survival for metastatic melanoma is only six to nine months and the 5°years survival rate of patients with conventional therapy is less than 5%. Our aim was to reveal the potential molecular mechanism in m6A modification of lncRNA and provide candidate prognostic biomarkers for metastatic SKCM.

Methods: lncRNAs expression level was obtained by re-annotation in TCGA and CCLE datasets. m6A-related lncRNAs were selected though correlation analysis. Univariate cox regression analysis was used to screen out independent prognostic factors. LASSO Cox regression was performed to construct an m6A-related lncRNA model (m6A-LncM). Univariate survival analysis and ROC curve were used to assess the prognostic efficacy of this model and candidate lncRNAs. Enrichment analysis was used to explore the candidate genes' functions.

Results: We obtained 1,086 common m6A-related lncRNAs after Pearson correlation analysis in both two datasets. 130 out of the 1,086 lncRNAs are independent prognostic factors. 24 crucial lncRNAs were filtered after LASSO Cox regression analysis. All the m6A-LncM and the 24 lncRNAs were related to overall survival. Stratified survival analysis of m6A-LncM showed that the model retains its prognostic efficacy in recurrence, radiation therapy and other subgroups. Enrichment analysis also found that these lncRNAs were immune associated.

Conclusion: Here, we obtained 24 crucial lncRNAs that may be potential biomarkers to predict survival of metastatic SKCM and may provide a new insight to improve the prognosis of it.

Keywords: prognostic biomarkers, lncRNAs, metastatic skin cutaneous melanoma, LASSO regression, N6-methylandenosine

INTRODUCTION

Based on the anatomical location, melanomas could be subdivided into limbic melanoma, skin cutaneous melanoma (SKCM), and mucosal melanoma (Curtin et al., 2005; Curtin et al., 2006), of which skin cutaneous melanoma is the major subtype of melanoma in Caucasians and the proportion of SKCM is roughly 20% of the Asian population (Chang et al., 1998; Chi et al., 2011). The incidence of SKCM has risen more rapidly than any other solid tumor in the past few decades (Eggermont et al., 2014). In recent years, researchers have made significant progress in understanding the biology, genetics, and treatment of SKCM. However, the prognosis remains poor due to the high rate of invasion and metastasis (Bsirini and Smoller, 2018). SKCM can metastasize extensively to the skin, subcutaneous, lymphatic system, lungs, and other non-pulmonary organs, so the patients with metastasis have poor survival rates and metastasis is also a major obstacle to improving prognosis (Balch et al., 2009; Domingues et al., 2018; Kremenovic et al., 2020). Therapies such as neoadjuvant treatment and targeted therapy have been shown to improve the prognosis of metastatic melanoma to some extent (Bai et al., 2019). The median survival for metastatic melanoma is only 6 to 9°months and the 5°years survival rate of patients with conventional therapy is less than 5% (Agarwala, 2009). Therefore, the search for effective biomarkers for prediction of prognosis and new therapeutic targets is urgent.

M6A is a mechanism of post-transcriptional modification of RNA prevalent in eukaryotic cells (Chen et al., 2019). m6A modification is involved in the degradation, translation, splicing, and other processes of mRNA (Liu et al., 2017; Chen et al., 2019; Liu and Gregory, 2019). m6A is frequently found in the 3′ UTR stop codon region and exon region, respectively (Dominissini et al., 2012; Meyer et al., 2012). The process of m6A modification is regulated by a variety of relevant factors (Jia et al., 2013). m6A regulators can be divided into three types based on previous research: 1) Writers that can recognize RNA and modify m6A, which includes *KIAA1429*, *METTL3*, *RBM15*, *METTL14*, *WTAP*, *METTL16*, and *ZC3H13* (Balacco and Soller, 2019); 2) Erasers, which are primarily responsible for the removal of m6A modifications, including *FTO* and *ALKNH5* (Liu et al., 2018; Pan et al., 2018) and; 3) Readers, including *HNRNPC*, *HNRNPA2B1*, *YTHDF1*, *YTHDC1*, *YTHDF2*, *YTHDC2*, and *YTHDF3* (Wang et al., 2018). The Readers could identify RNA methylation modifications and participate in RNA translation, degradation, and other processes (Ma et al., 2019). m6A affects many important life processes and is essential for cell division and proliferation, focal death and apoptosis (Zhou et al., 2019). Numerous studies have confirmed that aberrant m6A modifications play a key role in the genesis and development of a variety of tumors, including SKCM (He et al., 2021; Yu et al., 2021). For example, M6A methyltransferase *METTL3* is upregulated in melanoma and modulates melanoma cell invasiveness through *MMP2* (Dahal et al., 2019). *FTO* can promote bladder cancer by regulating the *miR-384*, *MALAT*, *MAL2* axis through m6A modifications (Tao et al., 2021). The *IGF2BP2*, *LINC00460*, and *DHX9* complex could promote colorectal cancer proliferation and metastasis by mediating the stability of *HMGA1* (Hou et al., 2021).

LncRNAs are transcripts with more than 200 nucleotides and with non-coding potential (Ramilowski et al., 2020). The function of lncRNAs remains largely unknown. LncRNA may be involved in the regulation of mRNA expression through translational regulation, histone modifications, or post-transcriptional (Huang et al., 2018). LncRNAs can influence various aspects of tumor cells including survival, proliferation, and migration though participating in gene regulation (Ramilowski et al., 2020). Aberrant expression of lncRNA is associated with tumor malignancy and has been shown to play a key role in the development of numerous cancers including SKCM. For instance, lncRNA *TTN-AS1* promotes SKCM development and metastasis by maintaining *TTN* expression (Wang et al., 2020). LncRNA *HCP5* inhibits the development of SKCM through regulation of *miR-12* expression (Wei et al., 2019).

In this article, we identified m6A modifications related lncRNAs and explored their prognostic ability by bioinformatic analysis and finally obtained potential biomarkers which can predict SKCM prognosis. We also established an important m6A related lncRNA-mRNA regulatory network to provide a new insight to investigate the mechanism of SKCM (**Figure 1**).

MATERIALS AND METHODS

Ethical Compliance

Public data was used for this study and there are no ethical issues.

Data Sources

Expression data of 484 SKCM patients and corresponding clinical characteristics were obtained from the GDC data portal using gdc-client (https://gdc-portal.nci.nih.gov/). As described in the Introduction, SKCM has a high rate of metastasis and a poor survival rate of patients with metastasis. Thus, our goal is developing the prognostic biomarkers that could predict the poor survival of metastatic SKCM. Then, we filtered 221 patients who with metastatic status for further analysis (**Table 1**). We also extracted the expression profile of 38 SKCM cell lines through CCLE database. All the gene expression profiles were quantified by FPKM and normalized though log2-based transformation. In addition, the expression level of 21 m6A-related genes (*FTO, ALKBH5, RBM15, RBM15B, METTL3, METTL14, METTL16, WTAP, VIRMA, ZC3H13, HNRNPC, HNRNPA2B1, IGF2BP1, IGF2BP2, IGF2BP3, RBMX, YTHDC1, YTHDC2*, *YTHDF1*, *YTHDF2*, and *YTHDF3*) (Zaccara et al., 2019) were constructed from the two datasets, respectively.

Re-Annotation of lncRNAs

Annotation information of lncRNAs were downloaded from GENECODE (https://www.gencodegenes.org/human/) and the genome annotation file (GRCh38) was downloaded from the UCSC database (http://hgdownload.cse.ucsc.edu/). Based on the

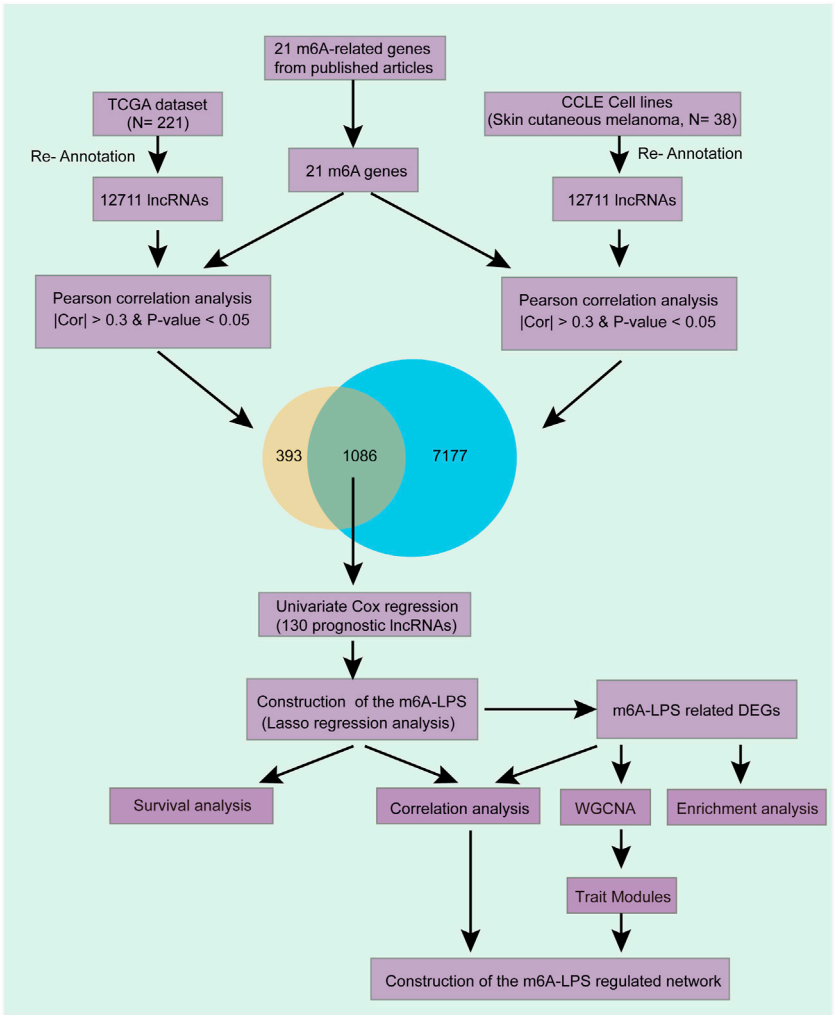


FIGURE 1 | Overview of the comprehensive analysis. We first obtained 12,711 lncRNAs expression level by re-annotation the TCGA and CCLE SKCM data. Then, 1,086 lncRNAs which correlated with at least one of the 21 m6A genes in both datasets after correlation analysis were filtered. Then, 130 out of the 1,086 lncRNAs are independent prognostic factors after Univariate Cox regression and had been selected for further analysis. We used LASSO Cox regression analysis to further filter a model that contained 24 crucial lncRNAs and have an excellent prognostic efficacy. Stratified survival analysis further testified the model's prognostic efficacy. Finally, enrichment analysis, WGCNA, correlation analysis were used to explored these lncRNAs' potential function in SKCM.

annotation information and Ensemble IDs, we obtained 12,711 lncRNAs in TCGA dataset and CCLE dataset.

Correlation Analysis Between lncRNAs and m6A-Related Genes

Pearson analysis was used to evaluate the correlation between these genes and lncRNAs based on the expression level of lncRNAs and the 21 m6A-related genes. LncRNAs with p value < 0.05 and an absolute Pearson correlation coeffcient (PCC) ≥ 0.3 were selected as m6A-related lncRNAs. We then also used Spearman correlation to double-check the correlation (p value < 0.05, absolute correlation coefficient ≥0.3). LncRNAs that were significantly associated with at least one of the 21 m6A genes in both data sets were selected for subsequent analysis.

Predict the Interactions Between lncRNAs and m6A Regulators and Predict m6A Modification Sites of lncRNAs

RNAInter (RNA Interactome Database) was used to predict the interactions between these lncRNAs and m6A regulators based on the RNA sequence (Lin et al., 2020). SRAMP database was used to predict m6A modification sites of the lncRNAs (Zhou et al., 2016).

Univariate Cox Regression Analysis and Lasso Analysis

The univariate cox regression analysis was used to screen out the prognostic lncRNAs. Subsequently, the R package *glmnet* (Friedman et al., 2010) was used to construct an m6A-related

lncRNA prognostic model (m6A-LncM) of SKCM patients by LASSO Cox regression. The riskscore of the LASSO regression model could be calculated as follow:

$$\text{Riskscore} = \sum_{i=1}^{n} \beta_i * E_i$$

where E_i is the expression value of the i gene in the model, and βi is the coefficient calculated by LASSO.

Differential Expression Analysis

Based on the riskscore of m6A-LncM, we classified the 221 SKCM patients into high or low risk scores groups. Then, differential expression analysis between the high and low groups was performed using the *limma* R package. The differentially expressed genes (DEGs) were identified with an adjusted p value < 0.05 and an absolute log2 fold change ≥0.585 (1.5 fold change).

Survival Analysis

Kaplan-Meier analysis was used to evaluate the prognostic efficiency of m6A-LncM and candidate lncRNAs. Survival curves reflect the relationship between the survival model or lncRNA expression level and SKCM patients' survival status though the survival R package. The *pROC* package was used to calculate the area under the ROC curve (AUC) of the m6A-LncM to assess the prognostic efficiency of it. We also compare the overall survival information between different subgroups of SKCM patients based on the riskscore of the model. The subgroups separated by the following features: age (≤57 or >57°years), neoadjuvant treatment (Yes or No), gender (male or female), melanoma clark level (I, II, III or IV, V), Recurrence (Yes or No), pathologic stage (I, II or III, IV), neoplasm cancer status (with tumor of tumor free), radiation therapy (Yes or No), and tumor issue site (regional or distant). *Caret* R package was used to randomly separate the 221 samples in two self-dependent test datasets (1:1) which were used to testify the survival efficacy of these lncRNAs.

WGCNA Network Construction and Module Identification

First, the co-expression network of m6A-LncM-related DEGs was constructed by an automatic network construction function in *WGCNA* R package. Second, co-expression modules were detected by the hierarchical clustering function. Then, modules were associated with clinical characteristics by calculating gene significance (GS) and module membership (MM). Finally, the candidate module with key genes were selected for further analysis.

Construction of the Co-Expression Network

The co-expression network between m6A related genes, m6A-LncM and riskscore-based DEGs was constructed based on the PCC that calculated by Pearson analysis. Those dysregulated lncRNA-mRNA pairs with an absolute PCC ≥ 0.3 and p value < 0.05 were selected to construct the co-expression network.

Function Enrichment Analysis

All the DEGs in the study were extracted for further functional enrichment by using the *clusterProfile* R package and Metascape software. The lncRNAs' potential functions were obtained from ImmLnc database (Li et al., 2020). Functions with a false discovery rate <0.05 were selected.

RESULTS

Screen m6A-Related lncRNAs in Metastatic SKCM Patients and Cell Lines

A total of 12,711 lncRNAs' expression levels were re-annotated in 221 SKCM patients and 38 SKCM cell lines. Then, we got the expression levels of 21 m6A-associated genes in the two datasets separately and performed Pearson analysis between the 21 genes and the 12,711 lncRNAs. The lncRNAs which exceed the threshold value (p value < 0.05 and |PCC| > 0.3) were defined as m6A-related lncRNAs. Finally, we obtained 1,479 lncRNAs that were significantly associated with m6A in 221 SKCM patients and 8,263 eligible ones in 38 SKCM cell lines. Finally, 1,086 lncRNAs that were significantly associated with at least one of the 21 m6A genes in both data sets were selected for subsequent analysis (**Figure 1** and **Supplementary Table S1**).

Identification of Potential Prognostic lncRNAs and Construct the m6A-LncM

Combined with clinical information, we used univariate Cox regression analysis to filter independent prognostic lncRNAs from the 1,086 lncRNAs related to m6A (p value < 0.05). We obtained 130 out of the 1,086 lncRNAs which were significantly associated with overall survival (OS) of metastatic SKCM patients (**Supplementary Table S2**).

Next, LASSO analysis was used on the 130 prognostic lncRNAs to generate an m6A-associated lncRNA model (m6A-LncM) which contains 24 lncRNAs (**Figures 2A,B**). The risk score of each sample in the dataset was calculated based on the coefficient of the 24 lncRNAs (**Figure 2C**). The 24 lncRNAs' PCC were showed in **Figure 2D**. We then also used Spearman correlation to double-check the correlation. The result showed that all the 24 lncRNAs were significantly associated with at least one of the 21 m6A genes (p value < 0.05, absolute correlation coefficient ≥0.3) (**Supplementary Figure S1A** and **Table 2**).

Then, based on the median risk score, we divided the patients into high- and low-risk score subgroups. The survival curve hinted that patients whose risk scores were high had a worse survival rate (**Figure 2E**). The ROC curves also hint that the m6A-LncM had a good efficacy to predict overall survival of metastatic SKCM patients (**Figure 2F**). We also performed multivariate survival analysis (cox proportional hazard models by combining the expression of the 24 lncRNAs and OS) to testify the survival efficacy of these lncRNAs. The result showed that they remained significant in multivariate survival models (**Supplementary Figure S1B**).

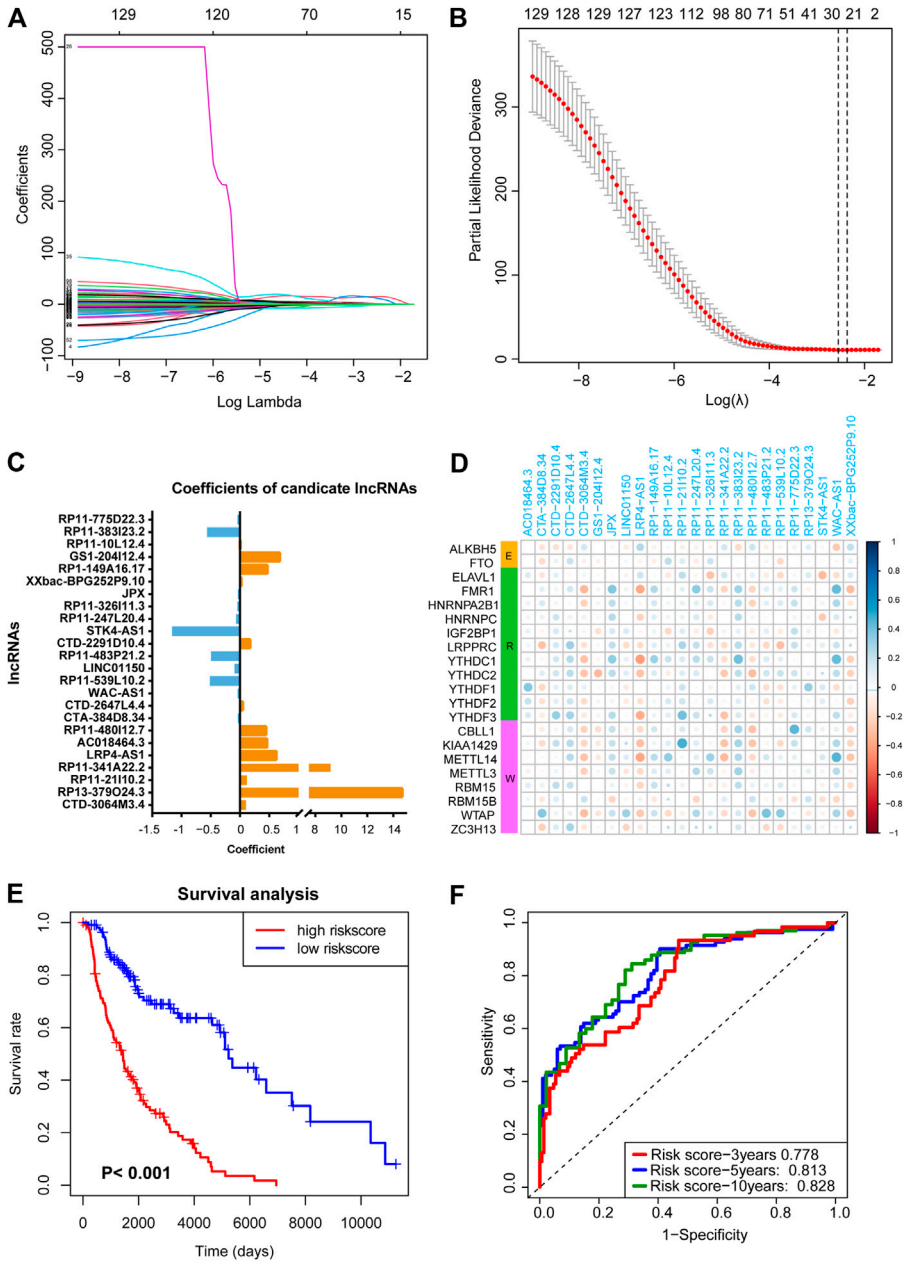


FIGURE 2 | LASSO regression analysis to find crucial survival model and lncRNAs. **(A,B)** The minimum criteria calculated though LASSO regression. **(C)** The coefficients of the survival model constructed by LASSO. **(D)** Heatmap showed the 24 prognostic m6A-related lncRNAs were correlated to the 21 m6A-related genes. **(E)** Survival curves hinted that the subgroup with highly risk score had worse OS rates compare with the low-risk subgroup. **(F)** AUC of the ROC analysis showed the predicted efficacy of m6A-LncM.

The 24 lncRNAs Have m6A Modification Sites and Have Interactions with m6A Regulators

We used RNAInter (RNA Interactome Database) to predict the interactions between these lncRNAs and m6A regulators based on the RNA sequence. We found that all the lncRNAs have interactions with m6A regulators (**Supplementary Table S3**). On the other hand, we also used SRAMP database to predict m6A modification sites based on the RNA sequences of these lncRNAs. We found that most of the lncRNAs have potential m6A modification sites except RP11-247L20.4 (**Supplementary Table S4**). The results of the two tools showed that the 24 lncRNAs are regulated by m6A modification and m6A regulators. It also illustrated that the correlation analysis that we performed is credible.

TABLE 1 | Clinicopathological characteristics of 221 Metastatic SKCM patients in TCGA dataset.

Clinical and pathological indices	Case No.	OS (%)	*p* value[a]
Specimens	221		
Mean age	57		
Age (years)			<0.001
≤57	111	43.2	
>57	110	34.5	
Gender			0.272
Male	139	36.0	
Female	82	43.9	
Pathologic stage			<0.001
I	62	56.5	
II	41	29.3	
III	85	41.2	
IV	10	40.0	
NA	23		
History of neoadjuvant treatment			0.881
Yes	15	40.0	
No	206	38.8	
Recurrence			<0.001
Yes	146	23.3	
No	75	69.3	
Radiation therapy			0.043
Yes	23	65.2	
No	198	35.9	

[a]*Log-rank test using the Kaplan Meier method; p value < 0.05 was considered significant.*

Survival Analysis of the Twenty-Four lncRNAs in m6A-LncM

The 24 crucial lncRNAs in this model were evaluated by survival analysis. The results showed that all these lncRNAs are survival associated (**Figure 3**). To be specific, 13 (LRP4-AS1, CTD-3064M3.4, RP11-341A22.2, GS1-204I12.4, XXbac-BPG252P9.10, RP11-480I12.7, RP1-149A16.17, CTD-2647L4.4, CTD-2291D10.4, RP13-379O24.3, RP11-10L12.4, AC018464.3, and RP11-21I10.2) out of the crucial 24 lncRNAs were risk factors of SKCM patients while the other 11 lncRNAs (STK4-AS1, RP11-326I11.3, LINC01150, JPX, RP11-247L20.4, RP11-539L10.2, CTA-384D8.34, RP11-483P21.2, RP11-383I23.2, WAC-AS1, and RP11-775D22.3) are protective factors (**Table 2**). The two self-dependent test datasets also verified their survival value. Both these 24 lncRNAs have good prognostic efficacy (**Supplementary Figures S2, S3**).

Stratified Survival Analysis of the m6A-LncM

To further assess the prognostic efficacy of m6A-LncM, we analyzed how the model's risk scores differed across various clinical traits and found that this model could distinguish melanoma clark level, recurrence, neoplasm cancer status, radiation therapy, and tumor issue site (**Figure 4A**). Therefore,

TABLE 2 | Basic information of 24 candidate lncRNAs.

Symbol	ENSG ID	Correlation with m6A regulators Zaccara et al. (2019) (Pearson)	Correlation with m6A regulators (Spearman)	Lasso model coefficients	Survival analysis	
					Role	*p* value
LRP4-AS1	ENSG00000247675	YTHDC1, METTL14, FMR1, YTHDF3	YTHDC1, FMR1	0.619	Risky	1.90E-07
CTD-3064M3.4	ENSG00000244998	FMR1, WTAP	FMR1	0.088	Risky	3.43E-05
RP11–341A22.2	ENSG00000225564	METTL14	METTL14	9.173	Risky	0.0041
STK4-AS1	ENSG00000227477	ELAVL1	ELAVL1	−1.148	Protective	0.0045
GS1-204I12.4	ENSG00000261729	YTHDC2	YTHDC2	0.679	Risky	0.00015
XXbac-BPG252P9.10	ENSG00000272273	FMR1	FMR1	0.019	Risky	0.0038
RP11–480I12.7	ENSG00000234996	YTHDC2	YTHDC2	0.444	Risky	0.0070
RP1-149A16.17	ENSG00000241954	YTHDC1	YTHDC1	0.466	Risky	0.032
RP11–326I11.3	ENSG00000271646	METTL14	METTL14	−0.034	Protective	0.0021
LINC01150	ENSG00000229671	WTAP	WTAP	−0.081	Protective	0.00017
CTD-2647L4.4	ENSG00000259366	LRPPRC	LRPPRC, ZC3H13	0.045	Risky	0.00013
CTD-2291D10.4	ENSG00000267886	YTHDF3	YTHDF3	0.170	Risky	1.52E-07
RP13-379O24.3	ENSG00000272050	YTHDF1	YTHDF1	14.678	Risky	0.0038
RP11-10L12.4	ENSG00000246560	METTL14	METTL14	0.005	Risky	0.0065
JPX	ENSG00000225470	YTHDC1, FMR1	YTHDC1, FMR1, METTL14	−0.023	Protective	0.0062
RP11–247L20.4	ENSG00000259071	FMR1	FMR1	−0.056	Protective	8.00E-04
RP11–539L10.2	ENSG00000246526	WTAP	WTAP	−0.502	Protective	5.57E-05
AC018464.3	ENSG00000234520	YTHDF1	YTHDF1, YTHDF2	0.460	Risky	0.0018
CTA-384D8.34	ENSG00000273272	WTAP	WTAP, LRPPRC	−0.010	Protective	5.56E-06
RP11–483P21.2	ENSG00000260228	WTAP	WTAP	−0.484	Protective	6.72E-05
RP11–21I10.2	ENSG00000249356	YTHDF3, KIAA1429	YTHDF3, KIAA1429,LRPPRC	0.105	Risky	0.00015
RP11–383I23.2	ENSG00000273374	YTHDC1	KIAA1429, YTHDC1,CBLL1, HNRNPA2B1,FMR1	−0.552	Protective	0.0043
WAC-AS1	ENSG00000254635	FMR1, YTHDC1,METTL14	FMR1, YTHDC1METTL14	−0.010	Protective	5.01E-05
RP11–775D22.3	ENSG00000240571	CBLL1	CBLL1	−0.031	Protective	0.0059

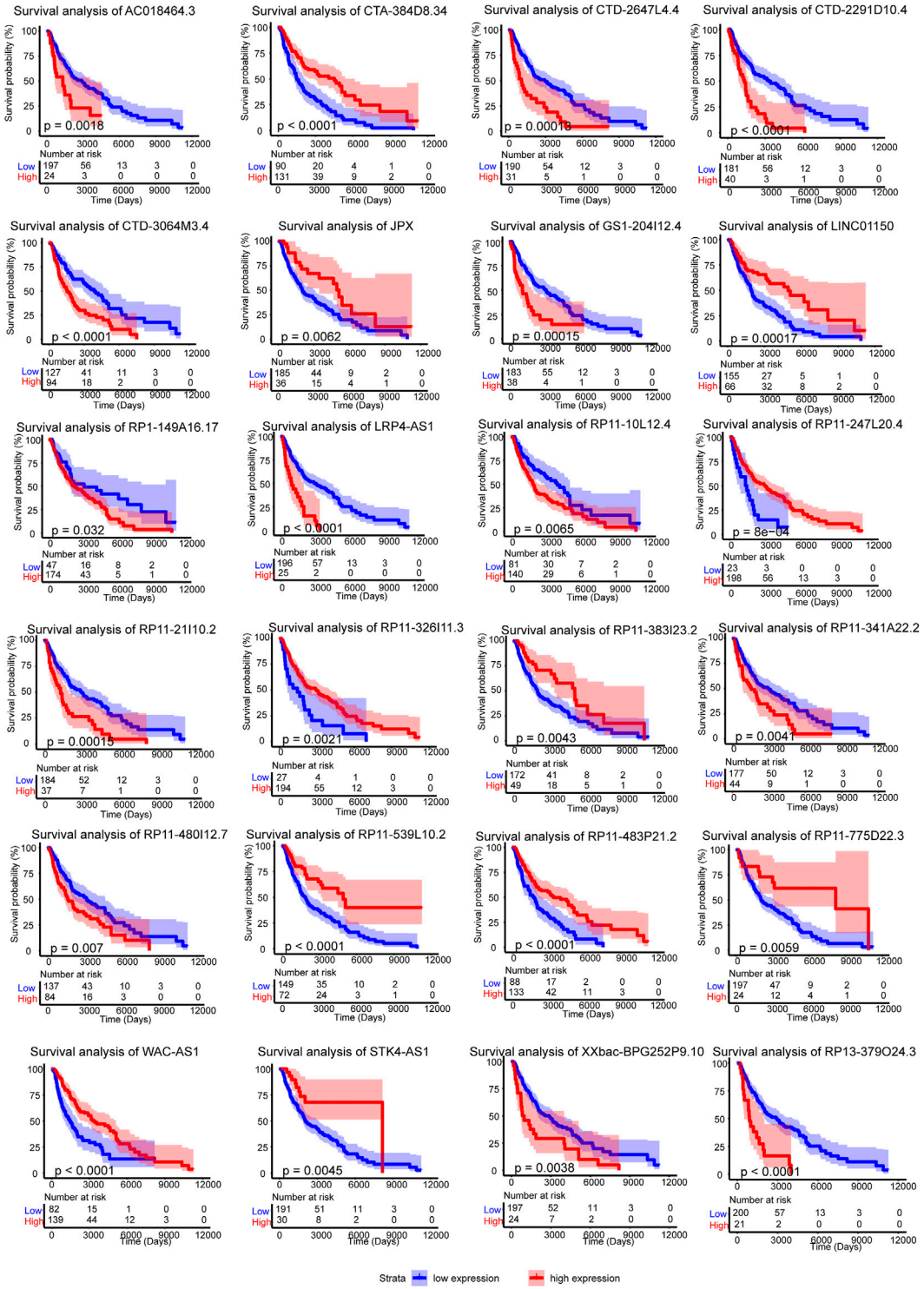


FIGURE 3 | Survival analysis of the 24 crucial lncRNAs in the m6A-LncM. Red line represented the high expression level patients when the blue is low expression. The light color areas represent 95% confidence intervals. The *p* value were calculated by Log-rank test.

we performed the stratified survival analysis to explore whether the model could predict OS in these subgroups. Interestingly, the higher risk metastatic SKCM patients had worse survival rate in all the melanoma clark levels (**Figures 4B,C**). Similarly, patients with high riskscores had a poor prognosis, regardless of recurrence (**Figures 4D,E**). When the neoplasm cancer status of patients is

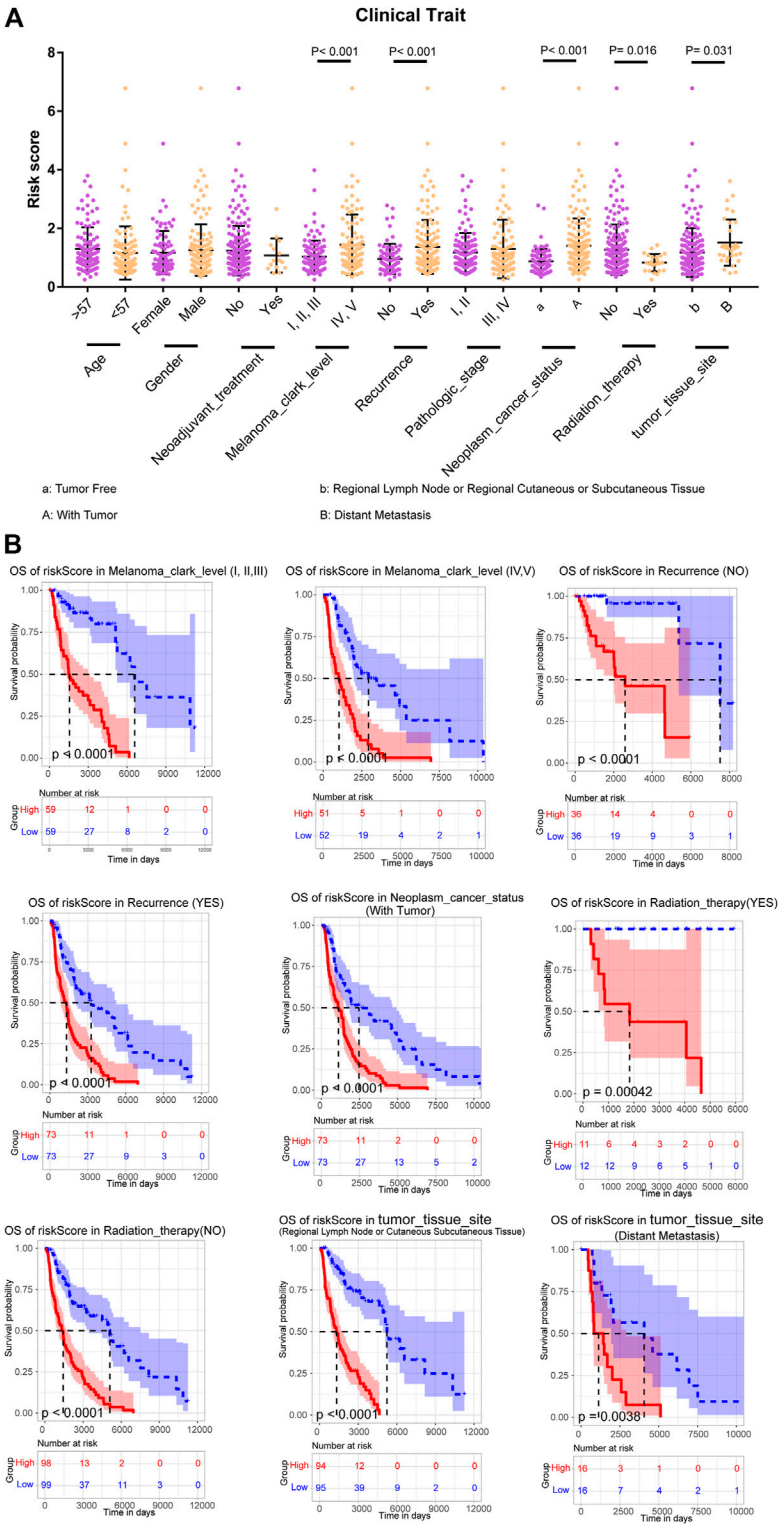


FIGURE 4 | Stratified survival analysis of the m6A-LncM. **(A)** The m6A-LncM could divide patients with different clinicopathological features into differential risk score subgroups. **(B)** The m6A-LncM still could predict the survival in multiple subgroups of metastatic SKCM patients. The red line represented the high riskscore patients when the blue represented low riskscore. The light color areas represent 95% confidence intervals. The *p* value were calculated by Log-rank test.

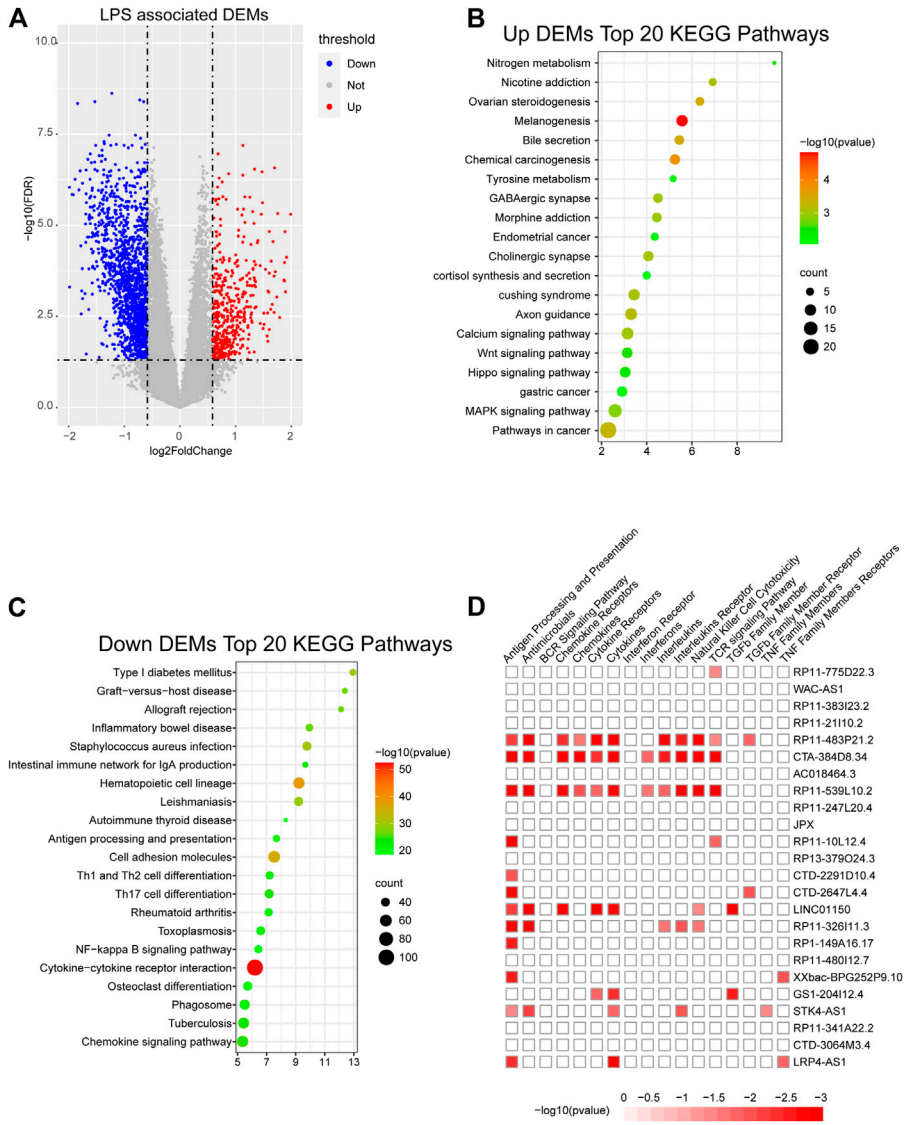


FIGURE 5 | Obtain m6A-LncM associated DEGs. **(A)** Volcano plot shows the differentially expressed genes (DEGs) between high riskscore samples and low riskscore samples. **(B, C)** The functional enrichment analysis of the up-regulated DEGs and down-regulated DEGs. **(D)** The functional enrichment analysis of 24 lncRNAs in m6A-LncM though ImmLnc database.

with tumor, the high riskscore was also associated with a lower survival rate (**Figure 4F**). We also confirmed that m6A-LncM retained its prognostic efficacy of patients with radiation therapy (**Figures 4G,H**). Patients with higher risk also had a worse survival rate in both the regional lymph node and distant metastasis subgroups (**Figure 4I**). These data proved that the model could be a novel predictor and performed an excellent prognostic efficacy in metastatic SKCM and may help to improve the prognosis of this cancer.

Enrichment Analysis of Key lncRNAs

To further confirm the potential functions of the m6A-LncM and crucial lncRNAs in SKCM, we classified the 221 patients into high- and low-risk score groups based on the model and obtained differentially expressed genes which are regulated by it (**Figure 5A**). The enrichment analysis showed that the upregulated DEGs were enriched in the Melanogenesis pathway, this further confirms that our model plays an important role in SKCM. Other cancer-related signaling pathways were also enriched, include Calcium signaling pathway, Wnt signaling pathway, Hippo signaling pathway, MAPK signaling pathway, and so on (**Figure 5B**). The downregulated DEGs were enriched in numerous immune-associated pathways such as Th1 and Th2 cell differentiation, intestinal immune network for IgA production and NF−kappa B signaling pathway (**Figure 5C**). Based on the result, we conjecture these lncRNAs may play a crucial role in the

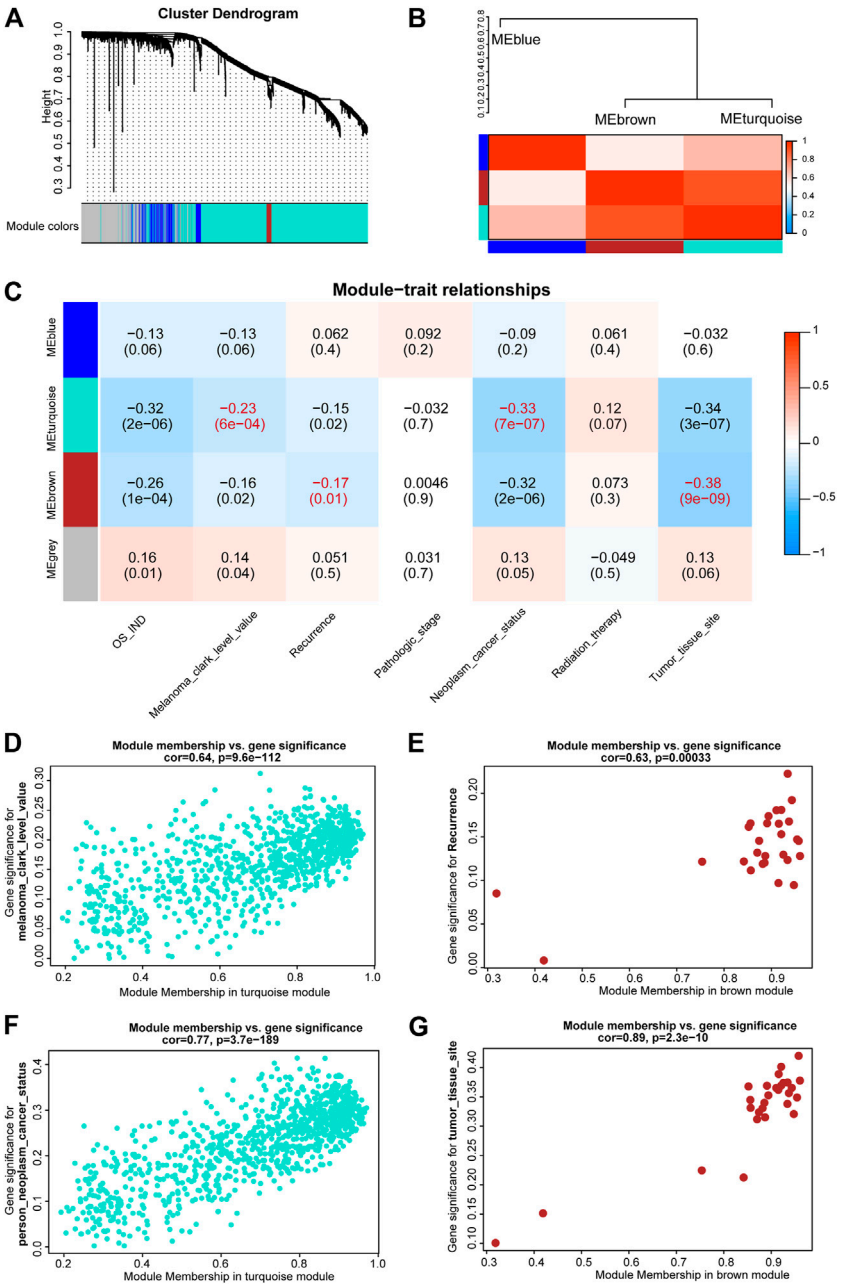


FIGURE 6 | WGCNA of m6A-LncM associated DEGs. **(A)** Clustering dendrogram of genes with dissimilarity based on topological overlap and assigned module colors. **(B)** The connectivity of eigengenes between these modules. **(C)** The relationships between Modules and clinical traits. The *p* value is showed in parentheses. **(D,F)** Scatter plot showed module membership between clinical traits (melanoma clark level and neoplasm cancer status) and turquoize module. **(E,G)** Scatter plot showed module membership between clinical traits (recurrence and tumor issue site) and brown module.

immune process of SKCM patients. Then, we used ImmLnc database to further testify whether the 24 lncRNAs are related to immune response. As showed in **Figure 5D** we confirmed these crucial lncRNAs are also enriched in multiple immune-associated pathways. All of these suggest that the m6A-LncM plays an important role in the prognosis of SKCM and the lncRNAs which we obtained also involved in immune response.

Identification of m6A-LncM-Regulated Gene Modules Associated with Clinical Traits

After differential analyses, we selected the m6A-LncM associated DEGs to construct a gene co-expression network by WGCNA. The soft threshold power was set at 6 for further analysis (**Supplementary Figure S4**). Next, we constructed the gene

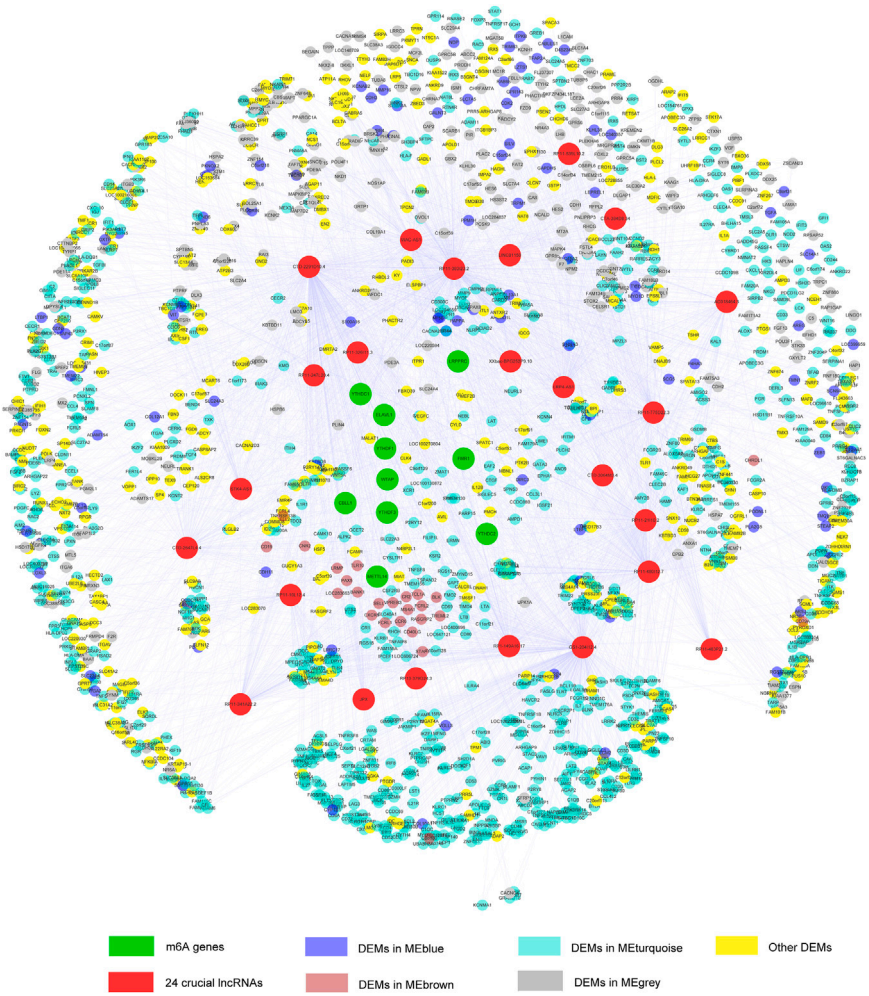


FIGURE 7 | Co-expression network constructed by m6A-related genes, m6A-LncM and DEGs.

network and identified modules using the one-step network construction function of the *WGCNA* R package. To cluster splitting, the minimum module size was set at 30 and three modules (blue, brown and turquoize) were generated (**Figure 6A**). We then analyzed the relationship between these modules. It hinted that there are two clusters over the three modules (**Figure 6B**).

Subsequently, we correlated modules with patients' clinical traits to find the crucial genes that associated with clinical characteristics (**Figure 6C**). The results showed that the turquoize module was significantly associated with melanoma clark level and neoplasm cancer status (**Figures 6D,F**). The brown module was associated with recurrence and tumor issue site (**Figures 6E,G**). The blue module was not associated with clinical phenotypes. Finally, we executed an enrichment analysis of the DEGs in the two functional modules and it showed that these key genes were also enriched in immune-associated pathways, such as activation of immune response, alpha-beta T cell activation and lymphocyte migration (**Supplementary Figure S5**). It may help us to further explore the potential regulatory mechanism of these critical lncRNAs in metastatic SKCM.

Construct a m6A-Regulated lncRNA-mRNA Co-Expressed Network

To better understand the m6A mediated regulation of lncRNA, we constructed a co-expressed network based on the PCC between the m6A genes, m6A-lncM, and m6A-LncM regulated DEGs. Ultimately, 21 m6Agenes, 24 lncRNAs, and 1,251 mRNAs were selected in the network (**Figure 7**). The colors of the nodes in the network represent their function in metastatic SKCM.

DISCUSSION

In recent years, integrate multiple omics analysis has been widely used in cancer research (Chen et al., 2020b). There are also many important studies involved in the prognosis of SKCM. Anjali et al. developed a webserver to predict the survival of SKCM patients

(Dhall et al., 2020). Li et al. also predicted the metastatic progression of melanoma based on mRNA expression status (Li et al., 2015). The machine learning model has also been performed to predict primary or metastatic SKCM patients (Bhalla et al., 2019). LncRNAs were also found to play a key role in SKCM. Several studies found that some lncRNAs could distinguish the subtypes of SKCM and predict their survival (Ma et al., 2017; Yang et al., 2018). Overexpressed lncRNA HCP5 could also decrease SKCM cell malignancy *in vitro* which may be upregulated by RARRES3 (Wei et al., 2019). m6A modification has been found that plays an important role in the occurrence and development of tumor. For example, sublethal heat treatment increases EGFR m6A modification near the 5′UTR region and promotes its binding to YTHDF1, thereby enhancing the translation of EGFR mRNA and promoting hepatocellular carcinoma progression (Su et al., 2021). METTL3 dependent m6A modification could regulate the differentiation of T follicle helper cells (Yao et al., 2021). WNT7B-m6A-TCF7L2 positive feedback loop could promote gastric cancer progression and metastasis (Gao et al., 2021). The m6A regulators have also been found by many studies to influence the progression and prognosis of cancers though regulating crucial lncRNAs. For instance, LINC00958 regulated by m6a modification can promote breast cancer tumorigenesis through miR-378a-3p and YY1 axis (Rong et al., 2021). A novel hypoxic long-stranded noncoding RNA KB-1980E6.3 maintains stemness in breast cancer stem cells by interacting with IGF2BP1 to promote the stability of c-Myc mRNA (Zhu et al., 2021). MALAT1 promotes thyroid cancer progression by binding to miR-204 and upregulating IGF2BP2, thereby affecting miR-204/IGF2BP2/m6A-MYC signaling (Ye et al., 2021). In summary, m6A modification of lncRNA is not negligible in tumor progression, however, there are no studies on m6A modification of lncRNAs in SKCM.

Here, we identified a prognostic model including 24 m6A-associated lncRNAs from 221 metastatic SKCM patients. RP11–480I12.5 was reported to promote breast cancer growth and tumorigenesis by inhibiting mir-29c-3p mediated degradation of AKT3 and CDK6 (Lou et al., 2020). LncRNA JPX can promote cervical cancer progression by regulating the miR-25–3p/SOX4 axis (Chen et al., 2020a). These lncRNAs were subsequently found to be closely related to immunity in our study. Interestingly, most of the 24 lncRNAs are novel to cancer research. Compared to the previous lncRNA studies in SKCM, the 24 lncRNAs identified in this study are novel and have the potential for application in clinic. Therefore, we hope that our findings will help to identify potential prognostic lncRNAs regulated by m6A and thus provide ideas to improve the poor prognosis of metastatic SKCM.

Limitations of the Study

Although we screened several novel m6A-related biomarkers, their specific relationship with m6A modifications needs to be verified experimentally. Meanwhile, due to the lack of lncRNA sequencing data and corresponding clinical information for SKCM, our model was not validated using an independent validation set. Subsequent studies may focus on the development of lncRNA sequencing in SKCM to provide strong evidence for screening more potential biomarkers.

CONCLUSION

This is the first study to explore the role of m6A-modified lncRNAs in the prognosis of metastatic SKCM. In this research, we identified the stable correlated m6A-lncRNA pairs in SKCM patients and SKCM cell lines. Then, we constructed a m6A-LncM by lasso regression and found potential prognostic lncRNAs by survival analysis. Twenty-four lncRNAs with independent prognostic efficacy were obtained in the analysis. Next, we comprehensively studied the relationship between these lncRNAs and clinical traits and explored their potential functions. Finally, we delineated a network mediated by these lncRNA, m6A genes and the m6A-LncM regulated differential genes. This study helps to identify potential prognostic targets for metastatic SKCM to improve its poor 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

YL, SL, and SH conceived and devised the study. SH, SL performed bioinformatic and statistical analysis. ZG, WZ, PW, and YS found testify data and analysis tools. YL, JH, and SH supervised research and wrote the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.687760/full#supplementary-material

REFERENCES

Agarwala, S. S. (2009). Current Systemic Therapy for Metastatic Melanoma. *Expert Rev. Anticancer Ther.* 9 (5), 587–595. doi:10.1586/era.09.25

Bai, X., Fisher, D. E., and Flaherty, K. T. (2019). Cell-state Dynamics and Therapeutic Resistance in Melanoma from the Perspective of MITF and IFNγ Pathways. *Nat. Rev. Clin. Oncol.* 16 (9), 549–562. doi:10.1038/s41571-019-0204-6

Balacco, D. L., and Soller, M. (2019). The m6A Writer: Rise of a Machine for Growing Tasks. *Biochemistry* 58 (5), 363–378. doi:10.1021/acs.biochem.8b01166

Balch, C. M., Gershenwald, J. E., Soong, S.-j., Thompson, J. F., Atkins, M. B., Byrd, D. R., et al. (2009). Final Version of 2009 AJCC Melanoma Staging and Classification. *Jco* 27 (36), 6199–6206. doi:10.1200/JCO.2009.23.4799

Bhalla, S., Kaur, H., Dhall, A., and Raghava, G. P. S. (2019). Prediction and Analysis of Skin Cancer Progression Using Genomics Profiles of Patients. *Sci. Rep.* 9 (1), 15790. doi:10.1038/s41598-019-52134-4

Bsirini, C., and Smoller, B. (2018). Histologic Mimics of Malignant Melanoma. *smedj* 59 (11), 602–607. doi:10.11622/smedj.2018041

Chang, A. E., Karnell, L. H., and Menck, H. R. (1998).The National Cancer Data Base Report on Cutaneous and Noncutaneous Melanoma. The American College of Surgeons Commission on Cancer and the American Cancer Society, *Cancer* 83 (8). 1664–1678. doi:10.1002/(sici)1097-0142(19981015)83:8<1664::aid-cncr23>3.0.co;2-g

Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2019). The Role of m6A RNA Methylation in Human Cancer. *Mol. Cancer* 18 (1), 103. doi:10.1186/s12943-019-1033-z

Chen, X., Yang, J., and Wang, Y. (2020a). LncRNA JPX Promotes Cervical Cancer Progression by Modulating miR-25-3p/SOX4 axis. *Cancer Cel Int* 20, 441. doi:10.1186/s12935-020-01486-3

Chen, Y., Liao, L.-D., Wu, Z.-Y., Yang, Q., Guo, J.-C., He, J.-Z., et al. (2020b). Identification of Key Genes by Integrating DNA Methylation and Next-Generation Transcriptome Sequencing for Esophageal Squamous Cell Carcinoma. *Aging* 12 (2), 1332–1365. doi:10.18632/aging.102686

Chi, Z., Li, S., Sheng, X., Si, L., Cui, C., Han, M., et al. (2011). Clinical Presentation, Histology, and Prognoses of Malignant Melanoma in Ethnic Chinese: a Study of 522 Consecutive Cases. *BMC Cancer* 11, 85. doi:10.1186/1471-2407-11-85

Curtin, J. A., Busam, K., Pinkel, D., and Bastian, B. C. (2006). Somatic Activation of KIT in Distinct Subtypes of Melanoma. *Jco* 24 (26), 4340–4346. doi:10.1200/JCO.2006.06.2984

Curtin, J. A., Fridlyand, J., Kageshita, T., Patel, H. N., Busam, K. J., Kutzner, H., et al. (2005). Distinct Sets of Genetic Alterations in Melanoma. *N. Engl. J. Med.* 353 (20), 2135–2147. doi:10.1056/NEJMoa050092

Dahal, U., Le, K., and Gupta, M. (2019). RNA m6A Methyltransferase METTL3 Regulates Invasiveness of Melanoma Cells by Matrix Metallopeptidase 2. *Melanoma Res.* 29 (4), 382–389. doi:10.1097/CMR.0000000000000580

Dhall, A., Patiyal, S., Kaur, H., Bhalla, S., Arora, C., and Raghava, G. P. S. (2020). Computing Skin Cutaneous Melanoma Outcome from the HLA-Alleles and Clinical Characteristics. *Front. Genet.* 11, 221. doi:10.3389/fgene.2020.00221

Domingues, B., Lopes, J., Soares, P., and Populo, H. (2018). Melanoma Treatment in Review. *Itt* Vol. 7, 35–49. doi:10.2147/ITT.S134842

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the Human and Mouse m6A RNA Methylomes Revealed by m6A-Seq. *Nature* 485 (7397), 201–206. doi:10.1038/nature11112

Eggermont, A. M., Spatz, A., and Robert, C. (2014). Cutaneous Melanoma. *The Lancet* 383 (9919), 816–827. doi:10.1016/S0140-6736(13)60802-8

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

Gao, Q., Yang, L., Shen, A., Li, Y., Li, Y., Hu, S., et al. (2021). A WNT7B-m6A-Tcf7l2 Positive Feedback Loop Promotes Gastric Cancer Progression and Metastasis. *Sig Transduct Target. Ther.* 6 (1), 43. doi:10.1038/s41392-020-00397-z

He, F., Yu, J., Yang, J., Wang, S., Zhuang, A., Shi, H., et al. (2021). m ^{6} A RNA Hypermethylation-Induced BACE2 Boosts Intracellular Calcium Release and Accelerates Tumorigenesis of Ocular Melanoma. *Mol. Ther.* S1525-0016 (21), 00081-2. doi:10.1016/j.ymthe.2021.02.014

Hou, P., Meng, S., Li, M., Lin, T., Chu, S., Li, Z., et al. (2021). LINC00460/DHX9/IGF2BP2 Complex Promotes Colorectal Cancer Proliferation and Metastasis by Mediating HMGA1 mRNA Stability Depending on m6A Modification. *J. Exp. Clin. Cancer Res.* 40 (1), 52. doi:10.1186/s13046-021-01857-2

Huang, Q., Yan, J., and Agami, R. (2018). Long Non-coding RNAs in Metastasis. *Cancer Metastasis Rev.* 37 (1), 75–81. doi:10.1007/s10555-017-9713-x

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA Adenosine Methylation in Biological Regulation. *Trends Genet.* 29 (2), 108–115. doi:10.1016/j.tig.2012.11.003

Kremenovic, M., Schenk, M., and Lee, D. J. (2020). Clinical and Molecular Insights into BCG Immunotherapy for Melanoma. *J. Intern. Med.* 288 (6), 625–640. doi:10.1111/joim.13037

Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., et al. (2020). Pan-cancer Characterization of Immune-Related lncRNAs Identifies Potential Oncogenic Biomarkers. *Nat. Commun.* 11 (1), 1000. doi:10.1038/s41467-020-14802-2

Li, Y., Krahn, J. M., Flake, G. P., Umbach, D. M., and Li, L. (2015). Toward Predicting Metastatic Progression of Melanoma Based on Gene Expression Data. *Pigment Cel Melanoma Res.* 28 (4), 453–463. doi:10.1111/pcmr.12374

Lin, Y., Liu, T., Cui, T., Wang, Z., Zhang, Y., Tan, P., et al. (2020). RNAInter in 2020: RNA Interactome Repository with Increased Coverage and Annotation. *Nucleic Acids Res.* 48 (D1), D189–D197. doi:10.1093/nar/gkz804

Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., and Pan, T. (2017). N 6-methyladenosine Alters RNA Structure to Regulate Binding of a Low-Complexity Protein. *Nucleic Acids Res.* 45 (10), 6051–6063. doi:10.1093/nar/gkx141

Liu, Q., and Gregory, R. I. (2019). RNAmod: an Integrated System for the Annotation of mRNA Modifications. *Nucleic Acids Res.* 47 (W1), W548–W555. doi:10.1093/nar/gkz479

Liu, Z.-X., Li, L.-M., Sun, H.-L., and Liu, S.-M. (2018). Link Between m6A Modification and Cancers. *Front. Bioeng. Biotechnol.* 6, 89. doi:10.3389/fbioe.2018.00089

Lou, W., Ding, B., Zhong, G., Yao, J., Fan, W., and Fu, P. (2020). RP11-480I12.5-004 Promotes Growth and Tumorigenesis of Breast Cancer by Relieving miR-29c-3p-Mediated AKT3 and CDK6 Degradation. *Mol. Ther. - Nucleic Acids* 21, 916–931. doi:10.1016/j.omtn.2020.07.022

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

Ma, X., He, Z., Li, L., Yang, D., and Liu, G. (2017). Expression Profiles Analysis of Long Non-coding RNAs Identified Novel lncRNA Biomarkers with Predictive Value in Outcome of Cutaneous Melanoma. *Oncotarget* 8 (44), 77761–77770. doi:10.18632/oncotarget.20780

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and Near Stop Codons. *Cell* 149 (7), 1635–1646. doi:10.1016/j.cell.2012.05.003

Pan, Y., Ma, P., Liu, Y., Li, W., and Shu, Y. (2018). Multiple Functions of m6A RNA Methylation in Cancer. *J. Hematol. Oncol.* 11 (1), 48. doi:10.1186/s13045-018-0590-8

Ramilowski, J. A., Yip, C. W., Agrawal, S., Chang, J. C., Ciani, Y., Kulakovskiy, I. V., et al. (2020). Functional Annotation of Human Long Noncoding RNAs via Molecular Phenotyping. *Genome Res.* 30 (7), 1060–1072.doi:10.1101/gr.254219.119

Rong, D., Dong, Q., Qu, H., Deng, X., Gao, F., Li, Q., et al. (2021). m6A-induced LINC00958 Promotes Breast Cancer Tumorigenesis via the miR-378a-3p/YY1 axis. *Cell Death Discov.* 7 (1), 27. doi:10.1038/s41420-020-00382-z

Su, T., Huang, M., Liao, J., Lin, S., Yu, P., Yang, J., et al. (2021). Insufficient Radiofrequency Ablation Promotes Hepatocellular Carcinoma Metastasis through M 6 A mRNA Methylation Dependent Mechanism. *Hepatology* doi:10.1002/hep.31766

Tao, L., Mu, X., Chen, H., Jin, D., Zhang, R., Zhao, Y., et al. (2021). FTO Modifies the m6A Level of MALAT and Promotes Bladder Cancer Progression. *Clin. Translational Med.* 11 (2), e310. doi:10.1002/ctm2.310

Wang, S., Chai, P., Jia, R., and Jia, R. (2018). Novel Insights on m6A RNA Methylation in Tumorigenesis: a Double-Edged Sword. *Mol. Cancer* 17 (1), 101. doi:10.1186/s12943-018-0847-4

Wang, Y., Li, D., Lu, J., Chen, L., Zhang, S., Qi, W., et al. (2020). Long Noncoding RNA TTN-AS1 Facilitates Tumorigenesis and Metastasis by Maintaining TTN Expression in Skin Cutaneous Melanoma. *Cell Death Dis* 11 (8), 664. doi:10.1038/s41419-020-02895-y

Wei, X., Gu, X., Ma, M., and Lou, C. (2019). Long Noncoding RNA HCP5 Suppresses Skin Cutaneous Melanoma Development by Regulating RARRES3 Gene Expression via Sponging miR-12. *Ott* Vol. 12, 6323–6335. doi:10.2147/OTT.S195796

Yang, S., Xu, J., and Zeng, X. (2018). A Six-Long Non-coding RNA Signature Predicts Prognosis in Melanoma Patients. *Int. J. Oncol.* 52 (4), 1178–1188. doi:10.3892/ijo.2018.4268

Yao, Y., Yang, Y., Guo, W., Xu, L., You, M., Zhang, Y.-C., et al. (2021). METTL3-dependent m6A Modification Programs T Follicular Helper Cell Differentiation. *Nat. Commun.* 12 (1), 1333. doi:10.1038/s41467-021-21594-6

Ye, M., Dong, S., Hou, H., Zhang, T., and Shen, M. (2021). Oncogenic Role of Long Noncoding RNAMALAT1 in Thyroid Cancer Progression through Regulation of the miR-204/IGF2BP2/m6A-MYC Signaling. *Mol. Ther. - Nucleic Acids* 23, 1–12. doi:10.1016/j.omtn.2020.09.023

Yu, J., Chai, P., Xie, M., Ge, S., Ruan, J., Fan, X., et al. (2021). Histone Lactylation Drives Oncogenesis by Facilitating m6A Reader Protein YTHDF2 Expression in Ocular Melanoma. *Genome Biol.* 22 (1), 85. doi:10.1186/s13059-021-02308-z

Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019). Reading, Writing and Erasing mRNA Methylation. *Nat. Rev. Mol. Cel Biol.* 20 (10), 608–624. doi:10.1038/s41580-019-0168-5

Zhou, J., Wang, J., Hong, B., Ma, K., Xie, H., Li, L., et al. (2019). Gene Signatures and Prognostic Values of m6A Regulators in Clear Cell Renal Cell Carcinoma - a Retrospective Study Using TCGA Database. *Aging* 11 (6), 1633–1647. doi:10.18632/aging.101856

Zhou, Y., Zeng, P., Li, Y.-H., Zhang, Z., and Cui, Q. (2016). SRAMP: Prediction of Mammalian N6-Methyladenosine (m6A) Sites Based on Sequence-Derived Features. *Nucleic Acids Res.* 44 (10), e91. doi:10.1093/nar/gkw104

Zhu, P., He, F., Hou, Y., Tu, G., Li, Q., Jin, T., et al. (2021). A Novel Hypoxic Long Noncoding RNA KB-1980E6.3 Maintains Breast Cancer Stem Cell Stemness via Interacting with IGF2BP1 to Facilitate C-Myc mRNA Stability. *Oncogene* 40, 1609–1627. doi:10.1038/s41388-020-01638-9

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.

Copyright © 2021 Huang, Lyu, Gao, Zha, Wang, Shan, He 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fmolb.2021.676138

Establishment of Novel Prostate Cancer Risk Subtypes and A Twelve-Gene Prognostic Model

*Enchong Zhang[1], Fujisawa Shiori[2], Mo Zhang[1], Peng Wang[1], Jieqian He[1], Yuntian Ge[1], Yongsheng Song[1] and Liping Shan[1]**

[1]Department of Urology, Shengjing Hospital of China Medical University, Shenyang, China, [2]Department of Breast Endocrine Surgery, Tohoku University Hospital, Sendai, Japan

Prostate cancer (PCa) is the most common malignancy among men worldwide. However, its complex heterogeneity makes treatment challenging. In this study, we aimed to identify PCa subtypes and a gene signature associated with PCa prognosis. In particular, nine PCa-related pathways were evaluated in patients with PCa by a single-sample gene set enrichment analysis (ssGSEA) and an unsupervised clustering analysis (i.e., consensus clustering). We identified three subtypes with differences in prognosis (Risk_H, Risk_M, and Risk_L). Differences in the proliferation status, frequencies of known subtypes, tumor purity, immune cell composition, and genomic and transcriptomic profiles among the three subtypes were explored based on The Cancer Genome Atlas database. Our results clearly revealed that the Risk_H subtype was associated with the worst prognosis. By a weighted correlation network analysis of genes related to the Risk_H subtype and least absolute shrinkage and selection operator, we developed a 12-gene risk-predicting model. We further validated its accuracy using three public datasets. Effective drugs for high-risk patients identified using the model were predicted. The novel PCa subtypes and prognostic model developed in this study may improve clinical decision-making.

Keywords: prostate cancer, predictive model, prognosis, RhoBTB2, TNFRSF10C, systems biology

OPEN ACCESS

Edited by:
Xin Wang,
City University of Hong Kong, China

Reviewed by:
Hongpan Zhang,
Capital Medical University, China
Huaide Qiu,
Nanjing Medical University, China

****Correspondence:***
Liping Shan
slpuro@163.com
shanlp@sj-hospital.org

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *06 April 2021*
Accepted: *14 May 2021*
Published: *28 May 2021*

Citation:
Zhang E, Shiori F, Zhang M, Wang P, He J, Ge Y, Song Y and Shan L (2021) Establishment of Novel Prostate Cancer Risk Subtypes and A Twelve-Gene Prognostic Model. Front. Mol. Biosci. 8:676138. doi: 10.3389/fmolb.2021.676138

INTRODUCTION

Prostate cancer (PCa) is highly intractable and incurable after metastasis. It is the leading type of noncutaneous cancer in males globally (Lee et al., 2017; Wang et al., 2018). Conventional therapeutic strategies for PCa are insufficient owing to tumor heterogeneity and complex molecular mechanisms of metastasis, leading to wide variation in outcomes (Wang et al., 2018). The clinical management of PCa includes surgery, androgen-deprivation therapy, ablative therapies, chemotherapy, radiation therapy, and immune therapy (Evans, 2018). Despite progress in therapeutic strategies, the treatment efficacy for advanced PCa is still low (Vlachostergios et al., 2017). In the context of precision medicine, the classification of PCa according to molecular features and prognosis will undoubtedly unlock effective targeted treatment strategies.

The mechanism underlying PCa heterogeneity and metastasis is highly complex; even within the same tumor, distinct phenotypes and characteristics exist (Meacham and Morrison, 2013). Multiple genomic changes contribute to PCa progression at the early stage and could define molecular subtypes. In our previous study (Zhang et al., 2020), we identified four subtypes of PCa based on immune-related gene sets. Labrecque et al., 2019 defined four novel subtypes of metastatic castration-resistant prostate cancer based on a 26-gene signature as well as distinct features of androgen

receptor responses and *NEUROI* and *NEUROII* gene expression levels (Labrecque et al., 2019). Multiple molecular mechanisms work together to influence the development, progression, and outcome of PCa and thus precise molecular characterization can improve the accuracy of clinical decision-making.

For patients with PCa, an elevated hypoxic status is related to a more aggressive and advanced disease; hypoxia reduction could increase immunity and the response to specific immunotherapies (Jayaprakash et al., 2018). Additionally, prostate is an androgen-dependent organ, and androgen interactions with androgen receptors play a key role in the progression of PCa. Endocrine therapy in PCa is aimed at lowering serum androgen levels and inhibiting androgen receptor; when this approach fails, PCa advances to a hormone-resistant state (Heinlein and Chang, 2004; Shafi et al., 2013). The PI3K-AKT-mTOR pathway interacts with multiple cellular cascades, further promoting PCa progression and aggression, and drugs targeting this pathway are employed in clinical settings (Shorning et al., 2020). E2F and MYC synergistically contribute to cell cycle regulation and are involved in tumorigenesis (Liu et al., 2015). Metabolic adaptation is pivotal for malignancy given the high energy demand, and glycolytic, fatty acid biosynthesis, and oxidative phosphorylation contribute to proliferation and worse outcomes in PCa (Schöpf et al., 2016; Xiao et al., 2018; Balusamy et al., 2020). Machine learning has become increasingly advantageous in cancer research in the era of big data, enhancing disease prediction and prognosis (Kourou et al., 2015; Qiu et al., 2020a; Qiu et al., 2020b). We classified samples into three subtypes with different patterns of pathway enrichment. We hypothesized that a multi-pathway approach could enable the subclassification of PCa with different phenotypes, functions, and clinical characteristics. Here, we exploited nine pivotal PCa-related pathways based on a literature review to divide PCa samples into three subtypes, Risk_H, Risk_M, and Risk_L, with high, middle, and low risks, respectively. Next, we explored the characteristics of subtypes with respect to the tumor microenvironment, proliferation, single nucleotide variation, and copy number variation. Then, we explored the factors contributing to the observed differences in prognosis. We constructed a risk-predicting model based on genes associated with the high-risk subtype to make the prognosis calculable. Finally, we validated the efficacy of the risk model in an internal and three external validation cohorts and predicted drugs with high sensitivity in patients with PCa classified as high risk.

TABLE 1 | The disease-related clinical information of patients with PCa included in the study.

Characteristics	Value
Patients (*n*)	484
Age (year), median (IQR)	62.0 (56.0–66.0)
PSA (ng/ml), median (IQR)	7.5 (5.1–11.3)
Pathological Gleason score, *n* (%)	
≤6	43 (9.0%)
7 (3+4)	143 (30.0%)
7 (4+3)	100 (21.0%)
8	56 (11.7%)
9–10	135 (28.3%)
Prior malignancy, *n* (%)	
No	450 (94.3%)
Yes	27 (5.7%)
Race, *n* (%)	
Asian	12 (2.5%)
Whit, American Indian or Alaska native	398 (83.4%)
Black or African American	55 (11.6%)
NA	12 (2.5%)
Residual tumor, *n* (%)	
R0	301 (63.1%)
R1	15 (3.1%)
R2	142 (29.8%)
Rx	5 (1.0%)
NA	14 (3.0%)
Clinical M, *n* (%)	
M0	437 (91.6%)
M1a or M1c	2 (0.4%)
NA	38 (8.0%)
Pathological T, *n* (%)	
T1c	2 (0.4%)
T2a	13 (2.7%)
T2b	10 (2.1%)
T2c	160 (33.5%)
T3a	151 (31.7%)
T3b	129 (27.0%)
T4	9 (1.9%)
NA	3 (0.7%)
Pathological N, *n* (%)	
N0	329 (69.0%)
N1	78 (16.4%)
NA	70 (14.6%)
Diagnostic CT or MRI, *n* (%)	
No evidence of extraprostatic extension	196 (41.1%)
Equivocal	6 (1.3%)
Extraprostatic extension localized	22 (4.6%)
Extraprostatic extension	9 (1.9%)
NA	244 (51.1%)

DFS, Disease-free survival; IQR, interquantile range; NA, not analyzed; PCa, prostate cancer; PSA, Prostate-specific antigen.

MATERIALS AND METHODS

The Cancer Genome Atlas Data Processing

RNA sequencing (RNA-seq) data (Workflow type: HTSeq-Counts), single nucleotide variants (SNV) (Workflow type: MuSE Variant Aggregation and Masking), copy number variants (CNV) (Data type: Masked Copy Number Segment), and clinical phenotypes for patients with PCa in TCGA were downloaded. RNA-seq data were normalized using the DESeq2 R package (Love et al., 2014). Tumor mutational burden (TMB) for each patient was determined from SNV data using the maftools R package. Patients with incomplete survival data or a follow-up duration of less than 30 days were excluded. In total, 484 patients were retained for the clustering analysis. The progression-free interval (PFI) was obtained from an integrated TCGA pan-cancer Clinical Data Resource (Liu et al., 2018). The clinical phenotypes of 484 patients are shown in **Table 1**. Survival outcomes are shown in **Supplementary Table S1**. The proliferation scores for patients in TCGA were obtained from Thorsson et al. (Thorsson et al., 2018). For the identification of a prognostic model, patients in TCGA were randomly divided into a training group and internal validation group using the caret R package (Kuhn,

2008). Furthermore, the AR activity scores and TMPRSS2–ERG fusion status of patients with PCa were obtained from cBioPortal (https://www.cbioportal.org/) and The Tumor Fusion Gene Data Portal (https://www.tumorfusions.org/) (Cerami et al., 2012; Gao et al., 2013; Yoshihara et al., 2015).

Validation Data Set Processing

Data sets from Gene Expression Omnibus (GEO) and cBioPortal for Cancer Genomics were used to validate the accuracy of the prognostic model (Barrett et al., 2009; Cerami et al., 2012; Gao et al., 2013). GSE70769 was obtained using the GEOquery R package from the GEO database (Davis and Meltzer, 2007; Ross-Adams et al., 2015). The datasets DKF2018 and MSKCC2010 were downloaded from the cBioPortal database. Patients with incomplete survival data or a follow-up duration of less than 30°days were excluded.

Single-Sample Gene Set Enrichment Analysis

Based on a literature review, nine gene sets associated with PCa were selected (Gann et al., 1994; Heinlein and Chang, 2004; Kaseb et al., 2007; Koh et al., 2010; Milosevic et al., 2012; Edlind and Hsieh, 2014; Ippolito et al., 2016; Xiao et al., 2018). HALLMARK_ANDROGEN_RESPONSE, HALLMARK_E2F_TARGETS, HALLMARK_FATTY_ACID_METABOLISM, 'HALLMARK_GLYCOLYSIS, HALLMARK_HYPOXIA, HALLMARK_MYC_TARGETS_V1, HALLMARK_MYC_TARGETS_V2, HALLMARK_OXIDATIVE_PHOSPHORYLATION, and HALLMARK_PI3K_AKT_MTOR_SIGNALING gene sets were downloaded from the Molecular Signatures Database v7.2 (Liberzon et al., 2015). ssGSEA based on these nine gene sets was performed using the GSVA R package (Hänzelmann et al., 2013). The parameter settings were as follows: method = "ssgsea," kcdf = "Gaussian," abs.ranking = TRUE. Patients from TCGA (n = 484) were evaluated using the ssGSEA algorithm and enrichment scores were obtained for each gene set.

Consensus Clustering

Consensus clustering was performed with the ssGSEA scores for patients (TCGA, n = 484) using the ConsensusClusterPlus R package (Wilkerson and Hayes, 2010). The number of subsamples was 100, proportion of items per sample was 0.8, and proportion of features per sample was 1. Hierarchical clustering was used. The adjacency distance matrix was determined as (1-Pearson correlation coefficient). Default settings were used for other parameters.

Principal Coordinate Analysis

RNA-seq data in Counts were normalized using the DESeq2 R package (Love et al., 2014) and used in a principal coordinate analysis (PCA). The normalized Counts matrix was transformed into a Bray–Curtis dissimilarity matrix using the vegan R package. Then, PCA was conducted using the ape R package. Owing to the large sample size, means and standard errors of principal coordinate values were used to display the separation among subtypes, as described previously (Wu et al., 2020). Finally, PERMANOVA with 10,000 permutations was performed to determine the statistical significance of the separation in PCA.

Deconvolution Algorithms

CIBERSORTx was used to analyze the immune composition in the microenvironment of PCa tissues from TCGA (Steen et al., 2020) assuming two modules. RNA-seq data in TPM format were uploaded as the mixture file. Impute Cell Fractions and LM22 (22 immune cell types) were selected for the signature matrix file. Additionally, 100 permutations were performed for the statistical analysis. Other parameters were set according to Tutorial 2 on the CIBERSORTx website.

xCell is a web-tool for cell type enrichment analyses of gene expression data for 64 immune and stroma cell types (Aran et al., 2017). According to the recommended guidelines, RNA-seq data were input in TPM format into xCell and "xCell (N = 64)" was selected as the gene signature.

The ESTIMATE algorithm can estimate tumor purity by calculating the ratio of stromal to immune cells based on gene expression data (Yoshihara et al., 2013). The Estimate R package was used to analyze. normalized RNA-seq data by this algorithm.

Weighted Correlation Network Analysis

A weighted correlation network analysis (WGCNA) can be used find phenotype-associated gene modules (Langfelder and Horvath, 2008; Li et al., 2019). RNA-seq data in TPM format were used as the input for a WGCNA. Twelve was set as the soft power threshold to construct a network that simultaneously satisfied a scale-free topology and high connectivity. Pearson correlation coefficients for the relationships between ssGSEA scores and gene modules were calculated. The correlations between the gene significance value and module membership of genes in a module were explored by a Pearson correlation analysis.

Least Absolute Shrinkage and Selection Operator Regression

LASSO regression was performed on data for patients in training group using the glmnet R package (Fonti and Belitser, 2017). Genes most highly related to E2F and MYC ssGSEA scores were used as inputs. During the selection of genes, the C-index after 10-fold cross-validation indicated the effect of different screening strategies. Genes with the maximal C-index values were included in the prognostic model with the following parameter settings: family = Cox, type.measure = C, parallel = TRUE, with default settings for other parameters.

Time-Dependent Receiver Operating Characteristic Curve Analysis

The accuracy of the prognostic model was determined by a tdROC analysis using the survivalROC R package. The endpoints were follow-up times of 1, 3, and 5°years. The area under the curve in the tdROC analysis was used to quantify accuracy. AUC values were obtained for the training group,

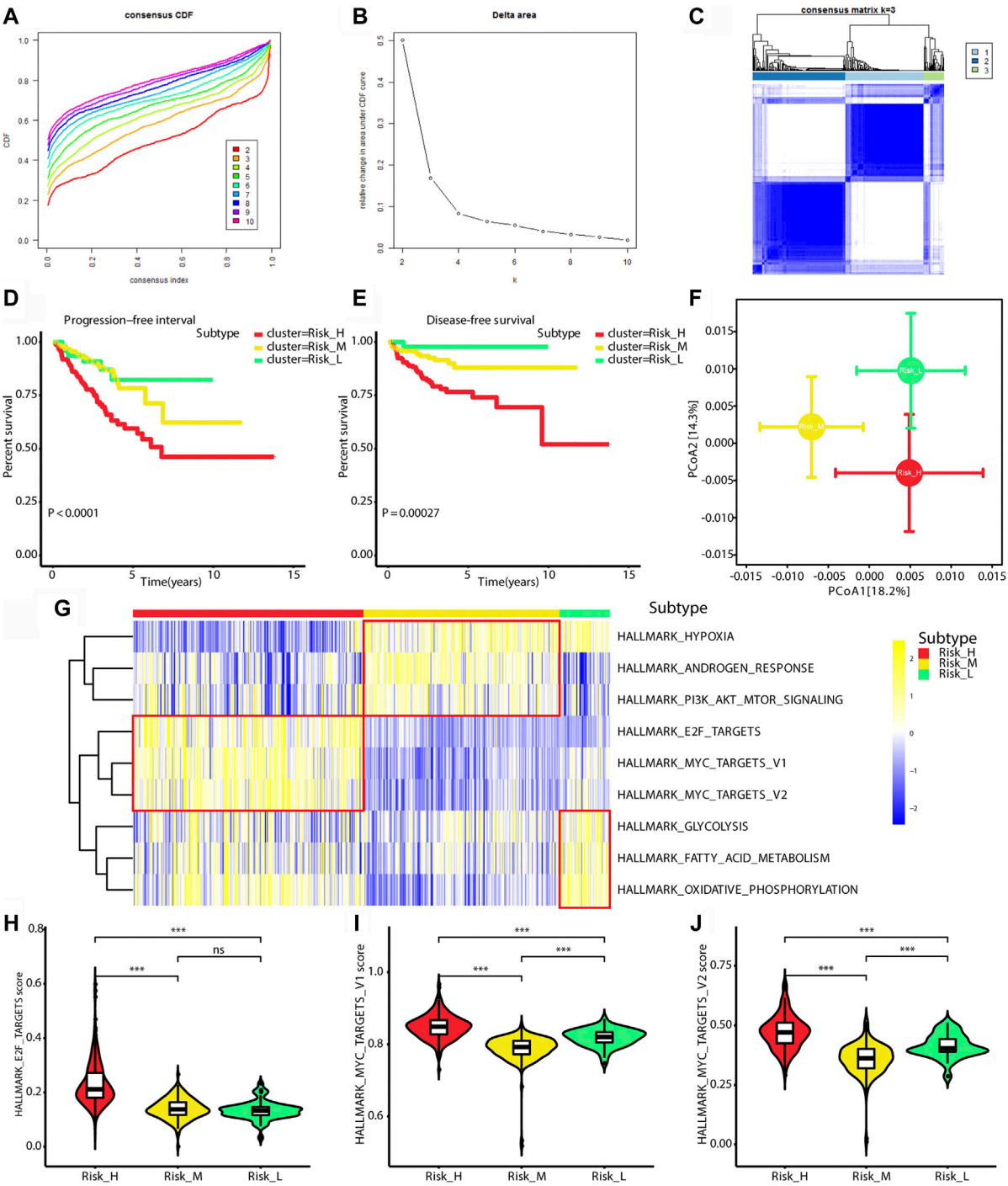


FIGURE 1 | Identification of three subtypes with different prognosis. **(A)** The CDF curve under different values of k. The value of k represents the number of clusters during the consensus cluster. When the optimal k value is reached, the area under the CDF curve will not significantly increase with the increase of k value. **(B)** Relative change in area under CDF curve under different values of k. **(C)** The consensus matrix obtained when k = 3. Consistency values range from 0 to 1, 0 means never clustering together (white), 1 means always clustering together (dark blue). **(D)** Survival curves for progression-free interval of patients in the different subtypes. **(E)** Survival curves for disease-free survival of patients in the different subtypes. **(F)** PCA of Bray-Curtis dissimilarities obtained for the transcriptional profiles. The circles and error bars indicate the mean and standard errors of the mean. PERMANOVA test with 10,000 permutations were performed to calculate *p* value. **(G)** The heatmap shows ssGSEA scores levels among three subtypes. **(H)** The violin plot shows ssGSEA score of HALLMARK_E2F_TARGETS is significantly highest in Risk_H subtype. **(I)** The violin plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V1 is significantly highest in Risk_H subtype. **(J)** The violin plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V2 is significantly highest in Risk_H subtype. (PCa, prostate cancer; CDF, cumulative distribution function; PCA, principal coordinate analysis. * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$, ns means $p > 0.05$, and $p < 0.05$ is defined as statistically significant).

TABLE 2 | The association between subtypes and clinicopathologic variables of prostate cancer.

clinicopathologic variables	Subtype			*P*
	Risk_L (*n* = 51)	Risk_M (*n* = 199)	Risk_H (*n* = 234)	
Age (year), median (IQR)	61.0 (56.0–67.5)	61.0 (56.0–66.0)	62.0 (57.0–66.0)	0.307[a]
PSA (ng/ml), median (IQR)	6.1 (4.2–10.0)	7.2 (5.0–10.8)	7.8 (5.2–12.8)	0.017[a]
Pathological Gleason score, *n* (%)				< 0.001[b]
≤6	8 (15.7%)	16 (8.0%)	19 (8.1%)	
7 (3+4)	18 (35.3%)	75 (37.7%)	49 (20.9%)	
7 (4+3)	10 (19.6%)	44 (22.1%)	47 (20.1%)	
8	3 (5.9%)	28 (14.1%)	30 (12.8%)	
9–10	12 (23.5%)	36 (18.1%)	89 (38.0%)	
Prior malignancy, *n* (%)				0.562[c]
No	50 (98.0%)	187 (94.0%)	219 (93.6%)	
Yes	1 (2.0%)	12 (6.0%)	15 (6.4%)	
Race, *n* (%)				0.804[c]
Asian	1 (2.0%)	3 (1.5%)	8 (3.4%)	
Whit, American Indian or Alaska native	43 (84.3%)	169 (84.9%)	192 (82.1%)	
Black or African American	5 (9.9%)	22 (11.1%)	27 (11.5%)	
NA	2 (3.8%)	5 (2.5%)	7 (3.0%)	
Residual tumor, *n* (%)				
R0	31 (60.8%)	138 (69.3%)	139 (59.4%)	0.035[b]
Rx/R1/R2	19 (37.3%)	53 (26.6%)	91 (38.9%)	
NA	1 (1.9%)	8 (4.1%)	4 (1.7%)	
Clinical M, *n* (%)				1.000[c]
M0	49 (96.1%)	182 (91.5%)	213 (91.0%)	
M1a or M1c	0 (0.0%)	1 (0.5%)	1 (0.4%)	
NA	2 (3.9%)	16 (8.0%)	20 (8.6%)	
Pathological T, *n* (%)				0.027[c]
T1c	0 (0.0%)	2 (1.0%)	0 (0.0%)	
T2a	2 (3.9%)	4 (2.0%)	7 (3.0%)	
T2b	2 (3.9%)	3 (1.5%)	5 (2.1%)	
T2c	18 (35.3%)	84 (42.2%)	59 (25.2%)	
T3a	16 (31.4%)	60 (30.2%)	81 (34.6%)	
T3b	12 (23.5%)	63 (21.6%)	75 (32.1%)	
T4	1 (2.0%)	2 (1.0%)	6 (2.6%)	
NA	0 (0.0%)	1 (0.5%)	1 (0.4%)	
Pathological N, *n* (%)				0.141[b]
N0	31 (60.8%)	149 (74.9%)	155 (66.2%)	
N1	7 (13.7%)	26 (13.1%)	46 (19.7%)	
NA	13 (25.5%)	24 (12.0%)	33 (14.1%)	
Outcome, *n* (%)				
DFS	1 (2.0%)	15 (7.5%)	41 (17.5%)	<0.001[b]
Disease free	50 (98.0%)	184 (92.5%)	193 (82.5.0%)	

p *values were calculated by the Kruskal test (a), Chi-square test (b) or Fisher's exact test (c). DFS, Disease-free survival; IQR, interquartile range; NA, not analyzed; PCa, prostate cancer; PSA, Prostate-specific antigen.*

internal validation group, and three external validation groups (GSE70769, DKF 2018, and MSKCC 2010).

Survival Analysis

The log-rank test and Cox regression for survival analyses were completed using the survival R package. The survival curve was plotted using the survminer R package. DFS, PFI, and biochemical recurrence-free survival were used as clinical outcomes. Follow-up time was evaluated in units of years. Finally, univariate and multivariate Cox regression analyses were used to explore whether the prognostic model is an independent predictor of prognosis.

Drug Target Prediction

Based on CTRP version 2 and PRISM databases, drug sensitivities of high-risk patients identified using the model were predicted by ridge regression based on gene expression levels. The analysis was implemented in the pRRophetic R package (Geeleher et al., 2014). Components with significantly lower areas under the dose–response curve (dr-AUC) in high-risk patients were selected first. Next, Spearman's correlation coefficients for the relationship between the dr-AUC and risk score were obtained. Components with significantly negative rho (less than −0.3) were retained.

Statistical Analysis

All statistical analyses were completed in R version 3.6.3. Chi-squared tests and Fischer's exact tests were used to compare discrete variables between or among groups. Continuous variables were compared using the Wilcoxon test (two groups) and the Kruskal–Wallis test (three or more groups). $p < 0.05$ was considered significant.

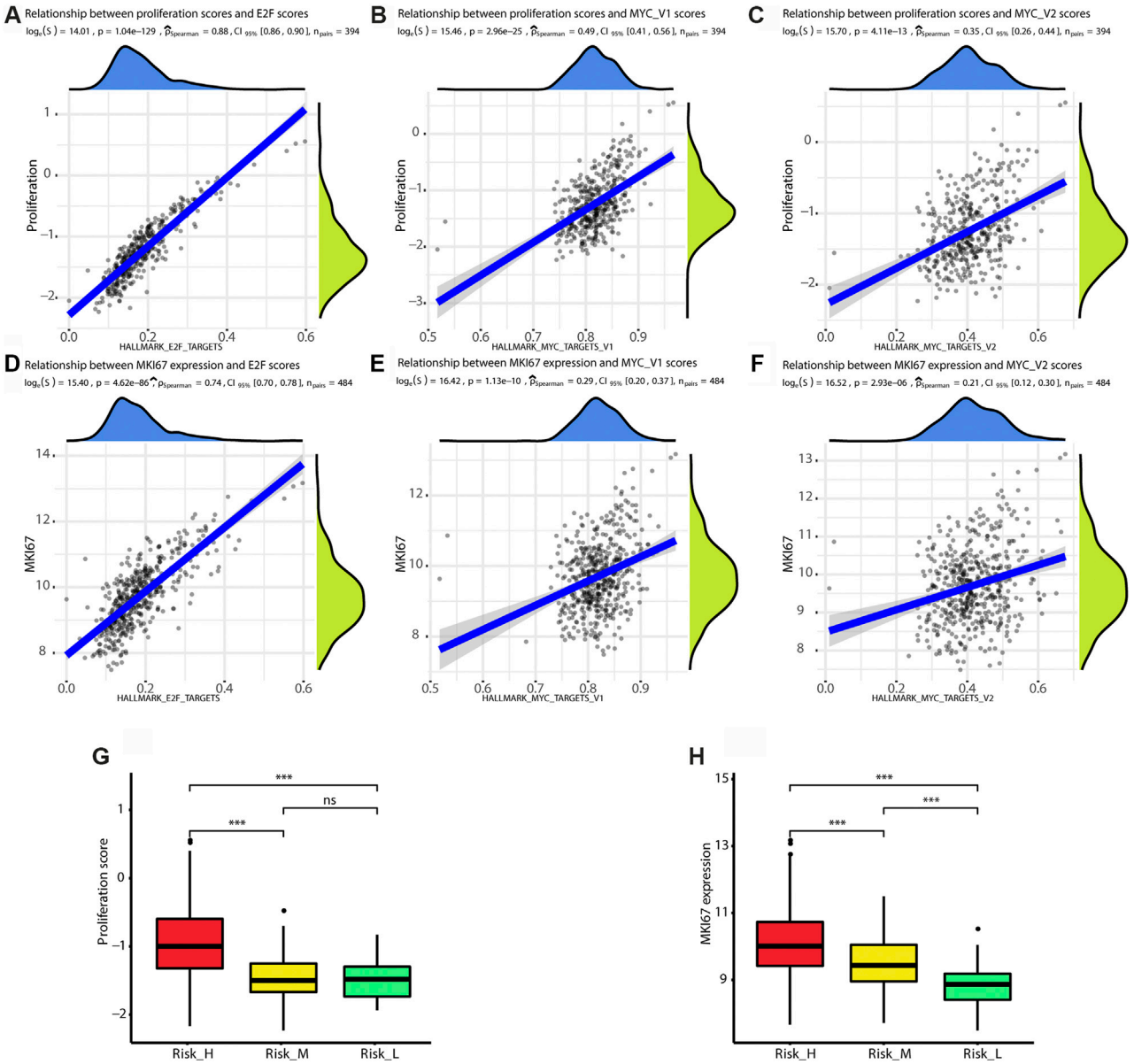


FIGURE 2 | Risk_H subtype shows high proliferation status. **(A)** The scatter plot shows ssGSEA score of HALLMARK_E2F_TARGETS is positively correlated with proliferation score (Spearman rho = 0.88, $p < 0.01$). **(B)** The scatter plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V1 is positively correlated with proliferation score (Spearman rho = 0.49, $p < 0.01$). **(C)** The scatter plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V2 is positively correlated with proliferation score (Spearman rho = 0.35, $p < 0.01$). **(D)** The scatter plot shows ssGSEA score of HALLMARK_E2F_TARGETS is positively correlated with expression of *MKI67* (Spearman rho = 0.74, $p < 0.01$). **(E)** The scatter plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V1 is positively correlated with expression of *MKI67* (Spearman rho = 0.29, $p < 0.01$). **(F)** The scatter plot shows ssGSEA score of HALLMARK_MYC_TARGETS_V2 is positively correlated with expression of *MKI67* (Spearman rho = 0.21, $p < 0.01$). **(G)** The boxplot shows the proliferation score is the highest in Risk_H subtype. **(H)** The boxplot shows the expression of *MKI67* is the highest in Risk_H subtype. (PCa, prostate cancer. * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$, ns means $p > 0.05$, and $p < 0.05$ is defined as statistically significant).

RESULTS

Identification of Three Subtypes With Different Risk Levels

Based on ssGSEA scores for nine PCa-associated gene sets, a consensus clustering analysis was performed for subtype identification. The cumulative distribution function (CDF) and relative change in the area under the CDF curve are shown in **Figures 1A,B**, respectively. According to Monti et al. (Qiu et al., 2020b), the optimal k-value is determined by a number of factors; one criterion is that when the optimal k-value is reached, the area under the CDF curve will not increase significantly with increases

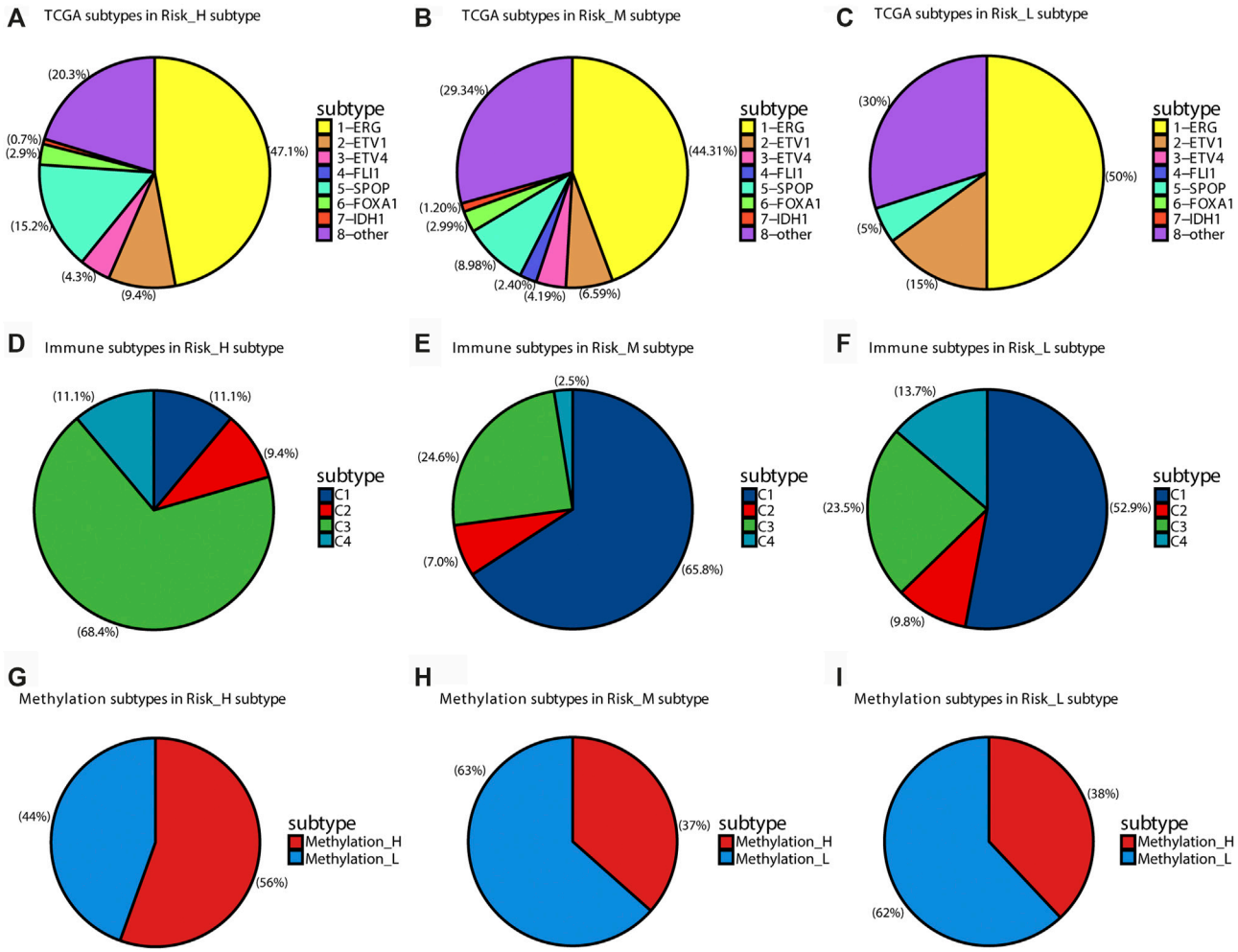


FIGURE 3 | The distribution of other TCGA subtypes in subtypes of this research. **(A)** The distribution of TCGA subtypes in Risk_H subtype. **(B)** The distribution of TCGA subtypes in Risk_M subtype. **(C)** The distribution of TCGA subtypes in Risk_L subtype. **(D)** The distribution of immune subtypes in Risk_H subtype. **(E)** The distribution of immune subtypes in Risk_M subtype. **(F)** The distribution of immune subtypes in Risk_L subtype. **(G)** The distribution of DNA methylation subtypes in Risk_H subtype. **(H)** The distribution of DNA methylation subtypes in Risk_M subtype. **(I)** The distribution of DNA methylation subtypes in Risk_L subtype.

in k. We first set k = 4, indicating that the cohort could be divided into up to four subtypes. However, one cluster consisted of only a single patient when k = 4. Additionally, the cluster-consensus value for each cluster was not large enough under k = 4 (**Supplementary Figure S1**). Therefore, we focused on k = 3, dividing patients into three clusters (**Figure 1C**). In particular, according to prognostic features shown in **Figures 1D,E**, the clusters were defined as a high-risk subtype (Risk_H), moderate-risk subtype (Risk_M), and low-risk subtype (Risk_L). In a PCA, there was significant separation among the three subtypes (**Figure 1F**, PERMANOVA $p < 0.05$). Since these subtypes were identified based on ssGSEA scores, the levels of nine PCa-associated gene sets in the three subtypes were displayed in a heatmap in **Figure 1G**. We found that HALLMARK_HYPOXIA, HALLMARK_ANDROGEN_RESPONSE, and HALLMARK_PI3K_AKT_MTOR_SIGNALING were enriched in Risk_M. HALLMARK_E2F_TARGETS, HALLMARK_MYC_TARGETS_V1, and HALLMARK_MYC_TARGETS_V2 were enriched in Risk_H. HALLMARK_GLYCOLYSIS, HALLMARK_FATTY_ACID_METABOLISM, and HALLMARK_OXIDATIVE_PHOSPHORYLATION were enriched in Risk_L. As shown in **Figures 1H–J**, the ssGSEA scores for HALLMARK_E2F_TARGETS, HALLMARK_MYC_TARGETS_V1, and HALLMARK_MYC_TARGETS_V2 were significantly higher in Risk_H than in Risk_L and Risk_M (Wilcoxon test $p < 0.001$). Furthermore, the PSA, Gleason score, residual tumor, pathological T, and survival outcome were associated with the subtype status (**Table 2**). Collectively, these data indicated that PCa could be divided into three subtypes based on the degree of enrichment of factors related to the androgen response, hypoxia, PI3K/AKT/MTOR signaling, E2F activity, MYC

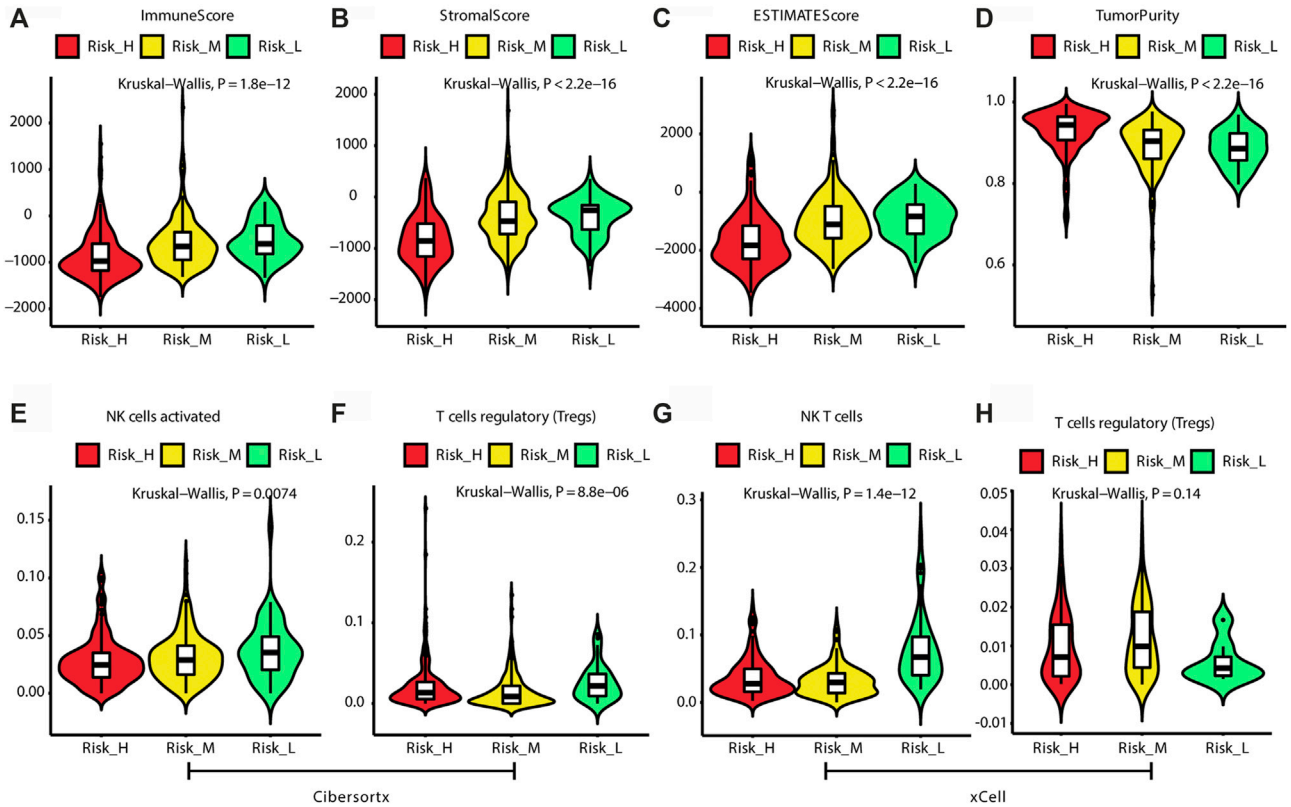


FIGURE 4 | The difference of tumor microenvironment among three subtypes. **(A)** The violin plot shows immune score is the lowest in Risk_H subtype. **(B)** The violin plot shows stromal score is the lowest in Risk_H subtype. **(C)** The violin plot shows ESTIMATE score is the lowest in Risk_H subtype. **(D)** The violin plot shows tumor purity is the highest in Risk_H subtype. **(E)** The violin plot shows content of NK cells activated is the lowest in Risk_H subtype. **(F)** The violin plot shows content of Treg cells is the lowest in Risk_H subtype. **(G)** The violin plot shows content of NK T cells is the lowest in Risk_H subtype. **(H)** The violin plot shows content of Treg cells is not significantly different among subtypes. ($p < 0.05$ is defined as statistically significant).

activity, glycolysis, fatty acid metabolism, and oxidative phosphorylation. The Risk_H subtype, with enrichment for E2F and MYC activity, showed a worse prognosis.

Risk_H Subtype is Associated With a Highly Proliferative State

We found that the ssGSEA scores for the E2F and MYC gene sets were positively correlated with proliferation scores (E2F score: rho = 0.88, $p < 0.01$; MYC_V1 score: rho = 0.49, $p < 0.01$; MYC_V2 score: rho = 0.35, $p < 0.01$), as shown in **Figures 2A–C**. Owing to the close relationship between KI67 expression and proliferation, the correlation between the expression of *MKI67* and ssGSEA scores was also explored. The levels of *MKI67* were positively correlated with ssGSEA scores for the E2F and MYC gene sets (E2F score: rho = 0.74, $p < 0.01$; MYC_V1 score: rho = 0.29, $p < 0.01$; MYC_V2 score: rho = 0.21, $p < 0.01$), as shown in **Figures 2D–F**. Finally, the proliferation scores and *MKI67* levels were highest in Risk_H among the three subtypes (Kruskal–Wallis test $p < 0.001$), as shown in **Figures 2G,H**. Taken together, tumors classified as Risk_H had higher levels of proliferation.

Known Subtypes Associated With a Poor Prognosis Were Overrepresented in the Risk_H Subtype

Seven subtypes were previously defined based on ETS fusions or mutations in *SPOP*, *FOXA1*, and *IDH1* (Abeshouse et al., 2015). We found that *SPOP mutations* were overrepresented in the Risk_H subtype (chi-square test $p < 0.05$, **Figures 3A–C**). Furthermore, we found that the Risk_L subtype had lower heterogeneity than other subtypes, including only four subtypes (**Figure 3C**). Furthermore, based on previously established immune-based subtypes, we found that the Risk_H subtype was mainly composed of the C1 subtype (Zhang et al., 2020), which is associated with a poor prognosis; however, Risk_M and Risk_L subtypes were mainly composed of the C3 subtype, associated with a relatively favorable prognosis (**Figures 3D–F**). In addition, based on subtypes differing in methylation patterns (Zhang et al., 2021), we found that the Risk_H subtype was mainly composed of the Methylation_H subtype, associated with a poor prognosis; however, Risk_M and Risk_L subtypes were mainly composed of the Methylation_L subtype, associated with a better prognosis (**Figures 3G–I**). Taken together, *SPOP* mutations, which are associated with a poor prognosis in PCa,

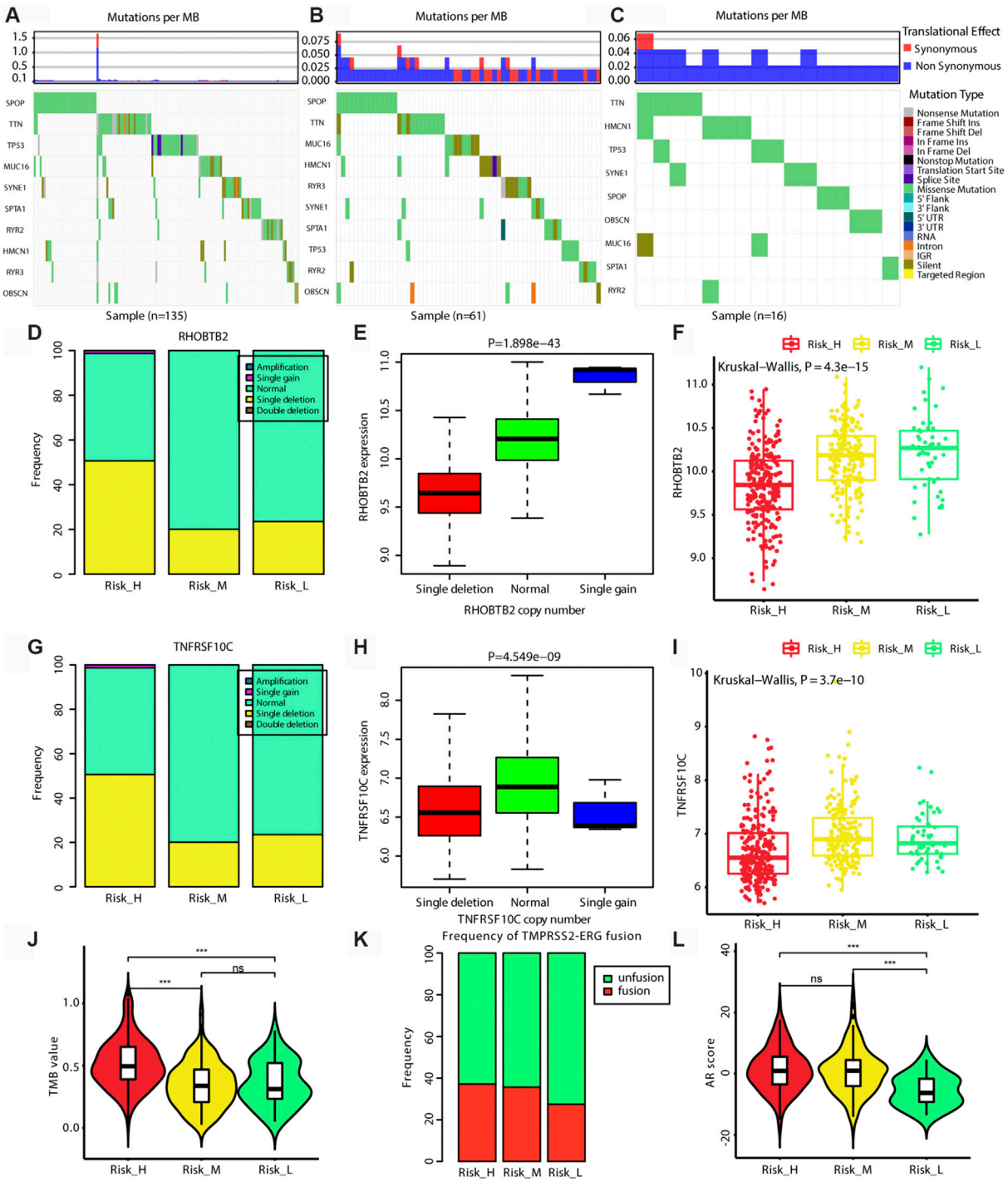


FIGURE 5 | The genomic differences among three subtypes. **(A)** The map of waterfall for the Risk_H subtype. **(B)** The map of waterfall for the Risk_M subtype. **(C)** The map of waterfall for the Risk_L subtype. **(D)** The frequency of CNV in *RHOBTB2* in Risk_H subtype is significantly higher than that in other subtype. **(E)** The expression level of *RHOBTB2* is significantly correlated with its CNV, and the expression level of *RHOBTB2* is decreased with single deletion. **(F)** The expression level of *RHOBTB2* is the lowest in Risk_H subtype. **(G)** The frequency of CNV in *TNFRSF10C* in Risk_H subtype is significantly higher than that in other subtype. **(H)** The expression level of *TNFRSF10C* is significantly correlated with its CNV, and the expression level of *TNFRSF10C* is decreased with single deletion. **(I)** The expression level of *TNFRSF10C* is the lowest in Risk_H subtype. **(J)** The violin plot shows TMB values of patients are the highest in Risk_H subtype. Wilcoxon *p* values were calculated. **(K)** The bar graph shows TMPRSS2–ERG fusion status among three subtypes. **(L)** The violin plot shows AR scores of patients are the higher in Risk_H and Risk_M subtypes. Wilcoxon *p* values were calculated. (PCa, prostate cancer; CNV, copy number variation; AR, androgen receptor. And $p < 0.05$ is defined as statistically significant).

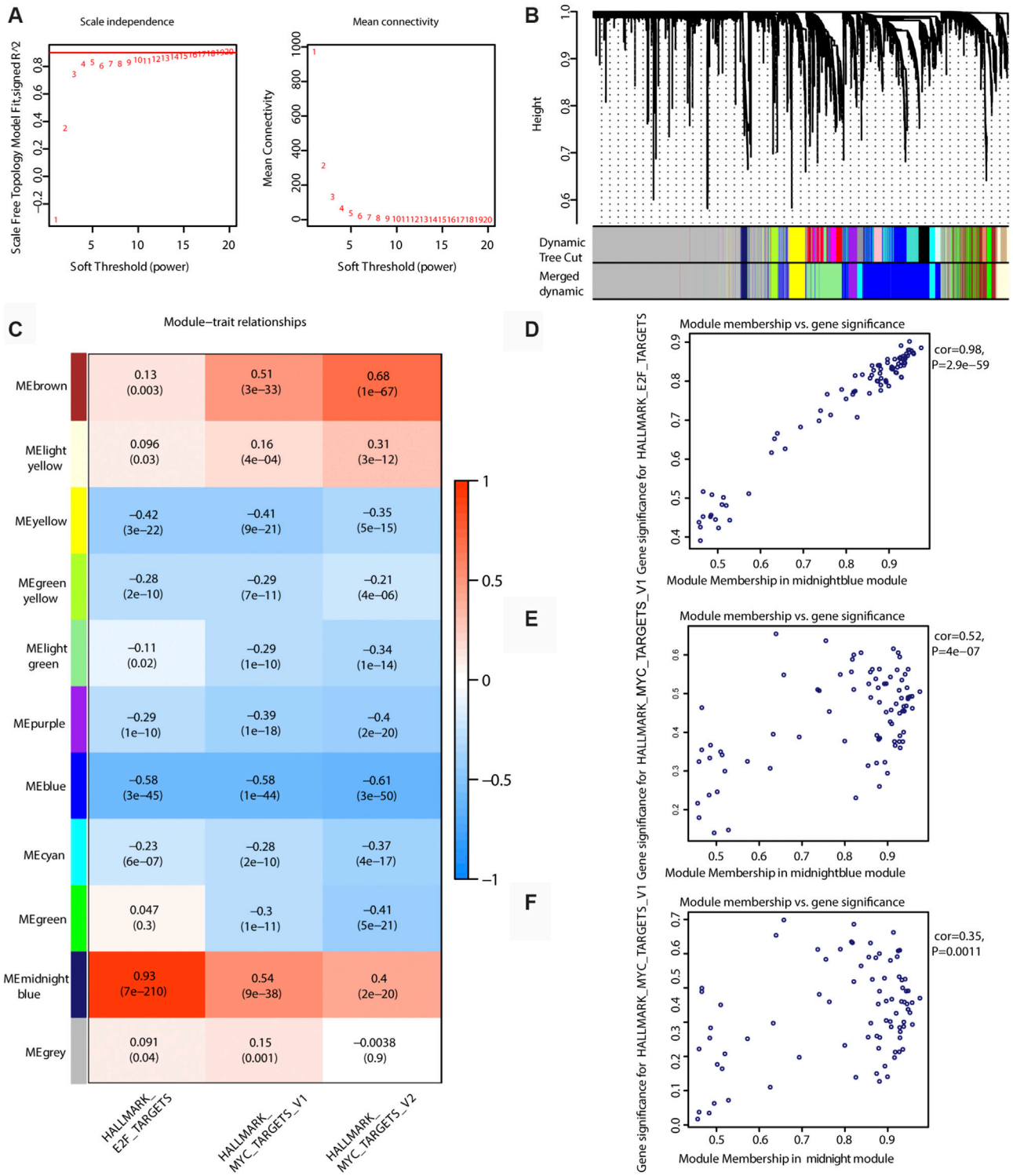


FIGURE 6 | WGCNA to find the genes for the development of the model. **(A)** The relationship of soft threshold and TOM-based dissimilarity **(left)**. The relationship of soft threshold and mean connectivity **(right)**. **(B)** After the dynamic of cut and merged, a total of 11 gene modules were finally generated. **(C)** Heat map for the correlation of gene modules and phenotypes. **(D)** The scatter plot shows the gene significance for E2F ssGSEA score is correlated with module membership in midnightblue module (Pearson rho = 0.98, $p < 0.01$). **(E)** The scatter plot shows the gene significance for MYC_V2 ssGSEA score is correlated with module membership in midnightblue module (Pearson rho = 0.35, $p < 0.01$). (WGCNA, weighted correlation network analysis; TOM, topological overlap matrix. And $p < 0.05$ is defined as statistically significant).

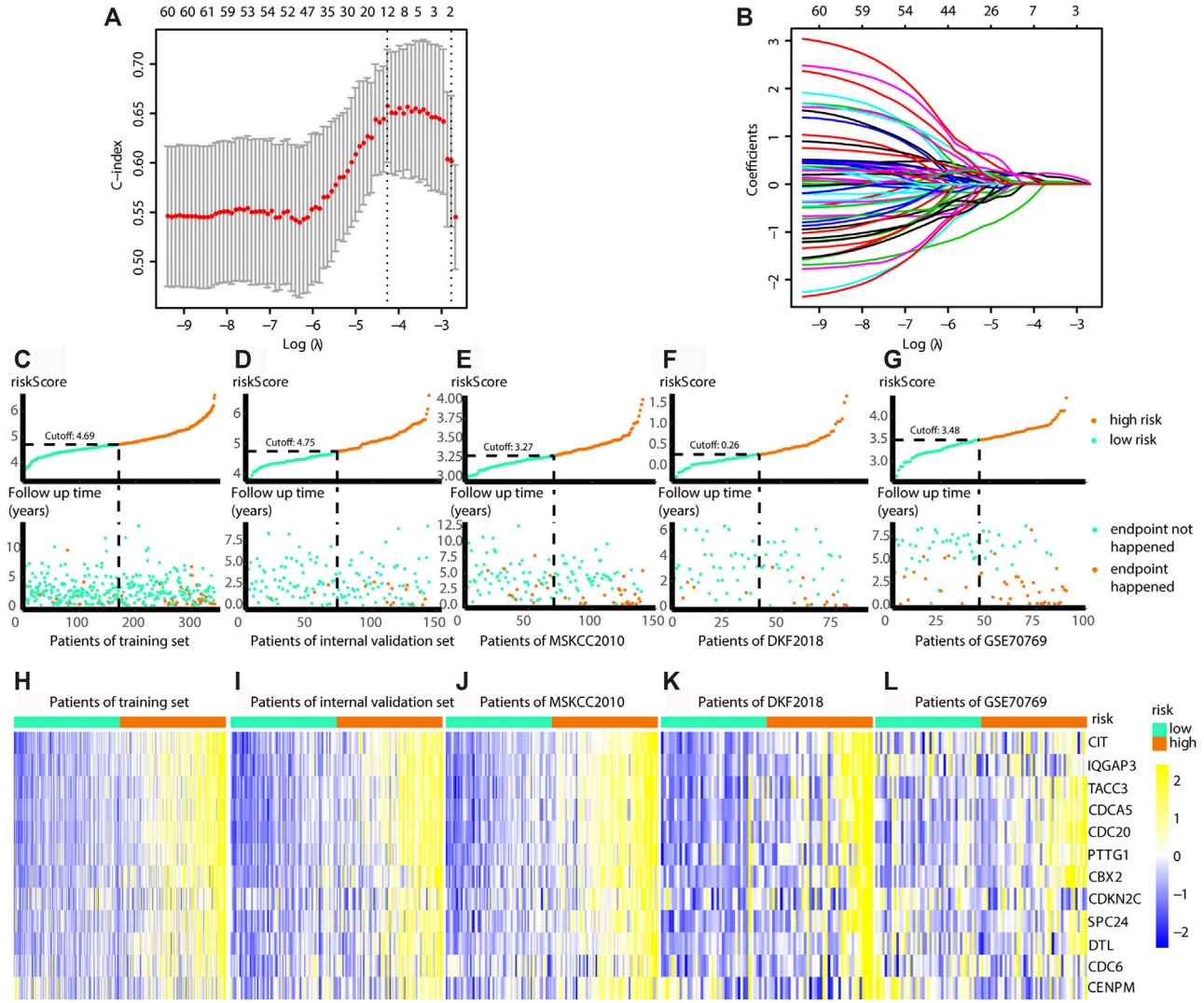


FIGURE 7 | Build the risk model by LASSO. **(A)** Cross validation based on C-index to determine the best choice of genes in the model. **(B)** Genes in the different combinations and their corresponding coefficients. **(C)** Patients of training set are arranged in the same ascending order of the risk score. **(D)** Patients of internal validation set are arranged in the same ascending order of the risk score. **(E)** Patients of MSKCC2010 data set are arranged in the same ascending order of the risk score. **(F)** Patients of DKF2018 data set are arranged in the same ascending order of the risk score. **(G)** Patients of GSE70769 data set are arranged in the same ascending order of the risk score. **(C–G)** Patients are divided into different risk levels according to the median of the risk scores in their respective data sets **(upward)**. The relationship between the survival outcome and risk levels of patients. Low-risk patients are shown on the left side of the dotted line and high-risk patients are shown on the right side **(downward)**. **(H–L)** Heatmaps show the expression levels of genes in the model, and indicate the model is robust in these data sets. (LASSO, least absolute shrinkage and selection operator. And $p < 0.05$ is defined as statistically significant).

were positively correlated with the Risk_H subtype; the *SPOP* mutation frequency in our subtypes decreased in the following order: Risk_H > Risk_M > Risk_L. Our previous studies all support the poor prognostic characteristics of the Risk_H subtype.

Risk_H Subtype Shows Greater Tumor Purity and Less Immune Cell Infiltration

According to the ESTIMATE algorithm, the Risk_H subtype had lower immune, stromal, and ESTIMATE scores (**Figures 4A,B**) and a higher tumor purity (**Figure 4D**). Based on the CIBERSORTx algorithm, which predicts the immune cell composition in the tumor microenvironment based on gene expression data for 22 kinds of immune cells, we found that activated NK cells and regulatory T cells (Tregs) were significantly less frequent in the Risk_H type than in the other types (Kruskal–Wallis test $p < 0.05$, **Figures 4E,F**). According to the xCell algorithm, we found that NK T cells were also significantly less frequent in the Risk_H type than in the other types (Kruskal–Wallis test $p < 0.05$, **Figure 4G**). However, we found that the frequency of Tregs did not differ significantly among subtypes (Kruskal–Wallis test $p > 0.05$, **Figure 4H**). Collectively, these findings indicated that the Risk_H subtype

TABLE 3 | LASSO coefficients of genes in model.

Gene name	Model coefficient
CIT	0.172
IQGAP3	0.045
TACC3	0.185
CDCA5	0.007
CDC20	0.24
PTTG1	0.193
CBX2	0.177
CDKN2C	0.017
SPC24	0.009
DTL	−0.026
CDC6	−0.022
CENPM	−0.451

had greater tumor purity and a smaller immune cell component. With respect to immune cell infiltration, activated NK cells and Tregs were reduced in the Risk_H type.

Mutational Landscape Across the Newly Established Risk Subtypes

As shown in **Figures 5A–C**, we found that the frequencies of SNVs in *SPOP* were higher in Risk_H and Risk_M than in Risk_L (chi-squared test $p < 0.05$). More single-copy deletion events were observed in the Risk_H subtype for *RHOBTB2* and *TNFRSF10C* (chi-squared test $p < 0.05$). As shown in **Figures 5E,H**, the single deletions of *RHOBTB2* and *TNFRSF10C* were associated with lower expression levels (Kruskal–Wallis test $p < 0.05$). Consistent with these findings, lower expression levels of *RHOBTB2* and *TNFRSF10C* were detected in the Risk_H subtype (Kruskal–Wallis test $p < 0.001$, **Figures 5F,I**). Furthermore, as shown in **Figure 5J**, TMB values for patients were highest in the Risk_H subtype. Although the frequency of the TMPRSS2-ERG fusion did not differ significantly among subtypes, it was highest in the Risk_H subtype (**Figure 5K**). Additionally, AR scores for patients were higher in the Risk_H and Risk_M subtypes than in the Risk-L subtype (**Figure 5L**, Wilcoxon $p < 0.001$). These data suggest that SNVs in *SPOP* and CNVs in *RHOBTB2* and *TNFRSF10C* are more common in the Risk_H subtype than in other subtypes. Low expression levels of *RHOBTB2* and *TNFRSF10C* in Risk_H could be associated with single deletion CNV events.

Identification of a Single Gene Interaction Network by WGCNA Associated With the Risk_H Subtype

As shown in **Figure 6A**, the soft threshold value was set to 8. Eleven gene interaction networks were finally defined (**Figure 6B**). The midnight blue network shown in **Figure 6C** was mostly correlated with ssGSEA scores for E2F and MYC, indicating that this gene network best represents the Risk_H subtype. As shown in **Figures 6D–F**, values for gene significance and module membership were significantly associated (E2F: rho = 0.98, $p < 0.001$; MYC_V1: rho = 0.52, $p < 0.001$; MYC_V2: rho = 0.35, $p = 0.0011$). Taken together, we identified a single group of genes that effectively reflects the Risk_H subtype.

Construction of a Prognostic Model Consisting of 12 Genes

Since the Risk_H subtype was associated with a poor prognosis, the midnight gene network was chosen to train a prognostic model *via* LASSO. In the training group, one 12-gene combination had the highest cross-validated C-index (**Figure 7A**). Changes in gene coefficients during the selection procedure are shown in **Figure 7B**. Risk scores were obtained for patients as follows: $Risk\ score = \sum_{n=1}^{12} (coefficient_n \times \text{expression of gene}_n)$. The coefficients for each gene are given in **Table 3**. Subsequently, patients in the training group, internal validation group, and three external validation groups (GSE70769, DKF 2018, and MSKCC 2010) were ranked in ascending order based on risk scores. Due to the batch effect across platforms, the median risk score in each group was selected as the cut-off value to divide patients into high-risk and low-risk groups (**Figures 7C–G**). We found that patients identified as high risk had a poorer prognosis than patients identified as low risk. The global expression levels of the 12 genes are shown in **Figures 7H–L**. Collectively, we developed a 12-gene prognostic model with robust global expression levels across all data sets.

Validation of the Model Accuracy

To validate the accuracy of the model, a tdROC analysis was performed. In the training group, 1°year AUC = 0.733, 3°year AUC = 0.713, and 5°year AUC = 0.714 (**Figure 8A**). In the internal validation group, 1°year AUC = 0.788, 3°year AUC = 0.778, 5°year AUC = 0.778 (**Figure 8B**); in the MSKCC2010 data set, 1°year AUC = 0.829, 3°year AUC = 0.748, 5°year AUC = 0.747 (**Figure 8C**); in the DKF2018 data set, 1°year AUC = 0.834, 3°year AUC = 0.698, 5°year AUC = 0.687 (**Figure 8D**); in the GSE70769 data set, 1°year AUC = 0.723, 3°year AUC = 0.788, 5°year AUC = 0.740 (**Figure 8E**). Subsequently, we found that the high-risk patients identified by this model had worse survival in the training group (log-rank test $p = 0.006$, Cox test $p < 0.001$), internal validation group (log-rank test $p = 0.005$, Cox test $p = 0.001$), MSKCC2010 data set (log-rank test $p = 0.024$, Cox test $p < 0.001$), DKF2018 data set (log-rank test $p = 0.019$, Cox test $p < 0.001$), and GSE70769 data set (log-rank test $p < 0.001$, Cox test $p < 0.001$), as shown in **Figures 8F–J**. Furthermore, we found that patients who died or experienced recurrence had higher risk scores in the training group (Wilcoxon test $p < 0.001$), internal validation group (Wilcoxon test $p < 0.001$), MSKCC2010 data set (Wilcoxon test $p < 0.001$), DKF2018 data set (Wilcoxon test $p < 0.05$), and GSE70769 data set (Wilcoxon test $p < 0.001$). According to univariate and multivariate Cox regression analyses (**Table 4**), this model and the Gleason grade were independent predictors of prognosis. Taken together, the prognostic model had high accuracy.

Target Drug Prediction for High-Risk Patients

Using compounds from the CTRP and PRISM databases, we predicted drug sensitivity for patients with high risk scores. As shown in **Figure 9A**, 3-CI-AHPC, CD-437, CR-1-31B,

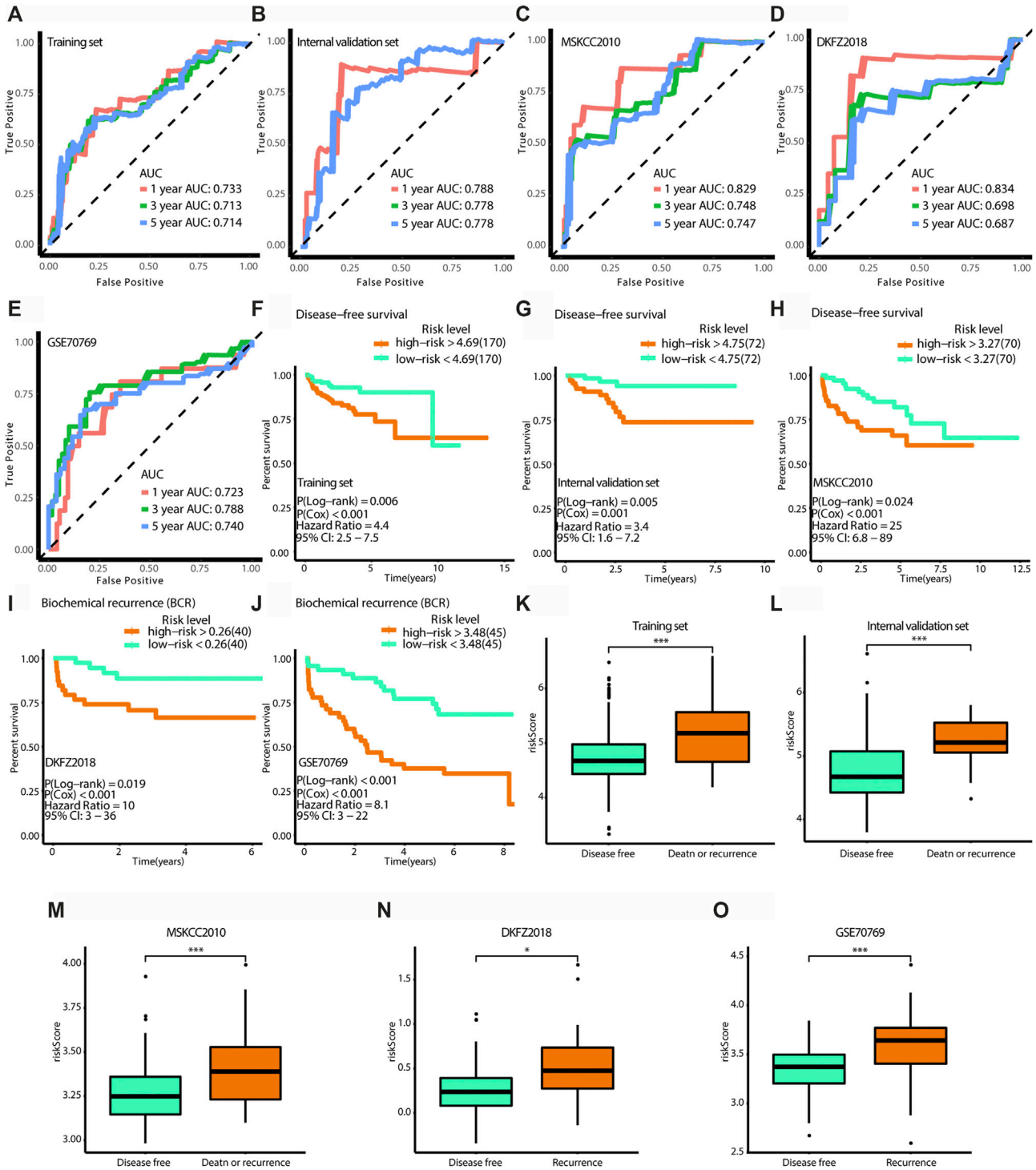


FIGURE 8 | Verification of the effectiveness of the model. **(A–E)** The ROC curves of 1-year, 3-year, and 5-year follow-up time. **(F–J)** Kaplan-Meier curves for survival analysis. **(K–O)** The boxplots show the difference of risk score between survival outcomes. **(A**,**F**,**K)** The results in the training set. **(B**,**G**,**L)** The results in the internal validation set. **(C**,**H**,**M)** The results in MSKCC 2010. **(D**,**I**,**N)** The results in DKFZ 2018. **(E**,**J**,**O)** The results in GSE70769. (AUC, area under curve. * means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$, ns means $p > 0.05$, and $p < 0.05$ is defined as statistically significant. And $p < 0.05$ is defined as statistically significant).

TABLE 4 | Results of univariate and multivariate Cox regression.

	Univariate analysis		Multivariate analysis	
	HR (95% CI)	***p* value**	**HR (95% CI)**	***p* value**
Race	0.787 (0.305–2.03)	0.62		
Age	0.998 (0.945–1.053)	0.931		
Pathological N	1.366 (0.643–2.903)	0.417		
Pathological T	4.279 (1.364–13.426)	0.013	1.113 (0.308–4.018)	0.87
Gleason grade	2.143 (1.42–3.235)	<0.001	1.792 (1.129–2.844)	0.013
Prior malignancy	0.767 (0.104–5.638)	0.794		
Diagnostic CT or MRI	1.38 (0.954–1.996)	0.087		
Residual tumor	1.272 (0.897–1.806)	0.177		
PSA	1.012 (0.991–1.034)	0.255		
Risk score	3.423 (1.862–6.293)	<0.001	2.099 (1.043–4.224)	0.038

PSA, prostate-specific antigen; and p < *0.05 is defined as statistically significant.*

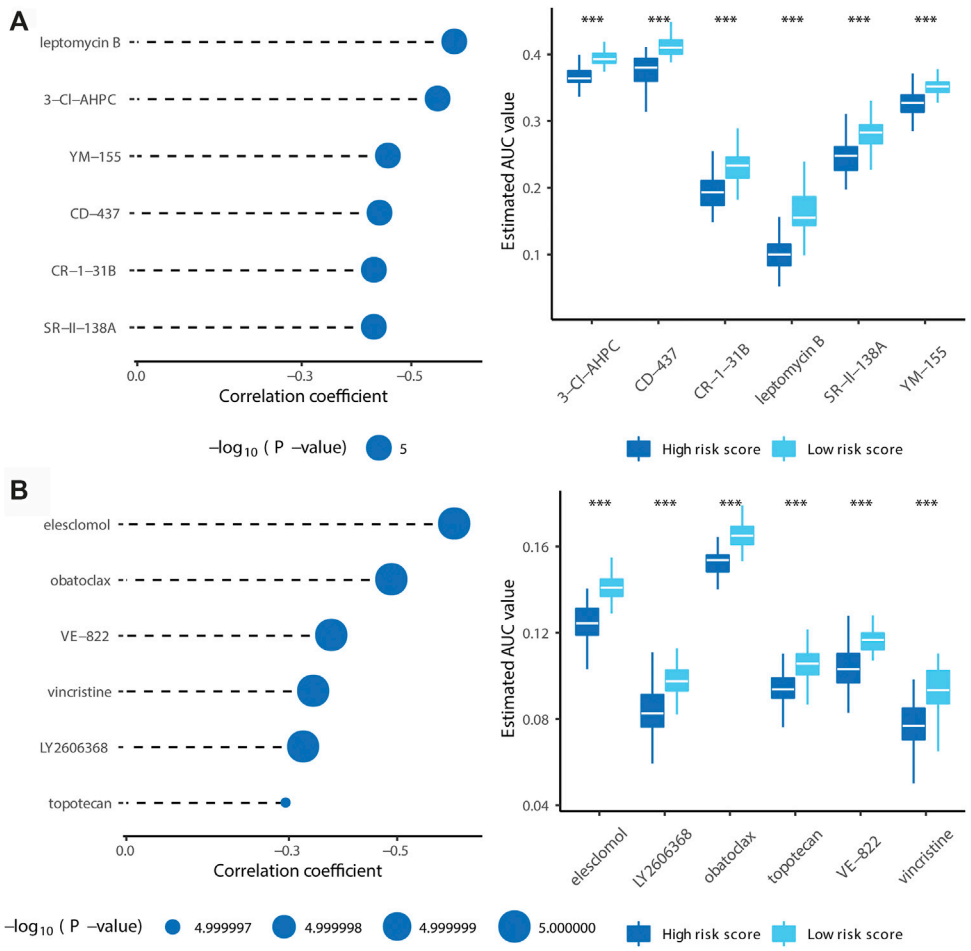


FIGURE 9 | Identification of candidate agents with higher drug sensitivity in patients with high risk score. **(A)** The results of Spearman's correlation analysis and differential drug response analysis of six CTRP-derived compounds. **(B)** The results of Spearman's correlation analysis and differential drug response analysis of six PRISM-derived compounds. Note that lower values on the y-axis of boxplots imply greater drug sensitivity. (* means $p < 0.05$, ** means $p < 0.01$, *** means $p < 0.001$, ns means $p > 0.05$, and $p < 0.05$ is defined as statistically significant. And $p < 0.05$ is defined as statistically significant).

leptomycin B, SR-II-138A, and YM-155 sensitivities were high for patients with high risk scores (Spearman correlation rho < −0.3, Spearman correlation test $p < 0.001$, and Wilcoxon test $p < 0.001$). As shown in **Figure 9B**, elesclomol, LY2606368, obatoclax, topotecan, VE-822, and vincristine sensitivities were high for patients with high risk scores (Spearman correlation test rho < −0.3, Spearman correlation test $p < 0.001$, and Wilcoxon test $p < 0.001$). Collectively, we identified ten target drugs predicted to be effective for high-risk patients, and leptomycin B, LY2606368, and vincristine showed particularly high effectiveness.

DISCUSSION

The mechanism underlying PCa progression is complex and cannot be explained by a single pathway. Accordingly, in this study, we used gene expression information for eight PCa-related pathways (i.e., hypoxia, androgen response, PI3K-AKT-mTOR signaling, E2F targets, MYC targets V1, MYC targets V2, glycolysis, fatty acid metabolism, and oxidative phosphorylation pathways) extracted from Molecular Signatures Database v7.2 and data for PCa cohorts from multiple platforms (TCGA, GSE70769, DKF 2018, and MSKCC 2010) to identify three PCa subtypes (Risk_H, Risk_M and Risk_L). These subtypes were then used to construct a risk-predicting model and drug sensitivity prediction was performed for the high-risk group.

For patients with PCa, an elevated hypoxic status is related to a more aggressive and advanced disease; hypoxia reduction could increase immunity and the response to specific immunotherapies (Jayaprakash et al., 2018). Additionally, prostate is an androgen-dependent organ, and androgen interactions with androgen receptors play a key role in the progression of PCa. Endocrine therapy in PCa is aimed at lowering serum androgen levels and inhibiting androgen receptor; when this approach fails, PCa advances to a hormone-resistant state (Heinlein and Chang, 2004; Shafi et al., 2013). The PI3K-AKT-mTOR pathway interacts with multiple cellular cascades, further promoting PCa progression and aggression, and drugs targeting this pathway are employed in clinical settings (Shorning et al., 2020). E2F and MYC synergistically contribute to cell cycle regulation and are involved in tumorigenesis (Liu et al., 2015). Metabolic adaptation is pivotal for malignancy given the high energy demand, and glycolytic, fatty acid biosynthesis, and oxidative phosphorylation contribute to proliferation and worse outcomes in PCa (Schöpf et al., 2016; Xiao et al., 2018; Balusamy et al., 2020). We classified samples into three subtypes with different patterns of pathway enrichment.

Among the three subtypes, the cluster with enrichment for the E2F and MYC pathways was identified as high-risk group (Risk_H), which was associated with the worst clinical outcomes. Further analyses of the proliferation scores and *MKI67* gene expression level support the highly proliferative feature of the Risk_H cohort. Additionally, the proportions of immune and stromal cells were highest in the Risk_L cohort. NK cells, which possesses important anti-cancer functions (Abel et al., 2018), were most abundant in the Risk_L group. Thus, the poor prognosis in the Risk_H group can be explained from the perspective of immune activity. Additionally, *RhoBTB2*, a candidate tumor suppressor, has been implicated in various cancers, such as breast cancer and lung cancer (McKinnon et al., 2008). However, little is known about its role in PCa. We found that the single copy deletion of *RhoBTB2* was most frequent in the Risk_H group, while its overall expression was highest in the Risk_L group. This finding may provide an entry point for future PCa research. The *TNFRSF10C* gene, also known as *TRAIL-R3*, is a decoy receptor for tumor necrosis factor-related apoptosis-inducing ligand, inducing tumor apoptosis in multiple malignancies (Almodóvar et al., 2004). We detected copy number variation distinguishing the Risk_H and Risk_L subtypes, and this may further explain the poor prognosis in the Risk_H group.

After establishing the prognostic value and properties of the subtypes, we constructed a twelve gene-based risk-prediction model. This model could supplement current strategies for clinical decision-making and prognostic predictions. Finally, we filtered twelve drugs expected to show high sensitivity in high-risk patients with PCa, 3-CI-HPC, CD-437, CR-1-31B, leptomycin B, SR-II-138A, YM-155, elesclomol, LY2606368, obatoclax, topotecan, VE-822, and vincristine.

As a complex and heterogeneous disease, PCa is difficult to manage by a universal treatment approach. In this study, we divided PCa into three clusters based on eight pivotal pathways, allowing for more innovative and objective results than those obtained by analyses of single pathways. Moreover, we translated the results into a clinically useful tool and identified potentially effective drugs for high-risk patients, providing direct guidance for clinical strategies aimed at precision medicine. However, our study had limitations. First, the results are based on retrospective investigations of cohorts from multiple platforms; prospective explorations are needed to validate our results. Further clinical studies of the drug candidates are needed. Despite these drawbacks, our results provide novel ideas for PCa management.

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

EZ, LS, and YS were responsible for the design and conception of the research project; EZ, FS, MZ, PW, JH, YG, and YS and contributed to data acquisition, or data analysis and data cleaning; EZ and LS participated in the drafting of the manuscript and the rigorous modification of the manuscript to clearly convey the research contents. All authors are responsible for the authenticity and reliability of this study and have no objection to the final submitted manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (No. 81802540), the Natural Science Foundation of Liaoning Province (No. 20180550985), the Shenyang Science and Technology Program for Young Innovative Talents (No. RC190386), and the 345 Talent Project of Shengjing Hospital of China Medical University. The 345 Talent Project includes the 30 project and the 50 project.

ACKNOWLEDGMENTS

We thank the Medical Research Center of Shengjing Hospital for providing us with laboratories.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.676138/full#supplementary-material

REFERENCES

Abel, A. M., Yang, C., Thakar, M. S., and Malarkannan, S. (2018). Natural Killer Cells: Development, Maturation, and Clinical Utilization. *Front. Immunol.* 9, 1869. doi:10.3389/fimmu.2018.01869

Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C. D., et al. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163 (4), 1011–1025. doi:10.1016/j.cell.2015.10.025

Almodóvar, C. R. d., López-Rivas, A., Redondo, J. M., and Rodríguez, A. (2004). Transcriptional Regulation of the TRAIL-R3 Gene. *Vitam Horm.* 67, 51–63. doi:10.1016/s0083-6729(04)67004-x

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

Balusamy, S. R., Perumalsamy, H., Veerappan, K., Huq, M. A., Rajeshkumar, S., Lakshmi, T., et al. (2020). Citral Induced Apoptosis through Modulation of Key Genes Involved in Fatty Acid Biosynthesis in Human Prostate Cancer Cells: In Silico and *In Vitro* Study. *Biomed. Res. Int.* 2020, 1–15. doi:10.1155/2020/6040727

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2009). NCBI GEO: Archive for High-Throughput Functional Genomic Data. *Nucleic Acids Res.* 37 (Suppl. l_1), D885–D890. doi:10.1093/nar/gkn764

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

Edlind, M. P., and Hsieh, A. C. (2014). PI3K-AKT-mTOR Signaling in Prostate Cancer Progression and Androgen Deprivation Therapy Resistance. *Asian J. Androl.* 16 (3), 378–386. doi:10.4103/1008-682X.122876

Evans, A. J. (2018). Treatment Effects in Prostate Cancer. *Mod. Pathol.* 31 (S1), 110–121. doi:10.1038/modpathol.2017.158

Fonti, V., and Belitser, E. (2017). Feature Selection Using Lasso. *VU Amsterdam Res. Paper Business Analytics* 30, 1–25.

Gann, P. H., Hennekens, C. H., Sacks, F. M., Grodstein, F., Giovannucci, E. L., and Stampfer, M. J. (1994). Prospective Study of Plasma Fatty Acids and Risk of Prostate Cancer. *JNCI J. Natl. Cancer Inst.* 86 (4), 281–286. doi:10.1093/jnci/86.4.281

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. Signaling* 6 (269), pl1. doi:10.1126/scisignal.2004088

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

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC bioinformatics* 14 (1), 1–15. doi:10.1186/1471-2105-14-7

Heinlein, C. A., and Chang, C. (2004). Androgen Receptor in Prostate Cancer. *Endocr. Rev.* 25 (2), 276–308. doi:10.1210/er.2002-0032

Ippolito, L., Marini, A., Cavallini, L., Morandi, A., Pietrovito, L., Pintus, G., et al. (2016). Metabolic Shift Toward Oxidative Phosphorylation in Docetaxel Resistant Prostate Cancer Cells. *Oncotarget* 7 (38), 61890–61904. doi:10.18632/oncotarget.11301

Jayaprakash, P., Ai, M., Liu, A., Budhani, P., Bartkowiak, T., Sheng, J., et al. (2018). Targeted Hypoxia Reduction Restores T Cell Infiltration and Sensitizes Prostate Cancer to Immunotherapy. *J. Clin. Invest.* 128 (11), 5137–5149. doi:10.1172/jci96268

Kaseb, A. O., Chinnakannu, K., Chen, D., Sivanandam, A., Tejwani, S., Menon, M., et al. (2007). Androgen Receptor- and E2F-1-Targeted Thymoquinone Therapy for Hormone-Refractory Prostate Cancer. *Cancer Res.* 67 (16), 7782–7788. doi:10.1158/0008-5472.can-07-1483

Koh, C. M., Bieberich, C. J., Dang, C. V., Nelson, W. G., Yegnasubramanian, S., and De Marzo, A. M. (2010). MYC and Prostate Cancer. *Genes & Cancer* 1 (6), 617–628. doi:10.1177/1947601910379132

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine Learning Applications in Cancer Prognosis and Prediction. *Comput. Struct. Biotechnol. J.* 13, 8–17. doi:10.1016/j.csbj.2014.11.005

Kuhn, M. (2008). Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* 28 (5), 1–26. doi:10.18637/jss.v028.i05

Labrecque, M. P., Coleman, I. M., Brown, L. G., True, L. D., Kollath, L., Lakely, B., et al. (2019). Molecular Profiling Stratifies Diverse Phenotypes of Treatment-Refractory Metastatic Castration-Resistant Prostate Cancer. *J. Clin. Invest.* 129 (10), 4492–4505. doi:10.1172/jci128212

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Lee, D. J., Mallin, K., Graves, A. J., Chang, S. S., Penson, D. F., Resnick, M. J., et al. (2017). Recent Changes in Prostate Cancer Screening Practices and Epidemiology. *J. Urol.* 198 (6), 1230–1240. doi:10.1016/j.juro.2017.05.074

Li, G., Jiang, Y., Lyu, X., Cai, Y., Zhang, M., Wang, Z., et al. (2019). Deconvolution and Network Analysis of IDH-Mutant Lower Grade Glioma Predict Recurrence and Indicate Therapeutic Targets. *Epigenomics* 11 (11), 1323–1333. doi:10.2217/epi-2019-0137

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, H., Tang, X., Srivastava, A., Pécot, T., Daniel, P., Hemmelgarn, B., et al. (2015). Redeployment of Myc and E2f1-3 Drives Rb-Deficient Cell Cycles. *Nat. Cel Biol* 17 (8), 1036–1048. doi:10.1038/ncb3210

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.e11. doi:10.1016/j.cell.2018.02.052

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), 1–21. doi:10.1186/s13059-014-0550-8

McKinnon, C. M., Lygoe, K. A., Skelton, L., Mitter, R., and Mellor, H. (2008). The Atypical Rho GTPase RhoBTB2 Is Required for Expression of the Chemokine CXCL14 in normal and Cancerous Epithelial Cells. *Oncogene* 27 (54), 6856–6865. doi:10.1038/onc.2008.317

Meacham, C. E., and Morrison, S. J. (2013). Tumour Heterogeneity and Cancer Cell Plasticity. *Nature* 501 (7467), 328–337. doi:10.1038/nature12624

Milosevic, M., Warde, P., Ménard, C., Chung, P., Toi, A., Ishkanian, A., et al. (2012). Tumor Hypoxia Predicts Biochemical Failure Following Radiotherapy

for Clinically Localized Prostate Cancer. *Clin. Cancer Res.* 18 (7), 2108–2114. doi:10.1158/1078-0432.ccr-11-2711

Qiu, H., Hu, X., He, C., Yu, B., Li, Y., and Li, J. (2020a). Identification and Validation of an Individualized Prognostic Signature of Bladder Cancer Based on Seven Immune Related Genes. *Front. Genet.* 11, 12. doi:10.3389/fgene.2020.00012

Qiu, H., Li, Y., Cheng, S., Li, J., He, C., and Li, J. (2020b). A Prognostic Microenvironment-Related Immune Signature via ESTIMATE (PROMISE Model) Predicts Overall Survival of Patients with Glioma. *Front. Oncol.* 10, 580263. doi:10.3389/fonc.2020.580263

Ross-Adams, H., Lamb, A. D., Dunning, M. J., Halim, S., Lindberg, J., Massie, C. M., et al. (2015). Integration of Copy Number and Transcriptomics Provides Risk Stratification in Prostate Cancer: A Discovery and Validation Cohort Study. *EBioMedicine* 2 (9), 1133–1144. doi:10.1016/j.ebiom.2015.07.017

Schöpf, B., Schäfer, G., Weber, A., Talasz, H., Eder, I. E., Klocker, H., et al. (2016). Oxidative Phosphorylation and Mitochondrial Function Differ between Human Prostate Tissue and Cultured Cells. *FEBS J.* 283 (11), 2181–2196. doi:10.1111/febs.13733

Shafi, A. A., Yen, A. E., and Weigel, N. L. (2013). Androgen Receptors in Hormone-dependent and Castration-Resistant Prostate Cancer. *Pharmacol. Ther.* 140 (3), 223–238. doi:10.1016/j.pharmthera.2013.07.003

Shorning, B. Y., Dass, M. S., Smalley, M. J., and Pearson, H. B. (2020). The PI3K-AKT-mTOR Pathway and Prostate Cancer: At the Crossroads of AR, MAPK, and WNT Signaling. *Int. J. Mol. Sci.* 21 (12), 4507. doi:10.3390/ijms21124507

Steen, C. B., Liu, C. L., Alizadeh, A. A., and Newman, A. M. (2020). Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol. Biol.* 2117, 135–157. doi:10.1007/978-1-0716-0301-7_7

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.e14. doi:10.1016/j.immuni.2018.03.023

Vlachostergios, P. J., Puca, L., and Beltran, H. (2017). Emerging Variants of Castration-Resistant Prostate Cancer. *Curr. Oncol. Rep.* 19 (5), 32. doi:10.1007/s11912-017-0593-6

Wang, G., Zhao, D., Spring, D. J., and DePinho, R. A. (2018). Genetics and Biology of Prostate Cancer. *Genes Dev.* 32 (17-18), 1105–1140. doi:10.1101/gad.315739.118

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

Wu, H., Tremaroli, V., Schmidt, C., Lundqvist, A., Olsson, L. M., Krämer, M., et al. (2020). The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cel Metab.* 32 (3), 379–390. doi:10.1016/j.cmet.2020.06.011

Xiao, H., Wang, J., Yan, W., Cui, Y., Chen, Z., Gao, X., et al. (2018). GLUT1 Regulates Cell Glycolysis and Proliferation in Prostate Cancer. *Prostate* 78 (2), 86–94. doi:10.1002/pros.23448

Xiao, H., Wang, J., Yan, W., Cui, Y., Chen, Z., Gao, X., et al. (2018). GLUT1 Regulates Cell Glycolysis and Proliferation in Prostate Cancer. *Prostate* 78 (2), 86–94. doi:10.1002/pros.23448

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), 1–11. doi:10.1038/ncomms3612

Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., et al. (2015). The Landscape and Therapeutic Relevance of Cancer-Associated Transcript Fusions. *Oncogene* 34 (37), 4845–4854. doi:10.1038/onc.2014.406

Zhang, E., He, J., Zhang, H., Shan, L., Wu, H., Zhang, M., et al. (2020). Immune-Related Gene-Based Novel Subtypes to Establish a Model Predicting the Risk of Prostate Cancer. *Front. Genet.* 11, 595657. doi:10.3389/fgene.2020.595657

Zhang, E., He, J., Zhang, H., Shan, L., Wu, H., Zhang, M., et al. (2020). Immune-Related Gene-Based Novel Subtypes to Establish a Model Predicting the Risk of Prostate Cancer. *Front. Genet.* 11, 595657. doi:10.3389/fgene.2020.595657

Zhang, E., Shiori, F., Mu, O. Y., He, J., Ge, Y., Wu, H., et al. (2021). Establishment of Novel DNA Methylation-Based Prostate Cancer Subtypes and a Risk-Predicting Eight-Gene Signature. *Front. Cell Dev. Biol.* 9 (303), 639615. doi:10.3389/fcell.2021.639615

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.

Copyright © 2021 Zhang, Shiori, Zhang, Wang, He, Ge, Song and Shan. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fmolb.2021.689037

Analyzing and Validating the Prognostic Value of a TNF-Related Signature in Kidney Renal Clear Cell Carcinoma

Wenhao Zhang[1,2†], Changjiu Li[3†], Fanding Wu[4], Ning Li[2], Yuwei Wang[1], Yixuan Hu[3], Tiantian Fang[1], Hui Yuan[1] and Huadong He[1,2]*

[1]The Fourth School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China, [2]Department of Urology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3]Department of Urology, Affiliated Hangzhou First People's Hospital, Nanjing Medical University, Hangzhou, China, [4]School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

Background: Kidney renal clear cell carcinoma (KIRC) has the highest incidence rate in renal cell carcinoma (RCC). Although bioinformatics is widely used in cancer, few reliable biomarkers of KIRC have been found. Therefore, continued efforts are required to elucidate the potential mechanism of the biogenesis and progression of KIRC.

Methods: We evaluated the expression of tumor necrosis factor (TNF) family genes in KIRC, and constructed a prognostic signature. We validated the signature by another database and explored the relationship between the signature and progression of KIRC. We assessed the prognostic value, immune infiltration, and tumor mutation burden (TMB) of the signature in KIRC.

Results: We selected four key genes (*TNFSF14*, *TNFRSF19*, *TNFRSF21*, and *EDA*) to construct the TNF-related signature. We divided the KIRC patients into high- and low-risk groups based on the signature. Patients with higher risk scores had shorter overall survival and worse prognosis. With another database, we validated the value of the signature. The signature was considered as an independent risk factor. A higher level of risk score was relevant to higher level of immune infiltration, especially T regulatory cells, $CD8^+$ T cells, and macrophages. The signature was also associated with TMB scores, and it may have an effect on assessing the efficacy of immunotherapy.

Conclusion: This is the first TNF-family-related signature of KIRC and we demonstrated its effectiveness. It played a significant role in predicting the prognosis of patients with KIRC. It also has the potential to become a powerful tool in guiding the immunotherapy of KIRC patients in clinical practice.

Keywords: signature, kidney renal clear cell carcinoma, biomarker, prognosis, immune

OPEN ACCESS

Edited by:
William C Cho, QEH, China

Reviewed by:
Shixiang Wang, ShanghaiTech University, China
Ramya Sivakumar, University of Washington, United States

***Correspondence:**
Huadong He
harry_lee99@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *31 March 2021*
Accepted: *10 May 2021*
Published: *28 May 2021*

Citation:
Zhang W, Li C, Wu F, Li N, Wang Y, Hu Y, Fang T, Yuan H and He H (2021) Analyzing and Validating the Prognostic Value of a TNF-Related Signature in Kidney Renal Clear Cell Carcinoma. Front. Mol. Biosci. 8:689037. doi: 10.3389/fmolb.2021.689037

Abbreviations: CI, confidence interval; DR6, death receptor 6; GO, gene ontology; HR, hazard ratios; HVEM, herpes virus entry mediator; ICGC, the International Cancer Genome Consortium; ICI, immune checkpoint inhibitors; JNK, c-Jun N-terminal kinase; KEGG, the Kyoto Encyclopedia of Genes and Genomes; KIRC, kidney renal clear cell carcinoma; LTβR, lymphotoxin-β receptor; MAPK, mitogen-activated protein kinase; NF-κB, nuclear factor kappa-B; NK cells, natural killer cells; OS, overall survival; RCC, renal cell carcinoma; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TMB, tumor mutation burden; TNF, the tumor necrosis factor; TNFRSF, the tumor necrosis factor receptor superfamily; TNFSF, the tumor necrosis factor ligands superfamily.

INTRODUCTION

Kidney carcinoma ranks as the 16th most common cause of cancer mortality worldwide (Znaor et al., 2015). Renal cell carcinoma (RCC) is the most common type of renal carcinoma and is responsible for up to 90% of cases (Ljungberg et al., 2019). The main pathological type of RCC is kidney renal clear cell carcinoma (KIRC), which accounts for 70–80% of cases (Nerich et al., 2014). With the further exploration of the tumor microenvironment and the development of immunotherapy, the interaction between tumor and immune system has been studied in depth (Galon and Bruni, 2019). However, the cell phenotypes and corresponding molecular mechanism of KIRC have not been established (Du et al., 2017). There are few reliable biomarkers to predict prognosis and immunotherapeutic response. Therefore, continued efforts are required to elucidate the potential mechanism of the biogenesis and progression of KIRC.

The tumor necrosis factor (TNF) family comprises the 19 TNF ligands superfamily (TNFSF) and 29 TNF receptor superfamily (TNFRSF), and is one of the best-studied protein families over the past 3 decades (Feng, 2005). Previous research has shown that TNF is an inflammatory regulator that can activate immune cells through TNFR1 and TNFR2 (Al-Lamki et al., 2010). Besides, recent studies have found that a variety of cancers, including KIRC, are closely related to the TNF family (Croft et al., 2013). We assumed that KIRC has a potential relationship with the TNF family and performed this study for validation.

In this study, we evaluated mRNA expression data, clinical information, and mutation data of KIRC patients from the TCGA database. We constructed a prognostic multi-gene signature with differentially expressed genes of the TNF family and verified the efficacy of the signature. Functional enrichment analysis, immune infiltration, and tumor mutation burden (TMB) were used to explore the underlying mechanisms of the signature.

MATERIALS AND METHODS

Acquisition of Patient Materials

The transcriptome profiles with HTSeq-FPKM format of KIRC patients were obtained from the TCGA database via the GDC portal (https://portal.gdc.cancer.gov/). The corresponding clinical data were also downloaded from the TCGA database. Data collected from the TCGA database were used as the training set. The transcriptome expression data and corresponding clinical information of KIRC patients obtained from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/projects) were considered as the validation set. All patients with incomplete data were excluded.

Construction of Signature and Survival Analysis

Cox's proportional hazards regression model was utilized to construct the signature. We used univariate Cox analysis to identify the prognostic genes. Genes considered significant with a cutoff point of $p < 0.05$ were selected to built a stepwise Cox regression model. According to the result, we applied the following formula to calculated the risk score:

$$\left[\text{Risk score} = \sum_{i=1}^{n} \text{Coef}(i) * x(i)\right]$$

Coef (i) and x (i) represented estimated regression value. Patients were divided into high- and low-groups by the median risk score. A Kaplan–Meier curve was drawn using the R package "survival" to compare the survival difference of the two groups. Receiver operating characteristic (ROC) curve was drawn by the R package "survivalROC" to assess the predictive effect of the signature on overall survival (OS).

Validation of the Signature

Validation data were downloaded from the ICGC database. The risk score of each patient was calculated with the same genes and coefficient score based on the signature. Kaplan–Meier and ROC curves were drawn to verify the predictive value of the signature.

The Signature Acts as an Independent Risk Factor

Univariate analysis and stepwise Cox regression model were performed to explore whether the TNF-related signature could be an independent risk factor of other clinical characteristics (including age, gender, grade, and stage) in the TCGA database. Patients in the TCGA database were classified into age ≤65 years and >65 years subgroups, female and male subgroups, G1/2 and G3/4 subgroups, stage I/II and III/IV subgroups, and high- and low-risk subgroups. OS analysis via R package "survival" was utilized in every subgroup.

Functional Enrichment Analysis

The KIRC patients were divided into high- and low-risk groups by the TNF-related signature. Then we applied gene ontology (GO) enrichment analysis to identify the biological processes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was utilized to establish the main signaling pathways regulated by the signature.

Immune Cell Infiltration Analysis

The transcriptome gene expression data of KIRC patients downloaded from the TCGA database was normalized via "limma" package. Then CIBERSORT algorithm was utilized to evaluate the immune infiltration. The CIBERSORT was based on the known reference set which containing 22 leukocyte subtypes (LM22). Wilcoxon rank-sum test was performed to calculate the infiltration difference between high- and low- risk groups, and the result was exhibited by "vioplot" package.

Tumor Mutation Burden Analysis

TMB was considered a measurement to calculate the total number of mutations in per million somatic genes. We downloaded the tumor mutation data of KIRC from the TCGA database and calculated the mutation rate of each

TABLE 1 | Characteristics of patients with KIRC.

Characteristics	Variable	Total	Percentages (%)
Age	≦65	352	65.55
	>65	185	35.45
Gender	Male	346	64.43
	Female	191	35.57
Grade	Grade 1	14	2.61
	Grade 2	230	42.83
	Grade 3	207	38.55
	Grade 4	78	14.52
	G X	5	0.93
	Unknown	3	0.56
Stage	Stage I	269	50.09
	Stage II	57	10.61
	Stage III	125	23.28
	Stage IV	83	15.46
	Unknown	3	0.56
T	T1	275	51.21
	T2	69	12.85
	T3	182	33.89
	T4	11	2.05
N	N0	240	44.69
	N1	17	3.17
	NX	280	52.14
M	M0	426	79.33
	M1	79	14.71
	MX	30	5.59
	Unknown	2	0.37
Survival rate	Survival	367	68.34
	Dead	170	31.66

sample via R package "maftools". We further assessed the mutation discrepancy between high- and low-groups by Wilcoxon test.

Statistical Analysis

All statistical analyses and generation of figures were performed by R software 4.0.2. $p < 0.05$ indicated significant effectiveness.

RESULTS

Construction of Tumor Necrosis Factor-Related Signature by the Cancer Genome Atlas Database

We chose 47 TNF family genes. **Table 1** showed the characteristics of all included patients. We performed a univariate Cox regression analysis and found 15 genes that contained seven TNF family genes and eight TNFRSF family genes (**Table 2**). We chose genes of $p < 0.05$ and built a stepwise Cox regression model to optimize the signature. Four genes were selected to construct the TNF-related signature: *TNFSF14* (TNFSF family), *TNFRSF19* (TNFRSF family), *TNFRSF21* (TNFRSF family), and *EDA* (TNFSF family). Among these four genes, *TNFSF14* was a high-risk factor (hazard ratio [HR] = 1.555, 95% confidence interval [CI] = 1.343–1.801), and *TNFRSF19* (HR = 0.678, 95% CI = 0.584–0.787), *TNFRSF21* (HR = 0.738, 95% CI = 0.638–0.853), and *EDA* (HR = 0.418, 95% CI = 0.303–0.577) were considered as low-risk factors. The risk score formula was formed by the expression level of the four genes and Cox coefficient: risk score = 0.25706 * TNFSF14 – 0.30544 * TNFRSF19– – 0.24573 * TNFRSF21 – 0.33039 * EDA. We divided the KIRC patients into high- and low-risk groups by the median risk score. The distribution characteristics of the four genes and the relevant risk score are shown in **Figures 1**, **2**. Kaplan–Meier analysis was applied to assess the value in predicting OS of KIRC patients. Patients in the low-risk group had better OS (**Figure 3A**). The ROC curve of 5-years OS was plotted to show the prognostic accuracy of the TNF-related signature (AUC = 0.712) (**Figure 3B**). We divided the KIRC patients into early stage (I and II) and advanced stage (III and IV), and applied the TNF-related signature. Higher risk contributed to worse OS regardless of clinical stage (**Figures 3C,D**).

TABLE 2 | 15 genes associated with patients' OS.

Gene	HR	Z	*p* value
CD27	1.125	1.832	0.067
CD70	1.056	1.204	0.229
EDA	0.492	–4.750	<0.001
EDA2R	0.615	–3.935	<0.001
FASLG	1.229	1.922	0.055
TNFRSF9	1.223	2.222	0.026
TNFRSF11B	0.838	–2.400	0.016
TNFRSF18	1.744	5.008	<0.001
TNFRSF19	0.588	–5.906	<0.001
TNFRSF21	0.700	–4.448	<0.001
TNFSF4	1.150	1.515	0.130
TNFSF9	1.138	1.454	0.146
TNFSF13	0.576	–4.855	<0.001
TNFSF13B	1.419	4.132	<0.001
TNFSF14	1.588	6.145	<0.001

HR, hazard ratio; Z, Z tezt.

Validation of the Tumor Necrosis Factor-Related Signature by International Cancer Genome Consortium Database

To validate the prognostic value of the TNF-related signature, we applied a new cohort downloaded from the ICGC database of KIRC patients. We used the same formula as the TCGA cohort to calculate the risk score of each patient and set the median score as the cutoff to divide the patients into high- and low-risk groups. The Kaplan–Meier curve, as expected, showed that the low-risk group had better OS (**Figure 4A**). The ROC curve of 5-year OS was plotted to validate the prognostic accuracy (AUC = 0.607) (**Figure 4B**).

Independence of the Tumor Necrosis Factor-Related Signature as a Risk Factor

To confirm the independence of the TNF-related signature as a risk factor, we utilized univariate and multivariate Cox analysis in the TCGA database (**Figures 5A,B**). The pathological features explored included age, gender, stage, grade, and risk score. The

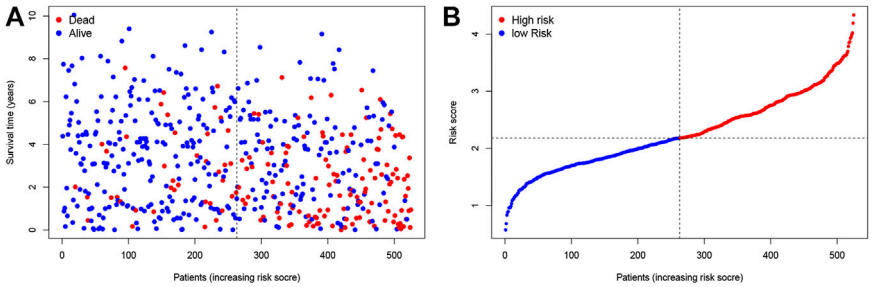


FIGURE 1 | Construction of the TNF-related signature by the TCGA database. **(A, B)** The contribution of risk score and survival status.

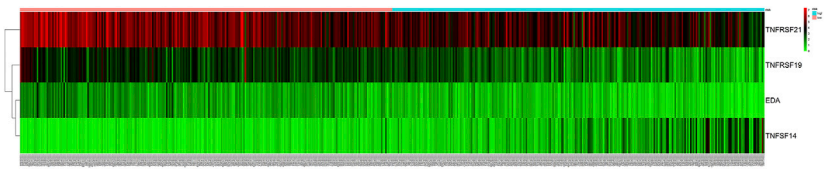


FIGURE 2 | The relationship between survival status and gene expression of the signature. TNFSF14 was highly expressed in high-risk group. TNFRSF19, TNFRSF21, and EDA were highly expressed in low-risk group.

high-risk group showed a difference in age (p = 0.011 HR = 1.495), grade (p = 0.013, HR = 1.610), stage (p < 0.001, HR = 3.144), and risk score (p < 0.001, HR = 1.751). The risk score was effective and the TNF-related signature was identified as an independent risk factor. The relationship between risk score and distribution of clinical characteristics were listed in the heatmap (**Supplementary Figure S1**).

Investigation of the Biological Pathway About the Tumor Necrosis Factor-Related Signature

To investigate further the potential functions of the TNF-related signature, we performed edgeR filtration (false discovery rate < 0.05, $|\log_2 FC > 1|$) and identified 544 differentially expressed genes (DEGs), including 129 negatively related and 415 positively related genes (**Supplementary Table S1**). We applied GO enrichment analysis and KEGG pathway enrichment analysis. In the BP category, the TNF-related signature was highly enriched in humoral immune response, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, and lymphocyte-mediated immunity (**Figure 6A**). In the CC category, the immunoglobulin complex and external side of the plasma membrane were markedly related to the signature. In the MF category, antigen binding, immunoglobulin receptor binding, receptor–ligand activity, and signaling receptor activator activity were highly enriched. KEGG analysis showed that the TNF-related signature was enriched in cytokine–cytokine receptor interaction, viral protein interaction with cytokines and cytokine receptors complement and coagulation cascades, and chemokine signaling pathway (**Figure 6B**).

Immune Infiltration of the Tumor Necrosis Factor-Related Signature

The results above illustrated that the TNF-related signature was significantly associated with immunity. To further explore the relationship between immune status and the signature, we quantified the immune infiltration between high- and low-risk groups through CIBERSORT. **Supplementary Figure S2** showed the fractions of the 22 immune cells in every KIRC patient. Wilcoxon rank-sum test displayed remarkable discrepancy between the two groups. Plasma cells, CD8 T cells, CD4 memory activated T cells, follicular helper T cells, regulatory T cells (Tregs), and M0 were positively associated with the risk score, while monocytes, M1, M2, resting mast cells, and eosinophils were negatively associated with the risk score (**Figure 7**).

Relationship Between Tumor Mutation Burden and Tumor Necrosis Factor-Related Signature

TMB is a novel measurement to calculate the mutations of tumor cells. It can be a specific biomarker to assess the value of cancer immunotherapy (Merino et al., 2020). We applied the TNF-related signature to TMB calculation to explore their correlations. Mutation information of the top 30 most mutated genes was displayed in the waterfall, VHL and PBRM1 took up the majority of the mutation (**Supplementary Figure S3**). The mutations were classified into variant types, with missense mutations making up the majority, single nucleotide polymorphism was the most frequent type and C >

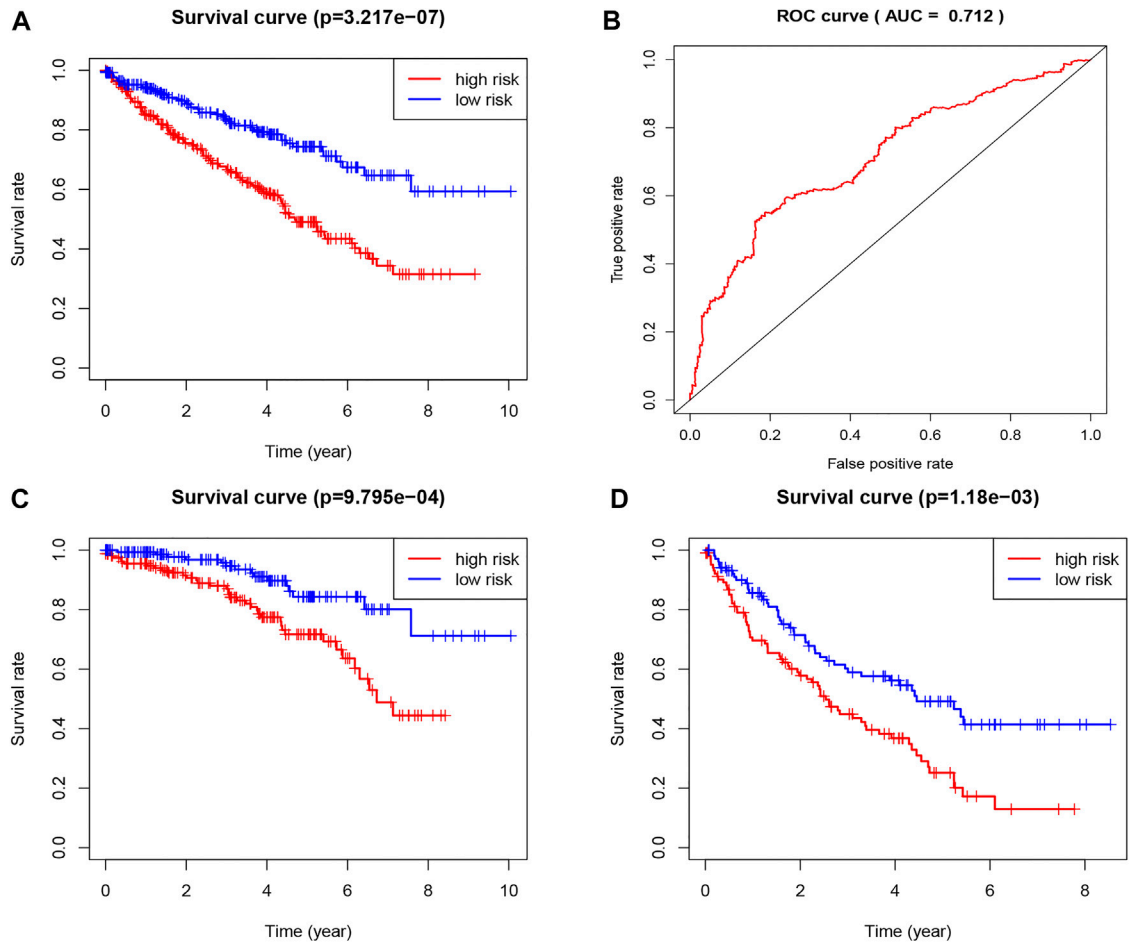


FIGURE 3 | Construction of the TNF-related signature by the TCGA database. **(A)** Kaplan-Meier survival curve of OS in total KIRC patients that classified by the TNF-related signature into high- and low-risk groups. **(B)** ROC curve showing the values of the signature for OS among KIRC patients. **(C)** Kaplan-Meier survival curve of OS in early stage (I and II) KIRC patients that classified by the TNF-related signature into high- and low-risk groups. **(D)** Kaplan-Meier survival curve of OS in advanced stage (III and IV) KIRC patients that classified by the TNF-related signature into high- and low-risk groups.

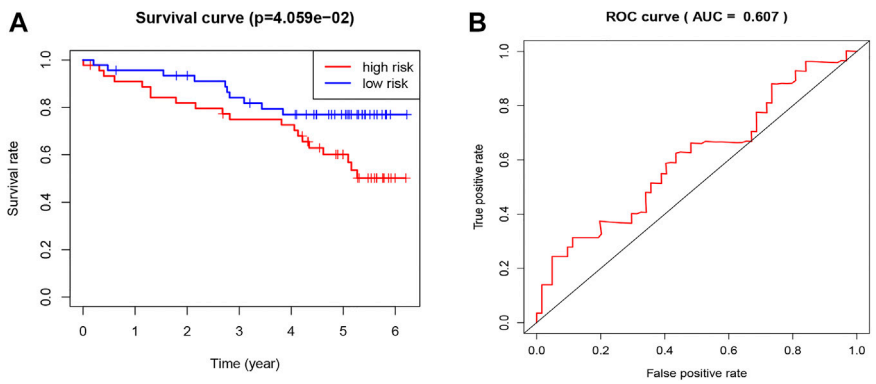


FIGURE 4 | Validation of the TNF-related signature by the ICGC database. **(A)** Kaplan-Meier survival curve of OS in KIRC patients from ICGC database that classified by the TNF-related signature into high- and low-risk groups. **(B)** ROC showing the values of the signature for OS among KIRC patients in the ICGC database.

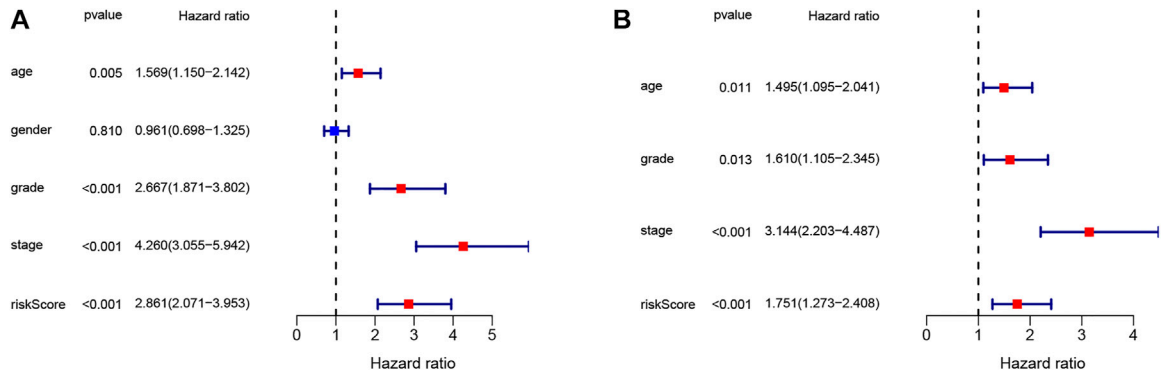


FIGURE 5 | The signature identified as an independent risk factor. **(A)** The result of univariate Cox regression analysis. **(B)** The result of multivariate Cox regression analysis.

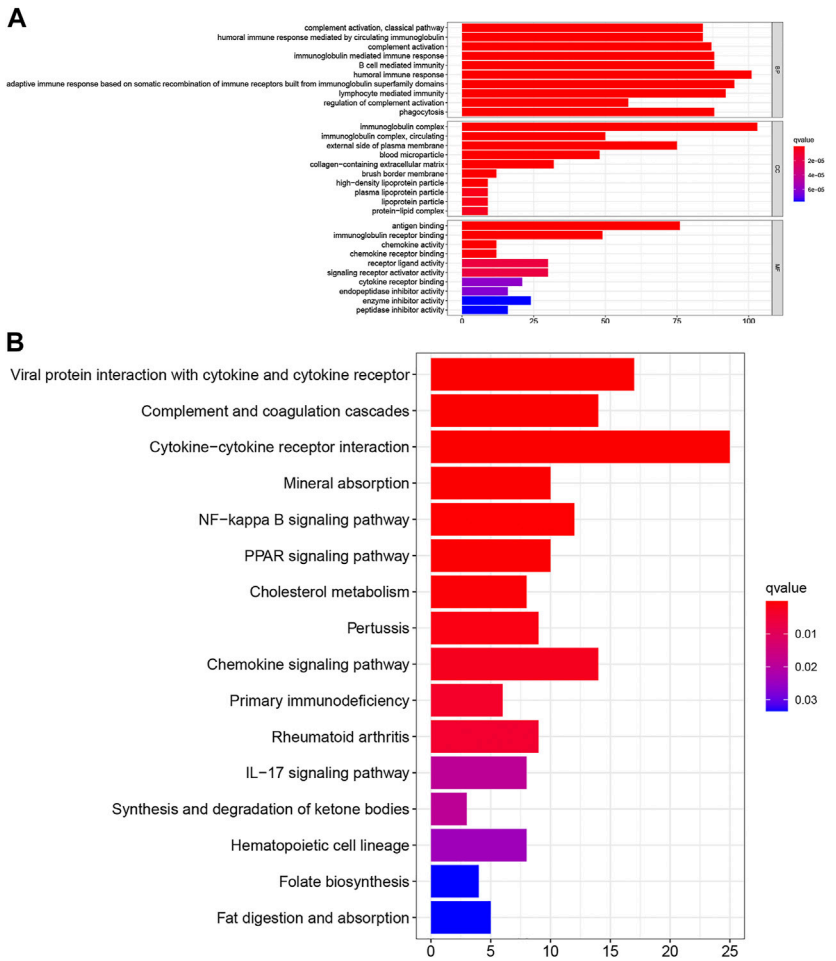


FIGURE 6 | Investigation of the biological pathway about TNF-related signature. **(A)** The top 30 enriched GO analysis of the corresponding genes. **(B)** The top 16 enriched KEGG pathways of the corresponding genes.

T was the most common type of single nucleotide variant. We combined the risk score of patients with their TMB and compared the difference (**Supplementary Figure S4**). Patients with higher risk scores had a higher TMB (**Figure 8**). This meant that relevant immunotherapy might be applied to KIRC patients who were sensitive to the TNF-related signature.

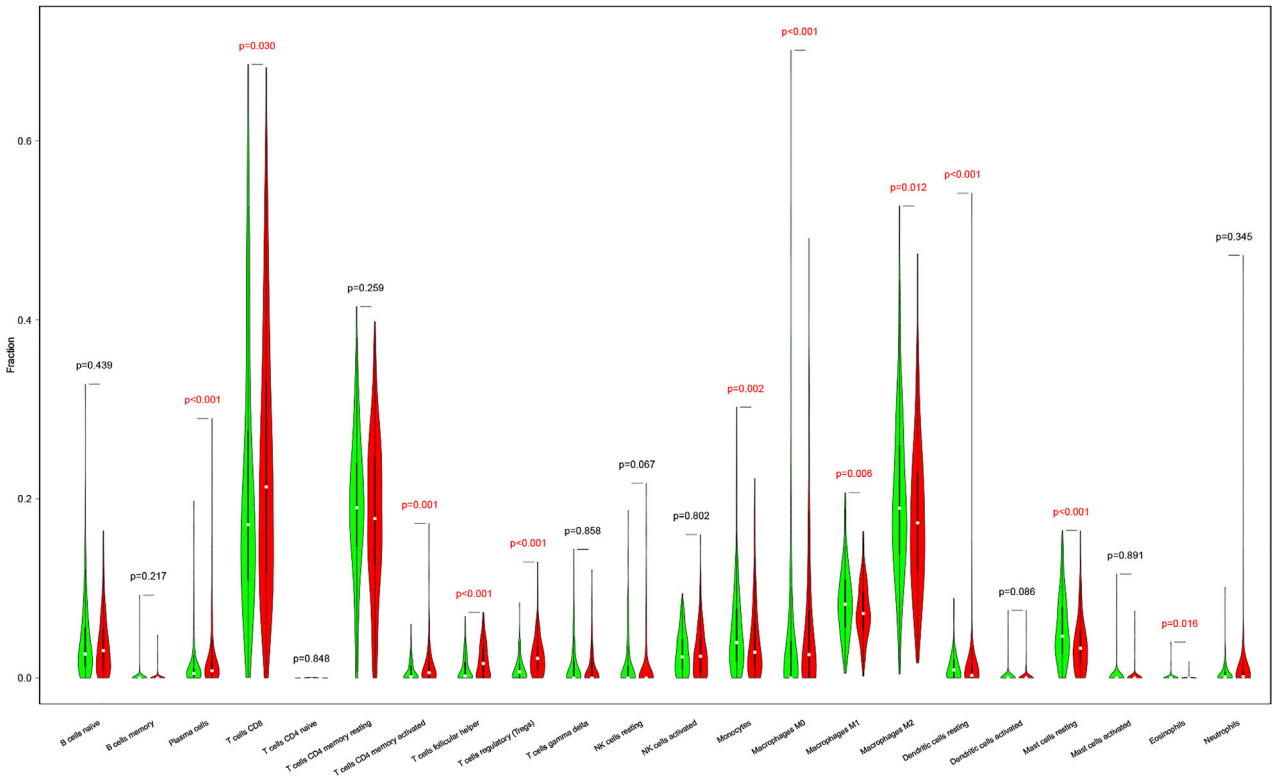


FIGURE 7 | The differences in immune cell infiltration abundances between high- and low-risk patients. Red for the high-risk patients. Green for low-risk patients.

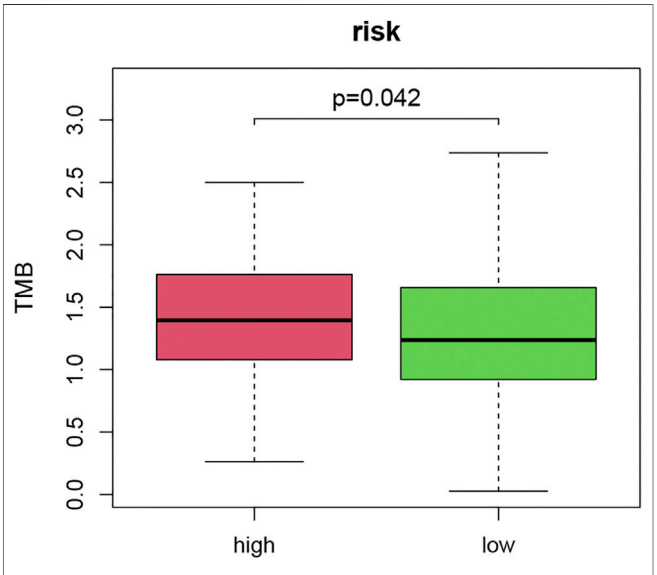


FIGURE 8 | The relationship between TMB score and high- and low-risk patients.

DISCUSSION

Due to the comprehensive research of high-throughput sequencing and bioinformatics technology, more accurate and effective biomarkers have been used in cancer (Sharma, 2018). However, reliable biomarkers for risk assessment are still rare. Exploring new specific biomarkers for KIRC patients is of importance.

The TNF family regulates cell proliferation, migration and apoptosis. Stimulation or inhibition of the TNF superfamily signaling pathway may affect tumor progression. Therefore, we supposed that TNF family members might play an important role in predicting the prognosis of patients with KIRC (Vanamee and Faustman, 2018). We systematically studied the prognostic value of the TNF family genes in the KIRC cohort and found that most were protective factors. This finding was in line with previous studies (Dostert et al., 2019). We created a signature containing four genes (*TNFSF14*, *TNFRSF19*, *TNFRSF21* and *EDA*) to evaluate the prognosis of KIRC patients. *TNFSF14* can activate the immune cells, including T cells, dendritic cells and natural killer cells, by combining herpes virus entry mediator (HVEM) and lymphotoxin-β receptor (Mauri et al., 1998; Rooney et al., 2000; Granger et al., 2001). Previous research has found that *TNFSF14* is part of a cytokine network that participates in the innate and adaptive immune system for immune enhancement (Brunetti et al., 2020). *TNFRSF19* is expressed in the ureteral bud during embryonic development. It is present in the stem cells of adult kidneys to maintain homeostasis and regeneration and to regulate nuclear factor (NF)-κB activity by combining with β-catenin (Schön et al., 2014; Schutgens et al., 2017). *TNFRSF21*, also known as *DR6*, has been revealed to play a

role in activating NF-κB and MAPK8/JNK and inducing apoptosis. It is an α-helical integral membrane receptor protein that inhibits the growth of blood vessels in tumor tissues (Pan et al., 1998). *EDA* regulates the structure and cell number during organ development. It regulates target genes by activating the downstream NF-κB pathway to suppress the proliferation of tumor cells (Sadier et al., 2014).

To explore the efficacy of the signature by combining the four genes above, the survival curve and ROC curve were utilized. The results demonstrated the good performance of the signature ($p < 0.0001$, ROC = 0.712). The ICGC cohort was used to validate the universality of the signature. The result indicated that the signature could be validated by different databases. When comparing our results to previous studies, it should be pointed out that our signature is more universal and efficient.

We discovered that the signature-related genes play critical roles in immunobiological pathways. We further revealed that patients in the high-risk group were in an immune-active state. The immune cells such as Tregs, $CD8^+$ T cells and macrophages were highly expressed in the high-risk group. Tregs play important roles in immune tolerance and immune homeostasis (Takeuchi and Nishikawa, 2016). Previous studies have indicated that in a variety of cancers, such as colon, breast and pancreas cancer, increased Tregs are associated with poor prognosis (Zhuo et al., 2015; Wang et al., 2017). M0 promotes cell proliferation and invasion (Qian and Pollard, 2010), and increased macrophages are associated with poor prognosis in RCC (Hajiran et al., 2020). $CD8^+$ T cells are considered to be the main antitumor cells and preferred targeted immune cells for treating cancer (Farhood et al., 2019). All these results were according to our expectation, indicating that our results proved the validity of the signature and provided a direction for further research, such as the possibility of immunotherapy for KIRC.

TMB has become an emerging biomarker of immunotherapy for many cancers (Chan et al., 2019). A few studies have reported that patients with higher TMB scores benefit more from immunotherapy. To further explore the correlation between the TNF-related signature and prognosis of immunotherapy, we analyzed the discrepancy in TMB score of the two groups. We discovered that the TMB score was significantly higher in the high-risk group, which made it possible to predict the efficacy of immunotherapy.

Combining the results of the TMB and immune infiltration, we found that high-risk patients had an elevated level of TMB, T cells, and B cells, which indicated that patients were in a state of immune activation. Due to the high intrinsic resistance to conventional chemo- and radiotherapies and the rapid development of resistance to targeted therapy, immune checkpoint inhibitors (ICIs) have been one of the few effective therapies for RCC. Previous studies have reported that higher TMB is closely related to better OS after ICI treatment (Valero et al., 2021). The new findings in our study indicated that the signature may be a predictive biomarker to predict the efficacy of immunotherapy.

This study had several limitations. All the samples used to establish and verify the signature were retrospective samples, therefore validation by prospective samples is necessary. Although our study found that the signature might be associated with immunotherapy, the efficacy of the signature cloud not been validated due to the lack of data, the potential mechanism and practical role in clinical practice need further exploration.

In summary, this is believed to be the first study of the TNF-family-based signature for KIRC and we demonstrated its value. It has the potential to become a powerful tool in guiding the immunotherapy of KIRC patients in clinical practice.

DATA AVAILABILITY STATEMENT

Public datasets were analyzed in this study. This data can be found here: TCGA database *via* the GDC portal (https://portal.gdc.cancer.gov/). International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/projects).

AUTHOR CONTRIBUTIONS

WZ and CL contributed equally to the literature research, drafting, interpretation, and writing of the manuscript. FW contributed to the supervision and writing of the manuscript. NL, YW, TF and YH contributed to the literature research of the manuscript. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.689037/full#supplementary-material

REFERENCES

Al-Lamki, R. S., Sadler, T. J., Wang, J., Reid, M. J., Warren, A. Y., Movassagh, M., et al. (2010). Tumor Necrosis Factor Receptor Expression and Signaling in Renal Cell Carcinoma. *Am. J. Pathol.* 177 (2), 943–954. doi:10.2353/ajpath.2010.091218

Brunetti, G., Belisario, D. C., Bortolotti, S., Storlino, G., Colaianni, G., Faienza, M. F., et al. (2020). LIGHT/TNFSF14 Promotes Osteolytic Bone Metastases in Non-small Cell Lung Cancer Patients. *J. Bone Miner Res.* 35 (4), 671–680. doi:10.1002/jbmr.3942

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

Croft, M., Benedict, C. A., and Ware, C. F. (2013). Clinical Targeting of the TNF and TNFR Superfamilies. *Nat. Rev. Drug Discov.* 12 (2), 147–168. doi:10.1038/nrd3930

Dostert, C., Grusdat, M., Letellier, E., and Brenner, D. (2019). The TNF Family of Ligands and Receptors: Communication Modules in the Immune System and beyond. *Physiol. Rev.* 99 (1), 115–160. doi:10.1152/physrev.00045.2017

Du, W., Zhang, L., Brett-Morris, A., Aguila, B., Kerner, J., Hoppel, C. L., et al. (2017). HIF Drives Lipid Deposition and Cancer in ccRCC via Repression of Fatty Acid Metabolism. *Nat. Commun.* 8 (1), 1769. doi:10.1038/s41467-017-01965-8

Farhood, B., Najafi, M., and Mortezaee, K. (2019). CD8+cytotoxic T Lymphocytes in Cancer Immunotherapy: A Review. *J. Cel Physiol* 234 (6), 8509–8521. doi:10.1002/jcp.27782

Feng, X. (2005). Regulatory Roles and Molecular Signaling of TNF Family Members in Osteoclasts. *Gene* 350 (1), 1–13. doi:10.1016/j.gene.2005.01.014

Galon, J., and Bruni, D. (2019). Approaches to Treat Immune Hot, Altered and Cold Tumours with Combination Immunotherapies. *Nat. Rev. Drug Discov.* 18 (3), 197–218. doi:10.1038/s41573-018-0007-y

Granger, S. W., Butrovich, K. D., Houshmand, P., Edwards, W. R., and Ware, C. F. (2001). Genomic Characterization of LIGHT Reveals Linkage to an Immune Response Locus on Chromosome 19p13.3 and Distinct Isoforms Generated by Alternate Splicing or Proteolysis. *J. Immunol.* 167 (9), 5122–5128. doi:10.4049/jimmunol.167.9.5122

Hajiran, A., Chakiryan, N., Aydin, A. M., Zemp, L., Nguyen, J., Laborde, J. M., et al. (2021). Reconnaissance of Tumor Immune Microenvironment Spatial Heterogeneity in Metastatic Renal Cell Carcinoma and Correlation with Immunotherapy Response. *Clin. Exp. Immunol.* 204, 96–106. doi:10.1111/cei.13567

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernández-Pello, S., et al. (2019). European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur. Urol.* 75 (5), 799–810. doi:10.1016/j.eururo.2019.02.011

Mauri, D. N., Ebner, R., Montgomery, R. I., Kochel, K. D., Cheung, T. C., Yu, G.-L., et al. (1998). LIGHT, a New Member of the TNF Superfamily, and Lymphotoxin α Are Ligands for Herpesvirus Entry Mediator. *Immunity* 8 (1), 21–30. doi:10.1016/s1074-7613(00)80455-0

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 (1). doi:10.1136/jitc-2019-000147

Nerich, V., Hugues, M., Nai, T., Stein, U., Nguyen Tan Hon, T., Montcuquet, P., et al. (2014). Clinical Impact of Targeted Therapies in Patients with Metastatic clear-cell Renal Cell Carcinoma. *Ott* 7, 365–374. doi:10.2147/ott.S56370

Pan, G., Bauer, J. H., Haridas, V., Wang, S., Liu, D., Yu, G., et al. (1998). Identification and Functional Characterization of DR6, a Novel Death Domain-Containing TNF Receptor. *FEBS Lett.* 431 (3), 351–356. doi:10.1016/s0014-5793(98)00791-1

Qian, B.-Z., and Pollard, J. W. (2010). Macrophage Diversity Enhances Tumor Progression and Metastasis. *Cell* 141 (1), 39–51. doi:10.1016/j.cell.2010.03.014

Rooney, I. A., Butrovich, K. D., Glass, A. A., Borboroglu, S., Benedict, C. A., Whitbeck, J. C., et al. (2000). The Lymphotoxin-β Receptor Is Necessary and Sufficient for LIGHT-Mediated Apoptosis of Tumor Cells. *J. Biol. Chem.* 275 (19), 14307–14315. doi:10.1074/jbc.275.19.14307

Sadier, A., Viriot, L., Pantalacci, S., and Laudet, V. (2014). The Ectodysplasin Pathway: from Diseases to Adaptations. *Trends Genet.* 30 (1), 24–31. doi:10.1016/j.tig.2013.08.006

Schön, S., Flierman, I., Ofner, A., Stahringer, A., Holdt, L. M., Kolligs, F. T., et al. (2014). β-Catenin Regulates NF-Kb activityviaTNFRSF19 in Colorectal Cancer Cells. *Int. J. Cancer* 135 (8), 1800–1811. doi:10.1002/ijc.28839

Schutgens, F., Rookmaaker, M. B., Blokzijl, F., van Boxtel, R., Vries, R., Cuppen, E., et al. (2017). Troy/TNFRSF19 marks Epithelial Progenitor Cells during Mouse Kidney Development that Continue to Contribute to Turnover in Adult Kidney. *Proc. Natl. Acad. Sci. U S A* 114 (52), E11190–e11198. doi:10.1073/pnas.1714145115

Sharma, A. K. (2018). Emerging Trends in Biomarker Discovery: Ease of Prognosis and Prediction in Cancer. *Semin. Cancer Biol.* 52 (Pt 1), iii–iv. doi:10.1016/j.semcancer.2018.05.008

Takeuchi, Y., and Nishikawa, H. (2016). Roles of Regulatory T Cells in Cancer Immunity. *Int. Immunol.* 28 (8), 401–409. doi:10.1093/intimm/dxw025

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

Vanamee, É. S., and Faustman, D. L. (2018). Structural Principles of Tumor Necrosis Factor Superfamily Signaling. *Sci. Signal.* 11 (511), eaao4910. doi:10.1126/scisignal.aao4910

Wang, H., Franco, F., and Ho, P.-C. (2017). Metabolic Regulation of Tregs in Cancer: Opportunities for Immunotherapy. *Trends Cancer* 3 (8), 583–592. doi:10.1016/j.trecan.2017.06.005

Zhuo, C., Xu, Y., Ying, M., Li, Q., Huang, L., Li, D., et al. (2015). FOXP3+ Tregs: Heterogeneous Phenotypes and Conflicting Impacts on Survival Outcomes in Patients with Colorectal Cancer. *Immunol. Res.* 61 (3), 338–347. doi:10.1007/s12026-014-8616-y

Znaor, A., Lortet-Tieulent, J., Laversanne, M., Jemal, A., and Bray, F. (2015). International Variations and Trends in Renal Cell Carcinoma Incidence and Mortality. *Eur. Urol.* 67 (3), 519–530. doi:10.1016/j.eururo.2014.10.002

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.

Copyright © 2021 Zhang, Li, Wu, Li, Wang, Hu, Fang, Yuan and He. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fmolb.2021.687576

A Comprehensive Analysis of the Downregulation of *miRNA-1827* and Its Prognostic Significance by Targeting *SPTBN2* and *BCL2L1* in Ovarian Cancer

Penghui Feng[1†], *Zhitong Ge*[2†], *Zaixin Guo*[1], *Lin Lin*[1,3] *and Qi Yu*[1*]

[1]*Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China,* [2]*Department of Ultrasound, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China,* [3]*Department of Obstetrics and Gynecology, The Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Beijing, China*

OPEN ACCESS

Edited by:
Hossain Shekhar,
University of Dhaka, Bangladesh

Reviewed by:
Udhaya Kumar S.,
Vellore Institute of Technology, India
Sanjay Mishra,
The Ohio State University, United States

***Correspondence:**
Qi Yu
yuqi2008001@sina.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *29 March 2021*
Accepted: *19 May 2021*
Published: *11 June 2021*

Citation:
Feng P, Ge Z, Guo Z, Lin L and Yu Q (2021) A Comprehensive Analysis of the Downregulation of miRNA-1827 *and Its Prognostic Significance by Targeting* SPTBN2 *and* BCL2L1 *in Ovarian Cancer. Front. Mol. Biosci. 8:687576. doi: 10.3389/fmolb.2021.687576*

Background: Previous studies demonstrated that *miRNA-1827* could repress various cancers on proliferation, angiogenesis, and metastasis. However, little attention has been paid to its role in ovarian cancer as a novel biomarker or intervention target, especially its clinical significance and underlying regulatory network.

Methods: A meta-analysis of six microarrays was adopted here to determine the expression trend of *miRNA-1827*, and was further validated by gene expression profile data and cellular experiments. We explored the functional annotations through enrichment analysis for the differentially expressed genes targeted by *miRNA-1827*. Subsequently, we identified two hub genes, *SPTBN2* and *BCL2L1*, based on interaction analysis using two online archive tools, miRWALK (it consolidates the resources of 12 miRNA-focused servers) and Gene Expression Profiling Interactive Analysis (GEPIA). Finally, we validated their characteristics and clinical significance in ovarian cancer.

Results: The comprehensive meta-analysis revealed that *miRNA-1827* was markedly downregulated in clinical and cellular specimens. Transfection of the *miRNA-1827* mimic could significantly inhibit cellular proliferation. Concerning its target genes, they were involved in diverse biological processes related to tumorigenesis, such as cell proliferation, migration, and the apoptosis signaling pathway. Moreover, interaction analysis proved that two hub genes, *SPTBN2* and *BCL2L1*, were highly associated with poor prognosis in ovarian cancer.

Conclusion: These integrated bioinformatic analyses indicated that *miRNA-1827* was dramatically downregulated in ovarian cancer as a tumor suppressor. The upregulation of its downstream modulators, *SPTBN2* and *BCL2L1*, was associated with an unfavorable prognosis. Thus, the present study has identified *miRNA-1827* as a potential intervention target for ovarian cancer based on our bioinformatic analysis processes.

Keywords: miRNA-1827, microRNA-1827, SPTBN2, BCL2L1, Ovarian Cancer

INTRODUCTION

Ovarian cancer (OV), as one of the three major gynecological malignancies, poses a severe threat to women's life quality and reproductive health (Siegel et al., 2019). Annually, there are ~300,000 new cases worldwide identified as OV, with the highest mortality rate among all gynecological malignancies (Bray et al., 2018). OV is classified into various histopathological types, and epithelial cancer is the most common type. Currently, nearly 65–75% of patients are already in the middle or advanced stages when they are first confirmed because the tumor lesions are always located deep in the pelvic cavity, and there are no practical approaches for screening and diagnosis for early onset. Although ~75% of patients can benefit from the initial treatment with impermanent remission, almost all terminal cases will develop into the recurrent and multidrug-resistant status, and the 5-year survival rate of OV patients is merely 30–50% (Salani et al., 2017; Allemani et al., 2018).

Recently, the development of bioinformatics has promoted the exploration of various kinds of cancers by integrating data resources and clinical information, which assists experimental biologists in carrying forward these findings into clinical validation and application (Fu et al., 2020; Udhaya Kumar et al., 2020a). Previous studies have determined the correlation between the aberrant expression of specific genes and oncogenesis and progression of OV, and unveiled the predictive potential of these biomarkers for prognosis (Zheng et al., 2019; Yang et al., 2020). As reported, a study has analyzed energy metabolism–associated characteristics to evaluate the prognosis of patients with OV by nonnegative matrix factorization clustering analysis; it finally established an eight-gene signature associated with metabolic genes (Wang and Li, 2020). Another study has also identified four core genes associated with the prognosis of OV using the Analyze Networks algorithm (Kumar et al., 2019). At the same time, mounting evidence has revealed the latent significance of miRNAs as new diagnostic and prognostic markers based on bioinformatic analyses for the individualized treatment of OV.

miRNAs are single-stranded, noncoding, small RNAs of approximately 22 nucleotides in length, generated from endogenous heparin loop miRNA precursors, which are enrolled in the regulation of gene expression at the transcriptional or posttranscriptional level (Kim, 2005; Graves and Zeng, 2012). Additionally, existing research recognizes the pivotal role of miRNAs in many biological behaviors of cancers, as penetratingly analyzed in the previous studies, including angiogenesis, tissue invasion or metastasis, deregulating cell energetics, and avoiding immune destruction (Bueno et al., 2008; Nicoloso et al., 2009; Hanahan and Weinberg, 2011; Fei et al., 2012; Wei et al., 2013). In particular, the aberrant expression of some ovary-specific miRNAs can be a crucial putative biomarker or indicator in the diagnosis, individualized treatment, and prognosis of OV (Pal et al., 2015). For example, *miR-182* was confirmed to be upregulated in OV cases using a public dataset from the Gene Expression Omnibus (GEO) database, and two hub genes, *MCM3* and *GINS2*, were indicated to be associated with the worse overall survival of patients (Li and Li, 2019). Based on these predictive hallmarks, miRNAs involved in tumorigenesis and progression have aroused wide attention.

microRNA-1827 or *miRNA-1827*, as a novel miRNA, acts with multiple biological functions in various types of cancers. *miRNA-1827* was initially identified in HeLa cells by miRDeep sequencing (Friedlander et al., 2008). A recent study demonstrated that it could tightly regulate the expression level and function of *TP53* to suppress the oncogenesis of human colorectal cancer by targeting *MDM2*, which could bind to and degrade the P53 protein under ubiquitylation (Zhang et al., 2016). Furthermore, similar research in ulcerative colitis patients also indicated that *miRNA-1827* in serum was decreased and was correlated with a higher risk of colorectal cancer (Polytarchou et al., 2015). Besides, *miRNA-1827* was a putative therapeutic target in lung cancer *in vitro* and *in vivo*, and it could modulate the migration, angiogenesis, or invasion directly by targeting the *CRKL* molecule (Ho et al., 2014; Ho et al., 2018). However, as much as it has been reported to be downregulated in most types of cancers as a tumor repressor, very few studies have investigated its role in OV. Therefore, it will be essential to verify its clinical significance as a new intervention target and reveal the underlying mechanisms by screening out the possible targets of *miRNA-1827* for further application.

This study assessed the influences of the differential expression of *miRNA-1827*, with all data collected from six GEO microarrays. To illuminate the potential modulatory mechanisms, the downstream regulators targeted by *miRNA-1827* were predicted and verified by 12 prophetic databases together with the Gene Expression Profiling Interactive Analysis (GEPIA) database (the latter is based on the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) data). Subsequently, Gene Oncology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG), and protein–protein interaction (PPI) network analysis were comprehensively performed.

MATERIALS AND METHODS

Gene Expression Omnibus Microarray Acquisition

To retrieve the microarray data on *miRNA-1827* in OV, we searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) with the keywords as listed: ((“ovarian” OR “ovary” OR “ovaries”) AND (“cancer” OR “cancerous” OR “carcinoma” OR “carcinomatous” OR “neoplasm” OR “neoplasms” OR “adenoma” OR “adenomas” OR “adenocarcinoma” OR “adenocarcinomas” OR “malignancy” OR “tumor” OR “tumors”) AND (“micro RNA” OR “microRNA” OR “miRNA” OR “non-coding RNA” OR “ncRNA” OR “small RNA”)). The retrieval expression was filtered by “*Homo sapiens*” of the organism. In addition, the inclusion criteria were set as follows: 1) patients diagnosed with OV and its subtypes were included; 2) the experiments were performed with cancerous samples and normal control (NC), and for at least three, biological

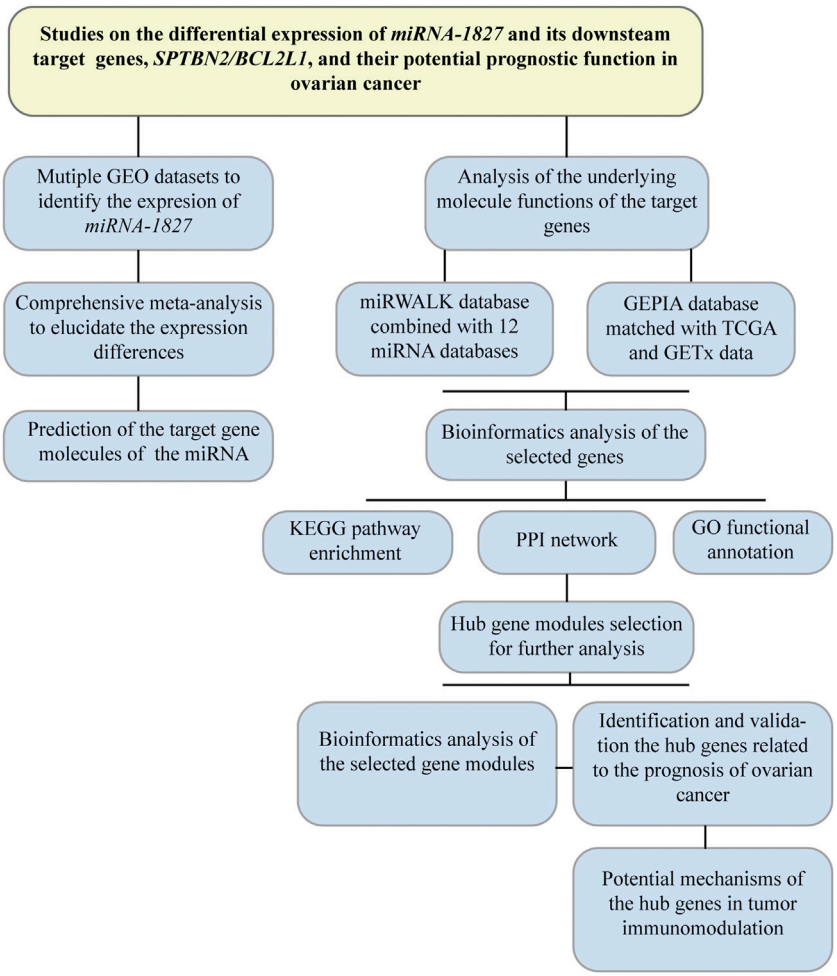


FIGURE 1 | Flow diagram of the research design. This flowchart presented a comprehensive meta-analysis to validate the expression among different microarrays in OV, and bioinformatic analysis was performed to investigate the putative target genes. Hub genes were filtered out for further research on their prognostic significance.

duplications were investigated; 3) the bio-specimens were isolated from tissue, serum, or urine; and 4) the expression profiling data of *miRNA-1827* were available. All data were processed according to the flow diagram, as shown in **Figure 1**.

Statistical Comparison and Comprehensive Meta-Analysis for Microarrays

The expression profiling results of *miRNA-1827* derived from the queried microarrays were aggregated to calculate the number, mean value (M), and standard deviation (SD) for comparisons between the two groups by the Student's t-test, and the expressional results were displayed using the ggpubr package. Then, a comprehensive meta-analysis by the meta package was performed to evaluate the data from all selected microarrays, for which the expression of *miRNA-1827* between groups was illustrated by the forest plot that displayed the standardized mean difference (SMD) and the 95% confidential interval. It should be mentioned that the DerSimonian–Laird estimator for τ^2 and I^2 statistics was applied to determine whether to adopt a random-effects model or fixed-effects model for the pooled estimate. The sensitivity analysis was conducted to assess the heterogeneity among different studies, and if it did exist, the subgroup analysis was carried out to ascertain the reason for the heterogeneity. To measure publication bias, Begg's test and funnel plots were used.

Roles of *miRNA-1827* in the Proliferation of Ovarian Cancer Cells

To verify the expression level of *miRNA-1827* in OV cells, five types of OV cell lines (OV1063, Caov3, SKOV3, OVCAR3, and A2780 cells) and two kinds of human normal ovarian epithelial cells (IOSE80 and HOSEpiC) were used for further experiments. As directed, these cells were cultured in the recommended medium in a humidified atmosphere supplemented with 5% CO_2. Cells were passaged by trypsin digestion at 80% confluence. To quantify the content of *miRNA-1827*, the total miRNAs were extracted by an miRNA Isolation Kit (Vazyme, China). The purified miRNAs were then pretreated with DNase I

TABLE 1 | Gene Expression Omnibus datasets involved.

Accession	GPL	Year	Ovarian cancer			Normal control			Source
			n	M	SD	*n*	M	SD	
GSE119056	GPL21572	2019	6	3.89	0.59	3	11.01	9.18	Tissue
GSE83693	GPL22079	2017	16	0.27	0.16	4	1.10	0.07	Tissue
GSE53829	GPL18138	2014	45	30.99	2.05	14	31.26	2.35	Tissue
GSE47841	GPL14613	2014	21	2.20	0.52	9	3.00	0.93	Tissue
GSE48485	GPL14943	2014	5	−1.82	0.72	5	−1.43	0.63	Serum
GSE58517	GPL18402	2015	5	−9.73	1.78	5	−5.92	7.64	Urine

M, mean; SD, standard deviation.

(5 U/μl) solution. The miRNA first-strand synthesis kit (Clontech, United States) was applied to convert miRNAs into cDNA. After that, a qRT-PCR reaction was carried out using the Mir-X miRNA qRT-PCR TB Green® kit (Clontech, United States). PCR primer sequences (*miRNA-1827* forward primer: GCAGTGAGGCAGTAGATTG, *miRNA-1827* reverse primer was provided by the kit; *U6* forward primer GGAACGATA CAGAGAAGATTAGC, *U6* reverse primer: TGGAACGCT TCACGAATTTGCG) were designed using miRprimer2 software (Busk, 2014). Cycling conditions for the PCR machine were set as follows: pre-denaturation at 95°C for 30 s, following amplification condition at 95°C for 5 s, and 60°C for 34 s for 40 cycles. Gene amplification levels were quantified by the delta–delta CT method and standardized to the reference gene.

Based on the quantification results, A2780 and OV1063 cells were seeded onto 25 cm^2 culture flasks and harvested after rinsing with PBS. Cells were then cultured in 96-well flat plates with 100 μl in each well. In order to determine the effects of *miRNA-1827* on the proliferation of OV cells, the riboFECT CP transfection kit (RiboBio, China) was used to transfect the *miRNA-1827* mimic or inhibitor into A2780 and OV1063 cells, respectively. In detail, a 50 nM mimic (sense strand: UGAGGCAGUAGAUUGAAU, antisense strand: ACUCCGUCAUCUAACUUA) or 100 nM inhibitor (sequence: mAmUmUmCmAmAmUmCmUm AmCmUmGmCmCmUmCmAm) was added into the corresponding medium mixed with 6 μl of buffer and 0.6 μl of transfection reagent. Cells were incubated for 48 h. Two hours before the end of the incubation, 10 μl of CCK-8 reagent was added (Vazyme, China) into each well. After that, the optical density was detected at 450 nm using a microplate reader.

Putative Targets of *miRNA-1827* in Ovarian Cancer

To locate the targets for *miRNA-1827* based on miRNA–gene interactions, data mining was processed using an online archived database, miRWALK (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/index.html) (Sticht et al., 2018). Twelve classical miRNA-focused servers, including miRWalk, MicroT4, miRanda, miRBridge, miRDB, miRMap, miRNAMap, PICTAR2, PITA, RNA22, RNAhybrid, and Targetscan, were taken into consideration. Target genes were selected only when they were projected by no less than six of the servers mentioned above. The upregulated genes in OV were hunted by the GEPIA database (http://gepia.cancer-pku.cn/) (Tang et al., 2019) with a fold change value of more than one and an adjusted p-value < 0.05. The putative target genes were finally recognized from the overlapped genes among the predictive targets from the miRWALK database and the overexpressed genes were acquired in the GEPIA database.

Functional Enrichment Analysis for Latent Target Genes

GO term enrichment and KEGG analyses have been widely applied to consolidate biology in compiling a disciplined, structured, and defined glossary for numerous genes. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/) (Dennis et al., 2003) was utilized to associate functional terms with the uploaded genes using the clustering algorithms. The enriched results were classified into the biological process, cellular component, and molecular function terms, as revealed by the bubble plots, and integrated functional pathways, as shown in the chordal graph by ggplot2 and GOplot packages.

Identification and Characteristics of Prognostic Hub Genes

To identify the hub genes that interacted with *miRNA-1827* and were relevant to the clinical prognosis, the latent target genes described above were explored by the PPI analysis *via* the online database, STRING (https://string-db.org/cgi/input.pl) (von Mering et al., 2003). The outcomes were presented by Cytoscape software (https://cytoscape.org/) (Shannon et al., 2003). On this basis, gene modules or clusters were filtered *via* the MCODE plugin with a K-core value > 2. Then, four clusters were screened out, and 34 genes were identified for subsequent bioinformatic analyses. Among these genes, two hub genes, *SPTBN2* and *BCL2L1*, were matched with prognostic significance and validated by the GEPIA survival and correlation analyses.

Validation of the Expression Patterns of the Hub Genes

To validate the expression patterns of the hub genes, gene expression profiles of OV samples were downloaded from the project of the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/) (Zhang et al., 2019). Besides, data

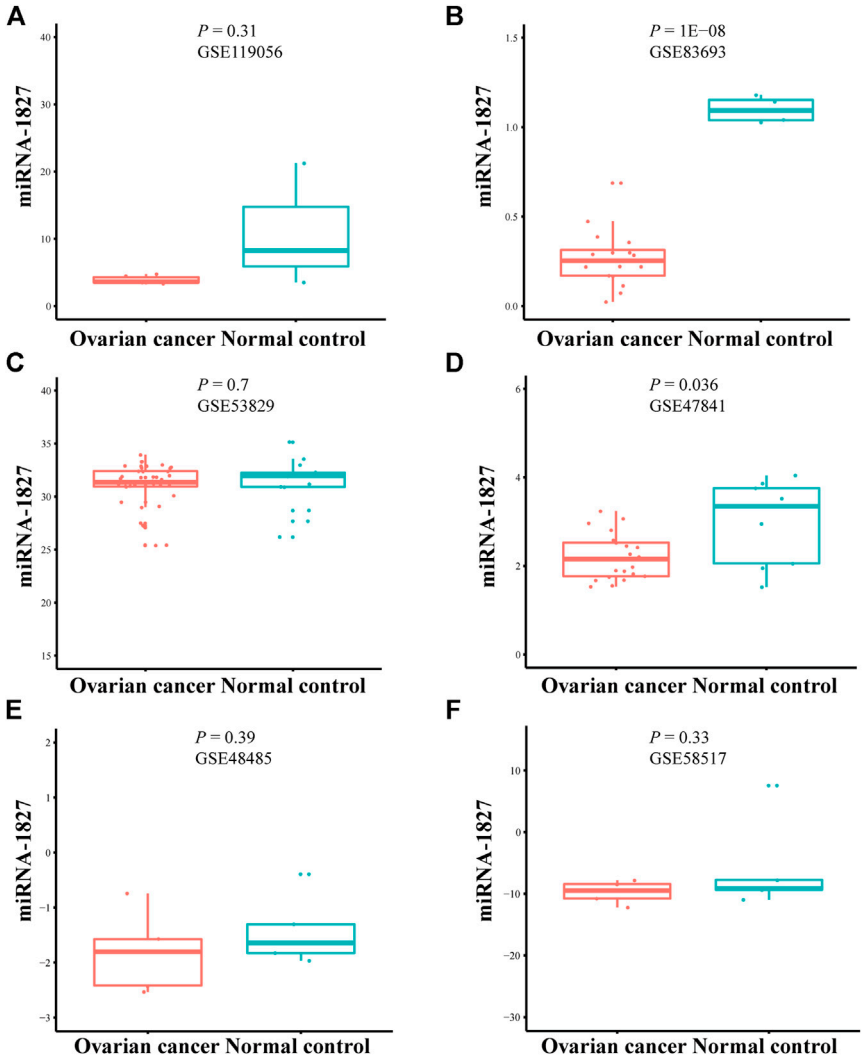


FIGURE 2 | Downregulation of *miRNA-1827* in the selected microarrays from the GEO datasets. **(A)** GSE119056. **(B)** GSE83693. **(C)** GSE53829. **(D)** GSE47841. **(E)** GSE48485. **(F)** GSE58517.

generated from normal tissues were obtained from the GEPIA database as controls. All the data after normalization were processed to identify the expression difference of *SPTBN2* and *BCL2L1* between groups. In order to better understand these two hub genes, the Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) (Uhlen et al., 2015) was further utilized to exploit the relevant information on the location and protein expression in the cancer specimens compared with controls through the cell, tissue, and pathology atlas modules of this tool for further qualitative analyses.

The Tumor Immunoregulatory Network Coordinated by the Hub Genes

To determine the potential mechanisms of the hub genes in the immunoregulatory system in OV, we explored the Tumor Immune Estimation Resource (TIMER) online tool (https://cistrome.shinyapps.io/timer/) (Li et al., 2017). We investigated the correlation between the hub gene expression and abundance of immune infiltration, and the association between clinical outcome and abundance of immune infiltration. The estimate algorithm was applied to determine the immune–stromal component in the tumor microenvironment (TME) of each sample utilizing the estimate package in R software (https://r-forge.r-project.org/). The limma package was applied to normalize the gene expression profile from the TCGA database to evaluate the proportion of tumor-infiltrating immune cells (TICs), and then a standardized gene expression profile was uploaded to CIBERSORT (Newman et al., 2015). The deconvolution algorithm was introduced here to estimate the TIC abundance. According to the average abundance of $CD4^+$ T cells or dendritic cells, Kaplan–Meier analysis by the log-rank test analysis was

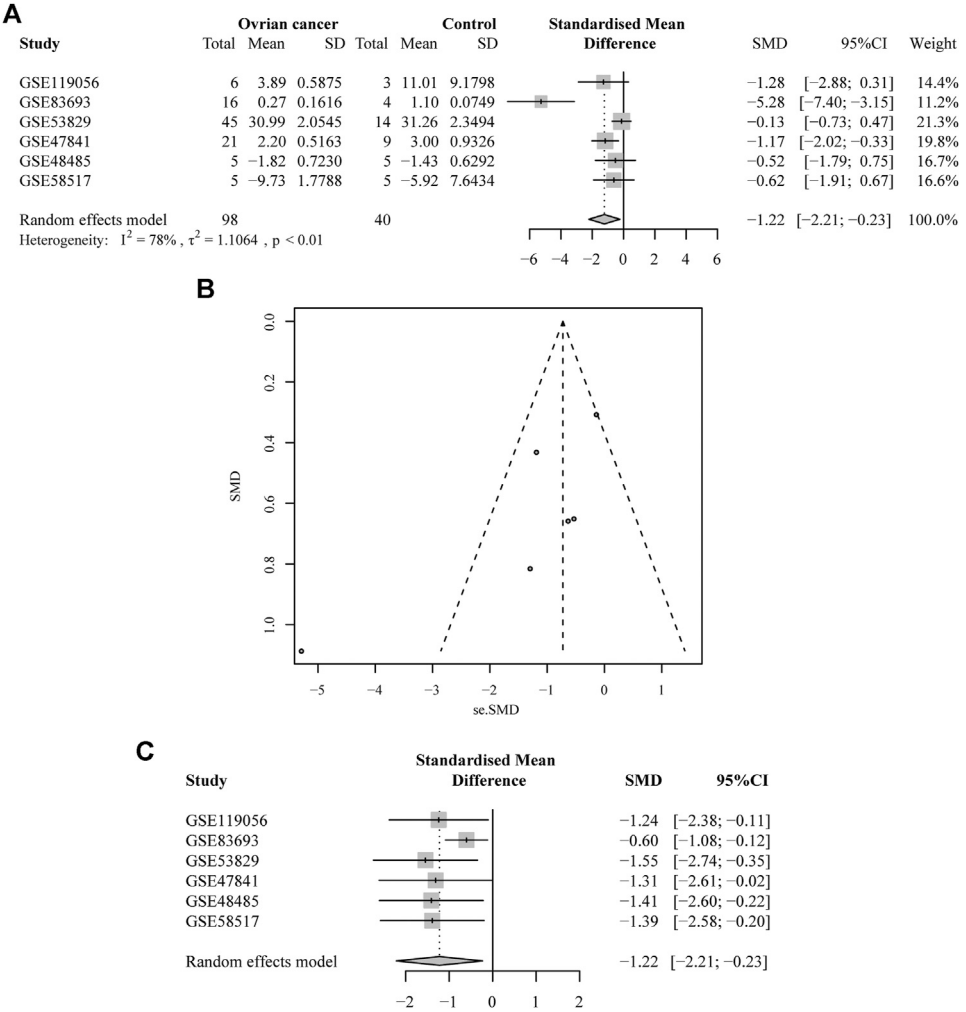


FIGURE 3 | A comprehensive meta-analysis of the expression of *miRNA-1827* from the collected microarrays. **(A)** The forest plot for the pooled SMD of −1.22 (95% CI: −2.21, −0.23) with a degree of heterogeneity (I^2 = 78%, p < 0.01). **(B)** The funnel plot for assessing the publication bias of the GEO datasets by Begg's test with p = 0.09. **(C)** The sensitivity analysis for the GEO microarray results.

carried out to assess survival possibility with the differential expression of *SPTBN2* and *BCL2L1*.

RESULTS

Confirmation of the Expression of *miRNA-1827* and the Meta-Analysis Results

There were six microarrays included in this study from the GEO database, as depicted in **Table 1**. In total, four datasets were collected from tissue samples (GSE119056, GSE83693, GSE53829, and GSE47841), and the rest were from either serum (GSE48485) or urine specimens (GSE58517). Concerning the expression data from those datasets of *miRNA-1827*, this miRNA was downregulated in the OV group compared with the NC group in all enrolled datasets, and two of them were expressed with significance, including GSE83693 (p = 1E-08) and GSE47841 (p = 0.036), as shown in **Figure 2**. On this basis, a comprehensive meta-analysis was performed to precisely quantify the expression of *miRNA-1827*. The calculation results are displayed in **Figure 3A**, for which the random-effects model was applied considering the existing heterogeneity (I^2 = 78%, τ^2 = 1.1064, p < 0.01). Results also indicated that *miRNA-1827* was remarkably downregulated in the OV group (SMD = −1.2239, 95% CI [−2.2145; −0.2333], Z = −2.42, p = 0.0155). To clarify the source of heterogeneity, publication bias was evaluated, as shown in **Figure 3B**. The symmetric funnel plot signified there was no publication bias (z = −1.6908, p-value = 0.09087). Later, a sensitivity test was generated to assess the significant heterogeneity, and no items were found to have a particular effect on the results (**Figure 3C**).

To further validate the heterogeneity, subgroup analysis was conducted according to the source types, divided into two groups (the non-tissue group or tissue group), as depicted in **Figure 4**. The results demonstrated that no heterogeneity was observed in the non-tissue group (SMD = −0.57, 95% CI [−1.47; −0.34], Q = 0.01, τ^2 = 0, I^2 = 0.0%). Conversely, the tissue group was found to

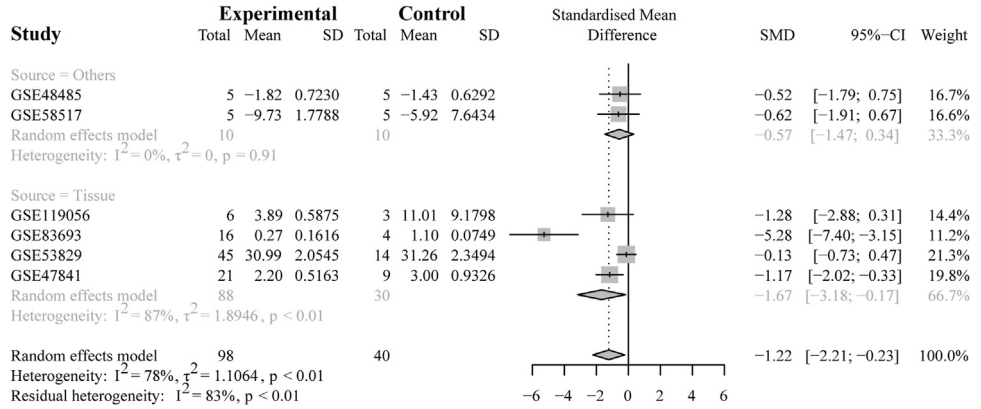


FIGURE 4 | Results of the subgroup analysis. The subgroup analysis based on sample sources (non-tissue or tissue). The result indicated that the tissue subgroup had obvious heterogeneity (I^2 = 87%, p < 0.01).

account for the heterogeneity (SMD = −1.67, 95% CI [−3.18, −0.17], Q = 22.90, τ^2 = 1.8946, I^2 = 87%). In this group, the study (GSE53829) might be the main cause with the largest number of samples included in the microarrays (45 from the OV samples and 14 from the NC samples).

Subsequently, we explored the expression of *miRNA-1827* in OV cells, and our results indicated that these cell lines exhibited consistent trends of declined expression in these five kinds of OV cells compared with the normal ovarian epithelial cells, as shown in **Supplementary Figure S1A**. Notably, there existed significant differences of *miRNA-1827* in A2780 cells. Similarly, the expression level of *miRNA-1827* in OVCAR3 cells differed from that of *miRNA-1827* in HOSEpiC cells. According to these results, two cells, A2780 and OV1063, were chosen to be transfected with the mimic or inhibitor of *miRNA-1827*, respectively. A CCK8 assay was performed after transfection to determine the cell viability, and we found that the proliferation of A2780 cells was dramatically inhibited after being treated with the *miRNA-1827* mimic (p < 0.01). However, the competitive inhibition of *miRNA-1827* seemed not to make a meaningful difference, which might be reasonably explained by its relatively low content in OV cells (**Supplementary Figure S1B**).

Functional Enrichment Results of the Target Genes and Their Core Modules

There were 334 overlapped target genes selected as interactive genes with *miRNA-1827,* which were collected from 4,166 genes from the miRWALK database based on 12 servers and 2,611 overexpressed genes from the GEPIA database (**Figure 5A**). After the intersection, these 334 target genes were processed by bioinformatic analysis, and the results were interpreted by GO functional annotation and KEGG analysis in the DAVID database. For GO analysis, it was performed by three modules, including biological process, cellular component, and molecular function. For the biological process part, the extracellular matrix organization, cell–cell adhesion, regulation of cell proliferation and migration, actin skeleton organization, and mitochondrial membrane permeability were the main processes (**Figure 5B**, left panel). As for the cellular component, the cell–cell adherent junctions, extracellular exosome, transport vesicle, actin cytoskeleton, and lysosome were the primary components (**Figure 5B**, middle panel). Regarding the molecular function, the cadherin binding involved in cell–cell adhesion, myosin V binding, protein kinase binding, and protein self-association were the top enriched functions (**Figure 5B**, right panel). Concerning KEGG pathways, cell lung cancer, hepatitis C, tight junction, and cell adhesion molecules were the major involved pathways (**Figure 5C**). PPI network filtered out four main modules according to their MCODE score (**Figure 6**). These four clusters consisted of 34 nodes or genes, as listed in **Table 2**. As seen, GO and KEGG enrichment of these 34 genes demonstrated that they were mainly involved in the activation of caspase activity by cytochrome c, cell cycle, and regulation of mitochondria membrane permeability for the biological process (**Figure 7A**). The chromosome, cytosol, and kinetochore were enriched for the cellular component (**Figure 7B**), and the protein heterodimerization activity, purine ribonucleotide binding, and ATP binding for molecular function were determined (**Figure 7C**). The pathways in cancers, tight junction, and cell adhesion molecules for KEGG were identified (**Figure 7D**). All the valuable information above gave us some tips and clues of the localization and molecular functions of these genes in OV, which could help us narrow the targets down to several specific hub genes.

Characteristics and Clinical Value of the Hub Genes

According to the clinical value retrieved in the GEPIA and HPA databases in OV, two hub genes, *SPTBN2* and *BCL2L1*, caught our attention. SPTBN2 and BCL2L1 were located in the cytosol or

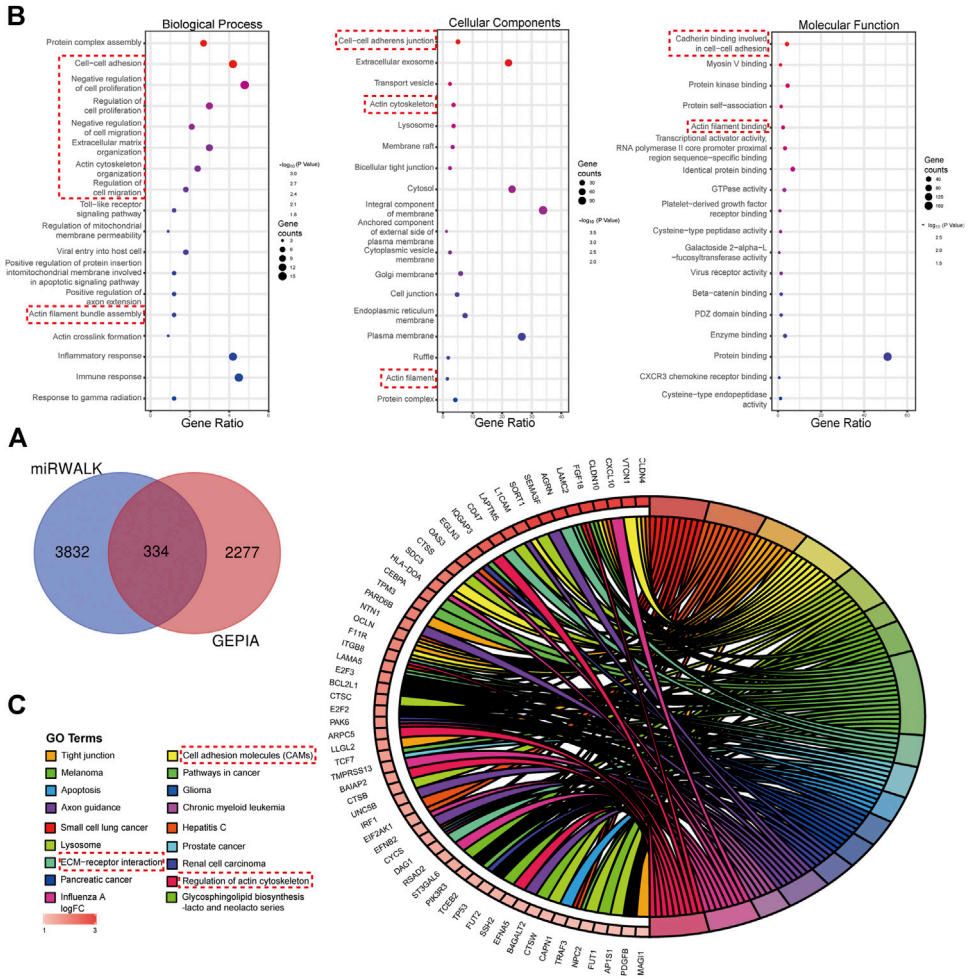


FIGURE 5 | Target genes interacted with *miRNA-1827* and the functional enrichment analysis for these genes. **(A)** The Venn diagram for the overlapped targets predicted by different databases, with 4,166 genes from the miRWALK tool based on 12 servers and 2,611 highly expressed genes from the GEPIA tool based on the TCGA and GTEx databases (adjusted $p < 0.05$). **(B)** GO enrichment analysis processed by the DAVID online tool including the biological process (left panel), cellular component (middle panel), and molecular function (right panel), as shown in the bubble plots. The size of the circles referred to the gene counts, and the X-axis pointed to the gene ratio. **(C)** KEGG pathway enrichment. The chordal diagram displayed the pathway terms enriched in the KEGG database. The input genes derived from the highly expressed predicted targets.

cell junctions (SPTBN2, **Figure 7E**, left) and mitochondria (BCL2L1, **Figure 7E**, right), respectively. Also, SPTBN2 and BCL2L1 were negatively expressed in normal ovarian tissues at the protein level, as depicted by the immunohistochemical results (**Figure 7F**). As we could see, *SPTBN2* and *BCL2L1* were overexpressed significantly in the OV group compared with the normal controls ($p < 0.05$), as shown in **Figure 8A**. The elimination of the inhibitory effect of miRNAs might account for the changes with the downregulation of *miRNA-1827*. To further verify the accuracy of these results, the gene expression profiles from the ICGC database were retrieved and processed. Consistent with the above findings, the contents of *SPTBN2* and *BCL2L1* were relatively overexpressed in these 111 cases of OV patients compared with the controls ($p < 0.01$) (**Supplementary Figures 2A,B**). In addition, there existed a moderate correlation between the expression of *SPTBN2* and *BCL2L1* with a significant difference ($p = 9.6E-59$, **Figure 8B**). Furthermore, OV patients in stage II, III, or IV shared similar gene expression patterns for *SPTBN2* (**Figure 8C**) or *BCL2L1* (**Figure 8D**) in cancer tissues. However, the upregulation of the two hub genes would lead to an elevated risk of death, as displayed in the survival curves (**Figure 8E**) (*SPTBN2*, upper panel, log-rank $p = 0.012$, HR = 1.4; *BCL2L1*, lower panel, log-rank $p = 0.034$, HR = 1.3). To further investigate the hub genes and their crucial roles in various types of cancers, we detected the gene and protein expression levels among different cancers. The images for qualitative analysis in this study were obtained from the HPA database. The results demonstrated that *SPTBN2* and *BCL2L1* were relatively overexpressed in most cancers compared with normal tissues, as delineated in **Figure 9A** at the gene level and **Figure 9B** at the protein level. These results indicated that

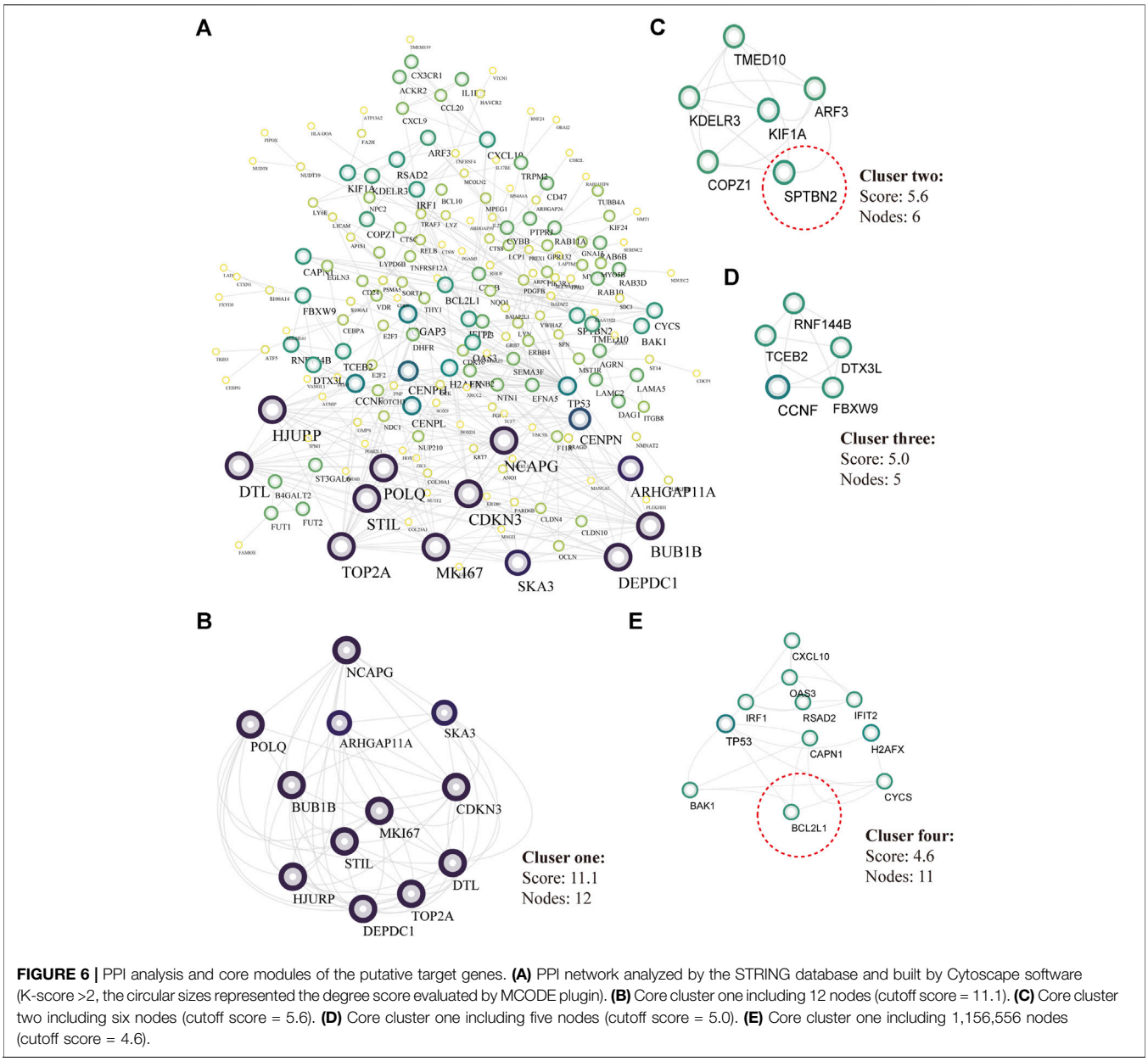


FIGURE 6 | PPI analysis and core modules of the putative target genes. **(A)** PPI network analyzed by the STRING database and built by Cytoscape software (K-score >2, the circular sizes represented the degree score evaluated by MCODE plugin). **(B)** Core cluster one including 12 nodes (cutoff score = 11.1). **(C)** Core cluster two including six nodes (cutoff score = 5.6). **(D)** Core cluster one including five nodes (cutoff score = 5.0). **(E)** Core cluster one including 1,156,556 nodes (cutoff score = 4.6).

the two hub genes might be suitable biomarkers not only in OV but also in other cancers, such as breast cancer, lung cancer, colon cancer, and prostate cancer.

The Potential Mechanisms of the Hub Genes Involved in Tumor Immunomodulatory Actions

To further decipher the function of the hub genes in the modulation of the tumor-infiltrating immune cells, the TIMER database was used to determine the changes of immunocytes in OV tissue and their interactive relationship with the hub genes. *SPTBN2* was negatively correlated, to a certain degree, with the infiltration of $CD4^+$ T cells (**Figure 10Ai**, partial.cor = −0.119, p = 9.10E-03) as well as dendritic cells (**Figure 10Aii**, partial.cor = −0.118, p = 9.53E-03). Similarly, *BCL2L1* was also negatively correlated, up to a point, with the infiltration of dendritic cells (**Figure 10Aiii**, partial.cor = −0.158, p = 5.17E-04). All those results suggested that the two hub genes might play a crucial role in adjusting $CD4^+$ T cells and dendritic cells in the OV microenvironment. The higher the expression of the hub genes, the lower the infiltration of the immunocytes. Lower immunocyte infiltration of $CD4^+$ T cells and dendritic cells meant poor prognosis as presented in **Figure 10B**. At the same time, we also explored the survival probability with the differential expression of *SPTBN2* and *BCL2L1* with the alteration of TIC abundance ($CD4^+$ T cells and dendritic cells), as shown in **Supplementary Figure S3**. Our results demonstrated that the

TABLE 2 | 34 target genes from the four clusters.

Names	Clusters	MCODE score	Names	Clusters	MCODE score	Names	Clusters	MCODE score
NCAPG	One	11.1	*TMED10*	Two	5.6	*CXCL10*	Four	4.6
POLQ	One	11.1	*KDELR3*	Two	5.6	*OAS3*	Four	4.6
ARHGAP11A	One	11.1	*KIF1A*	Two	5.6	*IRF1*	Four	4.6
SKA3	One	11.1	*ARF3*	Two	5.6	*RSAD2*	Four	4.6
BUB1B	One	11.1	*COPZ1*	Two	5.6	*IFIT2*	Four	4.6
MKI67	One	11.1	*SPTBN2*	Two	5.6	*TP5*	Four	4.6
CDKN3	One	11.1	*RNF144B*	Three	5.0	*CAPN1*	Four	4.6
STIL	One	11.1	*TCEB2*	Three	5.0	*H2AFX*	Four	4.6
DTL	One	11.1	*DTX3L*	Three	5.0	*BAK1*	Four	4.6
HJURP	One	11.1	*CCNF*	Three	5.0	*BCL2L1*	Four	4.6
TOP2A	One	11.1	*FBXW9*	Three	5.0	*CYCS*	Four	4.6
DEPDC1	One	11.1						

downregulation of *SPTBN2* in ovarian cancer cases would reduce the risk of death when they were enriched in $CD4^+$ T cells ($p < 0.05$) or dendritic cells ($p = 0.067$). These results were in line with the above findings. However, no significant difference was observed in patients with the differential expression of *BCL2L1* with the altered infiltration of dendritic cells. For further research on the chemokines related to the immunocytes and the two hub genes, the expression and correlation of *CCL27* and *CCR10* were explored. The results demonstrated that the overexpression of *SPTBN2* had a negative correlation with the expression of *CCL27* (**Figure 10Ci**, $p = 6.4E\text{-}15$, $R = -0.33$) and *CCR10* (**Figure 10Cii**, $p = 1.9E\text{-}11$, $R = -0.29$). Likewise, *BCL2L1* was negatively correlated with the expression of *CCL27* (**Figure 10Ciii**, $p = 4.6E\text{-}11$, $R = -0.29$) and *CCR10* (**Figure 10Civ**, $p = 5.3E\text{-}09$, $R = -0.25$). The expression profile of these molecules, *CCL27* and *CCR10*, indicated downregulated trends in OV tissue, as shown in **Figures 10Di,ii**, for each.

DISCUSSION

With the advent of high-throughput sequencing and microarray technologies, a striking upsurge in identifying novel biomarkers and the construction of molecular networks in various diseases is in the making based on the advances of bioinformatic analyses (Udhaya Kumar et al., 2020b; Mishra et al., 2021). At the same time, a growing body of evidence from gene expression profiles and microarrays has confirmed the significance of miRNAs in the tumorigenesis and progression of various types of cancers (Rupaimoole and Slack, 2017). The comparatively deep explorations on their roles have promoted the identification of novel biomarkers and application of these potential intervention targets toward the era of miRNA therapeutics (Hayes et al., 2014).

Previous studies have revealed that *miRNA-1827* functions as a tumor suppressor in various types of cancers since it is involved in tumor proliferation, invasion, migration, and angiogenesis (Ho et al., 2018). However, to date, no tangible proof has ever uncovered the relationship between *miRNA-1827* and its clinical significance in OV. Here, we performed systemic bioinformatic analyses to determine its pivotal role in OV and its intrinsic targets. The highlights of this study are the inclusion of a larger sample size by pooling similar studies and the application of bioinformatics and computational biology for information mining to determine the expression level of *miRNA-1827*. As reported, *miRNA-1827* is downregulated in most cancer samples, such as human lung adenocarcinoma cells and lymphoblastic leukemia clinical cases or cell lines, and the decreased level is correlated with tumor grade and lymph node metastasis (Guo et al., 2020; Naderi et al., 2020). In line with these findings, our research has also observed the reduced expression of *miRNA-1827* in all OV cases. The significantly differential expression was further confirmed by the comprehensive meta-analysis and qRT-PCR detection of seven types of OV cell lines or normal ovarian epithelial cells. Furthermore, a significant difference was noticed in the proliferation of OV cells after transfection of the *miRNA-1827* mimic. As previously described, *miRNA-1827* has been proved to suppress the development of lung adenocarcinoma by targeting oncogenic genes *MYC* and *FAM83F* (Fan et al., 2020). Another study also confirmed that a single-nucleotide polymorphism within the binding site of *miRNA-1827* on the 3' UTR of *MYCL1* could repress the expression level of the latter, triggering the progression of small-cell lung cancer (Xiong et al., 2011). These results signified that the overexpression of *miRNA-1827* could also act as a potential intervention pattern of OV. Subsequently, 334 candidate interactive genes targeted by *miRNA-1827* were screened out based on the miRWALK and GEPIA databases; GO and KEGG enrichment analyses were carried out to determine the characteristics and functions of these genes. According to the results, the most meaningful finding that emerged from the analyses was that these targets of miRNAs were involved in the regulation of cell proliferation and migration, extracellular matrix organization, cell–cell adhesion, actin skeleton organization, and actin filament bundle assembly. Intriguingly, *miRNA-1827* has been known to engage in biological activities chiefly, including the modulation of cellular proliferation, invasion, and metastasis in reproductive or nonproductive tumors (Wang et al., 2020; Shen et al., 2021). It is worth noting that altered cytoskeleton-associated proteins have been proven to produce an effect on the clinical outcomes of OV patients during malignant progression (Schiewek et al., 2018). In addition to this, cytoskeletal dynamics, such as those of the actin filament, play a pivotal role in enhanced cell motility to promote OV metastasis through

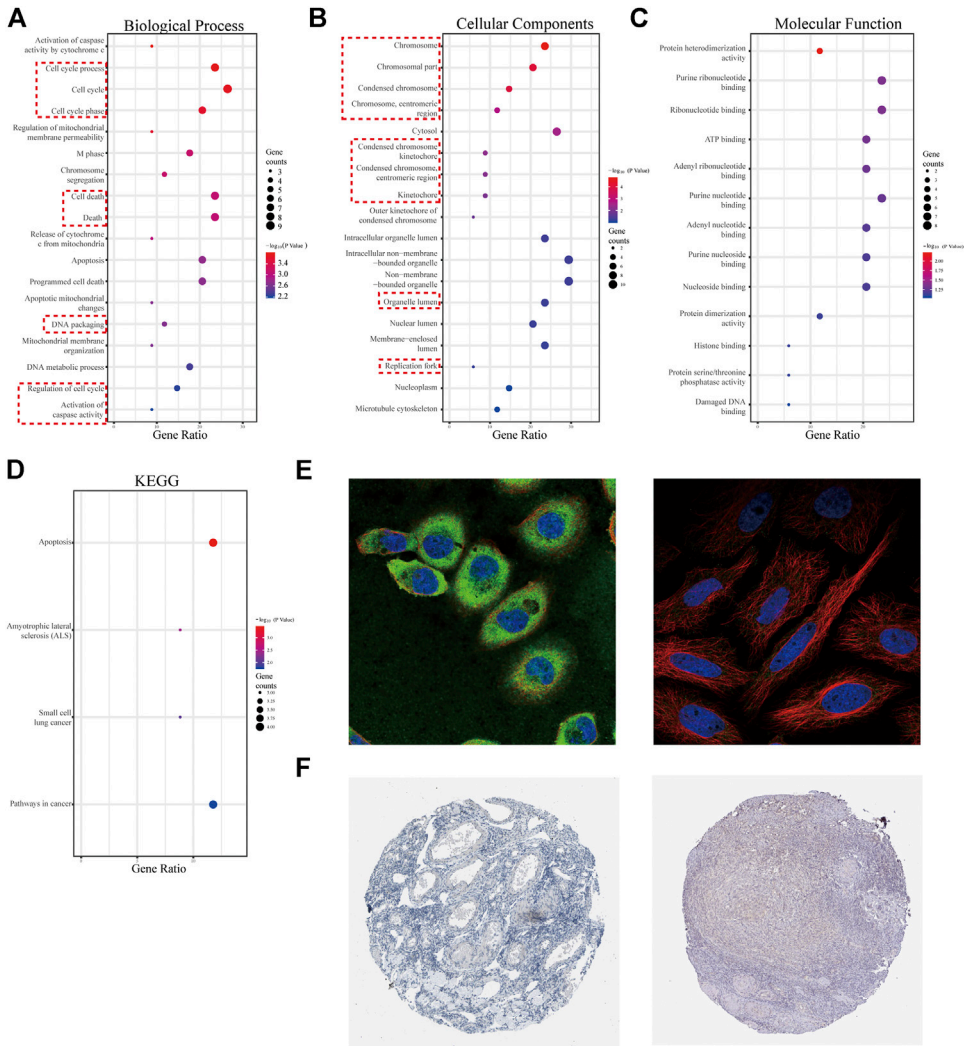


FIGURE 7 | GO analysis for 34 genes from four selected modules and characteristics of the promising hub genes, *SPTBN2* and *BCL2L1*. **(A)** Biological process for the 34 genes. **(B)** Cellular components for the 34 genes. **(C)** Molecular function for the 34 genes. **(D)** KEGG pathways enriched for the 34 genes. **(E)** Cellular localization from the cell atlas of the HPA database, with SPTNB2 localizing to the cytosol and cell junctions, stained in A-431 and MCF-7 cell lines **(left)**, and BCL2L1 localizing to mitochondria, stained in MCF-7 cell line **(right)**. **(F)** Decreased expression of SPTBN2 **(left)** and BCL2L1 **(right)** in normal ovarian tissue.

epithelial–mesenchymal transition (Lee et al., 2013). To elucidate the hub interactions among PPI networks, the MCODE plugin in Cytoscape software was applied, and finally, 34 node genes were identified. These genes were the main components of the chromosome (centromeric region), kinetochore, organelle lumen, and replication fork. As expected, these genes exerted considerable influence on the cell cycle process, chromosome segregation, DNA packaging, activation of caspase activity, and cell death. Based on this, two hub genes, *SPTBN2* and *BCL2L1*, were determined as target genes of *miRNA-1827* for further research according to the clinical value retrieved in the GEPIA and HPA databases in OV. *SPTBN2* is a protein-encoding gene, and its protein product belongs to the spectrin family. *SPTBN2* was previously reported in the development of cognition and motion or neurologic disorders (Forman et al., 2012; Lise et al., 2012). Additionally, it was recently identified as a signature gene, emerging in seven common cancers in the pathogenesis of cancers and pattern recognition (Wen et al., 2018). In the present study, *SPTBN2* was confirmed for the first time to be upregulated in OV at gene and protein levels. Most studies mainly focused on the role of *SPTBN2* in neurogenesis and degenerative diseases as a structural carrier for the stabilization and activation of membrane channels, receptors, and transporters (Yildiz Bolukbasi et al., 2017; Louis and Faust, 2020). In terms of cancer research, it was identified to be associated with the tumor progression and survival of colorectal cancer patients with m6A modifications (Zhang et al., 2021). Besides, it has also been reported to be a member of a novel, competing, endogenous RNA network for the prognosis of bladder cancer (Zhang et al., 2021). As for *BCL2L1*, it is an anti-apoptotic regulator of the *BCL-2* family and causes momentous effects on the integrity of the mitochondria membrane, autophagy, and cell survival

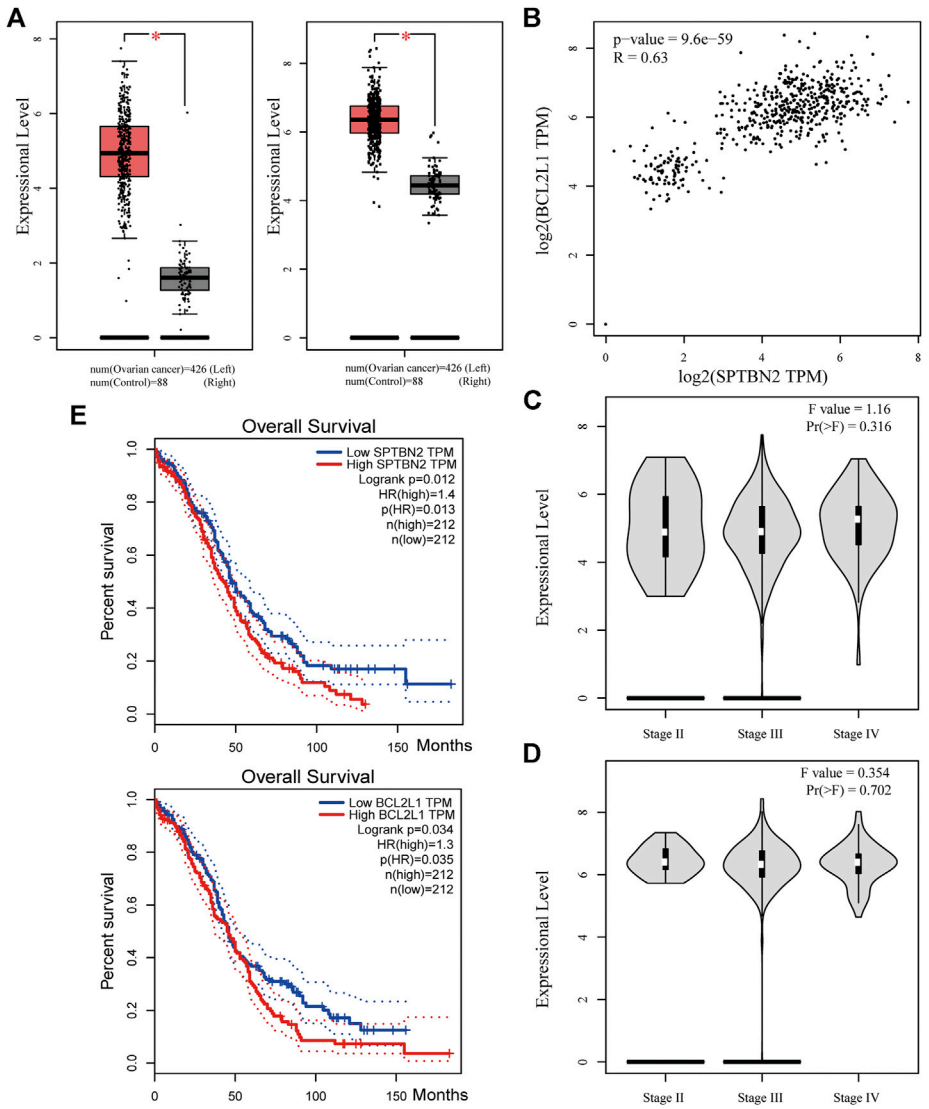


FIGURE 8 | Clinical value of the two hub genes. **(A)** Expressional levels between OV and NC groups for *SPTBN2* **(left)** and *BCL2L1* **(right)** (n = 426 for OV or 88 for NC, $p < 0.05$). **(B)** The correlation analysis for *SPTBN2* and *BCL2L1* (p = 9.6E-59, R = 0.63). **(C)** The overall survival for patients with high or low *SPTBN2* (upper panel, log-rank p = 0.012, HR = 1.4) and with high or low *BCL2L1* (upper panel, log-rank p = 0.034, HR = 1.3). **(D)** Expression of *SPTBN2* with different stages in OV (F = 1.16, Pr(>F) = 0.316). **(E)** Expression of *SPTBN2* with different stages in OV (F = 0.354, Pr(>F) = 0.702).

(Duan et al., 2019). *BCL2L1* has made significant differences in tumor targeting therapy, as a novel tumor promoter, for hepatocellular carcinoma, lung cancer, gastric cancer, and cervical cancer (Park et al., 2015; Cai et al., 2017). Here, we found that *BCL2L1* was also overexpressed in OV specimens. As a pro-oncogene, the aberrant expression of *BCL2L1* is related to a shortened disease-free survival, which accrues in patients with recurrence following chemotherapeutic interventions in OV (Williams et al., 2005). Preclinical research has also demonstrated that the ectopic expression of *BCL2L1* confers resistance to various pharmaceutical products, such as cisplatin, vincristine, and gemcitabine (Schniewind et al., 2004). Posttranslational modification by deamidation of *BCL2L1* has also been proven to be implicated in drug resistance to DNA-damaging medicaments (Deverman et al., 2002), and the pro-survival action of the oncogenic tyrosine kinase LCK following DNA damage may be mediated partly by restraining deamidation of *BCL2L1* as well (Zhao et al., 2004). All these findings have therefore specified the promising therapy of targeting *BCL2L1* in OV management. Besides, our investigation showed that these two hub genes were co-expressed and correlated with each other. The overexpression of these molecules would increase the risk of death or give rise to a poorer prognosis. During the progression of OV, they were persistently highly expressed across stages II, III, and IV. It should be mentioned that *SPTBN2* and *BCL2L1* were upregulated together not only in OV but also in various types of cancers, which were validated by various types of tumors and normal samples at both the gene and protein levels. Overall, these

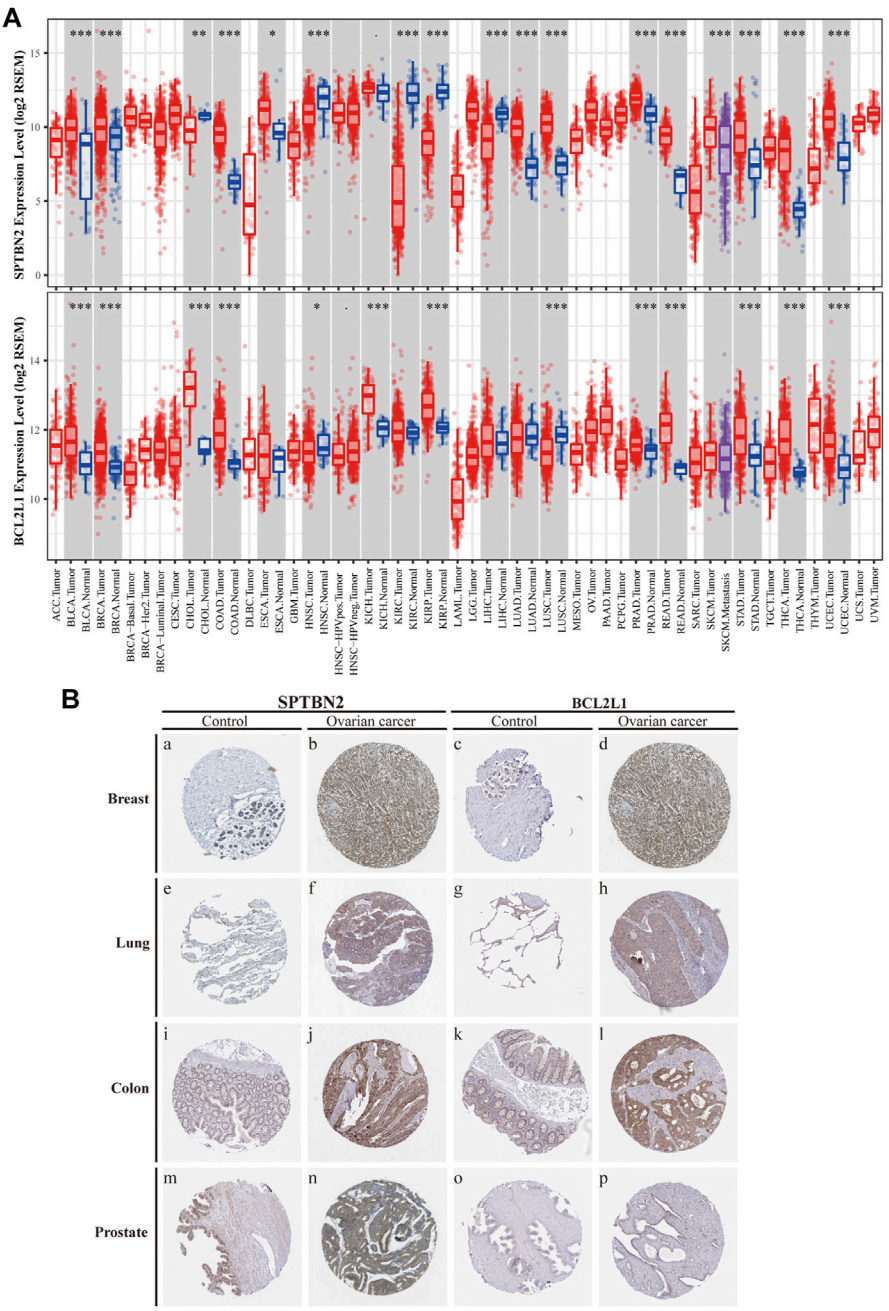


FIGURE 9 | Differential expression of the two hub genes in various types of tumors compared with NC. **(A)** Distributions of gene expression levels are displayed using box plots from the TIMER database by the Wilcoxon test for *SPTBN2* **(upper panel)** and *BCL2L1* **(lower panel)**. **(B)** Differential expression at protein levels between cancers and normal tissues, verified by the protein atlas in the HPA database including the breast, lung, colon, and prostate.

discoveries suggested the prognostic significance of *SPTBN2* and *BCL2L1* as therapeutic biomarkers targeted by *miRNA-1827* in OV.

In the following exploration, we have focused on the roles of *SPTBN2* and *BCL2L1* in the tumor microenvironment from the perspective of immune modulation by retrieving data from the TIMER database. Existing studies demonstrate that the interaction between tumor cells and host cells is indispensable for the oncogenesis and progression of tumors (Grivennikov et al., 2010). Within this environment and upon tumor-driven stimuli, tumors can form a tumor-permissive soil with reprogrammed host cells that exhibit tumor-supporting phenotypes (Binnewies et al., 2018). Intriguingly, this study revealed that *SPTBN2* and *BCL2L1* could generate adverse effects on the decreased infiltration of $CD4^+$ T cells and dendritic cells in OV, which seemed likely to bring about a more unsatisfactory prognostic outcome. The reduced abundance of $CD4^+$ T cells and dendritic cells in this study was

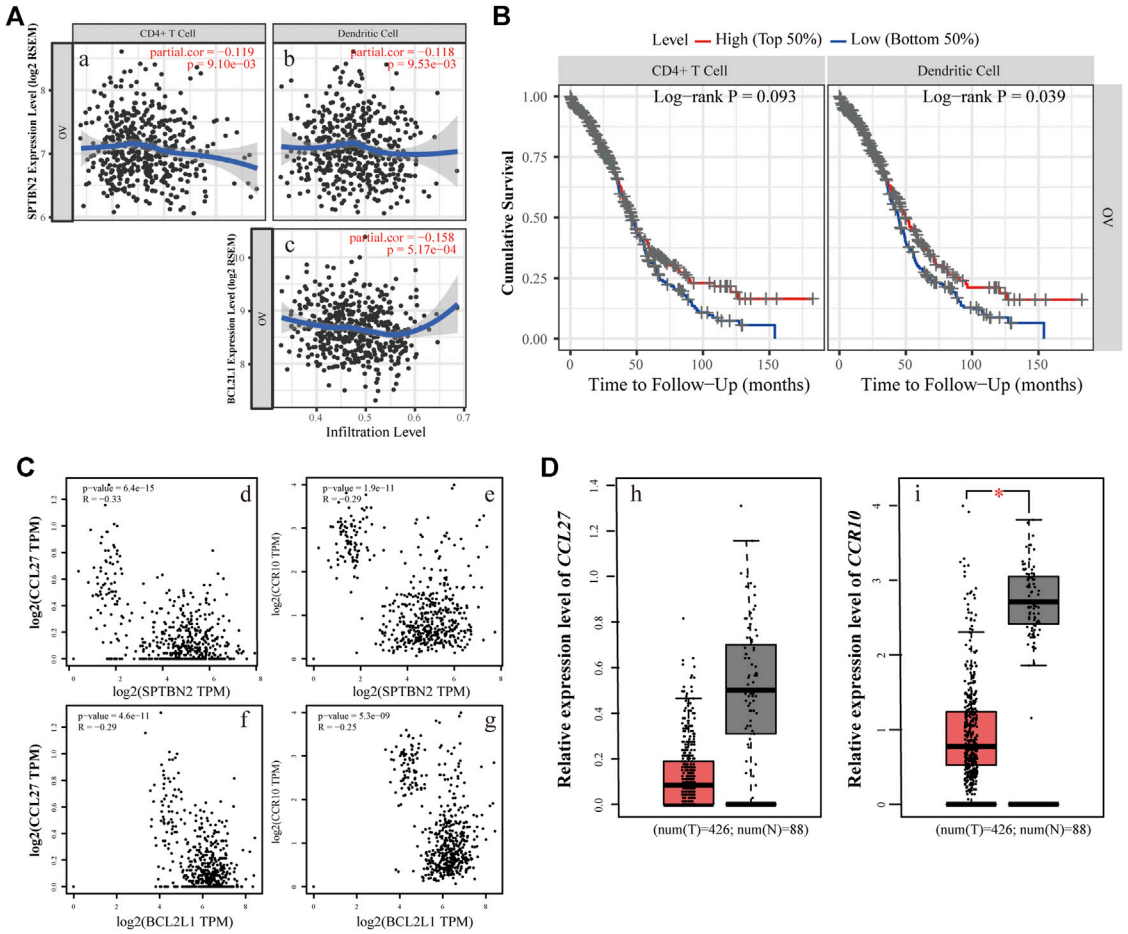


FIGURE 10 | Immunoregulatory relationship of the hub genes. **(A)** The relevance of the expression level and immunocyte infiltration, where *SPTBN2* modulated CD4$^+$ T cells **(i)** or dendritic cells **(ii)** and *BCL2L1* regulated dendritic cells **(iii)**. **(B)** Cumulative survival difference among OV patients with high or low CD4$^+$ T cells **(left panel)** and high or low dendritic cell infiltration **(right panel)**. **(C)** The correlation between chemokine family or receptors and the hub genes. The expression levels of *CCL27* and *CCR10* negatively correlated with *SPTBN2* **(i,ii)** and *BCL2L1* **(iii,iv)**, respectively. **(D)** Differential gene expression of *CCL27* **(i)** and *CCR10* **(ii)** with 426 OV samples vs. 88 NC samples.

confirmed to be associated with declined cumulative survival rate. In order to seek the reasons for the reduction of CD4$^+$ T cells and dendritic cells mediated by *SPTBN2* or *BCL2L1*, the related chemokine *CCL27* and the receptor *CCR10* were explored. In terms of the traits of the two molecules, *CCL27*, as a member of the chemokine ligand family, would participate in the homing of immunocytes by specifically binding to *CCR10* (Fujita et al., 2006). *CCL27* was once regarded as a sensitive serum biomarker to distinguish nasopharyngeal carcinoma patients from healthy donors (highly expressed in abnormal populations due to the activated immune system) (Mao et al., 2018). As reported previously, *CCL27* could inhibit tumor growth to a certain extent through recruiting natural killer cells or T cells locally (Gao et al., 2003) and could intensify the resistance to tumor formation as well (Okada et al., 2004). In addition, *CCL27* could enter into a chemotaxis role for the recruitment and activation of CD4$^+$ T cells. Our study found that these molecules were also negatively influenced by *SPTBN2* or *BCL2L1*. In accordance with our studies, people once reported (Khaiboullina et al., 2015) that the reduced chemokine of *CCL27* could help cancers to evade immunological surveillance, thereby resulting in immune escape (Pivarcsi et al., 2007). The results above delineate the alteration of *SPTBN2* and *BCL2L1* in view of tumor immune modulation mediated by *CCL27* and *CCR10*, resulting in a reduced abundance of CD4$^+$ T cells and dendritic cells.

The limitations of the current study need to be pointed out: the mutual interactions between *miRNA-1827* and its two hub genes should be further validated to ensure the target-binding action. Besides, the effects of *miRNA-1827* in OV should be verified in animal models to promote its clinical application.

CONCLUSION

In conclusion, the present study takes advantage of retrieved data from microarrays or gene expression profiles and integrative bioinformatic analyses. Therefore, we have identified that *miRNA-1827* as a novel biomarker is associated with the proliferation and prognosis of OV by targeting *SPTBN2* and *BCL2L1*. These hub genes are determined to be

involved in the tumor immunomodulatory process mediated by *CCL27* and *CCR10*. On the whole, *miRNA-1827* from computational biology analysis exerts promising effects as an intervention target in OV management in future applications.

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: GSE119056, GSE83693, GSE53829, GSE47841, GSE48485, and GSE58517.

AUTHOR CONTRIBUTIONS

PHF, ZTG, ZXG, and LL: collection and assembly of data, data analysis, and manuscript writing. QY: conception and design, administrative support, and final approval of the manuscript.

FUNDING

This study was funded by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant number 2020-PT320-003).

ACKNOWLEDGMENTS

We acknowledge GEO, GTEx, and TCGA databases for providing their platforms and contributors for uploading their meaningful datasets. The authors would like to thank Editage (www.editage.cn) for English language polishing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.687576/full#supplementary-material

REFERENCES

Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., et al. (2018). Global Surveillance of Trends in Cancer Survival 2000-14 (CONCORD-3): Analysis of Individual Records for 37 513 025 Patients Diagnosed with One of 18 Cancers From 322 Population-Based Registries in 71 Countries. *Lancet* 391 (10125), 1023–1075. doi:10.1016/S0140-6736(17)33326-3

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

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

Bueno, M. J., de Castro, I. P., and Malumbres, M. (2008). Control of Cell Proliferation Pathways by microRNAs. *Cell Cycle* 7 (20), 3143–3148. doi:10.4161/cc.7.20.6833

Busk, P. K. (2014). A Tool for Design of Primers for microRNA-Specific Quantitative RT-qPCR. *BMC Bioinformatics* 15, 29. doi:10.1186/1471-2105-15-29

Cai, J., Liu, T., Jiang, X., Guo, C., Liu, A., and Xiao, X. (2017). Downregulation of USP18 Inhibits Growth and Induces Apoptosis in Hepatitis B Virus-Related Hepatocellular Carcinoma Cells by Suppressing BCL2L1. *Exp. Cell Res.* 358 (2), 315–322. doi:10.1016/j.yexcr.2017.07.006

Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4 (5), P3. doi:10.1186/gb-2003-4-9-r60

Deverman, B. E., Cook, B. L., Manson, S. R., Niederhoff, R. A., Langer, E. M., Rosová, I., et al. (2002). Bcl-xL Deamidation is a Critical Switch in the Regulation of the Response to DNA Damage. *Cell* 111 (1), 51–62. doi:10.1016/s0092-8674(02)00972-8

Duan, S., Yu, S., Yuan, T., Yao, S., and Zhang, L. (2019). Exogenous Let-7a-5p Induces A549 Lung Cancer Cell Death through BCL2L1-Mediated PI3Kγ Signaling Pathway. *Front. Oncol.* 9, 808. doi:10.3389/fonc.2019.00808

Fan, G., Xu, P., and Tu, P. (2020). MiR-1827 Functions as a Tumor Suppressor in Lung Adenocarcinoma by Targeting MYC and FAM83F. *J. Cell Biochem.* 121 (2), 1675–1689. doi:10.1002/jcb.29402

Fei, X., Qi, M., Wu, B., Song, Y., Wang, Y., and Li, T. (2012). MicroRNA-195-5p Suppresses Glucose Uptake and Proliferation of Human Bladder Cancer T24 Cells by Regulating GLUT3 Expression. *FEBS Lett.* 586 (4), 392–397. doi:10.1016/j.febslet.2012.01.006

Forman, O. P., De Risio, L., Stewart, J., Mellersh, C. S., and Beltran, E. (2012). Genome-Wide mRNA Sequencing of a Single Canine Cerebellar Cortical Degeneration Case Leads to the Identification of a Disease Associated SPTBN2 Mutation. *BMC Genet.* 13, 55. doi:10.1186/1471-2156-13-55

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs From Deep Sequencing Data Using miRDeep. *Nat. Biotechnol.* 26 (4), 407–415. doi:10.1038/nbt1394

Fu, D., Zhang, B., Yang, L., Huang, S., and Xin, W. (2020). Development of an Immune-Related Risk Signature for Predicting Prognosis in Lung Squamous Cell Carcinoma. *Front. Genet.* 11, 978. doi:10.3389/fgene.2020.00978

Fujita, Y., Abe, R., Sasaki, M., Honda, A., Furuichi, M., Asano, Y., et al. (2006). Presence of Circulating CCR10+ T Cells and Elevated Serum CTACK/CCL27 in the Early Stage of Mycosis Fungoides. *Clin. Cancer Res.* 12 (9), 2670–2675. doi:10.1158/1078-0432.CCR-05-1513

Gao, J. Q., Tsuda, Y., Katayama, K., Nakayama, T., Hatanaka, Y., Tani, Y., et al. (2003). Antitumor Effect by Interleukin-11 Receptor Alpha-Locus chemokine/CCL27, Introduced Into Tumor Cells through a Recombinant Adenovirus Vector. *Cancer Res.* 63 (15), 4420–4425.

Graves, P., and Zeng, Y. (2012). Biogenesis of Mammalian microRNAs: A Global View. *Genomics Proteomics Bioinformatics* 10 (5), 239–245. doi:10.1016/j.gpb.2012.06.004

Grivennikov, S. I., Greten, F. R., and Karin, M. (2010). Immunity, Inflammation, and Cancer. *Cell* 140 (6), 883–899. doi:10.1016/j.cell.2010.01.025

Guo, X., Wang, Z., Sun, Q., Sun, C., Hua, H., and Huang, Q. (2020). The Inhibitory Effect of microRNA-1827 on Anoikis Resistance in Lung Adenocarcinoma A549 Cells via Targeting Caveolin-1. *Acta Biochim. Biophys. Sin (Shanghai)* 52 (10), 1148–1155. doi:10.1093/abbs/gmaa102

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

Hayes, J., Peruzzi, P. P., and Lawler, S. (2014). MicroRNAs in Cancer: Biomarkers, Functions and Therapy. *Trends Mol. Med.* 20 (8), 460–469. doi:10.1016/j.molmed.2014.06.005

Ho, C. S., Noor, S. M., and Nagoor, N. H. (2018). MiR-378 and MiR-1827 Regulate Tumor Invasion, Migration and Angiogenesis in Human Lung Adenocarcinoma by Targeting RBX1 and CRKL, Respectively. *J. Cancer* 9 (2), 331–345. doi:10.7150/jca.18188

Ho, C. S., Yap, S. H., Phuah, N. H., In, L. L. A., and Hasima, N. (2014). MicroRNAs Associated With Tumour Migration, Invasion and Angiogenic Properties in A549 and SK-Lu1 Human Lung Adenocarcinoma Cells. *Lung Cancer* 83 (2), 154–162. doi:10.1016/j.lungcan.2013.11.024

Khaiboullina, S. F., Gumerova, A. R., Khafizova, I. F., Martynova, E. V., Lombardi, V. C., Bellusci, S., et al. (2015). CCL27: Novel Cytokine With Potential Role in

Pathogenesis of Multiple Sclerosis. *Biomed. Res. Int.* 2015, 1–10. doi:10.1155/2015/189638

Kim, V. N. (2005). MicroRNA Biogenesis: Coordinated Cropping and Dicing. *Nat. Rev. Mol. Cell Biol.* 6 (5), 376–385. doi:10.1038/nrm1644

Kumar, S. U., Kumar, D. T., Siva, R., Doss, C. G. P., and Zayed, H. (2019). Integrative Bioinformatics Approaches to Map Potential Novel Genes and Pathways Involved in Ovarian Cancer. *Front. Bioeng. Biotechnol.* 7, 391. doi:10.3389/fbioe.2019.00391

Lee, S., Yang, Y., Fishman, D., Banaszak Holl, M. M., and Hong, S. (2013). Epithelial-Mesenchymal Transition Enhances Nanoscale Actin Filament Dynamics of Ovarian Cancer Cells. *J. Phys. Chem. B.* 117 (31), 9233–9240. doi:10.1021/jp4055186

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, Y., and Li, L. (2019). Prognostic Values and Prospective Pathway Signaling of MicroRNA-182 in Ovarian Cancer: A Study Based on Gene Expression Omnibus (GEO) and Bioinformatics Analysis. *J. Ovarian Res.* 12 (1), 106. doi:10.1186/s13048-019-0580-7

Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Parolin Schnekenberg, R., et al. (2012). Recessive Mutations in SPTBN2 Implicate β-III Spectrin in Both Cognitive and Motor Development. *Plos Genet.* 8 (12), e1003074. doi:10.1371/journal.pgen.1003074

Louis, E. D., and Faust, P. L. (2020). Essential Tremor Within the Broader Context of Other Forms of Cerebellar Degeneration. *Cerebellum* 19 (6), 879–896. doi:10.1007/s12311-020-01160-4

Mao, M.-J., Xue, N., Wang, X.-P., Chi, P.-D., Liu, Y.-J., Huang, Q., et al. (2018). Chemokine CCL27 is a Novel Plasma Biomarker for Identification the Nasopharyngeal Carcinoma Patients from the Epstein-Barr Virus Capsid Antigen-Specific IgA Seropositive Population. *BMC Cancer* 18 (1), 9. doi:10.1186/s12885-017-3718-2

Mishra, S., Shah, M. I., Udhaya Kumar, S., Thirumal Kumar, D., Gopalakrishnan, C., Al-Subaie, A. M., et al. (2021). Network Analysis of Transcriptomics Data for the Prediction and Prioritization of Membrane-Associated Biomarkers for Idiopathic Pulmonary Fibrosis (IPF) by Bioinformatics Approach. *Adv. Protein Chem. Struct. Biol.* 123, 241–273. doi:10.1016/bs.apcsb.2020.10.003

Naderi, T., Mohammadi-Yeganeh, S., Mohammadi-Hezaveh, N., Hadavi, R., Gharehbaghian, A., Vazifeh-Shiran, N., et al. (2020). Investigating the Inhibitory Effect of miR-34a, miR-449a, miR-1827, and miR-106b on Target Genes Including NOTCH1, C-Myc, and CCND1 in Human T Cell Acute Lymphoblastic Leukemia Clinical Samples and Cell Line. *Iran J. Basic Med. Sci.* 23 (3), 376–382. doi:10.22038/IJBMS.2019.40695.9615

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

Nicoloso, M. S., Spizzo, R., Shimizu, M., Rossi, S., and Calin, G. A. (2009). MicroRNAs - The Micro Steering Wheel of Tumour Metastases. *Nat. Rev. Cancer* 9 (4), 293–302. doi:10.1038/nrc2619

Okada, N., Gao, J.-Q., Sasaki, A., Niwa, M., Okada, Y., Nakayama, T., et al. (2004). Anti-Tumor Activity of Chemokine is Affected by Both Kinds of Tumors and the Activation State of the Host's Immune System: Implications for Chemokine-Based Cancer Immunotherapy. *Biochem. Biophys. Res. Commun.* 317 (1), 68–76. doi:10.1016/j.bbrc.2004.03.013

Pal, M. K., Jaiswar, S. P., Dwivedi, V. N., Tripathi, A. K., Dwivedi, A., and Sankhwar, P. (2015). MicroRNA: A New and Promising Potential Biomarker for Diagnosis and Prognosis of Ovarian Cancer. *Cancer Biol. Med.* 12 (4), 328–341. doi:10.7497/j.issn.2095-3941.2015.0024

Park, H., Cho, S.-Y., Kim, H., Na, D., Han, J. Y., Chae, J., et al. (2015). Genomic Alterations in BCL2L1 and DLC1 Contribute to Drug Sensitivity in Gastric Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 112 (40), 12492–12497. doi:10.1073/pnas.1507491112

Pivarcsi, A., Muller, A., Hippe, A., Rieker, J., van Lierop, A., Steinhoff, M., et al. (2007). Tumor Immune Escape by the Loss of Homeostatic Chemokine Expression. *Proc. Natl. Acad. Sci.* 104 (48), 19055–19060. doi:10.1073/pnas.0705673104

Polytarchou, C., Oikonomopoulos, A., Mahurkar, S., Touroutoglou, A., Koukos, G., Hommes, D. W., et al. (2015). Assessment of Circulating MicroRNAs for the Diagnosis and Disease Activity Evaluation in Patients with Ulcerative Colitis by Using the Nanostring Technology. *Inflamm. Bowel Dis.* 21 (11), 2533–2539. doi:10.1097/MIB.0000000000000547

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 (3), 203–222. doi:10.1038/nrd.2016.246

Salani, R., Khanna, N., Frimer, M., Bristow, R. E., and Chen, L.-M. (2017). An Update on Post-Treatment Surveillance and Diagnosis of Recurrence in Women with Gynecologic Malignancies: Society of Gynecologic Oncology (SGO) Recommendations. *Gynecol. Oncol.* 146 (1), 3–10. doi:10.1016/j.ygyno.2017.03.022

Schiewek, J., Schumacher, U., Lange, T., Joosse, S. A., Wikman, H., Pantel, K., et al. (2018). Clinical Relevance of Cytoskeleton Associated Proteins for Ovarian Cancer. *J. Cancer Res. Clin. Oncol.* 144 (11), 2195–2205. doi:10.1007/s00432-018-2710-9

Schniewind, B., Christgen, M., Kurdow, R., Haye, S., Kremer, B., Kalthoff, H., et al. (2004). Resistance of Pancreatic Cancer to Gemcitabine Treatment is Dependent on Mitochondria-Mediated Apoptosis. *Int. J. Cancer* 109 (2), 182–188. doi:10.1002/ijc.11679

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

Shen, A., Tong, X., Li, H., Chu, L., Jin, X., Ma, H., et al. (2021). TPPP3 Inhibits the Proliferation, Invasion and Migration of Endometrial Carcinoma Targeted With miR-1827. *Clin. Exp. Pharmacol. Physiol.* 48, 890–901. doi:10.1111/1440-1681.13456

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

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

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

Udhaya Kumar, S., Rajan, B., Thirumal Kumar, D., Anu, P. V., Abunada, T., Younes, S., et al. (2020a). Involvement of Essential Signaling Cascades and Analysis of Gene Networks in Diabesity. *Genes (Basel)* 11 (11), 1256. doi:10.3390/genes11111256

Udhaya Kumar, S., Thirumal Kumar, D., Siva, R., George Priya Doss, C., Younes, S., Younes, N., et al. (2020b). Dysregulation of Signaling Pathways Due to Differentially Expressed Genes From the B-Cell Transcriptomes of Systemic Lupus Erythematosus Patients - A Bioinformatics Approach. *Front. Bioeng. Biotechnol.* 8, 276. doi:10.3389/fbioe.2020.00276

Uhlen, M., Fagerberg, L., Hallstrom, 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

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: A Database of Predicted Functional Associations Between Proteins. *Nucleic. Acids Res.* 31 (1), 258–261. doi:10.1093/nar/gkg034

Wang, L., and Li, X. (2020). Identification of an Energy Metabolism Related Gene Signature in Ovarian Cancer Prognosis. *Oncol. Rep.* 43 (6), 1755–1770. doi:10.3892/or.2020.7548

Wang, Y., Gao, R., Li, J., Tang, S., Li, S., Tong, Q., et al. (2020). Circular RNA Hsa_circ_0003141 Promotes Tumorigenesis of Hepatocellular Carcinoma via a miR-1827/UBAP2 Axis. *Aging* 12 (10), 9793–9806. doi:10.18632/aging.103244

Wei, J., Wang, F., Kong, L.-Y., Xu, S., Doucette, T., Ferguson, S. D., et al. (2013). miR-124 Inhibits STAT3 Signaling to Enhance T Cell-Mediated Immune Clearance of Glioma. *Cancer Res.* 73 (13), 3913–3926. doi:10.1158/0008-5472.CAN-12-4318

Wen, J.-X., Li, X.-Q., and Chang, Y. (2018). Signature Gene Identification of Cancer Occurrence and Pattern Recognition. *J. Comput. Biol.* 25 (8), 907–916. doi:10.1089/cmb.2017.0261

Williams, J., Lucas, P. C., Griffith, K. A., Choi, M., Fogoros, S., Hu, Y. Y., et al. (2005). Expression of Bcl-xL in Ovarian Carcinoma Is Associated With Chemoresistance and Recurrent Disease. *Gynecol. Oncol.* 96 (2), 287–295. doi:10.1016/j.ygyno.2004.10.026

Xiong, F., Wu, C., Chang, J., Yu, D., Xu, B., Yuan, P., et al. (2011). Genetic Variation in an miRNA-1827 Binding Site in MYCL1 Alters Susceptibility to Small-Cell

Lung Cancer. *Cancer Res.* 71 (15), 5175–5181. doi:10.1158/0008-5472.CAN-10-4407

Yang, Y., Qi, S., Shi, C., Han, X., Yu, J., Zhang, L., et al. (2020). Identification of Metastasis and Prognosis-Associated Genes for Serous Ovarian Cancer. *Biosci. Rep.* 40 (6), BSR20194324. doi:10.1042/BSR20194324

Yildiz Bolukbasi, E., Afzal, M., Mumtaz, S., Ahmad, N., Malik, S., and Tolun, A. (2017). Progressive SCAR14 With Unclear Speech, Developmental Delay, Tremor, and Behavioral Problems Caused by a Homozygous Deletion of the SPTBN2 Pleckstrin Homology Domain. *Am. J. Med. Genet. A.* 173 (9), 2494–2499. doi:10.1002/ajmg.a.38332

Zhang, C., Liu, J., Tan, C., Yue, X., Zhao, Y., Peng, J., et al. (2016). microRNA-1827 Represses MDM2 to Positively Regulate Tumor Suppressor P53 and Suppress Tumorigenesis. *Oncotarget* 7 (8), 8783–8796. doi:10.18632/oncotarget.7088

Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* 37 (4), 367–369. doi:10.1038/s41587-019-0055-9

Zhang, Z., Wang, Q., Zhang, M., Zhang, W., Zhao, L., Yang, C., et al. (2021). Comprehensive Analysis of the Transcriptome-Wide m6A Methylome in Colorectal Cancer by MeRIP Sequencing. *Epigenetics* 16 (4), 425–435. doi:10.1080/15592294.2020.1805684

Zhao, R., Yang, F. T., and Alexander, D. R. (2004). An Oncogenic Tyrosine Kinase Inhibits DNA Repair and DNA-Damage-Induced Bcl-xL Deamidation in T Cell Transformation. *Cancer Cell* 5 (1), 37–49. doi:10.1016/s1535-6108(03)00333-7

Zheng, M. J., Li, X., Hu, Y. X., Dong, H., Gou, R., Nie, X., et al. (2019). Identification of Molecular Marker Associated With Ovarian Cancer Prognosis Using Bioinformatics Analysis and Experiments. *J. Cell Physiol.* 234 (7), 11023–11036. doi:10.1002/jcp.27926

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.

Copyright © 2021 Feng, Ge, Guo, Lin and 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.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 21 June 2021
doi: 10.3389/fmolb.2021.685174

BRCA1 and *BRCA2* Variation in Taiwanese General Population and the Cancer Cohort

*Jiasheng Chian, Siddharth Sinha, Zixin Qin and San Ming Wang**

Cancer Centre and Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau, China

BRCA1 and *BRCA2* (*BRCA*) play essential roles in maintaining genome stability. Rapidly evolving human *BRCA* generates oncogenic variants causing high cancer risk. *BRCA* variation is ethnic-specific in reflecting adaptation and/or effects of genetic drift. Taiwanese population of 23.8 million is an admixture of multiple ethnic origins; Taiwan's subtropical and tropical climate and geographically islandic location provide a unique natural environment. Therefore, Taiwanese population provides a unique model to study human *BRCA* variation. Through collecting, standardizing, annotating, and classifying publicly available *BRCA* variants derived from Taiwanese general population and the cancer cohort, we identified 335 *BRCA* variants, of which 164 were from 1,517 non-cancer individuals, 126 from 2,665 cancer individuals, and 45 from both types of individuals. We compared the variant data with those from other ethnic populations such as mainland Chinese, Macau Chinese, Japanese, Korean, Indian, and non-Asians. We observed that the sharing rates with other Asian ethnic populations were correlated with its genetic relationship. Over 60% of the 335 Taiwanese *BRCA* variants were VUS, unclassified variants, or novel variants, reflecting the ethnic-specific features of Taiwanese *BRCA* variation. While it remains challenging to classify these variants, our structural and *in silico* analyses predicted their enrichment of *BRCA* deleterious variants. We further determined the 3.8% prevalence of *BRCA* pathogenic variants in the Taiwanese breast cancer cohort, and determined 0.53% prevalence of the *BRCA* pathogenic variants in Taiwanese general population, with the estimated 126,140 *BRCA* pathogenic variant carriers. We identified *BRCA2* c.5164_5165delAG at *BRCA2* BRC6 motif as a potential founder mutation in Taiwanese population. Our study on *BRCA* variation in Taiwanese and other East Asian populations demonstrates that ethnic specificity is a common phenomenon for *BRCA* variation in East Asian population; the data generated from the study provide a reference for clinical applications in *BRCA*-related cancer in Taiwanese population.

Keywords: BRCA1, BRCA2, prevalence, population, ethnicity, pathogenic

OPEN ACCESS

Edited by:
Xin Wang,
City University of Hong, Hong Kong

Reviewed by:
Kajal Biswas,
National Cancer Institute, United States
Angelo Minucci,
Agostino Gemelli University Hospital Foundation, Italy

****Correspondence:***
San Ming Wang
sanmingwang@um.edu.mo

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *24 March 2021*
Accepted: *28 May 2021*
Published: *21 June 2021*

Citation:
Chian J, Sinha S, Qin Z and Wang SM (2021) BRCA1 *and* BRCA2 *Variation in Taiwanese General Population and the Cancer Cohort. Front. Mol. Biosci. 8:685174. doi: 10.3389/fmolb.2021.685174*

INTRODUCTION

BRCA1 and *BRCA2* (hereafter refer as *BRCA*) play essential roles in maintaining genome stability by repairing double-strand DNA damage through homologous recombination (Roy et al., 2011). *BRCA* is under positive selection in the humans, leading to high variability (Lou et al., 2014). While the majority of variants can be beneficial or neutral, those occurred at specific positions can damage the

function of *BRCA*, causing genome instability and increased risk of breast cancer, ovarian cancer, and other types of cancer (Kuchenbaecker et al., 2017). As *BRCA* variation is mostly of the germline nature, the later life stage of cancer occurrence provides a unique opportunity to prevent *BRCA* variation–caused cancer by early identification of the pathogenic variant carriers before cancer development (Burke et al., 1997). Furthermore, PARP inhibitors provide effective treatment of *BRCA* variant–caused cancer through synthetic lethal therapy (Jerez et al., 2020).

BRCA variation is well determined as highly ethnic specific in certain ethnic populations, such as the *BRCA1* 185delAG, 5382insC, and *BRCA2* 6174delT in Ashkenazi Jews population (Levy-Lahad et al., 1997). Restricted by the lack of *BRCA* variation data from non-Caucasian populations (Bhaskaran et al., 2019; Friebel et al., 2019), however, it remains unclear whether ethnic specificity is mainly in certain specific ethnic population or is a universal phenomenon across worldwide ethnic populations. Recently, we analyzed *BRCA* variation in Asian populations such as Indian, Chinese, Korean, and Japanese, and revealed that ethnic-specific *BRCA* variation is also widely present in these Asian populations (Bhaskaran et al., 2020). With a population size nearly 24 million, Taiwanese population consists of admixed ethnic origins across prehistory and current days. Although Taiwanese population included largely the ancestors from southern Han Chinese of Fujian and Guangdong regions of mainland China, it also included other ethnicities including the native Austronesians who also distributed to Pacific islands and Asian neighbors. Furthermore, the islandic location with subtropical and tropical climates in Taiwan Island provides a unique natural environment for Taiwanese population (Chen et al., 2016; **Figure 1**). Therefore, the Taiwanese population provides a unique model to study *BRCA* evolution and its impact on human health.

In the current study, we performed a systematic analysis for *BRCA* variation in the Taiwanese general population and the cancer cohort. Of the *BRCA* variants identified, we observed that forty percent *BRCA* variants were Taiwanese specific; using the identified *BRCA* pathogenic variants as the reference, we determined the prevalence of *BRCA* pathogenic variation in Taiwanese general population and the cancer cohort. Data from our study provide further evidence to demonstrate that ethnic specificity of *BRCA* variation is a common phenomenon in East Asian populations.

RESULTS

Data Collection

We collected a total of 335 *BRCA* variants derived from Taiwanese population, including 164 from general population, 126 from the Taiwanese cancer patient cohort, and 45 (19 in *BRCA1* and 26 in *BRCA2*) from both groups. For the variants from cancer patients, nearly all were from breast cancer and ovarian cancer (**Supplementary Table S1**). We performed standardization, annotation, and clinical classification for all *BRCA* variants (**Table 1**, **Supplementary Tables S2, S3**).

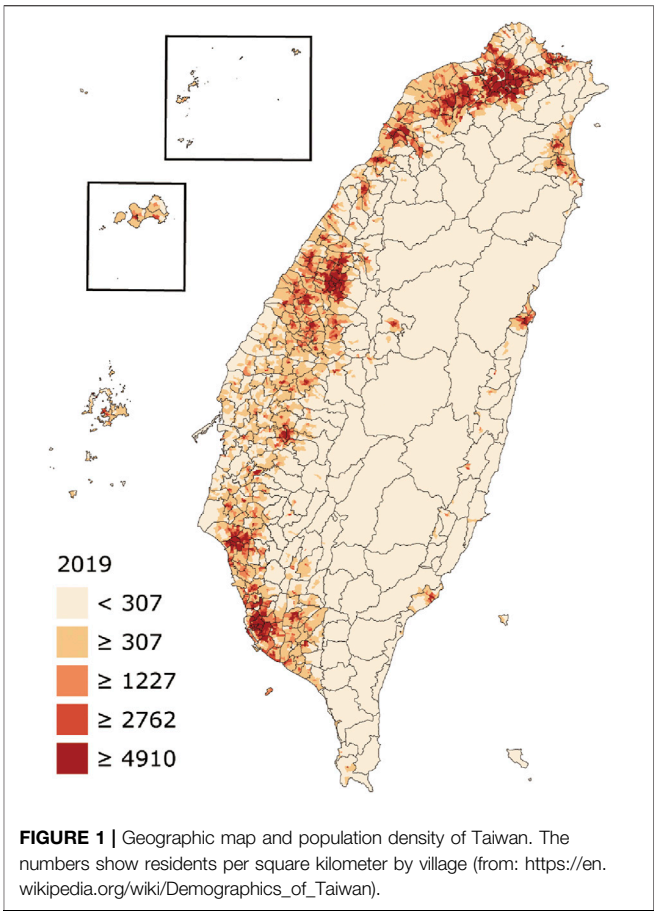


FIGURE 1 | Geographic map and population density of Taiwan. The numbers show residents per square kilometer by village (from: https://en.wikipedia.org/wiki/Demographics_of_Taiwan).

Similarity and Differences Between General Population and the Cancer Cohort

Data from both general population and cancer patients gave a unique opportunity to compare the similarity and differences of *BRCA* variation between the two groups with the same ethnic background. Although the total number of *BRCA* variants at the individual level was similar, significant differences existed between the two groups. The types of *BRCA* variation between the two groups were significantly different, including nonsynonymous SNV, synonymous SNV, stopgain, frameshift deletion, frameshift insertion, and splice site; and the frequency of nonsynonymous SNV and synonymous SNV in general population was higher than that in the cancer cohort (54.5% vs. 36.8% and 32.1% vs. 11.1%, $p < 0.001$ and 0.000, accordingly), whereas the frequency of stopgain, frameshift deletion/insertion, and splice variants was higher in the cancer cohort than that in the general population (**Table 1**). Significant differences in the clinical classification were also present in between. For example, 40.9% of *BRCA* variants in the cancer cohort were pathogenic variants, which was much higher than the value of 3.3% in general population ($p < 0.000$); VUS (variants of uncertain significance) and likely benign were significantly higher in general population than those in the cancer cohort (24.9% vs. 14% and 39.7% vs. 17%, $p < 0.009$, 0.000 accordingly) (**Table 1**).

TABLE 1 | BRCA variants identified in Taiwanese population.

Mutation type	General population		Total (%)	Cancer population		Total (%)	*p* Value
	***BRCA1* (%)**	***BRCA2* (%)**		***BRCA1* (%)**	***BRCA2* (%)**	—	—
A. Types of variation							
Nonsynonymous SNV	43 (53.1)	71 (55.5)	114 (54.5)	21 (31.3)	42 (40.4)	63 (36.8)	***0.001***
Synonymous SNV	27 (33.3)	40 (31.3)	67 (32.1)	8 (11.9)	11 (10.6)	19 (11.1)	***0.000***
Intron variant	7 (8.6)	10 (7.8)	17 (8.1)	7 (10.4)	2 (1.9)	9 (5.3)	0.270
Stopgain/nonsense	0 (0)	1 (0.8)	1 (0.5)	8 (11.9)	11 (10.6)	19 (11.1)	***0.000***
Frameshift deletion	2 (2.5)	4 (3.1)	6 (2.9)	19 (28.4)	28 (26.9)	47 (27.5)	***0.000***
Frameshift insertion	1 (1.2)	1 (0.8)	2 (1.0)	1 (1.5)	7 (6.7)	8 (4.7)	***0.024***
Non-frameshift deletion	1 (1.2)	1 (0.8)	2 (1.0)	0 (0)	1 (1.0)	1 (0.6)	0.683
5'/3' UTR	0 (0)	0 (0)	0 (0)	1 (1.5)	0 (0)	1 (0.6)	0.268
Splice site	0 (0)	0 (0)	0 (0)	2 (3.0)	2 (1.9)	4 (2.3)	***0.026***
B. Clinical classification							
Pathogenic	1 (1.2)	6 (4.7)	7 (3.3)	28 (41.8)	42 (40.4)	70 (40.9)	***0.000***
Likely pathogenic	1 (1.2)	0 (0)	1 (0.5)	3 (4.5)	1 (1.0)	4 (2.3)	0.113
Uncertain significance	20 (24.7)	32 (25)	52 (24.9)	9 (13.4)	15 (14.4)	24 (14.0)	***0.009***
Likely benign	34 (42)	49 (38.3)	83 (39.7)	9 (13.4)	20 (19.2)	29 (17.0)	***0.000***
Benign	20 (24.7)	35 (27.3)	55 (26.3)	13 (19.4)	18 (17.3)	31 (18.1)	0.058
Conflicting interpretations	2 (2.5)	2 (1.6)	4 (1.9)	0 (0)	0 (0)	0 (0)	0.069
Unclassified	3 (3.7)	4 (3.1)	7 (3.3)	5 (7.5)	8 (7.7)	13 (7.6)	0.065
Total	81	128	209	67	104	171	—

The bold-italic values indicate the numbers between general population and cancer population were statistically significant different.

Similarity and Differences From Other Ethnic Populations

We compared *BRCA* variants between Taiwanese population and other populations including mainland Chinese (Bhaskaran et al., 2019); Macau Chinese representing southern Chinese (Qin et al., 2020); Asian populations including Korean, Japanese, and Indian (Bhaskaran et al., 2020); and non-Asian populations, of which the majority were Caucasians (Dutil et al., 2015; Rebbeck et al., 2018). The results show that 35.5 and 37.6% of the Taiwanese variants were shared with Macau Chinese and mainland Chinese, respectively, 27.5% with Japanese, 20.3% with Korean, 11.3% with Indian, and 53.1% of entire non-Asian populations. The different sharing rates reflected the evolutionary relationship of Taiwanese population with non-Taiwanese populations (**Table 2**). We also compared with the *BRCA* variants from Fujian Chinese, which has the closest genetic tie with the Taiwanese population. Of the 18 *BRCA* variants available for comparison, 8 (44.4%) were matched by Taiwanese variants.

VUS, Unclassified Variants, and Novel Variants

Of the *BRCA* variants identified, 20.7% were VUS (52 in *BRCA1* and 24 in *BRCA2*, **Table 1**), 6.0% were unclassified variants (seven in *BRCA1* and 13 in *BRCA2*, **Table 1**), and 35.2% (118 *BRCA* variants) were absent in the *BRCA* data from worldwide ethnic populations (**Table 2**). The combination of VUS, unclassified, and novel variants accounted for 61.9% of all 335 *BRCA* variants identified in the Taiwanese population. Although the definitive classification for these variants remains to be solved, they may enrich with the Taiwanese-specific pathogenic *BRCA* variants. For example, 64.4% of the 118 *BRCA* variants were nonsynonymous SNV, frameshift insertion/deletion/substitution, stopgain, and non-frameshift deletion (**Table 2**, **Supplementary Tables S2, S3**).

To further test this possibility, we used the molecular dynamic simulation (MDS) to measure the impact of the four *BRCA1* unclassified variants (c.5068A > C p.Lys1690Gln, c.5347A > C p.Met1783Leu; c.5347A > G p.Met1783Val; c.5349G > A p.Met1783Ile) located at *BRCA1* BRCT repeat on BRCT structural stability, and use the information as the indication for their potential deleterious effects. Of the four unclassified variants, c.5068A > C p.Lys1690Gln and c.5347A > C p.Met1783Leu were predicted to be deleterious (**Figure 2**). Taking c.5347A > C p.Met1783Leu as an example, p.Met1783 is located within the α'1 helix at C terminal near the edge of the inter repeat interface of the native BRCT structure. While p.Met1783Leu by c.5347A > C was sterically stable without physical contact or clashes with adjoining residues, it unfolded the structure of BRCT and destabilized the hydrophobic interface, causing reposition between the two *BRCA1* BRCT repeats as reflected by the larger structure deviation and flexibility, reduced NH bond, and decreased structure compactness as measured by six different MDS programs (RMSD, RMSF, Rg, SASA, NH bond, and Covariance). The results showed that of the three missense variant-caused substitutions at the same position (p.Met1783Leu; p.Met1783Val; p.Met1783Ile), p.Met1783Leu was deleterious by disturbing BRCT structure stability.

We also used four different types of *in silico* prediction programs including SIFT, Polyphen2, LRT, and MutationTaster to predict the deleteriousness of the four unclassified variants. The results showed that the two deleterious variants (p.Lys1690Gln, p.Met1783Leu) predicted by MDS were also predicted as deleterious by at least three different programs. For example, p.Met1783Leu was predicted by all four programs as deleterious (**Table 3**).

The results from MDS and *in silico* prediction provide strong evidence for the enrichment of ethnic-specific deleterious variants in the unclassified variants.

TABLE 2 | Comparison between Taiwanese population and other populations.

Population	*BRCA1* Total	Matched	*BRCA2* Total	Matched	*BRCA* Total	Matched (%)	Unmatched (%)
Chinese							
Mainland Chinese	758	57	764	69	1,522	126 (37.6)	209 (62.4)
Macau Chinese	264	46	395	73	659	119 (35.5)	216 (64.5)
Non-Chinese Asian							
Japanese	415	37	649	55	1,064	92 (27.5)	243 (72.5)
Korean	281	35	300	33	581	68 (20.3)	267 (79.7)
Indian	274	23	244	15	518	38 (11.3)	297 (88.7)
Non-Asian							
BED	19,190	71	19,906	107	39,096	178 (53.1)	157 (46.9)
Total[a]	—	84		133	—	217 (64.8)	118 (35.2)

[a]*Distinct numbers by counting overlapped variants only once.*

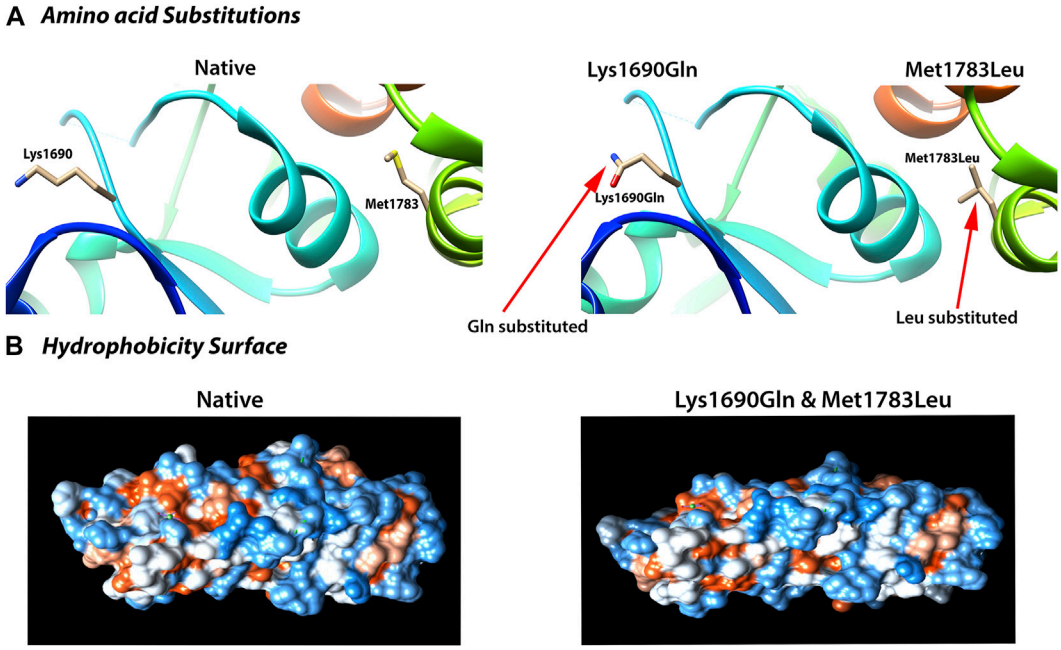


FIGURE 2 | Deleterious impact of unclassified variants (c.5068 A > C; p.Lys1690Gln; c.5347 A > C; p.Met1783Leu) on BRCA1 BRCT structural stability. **(A)** Amino acid substitution showing the variant-caused amino acid change from Lys and Met (left) in the native structure to Gln and Leu (right) at the position of 1,690 and 1783, respectively. **(B)** Deleterious effects reflected by the change in hydrophobicity surface in the mutant BRCT. Both Lys1690Gln and Met1783Leu caused nearly identical change as shown here. The results were from 40 ns simulation (see text for detailed explanation).

Pathogenic Variants and Prevalence

In the general population, we identified eight *BRCA* pathogenic and likely pathogenic variants, two in *BRCA1* with two carriers and six in *BRCA2* with six carriers. The eight pathogenic variant carriers in the 1,517 general individuals represent the prevalence of 0.53% *BRCA* pathogenic variants (0.13% in *BRCA1* and 0.40% in *BRCA2*) in Taiwanese population. The higher prevalence of *BRCA2* than *BRCA1* is consistent with the pattern in other Asian ethnic populations (Bhaskaran et al., 2020). With 0.53% prevalence, there are estimated 126,140 *BRCA* pathogenic variant carriers (30,940 in *BRCA1* and 95,200 in *BRCA2*) estimated in the Taiwanese population of 23.8 million or one *BRCA* pathogenic variant carrier in every 189 Taiwanese individuals. In the cancer cohort of 2,665 cases, we identified 74 *BRCA* pathogenic variants, 31 in *BRCA1* with 40 carriers (2.1% in 1,880 cases) and 43 in *BRCA2* with 61 carriers (2.5% in 2,417 cases), resulting in the prevalence of 3.8% in the Taiwanese cancer cohort of breast/ovarian cancer. Five pathogenic variants (two in *BRCA1* and three in *BRCA2*) were present only in general population with five carriers, and three *BRCA2* pathogenic variants were present in both general population and the cancer cohort with six carriers. *BRCA2* c.5164_5165delAG was a known pathogenic variant in Chinese cancer patients (Kwong et al., 2012). This pathogenic variant was present in the cancer cohort with 10 carriers but not in the general population. This variant is a potential founder mutation in the Taiwanese population and need to be validated by the haplotype test (**Table 4**).

TABLE 3 | Prediction of deleterious variants for the unclassified variants in BRCA1 BRCT repeats.

cDNA	Amino acid	MDS	*In silico* prediction programs				
			SIFT	Polyphen2_HDIV	LRT	MutationTaster	Total deleterious
c.5068A > C	p.Lys1690Gln	Deleterious	Deleterious	Probably damaging	Neutral	Disease causing	3 of 4
c.5347A > C	p.Met1783Leu	Deleterious	Deleterious	Probably damaging	Deleterious	Disease causing	4 of 4
c.5347A > G	p.Met1783Val	Tolerated	Deleterious	Probably damaging	Deleterious	Disease causing	4 of 4
c.5349G > A	p.Met1783Ile	Tolerated	Deleterious	Probably damaging	Neutral	Disease causing	3 of 4

DISCUSSION

It is well known that the number of benign variants is higher in the general population than in the cancer cohort, and the number of pathogenic mutations is higher in the cancer cohort than in the general population. Our current study aimed to obtain the detailed variant information including position, frequency, classification, and ethnic specificity in the Taiwanese healthy population and the cancer cohort in order to understand the genetic basis of *BRCA* variation in the population and to develop a precise reference to guide clinical applications.

Taiwanese population has its unique genetic features in reflecting its evolutionary and admixture history (Chen et al., 2016). With a population size of 23.8 million, *BRCA* variation information provides a unique source to understand its genetic variation in adaptation to the unique environment and the pathogenic variation causing cancer risk in the population. Data from our study provide an overview for *BRCA* variation and pathogenicity in this population, and further confirms the highly ethnic-specific nature of *BRCA* variation in eastern Asian population (Bhaskaran et al., 2020).

The availability of *BRCA* variant data from both general population and the cancer cohort allows comparison of the similarity and differences of *BRCA* variation between the two groups with the same ethnic background under the same geological environment. The higher rate of *BRCA* variation in its general population over other ethnic populations may reflect the rapidly evolving *BRCA* in Taiwanese population for better adaptation in Taiwan's natural environment (Chen et al., 2016). This could be a factor contributing to higher prevalence of pathogenic variation in Taiwanese general population by increased probability of generating more pathogenic variants. The prevalence of 0.53% of pathogenic variation in the general Taiwanese population is the highest in Asian ethnic populations, comparing to 0.26% in Japanese (Momozawa et al., 2018), 0.29% in southern Chinese (Qin et al., 2020), 0.38% in mainland Chinese (Dong et al., 2020), and 0.39% in Malaysia (Wen et al., 2018), and has reached the same level of 0.53% as in Caucasian populations (Kurian et al., 2019). One *BRCA* pathogenic variant carrier in every 189 Taiwanese individuals represents a serious threat for public health in Taiwanese population, justifying the inclusion of *BRCA*-related cancer diagnosis, treatment, and prevention in the healthcare system in Taiwan. Considering its impact on population health, further confirmation of the result with a larger sample size will be necessary to validate the observations. The prevalence of 3.8% in the cancer cohort was lower than that in other ethnic cancer patient groups, such as 5.4% in Caucasians (Sun et al., 2017) and 5.3% in mainland Chinese (Bhaskaran et al., 2019).

All the pathogenic variants identified in Taiwanese population are present in public *BRCA* databases. Similar situation exists for the pathogenic variants identified in other Asian populations (Bhaskaran et al., 2019; Bhaskaran et al., 2020; Dong et al., 2020; Qin et al., 2020; Zhang et al., 2020). In the meantime, 44.1% of the *BRCA* variants identified in Taiwanese population remain as novel, VUS, and unclassified variants. Our *BRCA* study across multiple ethnic Asian populations also showed that 30–50% of variants present in each population were novel, VUS, and unclassified variants. The distribution patterns of pathogenic and unclassified variants seem to suggest that pathogenic variants are universally shared between human populations, whereas non-pathogenic variants are largely ethnic specific. However, such assumption does not have a biological sense. Considering that *BRCA* variation is highly ethnic specific and a large portion of the *BRCA* variants identified in ethnic population remain unclassified, it will be logical to consider that ethnic-specific pathogenic variants should also exist, and these are likely enriched within the unclassified variants. The pathogenic variants highly shared between the human populations represent the common pathogenic variants inherited from their common ancestors. They are identifiable by referring to the current well-annotated *BRCA* pathogenic data predominately derived from Caucasian populations (Rebbeck et al., 2018; Bhaskaran et al., 2019). As these reference databases lack the pathogenic variant data from the non-Caucasian populations, the ethnic-specific pathogenic variants in the non-Caucasian populations are not identifiable by referring to these databases. The ethnic-specific pathogenic variants can be highly enriched within the ethnic-specific novel, VUS, and unclassified variants, as evidenced from our MDS and *in silico* analyses. However, it remains a challenge in cancer genetic study to develop extensive ethnic-specific pathogenic variant references.

In summary, the data generated from the study provide a comprehensive view for *BRCA* variation in the Taiwanese population and a reference for clinical applications in *BRCA*-related cancer in the Taiwanese population.

MATERIALS AND METHODS

Data Sources

Studies were selected by the following inclusion criteria: 1) cancer patients should be pathologically confirmed, 2) germline variants in *BRCA1/2* should be genotyped, and 3) studies on nonhuman or

TABLE 4 | *BRCA* pathogenic and likely pathogenic variants identified in Taiwanese population.

Variant impact	cDNA	Protein	Mutation type	Cases	Carrier	References[a]
BRCA1						
Pathogenic	c.66dupA	p.Glu23Argfs*17	Frameshift insertion	120	1	15
Pathogenic	c.303 T > A	p.Tyr101Ter	Stopgain	28	1	14
Pathogenic	c.470_471delCT	p.Ser157Ter	Stopgain	133	1	10
Pathogenic	c.726delT	p.Ser242Argfs*4	Frameshift deletion	99	1	9
Pathogenic	c.928C > T	p.Gln310Ter	Stopgain	271	3	3,14,15
Pathogenic	c.981_982delAT	p.Cys328Ter	Frameshift deletion	1,517	1	TWB
Pathogenic	c.1361delG	p.Ser454Ilefs*20	Frameshift deletion	161	2	10.14
Pathogenic	c.1934delC	p.Ser645Leufs*5	Frameshift deletion	133	1	10
Pathogenic	c.2188G > T	p.Glu730Ter	Stopgain	99	1	9
Pathogenic	c.2393delC	p.Pro798Glnfs*4	Frameshift deletion	68	1	11
Pathogenic	c.2679_2682delGAAA	p.Lys893Asnfs*105	Frameshift deletion	133	1	10
Pathogenic	c.3083delG	p.Arg1028Leufs*19	Frameshift deletion	28	1	14
Pathogenic	c.3228_3229delAG	p.Gly1077Alafs*7	Frameshift deletion	253	2	10.15
Pathogenic	c.3257 T > G	p.Leu1086Ter	Stopgain	120	1	15
Pathogenic	c.3472delG	p.Glu1158Lysfs*1	Frameshift deletion	68	1	11
Pathogenic	c.3607C > T	p.Arg1203Ter	Stopgain	201	3	10.11
Pathogenic	c.3637G > T	p.Glu1213Ter	Stopgain	68	1	11
Pathogenic	c.3644_3648delACTTA	p.Asn1215Ilefs*1	Frameshift deletion	480	1	12
Pathogenic	c.3770_3771delAG	p.Glu1257Glyfs*8	Frameshift deletion	253	2	10.15
Pathogenic	c.3858_3861delTGAG	p.Ser1286Argfs*19	Frameshift deletion	99	1	9
Pathogenic	c.4356delA	p.Ala1453Glnfs*1	Frameshift deletion	120	1	15
Pathogenic	c.4678_4679delGG	p.Gly1560Asnfs*12	Frameshift deletion	480	2	12
Pathogenic	c.5030_5033delCTAA	p.Thr1677Ilefs*1	Frameshift deletion	36	1	8
Pathogenic	c.5075-1G > A	—	Splice site	480	1	12
Pathogenic	c.5211_5212delAG	p.Gly1738Argfs*90	Frameshift deletion	99	1	9
Pathogenic	c.5332+1G > A	—	Splice site	167	2	9.11
Pathogenic	c.5335delC	p.Gln1779Asnfs*13	Frameshift deletion	36	1	8
Pathogenic	c.5470_5477delATTGGGCA	p.Ile1824Aspfs*2	Frameshift deletion	133	1	10
Pathogenic	c.5536C > T	p.Gln1846Ter	Stopgain	36	1	8
Likely pathogenic	c.122 A > G	p.His41Arg	Nonsynonymous SNV	658	1	13
Likely pathogenic	c.5072C > A	p.Thr1691Lys	Nonsynonymous SNV	36	1	8
Likely pathogenic	c.5288G > A	p.Gly1763Glu	Nonsynonymous SNV	1,517	1	TWB
Likely pathogenic	c.5396C > A	p.Thr1799Asn	Nonsynonymous SNV	120	1	15
BRCA2						
Pathogenic	c.-7_9del16	—	Frameshift deletion	480	1	12
Pathogenic	c.469_470delAA	p.Lys157Valfs*24	Frameshift deletion	480	2	12
Pathogenic	c.631G > C	p.Val211Leu	Nonsynonymous SNV	133	1	10
Pathogenic	c.750_753delGACA	p.Asp252Valfs*23	Frameshift deletion	1,517	1	TWB
Pathogenic	c.755_758delACAG	p.Asp252Valfs*23	Frameshift deletion	480	1	12
Pathogenic	c.773_774delAA	p.Glu260Serfs*13	Frameshift deletion	1,616	2	9,TWB
Pathogenic	c.774_775delAA	p.Glu260Serfs*13	Frameshift deletion	658	1	13
Pathogenic	c.857C > G	p.Ser286Ter	Stopgain	480	1	12
Pathogenic	c.1036_1037delAA	p.Asn346Profs*9	Frameshift deletion	68	1	11
Pathogenic	c.1058C > T	p.Ser353Leu	Nonsynonymous SNV	1,517	1	TWB
Pathogenic	c.1765_1766delAA	p.Lys589Valfs*6	Frameshift deletion	1,517	1	TWB
Pathogenic	c.2095C > T	p.Gln699Ter	Stopgain	480	1	12
Pathogenic	c.2339C > G	p.Ser780Ter	Stopgain	99	1	9
Pathogenic	c.2442delC	p.Met815Trpfs*9	Frameshift deletion	516	2	2.12
Pathogenic	c.2754delC	p.Asn918Lysfs*41	Frameshift deletion	480	1	12
Pathogenic	c.2808_2811delACAA	p.Ala938Profs*20	Frameshift deletion	613	2	10.12
Pathogenic	c.2845delT	p.Tyr949Metfs*10	Frameshift deletion	36	1	2
Pathogenic	c.2990 T > G	p.Leu997Ter	Stopgain	480	1	12
Pathogenic	c.3109C > T	p.Gln1037Ter	Stopgain	600	4	12.15
Pathogenic	c.3322 A > T	p.Lys1108Ter	Stopgain	480	1	12
Pathogenic	c.4914dupA	p.Val1639Serfs*2	Frameshift insertion	480	1	12
Pathogenic	c.5141_5144delATTT	p.Tyr1714Cysfs*9	Frameshift deletion	480	1	12
Pathogenic	c.5164_5165delAG	p.Ser1722Tyrfs*3	Frameshift deletion	740	10	9,10,12,14
Pathogenic	c.5574_5577delAATT	p.Ile1859Lysfs*2	Frameshift deletion	133	1	10
Pathogenic	c.5621_5624delTTAA	p.Ile1874Argfs*33	Frameshift deletion	480	1	12
Pathogenic	c.6275_6276delTT	p.Leu2092Profs*6	Frameshift deletion	480	1	12
Pathogenic	c.6448delA	p.Val2151Phefs*16	Frameshift deletion	133	1	10
Pathogenic	c.6468_6469delTC	p.Gln2157Ilefs*17	Frameshift deletion	36	1	2
Pathogenic	c.6484_6485delAA	p.Lys2162Thrfs*12	Frameshift deletion	28	1	14

(Continued on following page)

TABLE 4 | (*Continued*) *BRCA* pathogenic and likely pathogenic variants identified in Taiwanese population.

Variant impact	cDNA	Protein	Mutation type	Cases	Carrier	References[a]
Pathogenic	c.6490C > T	p.Gln2164Ter	Stopgain	480	1	12
Pathogenic	c.6645delC	p.Ser2216Profs*12	Frameshift deletion	133	1	10
Pathogenic	c.6800C > A	p.Ser2267Ter	Stopgain	480	1	12
Pathogenic	c.7409dupT	p.Thr2471Hisfs*3	Frameshift insertion	133	1	10
Pathogenic	c.7567_7568delCT	p.Leu2523Glufs*14	Frameshift deletion	68	1	11
Pathogenic	c.7977-1G > T	—	Splice site	68	1	11
Pathogenic	c.8009C > T	p.Ser2670Leu	Nonsynonymous SNV	2,175	2	13,TWB
Pathogenic	c.8234dupT	p.Thr2746Aspfs*17	Frameshift insertion	480	1	12
Pathogenic	c.8243G > A	p.Gly2748Asp	Nonsynonymous SNV	133	1	10
Pathogenic	c.8323delA	p.Met2775Cysfs1*	Frameshift deletion	133	1	10
Pathogenic	c.8485C > T	p.Gln2829Ter	Stopgain	480	1	12
Pathogenic	c.8488-1G > A	—	Splice site	99	1	9
Pathogenic	c.8531_8532delAA	p.Glu2846Glyfs2*1	Frameshift deletion	133	1	10
Pathogenic	c.8961_8964delGAGT	p.Ser2988Phefs*11	Frameshift deletion	480	1	12
Pathogenic	c.9227delG	p.Gly3076Aspfs6*	Frameshift deletion	480	1	12
Pathogenic	c.9739C > T	p.Gln3247Ter	Stopgain	2,175	2	13,TWB
Likely pathogenic	c.3883C > T	p.Gln1295Ter	Stopgain	480	1	12

[a]***Supplementary Table S1*** *lists each of the references. TWB: Taiwan Biobank.*

cell line were excluded. Thoroughly searching PubMed and Google Scholar using the keywords such as "*BRCA1*" "*BRCA2*" "Taiwan" "Taiwanese" and "cancer predisposition" we identified 15 publications reporting the *BRCA* data from Taiwanese cancer patients between 1997 and 2020 (Liu et al., 1997; Li et al., 1999; Wang et al., 2000; Chen et al., 2003; Lin et al., 2003; Chen et al., 2005; Chang et al., 2006; Kuo et al., 2012; Chao et al., 2016; Lin et al., 2016; Sung et al., 2017; Wang et al., 2018; Lin et al., 2019; Chao et al., 2020; Lin et al., 2020). Searching Taiwan Biobank (https://taiwanview.twbiobank.org.tw/index; accessed December 15, 2020) using the keywords "*BRCA1*" and "*BRCA2*" we obtained the *BRCA* variation data derived from Taiwanese general population.

Data Analysis

The following details were extracted from the filtered publications, including first author, year of publication, *BRCA* variants, mutation type of variants, study population, and the number of cases in the study. We standardized the collected *BRCA* variation data following the Human Genome Variation Society (HGVS) guidelines (den Dunnen et al., 2016). The following reference sequences were used for the mapping analysis: *BRCA1*: cDNA NM_007294.3, protein NP_009225.1, and genome hg19 NC_000017.10; *BRCA2*; cDNA NM_000059.3, protein NP_000050.2, and genome hg19 NC_000013.10. We annotated the variants using the ANNOVAR program (Wang et al., 2010). The population frequency was referred to East Asian variants (EAC) from the 1,000 Genome Project (Fairley et al., 2020), the Exome Aggregation Consortium (ExAC) (Lek et al., 2016), and the Genome Aggregation Database (gnomAD) (Karczewski et al., 2020). The variants were compared with the following two *BRCA* databases: the BRCA Exchange Database (BED, http://brcaexchange.org, accessed December 15, 2020) and ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/, accessed December 15, 2020). The variants present in *BRCA* databases were classified as known variants by referring to the existing classification of pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. The classes for those variants not present in existing *BRCA* databases were predicted using the InterVar program with default parameters (Li and Wang, 2017). The Fujianese *BRCA* variants were extracted from the whole genome sequences of Fujian individuals (Yang et al., 2020).

Molecular Dynamics Simulations and *in silico* Prediction

We utilized molecular dynamics simulations (MDS) to measure the impact of the four BRCA1 unclassified variants in the Taiwanese population (c.5068A > C p.Lys1690Gln, c.5347A > C p.Met1783Leu; c.5347A > G p.Met1783Val; and c.5349G > A p.Met1783Ile) on the stability of the *BRCA1* BRCT structure. The MDS system was developed for *BRCA1* BRCT variant classification as described in details (Sinha and Wang, 2020). In brief, the process included two major steps: 1) modeling mutant structure. Using the wild-type BRCT structure as the template, each mutant structure was constructed using the Modeller program (*version 9.22,* UCSF, CA, United States), and further evaluated using the PROCHECK (Sippl, 1993) and PROSA (Wiederstein et al., 2007) programs following the instructions; 2) analyzing the impact of variants on BRCT structural stability by using MDS (Karplus, 2002). Using the wild-type structure as the reference, MDS analyzes the trajectory of the mutant structure over a time period through multiple parameters including RMSD (root mean square deviation) to measure the average deviation in the backbone of Cα trace (Dong et al., 2018), RMSF (root mean square fluctuations) to measure the residue flexibility of the structure (Benson et al., 2012), Rg (radius of gyration) to measure the distance of the atoms of the structure from its center of gravity and axis for the compactness of each structure (Daidone et al., 2003), SASA (solvent accessible surface area) to measure the surface accessibility (Sheu et al., 2003), NH-bond (number of hydrogen bonds) to measure the overall change in the compactness of the mutant structures, and

covariance analysis to compare the overall protein motions (Amadei et al., 1993).

Four *in silico* prediction methods of SIFT (Kumar et al., 2009), Polyphen2_HDIV (Adzhubei et al., 2010), LRT (Chun et al., 2009), and MutationTaster (Schwarz et al., 2010) were used to predict the deleteriousness for the four unclassified variants in BRCT repeats following the default setting in each method.

Statistical Analysis

A chi-square test was used to compare the differences of *BRCA* variant data between different populations using SPSS (version 26.0, IBM, NY, United States). A p value lower than 0.05 was considered as statistically significant.

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

JC: data collection, annotation, analysis, and manuscript writing; SS: structural analysis; ZQ: data analysis; SW: funding, experimental design, data analysis and interpretation, and manuscript writing and revision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Macau Science and Technology Development Fund (No. 085/2017/A2, 0,077/2019/AMJ); the University of Macau (No. SRG2017-00097-FHS, MYRG2019-00018-FHS); and Faculty of Health Sciences, the University of Macau (No. FHSIG/SW/0007/2020P, startup fund) (SW).

ACKNOWLEDGMENTS

We thank Taiwan Biobank for providing the BRCA data from Taiwanese general population. We thank the Information and Communication Technology Office (ICTO), the University of Macau for providing the High-Performance Computing Cluster (HPCC) resource and facilities for the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.685174/full#supplementary-material

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* 7 (4), 248–249. doi:10.1038/nmeth0410-248

Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993). Essential Dynamics of Proteins. *Proteins* 17 (4), 412–425. doi:10.1002/prot.340170408

Benson, N. C., and Daggett, V. (2012). A Comparison of Multiscale Methods for the Analysis of Molecular Dynamics Simulations. *J. Phys. Chem. B* 116 (29), 8722–8731. doi:10.1021/jp302103t

Bhaskaran, S. P., Chandratre, K., Gupta, H., Zhang, L., Wang, X., Cui, J., et al. (2019). Germline Variation in BRCA1/2 Is Highly Ethnic-specific: Evidence from over 30,000 Chinese Hereditary Breast and Ovarian Cancer Patients. *Int. J. Cancer* 145 (4), 962–973. doi:10.1002/ijc.32176

Bhaskaran, S. P., Huang, T., Rajendran, B. K., Guo, M., Luo, J., Qin, Z., et al. (2020). Ethnic-specific BRCA1/2 Variation within Asia Population: Evidence from over 78 000 Cancer and 40 000 Non-cancer Cases of Indian, Chinese, Korean and Japanese Populations. *J. Med. Genet.* [Epub ahead of print]. doi:10.1136/jmedgenet-2020-107299

Burke, W., Daly, M., Garber, J., Botkin, J., Kahn, M. J., Lynch, P., et al. (1997). Recommendations for Follow-Up Care of Individuals with an Inherited Predisposition to Cancer. *JAMA* 277 (12), 997–1003. doi:10.1001/jama.1997.03540360065034

Chang, T.-W., Wang, S.-M., Guo, Y.-L. L., Tsai, P.-C., Huang, C.-J., and Huang, W. (2006). Glutathione S-Transferase Polymorphisms Associated with Risk of Breast Cancer in Southern Taiwan. *The Breast* 15 (6), 754–761. doi:10.1016/j.breast.2006.03.008

Chao, A., Chang, T.-C., Lapke, N., Jung, S.-M., Chi, P., Chen, C.-H., et al. (2016). Prevalence and Clinical Significance of BRCA1/2 Germline and Somatic Mutations in Taiwanese Patients with Ovarian Cancer. *Oncotarget* 7 (51), 85529–85541. doi:10.18632/oncotarget.13456

Chao, A., Lin, Y.-H., Yang, L.-Y., Wu, R.-C., Chang, W.-Y., Chang, P.-Y., et al. (2020). BRCA1/2mutation Status in Patients with Metachronous Breast and Ovarian Malignancies: Clues towards the Implementation of Genetic Counseling. *J. Gynecol. Oncol.* 31 (3), e24. doi:10.3802/jgo.2020.31.e24

Chen, C.-H., Yang, J.-H., Chiang, C. W. K., Hsiung, C.-N., Wu, P.-E., Chang, L.-C., et al. (2016). Population Structure of Han Chinese in the Modern Taiwanese Population Based on 10,000 Participants in the Taiwan Biobank Project. *Hum. Mol. Genet.* 25, ddw346–5331. doi:10.1093/hmg/ddw346

Chen, F.-M., Hou, M.-F., Chang, M.-Y., Wang, J.-Y., Hsieh, J.-S., Ou-Yang, F., et al. (2005). High Frequency of Somatic Missense Mutation of BRCA2 in Female Breast Cancer from Taiwan. *Cancer Lett.* 220 (2), 177–184. doi:10.1016/j.canlet.2004.10.024

Chen, S.-T., Chen, R.-A., Kuo, S.-J., and Chien, Y.-C. (2003). Mutational Screening of Breast Cancer Susceptibility Gene 1 from Early Onset, Bi-lateral, and Familial Breast Cancer Patients in Taiwan. *Breast Cancer Res. Treat.* 77 (2), 133–143. doi:10.1023/a:1021386026051

Chun, S., and Fay, J. C. (2009). Identification of Deleterious Mutations within Three Human Genomes. *Genome Res.* 19 (9), 1553–1561. doi:10.1101/gr.092619.109

Daidone, I., Amadei, A., Roccatano, D., and Nola, A. D. (2003). Molecular Dynamics Simulation of Protein Folding by Essential Dynamics Sampling: Folding Landscape of Horse Heart Cytochrome C. *Biophysical J.* 85 (5), 2865–2871. doi:10.1016/S0006-3495(03)74709-2

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., et al. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* 37 (6), 564–569. doi:10.1002/humu.22981

Dong, H., Chandratre, K., Qin, Y., Zhang, J., Tian, X., Rong, C., et al. (2020). Prevalence of BRCA1/BRCA2 Pathogenic Variation in Chinese Han Population. *J. Med. Genet.* [Epub ahead of print]. doi:10.1136/jmedgenet-2020-106970

Dong, Y.-w., Liao, M.-l., Meng, X.-l., and Somero, G. N. (2018). Structural Flexibility and Protein Adaptation to Temperature: Molecular Dynamics Analysis of Malate Dehydrogenases of marine Molluscs. *Proc. Natl. Acad. Sci. USA* 115 (6), 1274–1279. doi:10.1073/pnas.1718910115

Dutil, J., Golubeva, V. A., Pacheco-Torres, A. L., Diaz-Zabala, H. J., Matta, J. L., and Monteiro, A. N. (2015). The Spectrum of BRCA1 and BRCA2 Alleles in Latin America and the Caribbean: a Clinical Perspective. *Breast Cancer Res. Treat.* 154 (3), 441–453. doi:10.1007/s10549-015-3629-3

Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) Collection of Open Human Genomic Variation Resources. *Nucleic Acids Res.* 48 (D1), D941–D947. doi:10.1093/nar/gkz836

Friebel, T. M., Andrulis, I. L., Balmaña, J., Blanco, A. M., Couch, F. J., Daly, M. B., et al. (2019). BRCA1 and BRCA2 Pathogenic Sequence Variants in Women of African Origin or Ancestry. *Hum. Mutat.* 40 (10), 1781–1796. doi:10.1002/humu.23804

Jerez, Y., Márquez-Rodas, I., Aparicio, I., Alva, M., Martín, M., and López-Tarruella, S. (2020). Poly (ADP-Ribose) Polymerase Inhibition in Patients with Breast Cancer and BRCA 1 and 2 Mutations. *Drugs* 80 (2), 131–146. doi:10.1007/s40265-019-01235-5

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., et al. (2020). The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. *Nature* 581, 434–443. doi:10.1038/s41586-020-2308-7

Karplus, M. (2002). Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.* 35, 321–323. doi:10.1021/ar020082r

Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* 317 (23), 2402–2416. doi:10.1001/jama.2017.7112

Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the Effects of Coding Non-synonymous Variants on Protein Function Using the SIFT Algorithm. *Nat. Protoc.* 4 (7), 1073–1081. doi:10.1038/nprot.2009.86

Kuo, W.-H., Lin, P.-H., Huang, A.-C., Chien, Y.-H., Liu, T.-P., Lu, Y.-S., et al. (2012). Multimodel Assessment of BRCA1 Mutations in Taiwanese (Ethnic Chinese) Women with Early-Onset, Bilateral or Familial Breast Cancer. *J. Hum. Genet.* 57(2), 130–138. doi: doi:10.1038/jhg.2011.142

Kurian, A. W., Ward, K. C., Howlader, N., Deapen, D., Hamilton, A. S., Mariotto, A., et al. (2019). Genetic Testing and Results in a Population-Based Cohort of Breast Cancer Patients and Ovarian Cancer Patients. *Jco* 37, 1305–1315. doi:10.1200/JCO.18.01854

Kwong, A., Ng, E. K. O., Wong, C. L. P., Law, F. B. F., Au, T., Wong, H. N., et al. (2012). Identification of BRCA1/2 Founder Mutations in Southern Chinese Breast Cancer Patients Using Gene Sequencing and High Resolution DNA Melting Analysis. *PLoS One* 7 (9), e43994. doi:10.1371/journal.pone.0043994

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* 536, 285–291. doi:10.1038/nature19057

Levy-Lahad, E., Catane, R., Eisenberg, S., Kaufman, B., Hornreich, G., Lishinsky, E., et al. (1997). Founder BRCA1 and BRCA2 Mutations in Ashkenazi Jews in Israel: Frequency and Differential Penetrance in Ovarian Cancer and in Breast-Ovarian Cancer Families. *Am. J. Hum. Genet.* 60, 1059–1067.

Li, Q., and Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* 100 (2), 267–280. doi:10.1016/j.ajhg.2017.01.004

Li, S. S.-L., Tseng, H.-M., Yang, T.-P., Liu, C.-H., Teng, S.-J., Huang, H.-W., et al. (1999). Molecular Characterization of Germline Mutations in the BRCA1 and BRCA2 Genes from Breast Cancer Families in Taiwan. *Hum. Genet.* 104, 201–204. doi:10.1007/s004390050936

Lin, P.-C., Yeh, Y.-M., Wu, P.-Y., Hsu, K.-F., Chang, J.-Y., and Shen, M.-R. (2019). Germline Susceptibility Variants Impact Clinical Outcome and Therapeutic Strategies for Stage III Colorectal Cancer. *Sci. Rep.* 9, 3931. doi:10.1038/s41598-019-40571-0

Lin, P.-H., Kuo, W.-H., Huang, A.-C., Lu, Y.-S., Lin, C.-H., Kuo, S.-H., et al. (2016). Multiple Gene Sequencing for Risk Assessment in Patients with Early-Onset or Familial Breast Cancer. *Oncotarget* 7, 8310–8320. doi:10.18632/oncotarget.7027

Lin, P. H., Chen, M., Tsai, L. W., Lo, C., Yen, T. C., Huang, T. Y., et al. (2020). Using Next-generation Sequencing to Redefine BRCAness in Triple-negative Breast Cancer. *Cancer Sci.* 111 (4), 1375–1384. doi:10.1111/cas.14313

Lin, Y. P., Chen, Y. L., Chang, H. T., and Li, S. S. L. (2003). Nature of Genetic Variants in the BRCA1 and BRCA2 Genes from Breast Cancer Families in Taiwan. *Life Sci. J.* 6, 93–97. doi:10.7537/marslsj060309.15

Liu, F.-S., Ho, E. S.-C., Shih, R. T.-P., and Shih, A. (1997). Mutational Analysis of the BRCA1 Tumor Suppressor Gene in Endometrial Carcinoma. *Gynecol. Oncol.* 66 (3), 449–453. doi:10.1006/gyno.1997.4800

Lou, D. I., McBee, R. M., Le, U. Q., Stone, A. C., Wilkerson, G. K., Demogines, A. M., et al. (2014). Rapid Evolution of BRCA1 and BRCA2 in Humans and Other Primates. *BMC Evol. Biol.* 14 (1), 155. doi:10.1186/1471-2148-14-155

Momozawa, Y., Iwasaki, Y., Parsons, M. T., Kamatani, Y., Takahashi, A., Tamura, C., et al. (2018). Germline Pathogenic Variants of 11 Breast Cancer Genes in 7,051 Japanese Patients and 11,241 Controls. *Nat. Commun.* 9 (1), 4083. doi:10.1038/s41467-018-06581-8

Qin, Z., Kuok, C. N., Dong, H., Jiang, L., Zhang, L., Guo, M., et al. (2020). Can Population BRCA Screening Be Applied in Non-ashkenazi Jewish Populations? Experience in Macau Population. *J. Med. Genet.* [Epub ahead of print]. doi:10.1136/jmedgenet-2020-107181

Rebbeck, T. R., Friebel, T. M., Friedman, E., Hamann, U., Huo, D., Kwong, A., et al. (2018). Mutational Spectrum in a Worldwide Study of 29,700 Families with BRCA1 or BRCA2 Mutations. *Hum. Mutat.* 39, 593–620. doi:10.1002/humu.23406

Roy, R., Chun, J., and Powell, S. N. (2011). BRCA1 and BRCA2: Different Roles in a Common Pathway of Genome protection. *Nat. Rev. Cancer* 12 (1), 68–78. doi:10.1038/nrc3181

Schwarz, J. M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster Evaluates Disease-Causing Potential of Sequence Alterations. *Nat. Methods* 7 (8), 575–576. doi:10.1038/nmeth0810-575

Sheu, S.-Y., Yang, D.-Y., Selzle, H. L., and Schlag, E. W. (2003). Energetics of Hydrogen Bonds in Peptides. *Proc. Natl. Acad. Sci.* 100, 12683–12687. doi:10.1073/pnas.2133366100

Sinha, S., and Wang, S. M. (2020). Classification of VUS and Unclassified Variants in BRCA1 BRCT Repeats by Molecular Dynamics Simulation. *Comput. Struct. Biotechnol. J.* 18, 723–736. doi:10.1016/j.csbj.2020.03.013

Sippl, M. J. (1993). Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins* 17, 355–362. doi:10.1002/prot.340170404

Sun, J., Meng, H., Yao, L., Lv, M., Bai, J., Zhang, J., et al. (2017). Germline Mutations in Cancer Susceptibility Genes in a Large Series of Unselected Breast Cancer Patients. *Clin. Cancer Res.* 23 (20), 6113–6119. doi:10.1158/1078-0432.CCR-16-3227

Sung, P.-L., Wen, K.-C., Chen, Y.-J., Chao, T.-C., Tsai, Y.-F., Tseng, L.-M., et al. (2017). The Frequency of Cancer Predisposition Gene Mutations in Hereditary Breast and Ovarian Cancer Patients in Taiwan: From BRCA1/2 to Multi-Gene Panels. *PLoS One* 12, e0185615. doi:10.1371/journal.pone.0185615

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603

Wang, P.-H., Shyong, W. Y., Li, Y. F., Lee, H. H., Tsai, W. Y., Chao, H. T., et al. (2000). BRCA1 Mutations in Taiwanese with Epithelial Ovarian Carcinoma and Sporadic Primary Serous Peritoneal Carcinoma. *Jpn. J. Clin. Oncol.* 30 (8), 343–348. doi:10.1093/jjco/hyd092

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

Wen, W. X., Allen, J., Lai, K. N., Mariapun, S., Hasan, S. N., Ng, P. S., et al. (2018). Inherited Mutations in BRCA1 and BRCA2 in an Unselected Multiethnic Cohort of Asian Patients with Breast Cancer and Healthy Controls from Malaysia. *J. Med. Genet.* 55 (2), 97–103. doi:10.1136/jmedgenet-2017-104947

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* 35 (Web Server), W407–W410. doi:10.1093/nar/gkm290

Yang, M. A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.-C., et al. (2020). Ancient DNA Indicates Human Population Shifts and Admixture in Northern and Southern China. *Science* 369 (6501), 282–288. doi:10.1126/science.aba0909

Zhang, L., Bhaskaran, S. P., Huang, T., Dong, H., Chandratre, K., Wu, X., et al. (2020). Variants of DNA Mismatch Repair Genes Derived from 33,998 Chinese Individuals with and without Cancer Reveal Their Highly Ethnic-specific Nature. *Eur. J. Cancer* 125, 12–21. doi:10.1016/j.ejca.2019.11.004

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.

Copyright © 2021 Chian, Sinha, Qin 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 25 June 2021
doi: 10.3389/fmolb.2021.684050

Characterization of Hypoxia-Related Molecular Subtypes in Clear Cell Renal Cell Carcinoma to Aid Immunotherapy and Targeted Therapy *via* Multi-Omics Analysis

Weimin Zhong[1,2], *Hongbin Zhong*[1,3,4], *Fengling Zhang*[5], *Chaoqun Huang*[2], *Yao Lin*[6]* and *Jiyi Huang*[1,3,4]*

[1]Department of Clinical Medicine, Fujian Medical University, Fuzhou, China, [2]Central Laboratory, the Fifth Hospital of Xiamen, Xiamen, China, [3]Department of Nephrology, Xiang'an Branch, the First Affiliated Hospital of Xiamen University, Xiamen, China, [4]Department of Nephrology, The First Affiliated Hospital of Xiamen University, Xiamen, China, [5]Department of Traditional Chinese Medicine, the Fifth Hospital of Xiamen, Xiamen, China, [6]Central Laboratory at The Second Affiliated Hospital of Fujian Traditional Chinese Medical University, Collaborative Innovation Center for Rehabilitation Technology, Fujian University of Traditional Chinese Medicine, Fuzhou, China

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Anupam Nath Jha,
Tezpur University, India
Rupkatha Mukhopadhyay,
Johns Hopkins Medicine,
United States

***Correspondence:**
Yao Lin
yaolin@fjnu.edu.cn
Jiyi Huang
hjy0602@163.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 22 March 2021
Accepted: 14 June 2021
Published: 25 June 2021

Citation:
Zhong W, Zhong H, Zhang F, Huang C, Lin Y and Huang J (2021) Characterization of Hypoxia-Related Molecular Subtypes in Clear Cell Renal Cell Carcinoma to Aid Immunotherapy and Targeted Therapy via Multi-Omics Analysis. Front. Mol. Biosci. 8:684050. doi: 10.3389/fmolb.2021.684050

Objective: Tumor hypoxia is a key factor in resistance to anti-cancer treatment. Herein, this study aimed to characterize hypoxia-related molecular subtypes and assess their correlations with immunotherapy and targeted therapy in clear cell renal cell carcinoma (ccRCC).

Materials: We comprehensively analyzed copy number variation (CNV), somatic mutation, transcriptome expression profile and clinical information for ccRCC from TCGA and ICGC databases. Based on 98 prognosis-related hypoxia genes, samples were clustered using unsupervized non-negative matrix factorization (NMF) analysis. We characterized the differences between subtypes concerning prognosis, CNV, somatic mutations, pathways, immune cell infiltrations, stromal/immune scores, tumor purity, immune checkpoint inhibitors (ICI), response to immunotherapy and targeted therapy and CXC chemokines. Based on differentially expressed genes (DEGs) between subtypes, a prognostic signature was built by LASSO Cox regression analysis, followed by construction of a nomogram incorporating the signature and clinical features.

Results: Two hypoxia-related molecular subtypes (C1 and C2) were constructed for ccRCC. Differential CNV, somatic mutations and pathways were found between subtypes. C2 exhibited poorer prognosis, higher immune/stromal scores, and lower tumor purity than C1. Furthermore, C2 had more sensitivity to immunotherapy and targeted therapy than C1. The levels of CXCL1/2/3/5/6/8 chemokines in C2 were distinctly higher than in C1. Consistently, DEGs between subtypes were significantly enriched in cytokine-cytokine receptor interaction and immune responses. This subtype-specific signature can independently predict patients' prognosis. Following verification, the nomogram could be utilized for personalized prediction of the survival probability.

Conclusion: Our findings characterized two hypoxia-related molecular subtypes for ccRCC, which can assist in identifying high-risk patients with poor clinical outcomes and patients who can benefit from immunotherapy or targeted therapy.

Keywords: clear cell renal cell carcinoma, hypoxia, molecular subtypes, immunotherapy, targeted therapy

INTRODUCTION

Renal cell carcinoma (RCC) occupies approximately 2% of all adult cancers globally (Siegel et al., 2019). Clear cell RCC (ccRCC) is the main histological subtype of RCC (~80%), with extremely poor prognosis (Siegel et al., 2019). For local ccRCC, surgery is the preferred treatment, while immunotherapy, targeted therapy and chemotherapy have been approved for treatment of advanced or metastatic ccRCC (Chen et al., 2019). However, not all patients can respond to above treatments. Recent genomic research has uncovered a distinct complexity of intra- and inter-tumor heterogeneity in ccRCC, which has contribution to the varying prognosis of patients (Gerlinger et al., 2012; Cancer Genome Atlas Research Network, 2013). It is expected to achieve long-term survival of ccRCC patients by improving the ability to identify high-risk patients and further developing personalized treatment based on multi-omics.

Hypoxia is one of the signs of tumor microenvironment. It has been widely regarded as an active participator for ccRCC progression (Jing et al., 2019). Hypoxia-induced changes in gene expression exert critical effects on various cellular and physiological functions, thereby ultimately limiting the prognosis of patients (Vito et al., 2020). The behavior of tumor cells is highly influenced by their surrounding microenvironment. Under hypoxic conditions, tumor cells have remarkably restored their survival and proliferation (Riera-Domingo et al., 2020). For example, the acidic microenvironment induced by hypoxia can promote chemoresistance by inducing epithelial-mesenchymal transition and stem cell-like phenotypes (Damgaci et al., 2018). Especially, hypoxia can drive immune escape in the tumor microenvironment and hinder the success of immunotherapy (Riera-Domingo et al., 2020). Hence, a better understanding of hypoxia-related molecular characteristics may contribute to the progression of cancer immunotherapy research and provide a theoretical basis for clinical trials to help improve treatment effects (Zhang et al., 2020). In this study, we aimed to comprehensively characterize the hypoxia-related molecular subtypes and their clinical implications for immunotherapy and targeted therapy of ccRCC *via* multi-omics data.

MATERIALS AND METHODS

Hypoxia-Related Genes

The "HALLMARK_HYPOXIA" gene sets were downloaded from The Molecular Signatures Database v7.2 (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb) using Gene Set Enrichment Analysis (GSEA) v4.1.0 software (Subramanian et al., 2005), where there were 200 hypoxia genes that were up-regulated in response to hypoxia (**Supplementary Table 1**).

Data Collection and Preprocessing

Level 3 RNA sequencing (RNA-seq), somatic mutation data, copy number variation (CNV) data and corresponding clinical information (age, gender, grade, stage, survival status and follow-up information) for ccRCC were retrieved from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) or the International Cancer Genome Consortium (ICGC, www.icgc.org). Samples with survival time ≥30 days were retained. Consequently, 512 ccRCC samples from TCGA were enrolled as the training set, while 90 samples from ICGC database were included in the external validation set. The two datasets were integrated into the entire set and batch effects were corrected with the "ComBat" algorithm of sva package (Leek et al., 2012).

Clustering Analysis

Before clustering, univariate cox regression survival analysis was performed to evaluate the correlation between hypoxia genes and overall survival (OS) in TCGA-ccRCC cohort. Consequently, genes with $p < 0.05$ were retained for sample clustering analysis. Then, unsupervized non-negative matrix factorization (NMF) clustering was conducted *via* the NMF package in *R* on the TCGA and ICGC datasets, respectively (Gaujoux and Seoighe, 2010). The *k* value when cophenetic correlation coefficient started to decline was chosen as the optimal number of clusters. Principal components analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were presented to verify the classification performance on the basis of the transcriptome expression profile of above hypoxia-related genes. Kaplan-Meier overall survival (OS) curves were drawn using the survival package in *R*, followed by log-rank test.

Mutation Estimation

Amplification and deletion variations were evaluated using the Genomic Identification of Significant Targets in Cancer (GISTIC) v2.0 by the genePattern software. Furthermore, somatic mutation data were extracted and the mutation frequencies were counted *via* the MutSigCV algorithm.

Gene Set Variation Analysis

The GSVA algorithm was used to probe into the distinct signaling pathways between subtypes on the basis of transcriptomic expression profile (Hänzelmann et al., 2013). The gene set of "c2.cp.kegg.v7.1.symbols" was employed as the reference. The enrichment scores of pathways in each sample were calculated and their differences between subtypes were analyzed using the linear models for microarray data (limma) package (Ritchie et al.,

2015). Differential pathways were screened with the criteria of false discovery rate (FDR) < 0.05 and |log2 fold change (FC)| >0.2.

Cell Type Identification by Estimating Relative Subsets of RNA Transcripts

Using the CIBERSORT algorithm, the infiltration levels of 22 kinds of immune cells were estimated for each ccRCC sample in TCGA database. The differences in the immune infiltration levels between subtypes were calculated *via* the Wilcoxon rank-sum test. Infiltrating immune cells were clustered by hierarchical agglomerative clustering based on Euclidean distance and Ward's linkage.

Estimation of Stromal and Immune Cells in Malignant Tumors Using Expression Data

The levels of infiltrating stromal and immune cells in ccRCC tissues were estimated for each sample based on the gene expression profiles utilizing the ESTIMATE algorithm (Yoshihara et al., 2013). By combining stromal and immune scores, ESTIMATE scores were determined. Tumor purity of each sample was then calculated according to the ESTIMATE scores.

Assessment of Immune Checkpoint Inhibitors, Response to Immune Therapy and Tumor Mutation Burden Between Subtypes

The likehood of response to immunotherapy was assessed by the Tumor Immune Dysfunction and Exclusion (TIDE; http://tide.dfci.harvard.edu/login/) website. TMB was defined as the ratio of total count of variants and the whole length of exons. The differences in the expression levels of ICIs, TIDE scores and TMB levels were compared by the Wilcoxon rank-sum test.

Drug Sensitivity Prediction

The sensitivity of each sample to chemotherapy drugs was predicted by the Genomics of Drug Sensitivity in Cancer (GDSC; https://www.cancerrxgene.org/) database (Yang et al., 2013). The half maximal inhibitory concentration (IC50) was assessed through ride regression utilizing the pRRophetic package in *R*. Furthermore, the predictive accuracy was verified *via* ten-fold cross-verification in the TCGA-ccRCC cohort.

Differential Expression and Functional Annotation Analysis

Differentially expressed genes (DEGs) were filtered between two molecular subtypes *via* the egdeR package with the cutoff of FDR <0.05 and |log2 FC| ≥2. Their underlying functions were predicted through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis *via* the clusterProfiler package in *R* (Yu et al., 2012). The *p*-value was adjusted by Benjamini-Hochberg method. Adjusted $p < 0.05$ was considered significant.

Screening Small Molecule Drugs

The two gene lists of up- and down-regulated tags were uploaded into the Connectivity map (CMap; http://portals.broadinstitute.org/cmap/) database (Lamb et al., 2006). Candidate small molecular drugs were screened according to the enrichment value and permutation *p*-value. CMap mode-of-action (MoA) analysis was exploited to explore potential mechanisms of action.

Establishment of a Signature Based on DEGs in Two Molecular Subtypes

Prognosis-related DEGs with $p < 0.05$ were screened by univariate cox regression survival analysis. The least absolute shrinkage and selection operation (LASSO) Cox regression model was constructed *via* the glmnet package (Friedman et al., 2010). ccRCC patients from TCGA database were divided into high- and low-risk groups in line with the cutoff value of risk scores. Kaplan-Meir curves were portrayed to compare the differences in OS and disease-free survival (DFS) between the two groups *via* the survival package. Time-dependent receiver operating characteristic curves (ROCs) for one-, three- and five-years OS and DFS were conducted for assessment of the predictive power of the risk score using the survivalROC package. Multivariate Cox regression analysis was carried out to assess the independency of the risk score for OS and DFS. A forest plot containing the hazard ratio (HR) and 95% confidence interval (CI) of each variable was then drawn *via* survminer package.

Construction of a Nomogram Model

Clinical factors and the risk score were incorporated into a nomogram for predicting OS and DFS using the rms package in *R*. The scores of variables were given based on their regression coefficients. For each patient, a total score was calculated by adding up the corresponding individual scores of all variables. Then, using conversion function, the probability of outcome of each patient was calculated. The predictive efficacy of the nomogram was investigated by calibration plots.

Statistical Analysis

All statistical analysis was achieved *via R* language v4.0.2 (https://www.r-project.org/). Comparisons between two groups were presented *via* Wilcoxon rank-sum test. A two-tailed *p*-value <0.05 was considered statistically significant.

RESULTS

Characterization of Two Hypoxia-Related Molecular Subtypes with Distinct Clinical Implications for ccRCC

200 hypoxia genes were retrieved from the list "HALLMARK_HYPOXIA" gene set. In the TCGA-ccRCC (n = 512) cohort, a total of 98 genes were associated with ccRCC prognosis (all $p < 0.05$), while the other genes could not impact ccRCC prognosis (**Supplementary Table 2**). Based on the expression profiles of prognostic hypoxia genes, ccRCC samples from TCGA were clustered *via* the NMF package. The optimal value

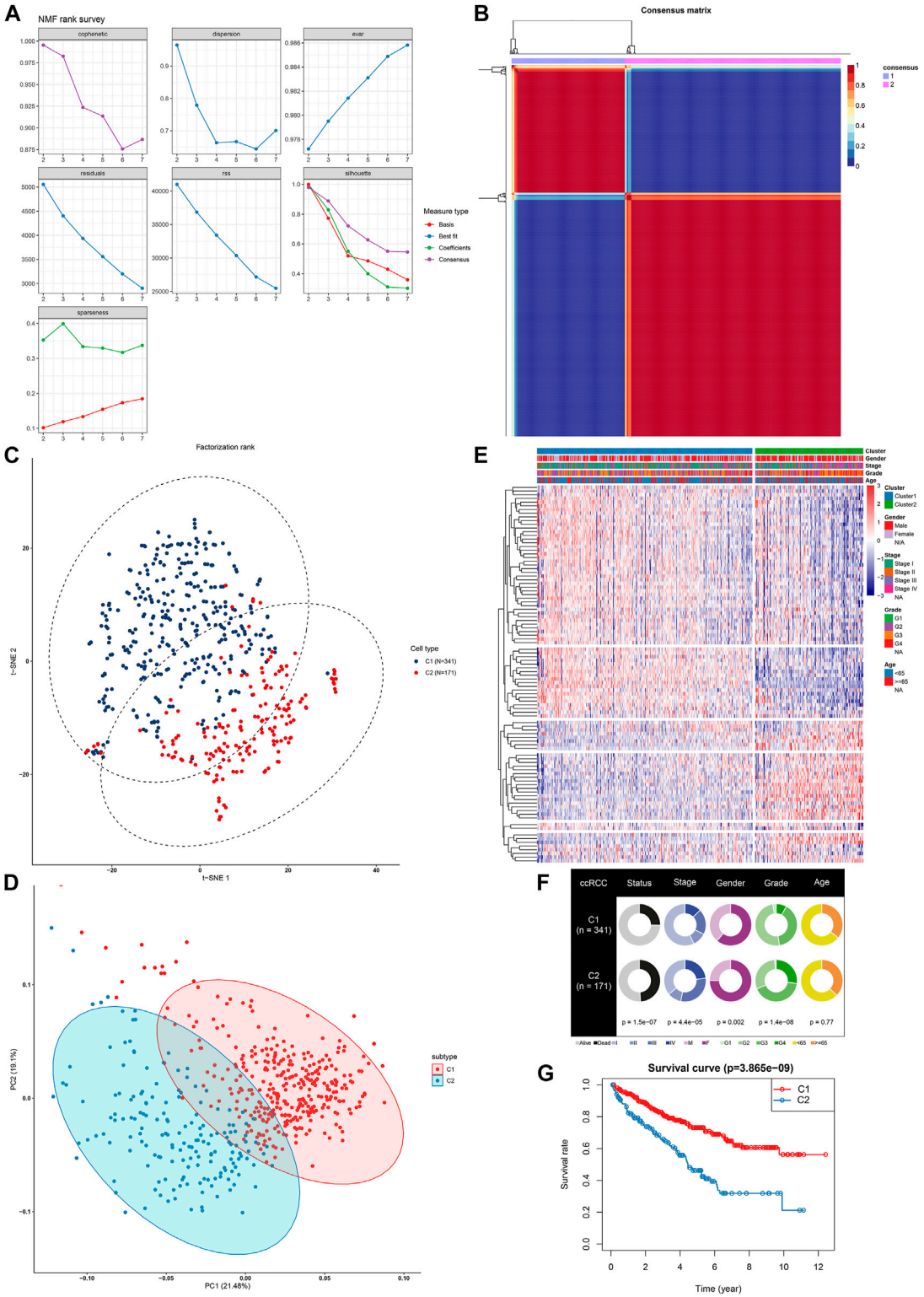


FIGURE 1 | NMF identifies two distinct hypoxia-related molecular subtypes for ccRCC in TCGA-ccRCC dataset **(A)** Factorization rank for k = 2–7 **(B)** The heat map of the consensus matrix when the consensus clustering k = 2. The value range is 0–1. The columns and rows are sorted through hierarchical clustering according to the Euclidean distance of the average link **(C)** The PCA and **(D)** t-SNE scatter plots are in support of the classification into two ccRCC molecular subtypes based on the gene expression profiles. The colors are indicative of samples from two molecular subtypes **(E)** The heatmap visualizing the expression patterns of hypoxia genes in the two subtypes. Samples are clustered according to different clinical features **(F)** Correlation between subtypes and clinical features **(G)** Kaplan-Meier OS curves for the two clusters in TCGA-ccRCC dataset. The assessment of difference was achieved by log-rank test.

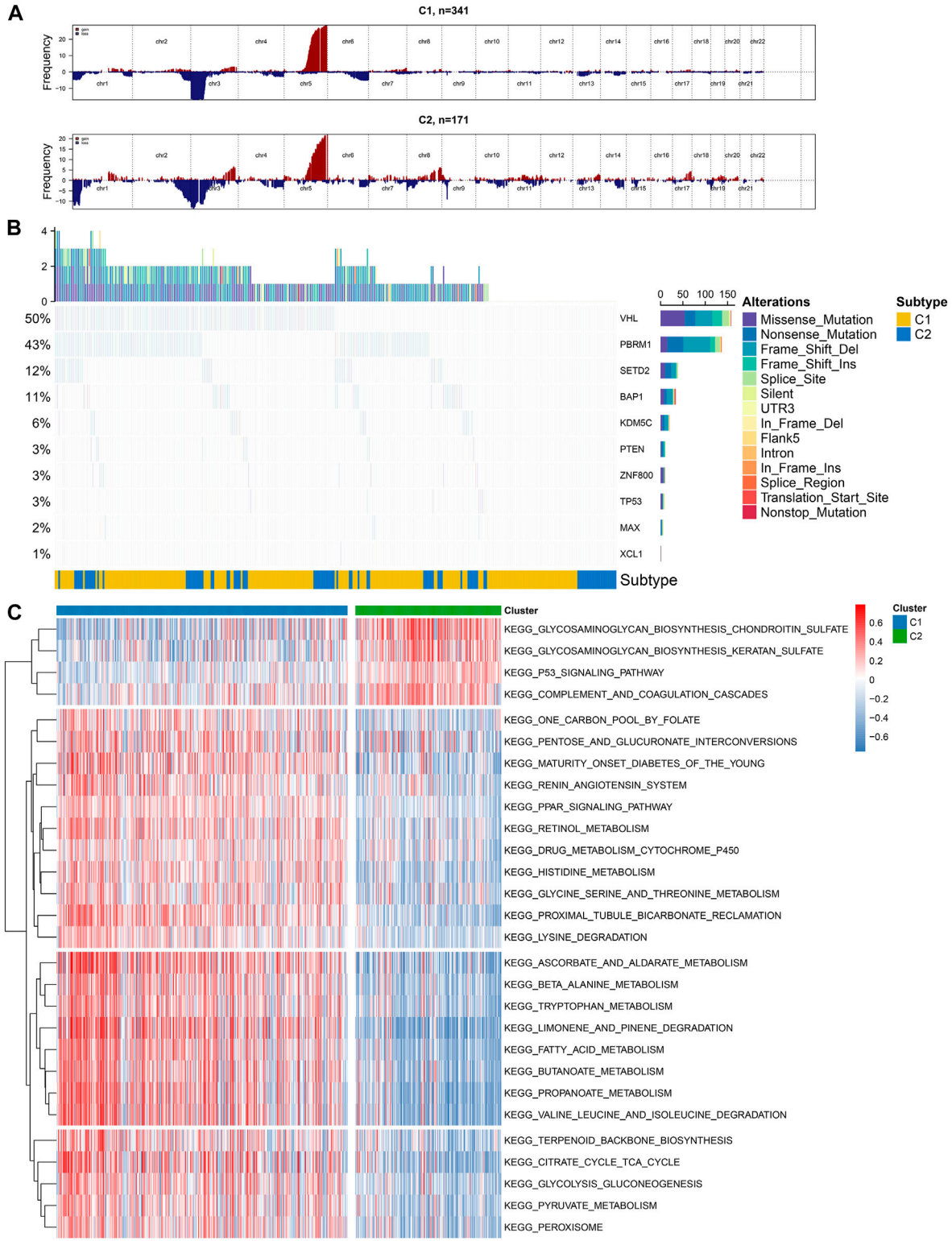


FIGURE 2 | Illustration of copy number and somatic variations and subtype-specific signaling pathways in two hypoxia-related molecular subtypes **(A)** The landscape of copy number alterations in C1 and C2 subtypes **(B)** Oncoplot visualizing the somatic landscape of ccRCC samples in the two subtypes. The top ten genes are ranked on the grounds of the mutation frequency. Different mutation types are annotated by different colors on the right **(C)** Heatmap showing subtype-specific signaling pathways.

of k was determined on the grounds of cophenetic correlation coefficient. When starting from k = 2, cophenetic correlation coefficient started to decrease (**Figure 1A**). The heatmap intuitively showed the consensus matrix when k = 2 (**Figure 1B**). Hence, ccRCC samples were clustered into two molecular subtypes C1 (n = 341) and C2 (171). The PCA (**Figure 1C**) and t-SNE (**Figure 1D**) supported the classification into two subtypes. As depicted in heatmap, there was a distinct difference in expression patterns of hypoxia genes between subtypes (**Figure 1E**). Furthermore, we found the significant differences in status (p = 1.5e-07), stage (p = 4.4e-05), gender (p = 0.002) and grade (p = 1.4e-08) between subtypes (**Figure 1F**). The significant prognostic difference was investigated in TCGA-ccRCC cohort, with shorter OS time in C2 than in C1 (p = 3.865e-09; **Figure 1G**). This classification was confirmed in the ICGC dataset (**Supplementary Figure 1A–E**). However, due to small sample size, the prognosis of patients was not significantly different between subtypes (**Supplementary Figure 1F**). Therefore, we integrated the samples from the TCGA and ICGC datasets into the entire set after removing batch effects (**Supplementary Figure 2A**). In the entire set, the two molecular subtypes with distinct prognosis and clinicopathological characteristics were confirmed (**Supplementary Figure 2B-F**).

Differential CNV and Somatic Variation Landscape and Subtype-specific Signaling Pathways Between Subtypes

We visualized the mutation frequencies of CNV in ccRCC samples from two subtypes (**Figure 2A**). The chromosome 5 occurred the most frequent amplification both in the C1 and C2 subtypes. Meanwhile, the chromosome 2 and 3 harbored the most frequent deletion sites. The frequency of amplification and deletion in C2 was more common than that in C1. The MutSigCV algorithm was applied to compare the frequency of somatic mutation between C1 and C2. Genetic alterations of ccRCC mainly consist of those that control cellular oxygen induction (such as VHL) as well as maintaining chromatin states (such as PBRM1). Consistently, among ccRCC samples, VHL exhibited the most frequent mutations (50%), followed by PBRM1 (43%) and SETD2 (12%; **Figure 2B**). We further probed into subtype-specific signaling pathways by GSVA (**Supplementary Table 3**). As depicted in **Figure 2C**, p53 signaling pathway was down-regulated C1 compared to C2, and metabolism-related pathways were up-regulated in C1 than C2.

More Sensitivity to Immunotherapy for Molecular Subtype 2

Immunotherapy has been approved for the treatment of ccRCC. However, which group of patients responds to immunotherapy is still unknown. Here, we firstly assessed the differential sensitivity to immunotherapy between the two hypoxia-related molecular subtypes. The infiltration levels of 22 kinds of immune cells for each sample were detected utilizing the CIBERSORT algorithm. As a result, C2 displayed the higher infiltration levels of T cells regulatory (Tregs; $p < 0.01$), macrophages M0 ($p < 0.001$), mast cells activated ($p < 0.05$), plasma cells ($p < 0.001$), T cells CD4 memory activated ($p < 0.001$), neutrophils ($p < 0.001$) compared to C1 (**Figure 3A**). Meanwhile, C1 exhibited distinctly higher levels of dendritic cells resting ($p < 0.001$), macrophages M1 ($p < 0.001$), mast cells resting ($p < 0.01$), monocytes ($p < 0.001$), T cells $CD8^+$ ($p < 0.05$) in comparison to C2. In **Figure 3B**, these immune cells were clustered into four cell clusters by hierarchical agglomerative clustering based on Euclidean distance and Ward's linkage. There was a complex interaction network between different immune cells, indicating the complexity of tumor immune microenvironment. For example, the infiltration levels of macrophages M2 were positively correlated with B cells naïve in ccRCC tissues. The patterns of stromal scores, immune scores, ESTIMATE scores and tumor purity in each ccRCC sample were evaluated *via* ESTIMATE algorithm. The samples in C2 subtype had the relatively high levels of stroma, immune and ESTIMATE scores in comparison to C1 (all $p < 0.001$; **Figure 3C**). Furthermore, we investigated the lower levels of tumor purity in C2 than in C1 ($p < 0.001$). These suggested that C2 was more likely to experience a worse prognosis than C1. ICIs have been used for the first-line therapy of metastatic ccRCC (Shah et al., 2019). Nevertheless, not all patients may benefit from it. The patients' response to immunotherapy was predicted by the TIDE algorithm. Higher expression levels of CD274 mRNA ($p < 0.001$) were found in C1 compared to C2 (**Figure 3D**). Meanwhile, LAG3 (p = 0.003), TIGIT (p = 0.035), IDO1 (p = 0.005) and CTLA4 (p = 0.05) mRNAs displayed higher expression levels in C2 than C1 (**Figure 3D**). Moreover, C2 displayed higher TIDE levels than C1 (**Figure 3E**; p = 2.4e-08). As such, our data showed that C2 was more likely to respond to immunotherapy compared to C1 (**Figure 3E**; p = 2.4e-08). High TMB usually indicates poor clinical outcomes and is a powerful predictor for immunotherapy response in ccRCC (Huang et al., 2020). However, there was no significant difference in TMB between subtypes (**Figure 3F**).

Evaluation of the Expression Levels of CXC Chemokines in Two ccRCC Molecular Subtypes

Herein, we assessed the expression levels of CXC chemokines in ccRCC samples between C1 and C2 (Zeng et al., 2019). As a result, C2 exhibited the higher expression levels of CXCL1 (**Figure 4A**), CXCL2 (**Figure 4B**), CXCL3 (**Figure 4C**), CXCL5 (**Figure 4D**), CXCL6 (**Figure 4E**) and CXCL8 (**Figure 4F**) in comparison to C1 (all $p < 0.001$). Among them, a previous study has showed that low expression of CXCL1/2/3/5 was in relationship with a better prognosis for RCC patients, indicating that these chemokines could contribute to poor clinical outcomes for patients in C2 (Zeng et al., 2019).

Differential Putative Chemotherapeutic Response Between Molecular Subtypes

Drug resistance has become a major challenge in chemotherapy, involving various mechanisms. Hypoxia, as a key factor, affects cell expression programs and induces treatment resistance (Jing et al., 2019). Herein, GDSC database was employed to assess the differences in the sensitivity between the two hypoxia-related molecular subtypes to eight common chemotherapy drugs

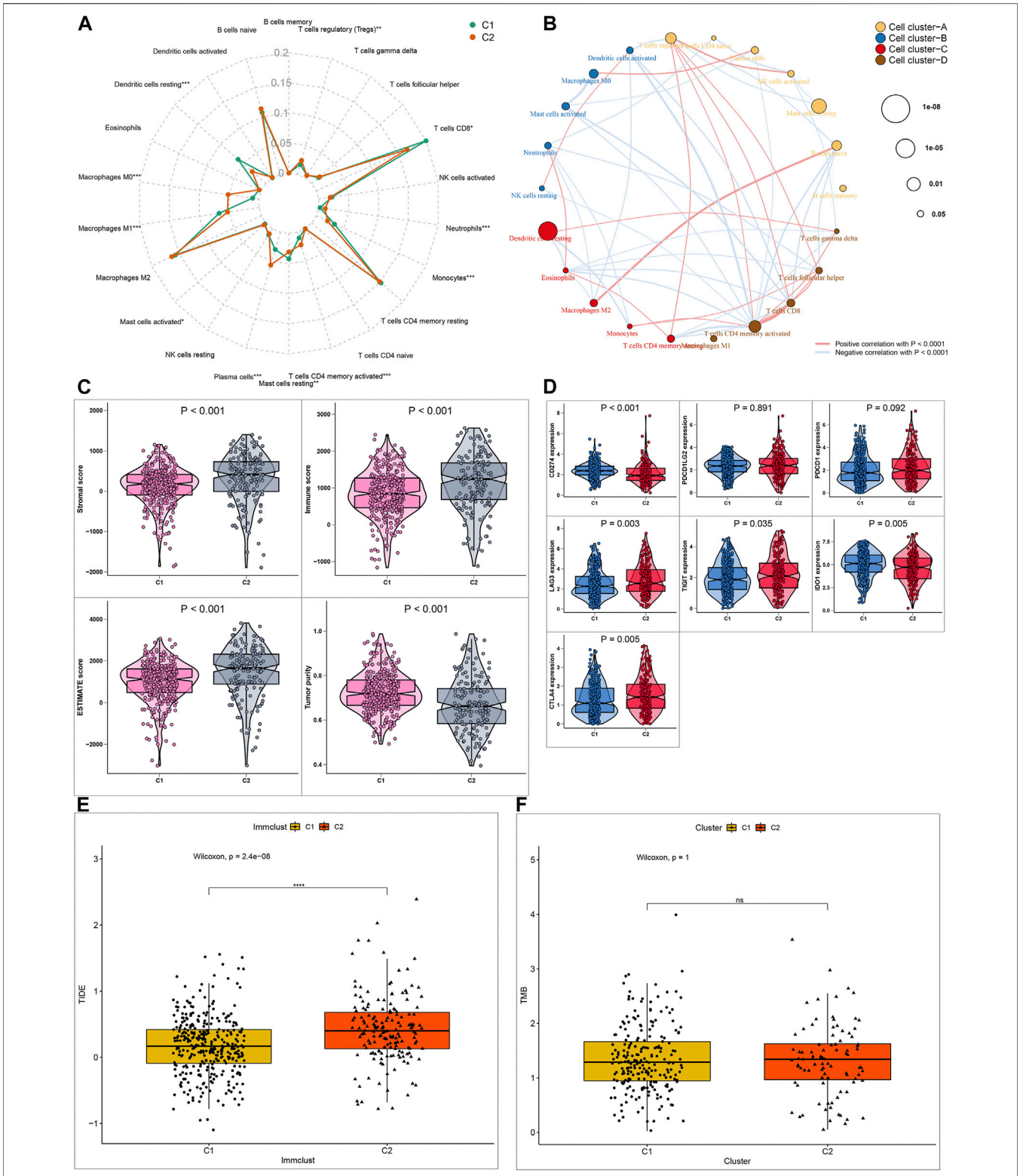


FIGURE 3 | Differential sensitivity to immunotherapy between two ccRCC molecular subtypes **(A)** Radar map showing the differences in proportions of 22 immune cell types between C1 and C2 molecular subtypes **(B)** A network of the correlation between the levels of 22 kinds of tumor-infiltrating immune cells in ccRCC samples. The size of bubble is inversely proportional to *p*-value **(C)** Patterns of stromal cell scores, immune cell scores, ESTIMATE scores and tumor purity between subtypes **(D)** Expression levels of immune checkpoint markers in the two subtypes **(E, F)** Box plots showing the correlation between TIDE/TMB levels and molecular subtypes. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; ns: no statistical significance.

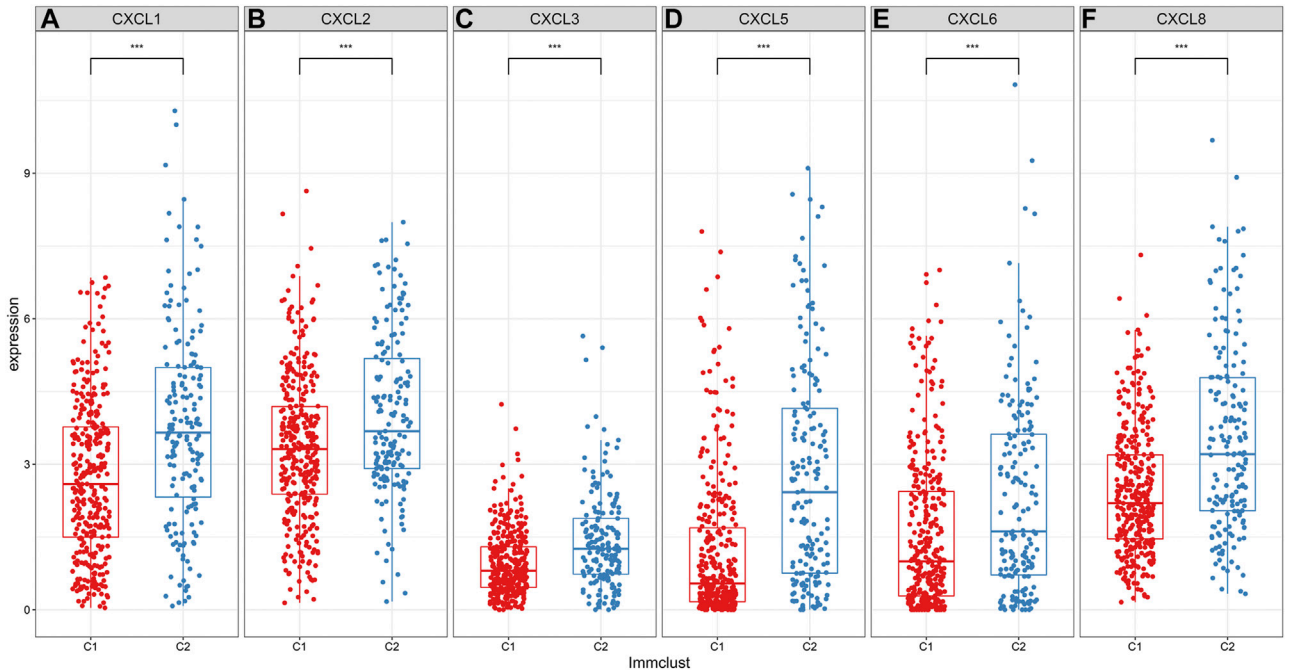


FIGURE 4 | Differences in expression patterns of CXC chemokines between two ccRCC molecular subtypes. As depicted in the box plots, the expression levels of **(A)** CXCL1 **(B)** CXCL2 **(C)** CXCL3 **(D)** CXCL5 **(E)** CXCL6 and **(F)** CXCL8 are visualized in ccRCC samples between C1 and C2. ***$p < 0.001$.

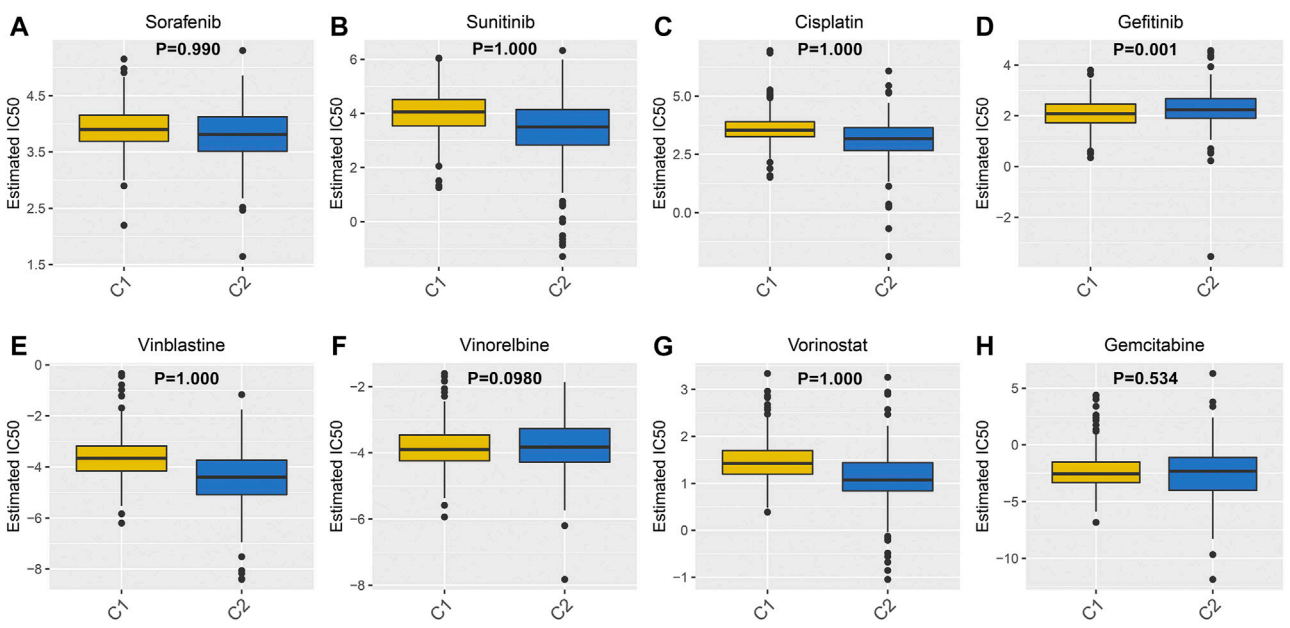


FIGURE 5 | Differences in sensitivity to chemotherapy drugs between two ccRCC molecular subtypes. The box plots depicting the estimated IC50 values for **(A)** Sorafenib **(B)** sunitinib **(C)** Cisplatin **(D)** gefitinib **(E)** Vinblastine **(F)** Vinorelbine **(G)** Vorinostat and **(H)** Gemcitabine in ccRCC samples from the two molecular subtypes.

including Sorafenib (**Figure 5A**), sunitinib (**Figure 5B**), Cisplatin (**Figure 5C**), gefitinib (**Figure 5D**), Vinblastine (**Figure 5E**), Vinorelbine (**Figure 5F**), Vorinostat (**Figure 5G**) and Gemcitabine (**Figure 5H**). Drug response was defined based on IC50 values. The data suggested that C2 subtype was more sensitive to most of chemotherapy drugs such as Sorafenib, Sunitinib, Cisplatin, Vinblastine and Vorinostat compared to C1 subtype, indicating that patients in C2 subtype were more

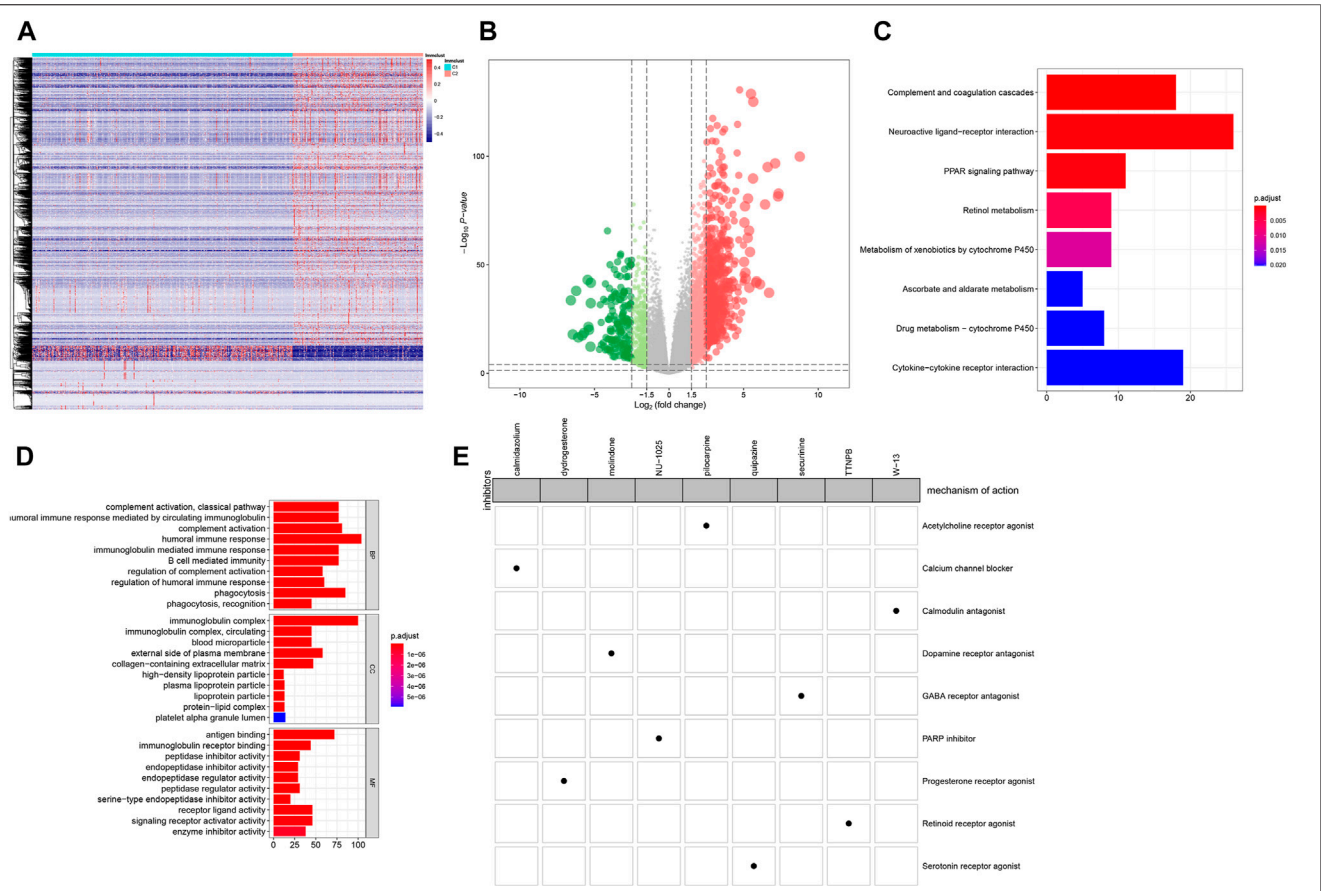


FIGURE 6 | CMap analysis identifies candidate small molecular inhibitors for ccRCC **(A)** Heatmap depicting all DEGs between C1 (blue) and C2 (red) **(B)** Volcano plots up-regulated genes (red bubbles) and down-regulated genes (green bubbles) in C1 compared to C2 **(C)** Bar plots of the top ten enriched KEGG signaling pathways **(D)** Bar plots of the top ten GO function annotation analysis results including biological processes (BP), cellular component (CC) and molecular function (MF) categories. Red suggests high enrichment and blue suggests low enrichment **(E)** Heatmap demonstrating each inhibitor (perturbagen) and its shared mechanisms of action (rows) *via* the CMap database.

likely to benefit from above chemotherapy drugs. Meanwhile, C1 subtype had the higher sensitivity to Gefitinib, Gemcitabine and Vinorelbine than C2 subtype, indicating that patients in C1 subtype might respond to these chemotherapy drugs.

CMap Analysis Identifies Candidate Inhibitors for ccRCC

1,472 DEGs were identified between C1 and C2 (**Figures 6A,B**). Among them, there were 1,203 up- and 269 down-regulated genes in C2 compared to C1 (**Supplementary Table 4**). As shown in KEGG enrichment analysis results, ccRCC-related signaling pathways such as complement and coagulation cascades, neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction and PPAR signaling pathway were significantly enriched by these DEGs (**Figure 6C**). GO function annotation analysis revealed that these DEGs could possess immune-related functions (**Figure 6D**). Through the CMap, we screened out nine small molecule inhibitors (pilocarpine, quipazine, calmidazolium, dydrogesterone, securinine, molindone, W-13, TTNPB and NU-1025). Among them, pilocarpine (enrichment = −0.865 and p = 0.00062) and quipazine (enrichment = −0.704 and p = 0.01578) could become candidate small molecule drugs for all ccRCC patients (**Figure 6E**). According to MoA analysis, nine mechanisms of actions (acetylcholine receptor agonist, serotonin receptor agonist, calcium channel blocker, progesterone receptor agonist, GABA receptor antagonist, dopamine receptor antagonist, calmodulin antagonist, retinoid receptor agonist and PARP inhibitor) shared the above small molecule inhibitors, indicating that nine small molecule inhibitors might suppress ccRCC progression through mediating these mechanisms of actions.

Development of a Subtype-specific Prognostic Five-Gene Signature for ccRCC

Totally, 902 prognosis-related DEGs between subtypes were identified for ccRCC. Under LASSO Cox regression analysis (**Figures 7A,B**), a five-gene signature was constructed, composed of COL7A1, ZIC2, AC116021.1, AC112715.1 and OTX1. The risk score of each sample was calculated and all ccRCC patients were separated into high- and low groups in accordance with the cutoff values of risk scores. The higher the risk score, the greater the number of patients with dead (**Figure 7C**)

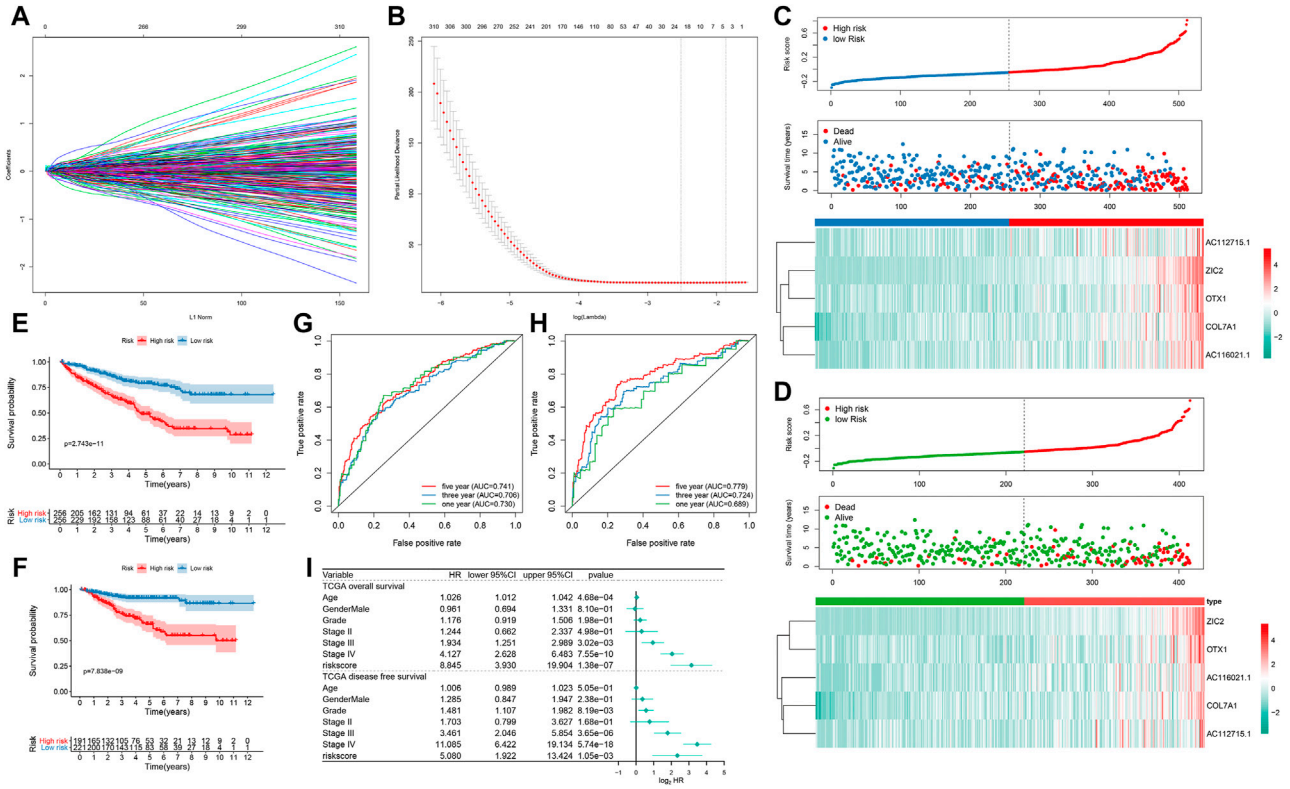


FIGURE 7 | Development of a prognostic five-gene signature for ccRCC in TCGA dataset **(A)** 20-time cross-validation for tuning parameter selection in the LASSO Cox model **(B)** Plots of the LASSO coefficients **(C)** The risk score rank (up), distribution of survival status (alive or dead; middle) and expression patterns of five genes in high- and low-risk groups **(D)** The risk score rank (up), distribution of survival status (diseased or disease-free; middle) and expression patterns of five genes (down) in high- and low-risk groups **(E, F)** Kaplan-Meier OS and DFS curve for high- and low-risk groups **(G)** Time-dependent ROC curves for one-, three- and five-years OS time **(H)** Time-dependent ROC curves for one-, three- and five-years DFS time **(I)** Forest plots showing the multivariate Cox regression analyses results of the risk score and clinical factors with OS and DFS.

or disease status (**Figure 7D**). High-risk group had the higher expression levels of these genes than low-risk group (**Figures 7C,D**). Patients with high-risk score exhibited a poorer OS (p = 2.743e-11; **Figure 7E**) and DFS (p = 7.838e-09; **Figure 7F**). As shown in ROC curves, AUCs for one-, three- and five-years OS were 0.730, 0.706, and 0.741, suggesting the well performance of the risk score for prediction of OS (**Figure 7G**). Also, AUCs for one-, three- and five-years DFS were 0.689, 0.724, and 0.779, confirming its predictive efficacy for DFS (**Figure 7H**). Our multivariate cox regression analysis demonstrated that this signature could independently predict ccRCC patients' OS and DFS (**Figure 7I**).

A Nomogram Integrating Subtype-specific Signature and Clinical Factors Improves Predictive Power for ccRCC Prognosis

We constructed a nomogram by combining the five-gene signature and clinical factors including age, grade, gender, and stage for predicting ccRCC patients' OS (**Figure 8A**) and DFS (**Figure 8B**). We further evaluated whether the integration of the five-gene signature and clinical factors could boost the predictive efficiency for ccRCC prognosis in TCGA dataset. Calibration plots confirmed that the nomogram-predicted probabilities of one- (**Figure 8C**), three- (**Figure 8D**) and five-years (**Figure 8E**) OS had high consistency with the actual survival. Moreover, the nomogram-predicted probabilities of one- (**Figure 8F**), three- (**Figure 8G**) and five-years (**Figure 8H**) DFS was close to the actual survival. Collectively, the nomogram integrating the five-gene signature, age, grade, gender, and stage could enhance the predictive power of ccRCC patients' prognosis.

DISCUSSION

Emerging first-line treatment options such as targeted drugs and immunotherapy have significantly improved the prognosis of ccRCC patients with high risk, for whom chemoradiotherapy has shown limited efficacy (Atkins and Tannir, 2018). It has been widely recognized the differences in response to therapy due to the molecular and histologic heterogeneity of ccRCC (Luo et al., 2019). In this study, we characterized two hypoxia-related molecular subtypes for ccRCC with distinct clinical outcomes and response to immunotherapy and targeted therapy based on multi-omics analysis.

Hypoxia is a key feature of the tumor microenvironment, driving tumor aggressiveness (Balamurugan, 2016). To adapt to hypoxia, the expression of hypoxia-related genes changes accordingly. Based on prognosis-related hypoxia genes, we characterized two molecular

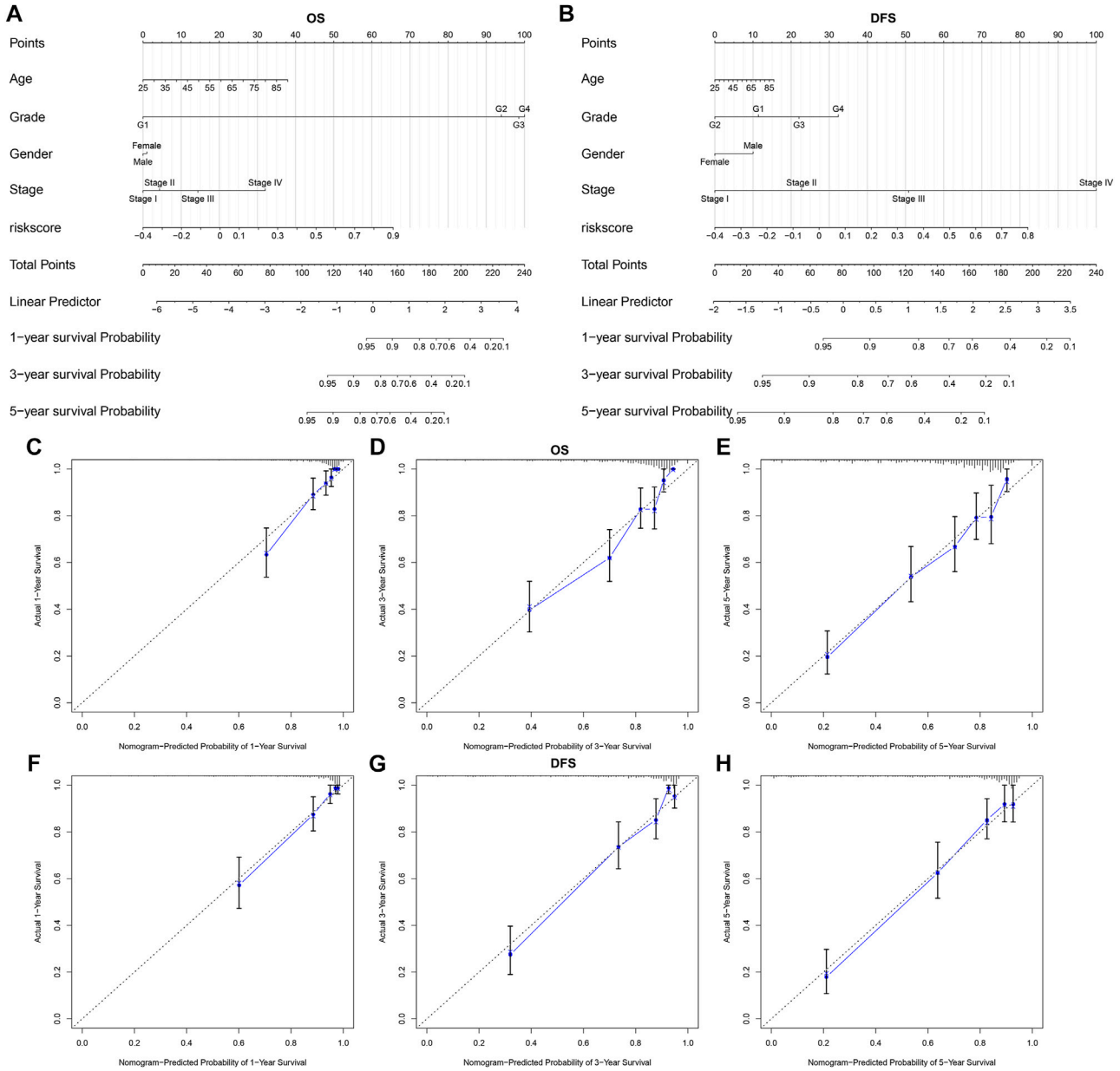


FIGURE 8 | A nomogram incorporating subtype-specific signature and clinical factors improves predictive efficacy for ccRCC prognosis **(A, B)** Construction of a nomogram combining the subtype-specific signature and clinical features for prediction of OS and DFS. Calibration plots displayed the actual and nomogram-predicted probability of one-, three- and five-years OS **(C–E)** and DFS **(F–H)**.

subtypes with distinct molecular subtypes in TCGA and ICGC databases. The hypoxia-related classifier may become a practical and reliable predictive tool, which could complement the current staging system for predicting ccRCC prognosis. The differences in survival status, stage, gender, and grade did not reduce the accuracy of the classifier in predicting patients' prognosis. Specifically, it has been acknowledged that male patients exhibit more aggressive characteristics as well as poorer OS compared to females (Brannon et al., 2012). Differential somatic mutations and CNVs were detected between subtypes. Consistent with previous study, VHL (50%) and PBRM1 (43%) mutations commonly occur in ccRCC (Carril-Ajuria et al., 2019). The loss of VHL tumor suppressor gene is the most common genetic feature of ccRCC, which improves the expression of target genes of hypoxia-inducible factors (HIFs), thereby affecting metabolism and signal transduction for ccRCC cells (Zhang et al., 2018).

Despite various gene mutations gain the incidence of ccRCC, the tumor microenvironment has a critical influence on tumor development and immune response. In the tumor immune microenvironment, there were distinct differences immune cell

infiltrations between subtypes. Tumor-infiltrating immune cells are linked to clinical outcomes as well as response to immunotherapy. Our characterized subtypes were associated with immune infiltration patterns in ccRCC. Particularly, ccRCC is the tumor type with the highest infiltration levels of T cells (Şenbabaoğlu et al., 2016). There were distinct differences in the infiltration of T cell subpopulations between subtypes. Immune status affects ccRCC patients' clinical outcomes. Functional enrichment analysis revealed that DEGs between subtypes could be involved in immune response. Compared to C1, C2 had higher stromal/immune scores and lower tumor purity. Furthermore, high stromal/immune scores and low tumor purity of ccRCC patients have been found to be significantly associated with poor prognosis (Xu et al., 2019). Targeted therapeutics, such as VEGF receptors and mTOR inhibitors, can distinctly prolong the survival time of metastatic ccRCC patients (Xu et al., 2019). Nevertheless, most of patients do not have targetable mutations. Immune checkpoint targets provide another promising treatment strategy. However, the hypoxic microenvironment of tumors can reduce immune activity. Here, we found that C2 subtype had higher levels of ICIs than C1. For example, LAG3+ T cells is a sign of T cell exhaustion that is a key factor for immunosuppressive properties and is associated with advanced ccRCC (Wang et al., 2019). Based on the TIDE algorithm, it was estimated that C2 possessed higher potential response of immune-checkpoint blockade (ICB) therapy. Thus, ICB therapy may be efficacious for C2 subtype of ccRCC patients.

Prognostic biomarkers related to the tumor immune microenvironment may provide promising prospects for identifying novel molecular targets and improving patients' clinical outcomes undergoing immunotherapy (Burger and Kipps, 2006). DEGs between subtypes were significantly enriched in chemokine-chemokine receptor interaction. Low expression of CXCL1/2/3/5 chemokines exhibits better clinical outcomes in RCC (Zeng et al., 2019). Our data showed that the expression of these chemokines in subtype C1 was significantly lower than that of subtype C2, indicating that chemokines could promote tumor escape of ccRCC of C2 subtype, thereby leading to poorer prognosis.

Treatment based on individual tumor characteristics provides the possibility to improve the different clinical outcomes of patients due to tumor heterogeneity in ccRCC (Hu et al., 2020). Changes in the cancer genome in response to hypoxia markedly affect the response to anticancer therapies (Ye et al., 2019). It is reasonable to predict the treatment response to chemotherapy, which can reduce the cost of treatment and improve the prognosis of patients. This study demonstrated that C2 subtype exhibited higher sensitivity to most of chemotherapeutic drugs being used [such as sorafenib and sunitinib have been approved for treating metastatic RCC (Hsieh et al., 2017)] or developed than C1 subtype, indicating that patients in C2 could be more suitable for above therapies, which can provide an available strategy to select patients who benefit from a particular therapy.

Signatures based on gene expression have not yet been incorporated into routine clinical practice for ccRCC. Compared with the traditional method using gene expression levels, LASSO algorithm eliminates the requirement for data preprocessing, which has been proven to produce reliable results including cancer classification (Li et al., 2017). This study constructed a subtype-specific signature using LASSO Cox regression analysis. Following validation, this signature could robustly and independently predict OS and DFS of ccRCC patients. Furthermore, we constructed the nomogram combining the signature and other clinical factors. The prediction system can guide the establishment of personalized examination procedures for ccRCC patients and boost the effective use of medical resources.

However, this study is based on the retrospective design. This classifier system will be verified in our future multi-center prospective research. In conclusion, the two hypoxia-related molecular subtypes as well as subtype-specific nomogram could be utilized to guide the current clinical application to maximize patients' benefit from immunotherapy or targeted therapy.

CONCLUSION

Collectively, we constructed two hypoxia-related molecular subtypes with distinct prognosis for ccRCC, which could assist manage risk assessment and provide valuable insights for the immunotherapy and targeted therapy strategies of ccRCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA, http://cancergenome.nih.gov/</b&gt.

AUTHOR CONTRIBUTIONS

YL and JH designed the study, FZ and CH collected the clinical information and gene expression data. WZ and HZ analysis data. WZ wrote the manuscript.

FUNDING

This study was supported by the Health Science Research Personnel Training Program of Fujian Province (2017-CXB-22), Fujian Science and Technology Plan Guiding Projects (2019D026) and Xiamen Medical Advantage Subspecialty Vascular Access Construction Fund (2018 296). Moreover, we also thank the support of the research program fund of the fifth hospital of Xiamen.

ACKNOWLEDGMENTS

We appreciated TCGA and ICGC database for providing the original study data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.684050/full#supplementary-material

REFERENCES

Atkins, M. B., and Tannir, N. M. (2018). Current and Emerging Therapies for First-Line Treatment of Metastatic clear Cell Renal Cell Carcinoma. *Cancer Treat. Rev.* 70, 127–137. doi:10.1016/j.ctrv.2018.07.009

Balamurugan, K. (2016). HIF-1 at the Crossroads of Hypoxia, Inflammation, and Cancer. *Int. J. Cancer* 138 (5), 1058–1066. doi:10.1002/ijc.29519

Brannon, A. R., Haake, S. M., Hacker, K. E., Pruthi, R. S., Wallen, E. M., Nielsen, M. E., et al. (2012). Meta-analysis of clear Cell Renal Cell Carcinoma Gene Expression Defines a Variant Subgroup and Identifies Gender Influences on Tumor Biology. *Eur. Urol.* 61 (2), 258–268. doi:10.1016/j.eururo.2011.10.007

Burger, J. A., and Kipps, T. J. (2006). CXCR4: a Key Receptor in the Crosstalk between Tumor Cells and Their Microenvironment. *Blood* 107 (5), 1761–1767. doi:10.1182/blood-2005-08-3182

Cancer Genome Atlas Research Network (2013). Comprehensive Molecular Characterization of clear Cell Renal Cell Carcinoma. *Nature* 499 (7456), 43–49. doi:10.1038/nature12222

Carril-Ajuria, L., Santos, M., Roldán-Romero, J. M., Rodriguez-Antona, C., and de Velasco, G. (2019). Prognostic and Predictive Value of PBRM1 in Clear Cell Renal Cell Carcinoma. *Cancers* 12 (1), 16. doi:10.3390/cancers12010016

Chen, V. J., Hernandez-Meza, G., Agrawal, P., Zhang, C. A., Xie, L., Gong, C. L., et al. (2019). Time on Therapy for at Least Three Months Correlates with Overall Survival in Metastatic Renal Cell Carcinoma. *Cancers* 11 (7), 1000. doi:10.3390/cancers11071000

Damgaci, S., Ibrahim-Hashim, A., Enriquez-Navas, P. M., Pilon-Thomas, S., Guvenis, A., and Gillies, R. J. (2018). Hypoxia and Acidosis: Immune Suppressors and Therapeutic Targets. *Immunology* 154 (3), 354–362. doi:10.1111/imm.12917

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

Gaujoux, R., and Seoighe, C. (2010). A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinformatics* 11, 367. doi:10.1186/1471-2105-11-367

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 366 (10), 883–892. doi:10.1056/NEJMoa1113205

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

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

Hu, J., Chen, Z., Bao, L., Zhou, L., Hou, Y., Liu, L., et al. (2020). Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, Which Results in Different Clinical Outcomes. *Mol. Ther.* 28 (7), 1658–1672. doi:10.1016/j.ymthe.2020.04.023

Huang, J., Li, Z., Fu, L., Lin, D., Wang, C., Wang, X., et al. (2020). RETRACTED ARTICLE: Comprehensive Characterization of Tumor Mutation burden in clear Cell Renal Cell Carcinoma Based on the Three Independent Cohorts. *J. Cancer Res. Clin. Oncol.* 147, 1745. doi:10.1007/s00432-020-03299-3

Jing, X., Yang, F., Shao, C., Wei, K., Xie, M., Shen, H., et al. (2019). Role of Hypoxia in Cancer Therapy by Regulating the Tumor Microenvironment. *Mol. Cancer* 18 (1), 157. doi:10.1186/s12943-019-1089-9

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* 313 (5795), 1929–1935. doi:10.1126/science.1132939

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, B., Cui, Y., Diehn, M., and Li, R. (2017). Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-small Cell Lung Cancer. *JAMA Oncol.* 3 (11), 1529–1537. doi:10.1001/jamaoncol.2017.1609

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 (16), 6029–6052. doi:10.18632/aging.102161

Riera-Domingo, C., Audigé, A., Granja, S., Cheng, W.-C., Ho, P.-C., Baltazar, F., et al. (2020). Immunity, Hypoxia, and Metabolism-the Ménage à Trois of Cancer: Implications for Immunotherapy. *Physiol. Rev.* 100 (1), 1–102. doi:10.1152/physrev.00018.2019

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

Şenbabaoğlu, Y., Gejman, R. S., Winer, A. G., Liu, M., Van Allen, E. M., de Velasco, G., et al. (2016). Tumor Immune Microenvironment Characterization in clear Cell Renal Cell Carcinoma Identifies Prognostic and Immunotherapeutically Relevant Messenger RNA Signatures. *Genome Biol.* 17 (1), 231. doi:10.1186/s13059-016-1092-z

Shah, A. Y., Kotecha, R. R., Lemke, E. A., Chandramohan, A., Chaim, J. L., Msaouel, P., et al. (2019). Outcomes of Patients with Metastatic clear-cell Renal Cell Carcinoma Treated with Second-Line VEGFR-TKI after First-Line Immune Checkpoint Inhibitors. *Eur. J. Cancer* 114, 67–75. doi:10.1016/j.ejca.2019.04.003

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

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

Vito, A., El-Sayes, N., and Mossman, K. (2020). Hypoxia-Driven Immune Escape in the Tumor Microenvironment. *Cells* 9 (4), 992. doi:10.3390/cells9040992

Wang, Q., Zhang, J., Tu, H., Liang, D., Chang, D. W., Ye, Y., et al. (2019). Soluble Immune Checkpoint-Related Proteins as Predictors of Tumor Recurrence, Survival, and T Cell Phenotypes in clear Cell Renal Cell Carcinoma Patients. *J. Immunotherapy Cancer* 7 (1), 334. doi:10.1186/s40425-019-0810-y

Xu, W.-H., Xu, Y., Wang, J., Wan, F.-N., Wang, H.-K., Cao, D.-L., et al. (2019). Prognostic Value and Immune Infiltration of Novel Signatures in clear Cell Renal Cell Carcinoma Microenvironment. *Aging* 11 (17), 6999–7020. doi:10.18632/aging.102233

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 (Database issue), D955–D961. doi:10.1093/nar/gks1111

Ye, Y., Hu, Q., Chen, H., Liang, K., Yuan, Y., Xiang, Y., et al. (2019). Characterization of Hypoxia-Associated Molecular Features to Aid Hypoxia-Targeted Therapy. *Nat. Metab.* 1 (4), 431–444. doi:10.1038/s42255-019-0045-8

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, 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

Zeng, Q., Sun, S., Li, Y., Li, X., Li, Z., and Liang, H. (2019). Identification of Therapeutic Targets and Prognostic Biomarkers Among CXC Chemokines in the Renal Cell Carcinoma Microenvironment. *Front. Oncol.* 9, 1555. doi:10.3389/fonc.2019.01555

Zhang, J., Wu, T., Simon, J., Takada, M., Saito, R., Fan, C., et al. (2018). VHL Substrate Transcription Factor ZHX2 as an Oncogenic Driver in clear Cell Renal Cell Carcinoma. *Science* 361 (6399), 290–295. doi:10.1126/science.aap8411

Zhang, Q., Huang, R., Hu, H., Yu, L., Tang, Q., Tao, Y., et al. (2020). Integrative Analysis of Hypoxia-Associated Signature in Pan-Cancer. *iScience* 23 (9), 101460. doi:10.1016/j.isci.2020.101460

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.

Copyright © 2021 Zhong, Zhong, Zhang, Huang, Lin 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.

GLOSSARY

ccRCC clear cell renal cell carcinoma

CI confidence interval

CMap Connectivity map

CNV copy number variation

DEGs differentially expressed genes

DFS disease-free survival

ESTIMATE Estimation of stromal and immune cells in malignant tumors using expression data

FC fold change

FDR false discovery rate

GDSC Genomics of Drug Sensitivity in Cancer

GISTIC Genomic Identification of Significant Targets in Cancer

GO Gene Ontology

GSEA Gene Set Enrichment Analysis

GSVA Gene set variation analysis

HR hazard ratio

IBERSORT Cell type identification by estimating relative subsets of RNA transcripts

IC50 the half maximal inhibitory concentration

ICGC International Cancer Genome Consortium

ICI immune checkpoint inhibitors

KEGG Kyoto Encyclopedia of Genes and Genomes

LASSO least absolute shrinkage and selection operation

limma linear models for microarray data

MoA CMap mode-of-action

MSigDB Molecular Signatures database

NMF non-negative matrix factorization

OS overall survival

PCA principal components analysis

RCC renal cell carcinoma

RNA-seq RNA sequencing

ROCs receiver operating characteristic curves

TCGA The Cancer Genome Atlas

TIDE Tumor Immune Dysfunction and Exclusion

TMB tumor mutation burden

t-SNE t-distributed stochastic neighbor embedding.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fmolb.2021.688274

Analysis of the Prognostic Value and Gene Expression Mechanism of *SHOX2* in Lung Adenocarcinoma

*Nanhong Li[1,2†], Yu Zeng[3†], Min Tai[4], Biyun Lin[2], Di Zhu[4], Yi Luo[4], Xinle Ren[4], Xiaoying Zhu[4], Lanlan Li[4], Hongrong Wu[2] and Jian Huang[2,4]**

[1]Institute of Nephrology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China, [2]The Center of Pathological Diagnosis and Research, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China, [3]Department of Respiration, The Second Affiliated Hospital of Guangdong Medical University, Zhanjiang, China, [4]Department of Pathology, Guangdong Medical University, Zhanjiang, China

Background: Detection of *SHOX2* methylation has been used to assist in the early diagnosis of lung cancer in many hospitals as *SHOX2* may be important in the tumorigenesis of lung cancer. However, there are few studies on the mRNA expression, methylation, and molecular mechanism of *SHOX2* in lung cancer. We aimed to explore the role of *SHOX2* in lung adenocarcinoma (LUAD).

Methods: First, we examined the differential expression of *SHOX2* mRNA and methylation in cancerous and normal tissues using databases. Second, we analyzed the relationship between *SHOX2* expression and common clinical parameters in LUAD patients. Third, we further explored the methylated level and its specific location of *SHOX2* and the mainly factors of *SHOX2* gene expression. Finally, we screened the correlatively expressed genes to analyze the pathways from the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes using DAVID.

Results: We found that the mRNA expression of *SHOX2* was higher in multiple cancers, including LUAD and lung squamous cell carcinoma (LUSC), than in normal tissues. Among LUAD patients, *SHOX2* expression was higher in patients of middle–young age, with smoking history, in advanced stages, and with nodal distant metastasis. In addition, our results showed that patients with high expression of *SHOX2* are prone to recurrence, poor differentiation, and poor prognosis. Thus, we identified that *SHOX2* might be an oncogene for LUAD progression. The main factor influencing the high expression of *SHOX2* mRNA may be DNA methylation, followed by copy number variation (CNV), but not by gene mutations in LUAD. Unexpectedly, we found that *SHOX2* undergoes hypomethylation in the gene body instead of hypermethylation in the promoter. Additionally, SHOX2 has cross talk in the PI3K–Akt signaling pathway and ECM–receptor interaction.

Conclusion: *SHOX2* is highly expressed in most cancers. *SHOX2* gene expression might be mainly regulated by methylation of its gene body in LUAD, and its high expression or hypomethylation indicates poor differentiation and poor prognosis. SHOX2 could be involved in PI3K–Akt and other important cancer-related signaling pathways to promote tumorigenesis.

Keywords: *SHOX2*, methylation, expression, clinical parameter, LUAD, bioinformatic platform analysis

OPEN ACCESS

Edited by:
Hossain Shekhar, University of Dhaka, Bangladesh

Reviewed by:
Magdalena Julita Orzechowska, Medical University of Lodz, Poland
Shilpita Karmakar, Independent Researcher, Rocky Hill, CT, United States

***Correspondence:**
Jian Huang
18665763598@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *30 March 2021*
Accepted: *26 May 2021*
Published: *28 June 2021*

Citation:
Li N, Zeng Y, Tai M, Lin B, Zhu D, Luo Y, Ren X, Zhu X, Li L, Wu H and Huang J (2021) Analysis of the Prognostic Value and Gene Expression Mechanism of SHOX2 in Lung Adenocarcinoma. Front. Mol. Biosci. 8:688274. doi: 10.3389/fmolb.2021.688274

INTRODUCTION

Much progress has been made in the research field of lung cancer occurrence and progression. However, epidemiological investigations have shown that lung cancer has higher morbidity and mortality than other malignant cancers in China (Chen et al., 2016; Wu et al., 2019). The alarming statistical results of lung cancer remain mainly attributed to the late detection, late diagnosis, late treatment, and unsatisfactory therapeutic effect. In lung cancer, driver gene alterations are critical in the whole process of tumorigenesis, recurrence, and metastasis. More representative driver genes as biomarkers for accurate early diagnosis and prognostic evaluation and therapeutic targeted molecules could improve treatment efficiency and increase mortality. Therefore, the identification of these molecular biomarkers for lung cancer is of great clinical significance.

Short stature homeobox 2 (*SHOX2*) is highly orthologous to murine Shox2 and human *SHOX* (Blaschke et al., 1998; De Baere et al., 1998; Marchini et al., 2016; Hu et al., 2018). According to current studies on *Shox2*, *SHOX*, and SHOX2, *SHOX2* is considered to play a critical role in idiopathic short stature and various types of cardiac arrhythmias (Mortensen et al., 2012; Marchini et al., 2016). However, in recent decades, many researchers have found that *SHOX2* also plays an important role in multiple cancers, including lung cancer (Schneider et al., 2011; Hong et al., 2014; Sun et al., 2019). Interestingly, methylation detection of *SHOX2* in sputum, blood, alveolar lavage fluid, and tissue of patients has been used to screen early lung cancer patients (Zhao et al., 2015; Weiss et al., 2017; Zhang et al., 2017; Shi et al., 2020). It is worth noting that there is little research on the potential prognostic influence and molecular mechanism of *SHOX2* in lung cancer, especially in non-small-cell lung cancer (NSCLC) [mainly lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)]. The clinical significance of *SHOX2* research as an early diagnostic gene is beyond doubt.

In our research, data from The Cancer Genome Atlas (TCGA) were mined using several online analysis tools to evaluate the *SHOX2* expression profile in NSCLC. The main factors of the *SHOX2* gene expression mechanism, including mutation, copy number, and methylation, were then investigated. The relationship between *SHOX2* expression and the clinical characteristics of LUAD was determined using publicly accessible databases. Finally, potential co-expressed genes and functional networks were analyzed to provide direction for further investigation into the mechanism of how *SHOX2* works in lung cancer.

METHODS

Oncomine Database Analysis

Oncomine is a large oncogene chip database that can be used to analyze and compare gene expression between tumors and normal tissues, as well as gene mutation profiles and their correlation with clinical characteristics (Rhodes et al., 2007). We used the Oncomine database to determine differences in the mRNA expression of the *SHOX2* gene between tumors and normal tissues in various cancers. The results of this analysis are addressed with a p-value of 0.05, a fold change of 2, all gene rankings, and mRNA data type.

GEPIA

Gene expression profiling interactive analysis (GEPIA) is a web tool that provides key interactive analysis and customization capabilities, including tumor/normal differential expression profilometry, profile mapping, pathological staging, patient survival analysis, similar gene assay analysis, and dimensionality reduction analysis (Tang et al., 2017). We used GEPIA to address the differential expression of *SHOX2* in all common cancers. We applied the following cut-off criteria: using the ANOVA method, $|\log2\,FC| > 1$, p-value < 0.01, and log2(TPM +1) for log-scale, matching TCGA and GTEx normal data, and adding all cancer tissue names.

GEO Microarray Analysis

The Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) is a database that stores chips, second-generation sequencing, and other high-throughput sequencing data worldwide (Edgar et al., 2002). An mRNA expression dataset (GSE33532) for early stage NSCLC was downloaded from the GEO database. GSE33532 stored mRNA expression information of 80 NSCLC tissue samples and 20 matched distant-normal samples from 20 patients. Microarray expression profiling was normalized by log2 transformation before analysis. These 20 patients were included in the validation dataset. The histological types of these NSCLC samples included adenocarcinoma, squamous cell carcinomas, and mixed carcinomas. Finally, we used GraphPad Prism 7 software to analyze and plot the collected data using unpaired t-test statistical methods. The p-value < 0.05 was considered statistically significant.

SurvivalMeth Analysis

SurvivalMeth (http://bio-bigdata.hrbmu.edu.cn/survivalmeth/) is a web server that is freely available to address the prognostic information of cancer-associated methylation, based on TCGA, CCLE, and GEO (Zhang et al., 2020). We used SurvivalMeth to investigate the effect of *SHOX2* DNA methylation–related functional elements on protein expression and LUAD prognosis. We applied the following cut-off criteria: the chosen disease of lung adenocarcinoma, the chosen experimental platform of 450 K (Illumine Infinium HumanMethylation450 BeadChip), the chosen gene symbol of *SHOX2*, transcript-related elements including "1st exon, 3′UTR, 5′UTR, gene body, TSS1500, and TSS200," CpG island–related elements including "Island, N_Shelf, N_Shore, S_Shelf, and S_Shore," unrestricted repeat element–related element and CTCF-binding region, statistical methods using t-test, a threshold value of 0.05, and median group strategy.

UCSC Xena Analysis

The University of California Santa Cruz Cancer Genomics Browser (UCSC Xena) (http://xena.ucsc.edu/) is an online

database of genomic, transcriptomic, and clinical and phenotypic data (Kent et al., 2002; Zweig et al., 2008; Navarro Gonzalez et al., 2021). To investigate the dominant factors influencing *SHOX2* expression in LUAD, UCSC Xena was used to analyze the relationship between the methylation level, copy number, mutation, Kaplan–Meier survival analysis, and mRNA expression of *SHOX2* in TCGA LUAD samples. We set the following conditions: sample types including solid tissue normal and primary tumor, gene expression using RNAseq-IlluminaHiSeq, DNA methylation using Methylation 450K, copy number using (gene-level)-gistic2, and somatic mutation using (SNP and INDEL)-MC3 public version. Additionally, we used this tool to analyze *SHOX2* expression patterns in different pathological types.

UALCAN Analysis

UALCAN (http://ualcan.path.uab.edu) is an online database that provides free visual figures of gene expression, survival analysis, correlation analysis, and gene DNA promoter region methylation data, grouped by clinicopathological features between normal and tumor tissues, based on TCGA data (Chandrashekar et al., 2017). To explore the relationship between *SHOX2* and the clinicopathological features of NSCLC patients, the expression and methylation levels of *SHOX2* in NSCLC and adjacent normal tissues were identified using UALCAN.

MethSurv Analysis

MethSurv (https://biit.cs.ut.ee/methsurv/) used methylation group data from the "Cancer Genome Map" to perform survival analysis on the methylation patterns of CpG to achieve a preliminary assessment of tumor biomarkers based on methylation (Modhukur et al., 2018). We used MethSurv to further explore the relationship between *SHOX2* methylation and several clinical characterizations in a heat map.

cBioPortal Database Analysis

The cBioPortal database (v3.5.4) (http://www.cbioportal.org) was used to perform the cancer genomic analysis (Gao et al., 2013; Jing et al., 2019). We chose two datasets, "Lung Adenocarcinoma (TCGA, Firehose Legacy)" and "Lung Adenocarcinoma (TCGA, Nature 2014)," to analyze the co-expressed *SHOX2* genes. We then set the following parameters: select genomic profiles [mutations, putative copy number alterations from GISTIC, and mRNA expression z-scores relative to diploid samples (RNA Seq V2 RSEM)], select patient/case set (all samples), and enter genes (user-defined list, *SHOX2*). Subsequently, we entered the co-expression interface and acquired the genes correlated with *SHOX2* in satisfactory samples from these datasets. These data were downloaded, and the next operating steps were followed.

Using the Venn Diagram Tool

We selected the first 300 *SHOX2* correlated genes from the previous steps with a higher Spearman correlation coefficient and $p < 0.05$. According to the applied cBioPortal database, one gene list was named "TCGA, Firehose Legacy" and the other gene list was named "TCGA, Nature 2014." We input the two lists into the upload lists. Then, we acquired the Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) of *SHOX2* co-expressed related genes.

Functional and Pathway Enrichment Analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (version 6.8; https://david.ncifcrf.gov/) is a biological information database that integrates biological data and analysis tools to provide systematic and comprehensive annotated biological function information for large-scale gene or protein lists to help users extract biological information from them (Huang da et al., 2009). We used DAVID to perform Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the selected common *SHOX2* co-expressed related genes screened by the two datasets. GO enrichment analysis consisted of cellular component (CC), biological process (BP), and molecular function (MF) analyses. Following these steps, we uploaded these genes, selected the identifier "OFFICIAL_GENE_SYMBOL," selected "Homo spaiens," selected "Gene List," submitted the list, and finally obtained functional annotation charts. We then acquired the GO functional enrichment and KEGG pathway enrichment results. Calculated *via* Fisher's exact test, a p-value < 0.05 and a count >2 were considered statistically significant.

RESULTS

SHOX2 mRNA Expression Profile in Multiple Cancers

To determine the expression level of *SHOX2* in various types of tumors, we examined the differential expression of *SHOX2* between tumor tissue and paired normal tissue using Oncomine and GEPIA online analysis tools. The Oncomine database has a total of 320 unique analyses for *SHOX2*. There are 52 significant unique analyses among them that showed *SHOX2* upexpression, while 15 analyses showed *SHOX2* downexpression (**Figure 1A**). *SHOX2* expression was higher in most cancers, including sarcoma, brain and CNS cancer, head and neck squamous carcinoma (HNSC), and lung cancer. Meanwhile, it was downregulated in breast cancer, leukemia, and many other cancers (**Figure 1A**). For further verification, we used the GEPIA web tool to compare the differential expression of *SHOX2* between cancerous and normal tissues. As shown in **Figure 1B**, higher *SHOX2* expression was observed in glioblastoma (GBM), HNSC, and LUSC. In addition, lower *SHOX2* expression was observed in breast invasive carcinoma (BRCA), acute myeloid leukemia (LAML), and testicular germ cell tumors (TGCTs). In summary, the results from the two databases were consistent, supporting the evidence that *SHOX2* should be an oncogene.

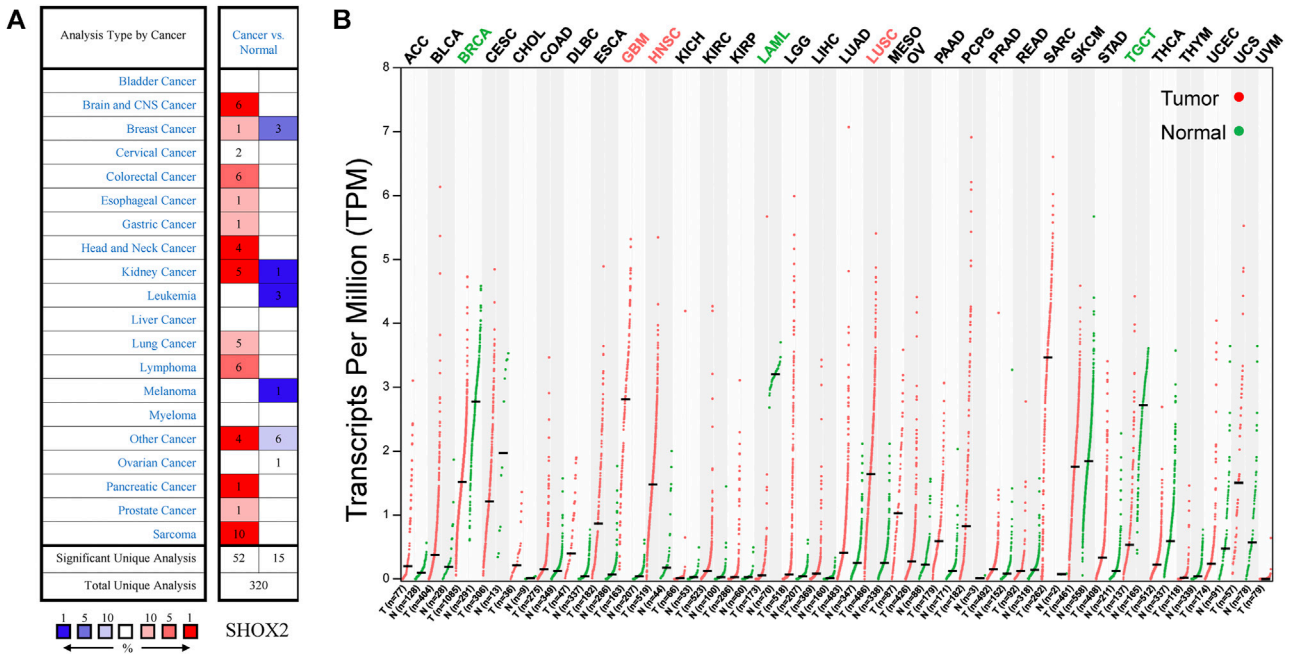


FIGURE 1 | *SHOX2* mRNA expression in multiple cancers. **(A)** The number of datasets with *SHOX2* mRNA differential expression between tumor tissues and normal tissues. The left column (red) represents overexpression, and the right column (blue) represents downexpression. This visualized graphic is available from Oncomine, and the threshold parameters were set as follows: *p*-value = 0.05, fold change = 2, all gene rankings, and mRNA data type. **(B)** *SHOX2* gene expression profile in various tumor samples and normal samples by GEPIA. *SHOX2*: short stature homeobox 2; GEPIA: gene expression profiling interactive analysis.

Analysis of *SHOX2* Expression in LUAD/LUSC and Normal Tissues

In order to explore *SHOX2* mRNA expression and its relationship with clinical features, we addressed the GSE33532 data using GraphPad Prism 7 and used UALCAN to analyze the relative clinical data of LUAD/LUSC patients from TCGA. The *SHOX2* mRNA expression profiling dataset of NSCLC showed that *SHOX2* is highly expressed in both LUAD and LUSC tissues compared to paired normal tissues (**Figure 2A**). The difference in *SHOX2* expression was statistically significant. This result was verified by an expanded sample size analysis using the online analysis database UALCAN. As shown in **Figure 2B**, we obtained consistent results.

The Relationship of *SHOX2* Expression With the Clinical Characteristics of LUAD Patients

In recent decades, the incidence of LUAD has been higher than that of LUSC. Moreover, there are few reports on the relationship between *SHOX2* expression and the clinicopathological features of LUAD. Therefore, we used the UALCAN online analysis tool to address *SHOX2* mRNA expression in different subgroups having multiple clinicopathological features from LUAD samples from TCGA. The *SHOX2* mRNA expression level was significantly higher in LUAD patients than in healthy people in subgroup analyses based on sex, age, smoking habits, stage, nodal metastasis, and TP-53 mutation status (**Figures 3A–F**). In addition, subgroup analyses showed that *SHOX2* expression was higher in patients aged 20–40 years, with a history of smoking, in stage 2, and with nodal distant metastasis, compared to other clinical features (**Figure 3**). The pathology of disease can help clinicians to clearly diagnose the disease and take reasonable and effective treatment measures. UCSC Xena was used to analyze the changes of *SHOX2* expression in different histological subtypes of invasive LUAD from a cohort of TCGA LUAD (n = 706) (**Figure 3G**). The results showed that *SHOX2* expression varied significantly among different histological subtypes of LUAD (one-way ANOVA, p = 0.0001271, f = 4.098). *SHOX2* expression was higher in subtypes prone to lymph node metastasis, prone to recurrence, with poor differentiation, and with poor prognosis.

Furthermore, to confirm the prognostic value of SHOX2 in LUAD, UCSC Xena was searched to investigate the effects of *SHOX2* expression, methylation, and CNV on overall survival (OS) and disease-specific survival (DSS). It was confirmed that high *SHOX2* mRNA expression was significantly associated with decreased OS and DSS time in LUAD ($p < 0.05$); DNA methylation and copy number variation had no significant effect on OS ($p > 0.05$), and a higher copy number indicated a shorter DSS ($p < 0.05$). The *SHOX2* expression level and DNA methylation level had no significant correlation with DSS ($p > 0.05$) (**Figure 4**).

Therefore, these results suggest that *SHOX2* might be a driving factor in the development of LUAD and may act as a potential diagnostic and prognostic indicator in LUAD.

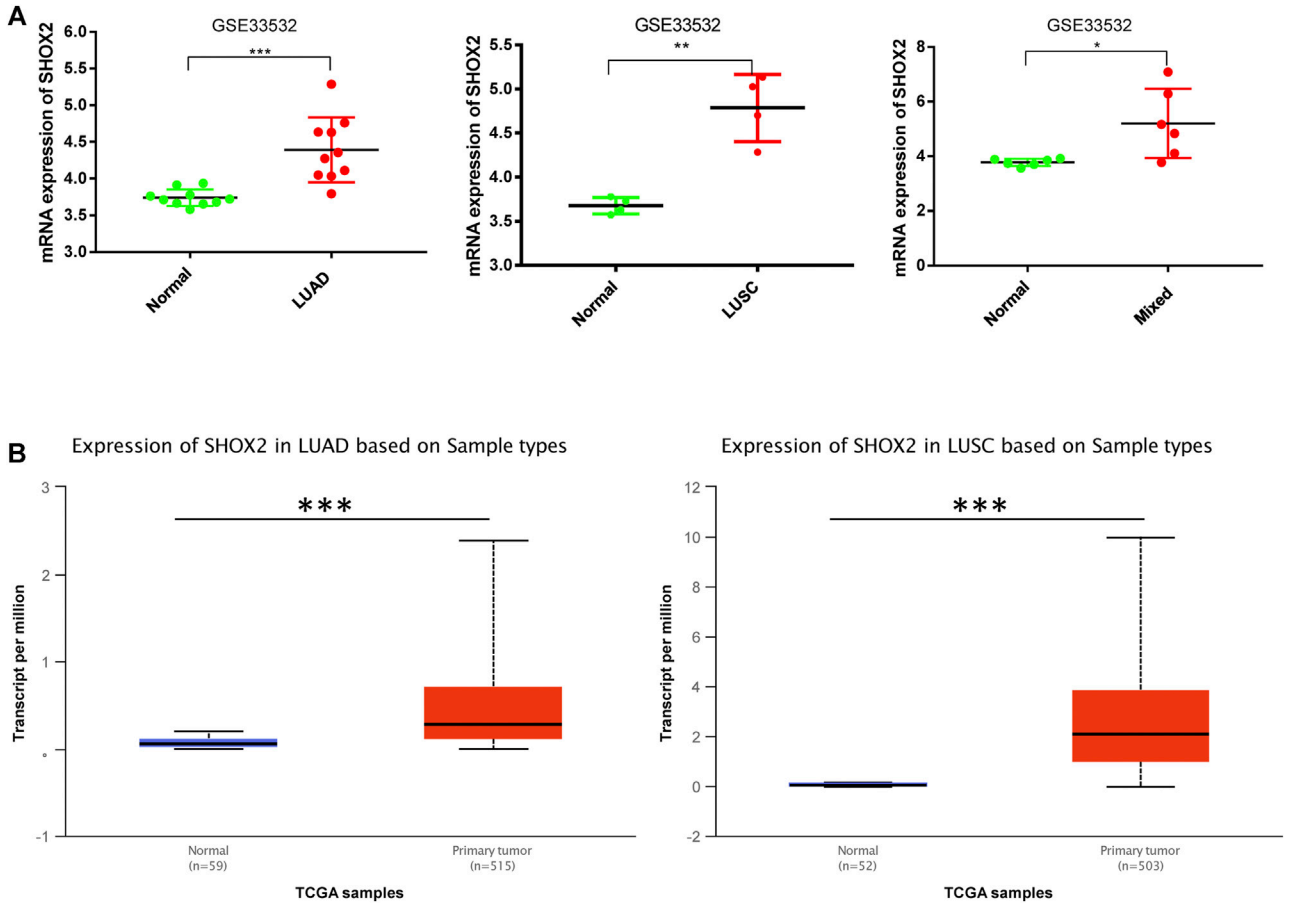


FIGURE 2 | Box plot showing the relative expression of *SHOX2* in normal and NSCLC samples. **(A)** *SHOX2* expression is shown by microarray analysis of 20 NSCLC patients in GSE33532 including 10 LUAD patients, 4 LUSC patients, and 6 mixed patients. These data are compared with those of paired normal tissues. The data are addressed by GraphPad Prism 7.0 and applied paired *t*-test. **(B)** The expression levels of *SHOX2* between normal and LUAD or LUSC tissues are shown in the box plot by UALCAN online analysis tool application. *** represents $p < 0.001$ (highly significant); ** represents $p < 0.01$ (very significant); * represents $p < 0.05$ (significant); LUAD: lung adenocarcinoma; LUSC: lung squamous carcinoma; NSCLC: non-small-cell lung cancer.

Influencing-Factor Analysis of *SHOX2* Gene Expression in LUAD

To determine the main influencing factors of *SHOX2* gene expression, we used several online tools performing mutation, methylation, and CNV alteration analysis. We used three methylation analysis tools (UCSC Xena, SurvivalMeth, and MethSurv) to probe the *SHOX2* methylation level in LUAD patients, from different perspectives. The methylation level of *SHOX2* was examined in TCGA LUAD patients, based on age, race, ethnicity, alive or dead events, *SHOX2* mRNA expression, CpG island–related elements, and transcript-related elements. From **Figure 5** and **Table 1**, we can conclude that DNA methylation of *SHOX2* was elevated in LUAD tissues compared with that in normal lung tissues. As shown in **Figure 5**, *SHOX2* DNA was only locally methylated. To explore the DNA-methylated specific site of *SHOX2* and verify the acquired results, we conducted an in-depth study using TCGA LUAD patient data in SurvivalMeth and MethSurv. In LUAD tissues, *SHOX2* DNA methylation mainly occurred in the gene body, at the hypomethylation level ($p < 0.05$) (**Figures 5** and **6**, **Tables 1** and **2**). According to previous research results, hypomethylation of the gene body leads to the high expression of oncogenes, which is opposite to the hypermethylation of the gene promoter for tumor suppressive genes (Yang et al., 2014). Therefore, hypomethylation of the *SHOX2* gene body might be one of the main factors causing mRNA expression. In conclusion, in LUAD tissue, the mRNA expression of *SHOX2* might be correlated with the DNA methylation level, followed by CNV, but not with gene mutation (**Figure 5**).

Enrichment Analysis of Correlated Genes With *SHOX2* in LUAD

Finally, to explore the possible signaling pathways from the genes correlated with *SHOX2*, we analyzed mRNA sequencing data from 816 LUAD patients in two TCGA research cohorts: Firehose Legacy (586 samples) and Nature 2014 (230 samples). As shown in the Venn diagram (**Figure 7A**), 139 significantly correlated

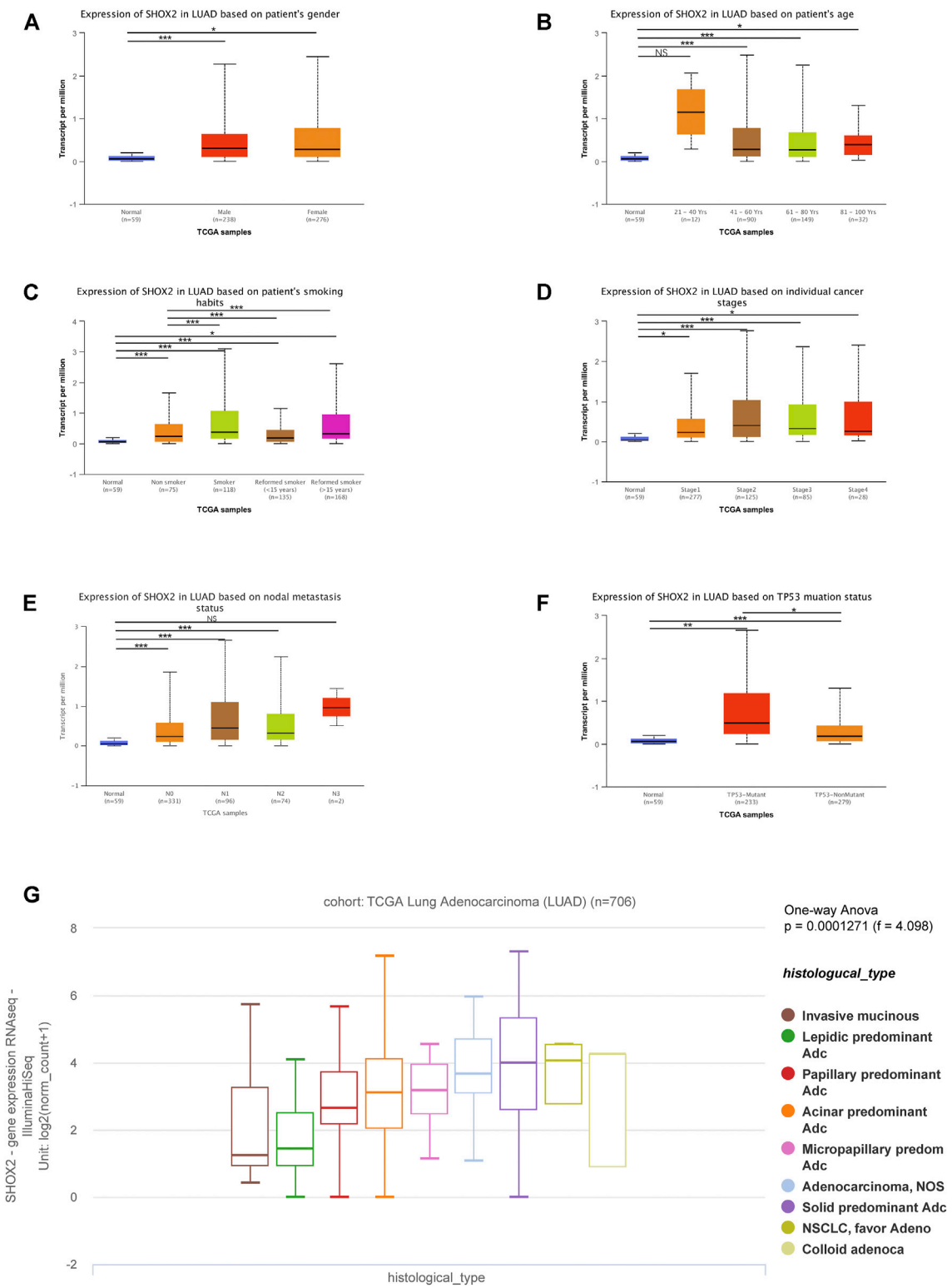


FIGURE 3 | Box plot showing the characteristic relationships of **(A)** gender, **(B)** age, **(C)** smoking habits, **(D)** stage, **(E)** nodal metastasis, **(F)** TP-53 mutation status, and **(G)** histological subtypes with *SHOX2* mRNA expression in LUAD analyzed using the online tools UALCAN and UCSC Xena. Data are represented as mean ± SE. *** represents $p < 0.001$ (highly significant); ** represents $p < 0.01$ (very significant); * represents $p < 0.05$ (significant); LUAD: lung adenocarcinoma.

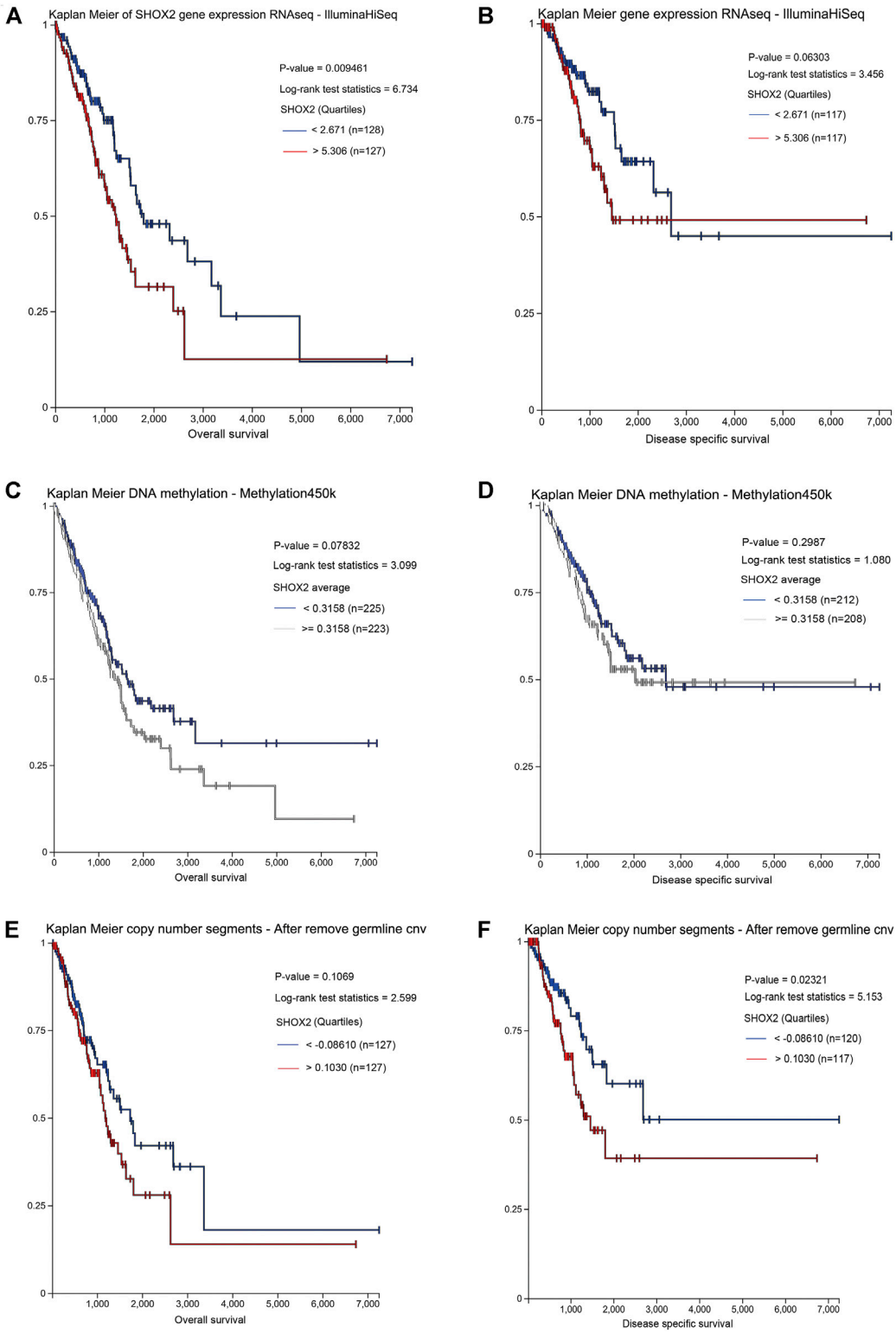


FIGURE 4 | Prognostic value of the *SHOX2* mRNA expression level in LUAD patients by UCSC Xena (Kaplan–Meier plotter). **(A)** Relationship between *SHOX2* expression and OS in LUAD. **(B)** Relationship between *SHOX2* expression and DSS in LUAD. **(C)** Relationship between *SHOX2* DNA methylation and OS in LUAD. **(D)** Relationship between *SHOX2* DNA methylation and DSS in LUAD. **(E)** Relationship between *SHOX2* copy number segments and OS in LUAD. **(F)** Relationship between *SHOX2* copy number segments and DSS in LUAD. $p < 0.05$ indicates statistical significance. LUAD: lung adenocarcinoma; OS: overall survival; DSS: disease-specific survival.

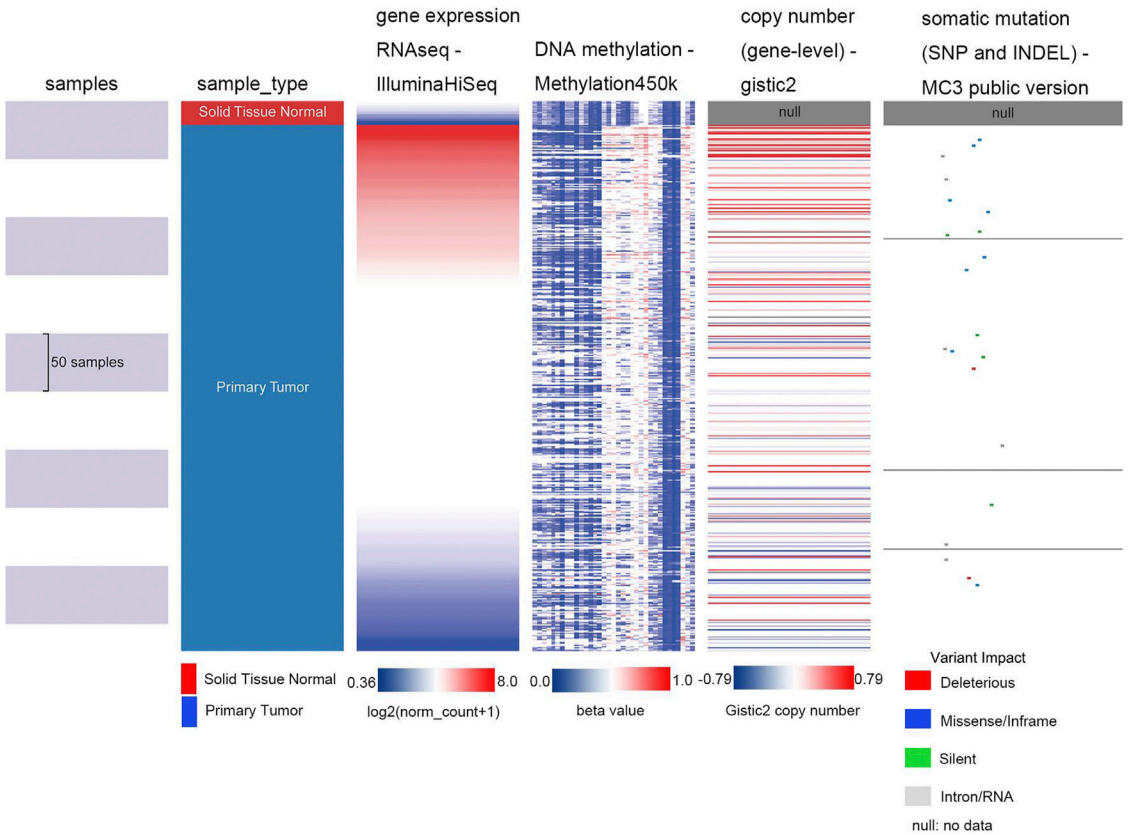


FIGURE 5 | Heat map of *SHOX2* mRNA expression, methylation, copy number, and somatic mutation in patients with primary LUAD and normal tissue. The data are obtained from TCGA-LUAD with a total of 706 samples. LUAD: lung adenocarcinoma.

genes with *SHOX2* appeared in both cohorts ($p < 0.05$). This suggests that *SHOX2* has a widespread impact on the transcriptome. We then performed functional enrichment analysis of *SHOX2* and these 139 common significantly correlated genes using the DAVID database. The top 10 GO terms of these genes, according to the gene counts, are shown in **Figure 7B**, including biological processes (BPs), cellular components (CCs), and molecular functions (MFs). The BP of *SHOX2* was mainly associated with positive regulation of cell migration, cell division, skeletal system development, and mitotic nuclear division. The CC of *SHOX2* was mainly located in the extracellular space. The MF of *SHOX2* mainly focuses on the structural constituents of the extracellular matrix and serine-type endopeptidase activity. Surprisingly, the significant KEGG pathways included the PI3K–Akt signaling pathway, ECM–receptor interaction, protein digestion and absorption, and amebiasis ($p < 0.05$).

DISCUSSION

Early detection, early diagnosis, and early treatment are of key importance in preventing disease development. At present, conventional diagnostic methods for lung cancer include computed tomography imaging and cellular/histopathological examination. Emerging molecular diagnostic methods are being increasingly applied in clinical practice because they are more sensitive and objective. DNA methylation alteration is considered an early predictor of cancer and can be detected during the early stages of tumorigenesis (Zhang Y. et al., 2016; Dirks et al., 2016; Wang et al., 2018; Srisuttee et al., 2020). Because of the hypermethylation of *SHOX2* in lung cancer, methylation detection has been applied to assist in the early diagnosis of unknown pulmonary nodules (Kneip et al., 2011; Konecny et al., 2016; Weiss et al., 2017; Shi et al., 2020). Many studies have reported that *SHOX2* is highly expressed in many cancers. The specific role of *SHOX2* in lung cancer patients and the molecular mechanism of its action are unknown.

As reported in detail in our previous review (Li et al., 2020), we synthesized the results of existing studies on *SHOX2* and its related homonymous genes in mice, lung cancer and other tumors and concluded that *SHOX2* might play an important role in tumorigenesis, metastasis, and recurrence in lung cancer. Here, we used bioinformatics analysis to verify the above conclusions.

In our study, we found that the mRNA expression of *SHOX2* was higher in multiple cancers than in paired normal tissues, including LUAD and LUSC tissues. Among the LUAD patients,

TABLE 1 | Differential methylation levels of the SHOX2 probe between LUAD and normal samples.

Probe ID	Gene location	CpG island regions	Average methylated data value of tumor samples	Average methylated data value of normal samples	Delta value	Fold change	*p*-Value
cg00591153	TSS1500	S_Shore	0.268614	0.180493	0.088121	1.488224	4.45580E–17
cg14961949	TSS1500	S_Shore	0.173536	0.065364	0.108171	2.654909	3.30727E–35
cg15548431	TSS1500	S_Shore	0.232887	0.171044	0.061843	1.36156	4.02456E–10
cg17191178	TSS1500	S_Shore	0.298609	0.18082	0.117789	1.651417	1.09924E–29
cg21503297	TSS1500	S_Shore	0.300121	0.227304	0.072816	1.320346	1.50735E–15
cg24317285	TSS1500	S_Shore	0.302047	0.237082	0.064965	1.274019	1.20478E–14
cg25460158	TSS1500	S_Shore	0.315503	0.266667	0.048836	1.183137	4.88871E–07
cg25694447	TSS200	S_Shore	0.31945	0.238284	0.081165	1.340622	4.16236E–16
cg26129769	TSS200	S_Shore	0.33075	0.241758	0.088992	1.368102	4.80355E–16
cg16703882	1st EXON	Island	0.249193	0.219976	0.029217	1.13282	4.32357E–03
cg00010946	BODY	N_Shore	0.616993	0.549898	0.067095	1.122014	9.46071E–08
cg00089486	BODY	S_Shore	0.116562	0.096583	0.019978	1.20685	7.74109E–05
cg01557547	BODY	Island	0.241769	0.093843	0.147926	2.576302	6.99470E–43
cg04521004	BODY	Island	0.512873	0.243548	0.269325	2.105839	3.27178E–51
cg04532033	BODY	Island	0.28867	0.137343	0.151326	2.101811	2.20417E–49
cg04913979	BODY	N_Shore	0.479174	0.188612	0.290561	2.540521	5.56975E–49
cg05512327	BODY	N_Shore	0.547018	0.372874	0.174143	1.467029	2.66119E–28
cg09095222	BODY	Island	0.1328	0.122822	0.009977	1.081235	3.59488E–02
cg09220088	BODY	N_Shore	0.424979	0.295727	0.129251	1.437061	4.68587E–21
cg09542210	BODY	Island	0.390491	0.135601	0.254889	2.879692	5.24557E–77
cg12993163	BODY	Island	0.389816	0.103129	0.286687	3.779884	9.59716E–121
cg15726154	BODY	S_Shore	0.350172	0.307982	0.04219	1.136988	2.95931E–03
cg18194945	BODY	N_Shore	0.542494	0.300898	0.241596	1.802915	2.95084E–28
cg18899952	BODY	Island	0.173432	0.107374	0.066058	1.615213	1.51266E–19
cg20501518	BODY	N_Shore	0.401894	0.183032	0.218862	2.19576	7.34036E–40
cg21437028	BODY	S_Shore	0.317038	0.267349	0.049689	1.185859	3.47849E–04
cg21847036	BODY	S_Shore	0.35551	0.304502	0.051008	1.167512	5.67448E–04
cg25506747	3′UTR	N_Shore	0.33401	0.169969	0.164041	1.965125	2.02697E–25
cg21552242	3′UTR	N_Shore	0.479399	0.276377	0.203021	1.734581	2.60443E–33
cg12315391	3′UTR	N_Shore	0.537819	0.425058	0.112761	1.265283	6.56708E–14
cg05401764	3′UTR	Island	0.045138	0.037674	0.007464	1.198133	3.45000E–02

Note: DNA methylation level is represented by data value (0 represents unmethylated, 0.25–0.3 represents hypomethylated, 0.5–0.7 represents hypermethylated, and 1 represents completely methylated).

SHOX2 expression was higher in patients of middle–young age, with smoking history, in advanced stages, and with nodal distant metastasis. In addition, our results showed that patients with high expression of *SHOX2* are prone to recurrence, poor differentiation, and poor prognosis. Here, we identified that *SHOX2* plays a negative role in LUAD progression and could act as an oncogene. Next, we performed *SHOX2* methylation analysis. Finally, we found that *SHOX2* undergoes hypomethylation in the gene body. The main factors for *SHOX2* mRNA expression are the DNA methylation and CNV, but not gene mutations in LUAD. Additionally, *SHOX2* has cross talk in the PI3K–Akt signaling pathway and in ECM–receptor interactions.

The potential role of *SHOX2* in patients with lung cancer may have been explained by many previous studies. We identified that the mRNA expression of *SHOX2* could assist in indicating subsets of LUAD with worse prognosis and survival. Zhang and Zhou also reported that *SHOX2* was an indicator for identifying subgroups with worse prognosis in lower-grade gliomas, which is consistent with our result (Zhang Y.A. et al., 2016). Additionally, there have been other similar reports of *SHOX2* methylation as a predictor of malignancy (Leiro et al., 2019; Xu et al., 2020).

Tumor cell genomes showed global hypomethylation and local hypermethylation. Local hypermethylation may occur on the promoter CpG island of a tumor suppressor gene to inhibit the expression of this gene, or on the gene body CpG island of an oncogene to induce high expression of this gene (Yang et al., 2014; Gagliardi et al., 2018). Downregulated TSGs or upregulated oncogenes are critical in tumorigenesis. According to recent research, *SHOX2* is considered an oncogene. Unexpectedly, the present majority of related reports declare that *SHOX2* methylation occurs on the promoter CpG island, which can seem incomprehensible. To explain the discrepancy of high methylation of *SHOX2* in cancer tissue and to identify whether methylation occurs at the gene promoter, we conducted further research to try to explain this phenomenon. We used UCSC XENA, SurvivalMeth, and MethSurv to obtain methylation data for all methylation sites of *SHOX2* and then analyzed them statistically using Excel. Interestingly, we observed that the *SHOX2* methylation is in a hypomethylated state, and methylation mainly occurs at the gene body instead of at the gene promoter. In summary, *SHOX2* is highly expressed, and its genomes are hypomethylated in LUAD, which might be the main mechanism of gene expression. Therefore, the current

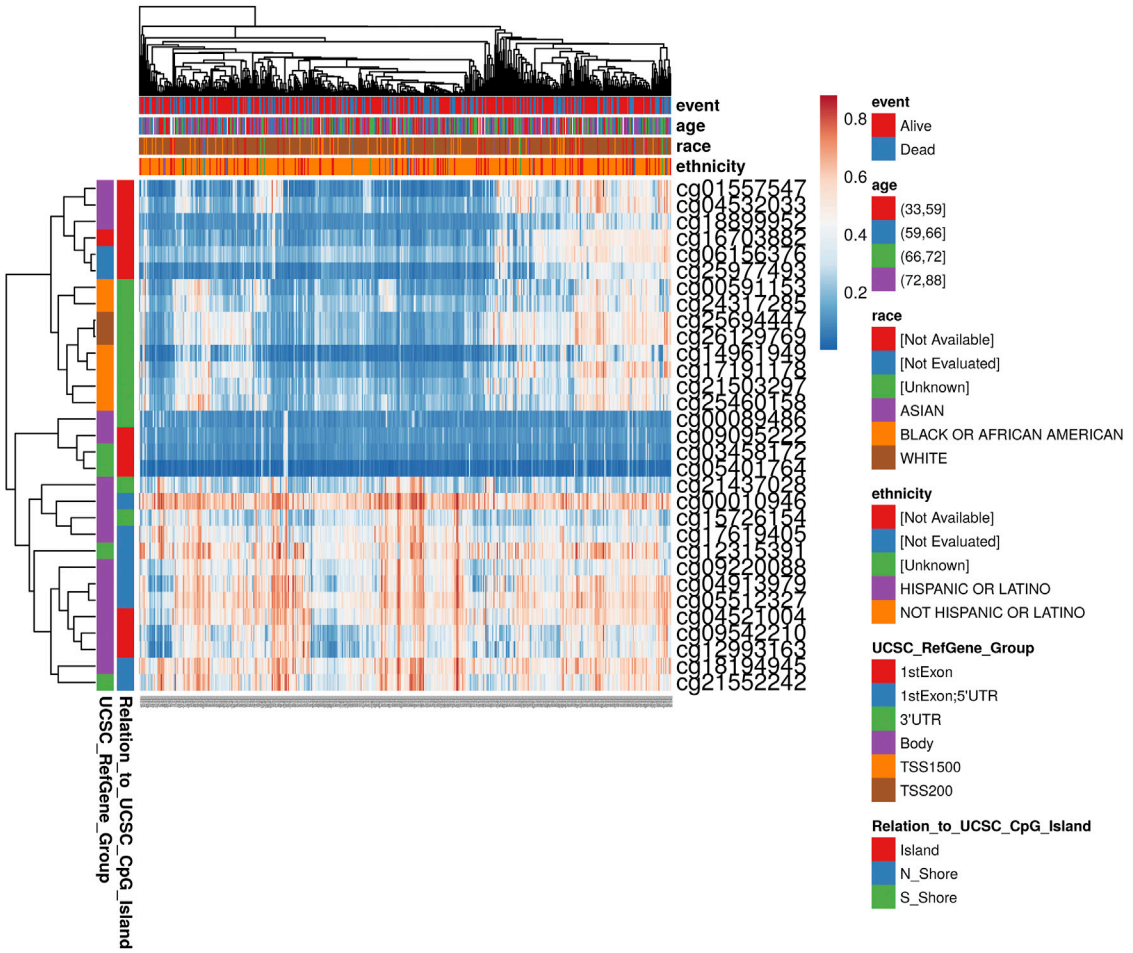


FIGURE 6 | Heat map of methylation levels of the *SHOX2* methylated probe. Red to blue: high expression to low expression. Various colorful side-boxes are used to characterize the event, age, race, ethnicity, UCSC_refGene_Group, and relation to UCSC_CpG_island.

TABLE 2 | Average data value of the SHOX2 methylated probe between LUAD and normal samples.

Gene location	Average data value of the methylated probe in tumor samples	Average data value of the methylated probe in normal samples
Promoter	0.282390778	0.200979556
1st exon	0.249193	0.219976
Body	0.369510882	0.224183353
3′UTR	0.3490915	0.2272695

Note: DNA methylation level is represented by data value (0 represents unmethylated, 0.25–0.3 represents hypomethylated, 0.5–0.7 represents hypermethylated, and 1 represents completely methylated).

description of *SHOX2* promoter hypermethylation may not be rigid.

The regulation of gene expression is the molecular basis of cell differentiation, morphogenesis, and ontogenesis *in vivo*. The influencing factors of gene expression include one or two of the following alterations in most cases: local mutation, CNV, DNA methylation, and the expression level of master transcription factors on another chromosome. We found that hypomethylation of the *SHOX2* gene body may be one of the main mechanisms driving the upregulation of *SHOX2*, and CNV is also one of them. Schneider et al. (2011) and Zhang Y.-A. et al. (2016) may agree with our findings.

According to our enrichment analysis of *SHOX2*, the BP of *SHOX2* was mainly associated with positive regulation of cell migration, cell division, skeletal system development, and mitotic nuclear division, and the significant KEGG pathways included the PI3K–Akt signaling pathway, ECM–receptor interaction, protein digestion and absorption, and amebiasis. These findings are

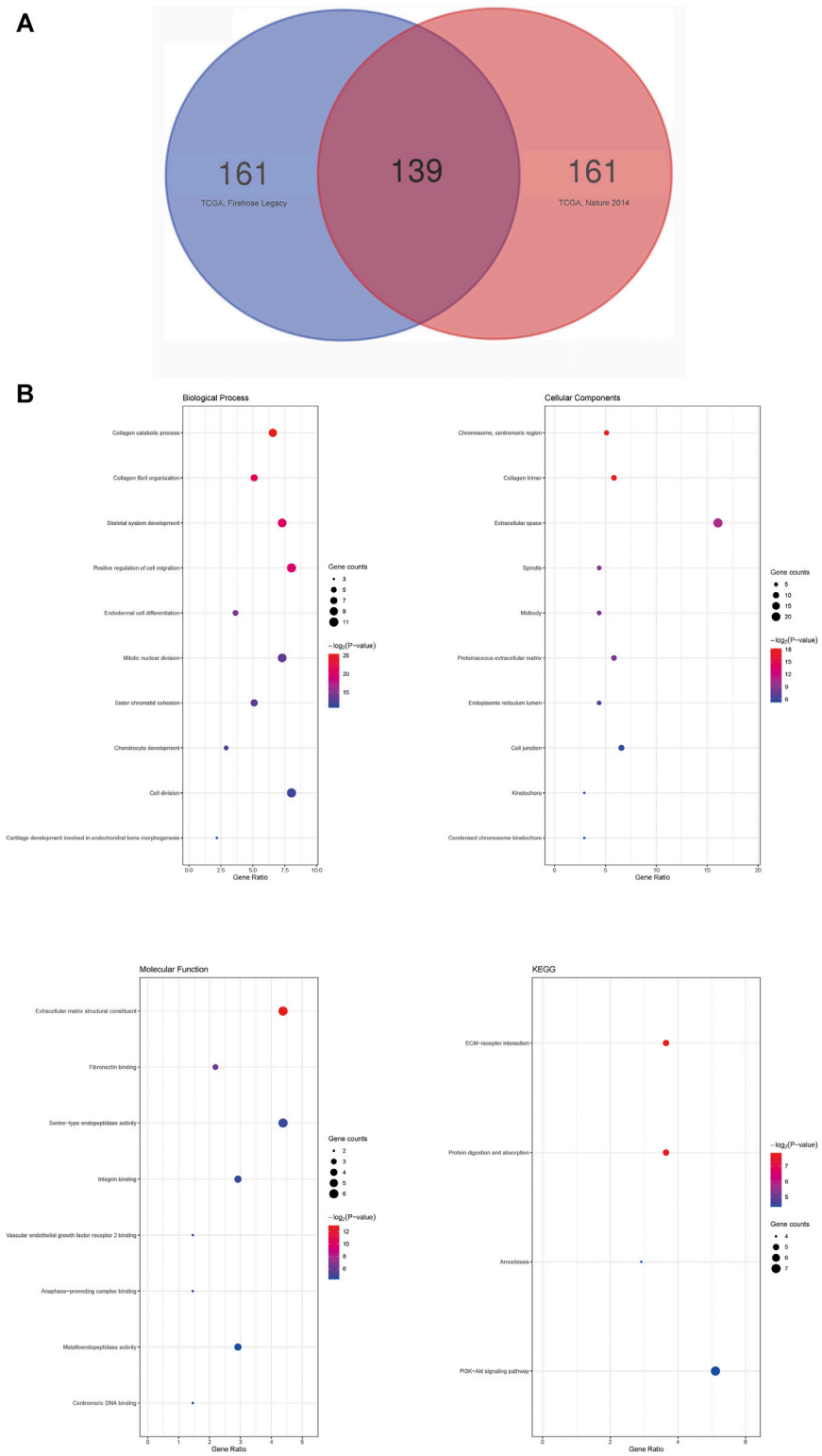


FIGURE 7 | Enrichment analysis of correlated genes with *SHOX2* in LUAD. **(A)** Venn diagram of *SHOX2* correlated genes in LUAD. The 139 significantly correlated genes with *SHOX2* appeared in two TCGA research cohorts, Firehose Legacy (586 samples) and Nature 2014 (230 samples). **(B)** GO and KEGG analyses of *SHOX2* and the significantly correlated genes by DAVID database. The node size represents the gene ratio; the node color represents the *p*-value. LUAD: lung adenocarcinoma; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

consistent with the fact that *SHOX2* is considered an oncogene and a diagnostic and predicted biomarker. This is critical for understanding how *SHOX2* expression alter and how *SHOX2* lead to cancers such as LUAD.

In this study, we combined multiple online bioinformatics data analysis platforms and tools to provide a statistical analysis of the correlation between *SHOX2* expression and clinical characteristic parameters, clinical prognosis, DNA methylation, CNV, mutations, co-expressed genes, and related signaling pathways in LUAD. *SHOX2* is highly expressed in most cancers. *SHOX2* gene expression may be mainly regulated by methylation of the gene body in LUAD, and its high expression or hypomethylation indicates poor differentiation and poor prognosis. *SHOX2* is involved in PI3K–Akt and other important cancer-related signaling pathways to promote tumorigenesis.

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

NL, YZ, MT, and JH conceptualized and designed the idea and analyzed and interpreted the data. JH provided administrative support. NL, YZ, MT, BL, DZ, YL, XR, XZ, LL, and HW contributed to the provision of study materials or patients and collected and assembled the data. NL and YZ wrote the manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81572610) and the "Yangfan Plan" for Outstanding Scholars in Guangdong Province (4YF16002G) which all are from JH.

ACKNOWLEDGMENTS

Thanks are due to Bo Zhao (Official Wechat Account: SCIPhD) of ShengXinZhuShou for suggestions on the manuscript.

REFERENCES

Blaschke, R. J., Monaghan, A. P., Schiller, S., Schechinger, B., Rao, E., Padilla-Nash, H., et al. (1998). SHOT, a SHOX-Related Homeobox Gene, Is Implicated in Craniofacial, Brain, Heart, and Limb Development. *Proc. Natl. Acad. Sci.* 95, 2406–2411. doi:10.1073/pnas.95.5.2406

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

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. Clin.* 66, 115–132. doi:10.3322/caac.21338

De Baere, E., Speleman, F., Van Roy, N., De Paepe, A., and Messiaen, L. (1998). Assignment1 of SHOX2 (Alias OG12X and SHOT) to Human Chromosome Bands 3q25→q26.1 by *In Situ* Hybridization. *Cytogenet. Genome Res.* 82, 228–229. doi:10.1159/000015108

Dirks, R. A., Stunnenberg, H. G., and Marks, H. (2016). Genome-wide Epigenomic Profiling for Biomarker Discovery, *Clin. Epigenetics* 8, 122. doi:10.1186/s13148-016-0284-4

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucl. Acids Res.* 30, 207–210. doi:10.1093/nar/30.1.207

Gagliardi, M., Strazzullo, M., and Matarazzo, M. R. (2018). DNMT3B Functions: Novel Insights from Human Disease, *Front Cell Dev. Biol.* 6, 140. doi:10.3389/fcell.2018.00140

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, pl1. doi:10.1126/scisignal.2004088

Hong, S., Noh, H., Teng, Y., Shao, J., Rehmani, H., Ding, H.-F., et al. (2014). SHOX2 Is a Direct miR-375 Target and a Novel Epithelial-To-Mesenchymal Transition Inducer in Breast Cancer Cells. *Neoplasia* 16, 279–290. e1-5. doi:10.1016/j.neo.2014.03.010

Hu, W., Xin, Y., Zhao, Y., and Hu, J. (2018). Shox2: The Role in Differentiation and Development of Cardiac Conduction System. *Tohoku J. Exp. Med.* 244, 177–186. doi:10.1620/tjem.244.177

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

Jing, X., Liang, H., Hao, C., Yang, X., and Cui, X. (2019). Overexpression of MUC1 Predicts Poor Prognosis in Patients with Breast Cancer, *Oncol. Rep.* 41, 801–810. doi:10.3892/or.2018.6887

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The Human Genome Browser at UCSC. *Genome Res.* 12, 996–1006. doi:10.1101/gr.229102

Kneip, C., Schmidt, B., Seegebarth, A., Weickmann, S., Fleischhacker, M., Liebenberg, V., et al. (2011). SHOX2 DNA Methylation Is a Biomarker for the Diagnosis of Lung Cancer in Plasma. *J. Thorac. Oncol.* 6, 1632–1638. doi:10.1097/jto.0b013e318220ef9a

Konecny, M., Markus, J., Waczulikova, I., Dolesova, L., Kozlova, R., Repiska, V., et al. (2016). The Value of SHOX2 Methylation Test in Peripheral Blood Samples Used for the Differential Diagnosis of Lung Cancer and Other Lung Disorders. *Neoplasma* 63, 246–253. doi:10.4149/210_150419n208

Leiro, V., De Chiara, L., Rodríguez-Girondo, M., Botana-Rial, M., Valverde, D., Núñez-Delgado, M., et al. (2019) Methylation Assessment for the Prediction of Malignancy in Mediastinal Adenopathies Obtained by Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration in Patients with Lung Cancer, *Cancers (Basel)* 11:1408. doi:10.3390/cancers11101408

Li, N., Zeng, Y., and Huang, J. (2020). Signaling Pathways and Clinical Application of RASSF1A and SHOX2 in Lung Cancer. *J. Cancer Res. Clin. Oncol.* 146, 1379–1393. doi:10.1007/s00432-020-03188-9

Marchini, A., Ogata, T., and Rappold, G. A. (2016). A Track Record on SHOX: From Basic Research to Complex Models and Therapy. *From Basic Res. Complex Models Ther. Endocr Rev* 37, 417–448. doi:10.1210/er.2016-1036

Modhukur, V., Iljasenko, T., Metsalu, T., Lokk, K., Laisk-Podar, T., and Vilo, J. (2018). MethSurv: a Web Tool to Perform Multivariable Survival Analysis Using DNA Methylation Data. *Epigenomics* 10, 277–288. doi:10.2217/epi-2017-0118

Mortensen, K. H., Andersen, N. H., and Gravholt, C. H. (2012). Cardiovascular Phenotype in Turner Syndrome-Integrating Cardiology, Genetics, and Endocrinology. *Endocr. Rev.* 33, 677–714. doi:10.1210/er.2011-1059

Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., et al. (2021). The UCSC Genome Browser Database: 2021 Update. *Nucl. Acids Res.* 49, D1046–d1057. doi:10.1093/nar/gkaa1070

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, 166–180. doi:10.1593/neo.07112

Schneider, K. U., Dietrich, D., Fleischhacker, M., Leschber, G., Merk, J., Schäper, F., et al. (2011). Correlation of SHOX2 Gene Amplification and DNA Methylation in Lung Cancer Tumors. *BMC Cancer.* 11, 102. doi:10.1186/1471-2407-11-102

Shi, J., Chen, X., Zhang, L., Fang, X., Liu, Y., Zhu, X., et al. (2020). Performance Evaluation of SHOX2 and RASSF1A Methylation for the Aid in Diagnosis of Lung Cancer Based on the Analysis of FFPE Specimen. *Front. Oncol.* 10, 565780. doi:10.3389/fonc.2020.565780

Srisuttee, R., Arayataweegool, A., Mahattanasakul, P., Tangjaturonrasme, N., Kerekhanjanarong, V., Keelawat, S., et al. (2020). Evaluation of NID2 Promoter Methylation for Screening of Oral Squamous Cell Carcinoma. *BMC Cancer* 20, 218. doi:10.1186/s12885-020-6692-z

Sun, C., Liu, X. H., and Sun, Y. R. (2019). MiR-223-3p Inhibits Proliferation and Metastasis of Oral Squamous Cell Carcinoma by Targeting SHOX2. *Eur. Rev. Med. Pharmacol. Sci.* 23, 6927–6934. doi:10.26355/eurrev_201908_18732

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. *Nucl. Acids Res.* 45, W98–w102. doi:10.1093/nar/gkx247

Wang, Y., Zhang, L., Yang, J., Li, B., and Wang, J. (2018). CDH13 Promoter Methylation Regulates Cisplatin Resistance of Non-small Cell Lung Cancer Cells. *Oncol. Lett.* 16, 5715–5722. doi:10.3892/ol.2018.9325

Weiss, G., Schlegel, A., Kottwitz, D., König, T., and Tetzner, R. (2017). Validation of the SHOX2/PTGER4 DNA Methylation Marker Panel for Plasma-Based Discrimination between Patients with Malignant and Nonmalignant Lung Disease. *J. Thorac. Oncol.* 12, 77–84. doi:10.1016/j.jtho.2016.08.123

Wu, C., Li, M., Meng, H., Liu, Y., Niu, W., Zhou, Y., et al. (2019). Analysis of Status and Countermeasures of Cancer Incidence and Mortality in China. *Sci. China Life Sci.* 62, 640–647. doi:10.1007/s11427-018-9461-5

Xu, Z., Wang, Y., Wang, L., Xiong, J., Wang, H., Cui, F., et al. (2020). The Performance of the SHOX2/PTGER4 Methylation Assay Is Influenced by Cancer Stage, Age, Type and Differentiation. *Biomarkers Med.* 14, 341–351. doi:10.2217/bmm-2019-0325

Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., and Liang, G. (2014). Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell* 26, 577–590. doi:10.1016/j.ccr.2014.07.028

Zhang, C., Yu, W., Wang, L., Zhao, M., Guo, Q., Lv, S., et al. (2017). DNA Methylation Analysis of the SHOX2 and RASSF1A Panel in Bronchoalveolar Lavage Fluid for Lung Cancer Diagnosis. *J. Cancer* 8, 3585–3591. doi:10.7150/jca.21368

Zhang, C., Zhao, N., Zhang, X., Xiao, J., Li, J., Lv, D., et al. (2020) SurvivalMeth: a Web Server to Investigate the Effect of DNA Methylation-Related Functional Elements on Prognosis, *Brief Bioinform.* 22(3):bbaa162. doi:10.5194/essd-2020-182-rc1

Zhang, Y.-A., Zhou, Y., Luo, X., Song, K., Ma, X., Sathe, A., et al. (2016). SHOX2 Is a Potent Independent Biomarker to Predict Survival of WHO Grade II-III Diffuse Gliomas. *EBioMedicine* 13, 80–89. doi:10.1016/j.ebiom.2016.10.040

Zhang, Y., Breitling, L. P., Balavarca, Y., Holleczek, B., Schöttker, B., and Brenner, H. (2016). Comparison and Combination of Blood DNA Methylation at Smoking-Associated Genes and at Lung Cancer-Related Genes in Prediction of Lung Cancer Mortality. *Int. J. Cancer* 139, 2482–2492. doi:10.1002/ijc.30374

Zhao, Q. T., Guo, T., Wang, H-E., Zhang, X-P., Zhang, H., Wang, Z-K., et al. (2015). Diagnostic Value of SHOX2 DNA Methylation in Lung Cancer: a Meta-Analysis, *Onco Targets Ther.* 8, 3433–3439. doi:10.2147/ott.S94300

Zweig, A. S., Karolchik, D., Kuhn, R. M., Haussler, D., and Kent, W. J. (2008). UCSC Genome Browser Tutorial. *Genomics* 92, 75–84. doi:10.1016/j.ygeno.2008.02.003

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.

Copyright © 2021 Li, Zeng, Tai, Lin, Zhu, Luo, Ren, Zhu, Li, Wu 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 in Molecular Biosciences

ORIGINAL RESEARCH
published: 01 July 2021
doi: 10.3389/fmolb.2021.700973

ANLN Enhances Triple-Negative Breast Cancer Stemness Through TWIST1 and BMP2 and Promotes its Spheroid Growth

Alishba Maryam[1] *and Y. Rebecca Chin*[1,2]*

[1]Department of Biomedical Sciences, City University of Hong Kong, Kowloon, Hong Kong, [2]Key Laboratory of Biochip Technology, Biotech and Health Centre, Shenzhen Research Institute, City University of Hong Kong, Shenzhen, China

ANLN is frequently upregulated in triple-negative breast cancer (TNBC) and its high expression in tumors are significantly associated with poor survival and recurrence, thereby it has been proposed to function as a prognostic marker for breast cancer. However, the specific function and molecular mechanisms by which ANLN promotes TNBC tumorigenesis remain elusive. Using multiomic profiling, we recently uncovered ANLN as a TNBC-specific gene driven by super-enhancer. Here, by Crispr/Cas9 editing, we showed that knockout of ANLN inhibits spheroid growth of TNBC. Interestingly, its effect on cell proliferation in 2D cultures is minimal. ANLN depletion inhibits mammosphere formation and clonogenicity potently, suggesting its important function in regulating cancer stem cells (CSCs). We screened a panel of stem cell-related genes and uncovered several CSC genes regulated by ANLN. We further identify TWIST1 and BMP2 as essential genes that mediate ANLN's function in stemness but not spheroid growth. These findings may contribute to search for effective targeted therapies to treat TNBC.

Keywords: Anilin, breast cancer, cancer stem cells, 3D culture, CRISPR/Cas9

OPEN ACCESS

Edited by:
Tianhui Chen,
Zhejiang Cancer Hospital of the University of Chinese Academy of Sciences, China

Reviewed by:
Tamer Saad Kaoud,
University of Texas at Austin, United States
Mohammad Imran Khan,
King Abdulaziz University, Saudi Arabia

***Correspondence:**
Y. Rebecca Chin
rebecca.chin@cityu.edu.hk

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *27 April 2021*
Accepted: *18 June 2021*
Published: *01 July 2021*

Citation:
Maryam A and Chin YR (2021) ANLN Enhances Triple-Negative Breast Cancer Stemness Through TWIST1 and BMP2 and Promotes its Spheroid Growth. Front. Mol. Biosci. 8:700973. doi: 10.3389/fmolb.2021.700973

INTRODUCTION

Breast cancer is the most prevalent cancer in women worldwide, with an estimated 2 million new cases diagnosed in 2020 (Sung et al., 2021). Conventionally, breast cancer is classified into three major subtypes: luminal, HER2 overexpressed, and basal-like. Approximately 70% of basal-like tumors are triple-negative breast cancer (TNBC; $ER^-/PR^-/HER2^-$) (Cancer Genome Atlas, 2012), which is more aggressive, enriched in cancer stem cells (CSCs) and prone to metastasis. Chemotherapy, usually with high toxicity, remains the major treatment option for TNBC, thus there is an urgent clinical need to identify novel therapeutic targets for this aggressive subtype. Over the past years, genomic and transcriptomic studies have revealed genetic features of different subtypes of breast cancer. However, epigenetic mechanisms in regulating oncogene expression in breast cancer subtypes remain poorly understood.

Enhancers are cis-regulatory regions that promote transcription of a target gene from a distance (Zhang et al., 2016). Super-enhancers contain clusters of enhancers that were originally shown to define tissue specificity (Herranz et al., 2014). They are characterized by high level of the histone modifications such as H3K27ac and coactivator binding. Increasing evidence has demonstrated the critical role of super-enhancers in promoting cancer development by driving

expression of oncogenes (Hnisz et al., 2013). For example, MYC-associated super-enhancer was shown to be focally amplified in multiple epithelial cancers to drive MYC expression in tumors (Zhang et al., 2016). In gastric cancer, super-enhancer landscape was reprogramed during tumorigenesis, underpinning the dysregulation of cancer-related genes (Ooi et al., 2016). The enrichment of super-enhancers in different subtypes of breast cancer, however, has not been determined. To address heterogeneity of super-enhancers and their functional role in breast cancer, we recently reported a multiomic study employing ChIPseq profiling, gene expression data and network-based analyses to uncover TNBC-specific super-enhancer landscape (Huang et al., 2021). Using Crispr/Cas9 editing, we further demonstrated the functional significance of super-enhancer in driving the expression of key oncogenes FOXC1 and MET in TNBC. Harnessing the power of exploring epigenomic features in TNBC, a number of novel super-enhancer-regulated TNBC-specific genes were emerged from the study, including Anilin (ANLN).

ANLN has been demonstrated to play a key role in cytokinesis. During interphase of the cell cycle, ANLN is resided in the nucleus. Whereas during telophase, ANLN accumulates in the cytoplasm, forming a contractile ring and cleavage furrow by interacting with various proteins including myosin, F-actin, RhoA and septin (Oegema et al., 2000; Piekny and Maddox, 2010). Upregulation of ANLN is observed in various cancers including lung, pancreatic, ovarian, colorectal, hepatic and breast cancer (Magnusson et al., 2016). Studies have demonstrated that ANLN promotes tumor cell proliferation by regulating cell cycle progression. For example, in ANLN-depleted non-small lung cancer and breast cancer lines, polynucleated cells were observed and cell proliferation was inhibited (Suzuki et al., 2005; Zhou et al., 2015; Magnusson et al., 2016). In addition, breast cancer patients with high expression of ANLN showed significantly poorer overall survival (Magnusson et al., 2016), recurrence as well as higher expression of proliferation genes (all 17 tested in (Thapa and Wilson, 2016)). ANLN is thereby proposed to serve as a prognostic marker for breast cancer. The specific function of ANLN in TNBC and the molecular mechanism by which ANLN is being regulated, however, is elusive. Our recent study provided direct evidence that TNBC-specific super-enhancer drives ANLN expression, and promotes TNBC clonogenicity (Huang et al., 2021), a sensitive indicator of undifferentiated CSCs. Owing to their clinical implications in metastasis, drug resistance and aggressiveness of tumor (Korkaya et al., 2012; Batlle and Clevers, 2017), it is important to dissect mechanisms of CSC maintenance. In the present study, we extend our effort to investigate the role of ANLN in TNBC spheroid growth and stemness. Using Crispr/Cas9 genomic editing, we delete ANLN and find a significant reduction of spheroid and clonogenic growth of TNBC, as well as impairment of mammosphere formation. Our findings further reveal the critical function of TWIST1 and BMP2 in mediating ANLN's effect on TNBC stemness.

MATERIALS AND METHODS

Cell Culture

HEK293T, BT-549 and Hs578T cells were obtained from ATCC. BT549 was maintained in RPMI 1640 medium (Gibco) supplemented with 10% FBS. HEK293T cells were cultured in Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% FBS. Hs578t cells were maintained in DMEM supplemented with 10% FBS and 10 μg/ml insulin. All cell lines obtained from cell banks listed above are tested for authentication using short tandem repeat profiling and passaged for fewer than 6 months, and routinely assayed for mycoplasma contamination.

Antibodies

Anti-ANLN (#AMAB90660) and anti-actin (#3700) antibodies were obtained from Cell Signaling Technology. Anti-HA tag antibody (#11867423001) was obtained from Roche. Anti-TWIST1 (#A3237), and anti-BMP2 (#A0231) antibodies were purchased from ABclonal. Horse peroxidase-conjugated anti-mouse and anti-rabbit immunoglobulin G (IgG) antibodies (AP307P, AP308P) were obtained from Millipore. Horse peroxidase-conjugated anti-rat IgG (#AS028) was obtained from ABclonal.

Plasmids

To knockout ANLN, Crispr/Cas9 inducible knockout system was used. FUCas9Cherry (#70182) and FgH1tUTG (#70183) were ordered from Addgene. Guide RNAs (gRNAs) were designed using online tools http://crisper.mit.edu/ and http://crisprscan.org. gRNAs oligos (**Supplementary Table S1**) with sticky end were synthesized by IDT company. The gRNAs were cloned into restriction BsmBI restriction sites of FgH1tUTG vector. For overexpression of exogenous BMP2-HA, CDS of BMP2 with HA tag at C-terminal was synthesized and cloned into vector CD532A-1 by GENEWIZ. To overexpress HA-TWIST1, CDS of TWIST1 with HA-tag at N-terminal were synthesized and cloned into vector CD532A-1 by GENEWIZ. For peak deletion of super-enhancer, a pair of gRNAs flanking the peak were designed using http://crisper.mit.edu/ (**Supplementary Table S1**). The pair of gRNAs were then inserted sequentially into BsaI and BbsI restriction sites of pX333 vector (Addgene #64073) that encodes spCas9 and two gRNA cassettes.

Lentivirus Infection

To prepare lentiviral supernatants, 6.3 μg (FgH1tUTG and FUCas9Cherry), 7.6 μg (HA-TWIST1) or 8.2 μg (HA-BMP2) of lentiviral vectors were co-transfected with 7 μg of psPAX2 and 2.4 μg of VSV-G vectors to HEK293T cells using polyethylenimine as transfection reagent. The lentiviruses were filtered and collected after 72 h of transfection, using 0.45 μm syringe filter (Thermo fisher 7232545). 0.5 ml of lentivirus with 2 μg/ml polybrene were added to breast tumor cells for 12–24 h in a well of 6-well plates. Cells were sorted by Florescent-activated cell sorting (FACS) with cell sorter (Sony) or selected with puromycin for 5–7 days.

Immunoblotting

Cells were washed with PBS at 4°C and lyzed in EBC buffer (0.5 NP-40, 120 mM NaCl, 50 mM Tris-HCl (pH 7.4), proteinase inhibitor cocktail, 50 mM calyculin, 1 mM sodium pyrophosphate, 20 mM sodium fluoride, 2 mM EDTA, 2 mM EGTA) for 25 min on ice. Cell extracts were pre-cleared by centrifugation at 13,000 × g for 10 min at 4°C and protein concentration was measured with Bio-Rad protein assay reagent using BioTek Synergy ™ H1 Microplate Reader. Lysates were then resolved on 10% acrylamide gels by SDS-PAGE and transferred electrophoretically to nitrocellulose membrane (Bio-Rad) at 160 mA for 80 min. The blots were blocked in TBST buffer (10 mM Tris-HCl, pH 8, 150 mM NaCl, 0.2% Tween 20) containing (w/v) non-fat dry milk 30 min, and then incubated with the specific primary antibody diluted in blocking buffer at 4°C overnight. Membranes were washed three times in TBST and incubated with horseradish peroxidase-conjugated secondary antibody for 1 h at room temperature. Membranes were washed 3 times and developed using enhanced chemiluminescence substrate (Pierce).

Clonogenic Growth Assays

Cells were seeded to 6-well plate at a density of 800 cells/well and cultured for 10 days. Medium was changed every 4 days. After 10 days, cells were fixed with 4% formaldehyde for 15 min at room temperature. 0.1% crystal violet was then used to stain colonies for 40 min followed by washing with PBS. Images were captured and the colony number was counted.

CellTiter-Glo® 3D and 2D Cell Viability Assays

3D cultures were prepared as previously described (Debnath et al., 2003). Briefly, 96-well plates (Corning #3610) were coated with growth factor-reduced Matrigel (BD Biosciences) and allowed to solidify for 30 min. The cells were seeded in assay medium on Matrigel-coated 96-well plate, with cell density of 2,500–4,000 cells per well. Assay medium contained DMEM/RPMI-1640 supplemented with 10% Tetracycline Free FBS and 2% Matrigel for BT549. Assay medium for Hs578T contained DMEM supplemented with 10% Tetracycline Free FBS, insulin (10 μg/ml) and 2% Matrigel. Cells coated with Matrigel were then allowed to interact with extracellular matrix and grow in 5% CO_2 humidified incubator at 37°C to form spheroids in 7–11 days. The assay medium was replaced every 4 days. Cells were treated with 100 ng/ml doxycycline every 2 days to induce ANLN knockdown. To quantify spheroid growth and viability, CellTiter-Glo® 3D Cell Viability Assay (Promega #G9682) was performed by following instructions on product manual. To assess cell viability in 2D culture, CellTiter- Glo® Luminescent Cell Viability Assay (Promega #G7571) was used. 3,000 cells per well were seeded to 96-well plate (Corning #3610) and cultured for 3 days 50 μL reagent was then added to each well and incubated on a shaker for 15 min. Signal was read by a Synergy™ H1 Microplate Reader (BioTek).

Mammosphere Formation Assay

Cells were seeded to ultra-low attachment 6-well plates (Corning 3,471) with cell density of 2,500 and 3,000 cells per well for BT549 and HS578t, respectively. Cells were cultured in mammosphere medium, containing DMEM/F12 supplemented with B27 (Gibco 12587010) and 20 ng/ml EGF (R&D 236-EG), for 7–8 days. Images of mammospheres were captured by Nikon NIS-Elements D software. Mammosphere with diameter >50 μm were counted using the Nikon NIS-Elements D software.

RT-qPCR

Mammosphere samples were cultured with mammosphere medium in ultra-low attachment 10-cm plates for 7 days. Total RNA from mammosphere cultures was extracted using RNeasy Plus Mini Kit (Qiagen #74134) following the manufacturer's instructions. Reverse transcription was performed using TaqMan Reverse transcription Reagents (Applied Biosystems, N8080234). Quantitative RT-PCR was performed using QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems).

RESULTS

Depletion of Anilin Inhibits Triple-Negative Breast Cancer Spheroid Growth

To explore the function of ANLN in growth and stemness of TNBC, we generated a panel of breast cancer lines (BT549 and Hs578T) with tet-on doxycycline (dox)-inducible Crispr/Cas9-mediated knockout of ANLN. TNBC cells were infected with lentiviral vectors, followed by florescent-activated cell sorting for double-positive cells containing both GFP (dox-inducible gRNA) and mCherry (spCas9). Upon dox administration, ANLN was depleted significantly with two distinct gRNAs (**Figure 1A**). We then investigated the consequence of ANLN knockout on TNBC cell proliferation. The 3D spheroid morphogenesis assay, that more accurately recapitulates phenotypes governing tumor growth *in vivo,* was performed. As shown in the representative images and bar graphs, depletion of ANLN in TNBC lines inhibited spheroid growth significantly in 3D cultures (BT549: gRNA1 93% inhibition, gRNA2 91% inhibition; Hs578T: gRNA1 63% inhibition, gRNA2 79% inhibition). Conversely, dox administration in vector-control cells had no effect (**Figure 1B**). We have also quantified viability of cells in 3D spheroids, and shown that knockout of ANLN resulted in significant reduction of cell viability (**Figure 1B**). Interestingly, ANLN depletion only led to 20% inhibition of BT549 cell viability in 2D, and it had no effect on Hs578T cells in 2D cultures (**Figure 1C**). As our recent paper demonstrated the regulation of ANLN expression by super-enhancer in TNBC, we examined the effect of super-enhancer on spheroid growth. Deletion of ANLN-associated super-enhancer by Crispr/Cas9 editing resulted in reduction of ANLN protein expression as well as impairment of spheroid growth (**Figure 1D**), agreeing with the functional significance of super-enhancer in promoting ANLN expression and its associated tumorigenic phenotype.

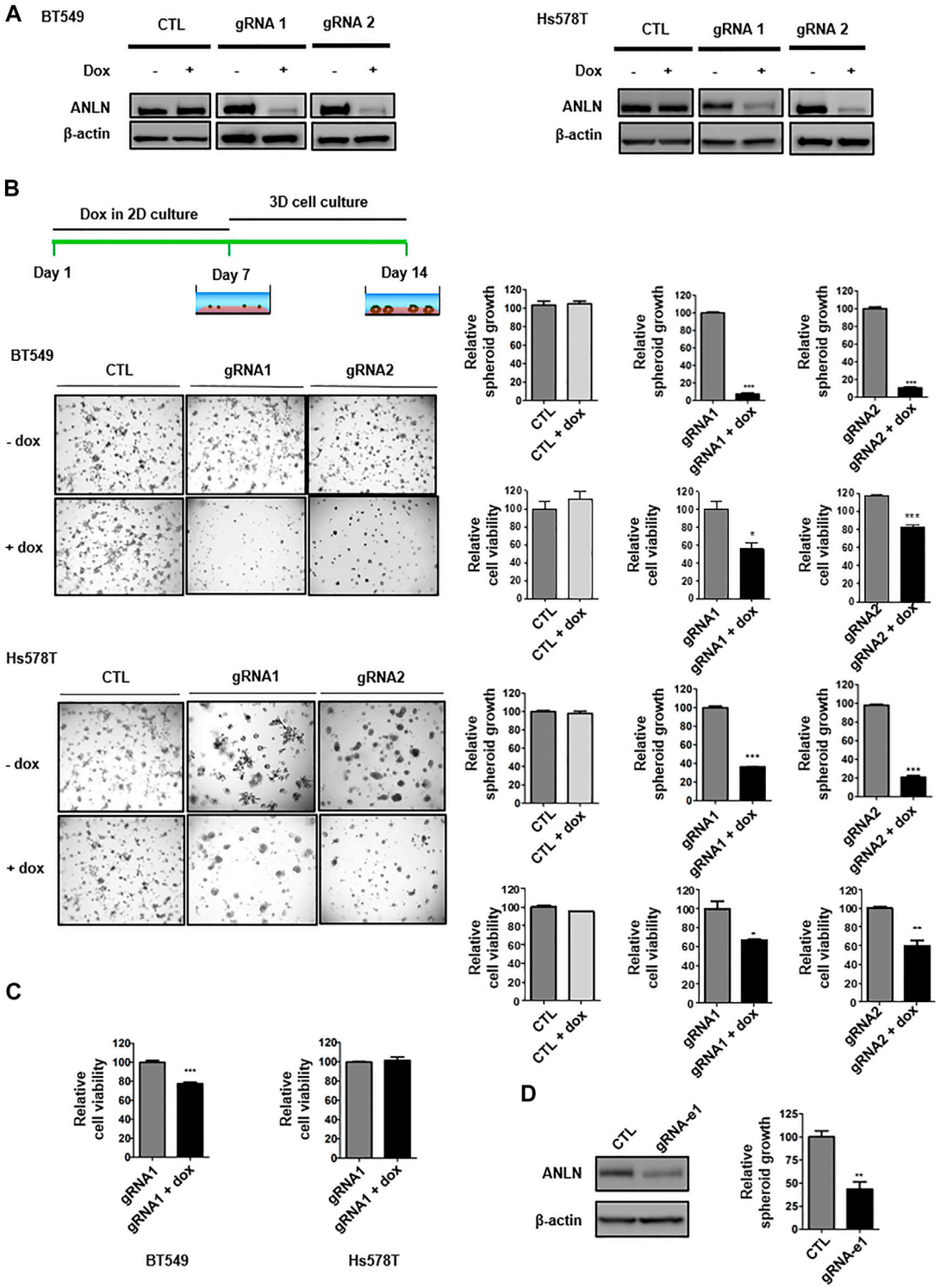


FIGURE 1 | ANLN depletion attenuates TNBC spheroid growth. **(A)** BT549 and Hs578T cells expressing tet-on ANLN gRNA or vector control (CTL) were treated with doxycycline (Dox; 100 ng/ml) for 7 days. Whole-cell lysates were subjected to immunoblotting. Experiments were repeated at least 3 times with similar results. **(B)** Schematics of dox treatment and 3D culture. BT549 and Hs578T cells were infected with tet-on ANLN or CTL gRNA. Cells were cultured in 3D for 7–9 days, and images were captured. Area of spheroids was measured using NIS-Elements D software. Bar graphs depict growth of spheroids with or without ANLN knockout. Error bars, mean ± SEM of 3 independent experiments. Cell viability was assessed by 3D cell titer-Glo assay, and depicted in the bar graphs ($n = 3$). Cell viability results are representative of 3 independent experiments. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. **(C)** BT549 and Hs578T cells were infected with tet-on ANLN or CTL gRNA. Cells were cultured in 2D for 3–5 days, followed by cell titer-Glo Luminescent cell viability assay; Bar graphs depict growth of BT549 and Hs578T cells with or without ANLN knockout. Error bars, mean ± SEM of 3 independent experiments. ***, $p < 0.001$. **(D)** Immunoblotting detection of ANLN in BT549 upon deletion of e1 super-enhancer of SSE256, experiment was repeated twice with similar results. Bar graph depicts growth of spheroid with and without deletion of e1 in BT549 cells. Error bars, mean ± SEM of 3 independent experiments. **, $p < 0.01$.

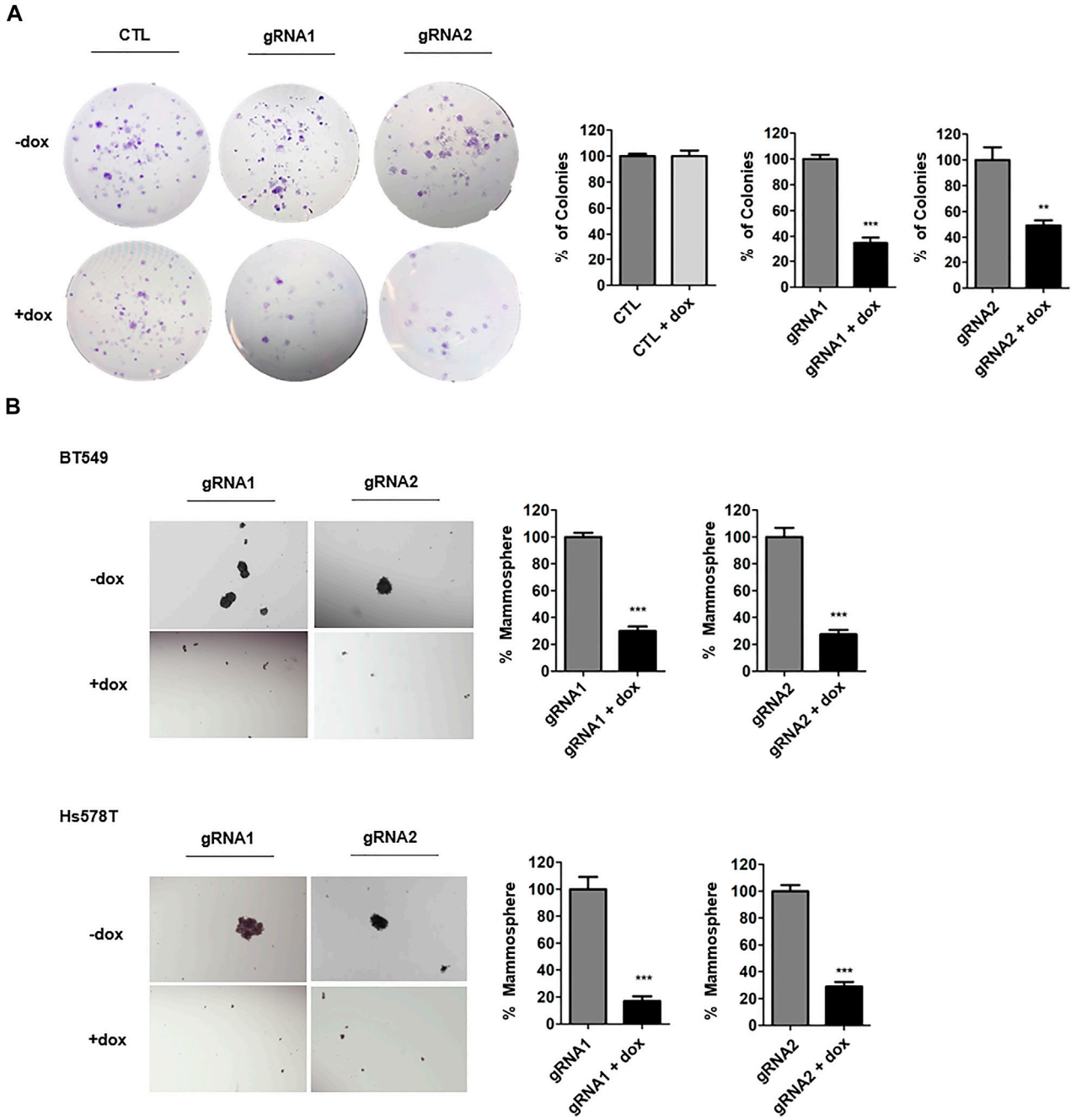


FIGURE 2 | ANLN promotes stem cell property of TNBC. **(A)** BT549 cells expressing tet-on ANLN or CTL gRNA were treated with dox (100 ng/ml) for 7 days. Cells were then cultured for colony formation assay for 10 days. Representative images are shown. Colony number was counted and depicted in bar graphs. Error bars, mean ± SEM of 3 independent experiments. ***, $p < 0.001$. **(B)** BT549 and Hs578T cells expressing tet-on ANLN or CTL gRNA were treated with dox (100 ng/ml) for 7 days. Cells were then seeded for mammosphere formation assay. Representative images are shown. Bar graphs depict the mammosphere number. Error bars, mean ± SEM of 3 independent experiments. ***, $p < 0.001$.

Knockout of Anilin Reduces Stemness of Triple-Negative Breast Cancer Cells

To examine progeny producing capability, clonogenic assays were performed. The ANLN-knockout TNBC cells were seeded at low density and then cells were allowed to grow for 10 days. Using crystal violet dye, cells were stained. We showed that ANLN knockout greatly reduced the colony formation abilities of BT549 and Hs578T cells (**Figure 2A**, **Supplementary Figure S1**). In cells containing control gRNA, dox treatment had no effect on colony numbers. Next, we

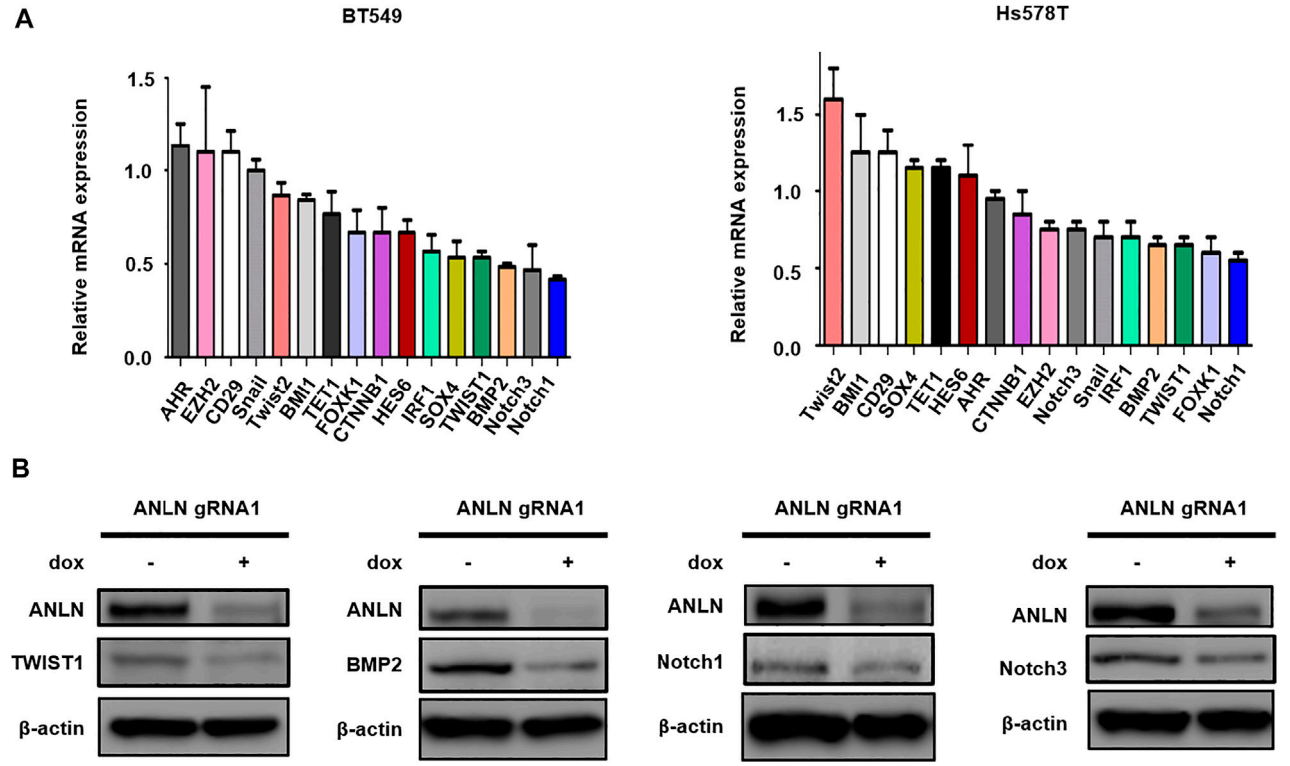


FIGURE 3 | Effect of ANLN on the expression of stemness-related genes. **(A)** Bar graphs depicting mRNA expression of stemness-related genes in dox-treated (ANLN-knockout) mammospheres compared to mammospheres without dox treatment. Error bars, mean ± SEM of 3 independent experiments. **(B)** Immunoblot showing expression level of Twsit1, Bmp2, Notch1 and Notch3 proteins in Hs578T cells with and without ANLN knockout. Experiments were repeated twice independently with similar results.

examined the functional contribution of ANLN on stemness of TNBC cells, by performing mammosphere formation assay. In mammosphere assay, only anoikis-resistant CSCs survive in suspension in the specific culture medium. The number of spheres present in culture reflects the number of cells which are capable of forming new tumor spheroids (Soule and McGrath, 1986; Dontu et al., 2003). TNBC cells were treated with dox for 7 days to induce ANLN knockout and then grown in mammosphere culture medium for 7–9 days. Spheres larger than 50 μm were considered as mammospheres. Using two independent gRNAs, we showed that ANLN-depleted samples (dox treated) had significantly lower number of mammospheres (**Figure 2B**; BT549: gRNA1 70% inhibition, gRNA2 72% inhibition; Hs578T: gRNA1 83% inhibition, gRNA2 71% inhibition). These data suggest that depletion of ANLN results in suppressed CSC properties of TNBC cells.

Anilin Promotes Mammosphere Formation via TWIST1 and BMP2

To dissect the mechanism by which ANLN regulates stemness in TNBC cells, a panel of stem cell-related genes were screened using RT-qPCR. **Figure 3A** and **Supplementary Table S2** show the mRNA levels of CSC-related genes in ANLN-depleted mammospheres. Among these, 4 of them (TWIST1, BMP2, Notch1, Notch3) were inhibited to the greatest extent in ANLN-depleted BT549 mammospheres. Decreased expression of these 4 proteins in ANLN-knockout cells was also confirmed by immunoblot analysis (**Figure 3B**). As TWIST1 and BMP2 have been shown to promote CSC properties in breast cancer (Vesuna et al., 2009; Li and Zhou, 2011; Huang et al., 2017), we overexpressed these two genes to determine the role of them in mediating ANLN's function. The overexpression of these proteins was confirmed by immunoblot analysis in TNBC cells (**Figure 4**). Next, we investigated the effect of TWIST1 on the regulation of CSC properties using mammosphere formation assay. Overexpression of HA-TWIST1 rescued the mammosphere formation ability in ANLN-depleted BT549 and Hs578T cells (**Figure 4A**). Similar results were observed for HA-BMP2 overexpression (**Figure 4B**). However, overexpression of TWIST1 or BMP2 did not rescue spheroid growth of ANLN-depleted cells (**Supplementary Figure S2**). These data indicated that TWIST1 and BMP2 mediate, at least in part, the function of ANLN in promoting stemness of TNBC cells.

DISCUSSION

By leveraging the TNBC-specific epigenomic data, we recently identified ANLN as a TNBC-specific gene

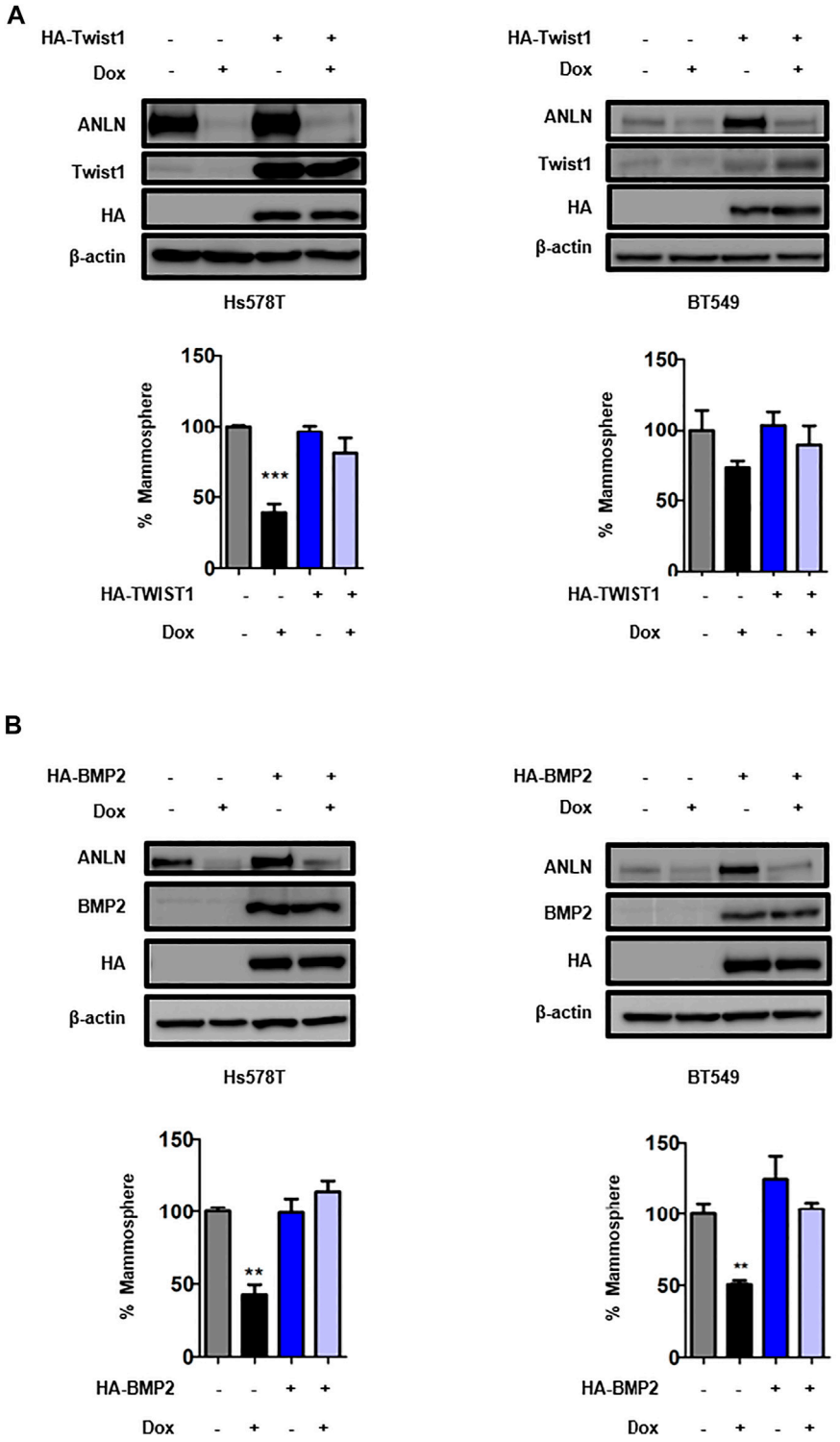


FIGURE 4 | TWIST1 and BMP2 are critical downstream effectors of ANLN in regulating TNBC CSCs. **(A)** Hs578t and BT549 cells expressing HA-Twist1 or control vector were infected with tet-on ANLN gRNA1. Cells were treated with dox (100 ng/ml) for 7 days, and then subjected to mammosphere formation assay. Error bars, mean ± SEM of 3 independent experiments. Whole cell lysates were subjected to Immunoblotting analysis. ***, $p < 0.001$. **(B)** Hs578t and BT549 cells expressing HA-Bmp2 or control vector were infected with tet-on ANLN gRNA1. Cells were treated with dox (100 ng/ml) for 7 days, and then subjected to mammosphere formation assay. Error bars, mean ± SEM of 3 independent experiments. Whole cell lysates were subjected to Immunoblotting analysis. **, $p < 0.01$.

regulated by super-enhancer (Huang et al., 2021). In the present study, we examined the functional role of ANLN in TNBC by knocking out the gene using Crispr/Cas9 approach. We first examined the effect of ANLN depletion on spheroid growth in 3D cultures. Our 3D system contains Matrigel, which is enriched in basement membrane components including laminin, collagen and entactin. These components resemble closely to the complex extracellular environment present in breast tumors (Badea et al., 2019), and our results demonstrated a potent impairment of spheroid growth upon ANLN depletion. Interestingly, when these ANLN-depleted cells were grown in 2D cultures, the effect of ANLN knockout on cell proliferation was minimal. Our data agree with a recent study on breast cancer, where neither ANLN knockdown nor overexpression affected cell proliferation of breast cancer cells grown in 2D *in vitro* (Wang et al., 2020). Similarly, previous studies by us and others have reported differences in proliferation and metabolic capability between 3D and 2D cultured cells. In prostate cancer, whereas silencing Akt2 in spheroids and xenografts induces robust apoptosis, knocking down Akt2 in 2D culture has minimal effect on cell survival (Chin et al., 2014). In colon cancer, 3D spheroids display low activities of mTOR, S6K and Akt signaling pathways compared to cells grown in 2D cultures (Riedl et al., 2017), suggesting distinct signaling rewiring in 3D environment. Since ANLN interacts with RhoA, which function has been shown to be modulated directly by interaction with the extracellular matrix (Lim et al., 2010), it is possible that the potent effect of ANLN on TNBC cell proliferation in 3D cultures is contributed by the interaction of extracellular matrix proteins with intracellular actin machinery and ANLN. In recent years, there have been advances in the development of 3D culture systems for mechanistic studies as well as drug screening. Commonly used models include scaffold-free, scaffold-based and hybrid 3D systems, each has their own merits and limitations (Langhans, 2018). Scaffold-free model, relying on self-aggregation of cells, allows formation of more uniform spheroids but not readily mimics cell-extracellular matrix interactions. On the other hand, scaffold-based system with Matrigel provides a good mimic of *in vivo* matrix environment yet may have the issue of batch-to-batch variability. Our laboratory routinely uses the Matrigel overlay method (cells embedded in Matrigel) as well as method involved seeding cells in medium/2% Matrigel to plates pre-treated with ultra-low attachment coating, and find no difference in spheroid morphology between the two methods (Chin et al., 2014). It would be interesting to perform co-culture 3D experiments in microplates which would allow assessment of ANLN function in a complex tumor microenvironment.

The role of ANLN in cancer stemness has not been well-studied. Nevertheless, ANLN is implicated in self-renewal of progenitor cells in developing zebrafish. A study reported that during asymmetric cell division of retinal ganglion cells, ANLN plays a critical role in progenitor cell self-renewal and balances the asymmetric and symmetric outcomes that are important for correct neurogenesis in the retina (Paolini et al., 2015). Recently, Wang et. al. has shown an important role of ANLN in promoting breast cancer stemness (Wang et al., 2020). The authors further performed RNAseq analysis in ANLN-depleted cells and showed that a number of known regulators of stemness and differentiation such as OVOL2, TBX18, FOXK1, SOX-9 and PBX1 were downregulated. The authors proposed that these genes may be involved in ANLN-mediated stemness, but there is lack of experimental support for the connections. We, therefore, first investigated the role of ANLN in CSC regulation in TNBC by performing mammosphere formation assay. Significant decrease in mammosphere numbers was observed in ANLN-depleted cells, indicating a prominent function of ANLN in CSC regulation. These results prompted us to explore the unknown mechanism of CSC regulation by ANLN. Using RT-qPCR, we identified several stem cell-related genes including TWIST1 and BMP2 that were suppressed in ANLN-depleted cells. In rescue experiments, we further demonstrated that TWIST1 and BMP2 successfully restored the mammosphere formation ability of ANLN-depleted TNBC cells. Interestingly, TWIST1 and BMP2 did not rescue spheroid growth in ANLN-depleted cells. As mammosphere assay medium enriches CSC survival and growth, whereas 3D culture with Matrigel allows TNBC cells grow and proliferate by interacting with extracellular matrix, our data support the role of TWIST1 and BMP2 as effectors of ANLN for promoting CSC properties that have not been reported previously. In addition to TWIST1 and BMP2, our finding suggests that ANLN regulates expression of other stemness-related genes including Notch1 and Notch3. Whether these genes mediate ANLN's function in CSCs and tumorigenesis await further investigation.

TWIST1 is a transcription factor that plays an important role in driving the process of epithelial mesenchymal transition (EMT) during development and in cancer (Yang et al., 2004). Various studies have also demonstrated the role of TWIST1 in cancer stemness. Overexpression of TWIST1 in breast cancer cell lines were shown to promote the ability of mammosphere to self-renew *in vitro* as well as promote tumor initiation ability in immunodeficient mice (Mani et al., 2008; Morel et al., 2008). Another study showed that transient activation of TWIST1 is responsible for promoting stemness without induction of EMT (Jung and Yang, 2015). Upstream of TWIST1, Metadherin (MTDH) was shown to epigenetically activate TWIST1 to promote stem-like traits in breast cancer (Liang et al., 2015). MTDH is transmembrane protein that contributes to the growth, metastasis, drug resistance and relapse of tumor. It facilitates histone 3 acetylation of the TWIST1 promotor, by interacting with and preventing degradation of histone acetyltransferase CBP. Interestingly, MTDH silencing in gastric cancer cells regulates actin cytoskeletal remodeling (Du et al., 2017). Given the link between ANLN and actin cytoskeleton, it is possible that ANLN interacts with MTDH, which in turn promotes the transcription of TWIST1 in TNBC.

Bone morphogenetic proteins (BMPs) belong to the superfamily of transformation growth factor TGF-β. They were originally identified for their function in osteogenesis and bone turnover. Since then, extensive studies have investigated their role in various cancers (Hogan, 1996; Hou et al., 2009; Liao et al., 2015). BMP signaling has been demonstrated to promote or inhibit cancer growth

and progression in different tumor contexts. In renal carcinoma and osteosarcoma, BMP2 inhibits cancer stemness (Wang et al., 2011; Wang et al., 2015). In contrast, the positive role of BMP2 in CSC regulation was observed in glioblastoma and colon cancer (Piccirillo and Vescovi, 2007; Kim et al., 2015). Our study indicates that BMP2 promotes TNBC stemness, which is concordant with another study which shows a critical function of BMP2 in promoting breast cancer EMT and stemness through Rb signaling pathway (Huang et al., 2017). The molecular mechanism by which BMP2 regulates CSCs in TNBC remains to be determined. In colon cancer, knockdown of STAT3 reversed BMP2-induced CSC formation (Kim et al., 2015). Results from our qPCR experiments show downregulation of STAT3 in ANLN-depleted TNBC cells. It will be interesting to test if ANLN regulates TNBC CSCs via BMP2-STAT3 signaling axis in future studies. Taken together, the present study identified an important function of ANLN in promoting spheroid growth of TNBC. In addition, we established an integrated mechanism by which ANLN induces TNBC stemness via TWIST1 and BMP2. These findings may be helpful for developing therapeutic strategies against TNBC.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

AM designed the experiments, performed the data collection and analysis, and wrote the manuscript. YC advised on the experimental design, assisted with manuscript writing and supervised the study.

FUNDING

This study was supported by the National Natural Science Foundation of China (81972781 to YC,) General Research Fund (11101517 to YC) and Research Impact Fund (R1020-18 to YC.) from the Research Grants Council of the Hong Kong Special Administrative Region, Shenzhen Science and Technology Innovation Commission (JCYJ20170818095453642 to YC.), and the City University of Hong Kong (7005517 to YC.)

ACKNOWLEDGMENTS

The authors thank members of the Chin laboratory for discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.700973/full#supplementary-material

REFERENCES

Badea, M. A., Balas, M., Hermenean, A., Ciceu, A., Herman, H., Ionita, D., et al. (2019). Influence of Matrigel on Single- and Multiple-Spheroid Cultures in Breast Cancer Research. *SLAS DISCOVERY: Advancing Sci. Drug Discov.* 24 (5), 563–578. doi:10.1177/2472555219834698

Batlle, E., and Clevers, H. (2017). Cancer Stem Cells Revisited. *Nat. Med.* 23 (10), 1124–1134. doi:10.1038/nm.4409

Cancer Genome Atlas, N. (2012). Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* 490 (7418), 61–70. doi:10.1038/nature11412

Chin, Y. R., Yuan, X., Balk, S. P., and Toker, A. (2014). PTEN-deficient Tumors Depend on AKT2 for Maintenance and Survival. *Cancer Discov.* 4 (8), 942–955. doi:10.1158/2159-8290.Cd-13-0873

Debnath, J., Muthuswamy, S. K., and Brugge, J. S. (2003). Morphogenesis and Oncogenesis of MCF-10A Mammary Epithelial Acini Grown in Three-Dimensional Basement Membrane Cultures. *Methods* 30 (3), 256–268. doi:10.1016/s1046-2023(03)00032-x

Dontu, G., Abdallah, W. M., Foley, J. M., Jackson, K. W., Clarke, M. F., Kawamura, M. J., et al. (2003). *In Vitro* propagation and Transcriptional Profiling of Human Mammary Stem/progenitor Cells. *Genes Dev.* 17 (10), 1253–1270. doi:10.1101/gad.1061803

Du, Y., Jiang, B., Song, S., Pei, G., Ni, X., Wu, J., et al. (2017). Metadherin Regulates Actin Cytoskeletal Remodeling and Enhances Human Gastric Cancer Metastasis via Epithelial-Mesenchymal Transition. *Int. J. Oncol.* 51 (1), 63–74. doi:10.3892/ijo.2017.4002

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S. A., Belver, L., Wendorff, A. A., et al. (2014). A NOTCH1-Driven MYC Enhancer Promotes T Cell Development, Transformation and Acute Lymphoblastic Leukemia. *Nat. Med.* 20 (10), 1130–1137. doi:10.1038/nm.3665

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the Control of Cell Identity and Disease. *Cell* 155 (4), 934–947. doi:10.1016/j.cell.2013.09.053

Hogan, B. L. (1996). Bone Morphogenetic Proteins: Multifunctional Regulators of Vertebrate Development. *Genes Dev.* 10 (13), 1580–1594. doi:10.1101/gad.10.13.1580

Hou, C.-H., Hsiao, Y.-C., Fong, Y.-C., and Tang, C.-H. (2009). Bone Morphogenetic Protein-2 Enhances the Motility of Chondrosarcoma Cells via Activation of Matrix Metalloproteinase-13. *Bone* 44 (2), 233–242. doi:10.1016/j.bone.2008.09.021

Huang, H., Hu, J., Maryam, A., Huang, Q., Zhang, Y., Ramakrishnan, S., et al. (2021). Defining Super-enhancer Landscape in Triple-Negative Breast Cancer by Multiomic Profiling. *Nat. Commun.* 12 (1), 2242. doi:10.1038/s41467-021-22445-0

Huang, P., Chen, A., He, W., Li, Z., Zhang, G., Liu, Z., et al. (2017). BMP-2 Induces EMT and Breast Cancer Stemness through Rb and CD44. *Cel Death Discov.* 3 (1), 1–12. doi:10.1038/cddiscovery.2017.39

Jung, H.-Y., and Yang, J. (2015). Unraveling the TWIST between EMT and Cancer Stemness. *Cell Stem Cell* 16 (1), 1–2. doi:10.1016/j.stem.2014.12.005

Kim, B. R., Oh, S. C., Lee, D.-H., Kim, J. L., Lee, S. Y., Kang, M. H., et al. (2015). BMP-2 Induces Motility and Invasiveness by Promoting colon Cancer Stemness through STAT3 Activation. *Tumor Biol.* 36 (12), 9475–9486. doi:10.1007/s13277-015-3681-y

Korkaya, H., Kim, G.-i., Davis, A., Malik, F., Henry, N. L., Ithimakin, S., et al. (2012). Activation of an IL6 Inflammatory Loop Mediates Trastuzumab Resistance in HER2+ Breast Cancer by Expanding the Cancer Stem Cell Population. *Mol. Cel* 47(4), 570–584. doi:10.1016/j.molcel.2012.06.014

Langhans, S. A. (2018). Three-Dimensional *In Vitro* Cell Culture Models in Drug Discovery and Drug Repositioning. *Front. Pharmacol.* 9, 6. doi:10.3389/fphar.2018.00006

Li, J., and Zhou, B. P. (2011). Activation of β-catenin and Akt Pathways by Twist Are Critical for the Maintenance of EMT Associated Cancer Stem Cell-like Characters. *BMC cancer* 11 (1), 1–11. doi:10.1186/1471-2407-11-49

Liang, Y., Hu, J., Li, J., Liu, Y., Yu, J., Zhuang, X., et al. (2015). Epigenetic Activation of TWIST1 by MTDH Promotes Cancer Stem-like Cell Traits in Breast Cancer. *Cancer Res.* 75 (17), 3672–3680. doi:10.1158/0008-5472.can-15-0930

Liao, A., Wang, W., Sun, D., Jiang, Y., Tian, S., Li, J., et al. (2015). Bone Morphogenetic Protein 2 Mediates Epithelial-Mesenchymal Transition via AKT and ERK Signaling Pathways in Gastric Cancer. *Tumor Biol.* 36 (4), 2773–2778. doi:10.1007/s13277-014-2901-1

Lim, S.-M., Kreipe, B. A., Trzeciakowski, J., Dangott, L., and Trache, A. (2010). Extracellular Matrix Effect on RhoA Signaling Modulation in Vascular Smooth Muscle Cells. *Exp. Cel. Res.* 316 (17), 2833–2848. doi:10.1016/j.yexcr.2010.06.010

Magnusson, K., Gremel, G., Rydén, L., Pontén, V., Uhlén, M., Dimberg, A., et al. (2016). ANLN Is a Prognostic Biomarker Independent of Ki-67 and Essential for Cell Cycle Progression in Primary Breast Cancer. *BMC Cancer* 16 (1), 904. doi:10.1186/s12885-016-2923-8

Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., et al. (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* 133 (4), 704–715. doi:10.1016/j.cell.2008.03.027

Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. *PLoS One* 3 (8), e2888. doi:10.1371/journal.pone.0002888

Oegema, K., Savoian, M. S., Mitchison, T. J., and Field, C. M. (2000). Functional Analysis of a Human Homologue of the Drosophila Actin Binding Protein Anillin Suggests a Role in Cytokinesis. *J. Cel Biol.* 150 (3), 539–552. doi:10.1083/jcb.150.3.539

Ooi, W. F., Xing, M., Xu, C., Yao, X., Ramlee, M. K., Lim, M. C., et al. (2016). Epigenomic Profiling of Primary Gastric Adenocarcinoma Reveals Super-enhancer Heterogeneity. *Nat. Commun.* 7, 12983. doi:10.1038/ncomms12983

Paolini, A., Duchemin, A. L., Albadri, S., Patzel, E., Bornhorst, D., González Avalos, P., et al. (2015). Asymmetric Inheritance of the Apical Domain and Self-Renewal of Retinal Ganglion Cell Progenitors Depend on Anillin Function. *Development* 142 (5), 832–839. doi:10.1242/dev.118612

Piccirillo, S. G. M., and Vescovi, A. L. (2007). Bone Morphogenetic Proteins Regulate Tumorigenicity in Human Glioblastoma Stem Cells. *Cancer Stem Cells*, 5, 59–81. doi:10.1007/2789_2007_044

Piekny, A. J., and Maddox, A. S. (2010). The Myriad Roles of Anillin during Cytokinesis. *Semin. Cel Dev. Biol.* 21 (9), 881–891. doi:10.1016/j.semcdb.2010.08.002

Riedl, A., Schlederer, M., Pudelko, K., Stadler, M., Walter, S., Unterleuthner, D., et al. (2017). Comparison of Cancer Cells in 2D vs 3D Culture Reveals Differences in AKT-mTOR-S6k Signaling and Drug Responses. *J. Cel Sci.* 130 (1), 203–218. doi:10.1242/jcs.188102

Soule, H. D., and McGrath, C. M. (1986). A Simplified Method for Passage and Long-Term Growth of Human Mammary Epithelial Cells. *In Vitro Cel Dev Biol.* 22 (1), 6–12. doi:10.1007/bf02623435

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

Suzuki, C., Daigo, Y., Ishikawa, N., Kato, T., Hayama, S., Ito, T., et al. (2005). ANLN Plays a Critical Role in Human Lung Carcinogenesis through the Activation of RHOA and by Involvement in the Phosphoinositide 3-kinase/AKT Pathway. *Cancer Res.* 65 (24), 11314–11325. doi:10.1158/0008-5472.Can-05-1507

Thapa, R., and Wilson, G. D. (2016). The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. *Stem Cell Int.* 2016, 1–15. doi:10.1155/2016/2087204

Vesuna, F., Lisok, A., Kimble, B., and Raman, V. (2009). Twist Modulates Breast Cancer Stem Cells by Transcriptional Regulation of CD24 Expression. *Neoplasia* 11 (12), 1318–1328. doi:10.1593/neo.91084

Wang, D., Naydenov, N. G., Dozmorov, M. G., Koblinski, J. E., and Ivanov, A. I. (2020). Anillin Regulates Breast Cancer Cell Migration, Growth, and Metastasis by Non-canonical Mechanisms Involving Control of Cell Stemness and Differentiation. *Breast Cancer Res.* 22 (1), 3. doi:10.1186/s13058-019-1241-x

Wang, L., Park, P., La Marca, F., Than, K. D., and Lin, C.-Y. (2015). BMP-2 Inhibits Tumor-Initiating Ability in Human Renal Cancer Stem Cells and Induces Bone Formation. *J. Cancer Res. Clin. Oncol.* 141 (6), 1013–1024. doi:10.1007/s00432-014-1883-0

Wang, L., Park, P., Zhang, H., La Marca, F., Claeson, A., Valdivia, J., et al. (2011). BMP-2 Inhibits the Tumorigenicity of Cancer Stem Cells in Human Osteosarcoma OS99-1 Cell Line. *Cancer Biol. Ther.* 11 (5), 457–463. doi:10.4161/cbt.11.5.14372

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., et al. (2004). Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis. *cell* 117 (7), 927–939. doi:10.1016/j.cell.2004.06.006

Zhang, X., Choi, P. S., Francis, J. M., Imielinski, M., Watanabe, H., Cherniack, A. D., et al. (2016). Identification of Focally Amplified Lineage-specific Super-enhancers in Human Epithelial Cancers. *Nat. Genet.* 48 (2), 176–182. doi:10.1038/ng.3470

Zhou, W., Wang, Z., Shen, N., Pi, W., Jiang, W., Huang, J., et al. (2015). Knockdown of ANLN by Lentivirus Inhibits Cell Growth and Migration in Human Breast Cancer. *Mol. Cel Biochem.* 398 (1), 11–19. doi:10.1007/s11010-014-2200-6

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.

Copyright © 2021 Maryam and Chin. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 06 July 2021
doi: 10.3389/fmolb.2021.682471

Hub Long Noncoding RNAs with m6A Modification for Signatures and Prognostic Values in Kidney Renal Clear Cell Carcinoma

Gaoteng Lin[1†], *Huadong Wang*[2†], *Yuqi Wu*[1], *Keruo Wang*[1] *and Gang Li*[1*]

[1]Department of Urology, Tianjin Institute of Urology, The Second Hospital of Tianjin Medical University, Tianjin, China,
[2]Department of Urology, Tianjin Baodi Hospital, Baodi Clinical College of Tianjin Medical University, Tianjin, China

Background: N6-methyladenosine (m6A)–modified long noncoding RNAs (m6A-lncRNAs) have been proven to be involving in regulating tumorigenesis, invasion, and metastasis for a variety of tumors. The present study aimed to screen lncRNAs with m6A modification and investigate their biological signatures and prognostic values in kidney renal clear cell carcinoma (KIRC).

Materials and Methods: lncRNA-seq, miRNA-seq, and mRNA-seq profiles of KIRC samples and the clinical characteristics of corresponding patients were downloaded from The Cancer Genome Atlas (TCGA). The R package "edgeR" was utilized to perform differentially expressed analysis on these profiles to gain DElncRNAs, DEmiRNAs, and DEmRNAs, respectively. The results of intersection of DElncRNAs and m6A-modified genes were analyzed by the weighted gene co-expression network analysis (WGCNA) to screen hub m6A-lncRNAs. Then, WGCNA was also used to construct an lncRNA-miRNA-mRNA (ceRNA) network. The Cox regression analysis was conducted on hub m6A-lncRNAs to construct the m6A-lncRNAs prognostic index (m6AlRsPI). Receiver operating characteristic (ROC) curve was used to assess the predictive ability of m6AlRsPI. The m6AlRsPI model was tested by internal and external cohorts. The molecular signatures and prognosis for hub m6A-lncRNAs and m6AlRsPI were analyzed. The expression level of hub m6A-lncRNAs in KIRC cell lines were quantified by qRT-PCR.

Results: A total of 21 hub m6A-lncRNAs associated with tumor metastasis were identified in the light of WGCNA. The ceRNA network for 21 hub m6A-lncRNAs was developed. The Cox regression analysis was performed on the 21 hub m6A-lncRNAs, screening two m6A-lncRNAs regarded as independent prognostic risk factors. The m6AlRsPI was established based on the two m6A-lncRNAs as follows: (0.0006066 × expression level of LINC01820) + (0.0020769 × expression level of LINC02257). The cutoff of m6AlRsPI was 0.96. KM survival analysis for m6AlRsPI showed that the high m6AlRsPI group could contribute to

OPEN ACCESS

Edited by:
Hossain Shekhar,
University of Dhaka, Bangladesh

Reviewed by:
Nan Liu,
Shenzhen University, China
Huiying Yang,
Sir Run Run Shaw Hospital, China

****Correspondence:***
Gang Li
797980@sina.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *18 March 2021*
Accepted: *04 May 2021*
Published: *06 July 2021*

Citation:
Lin G, Wang H, Wu Y, Wang K and Li G (2021) Hub Long Noncoding RNAs with m6A Modification for Signatures and Prognostic Values in Kidney Renal Clear Cell Carcinoma. Front. Mol. Biosci. 8:682471. doi: 10.3389/fmolb.2021.682471

Abbreviations: m6A, N6-methyladenosine; KIRC, kidney renal clear cell carcinoma; DElncRNAs, differentially expressed lncRNAs; WGCNA, weighted gene co-expression network analysis; m6AlRsPI, m6A-lncRNAs prognostic index; ceRNA network, lncRNA-miRNA-mRNA network; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; EMT, epithelial-mesenchymal transition; ROC, receiver operating characteristic

higher mortality. The area under ROC curve for m6AlRsPI for predicting 3- and 5-year survival was 0.760 and 0.677, respectively, and the m6AlRsPI was also tested. The mutation and epithelial–mesenchymal transition (EMT) analysis for m6AlRsPI showed that the high m6AlRsPI group had more samples with gene mutation and had more likely caused EMT. Finally, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed for mRNAs interacted with the two m6A-lncRNAs, showing they were involved in the process of RNA splicing and regulation of the mRNA surveillance pathway. qRT-PCR analysis showed that the two m6A-lncRNAs were upregulated in KIRC.

Conclusion: In the present study, hub m6A-lncRNAs were determined associated with metastasis in KIRC, and the ceRNA network demonstrated the potential carcinogenic regulatory pathway. Two m6A-lncRNAs associated with the overall survival were screened and m6AlRsPI was constructed and validated. Finally, the molecular signatures for m6AlRsPI and the two m6A-lncRNAs were analyzed to investigate the potential modulated processes in KIRC.

Keywords: kidney renal clear cell carcinoma, long noncoding RNA, N6-methyladenosine, prognostic index, WGCNA

INTRODUCTION

Renal cell carcinoma (RCC) is the third most common malignancy in the urogenital system (Siegel et al., 2019), with approximately 270,000 new diagnosed cases annually and 116,000 deaths around the world (Ljungberg et al., 2011). Among all renal tumors, 75% are kidney renal clear cell carcinoma (KIRC) (Rini et al., 2009), characteristically of high morbidity and mortality, for which chemotherapy, radiotherapy, and some other nonoperative treatments are difficult to obtain ideal outcomes (Rini et al., 2009). Patients with KIRC benefit majorly from surgery. There are still about 30% patients suffering from the risk of metastasis (Hsieh et al., 2017). Due to the complexity of carcinogenic regulatory network and the diversity of regulatory molecular modification, inhibition of known key targets cannot produce a marked effect. Other key factors that restrain tumor progression have become the major means for management of patients with metastatic KIRC. Therefore, it is crucial to identify other key regulatory factors for KIRC in an effort to improve the clinical outcome.

The accumulation of epigenetic alteration has become a momentous factor to advance development for tumors formatted by gene mutations. The aberrantly expressed long noncoding RNAs (lncRNAs) in specific cancers have been recognized as significant epigenetic regulatory molecules and as novel biomarkers for early diagnosis and therapy (Bach and Lee, 2018; Camacho et al., 2018). LncRNAs, defined as noncoding protein RNAs with more than 200 nucleotides are implicated in numerous bioprocesses of tumor cells, such as supporting proliferative signaling, eluding immune destruction and surveillance, capacitating replicative immortality, inducting angiogenesis, and activating invasion and metastasis (Gutschner and Diederichs, 2012; Bach and Lee, 2018). It has been found that upregulated TRPM2-AS contributes to tumor cell proliferation through TRPM2 stabilized by TAF15 in colorectal cancer (Pan et al., 2020). Qin et al. (2019) demonstrated that lncRNA MIR155HG acted as a tumor-promoting factor through inhibition of miR-802 that is a tumor suppressor in pancreatic cancer, providing a potential diagnostic and therapeutic target. Also in ovarian cancer, lncRNA HOTTIP boosts the expression of IL-6 to lead the immune escape of tumor cells by upregulating PD-L1 expression in neutrophils (Shang et al., 2019). In addition, the occurrence of tumor resistant to chemotherapy can also be caused by dysregulated lncRNA H19 through modulation of the glutathione metabolism pathway in advanced ovarian cancer (Zheng et al., 2016). Therefore, the regulatory role of lncRNAs to tumor is not less than that of oncogene or tumor suppressor gene, elucidating the significance of the landscape of lncRNAs in tumors.

N6-methyladenosine (m6A) refers to methylation at the N6 position of adenosine, which has been demonstrated as the most ubiquitous, abundant, and conserved internal dynamic posttranscriptional chemical modification within mRNA (Meyer and Jaffrey, 2014; Zhao et al., 2017). Functionally, m6A modulates mRNA maturation, splicing, translation, expression, and degradation and other major bioprocesses that have not been discovered yet (Meyer and Jaffrey, 2014; Zhao et al., 2017). It is not so surprising that accumulating evidence demonstrated that m6A was implicated in regulating hallmarks of tumors with tumorigenesis, proliferation, differentiation, invasion, and metastasis (Lin et al., 2016; Chen et al., 2019; Wang T. et al., 2020). The reversible processes of m6A modification were catalyzed by methyltransferases (termed as "writers"), demethylating enzymes (termed as "erasers"), and m6A-binding proteins (termed as "readers") (Meyer and Jaffrey, 2014). METTL3, classified as writers, inhibits SOX2 degradation for conducting the progression of colorectal carcinoma in an m6A-IGF2BP2–dependent manner (Li et al., 2019) and was linked with poor prognosis of gastric cancer due to

METTL3-mediated HDGF of methylation facilitating cancer progression (Wang Q. et al., 2020). ALKBH5, an m6A demethylase, is capable of attenuating WIF-1 RNA methylation and mediating the Wnt signaling pathway, to restrain the progression of pancreatic cancer, causing a favorable clinical outcome for patients (Tang et al., 2020). Readers are the factors that execute functions. YTHDF2, classified as readers, causes liver cancer metastasis by elevating the expression of OCT4 in an m6A RNA methylation manner (Zhang et al., 2020). Therefore, m6A modification offers a novel insight into the interplay between mRNAs and cancers.

m6A modification also occurs within lncRNAs, with the equally significant role in regulating bioprocesses of tumor (Yi et al., 2020). lncRNA NEAT1-1 with m6A modification is capable to regulate the Pol II ser2 phosphorylation and promotes the formation of the complex CYCLINL1/CDK19/NEAT1-1 contributing to bone metastasis of prostate cancer (Wen et al., 2020). In colorectal cancer, lncRNA LINRIS modulates the expression of MYC to influence the process of glycolysis, leading to the progression of cancer by stabilizing IGF2BP2 (Wang H. et al., 2019). However, the hub lncRNAs with m6A modification (m6A-lncRNAs) that are associated with the prognosis in KIRC remain unclear. The present study aims to identify m6A-lncRNAs in KIRC and investigate their biological signatures and prognostic values, providing promising targets for further research.

MATERIALS AND METHODS

Acquisition of Data About KIRC

The lncRNA-seq profiles, miRNA-seq profiles, and mRNA-seq profiles of KIRC samples and corresponding clinical characteristics including survival time, status, age, gender, grade, and TNM staging were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). External cohort (GSE40914) was gained from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) to validate the prognostic value of screened m6A-lncRNA. The identified genes with m6A modification were retrieved from the RMVar database (http://rmvar.renlab.org/index.html) (Luo et al., 2020). Data with missing or unknown values were deleted. Then, we randomly and equally divided the parental TCGA cohort into two subgroups. There was no discrepancy between them. A subgroup was regarded as the training dataset and another as the testing dataset.

Acquirement of the Hub m6A-lncRNAs

Production of Gene Modules

In the training dataset, the R package "edgeR" was utilized to perform differentially expressed analysis on the lncRNA-seq data in R software (version 4.0.2), with the statistical significance of $|\log2 \text{ fold change (FC)}| > 1$ and the false discovery rate (FDR) < 0.05. After intersecting DElncRNAs with m6A-mediated genes, the weighted gene co-expression network analysis (WGCNA) was conducted to identify hub m6A-lncRNAs in KIRC. First, the Pearson's correlation coefficient of the pairwise m6A-lncRNAs was calculated for constructing the similarity matrix. The similarity matrix was transformed into a weighted adjacency matrix after a soft threshold of β was identified. Next, the adjacency matrix was transformed into topological overlap matrix (TOM) measuring the connectivity between genes. Then, in terms of dissimilarity measure (1-TOM), average linkage hierarchical clustering was performed to cluster the genes with similar expression profiles for producing gene modules. The dissimilarity of module eigengenes (MEs) was calculated.

Identification of Significant Modules Linked With Clinical Traits

Gene significance (GS) obtained from the log10 transformation of the *p* value in a linear regression between the gene expression and clinical traits was computed to assess the significance of each module. Then, the average GS within a module was called as module significance (MS), which can be used to analyze the association between a module and clinical characteristics. The module correlated to a clinical trait was the one with the largest MS among all selected modules. The correlation between the MEs and clinical characteristics including grade and TNM staging was also calculated to determine the relevant module associated with clinical information. These processes were called as module–trait relationship (MTR) analysis (Langfelder and Horvath, 2008).

Obtaining Hub m6A-lncRNAs

A network was constructed according to the edges between two m6A-lncRNAs with weight >0.5 after screening m6A-lncRNAs correlated with clinical modules (the red and turquoise modules). On the basis of the plug-in unit "cytoHubba" within the Cytoscape software (version 3.7.2), m6A-lncRNAs in the network with Maximal Clique Centrality (MCC) ≥ 2 were identified as the hub m6A-lncRNAs.

Construction of an lncRNA-miRNA-mRNA Network Based on Hub m6A-lncRNAs

The R package "edgeR" was applied to conduct a differentially expressed analysis on the miRNA-seq and mRNA-seq of KIRC samples in R software. The selection criteria were the same as described previously. Then, weighted correlation network analysis (WGCNA) was performed on hub m6A-lncRNAs-DEmiRNAs profile to calculate the connectivity between molecules. The analysis process was described previously. An optimal soft threshold of β was equal to three. The edges between hub m6A-lncRNAs and DEmiRNAs with a threshold weight >0.3 were selected. Next, WGCNA was also performed on DEmiRNAs–DEmRNAs profile. The soft threshold of β was set to five. The edges between DEmiRNAs and DEmRNAs with a threshold weight >0.3 were opted. Finally, took the DEmiRNAs as a bridge, an lncRNA-miRNA-mRNA (ceRNA) network was constructed and visualized by the Cytoscape software (version 3.7.2). The Kaplan–Meier (KM) survival analysis along with the log-rank test on the regulatory molecules in the ceRNA network was performed.

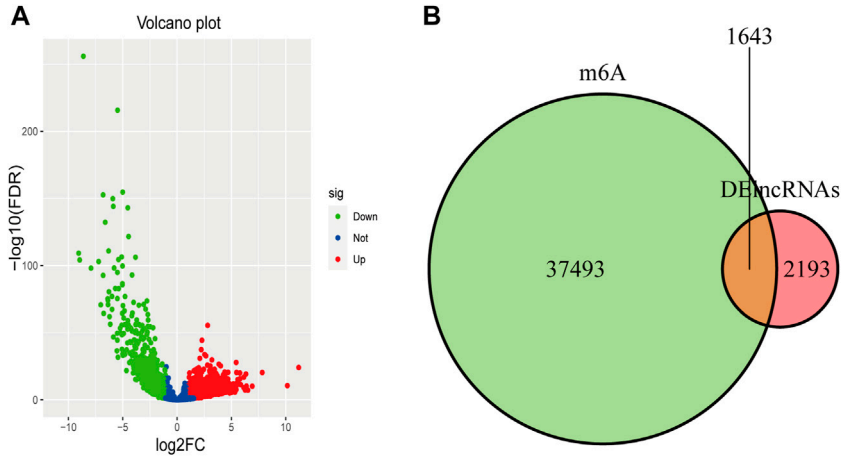


FIGURE 1 | Differentially expressed m6A-lncRNAs. **(A)** The volcano plot has shown the differentially expressed lncRNAs in the training cohort in KIRC. The green dots, blue dots, and red dots mean the downregulated, no differentially expressed, and upregulated lncRNAs, respectively. **(B)** The intersection of DElncRNAs with m6A-modified genes was illustrated by the Venn diagram.

Development of m6A-lncRNAs Prognostic Index

The univariate Cox regression analysis was conducted on the hub m6A-lncRNAs to screen the m6A-lncRNAs associated with overall survival. Based on the results of the univariate analysis, the multivariate Cox regression analysis was utilized to identify the independent prognostic factors. The m6A-lncRNAs prognostic index was constructed with the hub m6A-lncRNAs of $p < 0.05$ in the multivariate Cox regression model through multiplying the expression values of them by their coefficient in the model and then adding them together. The patients were then divided into high-m6AlRsPI and low-m6AlRsPI groups based on the median m6AlRsPI value. After screening the hub m6A-lncRNAs correlated with the overall survival through the Cox regression analysis, we performed the KM survival analysis on the relevant m6A-lncRNAs with statistical significance and the m6AlRsPI to estimate their prognostic power. The performance of m6AlRsPI for predicting 3- and 5-year overall survival of KIRC patients was analyzed by receiver operating characteristic (ROC) curve.

Validation for m6AlRsPI

Internal and external cohorts were applied to validate the prognostic signature of m6AlRsPI. In the testing dataset, the analysis processes were described previously. The m6AlRsPI was constructed and the prognostic and predictive signatures for m6AlRsPI in the testing cohort were estimated. External cohort (GSE40914) was also utilized to verify prognostic values of screened m6A-lncRNAs.

Analysis of Gene Mutation in Different m6AlRsPI Subgroups

The mutation analysis for KIRC samples was investigated to display the discrepant gene mutation in different m6AlRsPI subgroups. The major mutant genes and variant classifications in different m6AlRsPI subgroups were our concern.

Association of m6AlRsPI Subgroups With Epithelial–Mesenchymal Transition

The correlation between different m6AlRsPI subgroups and epithelial–mesenchymal transition (EMT) was explored to show the prognostic power in different m6AlRsPI subgroups, by analyzing the expression of CDH1, VIM, SNAI1, SNAI2, and CHD2 in high m6AlRsPI and low m6AlRsPI groups. The expression of PDCD1 in different m6AlRsPI subgroups was also calculated.

Molecular Signature Analysis for Screened Prognosis-Associated Hub m6A-lncRNAs

We looked for RNA-binding proteins (RBPs) in The Encyclopedia of RNA Interactomes (ENCORI) (http://starbase.sysu.edu.cn/index.php), RBPDB database (http://rbpdb.ccbr.utoronto.ca/), and catRAPID database (http://s.tartaglialab.com/page/catrapid_group) to obtain mRNAs interacted with the screened prognosis-associated hub m6A-lncRNAs. Then, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to exhibit the potential regulatory signaling pathways in KIRC.

Validation of the Expression Level of Screened Hub m6A-lncRNAs in KIRC by qRT-PCR

Total RNA was extracted from cell lines using TRIzol reagent (Invitrogen, China) in terms of the manufacturer's protocol. HK-2 is a normal kidney cell line, whereas ACHN, 769-P, and 786-O are KIRC cell lines. Reverse transcription was conducted on RNA into

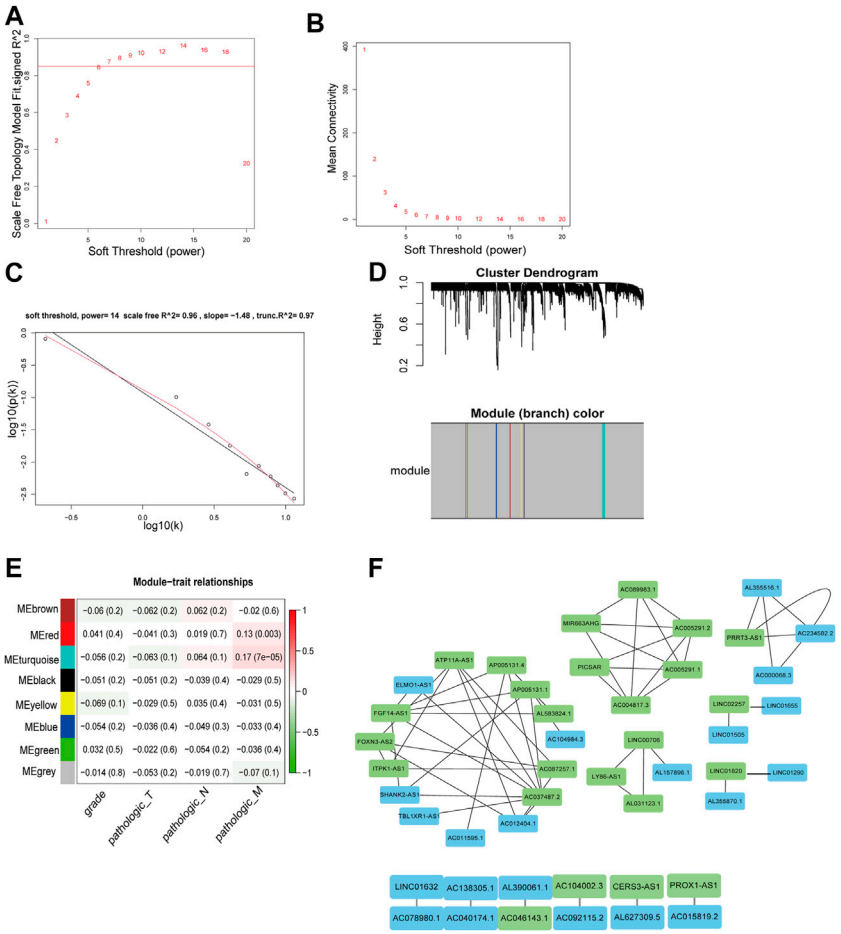


FIGURE 2 | The processes of screening the hub m6A-lncRNAs. **(A)** Analysis of the scale-free fit index for various soft thresholds, and the soft threshold was set as 14. **(B)** Analysis of the mean connectivity for various soft thresholds. **(C)** Checking the scale-free topology when the soft threshold = 14. **(D)** Dendrogram showing the clustered m6A-DElncRNAs. **(E)** Heat map illustrating the association of the MEs with clinical characteristics (including grade and TNM staging). The modules with $p < 0.05$ were selected. **(F)** Network for m6A-lncRNAs within the red and turquoise modules. The blue boxes mean hub-lncRNAs within the red module and the green boxes represent hub-lncRNAs within the turquoise module. And lines indicate that there is a correlation between m6A-lncRNAs.

cDNA with RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific; United States) according to the manufacturer's guidelines. Quantitative reverse transcription–polymerase chain reaction (qRT–PCR) was performed using FastStart Universal SYBR Green Master (ROX) (Roche; United States). GAPDH was used as the internal control and the relative expression levels of m6A-lncRNA were calculated by $2^{-\Delta\Delta Ct}$. The primer sequences utilized in this study were as follows: GAPDH-F: GGAAGGTGAAGGTCGGAGTCA, GAPDH-R: GTCATTGATGGCAACAATATCCACT; LINC01820-F: GGCCCACCCACATAGTTTAAAGCCA, LINC01820-R: GCACACTCACAGAACGCAAA; LINC02257-F: AGGTGGAGTCTCGCACTGTCATCCT, LINC02257-R: TTC ACTGGTTTGCTCTGCAATCCCA.

STATISTICAL ANALYSIS

The continuous variables were compared by independent *t*-test. Univariate and multivariate prognostic analysis were performed in the Cox regression model. The survival analysis was performed by the Kaplan–Meier survival analysis with the log-rank test. The R package "survival ROC" was utilized to plot the time-dependent ROC curve. The threshold of statistical significance was $p < 0.05$.

RESULTS

Obtaining m6A-lncRNAs for KIRC

Differentially expressed analysis performed on the training dataset (271 tumors vs. 35 normal samples), the sum of DElncRNAs was 3,836, including 1,221 downregulated lncRNAs and 2,615 upregulated lncRNAs, as shown in **Figure 1A**. A total of 1,643 differentially expressed m6A-lncRNAs were identified following intersection of 3,836 DElncRNAs with 39,136 m6A-modified genes, as shown in **Figure 1B**. After clearing invalid data, the data of clinical characteristics are presented in **Supplementary Table S1**.

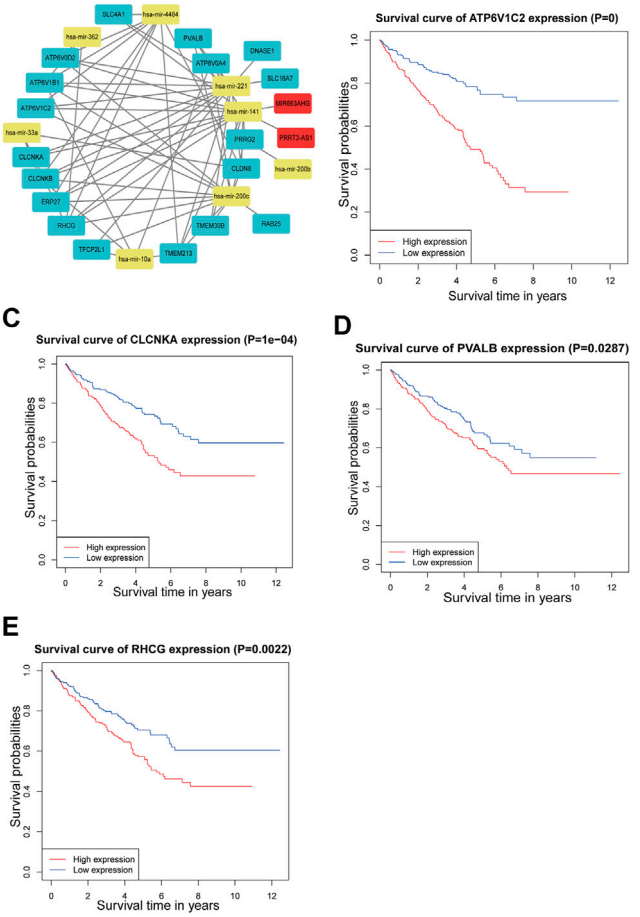


FIGURE 3 | ceRNA network and survival analysis. **(A)** ceRNA network based on the hub m6A-lncRNAs. There were two hub m6A-lncRNAs, eight DEmiRNAs, and 18 DEmRNAs in the ceRNA network. The red squares represent hub m6A-lnRNAs, yellow squares are miRNAs, and blue squares mean mRNAs. **(B–E)** The survival analysis for ATP6V1C2, CLCNKA, PVALB, and RHCG.

Identification of Hub m6A-lncRNAs Related to Clinical Traits

In the training cohort, WGCNA was conducted on 1,643 differentially expressed m6A-lncRNAs to obtain the hub m6A-lncRNAs in KIRC. The optimal soft threshold value was selected as 14 (scale free $R^2 = 0.96$) to ensure a scale-free network (**Figures 2A,B**). We further analyzed the reliability of the scale-free topology in the light of soft threshold equal to 14 (**Figure 2C**). The logarithm log10 (k) of the node with connectivity K was negatively associated with the logarithm log10 [p(k)] of the probability of the node. According to the similar expression pattern, a total of eight modules were determined by average linkage clustering (**Figure 2D**). Red and turquoise modules were found to have a correlation with tumor M stage by MTR analysis (**Figure 2E**). There were 22 m6A-lncRNAs within red module and 25 m6A-lncRNAs within turquoise module, with a threshold weight >0.5. Then, m6A-lncRNAs within the two modules were put into the Cytoscape software, to construct a network (**Figure 2F**) and to calculate the MCC value which is listed in **Supplementary Table S2**, and m6A-lncRNAs with MCC ≥2 were identified as the hub m6A-lncRNAs. Finally, the total of hub m6A-lncRNAs was equal to 21.

Construction of an lncRNA-miRNA-mRNA Network

After performing differential expression analysis on the miRNA-seq and mRNA-seq for KIRC samples, a total of 173 DEmiRNAs and 722 DEmRNAs were obtained, respectively. First, WGCNA was conducted on the hub m6A-lncRNA-DEmiRNAs profile to determine 19 hub m6A-lncRNAs associated with 107 DEmiRNAs. Similarly, 77 DEmiRNAs correlated with 459 DEmRNAs were obtained by performing WGCNA on DEmiRNAs-DEmRNAs profile. Regarding the common DEmiRNAs on both profiles as the connection point, the lncRNA-miRNA-mRNA network (ceRNA network) comprising of two hub m6A-lncRNAs, eight DEmiRNAs, and 18 DEmRNAs was constructed (**Figure 3A**). The potential regulatory pathways based on ceRNA network contain MIR663AHG/hsa-mir-141/SLC4A1 etc. and PRRT3-AS1/hsa-mir-141/SLC4A1 etc. Then, other lncRNA-miRNA correlation networks and miRNA-mRNA correlation networks are shown in **Supplementary Figures S1A,B**. The survival analysis on the molecules within the ceRNA network was conducted. However, the two hub m6A-lncRNAs (MIR663AHG and PRRT3-AS1) and miRNA (has-mir-141) had no impact on the overall survival of patients with KIRC (**Supplementary Figures S2A–C**). In addition, four mRNAs (ATP6V1C2, CLCNKA, PVALB, and RHCG) were found to be associated with the prognosis for patients (**Figures 3B–E**). Other mRNAs with no survival difference are shown in **Supplementary Figures S2D–M**.

Establishment of m6A-lncRNAs Prognostic Index Based on Cox Regression Analysis

The Cox regression analysis was conducted on the 21 hub m6A-lncRNAs to screen the prognostic factors for patients with KIRC. The outcome of the univariate Cox regression analysis was illustrated in **Figure 4A**. The hub m6A-lncRNAs with statistical significance were included in the multivariate Cox regression analysis (**Figure 4B**). It was found that LINC01820 (HR: 1.0006; 95% CI: 1.0003–1.0009; $p < 0.001$) and LINC02257 (HR: 1.0021; 95% CI: 1.0001–1.0041; $p = 0.0423$) were closely correlated with the poor prognosis of patients with KIRC (risky hub m6A-lncRNAs). The m6A-lncRNAs prognostic index (m6AlRsPI) was established based on the two m6A-lncRNAs as follows: m6AlRsPI = (0.0006066 × expression level of LINC01820) + (0.0020769 × expression level of LINC02257). The median value of m6AlRsPI was 0.96, as the cutoff to separate KIRC samples into the high and the low m6AlRsPI group. The Kaplan–Meier survival analysis for the two hub m6A-lncRNAs and m6AlRsPI

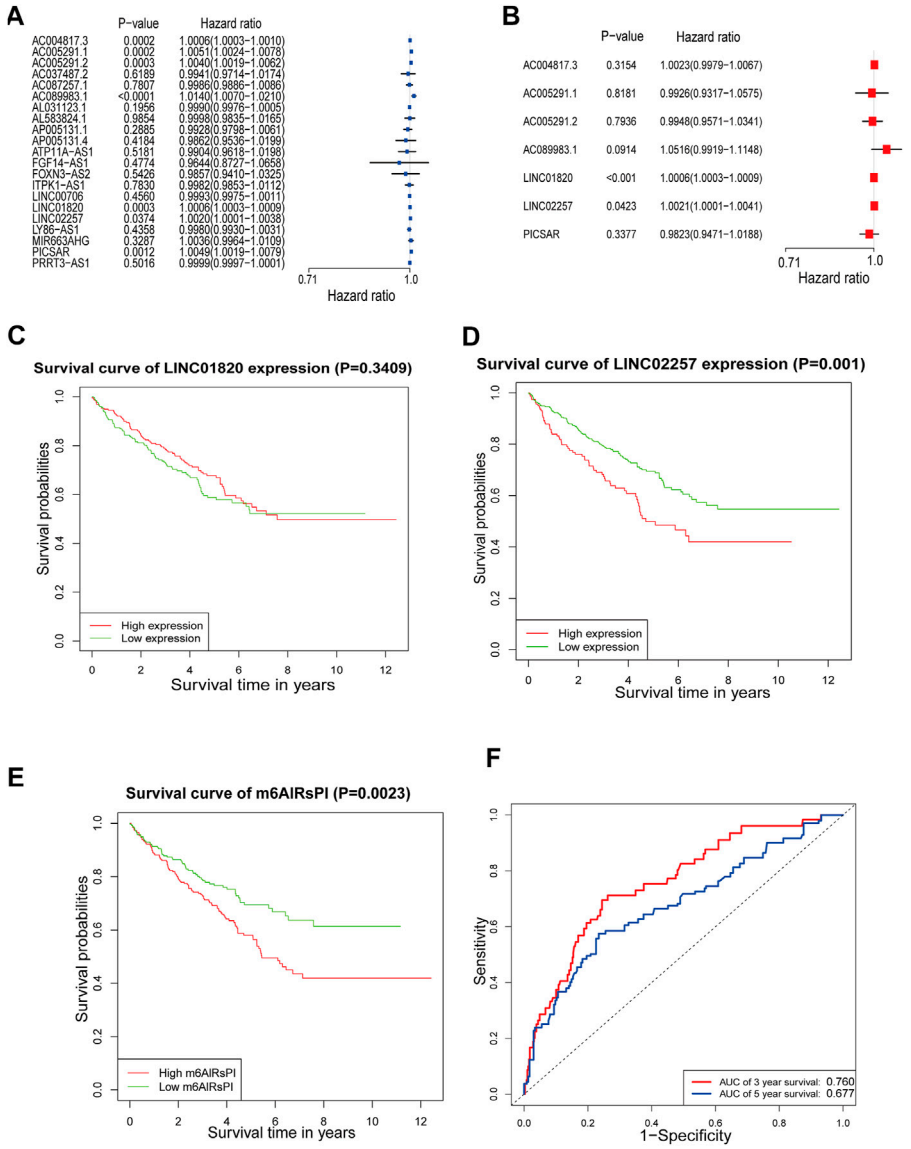


FIGURE 4 | Cox regression analysis and survival analysis. **(A)** The result of univariate Cox regression analysis for m6A-lncRNAs in red and turquoise modules shown by forest plot. **(B)** The outcome of multivariate Cox regression analysis for m6A-lncRNAs with $p < 0.05$ in univariate analysis. **(C–E)** Survival analysis for LINC01820, LINC02257, and m6AlRsPI in the training cohort **(F)** ROC curve for m6AlRsPI for predicting 3- and 5-year survival. The red curve means 3-year survival prediction, with the area under the curve of 0.760. The blue curve means 5-year survival prediction, with the area under the curve of 0.677.

in the training cohort were carried out to investigate the association with the overall survival. The results demonstrated that the high expression level of LINC02257 had an unfavorable overall survival as compared with low expression level of LINC02257 (**Figure 4D**). However, there was no survival discrepancy between low expression level and high expression level of LINC01820 (**Figure 4C**). The high m6AlRsPI group could contribute to higher mortality than that of the low m6AlRsPI group (**Figure 4E**). The area under ROC curve for m6AlRsPI for predicting 3- and 5-years survival was 0.760 and 0.677, respectively **Figure 4F**, demonstrated a moderate performance for predictive prognosis.

Validation of the Reliability for the m6AlRsPI

For internal validation, the testing cohort was used to validate the m6AlRsPI model developed by the training cohort (**Supplementary Table S1**). The m6A-lncRNAs in the testing cohort were screened by WGCNA as described previously (**Supplementary Figures S3A–D**). The hub m6A-lncRNAs with $p < 0.05$ within the univariate Cox regression analysis were taken into multivariate Cox regression analysis (**Figures 5A,B**) to identify hub m6A-lncRNAs correlated with prognosis in the testing cohort. Then, the m6AlRsPI model based on LINC02257 and LINC01820 was also constructed, and the KM survival analysis showed high m6AlRsPI group with poor prognosis (**Figure 5C**). Moreover, in the testing cohort, the area

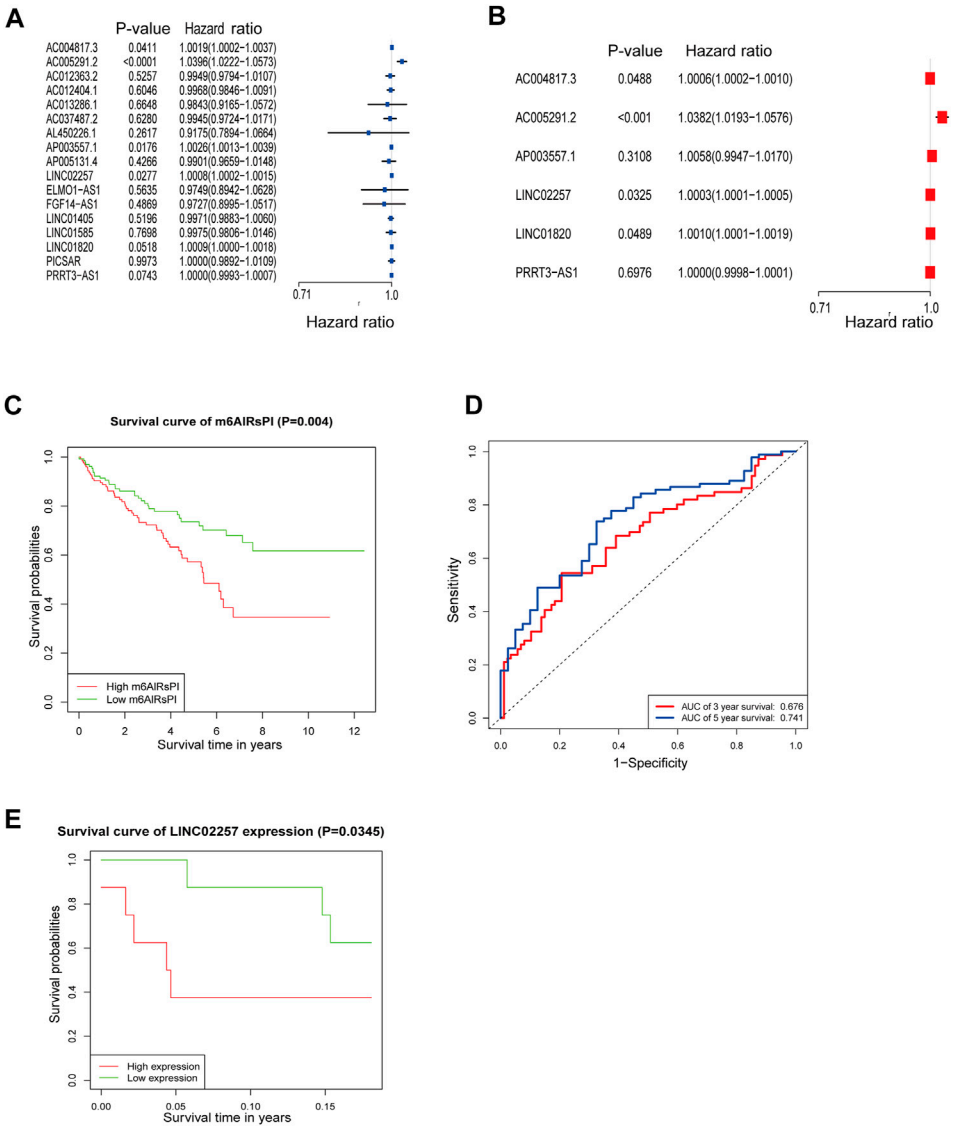


FIGURE 5 | Internal and external validation for m6AlRsPI. **(A)** The result of univariate Cox regression analysis for m6A-lncRNAs associated with M stage in gray module in the testing cohort, shown by forest plot. **(B)** The outcome of multivariate Cox regression analysis for hub lncRNAs with $p < 0.05$ within the univariate Cox regression analysis. **(C)** KM survival analysis for m6AlRsPI constructed by testing cohort. **(D)** In testing cohort, ROC curve for predicting 3- and 5-year survival. **(E)** Validation of the survival signature for LINC02257 within GEO dataset.

under ROC curve for predicting 3- and 5-years survival was 0.676 and 0.741, respectively (**Figure 5D**), with moderate predictive power, consistent with the result above. The KM survival analysis for LINC02257 within GEO cohort demonstrated a high expression level of LINC02257 which increased the mortality (**Figure 5E**) and no relevant external dataset was found to validate the prognostic value of LINC01820.

Gene Mutation Analysis for m6AlRsPI Subgroups

The gene mutation was analyzed to gain a novel insight into the molecular nature in m6AlRsPI subgroups. The gene mutation frequency in the high m6AlRsPI group was higher than that of the low m6AlRsPI group (89.22 vs. 85.45%). The top 10 mutational genes in the high m6AlRsPI group contain VHL, PBRM1, SETD2, TTN, BAP1, ATM, MTOR, MUC16, ADGRV1, and DNAH2 (**Figure 6A**). The mutation rate of more than 10% genes includes VHL, PBRM1, SETD2, TTN, and BAP1. The top 10 mutational genes in the low m6AlRsPI group comprise of VHL, PBRM1, TTN, SETD2, BAP1, DNAH9, HMCN1, KDM5C, MTOR, and MUC4 (**Figure 6B**). VHL, PBRM1, and TTN are the genes with the mutation rate of more than 10% genes. The most common mutation type in both groups is missense-mutation, followed by frameshift deletions and nonsense mutation (**Figures 6C,D**).

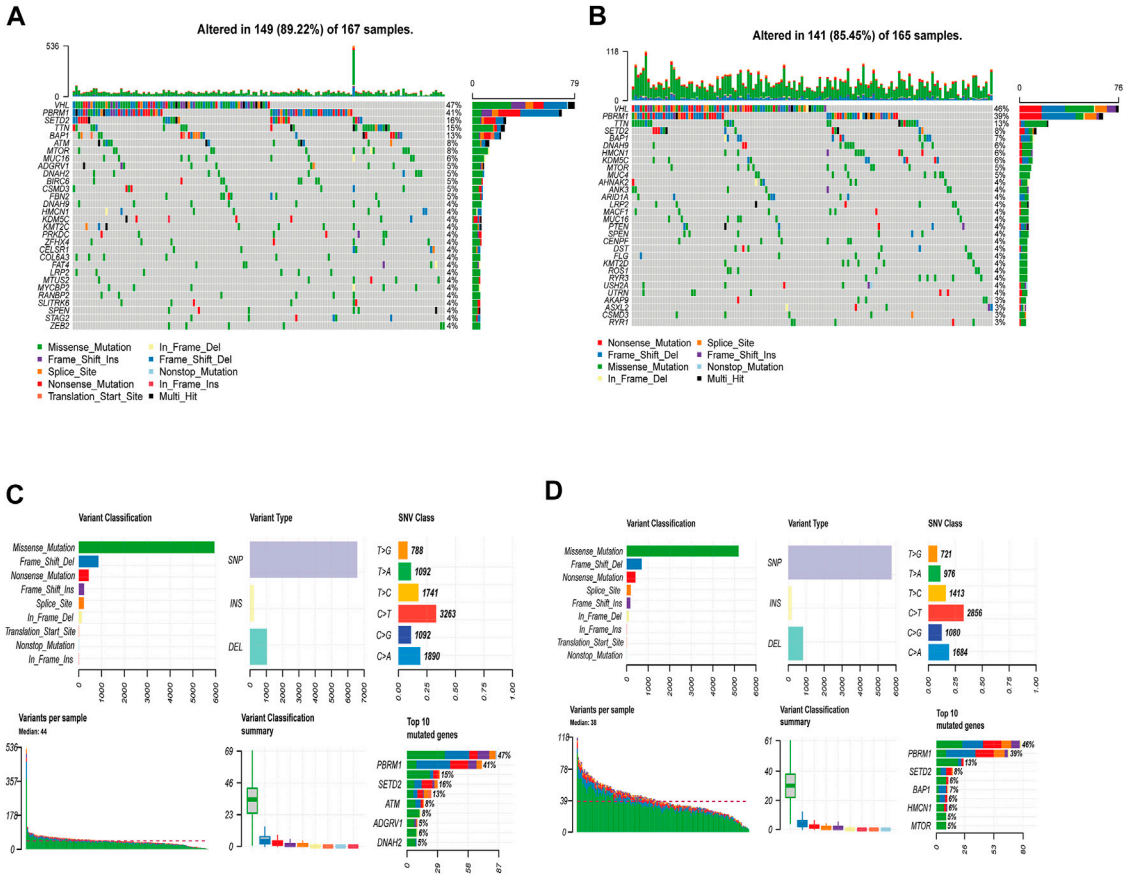


FIGURE 6 | Mutation analysis based on m6AlRsPI subgroups. **(A)** The waterfall plot demonstrated mutational genes in high m6AlRsPI group. **(B)** Mutational genes in low m6AlRsPI group. **(C)** Summary for mutation in high m6AlRsPI group. **(D)** Summary for mutation in low m6AlRsPI group.

Association of Epithelial–Mesenchymal Transition With m6AlRsPI Subgroups

As the hub m6A-lncRNAs are closely linked with the clinical trait of M stage, we made further analysis of the association of EMT with m6AlRsPI subgroups. E-cadherin is regarded as epithelial biomarker, which is coded by CDH1 gene. It was found that in the high m6AlRsPI group, the expression level of CDH1 gene decreased as compared with the low m6AlRsPI group (**Figure 7A**). Mesenchymal biomarkers include N-cadherin, vimentin, SNAI1, and SNAI2. N-cadherin and vimentin are coded by CDH2 and VIM gene, respectively. The analysis results showed the expression level of VIM, SNAI1, and SNAI2 in the high m6AlRsPI group were higher than that in the low m6AlRsPI group (**Figures 7B–D**). There was no statistical difference for the expression of CDH2 between the two groups (**Figure 7E**). In addition, the expression of PDCD1 between the two groups is statistically different (**Figure 7F**).

GO and KEGG Enrichment Analysis for Proteins Interacted with the Two Hub m6A-lncRNAs

After looking for RNA-binding proteins (RBPs) in The Encyclopedia of RNA Interactomes (ENCORI), RBPDB database, and catRAPID database, a total of 153 mRNAs interacted with the two hub m6A-lncRNAs were identified for GO and KEGG enrichment analysis. The RBP-lncRNA interaction in ENCORI was supported by CLIP-seq Data, and in RBPDB and catRAPID omics was predicted by RNA sequencing. The outcome from the GO analysis showed they were involved in processes of RNA splicing and mRNA metabolism (**Figures 8A,B**). KEGG analysis demonstrated they are available to modulate RNA transport and degradation and mRNA surveillance signaling pathway (**Figures 8C,D**).

The Expression Levels of Hub m6A-lncRNAs in KIRC

The expression levels of hub m6A-lncRNAs in KIRC were quantified by qRT-PCR. Compared to the normal kidney cell line (HK-2), the expression level of LINC01820 was elevated in KIRC cell lines (ACHN and 769-P), with statistical significance (**Figure 9A**). The expression level of LINC02257 was upregulated in 769-P and 786-O cell lines (**Figure 9B**). These results were consistent with our analysis results based on TCGA KIRC cohort, showing LINC01820 and LINC02257 were significant molecules in modulating KIRC progression.

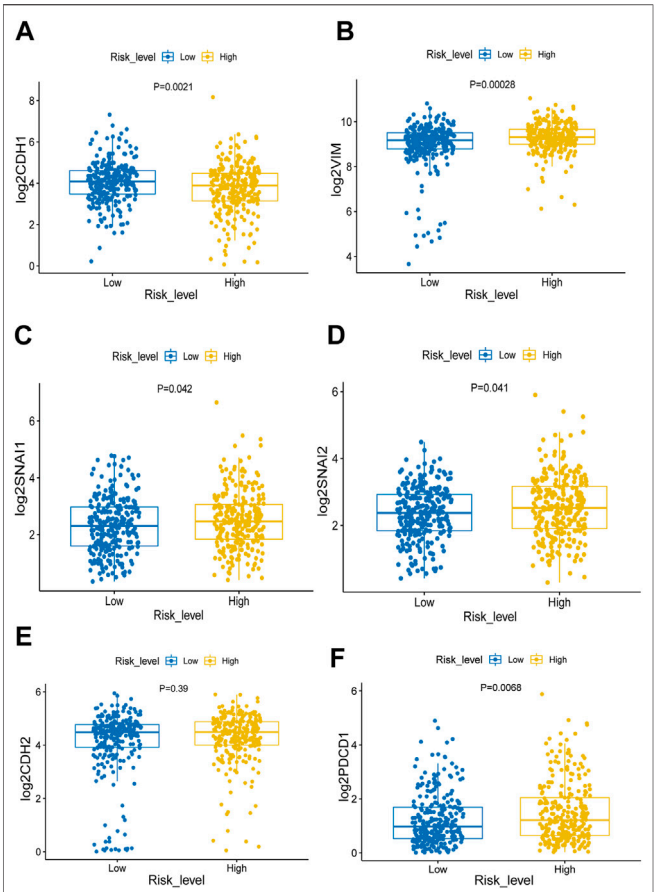


FIGURE 7 | Analyzing association with EMT and PDCD1 for m6AlRsPI subgroups. **(A)** The expression level of CDH1 in high m6AlRsPI group was decreased as compared with the low m6AlRsPI group. **(B–E)** Genes correlated with mesenchyme in high m6AlRsPI group were upregulated expression. **(F)** The expression level of PDCD1 was upregulated in high m6AlRsPI.

DISCUSSION

Metastatic KIRC is the major cause of poor clinical outcome and higher mortality for patients with kidney cancer. Although traditional targeted therapy and novel immunotherapy were utilized to treat patients with metastatic KIRC, no satisfactory lasting responses of drugs to tumors and 5-year overall survival rate for patients could not be obtained. For those patients who have lost the opportunity of surgery, inhibition of key regulatory targets for metastasis may be the most effective therapy (Thomas and Kabbinavar, 2015). Due to discovery of implicating in the diverse considerable bioprocesses for tumor, lncRNAs are regarded as a novel regulatory factor for investigating the mechanism of cancer metastasis (Gutschner and Diederichs, 2012). A variety of evidence has manifested the pivotal role of lncRNAs in regulating tumor metastasis. LncRNA MALAT1 has previously been regarded as a metastasis-promoting factor in various tumors, but it is a metastasis-suppressing factor in breast cancer through binding and inactivating the pro-metastatic transcription factor TEAD (Kim et al., 2018). LncRNA RP11-390F4.3 is induced by hypoxia/HIF-1α to facilitate EMT through modulation of multiple EMT-associated factors, leading to tumor metastasis (Peng et al., 2020). In addition, m6A has been identified as the most common chemical modification manner for various RNAs in eukaryotic cells, and it is involved in the progression of tumor through relative key regulatory factors with m6A modification (Wang Q. et al., 2020). The brain metastasis of lung cancer results from m6A-mediated matured miR-143-3p upregulating the expression of VASH1 to increase ubiquitylation of VEGFA (Wang Y. et al., 2019). BATF2 with m6A modification suppresses gastric tumor cell metastasis by stabilizing *p*53 protein to inhibit phosphorylation of ERK, causing a favorable prognosis (Xie et al., 2020). Furthermore, the process of EMT correlated with tumor metastasis can be triggered by m6A-related molecules. Li et al. (2020) demonstrated the overexpression of METTL3 induced EMT of cancer cells through modulation of expression and secretion of TGFβ1. Hua et al. (2018) found METTL3 promoted EMT by upregulating the receptor tyrosine kinase AXL in ovarian cancer. The crucial regulatory role of m6A modification in tumor metastasis ought to be paid attention. Therefore, the hub m6A-lncRNAs are vital targets for diagnosis, surveillance, and therapy of metastatic tumor, including KIRC. In colorectal cancer, m6A methylation can induce the upregulation of lncRNA RP11 to modulate Siah1-Fbxo45/Zeb1 complex to promote the dissemination of tumor cells (Wu et al., 2019). Zheng et al. (2019) demonstrated lncRNA FAM225A is an upregulated oncogenic lncRNA in nasopharyngeal carcinoma and m6A-mediated lncRNA FAM225A gives rise to tumor metastasis by sponging miR-590-3p and miR-1275 to upregulate ITGB3 and activate the FAK/PI3K/Akt signaling pathway. m6A-mediated LINC00470 facilitates gastric cancer cell distant metastasis by accelerating degradation of PTEN mRNA (Yan et al., 2020). In cervical cancer, lncRNA ZFAS1 with m6A modification conduces to unfavorable clinical outcome by suppressing miR-647 to lead to tumor metastasis (Yang et al., 2020). Currently, the significant biomarkers correlated with metastatic KIRC and relative regulatory mechanism have remained unknown. It is urgent to look for effective factors implicated in the modulation of metastasis to improve the overall survival of patients with KIRC. Additionally, hub m6A-lncRNAs can be regarded as extremely important factors for investigating the complex mechanism of metastasis as a result of abundant regulation functions.

Based on WGCNA, DElncRNAs with m6A modification were divided into modules with the correlation analysis. After calculating the correlation between modules and clinical characteristics, we found two modules that were linked with the clinical trait of M stage. Although they were weakly correlated, the hub m6A-lncRNAs with a potential regulatory role were of interest. Therefore, m6A-lncRNAs within the two modules were considered related with metastasis of KIRC and might be treated as potential major molecules for modulating tumor metastasis. Then, construction of a ceRNA network aims to investigate the potential regulation pathways of 21 hub m6A-lncRNAs. Cytoscape software was utilized to screen the hub m6A-lncRNAs, and the Cox regression analysis was performed to

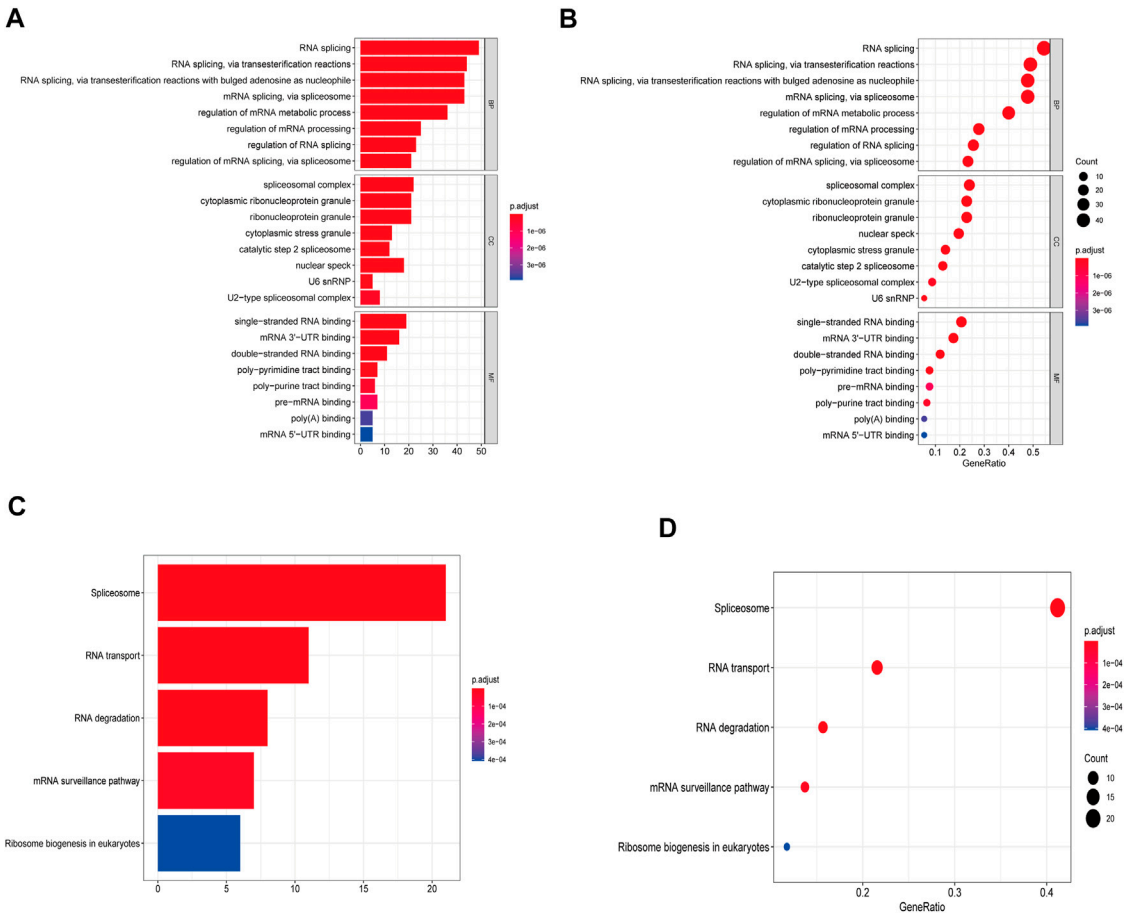


FIGURE 8 | GO and KEGG analysis. **(A,B)** The bar plot and dot plot of GO analysis for the top eight enrichment. **(C,D)** The bar plot and dot plot of KEGG analysis for the top five enrichment.

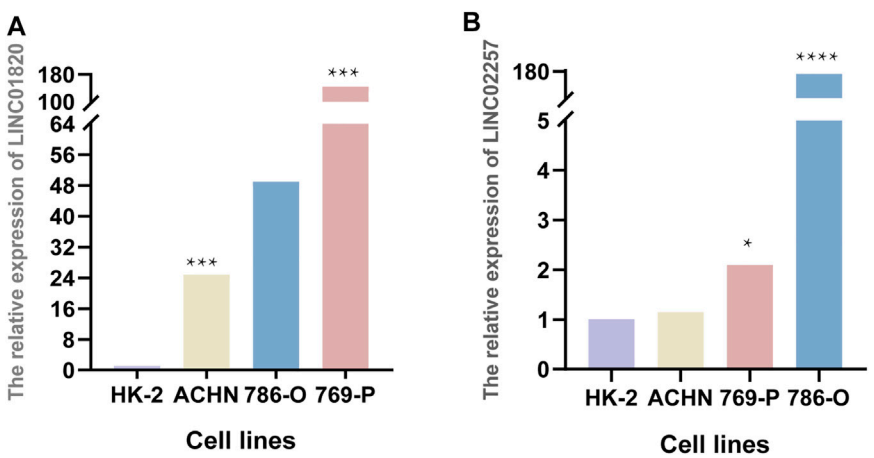


FIGURE 9 | Quantitative analysis for LINC01820 and LINC02257. **(A,B)** The expression levels of LINC01820 and LINC02257 in HK-2, ACHN, 786-O, and 769-P cell lines. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$.

identify hub m6A-lncRNAs correlated closely with the prognosis. However, prognosis-associated m6A-lncRNAs were not embodied in the ceRNA network. The m6A-lncRNAs within the ceRNA network were not correlated with overall survival by the KM survival analysis and Cox regression analysis. So it was shown that prognosis-associated m6A-lncRNAs may influence biological behaviors of KIRC through other manners (such as RNA-binding proteins manner) instead of lncRNA-miRNA-mRNA manner. Identified LINC02257 was shown to be associated with poor prognosis of patients with colorectal cancer based on bioinformatics analysis (Wang et al., 2018; Huang et al., 2020), but it was not clear whether m6A modification involved in the regulation of colorectal progression. In consideration of its unfavorable role in colorectal cancer, LINC02257 is considered as a novel significant regulator in metastatic KIRC, particularly in m6A manner. There was no survival discrepancy between low expression and high expression of LINC01820 through KM survival analysis. We thought KM survival analysis on LINC01820 may be disturbed by other confounding factors, leading to LINC01820 was not related with prognosis, and multivariate Cox regression analysis could solve the problem. The result from multivariate Cox regression analysis showed that LINC01820 was an independent factor associated with prognosis. Based on the result, we preferred to believe LINC01820 was a prognosis-associated m6A-lncRNA. Currently, there were no articles reporting that LINC01820 is correlated with prognosis in other tumors. ROC curve illustrated a moderate predictive signature for the m6AlRsPI. The m6AlRsPI developed by the training cohort was tested by self-validation and the prognostic signature of LINC0257 was validated by external GEO cohort. These results supported m6AlRsPI was a robust and reasonable model for predicting prognosis. The analysis of biological signatures for m6AlRsPI based on two hub m6A-lncRNAs showed that the high m6AlRsPI group had more samples with gene mutation, more likely to cause epithelial–mesenchymal transition, and led to worse prognosis. As LINC01820 and LINC02257 did not regulate tumor behavior by lncRNA-miRNA-mRNA manner, RNA-binding proteins which connected with the two hub m6A-lncRNAs were investigated. Then, we found that modulated by the two hub m6A-lncRNAs, the mRNA surveillance signaling pathway, termed nonsense-mediated mRNA decay (NMD), has been found to be associated with mutation in human cancer by NMD-triggering manner (Lindeboom et al., 2016). This signaling pathway may play an important role in tumorigenesis and development of KIRC. The over-expression of the two hub m6A-lncRNAs in KIRC cell lines validated by qRT-PCR further indicated that they may influence tumor behavior, contributing to poor prognosis. These findings demonstrated that the two hub m6A-lncRNAs and m6AlRsPI were crucial for patients with metastatic KIRC to predict the prognosis and investigate the complex metastatic regulatory mechanism. Accordingly, they were significantly promising targets for treatment of patients with metastatic KIRC.

However, some limitations in our study should be taken into consideration. First, WGCNA was only performed on differentially expressed m6A-lncRNAs, other lncRNAs were excluded, causing some m6A-lncRNAs loss. Second, the red and turquoise modules had a weak correlation with the M stage. Third, the prognostic values of LINC01820 and m6AlRsPI cannot be validated by the external cohort currently. Fourth, as the database on m6A gene is updating, the screening of m6A-lncRNAs in our study was based on the current m6A dataset. Fifth, downstream molecular mechanisms for the two hub m6A-lncRNAs are unknown. It is required to conduct relevant experiments on the two hub m6A-lncRNAs to elucidate regulatory signaling pathways for metastatic KIRC.

CONCLUSION

In the present study, hub m6A-lncRNAs in KIRC were screened based on WGCNA, the m6A-lncRNAs prognostic index (m6AlRsPI) was constructed and validated according to two hub m6A-lncRNAs linked with prognosis, and the biological signatures and prognostic values were investigated. The two hub m6A-lncRNAs can be looked upon as critical regulatory molecules in metastatic KIRC and as potential therapeutic targets.

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

Conception and design: GL, GTL, and HW. Data collection: GTL HW, YW, and KW. Data analysis and interpretation: GTL, HW, YW, and KW. Manuscript writing: GTL and HW. Final approval of manuscript: All authors.

FUNDING

This study was supported by National Natural Science Foundation of China (grant no.8180101195).

ACKNOWLEDGMENTS

We greatly appreciate the TCGA program, GEO database, the RMVar database, ENCORI, RBPDB database, and catRAPID database for providing the open-source data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.682471/full#supplementary-material

REFERENCES

Bach, D.-H., and Lee, S. K. (2018). Long Noncoding RNAs in Cancer Cells. *Cancer Lett.* 419, 152–166. doi:10.1016/j.canlet.2018.01.053

Camacho, C. V., Choudhari, R., and Gadad, S. S. (2018). Long Noncoding RNAs and Cancer, an Overview. *Steroids* 133, 93–95. doi:10.1016/j.steroids.2017.12.012

Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2019). The Role of m6A RNA Methylation in Human Cancer. *Mol. Cancer* 18 (1), 103. doi:10.1186/s12943-019-1033-z

Gutschner, T., and Diederichs, S. (2012). The Hallmarks of Cancer. *RNA Biol.* 9 (6), 703–719. doi:10.4161/rna.20481

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

Hua, W., Zhao, Y., Jin, X., Yu, D., He, J., Xie, D., et al. (2018). METTL3 Promotes Ovarian Carcinoma Growth and Invasion through the Regulation of AXL Translation and Epithelial to Mesenchymal Transition. *Gynecol. Oncol.* 151 (2), 356–365. doi:10.1016/j.ygyno.2018.09.015

Huang, X., Cai, W., Yuan, W., and Peng, S. (2020). Identification of Key lncRNAs as Prognostic Prediction Models for Colorectal Cancer Based on LASSO. *Int. J. Clin. Exp. Pathol.* 13 (4), 675–684.

Kim, J., Piao, H.-L., Kim, B.-J., Yao, F., Han, Z., Wang, Y., et al. (2018). Long Noncoding RNA MALAT1 Suppresses Breast Cancer Metastasis. *Nat. Genet.* 50 (12), 1705–1715. doi:10.1038/s41588-018-0252-3

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Li, J., Chen, F., Peng, Y., Lv, Z., Lin, X., Chen, Z., et al. (2020). N6-Methyladenosine Regulates the Expression and Secretion of TGFβ1 to Affect the Epithelial-Mesenchymal Transition of Cancer Cells. *Cells* 9 (2), 296. doi:10.3390/cells9020296

Li, T., Hu, P.-S., Zuo, Z., Lin, J.-F., Li, X., Wu, fnm., et al. (2019). METTL3 Facilitates Tumor Progression via an m6A-igf2bp2-dependent Mechanism in Colorectal Carcinoma. *Mol. Cancer* 18 (1), 112. doi:10.1186/s12943-019-1038-7

Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R. I. (2016). The M 6 A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cel.* 62 (3), 335–345. doi:10.1016/j.molcel.2016.03.021

Lindeboom, R. G. H., Supek, F., and Lehner, B. (2016). The Rules and Impact of Nonsense-Mediated mRNA Decay in Human Cancers. *Nat. Genet.* 48 (10), 1112–1118. doi:10.1038/ng.3664

Ljungberg, B., Campbell, S. C., Cho, H. Y., Jacqmin, D., Lee, J. E., Weikert, S., et al. (2011). The Epidemiology of Renal Cell Carcinoma. *Eur. Urol.* 60 (4), 615–621. doi:10.1016/j.eururo.2011.06.049

Luo, X., Li, H., Liang, J., Zhao, Q., Xie, Y., Ren, J., et al. (2020). RMVar: an Updated Database of Functional Variants Involved in RNA Modifications. *Nucleic Acids Res.* 49, D1405–D1412. doi:10.1093/nar/gkaa811

Meyer, K. D., and Jaffrey, S. R. (2014). The Dynamic Epitranscriptome: N6-Methyladenosine and Gene Expression Control. *Nat. Rev. Mol. Cel Biol* 15 (5), 313–326. doi:10.1038/nrm3785

Pan, L., Li, Y., Jin, L., Li, J., and Xu, A. (2020). TRPM2-AS Promotes Cancer Cell Proliferation through Control of TAF15. *Int. J. Biochem. Cel Biol.* 120, 105683. doi:10.1016/j.biocel.2019.105683

Peng, P.-H., Chieh-Yu Lai, J., Hsu, K.-W., and Wu, K.-J. (2020). Hypoxia-induced lncRNA RP11-390F4.3 Promotes Epithelial-Mesenchymal Transition (EMT) and Metastasis through Upregulating EMT Regulators. *Cancer Lett.* 483, 35–45. doi:10.1016/j.canlet.2020.04.014

Qin, Y., Liu, X., Pan, L., Zhou, R., and Zhang, X. (2019). Long Noncoding RNA MIR155HG Facilitates Pancreatic Cancer Progression through Negative Regulation of miR-802. *J. Cel Biochem* 120 (10), 17926–17934. doi:10.1002/jcb.29060

Rini, B. I., Campbell, S. C., and Escudier, B. (2009). Renal Cell Carcinoma. *The Lancet* 373 (9669), 1119–1132. doi:10.1016/S0140-6736(09)60229-4

Shang, A., Wang, W., Gu, C., Chen, C., Zeng, B., Yang, Y., et al. (2019). Long Non-coding RNA HOTTIP Enhances IL-6 Expression to Potentiate Immune Escape of Ovarian Cancer Cells by Upregulating the Expression of PD-L1 in Neutrophils. *J. Exp. Clin. Cancer Res.* 38 (1), 411. doi:10.1186/s13046-019-1394-6

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

Tang, B., Yang, Y., Kang, M., Wang, Y., Wang, Y., Bi, Y., et al. (2020). m6A Demethylase ALKBH5 Inhibits Pancreatic Cancer Tumorigenesis by Decreasing WIF-1 RNA Methylation and Mediating Wnt Signaling. *Mol. Cancer* 19 (1), 3. doi:10.1186/s12943-019-1128-6

Thomas, J. S., and Kabbinavar, F. (2015). Metastatic Clear Cell Renal Cell Carcinoma: A Review of Current Therapies and Novel Immunotherapies. *Crit. Rev. oncology/hematology* 96 (3), 527–533. doi:10.1016/j.critrevonc.2015.07.009

Wang, H., Deng, Q., Lv, Z., Ling, Y., Hou, X., Chen, Z., et al. (2019). N6-methyladenosine Induced miR-143-3p Promotes the Brain Metastasis of Lung Cancer via Regulation of VASH1. *Mol. Cancer* 18 (1), 181. doi:10.1186/s12943-019-1108-x

Wang, Q., Chen, C., Ding, Q., Zhao, Y., Wang, Z., Chen, J., et al. (2020). METTL3-mediated m6A Modification of HDGF mRNA Promotes Gastric Cancer Progression and Has Prognostic Significance. *Gut* 69 (7), 1193–1205. doi:10.1136/gutjnl-2019-319639

Wang, T., Kong, S., Tao, M., and Ju, S. (2020). The Potential Role of RNA N6-Methyladenosine in Cancer Progression. *Mol. Cancer* 19 (1), 88. doi:10.1186/s12943-020-01204-7

Wang, X., Zhou, J., Xu, M., Yan, Y., Huang, L., Kuang, Y., et al. (2018). A 15-lncRNA Signature Predicts Survival and Functions as a ceRNA in Patients with Colorectal Cancer. *Cmar* Vol. 10, 5799–5806. doi:10.2147/CMAR.S178732

Wang, Y., Lu, J.-H., Wu, Q.-N., Jin, Y., Wang, D.-S., Chen, Y. X., et al. (2019). LncRNA LINRIS Stabilizes IGF2BP2 and Promotes the Aerobic Glycolysis in Colorectal Cancer. *Mol. Cancer* 18 (1), 174. doi:10.1186/s12943-019-1105-0

Wen, S., Wei, Y., Zen, C., Xiong, W., Niu, Y., and Zhao, Y. (2020). Long Non-coding RNA NEAT1 Promotes Bone Metastasis of Prostate Cancer through N6-Methyladenosine. *Mol. Cancer* 19 (1), 171. doi:10.1186/s12943-020-01293-4

Wu, Y., Yang, X., Chen, Z., Tian, L., Jiang, G., Chen, F., et al. (2019). m6A-induced lncRNA RP11 Triggers the Dissemination of Colorectal Cancer Cells via Upregulation of Zeb1. *Mol. Cancer* 18 (1), 87. doi:10.1186/s12943-019-1014-2

Xie, J.-W., Huang, X.-B., Chen, Q.-Y., Ma, Y.-B., Zhao, Y.-J., Liu, L.-C., et al. (2020). m6A Modification-Mediated BATF2 Acts as a Tumor Suppressor in Gastric Cancer through Inhibition of ERK Signaling. *Mol. Cancer* 19 (1), 114. doi:10.1186/s12943-020-01223-4

Yan, J., Huang, X., Zhang, X., Chen, Z., Ye, C., Xiang, W., et al. (2020). LncRNA LINC00470 Promotes the Degradation of PTEN mRNA to Facilitate Malignant Behavior in Gastric Cancer Cells. *Biochem. biophysical Res. Commun.* 521 (4), 887–893. doi:10.1016/j.bbrc.2019.11.016

Yang, Z., Ma, J., Han, S., Li, X., Guo, H., and Liu, D. (2020). ZFAS1 Exerts an Oncogenic Role via Suppressing miR-647 in an m6A-dependent Manner in Cervical Cancer. *Ott* Vol. 13, 11795–11806. doi:10.2147/OTT.S274492

Yi, Y.-C., Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2020). Novel Insights into the Interplay between m6A Modification and Noncoding RNAs in Cancer. *Mol. Cancer* 19 (1), 121. doi:10.1186/s12943-020-01233-2

Zhang, C., Huang, S., Zhuang, H., Ruan, S., Zhou, Z., Huang, K., et al. (2020). YTHDF2 Promotes the Liver Cancer Stem Cell Phenotype and Cancer Metastasis by Regulating OCT4 Expression via m6A RNA Methylation. *Oncogene* 39 (23), 4507–4518. doi:10.1038/s41388-020-1303-7

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional Gene Regulation by mRNA Modifications. *Nat. Rev. Mol. Cel Biol* 18 (1), 31–42. doi:10.1038/nrm.2016.132

Zheng, Z.-G., Xu, H., Suo, S.-S., Xu, X.-L., Ni, M.-W., Gu, L.-H., et al. (2016). The Essential Role of H19 Contributing to Cisplatin Resistance by Regulating Glutathione Metabolism in High-Grade Serous Ovarian Cancer. *Sci. Rep.* 6, 26093. doi:10.1038/srep26093

Zheng, Z.-Q., Li, Z.-X., Zhou, G.-Q., Lin, L., Zhang, L.-L., Lv, J.-W., et al. (2019). Long Noncoding RNA FAM225A Promotes Nasopharyngeal Carcinoma Tumorigenesis and Metastasis by Acting as ceRNA to Sponge miR-590-3p/miR-1275 and Upregulate ITGB3. *Cancer Res.* 79 (18), 4612–4626. doi:10.1158/0008-5472.CAN-19-0799

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.

Copyright © 2021 Lin, Wang, Wu, Wang 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 07 July 2021
doi: 10.3389/fmolb.2021.682084

Comprehensive Molecular Analyses of a Novel Mutational Signature Classification System with Regard to Prognosis, Genomic Alterations, and Immune Landscape in Glioma

Zaoqu Liu[1,2,3†], *Taoyuan Lu*[4†], *Libo Wang*[5†], *Long Liu*[5], *Lifeng Li*[6*] and *Xinwei Han*[1,2,3*]

[1]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Interventional Institute of Zhengzhou University, Zhengzhou, China, [3]Interventional Treatment and Clinical Research Center of Henan Province, Zhengzhou, China, [4]Department of Cerebrovascular Disease, Zhengzhou University People's Hospital, Zhengzhou, China, [5]Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [6]Internet Medical and System Applications of National Engineering Laboratory, Zhengzhou, China

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China
Reviewed by:
Lukas Bunse, German Cancer Research Center (DKFZ), Germany
José Manuel Lopes, Universidade do Porto, Portugal

***Correspondence:**
Xinwei Han
fcchanxw@zzu.edu.cn
Lifeng Li
lilifeng0317@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 17 March 2021
Accepted: 28 June 2021
Published: 07 July 2021

Citation:
Liu Z, Lu T, Wang L, Liu L, Li L and Han X (2021) Comprehensive Molecular Analyses of a Novel Mutational Signature Classification System with Regard to Prognosis, Genomic Alterations, and Immune Landscape in Glioma. Front. Mol. Biosci. 8:682084. doi: 10.3389/fmolb.2021.682084

Background: Glioma is the most common malignant brain tumor with complex carcinogenic process and poor prognosis. The current molecular classification cannot fully elucidate the molecular diversity of glioma.

Methods: Using broad public datasets, we performed cluster analysis based on the mutational signatures and further investigated the multidimensional heterogeneity of the novel glioma molecular subtypes. The clinical significance and immune landscape of four clusters also investigated. The nomogram was developed using the mutational clusters and clinical characteristics.

Results: Four heterogenous clusters were identified, termed C1, C2, C3, and C4, respectively. These clusters presented distinct molecular features: C1 was characterized by signature 1, PTEN mutation, chromosome seven amplification and chromosome 10 deletion; C2 was characterized by signature 8 and FLG mutation; C3 was characterized by signature 3 and 13, ATRX and TP53 mutations, and 11p15.1, 11p15.5, and 13q14.2 deletions; and C4 was characterized by signature 16, IDH1 mutation and chromosome 1p and 19q deletions. These clusters also varied in biological functions and immune status. We underlined the potential immune escape mechanisms: abundant stromal and immunosuppressive cells infiltration and immune checkpoints (ICPs) blockade in C1; lack of immune cells, low immunogenicity and antigen presentation defect in C2 and C4; and ICPs blockade in C3. Moreover, C4 possessed a better prognosis, and C1 and C3 were more likely to benefit from immunotherapy. A nomogram with excellent performance was also developed for assessing the prognosis of patients with glioma.

Conclusion: Our results can enhance the mastery of molecular features and facilitate the precise treatment and clinical management of glioma.

Keywords: glioma, molecular subtype, prognosis, immunotherapy, mutational signature

INTRODUCTION

For Glioma is the most common malignant brain tumor associated with high heterogeneity and poor prognosis (Louis et al., 2016). The standardization regimen consisting of surgery resection combined with chemoradiotherapy is currently the dominant treatment for glioma. However, the overall therapeutic benefits remain unsatisfactory, especially glioblastoma, with a median survival of only 14.6 months after standardization therapy (Stupp et al., 2005). Hence, it is imperative to pursue new means to improve the treatment and management of glioma.

With the rapid development of bioinformatics and the rise of molecular diagnosis, precision therapy and immunotherapy have made it possible to emerge from the current dilemma (Reifenberger et al., 2017; Fecci and Sampson, 2019). The 2016 World Health Organization Classification of Tumors of the Central Nervous System incorporated molecular parameters into traditional histological diagnosis of glioma, dividing gliomas into distinct molecular phenotypes, such as IDH mutant and IDH wild gliomas, 1p/19q co-deletion and 1p/19q integrity gliomas (Louis et al., 2016). Accumulated evidence indicated that glioma patients with IDH mutation and those with 1p/19q co-deletion were relatively more sensitive to radiotherapy and chemotherapy, as well as had a favorable prognosis (Sabha et al., 2014). However, this classification only focuses on one or a few genomic alteration features, which lacks a global perspective, and fails to fully elucidate the high molecular heterogeneity of glioma. Therefore, a systematic exploration of genomic alterations in gliomas is necessary to reveal its molecular heterogeneity.

Over the past decade, immunotherapy has achieved great success in the treatment of tumors (Yang, 2015; Payandeh et al., 2019). In glioma, recent studies have reported that immunotherapy such as anti-PD-1 and anti-VEGFA could predominantly prolong the survival of some glioma patients, but the response population was not stable, only a subset of patients could benefit from immunotherapy (Sandmann et al., 2015; Lyon et al., 2017; Schalper et al., 2019). The immunotherapy limitation may be due to the high heterogeneity of gliomas and their complex immune escape mechanisms (Jackson et al., 2019; Wildes et al., 2020). Thus, investigating the molecular heterogeneity and potential immune escape mechanisms of gliomas may contribute to the development of immunotherapy.

Cancer is a complex disease arise from the constant accumulation of genomic alterations (Stratton et al., 2009). The 30 mutational signatures described by Alexandrov et al. systematically characterized the mutation accumulation leading to tumorigenesis, and linked the mutation process to DNA damage mechanisms and clinical characteristics, providing a new opportunity for in-depth analysis and understanding of the tumor molecular features (Alexandrov et al., 2015). But so far, there was no study have systematically analyzed genomic alterations and dissected mutational signatures of glioma.

Obviously, a deeper grasp of the molecular features and more refined molecular classification are essential for the precise treatment of gliomas, and the development of bioinformatics and the accumulation of broad data make it promising. In the present study, we performed molecular clustering based on the mutational signature profile of glioma patients, hoping to identify distinct molecular heterogeneous subtypes and better understand the biological characteristics of glioma. As a result, we successfully identified four heterogeneous subtypes with specific molecular characteristics, potential immune escape mechanism, and clinical outcomes in glioma. Combining the four subtypes and clinical features, we also constructed a nomogram with excellent performance to facilitate clinical prognosis management.

MATERIALS AND METHODS

Data Source

The glioma data (n = 736) were enrolled from The Cancer Genome Atlas (TCGA) cohorts TCGA-LGG (low grade glioma) and TCGA-GBM (glioblastoma), the details of baseline information was shown in **Supplementary Table S1**. Gene expression data and clinical information were retrieved from TCGA data portal (https://portal.gdc.cancer.gov/). The mutation data, copy number alteration data, and methylation 450K data were acquired from TCGA database. Three independent immunotherapeutic cohorts containing expression and clinical data were collected: (Louis et al., 2016) metastatic melanoma patients treated with cytotoxic T lymphocyte-associated protein 4 (CTLA-4) and PD-1 blockades (Roh cohort) (Roh et al., 2017; Stupp et al., 2005) melanoma patients treated with adoptive T cell therapy (ACT) (GSE100797) (Lauss et al., 2017; Reifenberger et al., 2017) melanoma patients treated with anti-PD-1 (GSE78220) (Hugo et al., 2016). According to the RECIST v1.1 criterion, patients with complete response (CR) or partial response (PR) and patients with stable disease (SD) or progressive disease (PD) were deemed as immunotherapy responders and nonresponders, respectively, and patients who were not evaluable (NE) were removed.

Identification of Mutational Signature Relevant Clusters

Somatic mutation data from TCGA-LGG and TCGA-GBM, removing silent mutations, were used for subsequent analysis. Mutational signatures described by Alexandrov et al. could be obtained from COSMIC website (Alexandrov et al., 2015). We calculated the proportion of 30 mutational signatures for each patient *via* DeconstructSigs package with signature cutoff = 0.06 and "exome2genome" normalization (Rosenthal et al., 2016). Next, the non-negative matrix factorization (NMF) algorithm was employed to perform consensus clustering and feature extraction in this study. Based on mutational signatures, consensus NMF clustering was performed *via* NMF package (Gaujoux and Seoighe, 2010) with the following parameters: potential ranks = 2–5, number of

runs to perform = 50, method = "lee". Ultimately, the optimal rank = 4 was determined by cophenetic coefficient (**Supplementary Figure S1I**). The "nmf" and "extractFeatures" function implemented in NMF package were utilized to extract the basis-specific features of each basis component. Out of 30 mutational signatures, 11 signatures were identified in above analysis based on method = "max" from Carmona-Saez et al. (2006). To investigate the importance of these signatures for each clinical characteristic, the decision tree C5.0 algorithm was performed with the C50 package, iterating 10 times. In addition, the APOBEC enrichment analysis described by Roberts et al. was further performed by the Maftool package (Mayakonda et al., 2018).

Genomic Alterations Analysis

The prediction of driver genes was performed by the OncodriveCLUST, an algorithm to identify candidate driver genes with a significant bias towards mutation clustering within the protein sequence (Tamborero et al., 2013). Genes with q values less than 0.05 and mutation frequency greater than 2% or genes with mutation frequency greater than 5% were considered as drivers in this study. The GISTIC2.0 (Mermel et al., 2011) was applied to examine recurrently amplified and deleted regions, and the regions altered in greater than 15% of the samples were included in further analysis.

Functional Annotation and Immune Infiltration Assessment

To investigate the biological behaviors among the four clusters, the gene set enrichment analysis (GSEA) was conducted based on the Hallmark gene sets ("h.all.v7.1. symbols.gmt"), and the biological function with FDR <0.05 was significance. We also explored the correlation between clusters and other related biological processes, using the gene sets proposed by Mariathasan et al. (2018) (**Supplementary Table S2**). Single sample gene set enrichment analysis (ssGSEA) algorithm implemented in GSVA package was applied to estimate the relative infiltration abundance of tumor microenvironment (TME) cells. The gene sets for marking 28 immune cell types were enrolled from Charoentong et al. (2017) (**Supplementary Table S3**). As endothelial cell and fibroblasts were also critical components of TME, we included another 40 genes from MCP-Counter gene list to mark these two cell types (Becht et al., 2016) (**Supplementary Table S3**).

Collection and Investigation of Immune Escape Indicators

A series of tumor immune-related indicators (**Supplementary Table S4**), including stromal and leukocyte fractions, nonsilent mutation rate, neoantigen burden, cancer testis antigens (CTA) score, aneuploidy score, intratumor heterogeneity, number of segments (Segs), number or fraction of segments with loss of heterozygosity (LOH), fraction altered, homologous recombination deficiency (HRD), TCR diversity (Shannon Entropy and Richness) score (Thorsson et al., 2018), microsatellite instability (MSI) score (Bonneville et al., 2017), cytolytic activity (Rooney et al., 2015), antigen processing and presenting machinery score (APS) (Wang et al., 2019) and the expression of immunomodulator molecules (Thorsson et al., 2018), were enrolled or calculated for the investigation of potential immune escape mechanisms in the four clusters. Moreover, multi-omics regulation of 75 immunomodulator molecules was further analyzed (**Supplementary Table S5**), including somatic mutation, copy number variation (CNV) and DNA methylation.

Clinical Relevance of the Four Clusters

Univariate and multivariate Cox regression analysis were performed to assess the prognostic significance of clusters and other vital clinical features. We included the features with multivariate Cox p value < 0.05 to construct a nomogram and further evaluated its performance by the calibration and receiver operating characteristic (ROC) curves. Subsequently, the package pRRophetic, which can predict the patients' chemotherapeutic response based on a ridge regression (Geeleher et al., 2014), was employed to estimate the sensitivity of the four clusters to gemcitabine and bortezomib. Drug sensitivity was quantified by half-maximal inhibitory concentration (IC50), the lower the IC50, the more sensitive. Two methods were further employed to predict the immunotherapeutic response of the four clusters. First, the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was utilized to predict the immunotherapeutic response of each patient (Jiang et al., 2018). Second, subclass mapping (Hoshida et al., 2007) was applied to assess the similarity of gene expression profiles between the four clusters and three published immunotherapy cohorts (GSE100797, GSE78220, and Roh cohorts), if a pair's expression profiles shared the more similarity, their clinical outcomes were more likely to be similar.

Statistical Analysis

The co-occurrence or exclusion of driver mutations were evaluated by Fisher exact test. Spearman correlation analyses were applied to compute the correlation coefficients of two variables. The comparisons of two groups were conducted by Wilcoxon rank-sum test, and when three or more groups, Kruskal–Wallis test was employed. The Kaplan–Meier method was applied to generate survival curves for prognosis analyses, and the log-rank test was used to define the significance of differences. The hazard ratios for variables were calculated by univariate Cox regression analyses, and multivariate Cox regression was employed to ascertain independent prognostic factors. The receiver operating characteristic (ROC) curves for survival variables were plotted by the timeROC R package. All heatmap in this study were plotted by the ComplexHeatmap package (Gu et al., 2016). All statistical p values were two-sided, and $p < 0.05$ was deemed as statistically significance. P-adjust value was obtained by Benjamini-Hochberg (BH) multiple test correction. All data processing was completed in R 3.6.3 software.

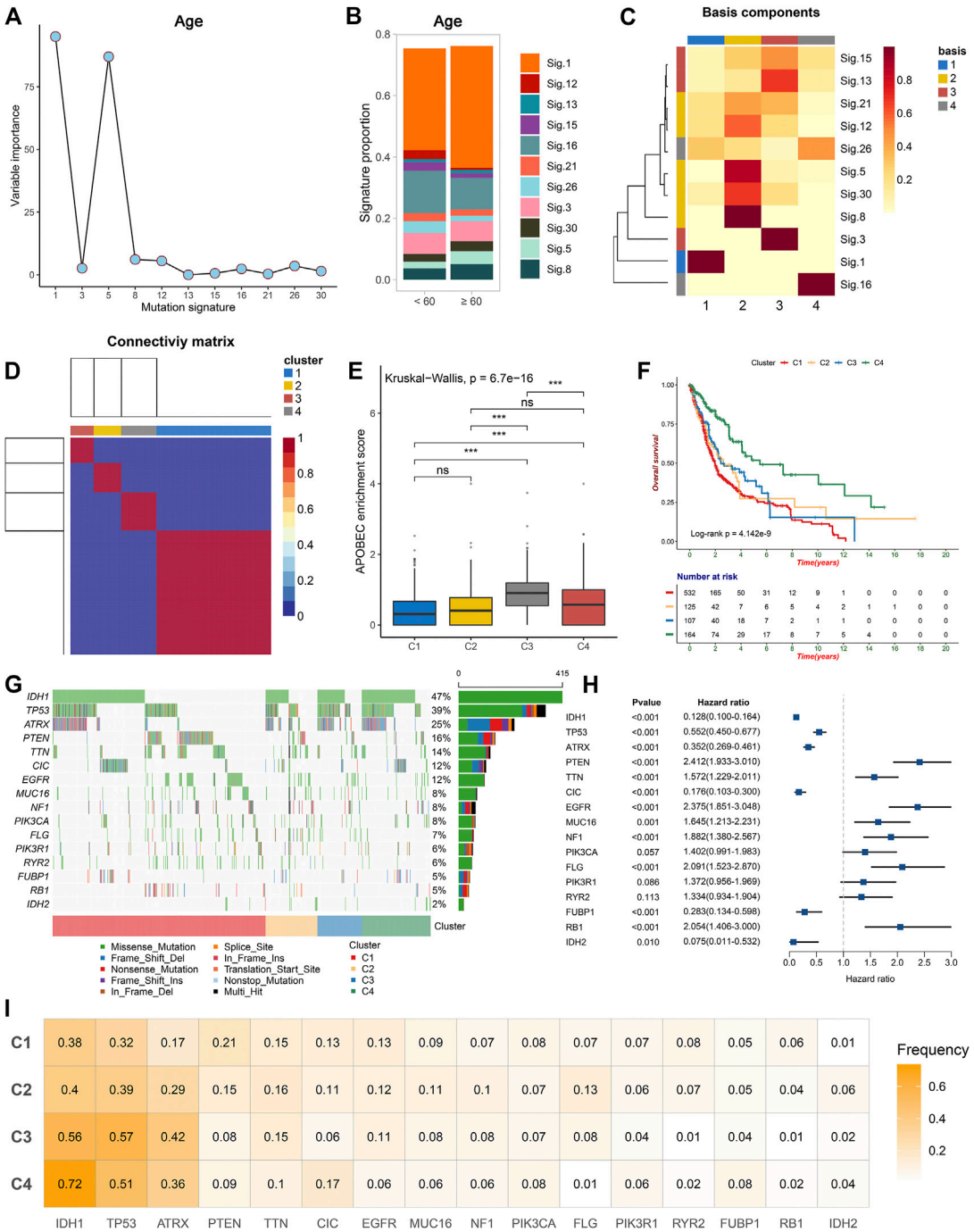


FIGURE 1 | The mutational signatures and driver genes of four clusters. **(A)** Importance of 11 extracted mutational signatures in distinguishing patients of different age groups. **(B)** Distribution of 11 signatures in different age groups. **(C)** Basis component of NMF with rank = 4 in TCGA glioma cohort. **(D)** Consensus matrix after clustering revealed four clusters with no overlap between clusters. **(E)** The distribution of APOBEC enrichment score among the four clusters. The asterisks represent the statistical *p* value (ns, $p > 0.05$; ***$p < 0.001$). **(F)** Kaplan–Meier curves for OS among the four clusters in TCGA glioma cohort **(G)** Mutation landscape of 16 candidate driver genes in the four clusters. **(H)** Univariate Cox regression analysis of 16 candidate driver genes. **(I)** Mutation frequency of 16 candidate driver genes in the four clusters.

RESULTS

Identification of Mutational Signatures Related Clusters

The somatic mutation landscape of 892 glioma patients was summarized in this study (**Supplementary Figure S1A**). A total of 56,369 somatic mutations were detected, and missense mutation occupied the dominant fraction. In single nucleotide variation (SNV), C > T displayed the highest frequency followed by T > C and C > A, which was consistent with the result of transition (Ti) and transversion (Tv) (**Supplementary Figure S1B**). In addition, we observed 14,852 mutant genes in total, and 26 genes with a mutation frequency of more than 3%. To better comprehend the contribution of these mutations to glioma, an in-depth exploration based on mutational signatures was conducted. Using the decision tree algorithm, we evaluated the importance of 11 extracted mutational signatures for various clinical characteristics in glioma (**Figures 1A,B**, **Supplementary Figures S1C–H**). There was no gender related mutational signature. Signature 1 and signature 5 were previously considered to correlate with age of cancer diagnosis, consistent with our findings in glioma (**Figures 1A,B**). We also observed that signature 1 was significantly associated with IDH mutation, TERT promoter mutation, MGMT promoter methylation, and the combination of chromosome seven gain and 10 loss (7+/10−). Based on previous studies, signature 3 was associated with homologous recombination deficiency (HRD), signature 13 was associated with the activation of apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC), and signature 16 was related to alcohol and transcription-coupled damage (Letouzé et al., 2017). In this study, we also found that signature 3 was linked to 7+/10−, signature 13 was linked to ATRX mutation and MGMT promoter methylation, and signature 16 was linked to grade in glioma. The proposed etiology of signature 8 was unknown yet, but we detected it was related to MGMT promoter methylation (**Supplementary Figures S1C–H**). These results suggested that different mutational signatures may be associated with specific etiology and clinical characteristics in glioma.

Consensus NMF clustering was further performed to identify heterogeneity clusters based on the fraction of 11 mutational signatures, and eventually four clusters were determined, termed as C1, C2, C3, and C4 (**Figures 1C,D**). C1 was characterized by signature 1, which was initiated by spontaneous deamination of 5-methylcytosine and correlated with age. C2 was characterized by signature 8, which associated with MGMT promoter methylation. C3 was characterized by signature 3 and signature 13, linked to HRD and the activation of APOBEC. The APOBEC enrichment results further revealed that C3 exhibited a higher score (**Figure 1E**). C4 was characterized by signature 16 associated with drinking and transcription-coupled damage (**Figure 1C**). Survival analysis revealed the different prognosis of four clusters, and C4 exhibited a favorable prognosis (**Figure 1F**).

Mutation Driver Genes

The tumor mutation burden (TMB) investigation displayed a decreasing trend from C1 to C4, although the difference between four clusters was not very pronounced (**Supplementary Figure S2A**). A total of 16 candidate driver genes were identified (**Supplementary Table S6**; **Supplementary Figure S2B**; **Figure 1G**). It was noted that there was a significant mutation co-occurrence between IDH1, ATRX, and TP53, which often appeared in astrocytoma. We also observed an exclusion between IDH1 and EGFR, IDH1 and PTEN, along with a co-occurrence between EGFR and PTEN, which often appeared in glioblastoma (**Supplementary Figure S2C**) (Diplas et al., 2018). We visualized the hotspots and other mutations in glioma *via* lollipop plots (**Supplementary Figure S2D**). For example, the IDH1 mutation focused on residue 132 of PTZ00435, consistent with previous cognition (Kloosterhof et al., 2011). Univariate cox regression and survival analysis further revealed the prognostic value of these 16 driver genes (**Figure 1H**; **Supplementary Figure S2E**). Out of these genes, IDH1, IDH2, TP53, ATRX, CIC, and FUBP1 mutations were favorable factors for prognosis, while others were poor factors. In addition, we also investigated the mutation frequency of the driver genes in each cluster. It was found that C1 was characterized by PTEN, C2 was characterized by FLG, C3 was characterized by TP53 and ATRX, and C4 was characterized by IDH1 (**Figure 1I**). In C4, genes that favor prognosis exhibited a relatively higher mutation frequency, while gene mutations that disfavor prognosis had a relatively low frequency, which may constitute an explanation for the better survival status of C4.

Significantly Altered Segments

The GISTIC algorithm revealed the landscape of significantly recurrent amplification and deletion in glioma (**Supplementary Tables S7,S8**; **Supplementary Figures S3A,B**). A total of 35 segments with alteration frequency more than 15% were selected for further analysis (**Figure 2A**), and univariate Cox regression assessed the prognostic significance of these segments (**Supplementary Figure S3C**). Of these, we revealed for the first time that in glioma, gains of 12q14.1 and losses of 6q22.31, 6q26, 13q14.2, 13q22.1, 15q14, and 22q13.32 were significantly linked to a poor prognosis whereas loss of 4q34.3 was linked to a favorable prognosis. C1 was characterized by the most frequent alterations encompassing five amplifications on chromosome seven and four deletions on chromosome 10, all linked to poor prognosis (**Figures 2B,C**; **Supplementary Figure S3C**). Oncogenes such as CDK6 (7q21.2) and MET (7q31.2) were appreciably amplified, whereas tumor suppressor genes (TSGs) such as CDKN2A/2B (9p21.3) and PTEN (10q23.31) were significantly deleted, which may contribute to poor outcomes in patients with chromosome seven gain, chromosome 9p loss or chromosome 10 loss (**Supplementary Tables S7,S8**; **Figures 2D–H**; **Supplementary Figures S3D–G**). The combined chromosome seven gain and chromosome 10 loss (7+/10−) was altered frequently in some gliomas (Stichel et al., 2018), we thus investigated the incidence of 7+/10− among four clusters and found that it was highest in C1 (p = 2.33e-12)

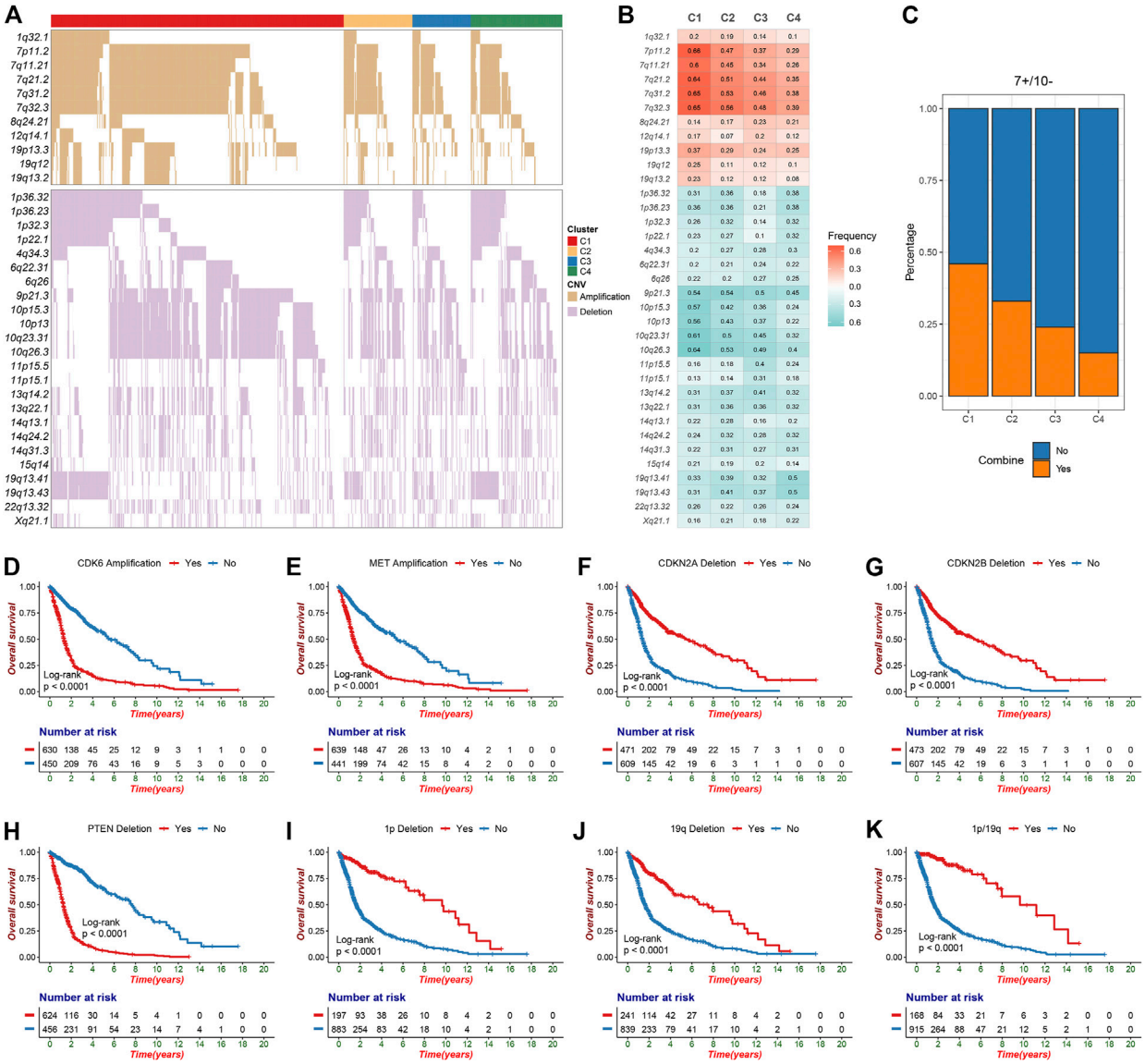


FIGURE 2 | The significant recurrent segments obtained from GISTIC algorithm in TCGA glioma cohort. **(A)** The oncoplot of 35 frequently segments in the four clusters. **(B)** Frequency of amplification (red) and deletion (green) among the four clusters. **(C)** The distribution of combined chromosome seven gain and 10 loss in the four clusters. (**D–H)** Kaplan-Meier survival analysis of CDK6 **(D)** and MET **(E)** amplifications as well as CDKN2A (**F**), CDKN2B **(G)** and PTEN **(H)** deletions. **(I-K)** Chromosome 1p deletion **(I)**, chromosome 19q deletion **(J)** or 1p/19q co-deletion **(K)** were associated with poor overall survival in glioma patients.

(**Figure 2C**). C3 exhibited higher frequency of 11p15.1, 11p15.5, and 13q14.2 deletions in contrast to the other clusters. Four deletions on 1p and two on 19q which linked to a favorable prognosis were most frequently altered in C4 (**Figure 2B**; **Supplementary Figure S3C**). Of note, patients with TSG-associated deletions including 1p32.3 (CDKN2C), 1p36.23 (ERRFI1), 1p36.32 (AJAP1 and HES3), and 19q13.41 (PPP2R1A) had a survival advantage over no deletions (**Figures 2I–K**; **Supplementary Table S8**). This can be explained by the fact that 1p/19q co-deletion was a driving event in oligodendroglioma but could increase patient sensitivity to chemoradiotherapy (Kaloshi et al., 2007; Barthel et al., 2018). In addition, another deletion significantly associated with a favorable prognosis, 4q34.3, also had the highest alteration frequency in C4.

Biological Characteristics of the Four Clusters

To explore and characterize the biological behaviors among the four clusters, we performed GSEA enrichment analysis. C1 was enriched in stromal and immune activation relevant pathways, such as angiogenesis, epithelial mesenchymal transition, hypoxia, IL-6 JAK STAT3 signaling, and IFN-γ response (**Figure 3A**). C2 presented pathways involved in promoting proliferation such as

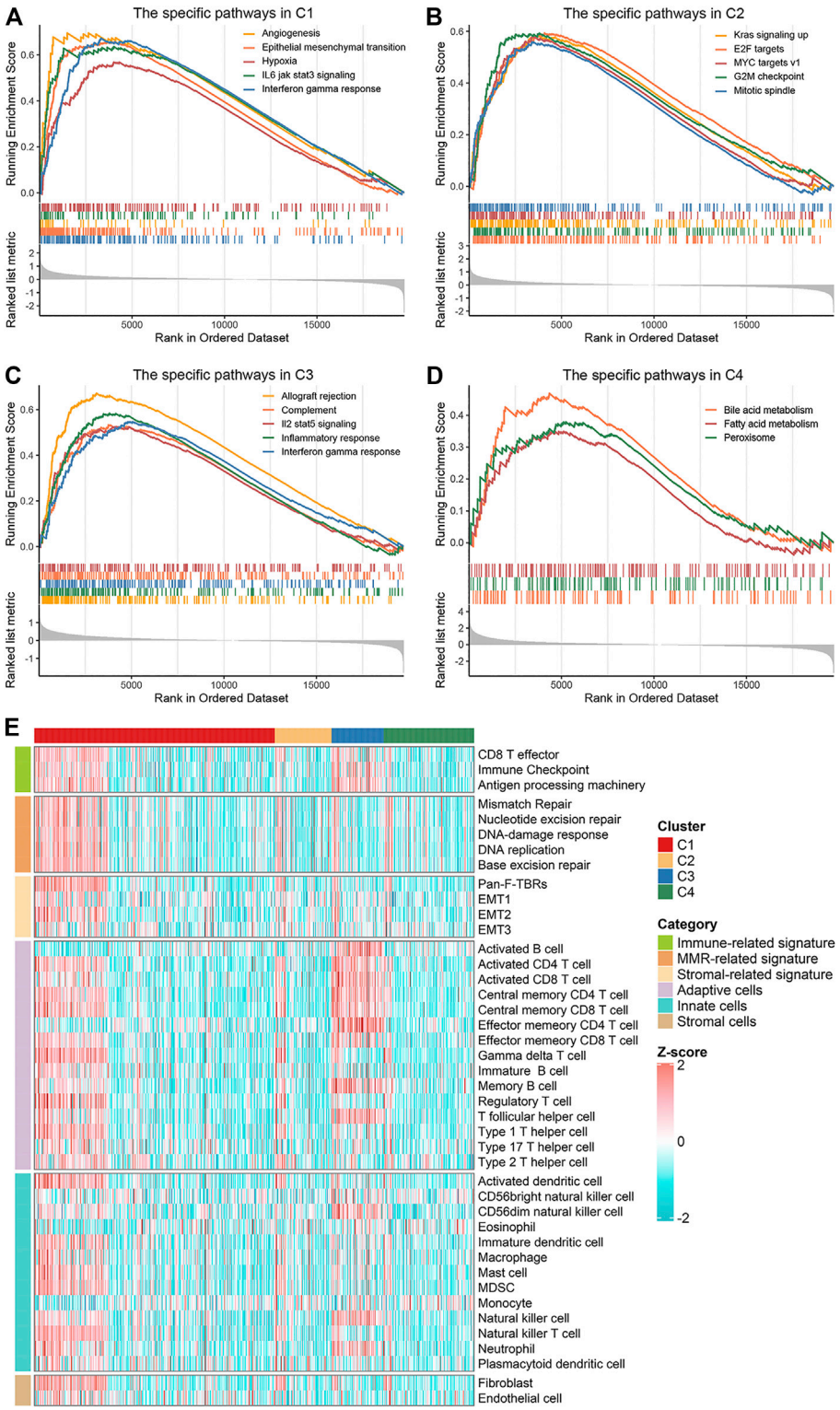


FIGURE 3 | Biological characteristics of the four clusters. **(A–D)** GSEA enrichment analysis revealed activated Hallmark pathways of C1 **(A)**, C2 **(B)**, C3 **(C)** and C4 **(D)**, the FDR of the biological function was <0.05. **(E)** The distribution of known signatures (immune-relevant signatures, mismatch-relevant signatures, and stromal-relevant signatures) and TME cells assessment (adaptive immune cells, innate immune cells and stromal cells) in the four clusters.

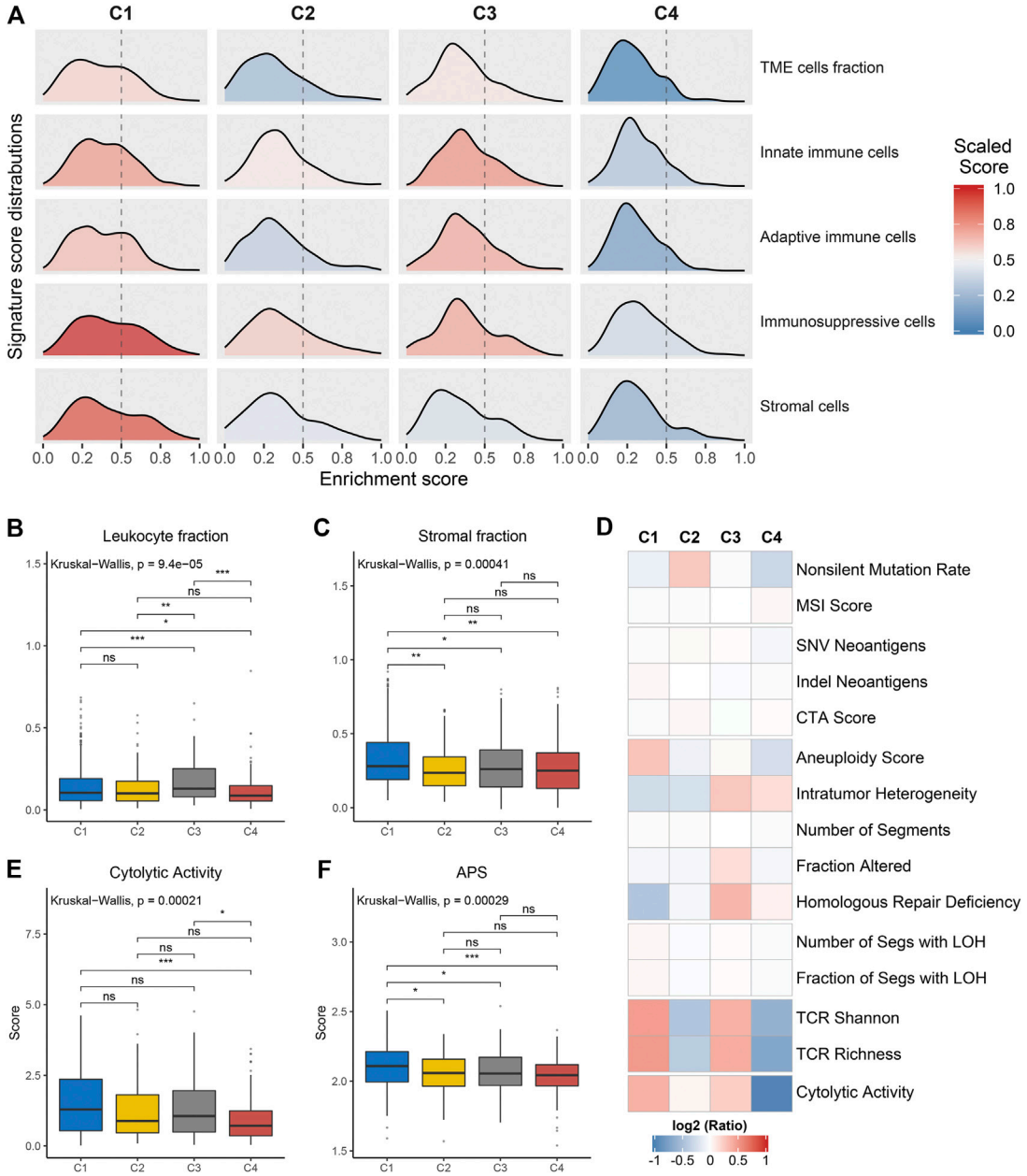


FIGURE 4 | Potential extrinsic immune escape mechanisms of each cluster. **(A)** The scaled signature score distributions of five cell subsets among the four clusters. **(B,C)** Comparison of leukocyte fraction **(B)** and stromal fraction **(C)** among the four clusters. **(D)** Comparison of 14 immunogenicity associated indicators among the four clusters, the cell represented by the mean value of corresponding cluster divided by the overall mean value. **(E,F)** Comparison of cytolytic activity **(E)** and APS **(F)** among the four clusters. For all boxplots, the asterisks represent the statistical *p* value (ns, $p > 0.05$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

KRAS signaling, E2F targets and G2M checkpoint (**Figure 3B**). C3 displayed intense pathways associated with immune activation encompassing allograft rejection, complement, IL-2 STAT5 signaling, inflammatory response, and IFN-γ response (**Figure 3C**). C4 was enriched in pathways pertinent to metabolism such as bile acid metabolism, fatty acid metabolism, and peroxisome (**Figure 3D**). We also examined known signatures in four clusters to better understand the functionality of them. Stroma-related signatures such as epithelial-mesenchymal transition (EMT) and pan-fibroblast TGF beta response (Pan-F-TBR), and mismatch repair (MMR)-related signatures such as base excision repair, were markedly enhanced in C1; immune-related signatures such as antigen processing machinery, CD8 T effector and immune-checkpoint were appreciably enhanced in C3 (**Figure 3E**, **Supplementary Figure S4A**). This confirmed the findings in GSEA. Of note, C1 exhibited not only stromal activation but also immune activation. Analysis of TME cell infiltration

demonstrated that immune cells such as activated $CD4^+$ T cell and activated $CD8^+$ T cell were most abundant in C3, followed by C1; stromal activation-associated cells such as endothelial cell and fibroblasts were most enriched in C1 (**Figure 3E**, **Supplementary Figure S4B**). This confirmed again that C1 was characterized by stromal and immune dual activation, while C3 was characterized by immune activation. Moreover, C1 also displayed the highest infiltration of MDSC and regulatory T cell (**Figure 3E**, **Supplementary Figure S4B**). We speculated that with the stromal activation, C1 may progress from an immune activation state similar to C3, towards an immunosuppressive state. Overall, C1 was classified as stromal and immune dual activation phenotype, C2 was classified as proliferation phenotype, C3 was classified as immune activation phenotype, and C4 was classified as metabolism phenotype.

Potential Extrinsic Immune Escape mechanisms

We further investigated the specific immune escape mechanisms of each subtype. Extrinsic immune escape mainly encompassed absence of immune cells, emergence of immunosuppressive cells, and high abundance of stromal cells (Schumacher and Schreiber, 2015; Wang et al., 2019). We pooled together the relative abundance of TME cells among four clusters. C1 and C3 exhibited a high level of TME cells, innate immune cells and adaptive immune cells, which were considered as immune-hot phenotypes. Whereas C2 and C4 demonstrated overall low TME cell levels, which were considered as immune-cold phenotypes (**Figure 4A**, **Supplementary Figures S5A–C**). In addition, C1 also displayed significantly superior levels of immunosuppressive cells and stromal cells (**Supplementary Figures S5D,E**), suggesting an immune-hot but suppressive microenvironment. This phenomenon may contribute to the extrinsic immune escape of C1. The stromal and leukocyte fractions from Thorsson et al. study also indicated that C1 was characterized by high stromal fraction while C3 was characterized by high leukocyte fraction, further validating the above findings (**Figures 4B,C**). In term of C2 and C4, the lack of immune cells implied an inability to immunologically eliminate tumor. Molecules associated with initiation of innate immunity, such as CLEC7A, PYCARD and TLR2, were also relatively low expressed in these two clusters (**Supplementary Figure S5F**). These results illustrated that absence of immune cells may be an extrinsic immune escape mechanism for C2 and C4.

Potential Intrinsic Immune Escape mechanisms

The exploration of intrinsic immune escape mechanism mainly included three major aspects: tumor immunogenicity, antigen presentation capacity and immune checkpoint molecules expression (Schumacher and Schreiber, 2015). We first assessed a series of factors associated with tumor antigenicity, including mutations, MSI, neoantigens, CTA, and CNV-related indicators (**Figure 4D**). C4 exhibited a lower rate of nonsilent mutation compared to C1 and C2 ($p < 0.05$; **Figure 4D**, **Supplementary Figure S6A**). MSI score displayed a decreasing trend from C4 to C1, although it was not significant (**Figure 4D**, **Supplementary Figure S6B**). Neoantigens and CTA were also vital source of tumor-specific antigens, but they were not significantly different between the four clusters (**Figure 4D**, **Supplementary Figures S6C–E**). C1 presented higher aneuploidy score and fraction of segments with LOH, in contrast to the other three clusters ($p < 0.05$; **Figure 4D**, **Supplementary Figure S6F**, **Supplementary Figure S6L**). C3 exhibited a high level of homologous recombination deficiency, consistent with its mutation cluster characteristics (mutational signature 3) (**Figure 4D**, **Supplementary Figure S6J**). In addition, TCR diversity and cytolytic activity were applied to further assess tumor immunogenicity (Rooney et al., 2015). C2 and C4 exhibited a lack of TCR diversity and low cytolytic activity, as opposed to C1 and C3 (**Figures 4D,E**, **Supplementary Figures S6M,N**). Overall, C2 and C4 displayed lower immunogenicity, which may be an intrinsic immune escape mechanism for these two clusters. In term of antigen processing and presenting machinery, C1 exhibited the highest APS while C2 and C4 were quite the opposite ($p < 0.05$; **Figure 4F**). Expression of MHC molecular were also relatively low in C2 and C4 (**Figure 5A**). Of note, corresponding to MHC loss, TCR diversity was also lacking in C2 and C4. The absence of MHC stimulation may be responsible for the scarcity of TCR diversity in these two clusters (**Figure 4D**). Therefore, we believed the defect of antigen presentation capacity may be another intrinsic immune escape mechanism for C2 and C4.

Multi-Omics Analysis of Immunomodulators

The expression and regulation of immune checkpoint molecules were also a crucial intrinsic immune escape mechanism (**Figure 5A**). In this research, the expression of immunomodulators varied across the clusters, with the vast majority being highly expressed in C1 and C3, but quite low in C4 (**Figure 5A**). C1 had higher expression of many stimulators (e.g., CD80, CCL5, CD70, and PRF1) and inhibitors (e.g., CD274 and VEGFA) compared with the other three clusters. C3 exhibited markedly high expression of stimulators such as TLR4, and inhibitors such as C10orf54, CTLA4 and HAVCR2. These results hinted that C1 and C3 may escape immune elimination by overexpressing immune inhibitors after stimulating immune activation, implying an intrinsic immune escape mechanism for these two clusters.

To advance this investigation, we further analyzed the multi-omics features of the immunomodulators among the four clusters (**Figure 5A**). Most immunomodulators presented rare somatic mutations. In term of CNVs, C1 exhibited frequent amplification and deletion of many immunomodulator genes such as CD70, ICAM1, C10orf54, etc., in line with the high genomic instability of C1. Of note, we found the expression levels of CD70, CD40, and ICAM1 with amplification were higher than those without amplifications, while C10orf54 with deletion displayed lower expression level relative to no deletion (**Figures 5A–C**, **Supplementary Figures S6O,P**). This phenomenon indicated that CNVs played a non-negligible role in regulating the expression of certain immunomodulators. We also detected

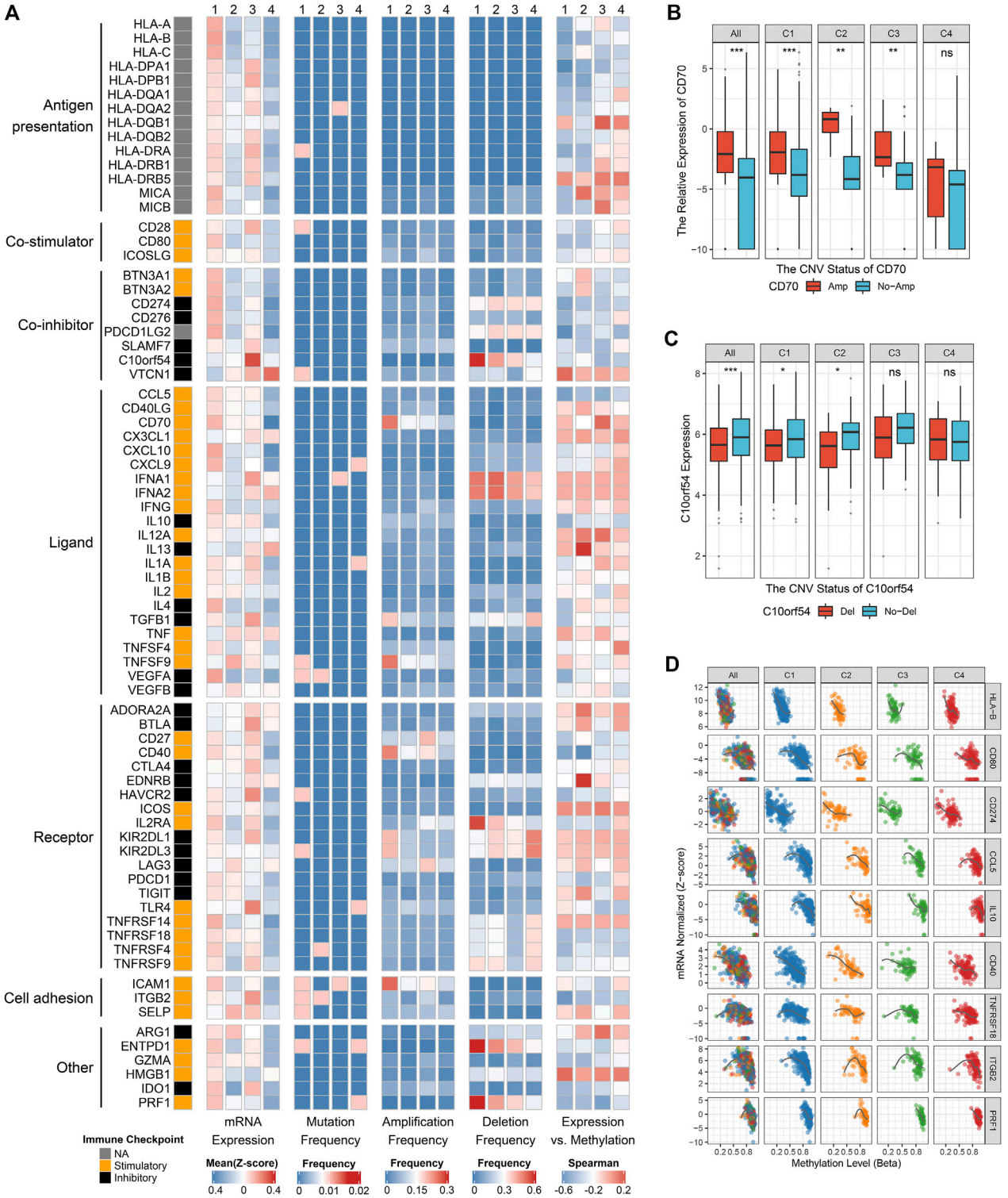


FIGURE 5 | Multi-omics analysis of 75 immunomodulators in glioma. **(A)** From left to right: mRNA expression (z-score), mutation frequency, amplification frequency, deletion frequency, and expression vs. methylation (gene expression correlation with DNA-methylation beta-value) of 75 immunomodulators in the four clusters. **(B)** Comparison of CD70 relative expression between amplification and non-amplification groups. **(C)** Comparison of C10orf54 relative expression between deletion and non-deletion groups. **(D)** Correlation analysis of DNA methylation levels and mRNA expression levels for HLA-B, CD80, CD274, CCL5, IL10, CD40, TNFRSF18, ITGB2, and PRF1. For boxplot, the asterisks represent the statistical *p* value (ns, $p > 0.05$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

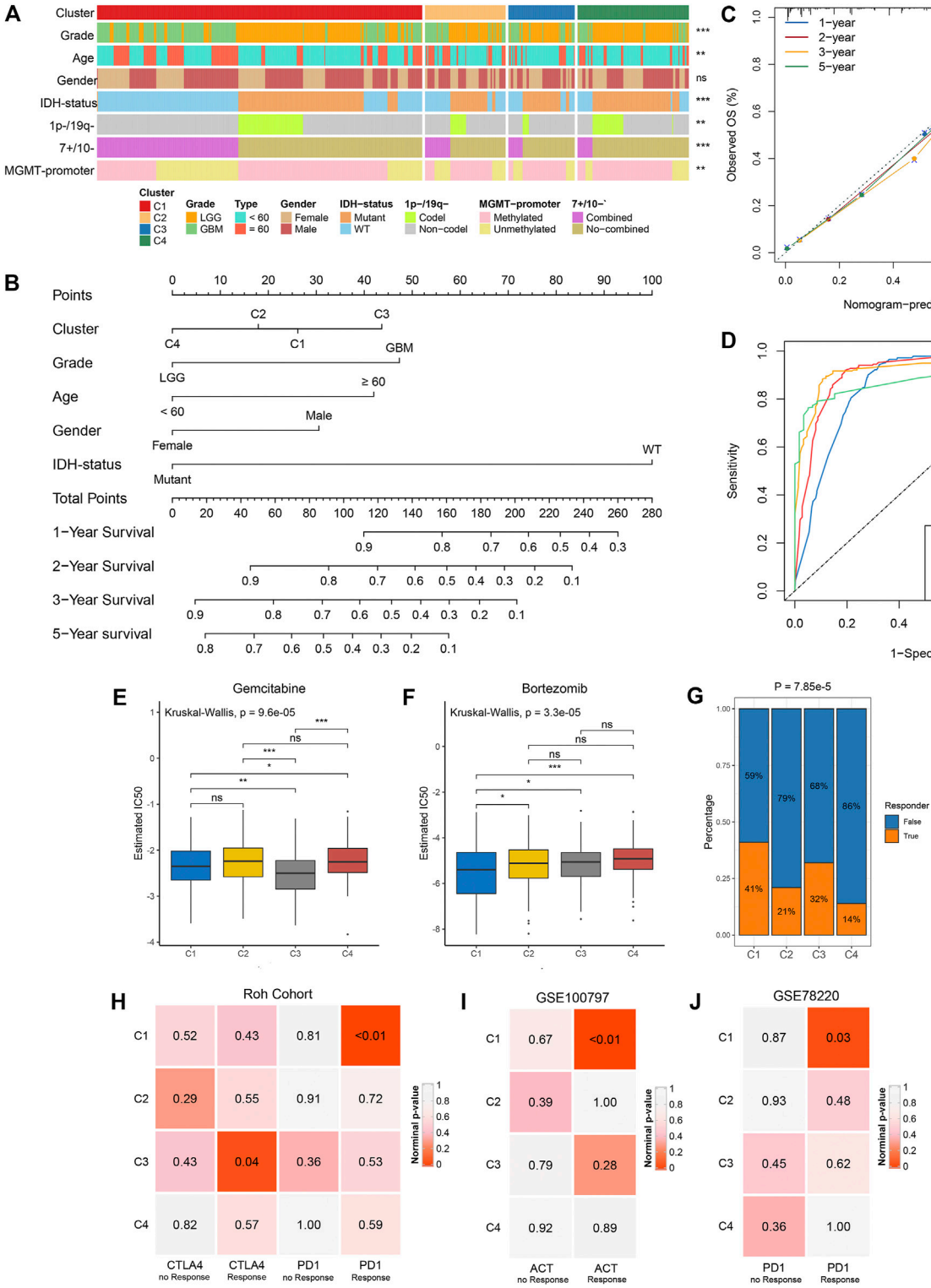


FIGURE 6 | Clinical relevance of the four clusters in TCGA glioma cohort. **(A)** Distribution of grade, age, gender, IDH status, 1p-/19q–, 7+/10– and MGMT-promoter methylation in the four clusters. The right asterisks represent the statistical *p* value (ns, $p > 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$) for significance of the difference among clusters. **(B)** The nomogram for predicting the 1-, 2-, 3- ,and 5-years survival possibility of individuals. **(C,D)** Calibration curve **(C)** and ROC curve **(D)** for evaluating the performance of nomogram. **(E,F)** The estimated IC50 of gemcitabine **(E)** and bortezomib **(F)** among the four clusters. The asterisks represent the statistical *p* value (ns, $p > 0.05$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$). **(G)** The distribution of the immunotherapy responders predicted by TIDE algorithm in the four clusters. **(H)** Submap analysis of the four clusters and Roh cohort with detailed anti-PD1 and anti CTLA4 therapy information. **(I)** Submap analysis of the four clusters and GSE100797 with detailed ACT information. **(J)** Submap analysis of the four clusters and GSE78220 with detailed anti-PD1 therapy information. For submap analysis, a smaller *p*-value implied a more similarity of paired expression profiles.

that DNA methylation levels of many immunomodulator genes, such as HLA-B, CD80, CD274, CCL5, IL10, CD40, TNFRSF18, ITGB2, and PRF1, were inversely correlated with their gene expression levels, implying their essential role by epigenetic silencing (**Figures 5A,D**). In summary, CNVs and methylation modification were prominent participants in the regulation of immunomodulators, which suggested a novel orientation for the development of immune therapy.

Distinct Clusters Associated With Different Clinical Outcomes

We examined the distribution of clinical characteristics including grade, age, gender, IDH-status, 1p−/19q−, 7+/10− and MGMT-promoter methylation in the four clusters (**Figure 6A**). The percentage of elderly patients and senior grade glioma patients displayed a decreasing trend from C1 to C4 (**Supplementary Figures S7A,B**). There was no significant difference in gender distribution among the four clusters (**Figure 6A**; **Supplementary Figure S7C**). Of note, C4 had the highest percentage of IDH mutation, 1p/19q codeletion and MGMT-promoter methylation (**Supplementary Figures S7D–F**; **Figure 2C**). Univariate and multivariate Cox regression analysis further revealed the prognosis value of these characteristics, and then we identified five independent prognostic factors encompassing the clusters, grade, age, gender and IDH-status (**Supplementary Table S9**). Based on these five factors, we developed a nomogram to assess the 1-year, 2-years, 3-years, and 5-years survival of individual patients (**Figure 6B**). The calibration curve demonstrated good agreement between nomogram-predictions and observations (**Figure 6C**). The Area Under the Curve (AUC) of ROC curve for 1-year, 2-years, 3-years and 5-years were 0.859, 0.910, 0.925, and 0.888, respectively (**Figure 6D**). These results suggested that the nomogram had an excellent performance. In addition, the chemotherapy and immunotherapy sensitivity of each cluster was further evaluated. We first predicted the response of the four clusters to two chemotherapeutic drugs: gemcitabine and bortezomib, which can benefit glioma patients in combination with standard chemotherapeutic drug temozolomide (Bastiancich et al., 2018; Kong et al., 2018). In contrast to the other clusters, C1 and C3 were more sensitive to bortezomib and gemcitabine, respectively (all p-value <0.05) (**Figures 6E,F**). The previous results suggested that C1 and C3 were considered as immune-hot subtypes, while C2 and C4 were considered as immune-cold subtypes. Therefore, we further investigated the sensitivity of each cluster to immunotherapy. The TIDE algorithm was utilized to assess the immunotherapeutic response of patients in each cluster, and it indicated that C1 (41%) and C3 (32%) had a higher response rate relative to C2 (21%) and C4 (14%) (**Figure 6G**). Subclass mapping analysis in Roh cohort (Roh et al., 2017) revealed that C1 displayed high similarity with patients who responded to anti-PD1 therapy (p-value <0.01), and C3 was significantly similar with anti-CTLA4 treatment responders (p-value = 0.04), implying that C1 and C3 were more prospective to respond anti-PD1 and anti-CTLA4 immunotherapy, respectively (**Figure 6H**). This phenomenon precisely corresponded to the high expression level of CD274 (PD-L1) in C1 and CTLA4 in C3 (**Supplementary Figures S7G,H**). In another two cohorts, GSE100797 and GSE78220, C1 was more likely responded to adoptive cell therapy (ACT, p-value <0.01) and anti-PD1 treatment (p-value = 0.03), further demonstrating that C1 was more promising to benefit from immunotherapy (**Figures 6I,J**).

DISCUSSION

Glioma is characterized by high heterogeneity and complex immune escape mechanism, which are increasingly recognized as critical factors that limit the progress of glioma treatment (Reardon and Wen, 2015; Jackson et al., 2019). Specific genomic alterations drive the formation of multidimensional heterogeneity in gliomas (Barthel et al., 2018). Mutational signatures that characterize different mutational processes play a crucial role in the investigation of genomic variation. Our study identified four distinct clusters based on mutational signatures, evidencing the intertumoral molecular variability in glioma. These clusters varied regarding genomic variation, biological characteristics, underlying immune escape mechanisms and clinical characteristics (**Supplementary Table S10**). To the best of our knowledge, the present study is the first to dissect the mutational signatures of glioma, and systematically investigate molecular heterogeneity of glioma from the perspective of genomic variation and immune escape. Meanwhile, we revealed plenty of prognosis relevant genomic driver events. In addition, the nomogram was developed to serve as a robust and promising tool for predicting the prognosis of glioma patients.

Basically, the four clusters enriched in specific mutational processes with different DNA damage mechanisms. C1 was characterized by signature 1 and related to spontaneous deamination of 5-methylcytosine, which was reported to be able to mediate a high incidence of C > T transition in some tumor-suppressor genes and play a role in carcinogenesis of human tumors (Laird and Jaenisch, 1994). Signature 8 was the characteristic of C2, but its proposed etiology was unknown yet. Based on the results of decision tree analysis, we hypothesized that signature 8 was associated with MGMT promoter methylation, which can epigenetically silence the DNA mismatch repair enzyme MGMT (Hegi et al., 2005). In C3, the characteristics were signature 3, associated with homologous recombination deficiency, and signature 13, linked to APOBEC activation. Cytosine deamination of genomic DNA catalyzed by APOBEC family members is a mechanism fueling cancer heterogeneity and evolution (Swanton et al., 2015). In C4, the characteristic was signature 16, associated with alcohol consumption and transcription-coupled damage according to a recent study (Letouzé et al., 2017). DNA damage caused by various mechanisms drives genomic instability and ultimately the cancer process (Lord and Ashworth, 2012). Therefore, the distinct mutational processes may translate into different molecular and clinical features among the four clusters.

Unsurprisingly, the present study detected a significant heterogeneity among the four clusters in genomic variation. A total of 16 significantly mutated genes were identified as drivers involved in the tumorigenesis and evolution of gliomas. Of these,

PTEN mutation, co-occurring with EGFR but repelling with IDH1 mutation, was more specific to C1; FLG mutation, repulsive to IDH mutation, occurred more frequently in C2; ATRX and TP53 mutations were significantly enriched in C3; and IDH1 mutation was more specific to C4. PTEN, EGFR, ATRX, TP53 and IDH mutations were all oncogenic drivers, predominantly related to molecular diagnosis and different prognosis in glioma (Louis et al., 2016; Diplas et al., 2018). FLG mutation, associated with ichthyosis vulgaris and atopic eczema (Akiyama, 2010), was first identified as a biomarker linked to a poor prognosis in glioma. For copy number variation, the clusters also exhibited distinct characteristics, as summarized in **Supplementary Table S10**. The most frequent alterations were located on chromosomes 1p, 7, 10, and 19q, in line with the focus of earlier studies (Louis et al., 2016; Stichel et al., 2018). Copy number variation can lead to oncogene activation or tumor suppressor gene (TSG) inactivation in cancer. In the present study, it was detected that 7q21.2 (CDK6) and 7q31.2 (MET), more specific to C1, were appreciably amplified; 10q23.31 (PTEN), occurred more frequently in C1, and 1p32.3 (CDKN2C), 1p36.23 (ERRFI1), 1p36.32 (AJAP1 and HES3), and 19q13.41 (PPP2R1A), more specific to C4, were appreciably deleted. These oncogene-relevant amplifications and TSG-relevant deletions may contribute to the tumorigenesis and progression of glioma. In addition, the present study also revealed that 12q14.1 amplification, and 6q22.31, 6q26, 13q14.2, 13q22.1, 15q14, 22q13.32, and 4q34.3 deletions were significantly related to prognosis in glioma, implying that these alterations may be able to serve as novel prognostic biomarkers.

The heterogeneity among the four clusters was also reflected in biological function and immune status. As described, C1 was characterized by activation of stroma and immunity and high infiltration of immune and stromal cells, corresponding to stromal and immune dual activation phenotype; C2 was characterized by proliferation promotion, corresponding to proliferation phenotype; C3 was characterized by activation of immunity and high immune cells infiltration, corresponding to immune activation phenotype; and C4 was characterized by metabolism-related pathways, corresponding to metabolism phenotype. Further, we summarized and underlined the potential immune escape mechanisms of each cluster: abundant stromal and immunosuppressive cells infiltration and immune checkpoint blockade in C1; lack of immune cells, low immunogenicity and antigen presentation defect in C2 and C4; and immune checkpoint blockade in C3. The comprehensive understanding of distinct biological characteristics and potential immune escape mechanisms could guide more effective personalized therapy. Moreover, we detected that CNVs and methylation modification were prominent participants in the regulation of immunomodulators, which suggested a novel orientation for the development of glioma immunotherapy.

The present study had important implications for clinical translations and application. Synchronization with heterogeneous molecular features, the four clusters also varied in clinical characteristics, as summarized in **Supplementary Table S10**. First, the novel molecular subtypes had prognostic significance. C4 displayed a better OS in contrast to the other clusters. Consistently, this cluster also exhibited a higher percentage of IDH mutation, 1p/19q codeletion and MGMT-promoter methylation, which were all reported being associated with a favor prognosis in glioma (Hegi et al., 2005; Sabha et al., 2014). Based on the clusters, grade, age, gender and IDH-status, an accurate nomogram was developed to predict the 1-year, 2-years, 3-years, and 5-years survival of individual patients. Second, our results can provide a reference for the selection of suitable patients for chemotherapy or immunotherapy. The present study deciphered that C1 and C3 were more sensitive to bortezomib and gemcitabine, respectively. The combination of the standard chemotherapy drug, temozolomide, and these two drugs can achieve better treatment in gliomas (Bastiancich et al., 2018; Kong et al., 2018). Further, according to the biological function and immune pattern analysis, we observed C1 and C3 belonged to the "immune-hot" subtype, and they also exhibited high expression of immune checkpoint molecules, which were the promising targets of immunotherapy. Thus, we suspected that C1 and C3 may be more sensitive to immunotherapy. Meanwhile, the results of TIDE and subclass mapping analyses evidenced our speculation. C1 was more sensitive to anti-PD1 therapy and ACT, and C3 was more sensitive to anti-CTLA4 therapy. Consistent with this, PD-L1 and CTLA4 were up-regulated in C1 and C3, respectively. In addition, combining the four clusters and clinical features, we developed an excellent nomogram for prognostic evaluation, which could guide more effective clinical management.

The present study also has some limitations. First, candidate genomic carcinogenic drivers and abundance of TME cells require further experimental verification. Second, patients suitable for bortezomib and gemcitabine have been predicted by bioinformatics algorithms, but further clinical validation is also required. Finally, intra-tumor heterogeneity was not considered due to the lack of relevant data.

CONCLUSIONS

The present study revealed four heterogeneous glioma clusters with distinct genomic variants, functional phenotype, immune escape mechanism, and clinical characteristics. The nomogram with excellent performance was developed to serve as a powerful prognostic predictor. These results could enhance the mastery of molecular features and promote the precise therapy and clinical management of glioma.

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

ZL, XH, and LiL designed this work. ZL and TL integrated and analyzed the data. ZL, TL, and LW wrote this manuscript. ZL,

LoL, and LiL edited and revised the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by The National Natural Science Foundation of China (Grant No. 82002433), Henan Province Medical Science and Technology Research Project Joint Construction Project (Grant No. LHGJ20190003), and Science and Technology Project of Henan Provincial Department of Education (Grant No. 21A320036).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.682084/full#supplementary-material

REFERENCES

Akiyama, M. (2010). FLG Mutations in Ichthyosis Vulgaris and Atopic Eczema: Spectrum of Mutations and Population Genetics. *Br. J. Dermatol.* 162 (3), 472–477. doi:10.1111/j.1365-2133.2009.09582.x

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., et al. (2015). Clock-like Mutational Processes in Human Somatic Cells. *Nat. Genet.* 47 (12), 1402–1407. doi:10.1038/ng.3441

Barthel, F. P., Wesseling, P., and Verhaak, R. G. W. (2018). Reconstructing the Molecular Life History of Gliomas. *Acta Neuropathol.* 135 (5), 649–670. doi:10.1007/s00401-018-1842-y

Bastiancich, C., Bastiat, G., and Lagarce, F. (2018). Gemcitabine and Glioblastoma: Challenges and Current Perspectives. *Drug Discov. Today* 23 (2), 416–423. doi:10.1016/j.drudis.2017.10.010

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

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H.-Z., et al. (2017). Landscape of Microsatellite Instability across 39 Cancer Types. *JCO Precision Oncol.* 2017, 1–15. doi:10.1200/PO.17.00073

Carmona-Saez, P., Pascual-Marqui, R. D., Tirado, F., Carazo, J. M., and Pascual-Montano, A. (2006). Biclustering of Gene Expression Data by Non-smooth Non-negative Matrix Factorization. *BMC Bioinformatics* 7, 78. doi:10.1186/1471-2105-7-78

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. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Diplas, B. H., He, X., Brosnan-Cashman, J. A., Liu, H., Chen, L. H., Wang, Z., et al. (2018). The Genomic Landscape of TERT Promoter Wildtype-IDH Wildtype Glioblastoma. *Nat. Commun.* 9 (1), 2087. doi:10.1038/s41467-018-04448-6

Fecci, P. E., and Sampson, J. H. (2019). The Current State of Immunotherapy for Gliomas: an Eye toward the Future. *J. Neurosurg.* 131 (3), 657–666. doi:10.3171/2019.5.JNS181762

Gaujoux, R., and Seoighe, C. (2010). A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinformatics* 11, 367. doi:10.1186/1471-2105-11-367

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

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

Hegi, M. E., Diserens, A.-C., Gorlia, T., Hamou, M.-F., de Tribolet, N., Weller, M., et al. (2005). MGMTGene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* 352 (10), 997–1003. doi:10.1056/NEJMoa043331

Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2007). Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets. *PLoS One* 2 (11), e1195. doi:10.1371/journal.pone.0001195

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

Jackson, C. M., Choi, J., and Lim, M. (2019). Mechanisms of Immunotherapy Resistance: Lessons from Glioblastoma. *Nat. Immunol.* 20 (9), 1100–1109. doi:10.1038/s41590-019-0433-y

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

Kaloshi, G., Benouaich-Amiel, A., Diakite, F., Taillibert, S., Lejeune, J., Laigle-Donadey, F., et al. (2007). Temozolomide for Low-Grade Gliomas: Predictive Impact of 1p/19q Loss on Response and Outcome. *Neurology* 68 (21), 1831–1836. doi:10.1212/01.wnl.0000262034.26310.a2

Kloosterhof, N. K., Bralten, L. B., Dubbink, H. J., French, P. J., and van den Bent, M. J. (2011). Isocitrate Dehydrogenase-1 Mutations: a Fundamentally New Understanding of Diffuse Glioma? *Lancet Oncol.* 12 (1), 83–91. doi:10.1016/S1470-2045(10)70053-X

Kong, X.-T., Nguyen, N. T., Choi, Y. J., Zhang, G., Nguyen, H. N., Filka, E., et al. (2018). Phase 2 Study of Bortezomib Combined with Temozolomide and Regional Radiation Therapy for Upfront Treatment of Patients with Newly Diagnosed Glioblastoma Multiforme: Safety and Efficacy Assessment. *Int. J. Radiat. Oncology*Biology*Physics* 100 (5), 1195–1203. doi:10.1016/j.ijrobp.2018.01.001

Laird, P., and Jaenisch, R. (1994). DNA Methylation and Cancer. *Hum. Mol. Genet.* 3, 1487–1495. Spec No. doi:10.1093/hmg/3.suppl_1.1487

Lauss, M., Donia, M., Harbst, K., Andersen, R., Mitra, S., Rosengren, F., et al. (2017). Mutational and Putative Neoantigen Load Predict Clinical Benefit of Adoptive T Cell Therapy in Melanoma. *Nat. Commun.* 8 (1), 1738. doi:10.1038/s41467-017-01460-0

Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.-F., Tubacher, E., et al. (2017). Mutational Signatures Reveal the Dynamic Interplay of Risk Factors and Cellular Processes during Liver Tumorigenesis. *Nat. Commun.* 8 (1), 1315. doi:10.1038/s41467-017-01358-x

Lord, C. J., and Ashworth, A. (2012). The DNA Damage Response and Cancer Therapy. *Nature* 481 (7381), 287–294. doi:10.1038/nature10760

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Lyon, J. G., Mokarram, N., Saxena, T., Carroll, S. L., and Bellamkonda, R. V. (2017). Engineering Challenges for Brain Tumor Immunotherapy. *Adv. Drug Deliv. Rev.* 114, 19–32. doi:10.1016/j.addr.2017.06.006

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

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 (11), 1747–1756. doi:10.1101/gr.239244.118

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 Facilitates Sensitive and Confident Localization of the Targets of Focal Somatic Copy-Number Alteration in Human Cancers. *Genome Biol.* 12 (4), R41. doi:10.1186/gb-2011-12-4-r41

Payandeh, Z., Yarahmadi, M., Nariman-Saleh-Fam, Z., Tarhriz, V., Islami, M., Aghdam, A. M., et al. (2019). Immune Therapy of Melanoma: Overview of Therapeutic Vaccines. *J. Cel Physiol* 234, 14612–14621. doi:10.1002/jcp.28181

Reardon, D. A., and Wen, P. Y. (2015). Unravelling Tumour Heterogeneity-Implications for Therapy. *Nat. Rev. Clin. Oncol.* 12 (2), 69–70. doi:10.1038/nrclinonc.2014.223

Reifenberger, G., Wirsching, H.-G., Knobbe-Thomsen, C. B., and Weller, M. (2017). Advances in the Molecular Genetics of Gliomas - Implications for Classification and Therapy. *Nat. Rev. Clin. Oncol.* 14 (7), 434–452. doi:10.1038/nrclinonc.2016.204

Roh, W., Chen, P.-L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., et al. (2017). Integrated Molecular Analysis of Tumor Biopsies on Sequential CTLA-4 and PD-1 Blockade Reveals Markers of Response and Resistance. *Sci. Transl. Med.* 9 (379), eaah3560. doi:10.1126/scitranslmed.aah3560

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

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S., and Swanton, C. (2016). DeconstructSigs: Delineating Mutational Processes in Single Tumors Distinguishes DNA Repair Deficiencies and Patterns of Carcinoma Evolution. *Genome Biol.* 17, 31. doi:10.1186/s13059-016-0893-4

Sabha, N., Knobbe, C. B., Maganti, M., Al Omar, S., Bernstein, M., Cairns, R., et al. (2014). Analysis of IDH Mutation, 1p/19q Deletion, and PTEN Loss Delineates Prognosis in Clinical Low-Grade Diffuse Gliomas. *Neuro-Oncology* 16 (7), 914–923. doi:10.1093/neuonc/not299

Sandmann, T., Bourgon, R., Garcia, J., Li, C., Cloughesy, T., Chinot, O. L., et al. (2015). Patients with Proneural Glioblastoma May Derive Overall Survival Benefit from the Addition of Bevacizumab to First-Line Radiotherapy and Temozolomide: Retrospective Analysis of the AVAglio Trial. *Jco* 33 (25), 2735–2744. doi:10.1200/JCO.2015.61.5005

Schalper, K. A., Rodriguez-Ruiz, M. E., Diez-Valle, R., López-Janeiro, A., Porciuncula, A., Idoate, M. A., et al. (2019). Neoadjuvant Nivolumab Modifies the Tumor Immune Microenvironment in Resectable Glioblastoma. *Nat. Med.* 25 (3), 470–476. doi:10.1038/s41591-018-0339-5

Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in Cancer Immunotherapy. *Science* 348 (6230), 69–74. doi:10.1126/science.aaa4971

Stichel, D., Ebrahimi, A., Reuss, D., Schrimpf, D., Ono, T., Shirahata, M., et al. (2018). Distribution of EGFR Amplification, Combined Chromosome 7 Gain and Chromosome 10 Loss, and TERT Promoter Mutation in Brain Tumors and Their Potential for the Reclassification of IDHwt Astrocytoma to Glioblastoma. *Acta Neuropathol.* 136 (5), 793–803. doi:10.1007/s00401-018-1905-0

Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The Cancer Genome. *Nature* 458 (7239), 719–724. doi:10.1038/nature07943

Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J. B., et al. (2005). Radiotherapy Plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* 352 (10), 987–996. doi:10.1056/NEJMoa043330

Swanton, C., McGranahan, N., Starrett, G. J., and Harris, R. S. (2015). APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* 5 (7), 704–712. doi:10.1158/2159-8290.CD-15-0344

Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2013). OncodriveCLUST: Exploiting the Positional Clustering of Somatic Mutations to Identify Cancer Genes. *Bioinformatics* 29 (18), 2238–2244. doi:10.1093/bioinformatics/btt395

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–e14. doi:10.1016/j.immuni.2018.03.023

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. doi:10.7554/eLife.49020

Wildes, T. J., Dyson, K. A., Francis, C., Wummer, B., Yang, C., Yegorov, O., et al. (2020). Immune Escape after Adoptive T-Cell Therapy for Malignant Gliomas. *Clin. Cancer Res.* 26, 5689–5700. doi:10.1158/1078-0432.CCR-20-1065

Yang, Y. (2015). Cancer Immunotherapy: Harnessing the Immune System to Battle Cancer. *J. Clin. Invest.* 125 (9), 3335–3337. doi:10.1172/JCI83871

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.

Copyright © 2021 Liu, Lu, Wang, Liu, Li 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 in Molecular Biosciences

ORIGINAL RESEARCH
published: 08 July 2021
doi: 10.3389/fmolb.2021.684836

Plasma Levels of Heat Shock Protein 90 Alpha Associated With Colorectal Cancer Development

Wene Wei[1†], Jiahui Zhou[1†], Lipeng Chen[2], Haizhou Liu[1], Fuyong Zhang[3], Jilin Li[1], Shufang Ning[1], Shirong Li[1], Chen Wang[1], Yi Huang[1], Chang Zou[2] and Litu Zhang[1]**

[1]Department of Research, Guangxi Medical University Cancer Hospital, Nanning, China, [2]Shenzhen People's Hospital, The Second Clinical Medical College, Jinan University, The First Affiliated Hospital, Southern University of Science and Technology, Shenzhen, China, [3]Department of Laboratory, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China

Reviewed by:
Salva Mena-Mollá,
University of Valencia, Spain
Prihantono Prihantono,
Hasanuddin University, Indonesia
Bo Wei,
Third Affiliated Hospital of Sun Yat-sen University, China

***Correspondence:**
Chang Zou
zou.chang@szhospital.com
Litu Zhang
zhanglitu@gmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 24 March 2021
Accepted: 25 June 2021
Published: 08 July 2021

Citation:
Wei W, Zhou J, Chen L, Liu H, Zhang F, Li J, Ning S, Li S, Wang C, Huang Y, Zou C and Zhang L (2021) Plasma Levels of Heat Shock Protein 90 Alpha Associated With Colorectal Cancer Development. Front. Mol. Biosci. 8:684836. doi: 10.3389/fmolb.2021.684836

Aim: The role of plasma heat shock protein 90 alpha (HSP90α) in colorectal cancer patients remains unclear. This study aimed to evaluate the relationship between HSP90α and the occurrence and development of colorectal cancer through diagnosis and prognosis value.

Methods: 635 colorectal cancer patients and 295 healthy controls were recruited. The HSP90α was measured by using the ELISA kit in all objects and the immune cells and common biomarkers as CEA, AFP, CA125, CA153 and CA199 were measured in all colorectal cancer patients. The relationship between plasma HSP90α with clinical features, common tumor markers and immune cells were also conducted. The survival analysis endpoint was progression-free survival (PFS).

Results: The levels of plasma HSP90α were significantly higher in colorectal cancer patients compared to healthy controls [51.4 (ng/ml) vs. 43.7 (ng/ml), $p < 0.001$]. In additional, the levels of plasma HSP90α were associated with gender and disease progress as stage, lymphatic and distant metastasis. Furthermore, plasma HSP90α was closed correlation with CEA, CA125, CA199 and percentage of B cells. However, the initial expression level of plasma HSP90α failed to show a prognostic value for progression-free survival in colorectal cancer.

Conclusion: The plasma Hsp90α was remarkable higher in colorectal cancer and correlated with common tumor biomarkers and immune cells. Plasma Hsp90α levels were associated with disease progress but a poor diagnosis performance and also failed to show a prognostic value in colorectal cancer.

Keywords: heat shock protein 90 alpha, colorectal cancer, metastasis, diagnosis, prognosis

INTRODUCTION

Colorectal cancer (CRC) is one of the most common tumors worldwide. Approximately 1.8 million people were diagnosed with colorectal cancer and 861,000 died from colorectal cancer in 2018, which equates to 4,931 people diagnosed with colorectal cancer and 2,358 people dying from colorectal cancer every day (Bray et al., 2018). Metastasis and recurrence have always been difficult points in cancer treatment. The liver is the most common site of distant metastases and accounts for 20% of

CRC patients who arrive in hospitals with synchronous disease, and approximately 50% of CRC patients develop liver metastases at some point in the course of their disease (Leonard et al., 2005; Manfredi et al., 2006). With the development of medical technology, the treatments for colorectal cancer patients include endoscopic and surgical local excision, downstaging preoperative radiotherapy and systemic therapy, extensive surgery for locoregional and metastatic disease, local ablative therapies for metastases, and palliative chemotherapy, targeted therapy, and immunotherapy (Hurwitz et al., 2004; André et al., 2009; Van Cutsem et al., 2009; Haller et al., 2011; Ahlenstiel et al., 2014; Baratti et al., 2016; Ma et al., 2017; Abdel-Rahman and Cheung, 2018). Distant metastatic CRC (mCRC) patients have historically been associated with poor survival. Without treatment, the median survival time of mCRC patients is < 12 months, and the 5 years survival rate is <10% (Mohammad and Balaa, 2009; Morris and Treasure, 2018). Undergoing bevacizumab in combination with oxaliplatin or capecitabine chemotherapy therapy, patients with colorectal cancer have a significantly improved prognosis, with a median progression-free survival (PFS) of approximately 9 months (Saltz et al., 2008; Cunningham et al., 2013). Patients with metastatic colorectal cancer have a good survival benefit from prompt and aggressive treatment. Therefore, timely detection of metastasis is particularly important for colorectal cancer patients.

Traditionally, computed tomography (CT) scan is the common modality for initial diagnosis of distant metastasis and the modality of choice in order to discover those patients for evidence of a recurrence of the disease (Benson et al., 2013). However, preoperative chemotherapy or neo-adjuvant therapy would generally decrease the sensitivity of CT scans as a method for monitoring metastasis in CRC patients (van Kessel et al., 2012; Rojas Llimpe et al., 2014). Therefore, an optimal strategy for early detection and diagnosis distant metastasis for CRC patients is essential. Meanwhile, an appropriate way to assess prognosis is also urgently needed.

Due to the accessibility and noninvasiveness advantages, peripheral blood samples are commonly used for biomarker determinations and are widely accepted as an acceptable intervention by patients and health workers. In addition to commonly used tumor markers such as CEA, AFP, CA125, CA153 and CA199, heat shock protein 90 alpha (HSP90α) is a broad-spectrum tumor marker worthy of attention in recent years. HSP90α is an intracellular molecular chaperone which is exposed to the extracellular space. It has been documented that the overexpression of HSP90α was associated with tumor development, progression, invasiveness, metastatic potentials and chemo-resistance in various types of cancers (Passarino et al., 2003; Eustace et al., 2004; Tsutsumi et al., 2008; Chang et al., 2009; Zuehlke et al., 2015). Previous studies showed that HSP90α levels with or without AFP can act as excellent diagnostic markers for liver cancer (Fu et al., 2017; Wei et al., 2020). In addition, HSP90α also showed a good diagnostic performance for lung cancer and early CRC patients (Shi et al., 2014; Kasanga et al., 2018). In China, HSP90α was approved for clinical application as a broad-spectrum tumor marker in 2016. However, the diagnostic and prognostic efficacies of plasma HSP90α in patients with colorectal cancer has not been thoroughly confirmed. In the current study, we will assess the diagnostic and prognostic value of HSP90α for CRC patients.

MATERIAL AND METHODS

Patients

635 CRC patients were recruited from the Department of Gastrointestinal Surgery, Guangxi Medical University Cancer Hospital from 1, Jan, 2018 to 31, Aug, 2019 with a median age 60 years. The inclusion criteria for all patients that there was: 1) confirmation of CRC by clinical manifestation and histopathological examination associated with imaging diagnosis; 2) the state of distant metastases was assessed; 3) no anti-tumor treatment or surgical resection were performed at the time of diagnosis; 4) the tumor markers, immune cells and HSP90α in peripheral blood were evaluated before any treatment; 5) complete information for clinical features were available. The exclusion criteria for all patients were if they had: 1) CRC combined with other cancers; 2) a history of malignant tumors in other organs; 3) radiation or chemotherapy prior to admission; 4) distant metastases which could not be evaluated. The control group included 170 patients from The First Affiliated Hospital of Guangxi Medical University who had received health examinations during the same period with a median age 37.5 years. Stages for patients with colorectal cancer were classified according to the American Joint Committee on Cancer Classification (the 7th edition). Patients were followed up for tumor assessments every 12 weeks. Disease progression assessed by using RECIST 1.1 (Eisenhauer et al., 2009).

Methods

Detection of peripheral blood markers was performed as follows. Fasting elbow venous blood was collected in EDTA anticoagulant tubes to prepare plasma samples and in dried tubes to prepare serum samples. All samples were separated by centrifugation at 3,000 rpm for 10 min. The levels of plasma HSP90α were measured by using the ELISA kit for HSP90α protein (Yantai Protgen Biotechnology Development Co., Ltd., Yantai, China). All operations followed the manufacturer's instructions. The kit was pre-incubated at 37°C for 30 min and the plasma samples were diluted 20 times with diluent solution provided in the kit. The standards were loaded together with the quality controls and the prepared samples (50 μL of each) added into 96-well plates followed by addition of 50 μL of anti–Hsp90aHRP-conjugated antibody. Then, the plate was incubated at 37°C for 1 h after gentle shaking. Next, the color reaction step was carried out after six washes. 50 μL of peroxide and 50 μL of 3, 3′, 5, 5′ -tetramethylbenzidine solution was added and the mixture was incubated at 37°C for a further 20 min followed by termination of the reaction with an acid stop buffer. Finally, the optical density was measured using a spectrophotometer at 450 nm for the detection wavelength, with 620 nm as the reference wavelength. The concentration of HSP90α protein in each sample was calculated according to a standard curve of optical density values. Meanwhile, the expression status of HSP90α in colon adenocarcinoma (COAD) was also evaluated by using data retrieved from The Cancer Genome Atlas (TCGA).

The levels of serum CEA, AFP, CA125, CA153 and CA19-9 were measured by chemiluminescence microparticle immuno assays (CMIA) using the Architect i2000SR analyzer and the corresponding reagent kits which were purchased from Architect Diagnostics, America. All operations followed the manufacturer's instructions. The immune cells in this study including T cells, helper T lymphocytes (Th cells), suppressor T lymphocytes (Ts cells), Natural killer cells (NK cells) and B cells were defined as CD45+/CD3+ lymphocytes, CD45+/CD3+/CD4+ lymphocytes, CD45+/CD3+/CD8+ lymphocytes, CD45+/CD3-/CD16+/CD56 + lymphocytes and CD45+/cCD19 + lymphocytes, respectively. Flow cytometry (BD Biosciences, Franklin Lakes, NJ, United States) was performed to detect labeled cells and analyze the results.

Statistical Analysis

Levels of the HSP90α, CEA, AFP, CA125, CA153 and CA19-9 were assessed by SPSS STATISTICS 23.0 (IBM, Chicago, IL, United States) to ascertain normal distribution and the data are presented as medians and ranges. Nonparametric Kruskal-Wallis H-test was used to compare the differences between groups. Wilcoxon rank sum test was used to analyze HSP90α expression in COAD samples from The Cancer Genome Atlas (TCGA). Correlation between the indicators was analyzed using Pearson analysis. The diagnostic value was analyzed using receiver operating characteristic (ROC) curves, with the area under the curve (AUC). The optimal cut-off values for ROC curves were established using the Youden Index (YI = sensitivity + specificity −1). Progression-free survival (PFS) was the primary study end point, and was defined as the time from initial diagnosis to the date of disease recurrence and was censored at the last follow-up or at the time of death from any cause. The survival curves were generated using the Kaplan-Meier curve and survival differences were estimated by a log-rank test. $p < 0.05$ was considered statistically significant.

RESULTS

Basic Profile of Tumor Markers and Immune Cells in Patients With Colorectal Cancer

This study recruited 635 cases of colorectal cancer patients and 295 cases of healthy controls. Routine tumor markers and immune cells were assessed in all patients and the results were showed in **Table 1**. The plasma levels of HSP90α protein were significantly higher in patients with colorectal cancer than healthy controls [51.4 (33.8, 80.3) ng/ml vs. 43.7 (34.3, 54.8) ng/ml; $p < 0.001$, **Figure 1**]. There was no significant difference in plasma HSP90α protein levels among healthy controls with different age distribution (all $p > 0.05$, **Figure 1**).

Correlations Between the Plasma Levels of HSP90α Protein and Clinic-Pathological Characteristics in Colorectal Cancer Patients

The associations of plasma levels of HSP90α protein and clinic-pathological characteristics in colorectal cancer patients were showed in **Table 2**. Male patients with colorectal cancer were showed a significantly higher plasma level of HSP90α protein ($p = 0.002$). Meanwhile, the plasma levels of HSP90α protein were correlated with disease stage, lymphatic metastasis, and distant metastasis (all $p < 0.001$). There was no difference in plasma HSP90α protein expression among different pathological differentiation levels ($p = 0.153$).

Correlations Between the Plasma Levels of HSP90α Protein With Routine Tumor Markers and Immune Cells

The correlations between tumor markers and immune cells percentage were showed in **Figure 2**. There was no statistical correlation between common tumor markers CEA, AFP, CA125, CA153, CA199, TK1 and immune cells in patients with colorectal cancer (all $p > 0.05$, **Figure 2**). Plasma HSP90α protein levels were positively related to CEA, CA125, CA199 and negatively related to B lymphocyte percentage (all $p < 0.05$, **Figure 2**).

Univariate and Multivariate Analysis for Factors Associated With Metastasis in Colorectal Cancer

The univariate and multivariate analysis for factors associated with metastasis in colorectal cancer were showed in **Table 3**. The univariate analysis showed that the biomarkers as CA125, CEA, CA199 and HSP90α were associated with metastasis in colorectal cancer patients, and the multivariate analysis revealed that the biomarkers as CEA, CA199 and HSP90α were independent risk factor for distant metastasis in patients with colorectal cancer (all $p < 0.001$). A nomogram for predicting the presence of metastasis in patients with colorectal cancer was showed in **Figure 3**.

TABLE 1 | Basic profile of tumor markers and immune cells in patients with colorectal cancer.

Parameters	Colorectal cancer	Healthy control	*p* value
CEA (ng/ml)	5.3 (2.4, 17.8)	0.9 (0.6, 1.4)	*<0.001*
AFP (ng/ml)	2.5 (1.9, 3.4)	2.1 (1.6, 2.8)	*<0.001*
CA125 (U/ml)	11.1 (7.7, 18.7)	11.7 (8.3, 17.2)	*0.952*
CA153 (U/ml)	9.3 (6.2, 13.3)	16.0 (13.0, 22.7)	*<0.001*
CA199 (U/ml)	9.1 (3.5, 41.4)	12.9 (7.4, 17.9)	*0.090*
HSP90α (ng/ml)	51.4 (33.8, 80.3)	43.7 (34.3, 54.8)	*<0.001*
TK1 (ng/ml)	0.8 (0.4, 1.3)	—	—
T cells (%)	66.1 (58.7, 72.13)	—	—
Th cells (%)	39.7 (33.9, 45.5)	—	—
Ts cells (%)	19.3 (15.2, 24.2)	—	—
Ratio	2.0 (1.6, 2.7)	—	—
NK (%)	13.1 (8.7, 18.8)	—	—
B cells (%)	12.0 (9.3, 15.7)	—	—

Ratio: the ratio of the Th cells to Ts cells. All values are presented in median and quartile intervals. The p values were shown in italics.

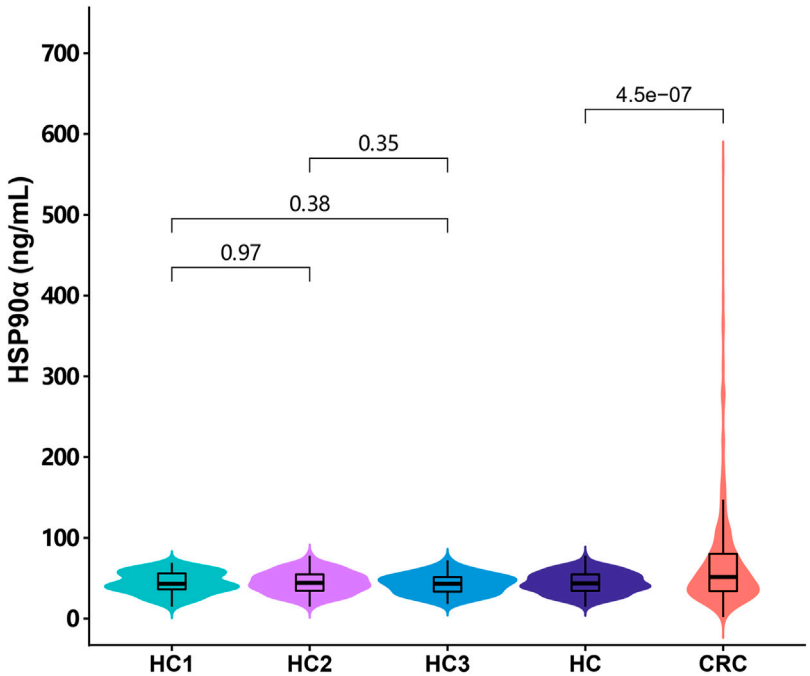


FIGURE 1 | The distribution of the plasma HSP90α protein expression in colorectal cancer patients and healthy controls. HC1: subgroup under 40 years of age in healthy control group; HC2: subgroup of healthy controls aged 40–59 years; HC3: subgroup over 60 years of age in healthy control group; HC: all healthy controls, CRC: colorectal cancer.

TABLE 2 | Correlations of the plasma HSP90α with clinicopathologic features in colorectal cancer.

Parameters	Frequency	HSP90α (ng/ml)	*p* value
Age (<60 years, ≥60 years)	315/320	51.4 (33.8, 80.3)/48.2 (32.5, 72.6)	*0.073*
Gender (male, female)	414/221	54.5 (35.9, 84.7)/43.7 (30.2, 71.7)	***0.002***
Differentiation (well, moderately, poor)	25/545/65	48.1 (31.0, 72.3)/50.5 (33.3, 78.3)/58.0 (38.5, 90.8)	*0.153*
Stage (I-II, III-IV)	201/434	43.5 (30.1, 63.0)/55.6 (35.4, 88.1)	***<0.001***
Lymphatic metastasis (absent/present)	233/402	43.3 (30.7, 62.5)/57.1 (35.7, 92.8)	***<0.001***
Metastasis (absent/present)	470/165	45.3 (30.9, 70.6)/72.6 (43.3, 131.0)	***<0.001***

All values are presented in median and quartile intervals. The results of $p < 0.05$ *were highlighted in bold italics.*

The Diagnostic Efficacy of Plasma HSP90α in Colorectal Cancer and its Ability to Distinguish Colorectal Cancer Patients With Distant Metastasis

ROC analysis was conducted to assess the diagnostic ability of plasma HSP90α for colorectal cancer and results were showed in **Figure 4**. Using serum CEA, CA199 and plasma HSP90α levels of healthy donors as control group, the serum CEA has significant advantages in colorectal cancer diagnosis with a cut-off value 1.77 ng/ml (AUC = 0.944, sensitivity 86.3%, specificity 87.12%, **Figure 3A**). However, the plasma HSP90α levels showed a poor performance in colorectal cancer diagnosis with a cut-off value 69.1 ng/ml (AUC = 0.602, sensitivity 33.1%, specificity 99.0%, **Figure 3A**). Patients with colorectal cancer were divided into M0 and M1 groups according to the absence or presence of distant metastasis. Using the M0 group as controls, the AUCs of CEA, CA19-9, HSP90α and the panel for distinguishing the presence of distant metastasis from CRC patients were 0.776, 0.755, 0.690, 0.819, respectively (**Figure 3B**). A nomogram for predicting the presence of metastasis in patients with colorectal cancer was showed in **Figure 4**, higher total score based on the sum of the assigned numbers for each of the factors in the nomogram was associated with a high risk of mCRC.

Correlations Between Concentrations of HSP90α, CEA and CA19-9 and PFS in CRC Patients

Of 635 colorectal cancer patients, 454 cases had complete follow-up data and received standardized treatment. The median follow-up time was 7 months (range from 1 to 32 months). Survival analysis was conducted based on the median of CEA, CA19-9 and HSP90α levels in all colorectal cancer patients and the cutoff values calculated according

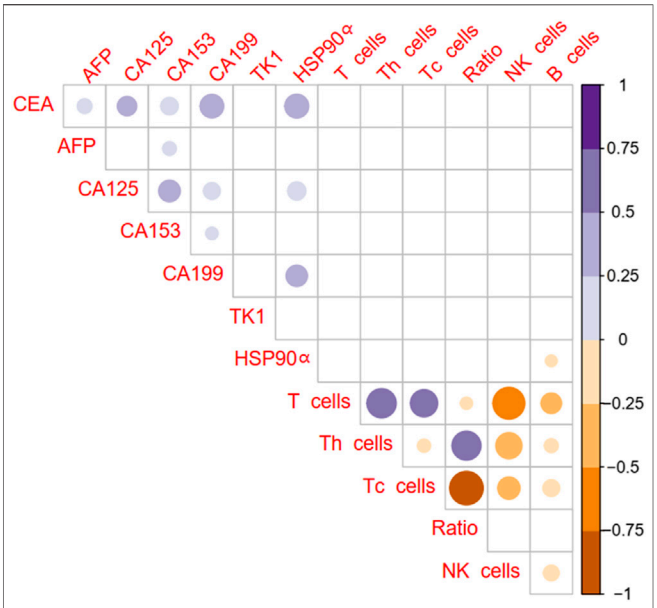


FIGURE 2 | Correlation of plasma HSP90α protein levels with routine tumor markers and immune cells in colorectal cancer patients (Pearson analysis). Blue color represents a positive correlation, brown color represents a negative correlation, and the shade of color represents the degree of correlation. The results with $p < 0.05$ were filled with color and display in the diagram.

to Yuden index respectively, and the results are shown in **Figure 5**. The initial concentrations of CEA, CA19-9 and HSP90α were not correlated with PFS in colorectal cancer (all $p > 0.05$, **Figure 5**).

HSP90α Expression in TCGA Colorectal Cancer

To further explore the expression status of HSP90α in colon adenocarcinoma (COAD), we analyzed the expression data of HSP90α in COAD samples using data retrieved from The Cancer Genome Atlas (TCGA), and found that the expression levels of HSP90α in tumor samples were significantly higher than these from normal samples ($p < 0.001$), as shown in **Figure 6A**. The expression levels of HSP90α were further compared in samples from different stages of tumor, but no significant difference was found, although the expression levels of HSP90α in samples from all stages were higher than these from normal samples, as shown in **Figure 6B**.

DISCUSSION

HSP90α, which has evolved for almost 3.5 billion years, is a member of the HSP90 family, a conserved and essential molecular chaperone, can be translocated to the cell surface and secreted into the extracellular space by cancer cells (Frydman, 2001; Eustace et al., 2004). The secreted HSP90α was associated with tumor development and prognosis, especially with cancer metastasis (Passarino et al., 2003; Eustace et al., 2004; Chang et al., 2009). Previously studies have showed that plasma HSP90α is an excellent biomarker for the diagnosis of lung cancer and liver cancer (Jhaveri et al., 2014; Shi et al., 2014; Fu et al., 2017). Meanwhile, plasma HSP90α levels were significantly higher than healthy controls in other cancers, but the diagnostic efficiency was insufficient, such as gastric cancer, breast cancer, nasopharyngeal carcinoma and colorectal cancer (Kasanga et al., 2018; Liu et al., 2019; Lin et al., 2020; Zhang et al., 2020). In current study, the plasma HSP90α levels was significantly higher in CRC patients than in healthy controls and consistent with previously reported results (Kasanga et al., 2018). Additionally, this current study finds that plasma HSP90α levels were associated with gender and disease progression as stage, lymphatic metastasis and distant metastasis in colorectal cancer patients. Plasma HSP90α levels were also observed in breast, liver and nasopharyngeal cancers to be closely correlated with disease stage and distant metastasis, but not with gender (Fu et al., 2017; Wei et al., 2020; Mao et al., 2021). Therefore, we speculate that the gender differences in HSP90α levels may be influenced by hormones. Multivariate statistical analysis also showed that the serum CEA, CA199 and plasma HSP90α are independent risk factors for distant metastasis,

TABLE 3 | Univariate and multivariate analysis for factors associated with metastasis in colorectal cancer.

	Univariate analysis OR (95%CI)	*p* value	Multivariate analysis OR (95%CI)	*p* value
Gender	0.789 (0.540,1.154)	*0.223*	—	—
Age	0.902 (0.633, 1.286)	*0.569*	—	—
T cells	0.988 (0.970, 1.005)	*0.174*	—	—
Th cells	0.994 (0.973, 1.015)	*0.564*	—	—
Ts cells	0.988 (0.962, 1.016)	*0.399*	—	—
Ratio	1.011 (0.842, 1.213)	*0.909*	—	—
NK	1.002 (0.980, 1.024)	*0.872*	—	—
B cells	0.999 (0.965, 1.034)	*0.940*	—	—
CA153	1.024 (0.999, 1.050)	*0.063*	—	—
AFP	1.063 (0.991, 1.141)	*0.089*	—	—
TK1	0.997 (0.876, 1.135)	*0.964*	—	—
CA125	1.010 (1.005, 1.016)	*<0.001*	1.695 (1.000, 1.008)	*0.090*
CEA	1.008 (1.005, 1.011)	*<0.001*	3.647 (1.003, 1.008)	*<0.001*
CA199	1.003 (1.002, 1.004)	*<0.001*	5.454 (1.001, 1.003)	*<0.001*
HSP90α	1.013 (1.010, 1.017)	*<0.001*	3.935 (1.004, 1.012)	*<0.001*

The p *values were shown in italics.*

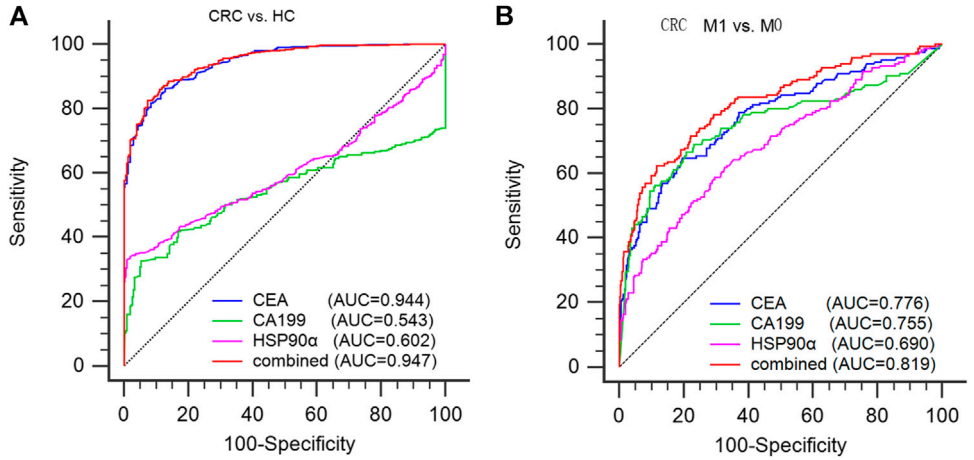


FIGURE 3 | diagnostic ability of serum CEA, CA199 and plasma HSP90α in colorectal cancer. **(A)** Using healthy donors as controls to evaluate the efficacy of CEA, CA199, HSP90α and the panel in the diagnosis of colorectal cancer; **(B)** Using M0 patients as controls to evaluate the efficacy of CEA, CA199, HSP90α and the panel for distinguishing the presence of distant metastasis from CRC patients. M0: no distant metastasis; M1: distant metastasis.

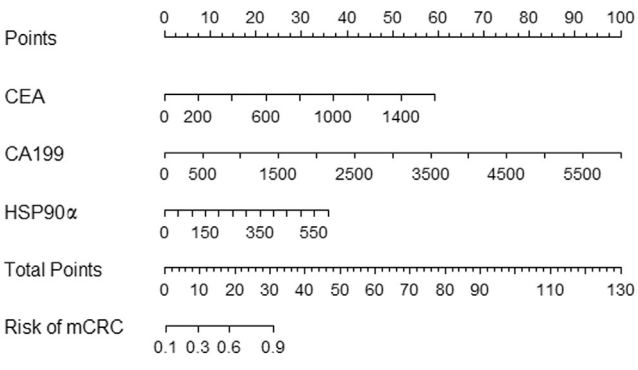


FIGURE 4 | Nomogram for predicting the presence of metastasis in patients with colorectal cancer.

respectively. The Nomogram prediction of distant metastases in combination with CEA, CA199, and HSP90α can provide evidence for clinical treatment plans that may lead to better patient outcomes.

In correlation analysis, the common tumor makers CEA, AFP, CA125, CA153, CA199 showed no statistical correlation with immune cells. But it is worth noting that the plasma HSP90α levels was highly correlated with common tumor biomarkers as CEA, CA125, CA199 and immune cells as B lymphocyte that indicates that the expression of plasma HSP90α is closely related to the regulation of immune function. However, the mechanism of the association between plasma HSP90α and immune cells in the occurrence and progression of cancer is still unknown. It is well known that cancer is also an immune-related disease. At present, the detection of immune cells is one of the commonly used methods to monitor the body immunity in clinical practice. Previous studies have reported that NK cells percentage and B lymphocyte are associated with survival in CRC patients (Berntsson et al., 2016; Tang et al., 2020). However, the plasma HSP90α levels were failed to find the significant prognosis value in colorectal cancer patients in this study. It's worth noting that the endpoint of most of the previous studies was overall survival (OS), while the endpoint of this study was PFS. Therefore, we speculated that although plasma HSP90α levels is not associated with PFS, it may be associated with OS in colorectal cancer patients and further research is needed to confirm this hypothesis.

Plasma HSP90α is overexpress and increases multi-malignant phenotypes including chemoresistance to cisplatin as well as metastatic potentials in various types of cancers (Eustace et al., 2004; Chang et al., 2009; Zuehlke et al., 2015). Therefore, targeted treatment with HSP90α inhibitors offer interesting perspectives for the treatment of cancers. Previous studies have showed that cellular secretion of HSP90α from colorectal cancer cells was enhanced after serum starvation, and secreted HSP90α could be used to stimulate migration and invasion of other non-serum-starved cells, and the mechanism might be that secreted HSP90α acts through TCF12 expression to enhance CRC cell spreading (Chen et al., 2010; Chen et al., 2013). Currently, approximately 20 of these inhibitors have undergone clinical evaluation (Jhaveri et al., 2012; Jhaveri et al., 2014). A recent study reported that HSP90α inhibition sensitized immune-refractory tumor to adoptive T cell transfer as well as PD-1 blockade, and re-invigorated the immune cycle of tumor-reactive T cells (Song et al., 2020). From the results of this study, we know that the expression of plasma HSP90α is related to immune cells. Therefore, the use of HSP90α inhibitors alone or in combination with chemoradiotherapy or immunotherapy may be one of the treatment methods to improve the prognosis of metastatic colorectal cancer. At the same time, dynamic observation of changes of in plasma HSP90α levels may be one of the monitoring methods for the therapeutic efficacy of HSP90α inhibitors.

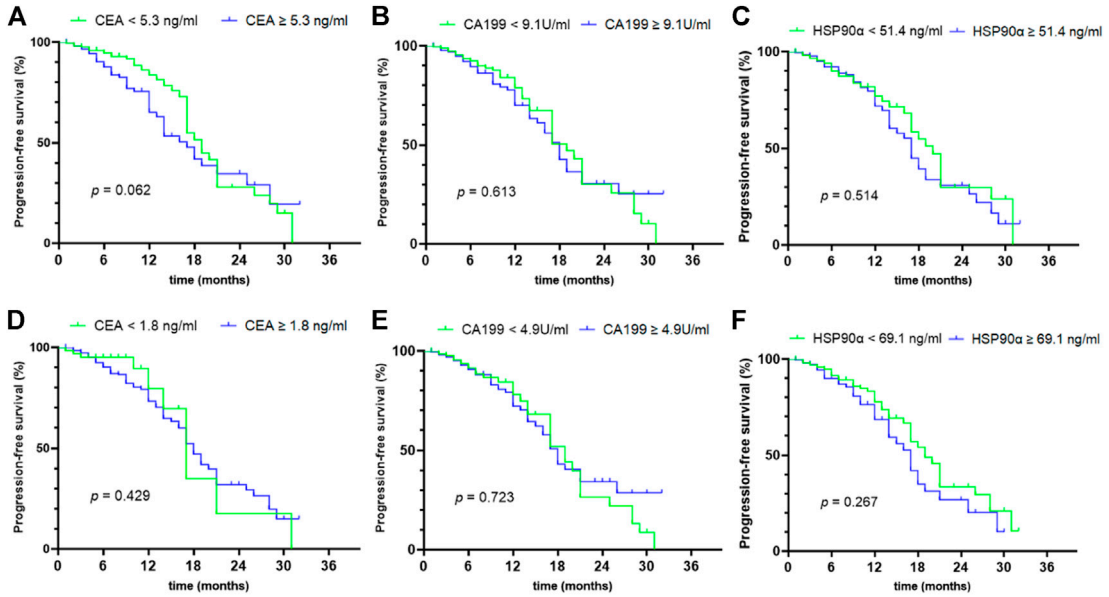


FIGURE 5 | prognostic value of initial concentrations of CEA, CA199 and HSP90α in colorectal cancer. **(A–C)** The prognostic values of initial concentrations of CEA, CA199 and HSP90α were analyzed by grouping based on median value; **(D–F)** The prognostic values of initial concentrations of CEA, CA199 and HSP90α were analyzed by grouping based on the cut-off value of ROC analysis in **Figure 3A**.

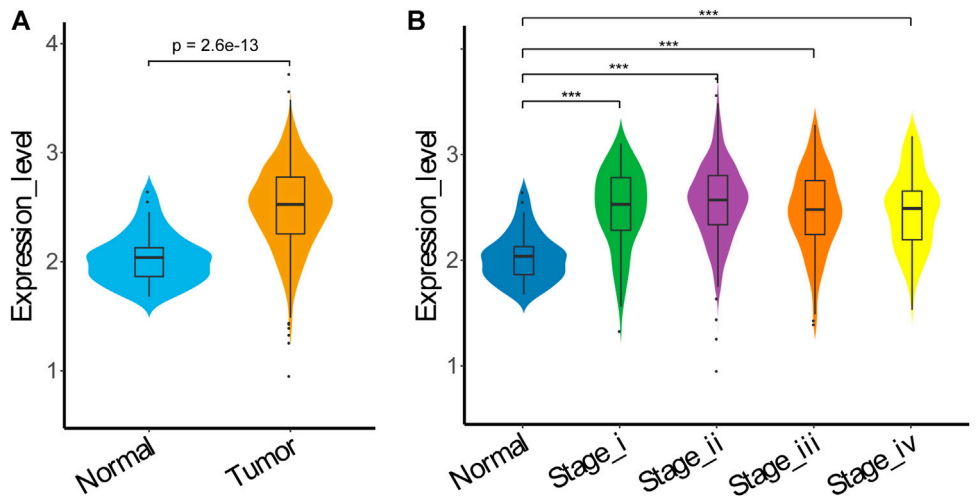


FIGURE 6 | HSP90α expression in TCGA COAD. **(A)** HSP90α expression in TCGA COAD samples from normal tissue (*n* = 40) or tumor tissue (*n* = 458); **(B)** HSP90α expression in TCGA COAD samples from normal tissue (*n* = 40) or different tumor stage tissue (stageI: *n* = 78, stageII: *n* = 183, stage III: *n* = 132, stage IV: *n* = 65). Data were analyzed using Wilcoxon Rank Sum test.

CONCLUSION

The present study demonstrated that plasma Hsp90α was remarkable higher in colorectal cancer and correlated with common tumor biomarkers and immune cells. Plasma Hsp90α levels were associated with gender, stage, lymphatic metastasis and distant metastasis but a poor diagnosis performance in colorectal cancer. Meanwhile, the current study failed to show a prognostic value for plasma HSP90α in colorectal cancer.

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 Local Ethics Committee of Guangxi Medical University Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WW, LZ, CZ, and HL conceived and designed the experiments; JZ and HL collected the patients' samples; WW, JZ, JL, SN, SL, and CW performed the experiments; WW, LC, and YH analyzed the data; WW wrote the paper; LZ, HL, and CZ revised the paper; All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Science Foundation of Guangxi Province of China (Grant No. 2017GXNSFAA198155); Guangxi Scientific Research and Technical Planning Project (Gui ke AB19110018); and the Scientific Research and Technical Development Project of Qingxiu District, Nanning City, Guangxi Province (Nos. 2017034 and 2016051).

ACKNOWLEDGMENTS

The authors would like to thank Dev Sooranna, Imperial College London, for editing the manuscript.

REFERENCES

Abdel-Rahman, O., and Cheung, W. Y. (2018). Integrating Systemic Therapies into the Multimodality Treatment of Resectable Colorectal Liver Metastases. *Gastroent Res. Pract.* 2018, 4326082. doi:10.1155/2018/4326082

Ahlenstiel, G., Hourigan, L. F., Brown, G., Zanati, S., Williams, S. J., Singh, R., et al. (2014). Actual Endoscopic versus Predicted Surgical Mortality for Treatment of Advanced Mucosal Neoplasia of the colon. *Gastrointest. Endosc.* 80 (4), 668–676. doi:10.1016/j.gie.2014.04.015

André, T., Boni, C., Navarro, M., Tabernero, J., Hickish, T., Topham, C., et al. (2009). Improved Overall Survival with Oxaliplatin, Fluorouracil, and Leucovorin as Adjuvant Treatment in Stage II or III Colon Cancer in the MOSAIC Trial. *Jco* 27 (19), 3109–3116. doi:10.1200/jco.2008.20.6771

Baratti, D., Kusamura, S., Pietrantonio, F., Guaglio, M., Niger, M., and Deraco, M. (2016). Progress in Treatments for Colorectal Cancer Peritoneal Metastases during the Years 2010-2015. A Systematic Review. *Crit. Rev. Oncol. Hematol.* 100, 209–222. doi:10.1016/j.critrevonc.2016.01.017

Benson, A. B., 3rd, Bekaii-Saab, T., Chan, E., Chen, Y.-J., Choti, M. A., Cooper, H. S., et al. (2013). Metastatic Colon Cancer, Version 3.2013. *J. Natl. Compr. Canc Netw.* 11 (2), 141–152. doi:10.6004/jnccn.2013.0022

Berntsson, J., Nodin, B., Eberhard, J., Micke, P., and Jirström, K. (2016). Prognostic Impact of Tumour-Infiltrating B Cells and Plasma Cells in Colorectal Cancer. *Int. J. Cancer* 139 (5), 1129–1139. doi:10.1002/ijc.30138

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, W., Ma, L., Lin, L., Gu, L., Liu, X., Cai, H., et al. (2009). Identification of Novel Hub Genes Associated with Liver Metastasis of Gastric Cancer. *Int. J. Cancer* 125 (12), 2844–2853. doi:10.1002/ijc.24699

Chen, J.-S., Hsu, Y.-M., Chen, C.-C., Chen, L.-L., Lee, C.-C., and Huang, T.-S. (2010). Secreted Heat Shock Protein 90α Induces Colorectal Cancer Cell Invasion through CD91/LRP-1 and NF-Kb-Mediated Integrin αV Expression. *J. Biol. Chem.* 285 (33), 25458–25466. doi:10.1074/jbc.m110.139345

Chen, W.-S., Chen, C.-C., Chen, L.-L., Lee, C.-C., and Huang, T.-S. (2013). Secreted Heat Shock Protein 90α (HSP90α) Induces Nuclear Factor-Kb-Mediated TCF12 Protein Expression to Down-Regulate E-Cadherin and to Enhance Colorectal Cancer Cell Migration and Invasion. *J. Biol. Chem.* 288 (13), 9001–9010. doi:10.1074/jbc.m112.437897

Cunningham, D., Lang, I., Marcuello, E., Lorusso, V., Ocvirk, J., Shin, D. B., et al. (2013). Bevacizumab Plus Capecitabine versus Capecitabine Alone in Elderly Patients with Previously Untreated Metastatic Colorectal Cancer (AVEX): an Open-Label, Randomised Phase 3 Trial. *Lancet Oncol.* 14 (11), 1077–1085. doi:10.1016/s1470-2045(13)70154-2

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

Eustace, B. K., Sakurai, T., Stewart, J. K., Yimlamai, D., Unger, C., Zehetmeier, C., et al. (2004). Functional Proteomic Screens Reveal an Essential Extracellular Role for Hsp90α in Cancer Cell Invasiveness. *Nat. Cel Biol* 6 (6), 507–514. doi:10.1038/ncb1131

Frydman, J. (2001). Folding of Newly Translated Proteins *In Vivo*: The Role of Molecular Chaperones. *Annu. Rev. Biochem.* 70, 603–647. doi:10.1146/annurev.biochem.70.1.603

Fu, Y., Xu, X., Huang, D., Cui, D., Liu, L., Liu, J., et al. (2017). Plasma Heat Shock Protein 90alpha as a Biomarker for the Diagnosis of Liver Cancer: An Official, Large-Scale, and Multicenter Clinical Trial. *Ebiomedicine* 24, 56–63. doi:10.1016/j.ebiom.2017.09.007

Haller, D. G., Tabernero, J., Maroun, J., de Braud, F., Price, T., Van Cutsem, E., et al. (2011). Capecitabine Plus Oxaliplatin Compared with Fluorouracil and Folinic Acid as Adjuvant Therapy for Stage III Colon Cancer. *Jco* 29 (11), 1465–1471. doi:10.1200/jco.2010.33.6297

Hurwitz, H., Fehrenbacher, L., Novotny, W., Cartwright, T., Hainsworth, J., Heim, W., et al. (2004). Bevacizumab Plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. *N. Engl. J. Med.* 350 (23), 2335–2342. doi:10.1056/nejmoa032691

Jhaveri, K., Taldone, T., Modi, S., and Chiosis, G. (2012). Advances in the Clinical Development of Heat Shock Protein 90 (Hsp90) Inhibitors in Cancers. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1823 (3), 742–755. doi:10.1016/j.bbamcr.2011.10.008

Jhaveri, K., Ochiana, S. O., Dunphy, M. P., Gerecitano, J. F., Corben, A. D., Peter, R. I., et al. (2014). Heat Shock Protein 90 Inhibitors in the Treatment of Cancer: Current Status and Future Directions. *Expert Opin. Investig. Drugs* 23 (5), 611–628. doi:10.1517/13543784.2014.902442

Kasanga, M., Liu, L., Xue, L., and Song, X. (2018). Serum Heat Shock Protein 90α Have an Advantage in Diagnosis of Colorectal Cancer at Early Stage. *Biomarkers Med.* 12 (8), 881–890. doi:10.2217/bmm-2018-0155

Leonard, G. D., Brenner, B., and Kemeny, N. E. (2005). Neoadjuvant Chemotherapy before Liver Resection for Patients with Unresectable Liver Metastases from Colorectal Carcinoma. *Jco* 23 (9), 2038–2048. doi:10.1200/jco.2005.00.349

Lin, T., Qiu, Y. Q., Peng, W. Y., and Peng, L. S. (2020). Heat Shock Protein 90 Family Isoforms as Prognostic Biomarkers and Their Correlations with Immune Infiltration in Breast Cancer. *Biomed. Res. Int.*, 2020, 2148253. doi:10.1155/2020/2148253

Liu, F. X., Wang, L. L., Yi, S. J., Liu, Q. H., Xu, X. X., and Su, M. (2019). Clinical and Biological Significances of Heat Shock Protein 90 (Hsp90) in Human Nasopharyngeal Carcinoma Cells and Anti-cancer Effects of Hsp90 Inhibitor. *Biomed. Pharmacother.* 120, 109533. doi:10.1016/j.biopha.2019.109533

Ma, B., Gao, P., Wang, H., Xu, Q., Song, Y., Huang, X., et al. (2017). What Has Preoperative Radio(chemo)therapy Brought to Localized Rectal Cancer Patients in Terms of Perioperative and Long-Term Outcomes over the Past

Decades? A Systematic Review and Meta-Analysis Based on 41,121 Patients. *Int. J. Cancer* 141 (5), 1052–1065. doi:10.1002/ijc.30805

Manfredi, S., Lepage, C. m., Hatem, C., Coatmeur, O., Faivre, J., and Bouvier, A.-M. (2006). Epidemiology and Management of Liver Metastases from Colorectal Cancer. *Ann. Surg.* 244 (2), 254–259. doi:10.1097/01.sla.0000217629.94941.cf

Mao, M., Wang, X., Sheng, H., Li, H., Liu, W., Han, R., et al. (2021). Heat Shock Protein 90α Provides an Effective and Novel Diagnosis Strategy for Nasopharyngeal Carcinoma. *Adv. Ther.* 38 (1), 413–422. doi:10.1007/s12325-020-01518-4

Mohammad, W., and Balaa, F. (2009). Surgical Management of Colorectal Liver Metastases. *Clin. colon rectal Surg.* 22 (4), 225–232. doi:10.1055/s-0029-1242462

Morris, E., and Treasure, T. (2018). Surgical Management and Outcomes of Colorectal Cancer Liver Metastases. *Cancer Epidemiol.* 52, 160–161. doi:10.1016/j.canep.2017.12.010

Passarino, G., Cavalleri1, G. L., Stecconi, R., Franceschi, C., Altomare, K., Dato, S., et al. (2003). Molecular Variation of Human HSP90α and HSP90β Genes in Caucasians. *Hum. Mutat.* 21 (5), 554–555. doi:10.1002/humu.9141

Rojas Llimpe, F. L., Di Fabio, F., Ercolani, G., Giampalma, E., Cappelli, A., Serra, C., et al. (2014). Imaging in Resectable Colorectal Liver Metastasis Patients with or without Preoperative Chemotherapy: Results of the PROMETEO-01 Study. *Br. J. Cancer* 111 (4), 667–673. doi:10.1038/bjc.2014.351

Saltz, L. B., Clarke, S., Díaz-Rubio, E., Scheithauer, W., Figer, A., Wong, R., et al. (2008). Combination with Oxaliplatin-Based Chemotherapy as First-Line Therapy in Metastatic Colorectal Cancer: A Randomized Phase III Study. *J. Clin. Oncol.* 26 (12), 2013–2019. doi:10.1200/JCO.2007.14.9930

Shi, Y., Liu, X., Lou, J., Han, X., Zhang, L., Wang, Q., et al. (2014). Plasma Levels of Heat Shock Protein 90 Alpha Associated with Lung Cancer Development and Treatment Responses. *Clin. Cancer Res.* 20 (23), 6016–6022. doi:10.1158/1078-0432.ccr-14-0174

Song, K. H., Oh, S. J., Kim, S., Cho, H., Lee, H. J., Song, J. S., et al. (2020). HSP90A Inhibition Promotes Anti-tumor Immunity by Reversing Multi-Modal Resistance and Stem-like Property of Immune-Refractory Tumors. *Nat. Commun.* 11 (1), 562. doi:10.1038/s41467-019-14259-y

Tang, Y. P., Xie, M. Z., Li, K. Z., Li, J. L., Cai, Z. M., and Hu, B. L. (2020). Prognostic Value of Peripheral Blood Natural Killer Cells in Colorectal Cancer. *Bmc Gastroenterol.* 20 (1), 31. doi:10.1186/s12876-020-1177-8

Tsutsumi, S., Scroggins, B., Koga, F., Lee, M.-J., Trepel, J., Felts, S., et al. (2008). A Small Molecule Cell-Impermeant Hsp90 Antagonist Inhibits Tumor Cell Motility and Invasion. *Oncogene* 27 (17), 2478–2487. doi:10.1038/sj.onc.1210897

Van Cutsem, E., Köhne, C.-H., Hitre, E., Zaluski, J., Chang Chien, C.-R., Makhson, A., et al. (2009). Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. *N. Engl. J. Med.* 360 (14), 1408–1417. doi:10.1056/nejmoa0805019

van Kessel, C. S., Buckens, C. F. M., van den Bosch, M. A. A. J., van Leeuwen, M. S., van Hillegersberg, R., and Verkooijen, H. M. (2012). Preoperative Imaging of Colorectal Liver Metastases after Neoadjuvant Chemotherapy: a Meta-Analysis. *Ann. Surg. Oncol.* 19 (9), 2805–2813. doi:10.1245/s10434-012-2300-z

Wei, W., Liu, M. S., Ning, S. F., Wei, J., Zhong, J. H., Li, J. L., et al. (2020). Diagnostic Value of Plasma HSP90 Alpha Levels for Detection of Hepatocellular Carcinoma. *Bmc Cancer* 20 (1). doi:10.1186/s12885-019-6489-0

Zhang, Y., Ni, L., Li, Q., Li, M., Zhu, J., Zhang, F., et al. (2020). Diagnostic, Clinicopathologic, Therapeutic and Prognostic Value of Plasma Heat Shock Protein 90 Levels in Patients with Advanced Gastrointestinal Carcinoma. *J. Cancer* 11 (20), 5890–5899. doi:10.7150/jca.46343

Zuehlke, A. D., Beebe, K., Neckers, L., and Prince, T. (2015). Regulation and Function of the Human HSP90AA1 Gene. *Gene* 570 (1), 8–16. doi:10.1016/j.gene.2015.06.018

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 handling editor declared a past collaboration with one of the authors CZ.

Copyright © 2021 Wei, Zhou, Chen, Liu, Zhang, Li, Ning, Li, Wang, Huang, Zou 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fmolb.2021.686803

Phenotypic and Functional Analyses of B7S1 in Ovarian Cancer

Dongli Cai[1†], *Fang Wang*[2†], *Changgang Wang*[3*] *and Liping Jin*[1*]

[1]*Clinical and Translational Research Center, Shanghai First Maternity and Infant Hospital, School of Medicine,Tongji University, Shanghai, China,* [2]*Department of Gynaecology, Shanghai East Hospital, School of Medicine,Tongji University, Shanghai, China,* [3]*Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China*

Background: Although programmed death (PD) ligand 1 (PD-L1)/PD-1 inhibitors show potent and durable antitumor effects in a variety of tumors, their efficacy in patients with OvCa is modest. Thus, additional immunosuppressive mechanisms beyond PD-L1/PD-1 need to be identified.

Methods: The mRNA expression profiles of OvCa patients were obtained from The Cancer Genome Atlas (TCGA) database. The expression and clinical characteristics of *VTCN1* (encoding B7S1) in OvCa were analyzed. The molecular interaction network, Gene Ontology (GO) analysis and Gene set enrichment analysis (GSEA) were used to functionally annotate and predict signaling pathways of *VTCN1* in OvCa. Moreover, 32 treatment-naïve patients with OvCa were recruited to assess B7S1 expression. The cytotoxic immune phenotypes in distinct subgroups were analyzed.

Results: B7S1 expression was increased in tumor sections compared with that in normal tissues from OvCa patients at both the mRNA and protein levels. *VTCN1* expression was significantly correlated with the mRNA expression levels of several other co-inhibitory immune checkpoints. B7S1 protein was found to be highly expressed in $CD45^+CD68^+$ myeloid cells, whereas its putative receptor was expressed in $CD8^+$ tumor-infiltrating lymphocytes (TILs). Furthermore, expression of B7S1 in antigen-presenting cells (APCs) was significantly correlated with the cytolytic function of $CD8^+$ TILs. Functional annotations indicated that *VTCN1* was involved in regulating T cell-mediated immune responses and participated in the activation of a variety of classic signaling pathways related to the progression of human cancer.

Conclusion: In OvCa, B7S1 was highly expressed and may initiate dysfunction of $CD8^+$ TILs, which could be targeted for cancer immunotherapy.

Keywords: ovarian cancer, immune checkpoint, B7S1, tumor-infiltrating lymphocytes, immunotherapy

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China

Reviewed by:
Wenlong Lin,
Zhejiang University, China
Kongyang Ma,
Sun Yat-Sen University, China

****Correspondence:***
Liping Jin
jinlp01@163.com
Changgang Wang
changgangwang@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *28 March 2021*
Accepted: *28 June 2021*
Published: *09 July 2021*

Citation:
Cai D, Wang F, Wang C and Jin L (2021) Phenotypic and Functional Analyses of B7S1 in Ovarian Cancer. Front. Mol. Biosci. 8:686803. doi: 10.3389/fmolb.2021.686803

INTRODUCTION

Ovarian cancer (OvCa) is one of the most lethal gynecologic malignancies, with increasing global incidence. Global Cancer Statistics estimated that over 295,414 new OvCa cases and 184,799 OvCa-related deaths occurred worldwide in 2018 (Bray et al., 2018). Owing primarily to the lack of effective screening strategies and the absence of early specific symptoms, more than 75% patients with OvCa present disease stage III or IV at diagnosis (Hung et al., 2016), thereby contributing to a poor overall survival. Although aggressive frontline treatments with taxane/platinum-based chemotherapy and

cytoreductive surgery, more than 70% of patients with advanced-stage cancer relapse within 5 years and become resistant to chemotherapy; the 5-year overall survival rate remains ominously low at 45% (Capriglione et al., 2017). Thus, there is an urgent need for the development of novel therapeutics for advanced OvCa.

Recent studies have shown that, like many solid tumors, OvCa is immunogenic (Zhang et al., 2003) and can elicit a spontaneous antitumor immune response (Cai et al., 2020). Engaging the immune system is a critical component of optimal OvCa therapy, and the quality of tumor-infiltrating T cells (TILs) is a critical determinant of outcomes in patients with OvCa (Zhang et al., 2003; Hamanishi et al., 2007; Hwang et al., 2012). T cell-mediated immune responses are regulated by several costimulatory and co-inhibitory signals (Sun et al., 2018). Evidence in several cancer systems has shown that inhibitory immune checkpoint receptors expressed on T cells promote T-cell exhaustion, dampen host immunity, and assist tumor evasion (Thommen et al., 2018). Blockade of these inhibitory checkpoint receptors with specific antibodies has been reported to reinstate antitumor responses at different levels (Pardoll, 2012).

Programmed death (PD)-1 is a prototypical co-inhibitory checkpoint receptor that has emerged as a critical intrinsic modulatory mechanism for impairing natural antitumor immunity (Gaillard et al., 2016). Interfering with PD-1 or its ligand PD-L1 has been shown to enhance antineoplastic immune responses through the recovery of T-cell function in a wide spectrum of tumors (Mariathasan et al., 2018), including melanoma (Patel and Kurzrock, 2015), non-small cell lung cancer (Patel and Kurzrock, 2015) and renal cell carcinomas (McDermott et al., 2016). However, although the PD-1/PD-L1 pathway blockade produces durable clinical responses in preclinical tests for OvCa, the best overall response rate in 26 PD-L1-positive patients treated with pembrolizumab (anti-PD-1) was only 11.5% (Disis et al., 10718). Since immune checkpoint molecules function nonredundantly and cooperatively to fine-tune immune responses and promote T-cell exhaustion (Li et al., 2018), additional immunosuppressive mechanisms need to be identified.

B7S1 (VTCN1, B7-H4, B7x) is a member of the B7 superfamily and shows 25% amino acid identity with other B7 family members (Hansen et al., 2009). B7S1 mRNA is broadly expressed; however, the restricted pattern of protein expression in normal tissue suggests post-transcriptional regulation (Choi et al., 2003). Aberrant B7S1 expression has been reported in a vast majority of human malignancies, including melanoma (Quandt et al., 2011), renal cell carcinoma (Krambeck et al., 2006), breast cancer (Tringler et al., 2005), non-small cell lung cancer (Sun et al., 2006), thyroid cancer (Zhu et al., 2013), and hepatocellular carcinoma (Kang et al., 2017). B7S1 expression in multiple solid tumors is positively correlated with malignant phenotypes, such as advanced lymph node metastasis, high tumor stage, poor differentiation, early recurrence, and is reversely related to the infiltration intensity of TILs, as well as the overall survival rate (Wang and Wang, 2020).

B7S1 expression is restricted to antigen-presenting cells (APCs), and its putative receptor is induced on activated T cells to inhibit their proliferation, cytokine production, and cytotoxicity (Kryczek et al., 2006; Kryczek et al., 2007; Li et al., 2018). In human hepatocellular carcinoma, B7S1 expression is reportedly upregulated on APCs, and its putative receptor is co-expressed with PD-1 on activated early $CD8^+$ TILs, to promote T cell exhaustion and depress anti-tumor immune response via upregulating Eomes (Li et al., 2018). Studies have shown that B7S1 is closely relevant to classical signaling pathways. In esophageal squamous cell carcinoma, by activating JAK2/STAT3 pathway, B7S1 stimulated the secretion of IL-6, which in turn upregulated the expression of B7S1, thereby mutually enhancing the growth and tumorigenicity of cancer cells (Chen et al., 2016). Incubated T cells with B7S1-Ig fusion protein *in vitro* significantly inhibits the proliferation and cytotoxic activity of T cells, by interfering with the activation of ERK, JNK, and AKT (Wang et al., 2012; Wang and Wang, 2020). B7S1 silencing enhanced drug-induced apoptosis by inhibiting the PTEN/PI3K/AKT pathway in triple-negative breast cancer cells (Wang et al., 2018). Furthermore, blockade of B7S1 down-regulated the transcription of CXCL12/CXCR4. By activating ERK1/2, AKT, PI3K and the other signaling pathways, CXCL12/CXCR4 axis is widely involved in the proliferation, invasion and metastasis of tumor cells (Peng et al., 2015).

B7S1 is highly expressed in primary and metastatic serous, endometrioid, clear cell and epithelial ovarian carcinomas, but is low in mucinous and non-epithelial ovarian carcinomas (Wang and Wang, 2020). According to a recent study, B7S1 was predominantly expressed by ovarian cancer cells, and this alteration is positively correlated with the expression of C-X-C motif chemokine ligand 17 and the proportion of infiltrating mature APCs (MacGregor et al., 2019). However, in 2006, a suppressive macrophage population with B7S1 expression has been detected in human ovarian carcinoma. The $B7S1^+$ macrophages express CD86 at a higher level and have stronger inhibitory activity than $B7S1^-$ macrophages (Kryczek et al., 2006; Kryczek et al., 2007). Due to high expression level of B7S1 in OvCa was found significantly associated with tumor stage (Liang et al., 2016) and a worse progression-free survival (PFS) (Ye et al., 2018), B7S1 may serve as a promising candidate target for OvCa immunotherapy.

Accordingly, in this study, we aimed to determine the expression and roles of B7S1 in OvCa, with a focus on its relation to cancer-associated immune responses. Our findings demonstrated that B7S1 suppressed antitumor immunity and supported the applications of B7S1 as a promising target for immunotherapy in OvCa.

MATERIALS AND METHODS

Analysis of the Public Dataset

RNA sequencing-based gene expression data of samples from patients with OvCa were obtained from Gene Expression Profiling for Interactive Analysis (GEPIA) for Cancer

Genomics (http://gepia.cancer-pku.cn/) (Tang et al., 2017) and TISIDB (http://cis.hku.hk/TISIDB/) (Ru et al., 2019).

Human Specimens

Fresh tumor tissues, malignant ascites, and matched blood were collected from 32 patients with OvCa undergoing primary surgical treatment without chemotherapy at Shanghai First Maternity and Infant Hospital. All experimental protocols were approved by the Ethical Committee of the Shanghai First Maternity and Infant Hospital (IEC approval NO. 2017-100), and informed consent was obtained from patients prior to their enrollment in the study.

Isolation of Peripheral Blood Mononuclear Cells and TILs from Tumors or Ascites

Blood and ascites from patients with OvCa were drawn into heparinized tubes and centrifuged on Ficoll-Hypaque density gradients (cat. no. 17-1440-02; GE Healthcare Life Sciences). Fresh tumor tissues from patients with OvCa were digested in RPMI-1640 medium supplemented with 1 mg/ml collagenase IV (cat. no. 17104019; Gibco) for 30 min at 37°C prior to Ficoll-Hypaque density gradient centrifugation. This method has been described previously (5).

Immunofluorescence

Paraffin sections of human OvCa specimens were dewaxed in xylene, dehydrated in ethanol, subjected to heat-induced epitope retrieval, and then incubated with primary antibodies against human CD45 (cat. no. ab10559; Abcam, Cambridge, United Kingdom) and B7S1 (MIH43; cat. no. ab110221; Abcam, Cambridge, United Kingdom) at 4°C overnight. AffiniPure F (ab')2 Fragment donkey anti-rabbit immunoglobulin (cat. no. 711-546-152; Jackson Immuno Research, West Grove, PA, United States) and AffiniPure F (ab')2 Fragment donkey anti-mouse immunoglobulin (cat. no. 715-166-150; Jackson Immuno Research, West Grove, PA, United States) were chosen as the secondary antibodies. All slides were incubated with mounting medium containing 4′,6-diamidino-2-phenylindole for 20 min. Images were obtained using a Zeiss fluorescence microscope. Quantification analysis was performed using ImageJ software (National Institutes of Health, Bethesda, MD, United States). The method has been described previously (Cai et al., 2020).

Flow Cytometry

The following fluorescent dye-conjugated antibodies were used: anti-CD45-PerCP-CY5.5 (HI30; cat. no.304028, Biolegend), anti-lineage-fluorescein isothiocyanate-FITC (UCHT1, HCD14, 3G8, HIB19, 2H7, HCD56; cat. no. 348701, Biolegend), anti-human leukocyte antigen (HLA)-DR-AF700 (L243; cat. no. 307626Biolegend), anti-CD14-allophycocyanin (APC)-CY7 (M4P9; cat. no. 557831, BD), anti-CD15-phycoerythrin (PE)-CY5 (W6D3; cat. no. 323014, Biolegend), anti-B7S1-PE-CY7 (MIH43; cat. no. 358106, Biolegend), anti-CD3-PerCP-CY5.5 (OKT3; cat. no. 300328, Biolegend), anti-CD56-FITC (HCD56; cat. no. 318304, Biolegend), anti-CD4-PE-CF594 (L200; cat. no. 562402, BD), anti-CD8-AF700 (SK1; BD), anti-FoxP3-PE-CF594 (259D/C7; cat. no. 563955, BD), anti-T cell immunoglobulin and mucin domain-containing protein 3 (TIM-3)-PE (F38-2E2; cat. no. 345006, Biolegend), anti-CD27-PerCP-CY5.5 (O323; cat. no. 302820, Biolegend), anti-Ki-67-AF700 (B56; BD), anti-tumor necrosis factor (TNF)-α-PE-CF594 (Mab11; cat. no. 502946, Biolegend), and anti-interferon (IFN)-γ-PE-CY7 (B27; cat. no. 506518, Biolegend). Dead cells were excluded by viability dye staining (Fixable viability dye eF506; cat. no. 65-0866-18, eBioscience), as described previously (Cai et al., 2020).

To detect the expression of human B7S1 receptor, cell suspensions were incubated with 10 μg/ml biotin-labeled human B7S1-mouse IgG2a Fc fusion protein (generated in-house) at 4°C for 40 min and then with 0.5 μg/ml streptavidin Brilliant Violet 421 (cat. no. 405225 Biolegend) together with surface antibodies. Cells were acquired using an LSRFortessa flow cytometer, and data were analyzed using FlowJo.X software (Tree Star, Ashland, OR, United States).

Protein–Protein Interaction (PPI) Network Construction

The STRING (https://string-db.org/) online database was used to analyze the functional interactions between proteins. Cytoscape (version 3.5) was employed to visualize the molecular interaction network.

Functional Annotations

mRNA expression profiles of OvCa were obtained from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) online database. Gene Ontology (GO) enrichment analysis for hub genes was performed and visualized using ClueGo (version 2.5.7) and CluePedia (version 1.5.7). Gene set enrichment analysis (GSEA) was used to predict potential hallmarks of *VTCN1* in OvCa.

Statistics

Statistical analyses were conducted with Prism 6.0 software (GraphPad), using the appropriate tests as indicated in the legends. All values are expressed as means ± standard errors of the means. Results with p values less than 0.05 were considered statistically significant.

RESULTS

VTCN1 was Highly Expressed in Human OvCa and was Correlated With Multiple Co-inhibitory Immune Checkpoint Genes

The low response rate to PD-1 blockade in OvCa may be associated with co-expression of other co-inhibitory immune checkpoint molecules in the tumor microenvironment (TME). To identify potential targets in patients with OvCa, we examined the mRNA expression levels of several checkpoint molecules in various tumors. RNA-sequencing data from TCGA revealed high upregulation of *VTCN1* (encoding B7S1) in several solid tumors,

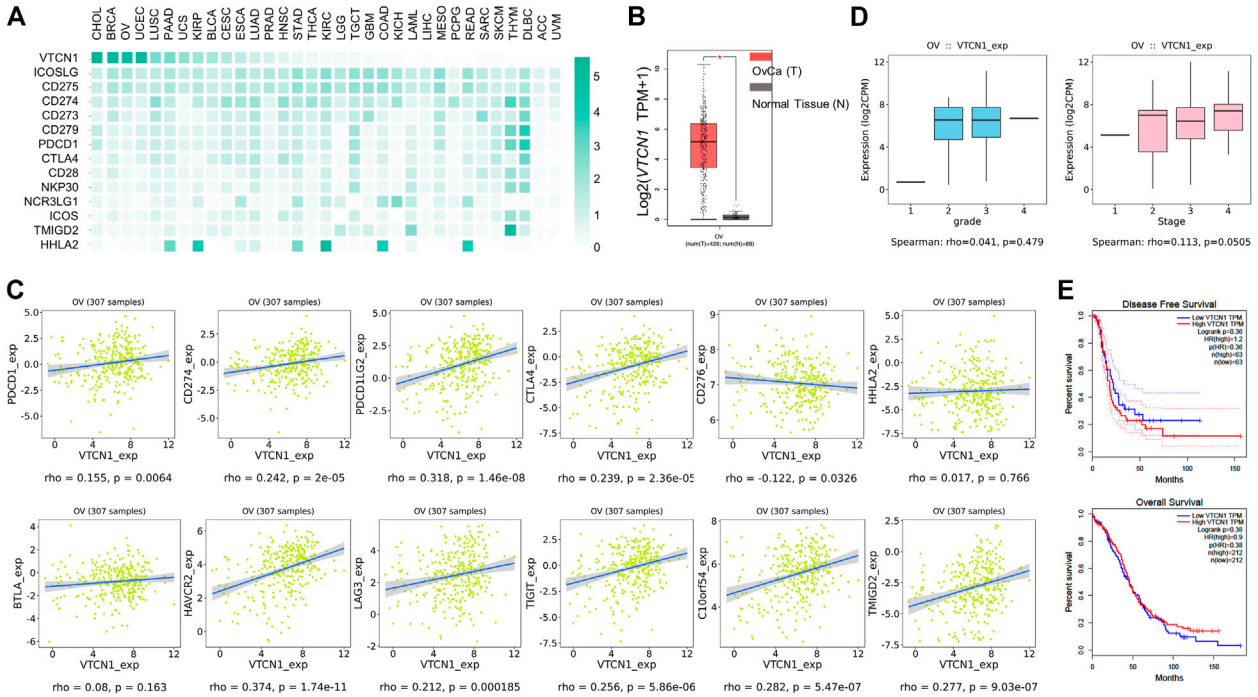


FIGURE 1 | *VTCN1* was overexpressed in OvCa and correlated with multiple co-inhibitory immune checkpoint genes. **(A)** Heatmap analysis of the mRNA expression of B7/CD28 family genes in various human tumors, shown as scaled log2-fold changes (GEPIA data). **(B)** The expression levels of *VTCN1* in human OvCa and normal controls. The data were derived from TCGA database and are shown on a log2 (TPM +1) scale. TPM: transcripts per million. The *p* value cutoff was 0.01. $^{*}p < 0.05$. **(C)** Correlations of *VTCN1* expression with tumor grade (left) and stage (right) in patients with OvCa. **(D)** Correlations of *VTCN1* expression with the mRNA expression levels of several immune checkpoint proteins. **(E)** Correlations of *VTCN1* expression with disease-free survival (upper panel) and overall survival (bottom panel) in patients with OvCa.

including human OvCa, as compared with known B7/CD28 family members (**Figure 1A**). Upregulation of *VTCN1*, i.e., a $\log_2$ (TPM+1) fold increase of more than 32, was observed in OvCa compared with normal tissues (**Figure 1B**). Moreover, the expression of *VTCN1* was significantly positively correlated with the expression of some other checkpoint genes, such as *CTLA4*, *HAVCR2*, *LAG3*, *TIGIT*, and *C10orf54* (**Figure 1C**). In addition, *VTCN1* expression was not significantly correlated with tumor grade or tumor stage (**Figure 1D**). Stratification of patients with OvCa by *VTCN1* expression revealed no significant differences in disease-free survival or overall survival, although patients with high *VTCN1* expression had a high hazard ratio (HR; HR = 1.2) for relapse (**Figure 1E**). The highly expressed B7S1 in OvCa has been reported significantly associated with tumor stage (Liang et al., 2016) and a worse progression-free survival (Ye et al., 2018). The inconsistency between mRNA level and protein level may be related to the complex mechanism of post-transcriptional regulation. Taken together, these results implied that B7S1 may contribute to antitumor immunosuppression in OvCa.

B7S1 Protein Was Mainly Expressed by Intratumoral Myeloid Cells in Human OvCa

Because *VTCN1* is differentially expressed in OvCa subtypes (Chen et al., 2018), we next evaluated the protein levels of B7S1 in tumor samples from primary debulking surgeries of 32 treatment-naïve patients with epithelial OvCa (**Table. 1**) by immunofluorescence and flow cytometry (Cai et al., 2020). Compared with normal ovary tissues, B7S1 was upregulated in tumor sections, as detailed in our previous study (Cai et al., 2020). Contrary to reported data demonstrating that B7S1 expression is restricted to the tumor cell compartment (MacGregor et al., 2019), our study showed that B7S1 was expressed on both $CD45^-$ and $CD45^+$ cells (**Figure 2A**, indicated by the white arrows). In $CD45^+$ cells, B7S1 was predominantly detected within the $CD68^+$ myeloid compartment (**Figure 2A**).

Given that B7S1 could be detected on tumor-infiltrating $CD45^+$ cells, we sought to identify which subsets of myeloid cells expressed this target. TILs from tumor tissues and tumor-associated lymphocytes from ascites were isolated by density-gradient centrifugation. PBMCs from the same patients were used as controls. B7S1 was predominantly expressed on myeloid dendritic cells (mDCs), $CD14^+HLA\text{-}DR^{hi}$ monocytes/macrophages (Mø/Mo), and $CD14^+HLA\text{-}DR^{low/-}$ monocytic myeloid-derived suppressed cells (mMDSCs). The percentages of B7S1-expressing APCs in the TME and ascites were significantly higher than those in peripheral blood (**Figures 2B,C**). We did not find a significant difference in B7S1 expression on $CD15^+$ $HLA\text{-}DR^{low/-}$ granulocytic myeloid-derived suppressed cells (gMDSCs) among tumors, ascites, and PBMCs (**Figures 2B,C**).

TABLE 1 | Clinical and pathological characteristics of patients with OvCa.

Patient no	Age	FIGO Stage	Pathology	Chemotherapy
P#1	32	IV	Low-grade adenocarcinoma	No
P#2	68	IIIc	Low-grade adenocarcinoma	No
P#3	59	IIIc	Endometrioid adenocarcinoma	No
P#4	55	IIIb	Low-grade adenocarcinoma	No
P#5	50	IIIc	Low-grade adenocarcinoma	No
P#6	42	IV	Serous papillary adenocarcinoma	No
P#7	54	IIc	Endometrioid adenocarcinoma	No
P#8	63	IIa	High-grade adenocarcinoma	No
P#9	28	IIIc	Mucinous adenocarcinoma	No
P#10	51	Ic	Adenocarcinoma	No
P#11	54	IIa	Endometrioid adenocarcinoma	No
P#12	45	Ic	Adenocarcinoma	No
P#13	67	IIIa	High-grade adenocarcinoma	No
P#14	48	IIIc	High-grade adenocarcinoma	No
P#15	49	IIIc	High-grade adenocarcinoma	No
P#16	54	IIIc	Low-grade adenocarcinoma	No
P#17	61	IIa	Endometrioid adenocarcinoma	No
P#18	60	-	Low-grade adenocarcinoma	No
P#19	67	IIb	High-grade adenocarcinoma	No
P#20	39	Ia	Mucinous carcinoma	No
P#21	66	IIb	High-grade adenocarcinoma	No
P#22	60	IIa	High-grade adenocarcinoma	No
P#23	48	Ib	Adenocarcinoma	No
P#24	61	IIIc	High-grade adenocarcinoma	No
P#25	55	-	Adenocarcinoma	No
P#26	51	IIc	Adenocarcinoma	No
P#27	-	-	High-grade serous carcinoma	No
P#28	67	IIb	Adenocarcinoma	No
P#29	49	IIIc	High-grade adenocarcinoma	No
P#30	51	IIIb	Low-grade adenocarcinoma	No
P#31	36	Ia	Clear cell carcinoma	No
P#32	46	IIc	High-grade serous carcinoma	No

B7S1 is a negative immune checkpoint protein that binds to activated T cells and inhibits their proliferation and function (Li et al., 2018). To identify the potential targets of B7S1 in OvCa, we utilized a biotin-labeled human B7S1-mouse IgG2a Fc fusion protein to assess the expression of the putative B7S1 receptor (pB7S1R). pB7S1R was detected on $CD4^+$ T, $CD8^+$ T, natural killer (NK), and NK T cells in tumor tissues (**Figures 2D,E**). Because B7S1 expressed by tumor-infiltrating myeloid cells has been reported to induce dysfunction of antitumor $CD8^+$ T cells in liver cancer (Li et al., 2018), the expression patterns of B7S1 and pB7S1R strongly suggested that B7S1 expression in OvCa may inhibit $CD8^+$ T cell function.

B7S1 Expression was Inversely Correlated With the Cytolytic Function of $CD8^+$ TILs

To investigate the relevance of B7S1 expression in the infiltration and function of $CD8^+$ TILs, patients with OvCa were divided into two groups ($B7S1^{hi}$ and $B7S1^{low}$) based on the average frequency of $B7S1^+$ cells in all $CD14^+$ myeloid cells (Hong et al., 2019) (cutoff = 22.9%). Compared with $B7S1^{hi}$ patients, significantly lower frequencies of $CD4^+$ T cells and $CD4^+Foxp3^+$ cells were found in $B7S1^{low}$ patients. In contrast, $B7S1^{low}$ patients had markedly higher fractions of $CD8^+$ TILs and CD8/CD4 ratios (**Figures 3A,B**). Moreover, $CD8^+$ TILs isolated from $B7S1^{hi}$ patients displayed the characteristic exhausted T-cell phenotypes, including higher levels of the co-inhibitory molecule PD-1, decreased surface expression of the co-stimulatory molecule CD27 and proliferation marker Ki67, and decreased levels of TNF-α upon phorbol myristate acetate/ionomycin stimulation in comparison with those from $B7S1^{low}$ patients (**Figures 3C,D**). $CD8^+$ TILs isolated from $B7S1^{hi}$ patients tended to show increased TIM-3 expression (**Figures 3C,D**); the lack of a significant difference between the two groups may have been related to the small sample size in this study. In addition, although there was no statistically significant difference in the proportion of IFN-γ^+ CD8 TILs (data not shown) between the two groups, which may be due to the lack of NK cells in some patients, the frequency of TNF-α^+IFN-γ^+ CD8 TILs was found to be increased in $B7S1^{low}$ patients (**Figures 3C,D**). Since almost half of $CD4^+$ T cells expressed pB7S1R as shown in **Figure 1D**, we also analyzed the expression of PD-1 and IFN-γ in $CD4^+$ TILs. It was found that there were no obvious differences between $B7S1^{hi}$ and $B7S1^{low}$ groups (**Figure 3E**), which could be due to the limited sample size. Collectively, these results indicated that B7S1 expression on APCs was inversely correlated with the infiltration and cytolytic function of $CD8^+$ TILs in OvCa.

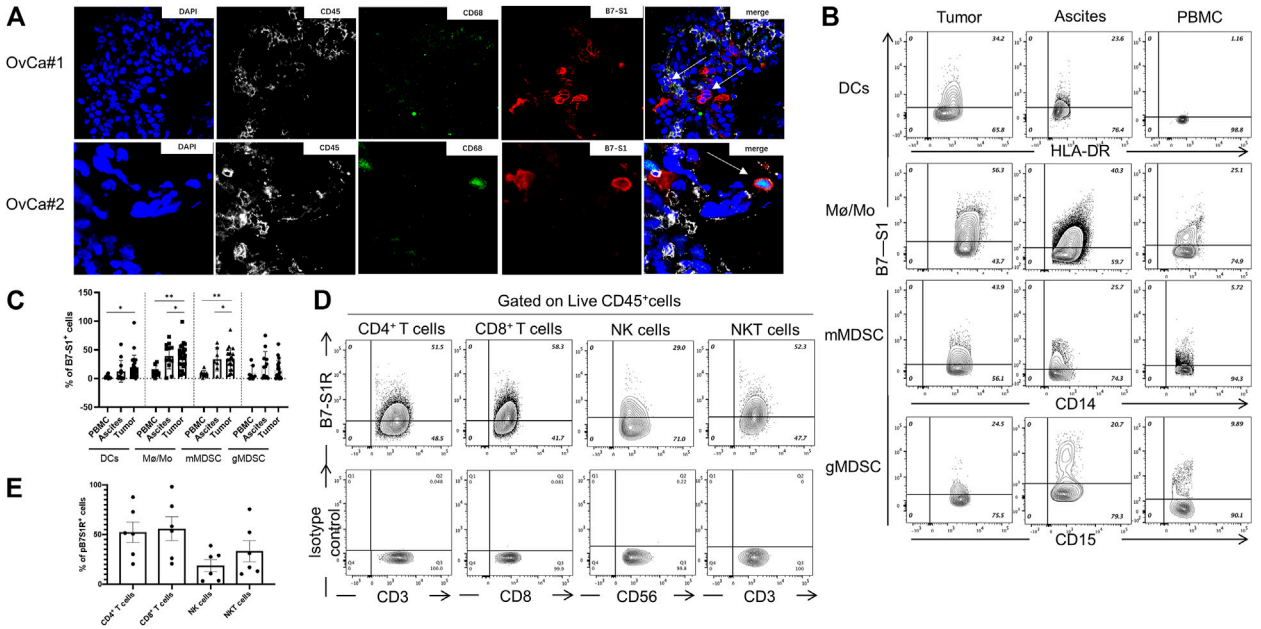


FIGURE 2 | B7S1 protein was mainly expressed in intratumoral myeloid cells in human OvCa. **(A)** Representative immunofluorescence images of B7S1 in human OvCa. Original magnification: 40x (OvCa#1), 100x(OvCa#2). **(B,C)** Representative figures and summarized data showing the percentages of B7S1-positive DCs, Mø/Mo, mMDSCs, and gMDSCs in tumors, ascites, and PBMCs from patients with OvCa. Statistical analysis was performed using one-way analysis of variance followed by Tukey's multiple comparison tests. The data are presented as means ± standard errors of the means; $^{ns}P > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. **(D,E)** Representative figures and summarized data of pB7S1R expression on $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and NK T cells in OvCa tumors detected by a biotinylated hB7S1-mIgG2a Fc fusion protein.

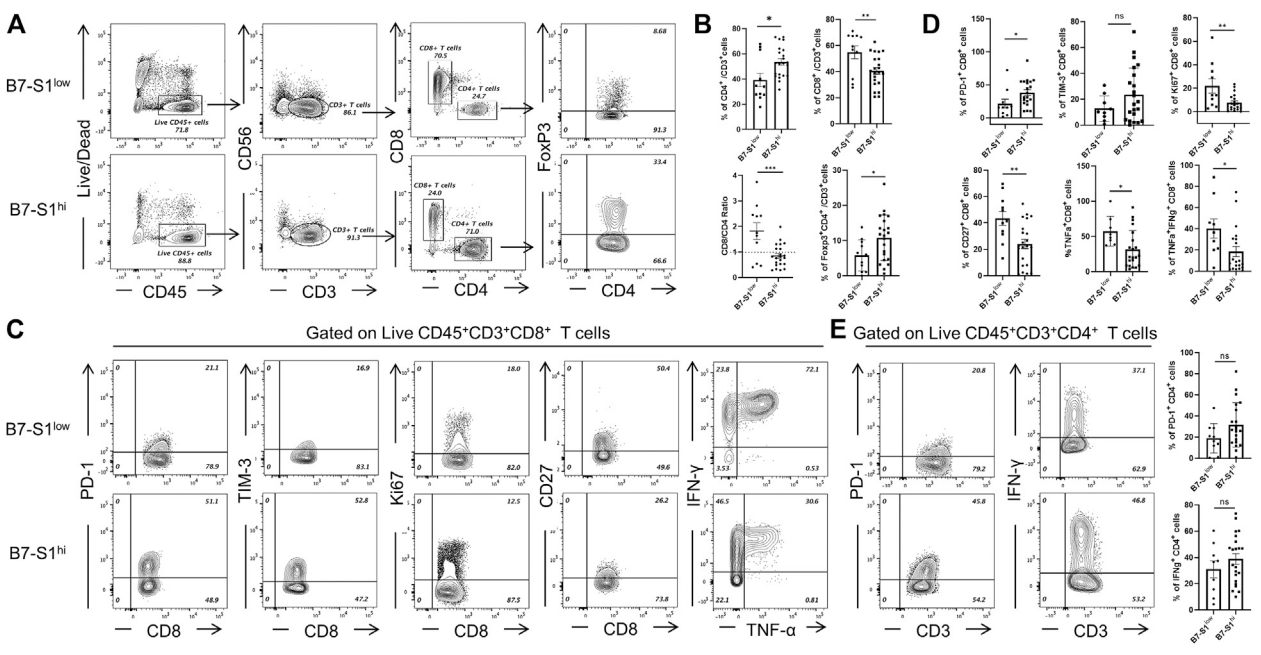


FIGURE 3 | B7S1 expression was inversely correlated with the cytolytic function of CD8+ TILs. **(A,B)** Representative figures and summarized data of infiltrated T cells in tumor tissues of patients with OvCa in $B7S1^{hi}$ and $B7S1^{low}$ groups. **(C,D)** Representative figures and summarized data displaying PD-1, Ki67, CD27, TNFα, and IFNγ expression in $CD8^+$ T cells in tumors. **(E)** Representative figures and summarized data displaying PD-1 and IFNγ expression in $CD4^+$ T cells in tumors. The data are presented as means ± standard errors of the means; *p* value was computed by *t*-test. $^{ns}P > 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

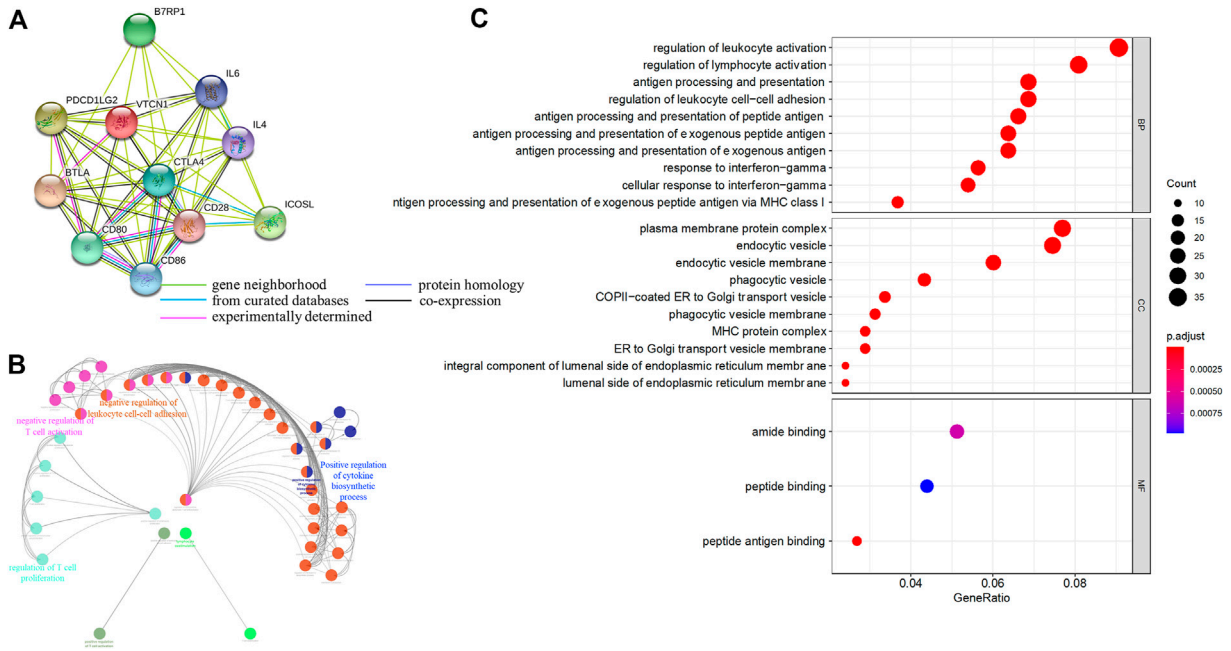


FIGURE 4 | Functional annotations and predicted signaling pathway of *VTCN1* in OvCa. **(A)** The PPI network of *VTCN1* constructed using STRING. **(B)** Functional annotation of the *VTCN1*-related module genes constructed by ClueGo. **(C)** Significantly enriched GO annotations of *VTCN1* in OvCa. BP: biological processes. CC: cellular components; MF: molecular functions.

Functional Annotations and Predicted Signaling Pathways

Next, we systematically analyzed the biological functions of *VTCN1* in OvCa. A PPI network of potential targets was constructed (**Figure 4A**), and the most significant modules were acquired using Cytoscape (**Figure 4B**). Functional analyses of *VTCN1* demonstrated that genes in this module were mainly enriched in the regulation of T-cell activation, T-cell proliferation, leukocyte cell-cell adhesion and cytokine biosynthetic process (**Figure 4B**). We selected genes that were co-expressed with *VTCN1* in OvCa from TCGA database, and the top 500 ($p < 0.01$ and correlation >0.3) positively and negatively correlated genes were processed *via* GO functional analysis using the cluster-filer package of R software (Yu et al., 2012).We found that the significant GO terms in the biological process category were tightly associated with the regulation of lymphocyte activation, antigen processing and presentation, and response to IFN-γ (**Figure 4C**).

To identify the *VTCN1*-related pathways activated in OvCa, we conducted GSEA between low and high *VTCN1* expression datasets. Significant differences (false-discovery rate <0.25, $p < 0.05$) in the enrichment of the Molecular Signature database Collection are shown in **Figure 5**. The results showed that *VTCN1* was mainly involved in IL-2/signal transducer and activator of transcription (STAT)5 signaling, p53 pathway, mammalian target of rapamycin complex 1 (mTORC1) signaling, apoptosis, TNF-α signaling via nuclear factor (NF)-κB, inflammatory response, IFN-γ response, IFN-α response, IL-6/Janus kinase (JAK)/STAT3 signaling, reactive oxygen species signaling, WNT/β-catenin signaling, and KRAS signaling. These results showed that high expression of *VTCN1* was closely associated with antitumor immune responses and malignancy in OvCa.

DISCUSSION

In this study, we demonstrated that B7S1 was highly expressed in OvCa tumor tissues compared with that in non-tumor tissues at both the mRNA and protein levels. B7S1 was mainly expressed by tumor-infiltrating APCs, and its putative receptor was expressed by $CD8^+$ TILs in human OvCa. The expression of B7S1 in the TME was strongly correlated with poor $CD8^+$ T-cell responses. Functional enrichment and GSEA analyses illustrated that *VTCN1* was significantly involved in T-cell regulation, cancer-related pathways, and hallmarks in OvCa. These results indicated that B7S1 could serve as a novel biomarker for diagnosis and could be considered as a potential immunotherapeutic target.

High expression of B7S1 in OvCa has been described, and the overexpression of B7S1 has been identified to be associated with tumor stage (Liang et al., 2016) and a worse progression-free survival (Ye et al., 2018). In our study, we further demonstrated robust expression of B7S1 in human OvCa. However, *VTCN1* expression was not significantly correlated with tumor grade and disease-free survival, an observation not consistent with protein levels, suggesting that it may be related to the mechanism of post-transcriptional regulation. In contrast to a previous study showing that B7S1 is predominantly expressed by tumor cells, we found that B7S1 could be detected in both $CD45^-$ and $CD45^+$

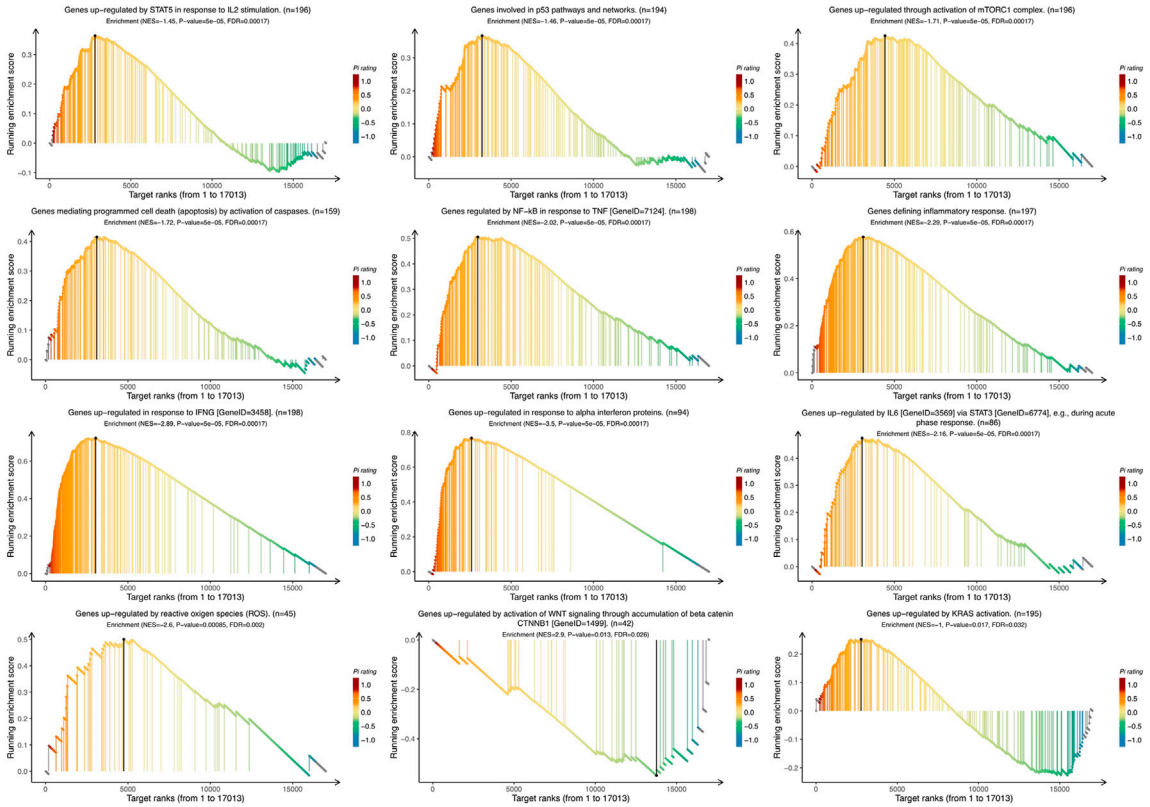


FIGURE 5 | The Gene Set Enrichment Analysis (GSEA) of the relationship between the expression level of VTCN1 in the TCGA ovarian cancer dataset. The most involved significant hallmark pathways which were closely correlated with *VTCN1* in OvCa obtained by GSEA. NES: normalized enrichment score; FDR: false discovery rate.

cells in the OvCa TME. In $CD45^+$ cells, B7S1 was mainly detected within the $CD68^+$ myeloid compartment, particularly mDCs, Mø/Mo, and mMDSCs. Moreover, the expression levels of B7S1 in tumor sections and ascites were significantly higher than those in PBMCs. According to previous reports (Motzer et al., 2015; Patel and Kurzrock, 2015; Yi et al., 2018; Cai et al., 2020), high PD-L1 expression in the TME is favorably correlated with increased response rates and clinical benefits in PD-L1/PD-1 blockade therapies. However, in OvCa, PD-L1 is observed rather low expressed on tumor cells, and the percentage of PD-L1-expressing APCs is not high in tumor sections or ascites (5). The low expression levels of PD-L1 may contribute to the low overall response rates in patients with OvCa after anti-PD-L1/PD-1 therapies. The high levels of B7S1 in the TME and the distinct expression patterns from PD-L1 indicate that B7S1 may be a potential therapeutic target in patients with OvCa who are insensitive to PD-L1/PD-1 inhibition.

The expression level of B7S1 is negatively correlated with the density of TILs (Wang and Wang, 2020). In our study, compared with $B7S1^{hi}$ patients, $B7S1^{low}$ patients displayed decreased frequencies of $CD4^+$ T cells and $CD4^+Foxp3^+$ cells, but increased fractions of $CD8^+$ TILs and CD8/CD4 ratios. In hepatocellular carcinoma, B7S1 has been found to be upregulated in APCs and to be related to T-cell exhaustion *via* its receptor expressed on early activated $CD8^+$ TILs. B7S1 blockade was found to promote $CD8^+$ T cell-mediated antitumor immunity in a murine cancer model (Li et al., 2018). Consistent with this, in OvCa, pB7S1R has been detected in $CD4^+$ T, $CD8^+$ T, NK, and NK T cells. The expression patterns of B7S1 and pB7S1R strongly suggest that B7S1 has an important role in regulating T cells in antitumor immunity in OvCa. Indeed, compared with $B7S1^{low}$ patients, $B7S1^{hi}$ patients displayed more severe immunosuppression in the TME, with a higher percentage of regulatory T cells and a lower fraction of $CD8^+$ TILs. Moreover, $CD8^+$ TILs isolated from $B7S1^{hi}$ patients displayed characteristic exhausted T-cell phenotypes, including PD-1 expression, decreased CD27 and Ki67 expression, and TNF-α secretion. Thus, through cross-presentation between APCs and $CD8^+$ T cells, B7S1 may facilitate the suppression of antitumor immunity in OvCa.

Using publicly available clinical data from TCGA, we focused on the biological functions and underlying mechanisms of *VTCN1* in the TME *via* GO enrichment analysis and GSEA. Several biological processes correlated with immune regulation were identified, including regulation of leukocyte activation, regulation of lymphocyte activation, antigen processing and presentation, regulation of leukocyte cell–cell adhesion, and response to IFN-γ. Importantly, B7S1 can be detected on

APCs isolated from OvCa, and high expression of B7S1 is associated with decreased cytotoxic activity of $CD8^+$ TILs. Moreover, B7S1 has been reported to inhibit the activation and function of T cells, potently suppressing the proliferation, cytokine production, and cytotoxicity of activated T cells (Sica et al., 2003). When evaluating the underlying mechanisms of *VTCN1*, we found that *VTCN1* was significantly involved in most significant hallmark pathways, including IL-2/STAT5 signaling, p53 pathway, mTORC1 signaling, apoptosis, TNF-α signaling *via* NF-κB, inflammatory response, IFN-γ response, IFN-α response, IL-6/JAK/STAT3 signaling, reactive oxygen species signaling, WNT/β-catenin signaling, and KRAS signaling, in OvCa. These hallmark pathways are related to inflammation, immune regulation, tumor suppressor mutations, reactive oxygen species, cell cycle regulation, and apoptosis. Thus, our findings highlighted the specific associations that may trigger carcinogenesis.

Taken together, our results showed that overexpression of B7S1 in OvCa was positively correlated with antitumor immunosuppression. We also identified a potential immunotherapeutic target for OvCa. Further studies are needed to confirm these results and elucidate the underlying mechanisms.

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 Ethical Committee of the Shanghai First Maternity and Infant Hospital (IEC approval NO. 2017-100). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DC performed the experiments. CW conducted the bioinformatic analysis. FW supervised and collected clinical specimens. DC, FW, and CW wrote and revised the article. LJ supervised the study.

FUNDING

This project was funded in part by the National Natural Science Foundation of China (82071653, 81730039,81671460), the National Key Research and Development Program of China (2017YFC1001401), and Shanghai Municipal Medical and Health Discipline Construction Projects (2017ZZ02015) to LJ.

REFERENCES

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

Cai, D., Li, J., Liu, D., Hong, S., Qiao, Q., Sun, Q., et al. (2020). Tumor-expressed B7-H3 Mediates the Inhibition of Antitumor T-Cell Functions in Ovarian Cancer Insensitive to PD-1 Blockade Therapy. *Cell Mol Immunol* 17 (3), 227–236. doi:10.1038/s41423-019-0305-2

Capriglione, S., Luvero, D., Plotti, F., Terranova, C., Montera, R., Scaletta, G., et al. (2017). Ovarian Cancer Recurrence and Early Detection: May HE4 Play a Key Role in This Open challenge? A Systematic Review of Literature. *Med. Oncol.* 34 (9), 164. doi:10.1007/s12032-017-1026-y

Chen, G. M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D. M. A., et al. (2018). Consensus on Molecular Subtypes of High-Grade Serous Ovarian Carcinoma. *Clin. Cancer Res.* 24 (20), 5037–5047. doi:10.1158/1078-0432.CCR-18-0784

Chen, X., Wang, L., Wang, W., Zhao, L., and Shan, B. (2016). B7-H4 Facilitates Proliferation of Esophageal Squamous Cell Carcinoma Cells through Promoting Interleukin-6/signal Transducer and Activator of Transcription 3 Pathway Activation. *Cancer Sci.* 107 (7), 944–954. doi:10.1111/cas.12949

Choi, I. H., Zhu, G., Sica, G. L., Strome, S. E., Cheville, J. C., Lau, J. S., et al. (2003). Genomic Organization and Expression Analysis of B7-H4, an Immune Inhibitory Molecule of the B7 Family. *J. Immunol.* 171 (9), 4650–4654. doi:10.4049/jimmunol.171.9.4650

Disis, M. L. P. M., S. P., E. P. H., A. C. L., and Avelumab, al. e. (0010). Anti- PD-L1) in Patients with Recurrent/refractory Ovarian Cancer from the JAVELIN Solid Tumor Phase Ib Trial: Safety and Clinical Activity. *J. Clin. Oncol.* 2016 (34). doi:10.1016/j.immuni.2016.05.001

Gaillard, S. L., Secord, A. A., and Monk, B. (2016). The Role of Immune Checkpoint Inhibition in the Treatment of Ovarian Cancer. *Gynecol. Oncol. Res. Pract.* 3, 11. doi:10.1186/s40661-016-0033-6

Hamanishi, J., Mandai, M., Iwasaki, M., Okazaki, T., Tanaka, Y., Yamaguchi, K., et al. (2007). Programmed Cell Death 1 Ligand 1 and Tumor-Infiltrating CD8+ T Lymphocytes Are Prognostic Factors of Human Ovarian Cancer. *Proc. Natl. Acad. Sci. U S A.* 104 (9), 3360–3365. doi:10.1073/pnas.0611533104

Hansen, J. D., Du Pasquier, L., Lefranc, M. P., Lopez, V., Benmansour, A., and Boudinot, P. (2009). The B7 Family of Immunoregulatory Receptors: a Comparative and Evolutionary Perspective. *Mol. Immunol.* 46 (3), 457–472. doi:10.1016/j.molimm.2008.10.007

Hong, S., Yuan, Q., Xia, H., Zhu, G., Feng, Y., Wang, Q., et al. (2019). Analysis of VISTA Expression and Function in Renal Cell Carcinoma Highlights VISTA as a Potential Target for Immunotherapy. *Protein Cell* 10 (11), 840–845. doi:10.1007/s13238-019-0642-z

Hung, H. I., Klein, O. J., Peterson, S. W., Rokosh, S. R., Osseiran, S., Nowell, N. H., et al. (2016). PLGA Nanoparticle Encapsulation Reduces Toxicity while Retaining the Therapeutic Efficacy of EtNBS-PDT *In Vitro*. *Sci. Rep.* 6, 33234. doi:10.1038/srep33234

Hwang, W. T., Adams, S. F., Tahirovic, E., Hagemann, I. S., and Coukos, G. (2012). Prognostic Significance of Tumor-Infiltrating T Cells in Ovarian Cancer: a Meta-Analysis. *Gynecol. Oncol.* 124 (2), 192–198. doi:10.1016/j.ygyno.2011.09.039

Kang, F. B., Wang, L., Sun, D. X., Li, H. J., Li, D., Wang, Y., et al. (2017). B7-H4 Overexpression Is Essential for Early Hepatocellular Carcinoma Progression and Recurrence. *Oncotarget* 8 (46), 80878–80888. doi:10.18632/oncotarget.20718

Krambeck, A. E., Thompson, R. H., Dong, H., Lohse, C. M., Park, E. S., Kuntz, S. M., et al. (2006). B7-H4 Expression in Renal Cell Carcinoma and Tumor Vasculature: Associations with Cancer Progression and Survival. *Proc. Natl. Acad. Sci. U S A.* 103 (27), 10391–10396. doi:10.1073/pnas.0600937103

Kryczek, I., Wei, S., Zhu, G., Myers, L., Mottram, P., Cheng, P., et al. (2007). Relationship between B7-H4, Regulatory T Cells, and Patient Outcome in Human Ovarian Carcinoma. *Cancer Res.* 67 (18), 8900–8905. doi:10.1158/0008-5472.CAN-07-1866

Kryczek, I., Zou, L., Rodriguez, P., Zhu, G., Wei, S., Mottram, P., et al. (2006). B7-H4 Expression Identifies a Novel Suppressive Macrophage Population in Human Ovarian Carcinoma. *J. Exp. Med.* 203 (4), 871–881. doi:10.1084/jem.20050930

Li, J., Lee, Y., Li, Y., Jiang, Y., Lu, H., Zang, W., et al. (2018). Co-inhibitory Molecule B7 Superfamily Member 1 Expressed by Tumor-Infiltrating Myeloid Cells Induces Dysfunction of Anti-tumor CD8(+) T Cells. *Immunity* 48 (4), 773–786.e5. doi:10.1016/j.immuni.2018.03.018

Liang, L., Jiang, Y., Chen, J. S., Niu, N., Piao, J., Ning, J., et al. (2016). B7-H4 Expression in Ovarian Serous Carcinoma: a Study of 306 Cases. *Hum. Pathol.* 57, 1–6. doi:10.1016/j.humpath.2016.06.011

MacGregor, H. L., Garcia-Batres, C., Sayad, A., Elia, A., Berman, H. K., Toker, A., et al. (2019). Tumor Cell Expression of B7-H4 Correlates with Higher Frequencies of Tumor-Infiltrating APCs and Higher CXCL17 Expression in Human Epithelial Ovarian Cancer. *Oncoimmunology* 8 (12), e1665460. doi:10.1080/2162402X.2019.1665460

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFbeta Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

McDermott, D. F., Sosman, J. A., Sznol, M., Massard, C., Gordon, M. S., Hamid, O., et al. (2016). Atezolizumab, an Anti-programmed Death-Ligand 1 Antibody, in Metastatic Renal Cell Carcinoma: Long-Term Safety, Clinical Activity, and Immune Correlates from a Phase Ia Study. *J. Clin. Oncol.* 34 (8), 833–842. doi:10.1200/JCO.2015.63.7421

Motzer, R. J., Rini, B. I., McDermott, D. F., Redman, B. G., Kuzel, T. M., Harrison, M. R., et al. (2015). Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. *J. Clin. Oncol.* 33 (13), 1430–1437. doi:10.1200/JCO.2014.59.0703

Pardoll, D. M. (2012). The Blockade of Immune Checkpoints in Cancer Immunotherapy. *Nat. Rev. Cancer* 12 (4), 252–264. doi:10.1038/nrc3239

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

Peng, H. X., Wu, W. Q., Yang, D. M., Jing, R., Li, J., Zhou, F. L., et al. (2015). Role of B7-H4 siRNA in Proliferation, Migration, and Invasion of LOVO Colorectal Carcinoma Cell Line. *Biomed. Res. Int.* 2015, 326981. doi:10.1155/2015/326981

Quandt, D., Fiedler, E., Boettcher, D., Marsch, W., and Seliger, B. (2011). B7-h4 Expression in Human Melanoma: its Association with Patients' Survival and Antitumor Immune Response. *Clin. Cancer Res.* 17 (10), 3100–3111. doi:10.1158/1078-0432.CCR-10-2268

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

Sica, G. L., Choi, I. H., Zhu, G., Tamada, K., Wang, S. D., Tamura, H., et al. (2003). B7-H4, a Molecule of the B7 Family, Negatively Regulates T Cell Immunity. *Immunity* 18 (6), 849–861. doi:10.1016/s1074-7613(03)00152-3

Sun, C., Mezzadra, R., and Schumacher, T. N. (2018). Regulation and Function of the PD-L1 Checkpoint. *Immunity* 48 (3), 434–452. doi:10.1016/j.immuni.2018.03.014

Sun, Y., Wang, Y., Zhao, J., Gu, M., Giscombe, R., Lefvert, A. K., et al. (2006). B7-H3 and B7-H4 Expression in Non-small-cell Lung Cancer. *Lung Cancer* 53 (2), 143–151. doi:10.1016/j.lungcan.2006.05.012

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

Thommen, D. S., Koelzer, V. H., Herzig, P., Roller, A., Trefny, M., Dimeloe, S., et al. (2018). A Transcriptionally and Functionally Distinct PD-1(+) CD8(+) T Cell Pool with Predictive Potential in Non-small-cell Lung Cancer Treated with PD-1 Blockade. *Nat. Med.* 24 (7), 994–1004. doi:10.1038/s41591-018-0057-z

Tringler, B., Zhuo, S., Pilkington, G., Torkko, K. C., Singh, M., Lucia, M. S., et al. (2005). B7-h4 Is Highly Expressed in Ductal and Lobular Breast Cancer. *Clin. Cancer Res.* 11 (5), 1842–1848. doi:10.1158/1078-0432.CCR-04-1658

Wang, J. Y., and Wang, W. P. (2020). B7-H4, a Promising Target for Immunotherapy. *Cell Immunol* 347, 104008. doi:10.1016/j.cellimm.2019.104008

Wang, L., Yang, C., Liu, X. B., Wang, L., and Kang, F. B. (2018). B7-H4 Overexpression Contributes to Poor Prognosis and Drug-Resistance in Triple-Negative Breast Cancer. *Cancer Cel Int* 18, 100. doi:10.1186/s12935-018-0597-9

Wang, X., Hao, J., Metzger, D. L., Ao, Z., Chen, L., Ou, D., et al. (2012). B7-H4 Treatment of T Cells Inhibits ERK, JNK, P38, and AKT Activation. *PLoS One* 7 (1), e28232. doi:10.1371/journal.pone.0028232

Ye, Y., Wang, J. J., Li, S. L., Wang, S. Y., and Jing, F. H. (2018). Does B7-H4 Expression Correlate with Clinicopathologic Characteristics and Survival in Ovarian Cancer?: A Systematic Review and PRISMA-Compliant Meta-Analysis. *Medicine (Baltimore)* 97 (32), e11821. doi:10.1097/MD.0000000000011821

Yi, M., Jiao, D., Xu, H., Liu, Q., Zhao, W., Han, X., et al. (2018). Biomarkers for Predicting Efficacy of PD-1/pd-L1 Inhibitors. *Mol. Cancer* 17 (1), 129. doi:10.1186/s12943-018-0864-3

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, 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

Zhu, J., Chu, B. F., Yang, Y. P., Zhang, S. L., Zhuang, M., Lu, W. J., et al. (2013). B7-H4 Expression Is Associated with Cancer Progression and Predicts Patient Survival in Human Thyroid Cancer. *Asian Pac. J. Cancer Prev.* 14 (5), 3011–3015. doi:10.7314/apjcp.2013.14.5.3011

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.

Copyright © 2021 Cai, Wang, Wang and Jin. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fmolb.2021.690151

Identification of Arp2/3 Complex Subunits as Prognostic Biomarkers for Hepatocellular Carcinoma

Shenglan Huang[1,2], *Dan Li*[1,2], *LingLing Zhuang*[2,3], *Liying Sun*[1,2] *and Jianbing Wu*[1,2]*

[1]Department of Oncology, The Second Affiliated Hospital of Nanchang University, Nanchang, China, [2]Jiangxi Key Laboratory of Clinical and Translational Cancer Research, Nanchang, China, [3]Department of Gynaecology, The Second Affiliated Hospital of Nanchang University, Nanchang, China

The actin-related protein 2/3 complex (Arp2/3) is a major actin nucleator that has been widely reported and plays an important role in promoting the migration and invasion of various cancers. However, the expression patterns and prognostic values of Arp2/3 subunits in hepatocellular carcinoma (HCC) remain unclear. In this study, The Cancer Genome Atlas (TCGA) and UCSC Xena databases were used to obtain mRNA expression and the corresponding clinical information, respectively. The differential expression and Arp2/3 subunits in HCC were analyzed using the "limma" package of R 4.0.4 software. The prognostic value of each subunit was evaluated using Kaplan–Meier survival analysis and Cox proportional hazards regression analyses. The results revealed that mRNA expression of Arp2/3 members (ACTR2, ACTR3, ARPC1A, APRC1B, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L) was upregulated in HCC. Higher expression of Arp2/3 members was significantly correlated with worse overall survival (OS) and shorter progression-free survival (PFS) in HCC patients. Cox proportional hazards regression analyses demonstrated that ACTR3, ARPC2, and ARPC5 were independent prognostic biomarkers of survival in patients with HCC. The relation between tumor immunocyte infiltration and the prognostic subunits was determined using the TIMER 2.0 platform and the GEPIA database. Gene set enrichment analysis (GSEA) was performed to explore the potential mechanisms of prognostic subunits in the carcinogenesis of HCC. The results revealed that ACTR3, ARPC2, and ARPC5 were significantly positively correlated with the infiltration of immune cells in HCC. The GSEA results indicated that ACTR3, ARPC2, and ARPC5 are involved in multiple cancer-related pathways that promote the development of HCC. In brief, various analyses indicated that Arp2/3 complex subunits were significantly upregulated and predicted worse survival in HCC, and they found that ACTR3, ARPC2, and ARPC5 could be used as independent predictors of survival and might be applied as promising molecular targets for diagnosis and therapy of HCC in the future.

Keywords: hepatocellular carcinoma (HCC), Arp2/3 complex subunits, prognosis, immune infiltration, clinical features

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Anupam Nath Jha,
Tezpur University, India
Vikram Srivastava,
Iowa State University, United States

***Correspondence:**
Jianbing Wu
hhgwjb@163.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *02 April 2021*
Accepted: *14 June 2021*
Published: *09 July 2021*

Citation:
Huang S, Li D, Zhuang L, Sun L and Wu J (2021) Identification of Arp2/3 Complex Subunits as Prognostic Biomarkers for Hepatocellular Carcinoma. Front. Mol. Biosci. 8:690151. doi: 10.3389/fmolb.2021.690151

Abbreviations: Arp2/3, actin-related protein 2/3 complex; CCLC, Broad Institute Cancer Cell Line Encyclopedia; FPKM, fragments per kilobase of transcript per million mapped reads; GSEA, gene set enrichment analysis; HCC, hepatocellular carcinoma; HPA, The Human Protein Atlas; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival (OS); PFS, progress free survival; TCGA, The Cancer Genome Atlas; TIICs, tumor-infiltrating immune cells.

INTRODUCTION

Liver cancer is the fourth leading cause of cancer-related mortality and ranks sixth in terms of incidence rate, and it is estimated that more than one million patients will die from liver cancer in 2030 (Akinyemiju et al., 2017; Villanueva, 2019). Hepatocellular carcinoma (HCC), which accounts for 75–80% of liver cancers, is the most common liver cancer, with its morbidity and prevalence increasing annually (Tan et al., 2003). Approximately 700,000 patients are newly diagnosed with HCC every year, with over half of the cases occurring in developing countries, and Asian countries account for three-quarters of HCC-related deaths (Asia-Pacific Working Part, 2010). This high fatality rate is mainly due to the low early diagnosis rate of HCC, rapid progress, fewer treatments for advanced cancer, particularly with high heterogeneity in cancer, undefined molecular mechanisms, and lack of early prognostic indicators. Therefore, it is imperative to search for highly sensitive and specific prognostic markers and potential drug targets to clarify the molecular mechanisms and improve the prognosis of HCC patients.

The actin-related protein 2/3 complex (Arp2/3) is a major actin nucleator that is responsible for promoting the nucleation of microfilaments, facilitating the assembly process of intracellular actin monomers into microfilaments, which forms the cellular structure and promotes processes involved in the formation of cell–cell junctions, motility of pathogens, and transport of vesicles (García-Ponce et al., 2015). The process of actin filament nucleation also plays an important role in the formation of invasive pseudopodia in cancer cells (Firat-Karalar and Welch, 2011). Abnormal migration and invasion are critical factors for tumor metastasis. Kiuchi et al. (Kiuchi et al., 2011) have shown that Arp2/3 is related to the formation of pseudopodia and the movement of bladder cancer cells. It has also been reported that high expression of Arp2/3 positively correlates with the malignancy of glioma specimens and that Arp2/3 system deregulation promotes cancer progression and directly impacts patient survival (Liu et al., 2013; Molinie and Gautreau, 2018). Thus, Arp2/3 plays a crucial role in tumor invasion and metastasis.

The actin-related protein 2/3 complex family (Arp2/3) consists of seven evolutionarily conserved subunits, including two actin-associated proteins, the Arp2 and Arp3 subunits (ACTR2 and ACTR3), and five accessory subunits, ARPC1, ARPC2, ARPC3, ARPC4, and ARPC5. ARPC1 has two subtypes in humans: ARPC1A and ARPC1B. ARPC5 subunits are classified into ARPC5 and ARPC5L (Abella et al., 2016). The center of the Arp2/3 complex is composed of ARPC2 and ARPC4, forming a C-type structure, while the other subunits interact around the center, forming a stable Arp2/3 complex, and ACTR2 and ACTR3 are in contact with the pointed end of the new daughter filament (Pollard and Beltzner, 2002). Many studies have shown that abnormal expression of Arp2/3 subunits is associated with the proliferation and invasion of various cancers, including pancreatic cancer (Rauhala et al., 2013), gastric cancer (Zhang et al., 2017a), colorectal cancer (Su et al., 2018), breast cancer (Chen et al., 2019; Cheng et al., 2019), bladder cancer (Chen et al., 2019; Cheng et al., 2019), gliomas (Liu et al., 2013), lung squamous cell carcinoma (lung SCC) (Moriya et al., 2012), and head and neck squamous cell carcinoma (HNSCC) (Kinoshita et al., 2012). However, the significance of the whole Arp2/3 subunit expression and the prognostic value of HCC has not yet been determined. Few studies have reported the correlation between the mRNA expression of Arp2/3 subunits and immune infiltration in HCC.

In this study, the mRNA expression and the prognostic values of the Arp2/3 family members were comprehensively evaluated in HCC according to updated public resources and multiple bioinformatics analyses. Furthermore, we investigated the potential correlation between Arp2/3 family members and immune cell infiltration levels in HCC.

MATERIALS AND METHODS

Data Acquisition

The high-throughput sequencing (HTSeq) fragments per kilobase of transcript per million mapped reads (FPKM) data of hepatocellular carcinoma (HCC) tissues, including 374 tumor samples and 50 normal control samples, were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/), which is a free and available reference database for cancer research covering 33 cancer types, 20,000 primary cancer samples, and matched normal samples (Tomczak et al., 2015). The corresponding clinical information was obtained from the UCSC Xena database (http://xena.ucsc.edu/), which included the survival status, survival time (days), sex, age, histological grade, TNM stage, and progression-free survival (PFS) time of HCC patients. The mRNA expression data of HCC cell lines were obtained from the Cancer Cell Line Encyclopedia (CCLE) (https://portals.broadinstitute.org/ccle) (Barretina et al., 2012).

Differential Expression of Arp2/3 Subunits at mRNA Levels in Pan-Cancers and Hepatocellular Carcinoma

First, the mRNA differential expression of Arp2/3 members in pan-cancers and corresponding normal tissues was analyzed using the Oncomine 4.5 database (https://www.oncomine.org/), which is an online large data–mining platform and integrated oncogene microarray database covering 715 datasets and 86,733 samples (Rhodes et al., 2007). In this study, the cell color is determined by the best gene rank percentile for the analyses within the cell; Student's *t*-test was applied to calculate the *p* value for expression differences of Arp2/3 subunits between cancer and normal controls. The threshold parameters were set as follows: *p*-value < 0.0001, fold change = 1.5, and gene rank = 10%.

Thereafter, according to the RNA-Seq FPKM data of HCC downloaded from TCGA, the Arp2/3 subunits' differential expression in HCC tissues compared to that in normal tissues was analyzed using the "limma" package of R 4.0.4 software (http:///www.r-project.org/). After removing some samples, the remaining 371 tumor samples and 50 normal samples were included in differential expression analysis. The Wilcox test

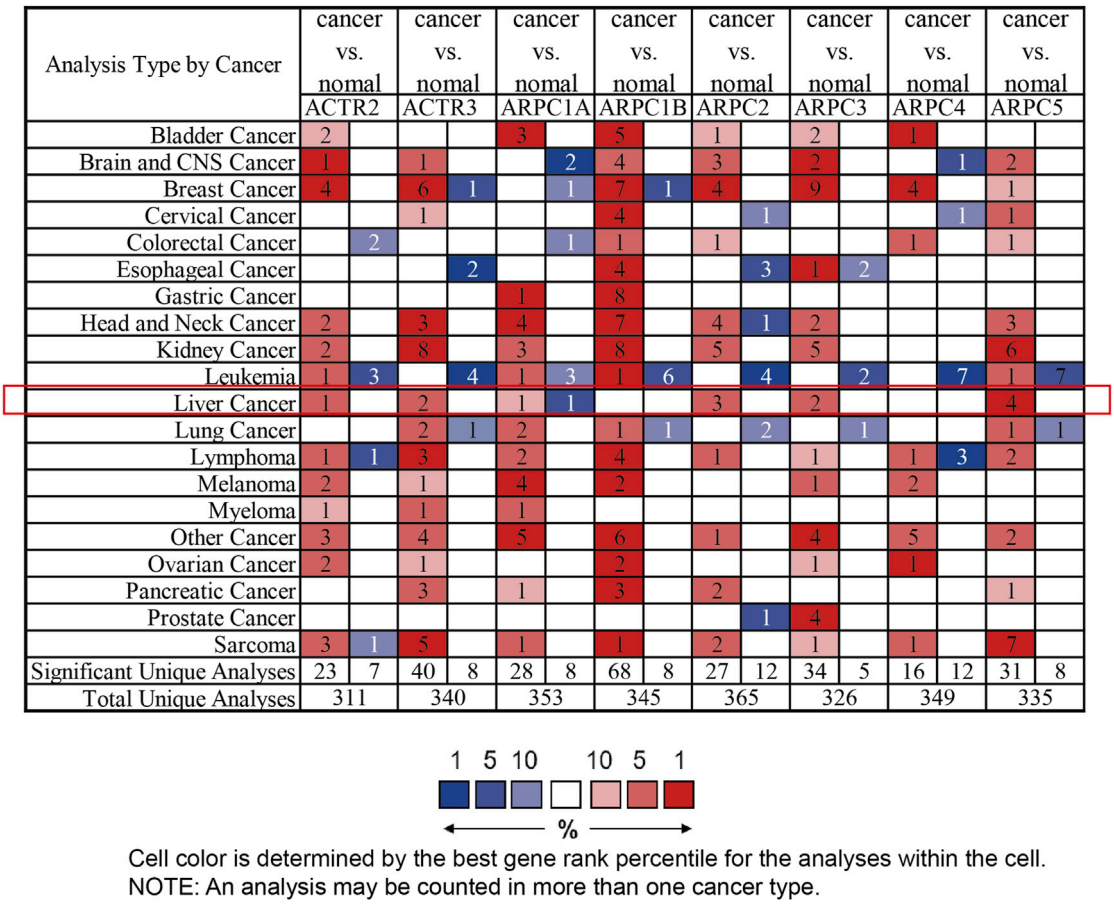


FIGURE 1 | mRNA expression level of Arp2/3 complex members in different types of cancer from multiple datasets (the Oncomine database). Difference of transcriptional expression was compared using Student's *t*-test and cutoff of parameters: *p*-value < 0.0001, fold change = 1.5, and gene rank = 10%. The red color in the cell represents overexpression, and green represents down-expression. The number in each cell presents the amount of datasets.

was applied to generate *p*-values, statistical significance was set at $p < 0.05$, and the results were visualized using heatmaps. Moreover, UALCAN (Chandrashekar et al., 2017) (http://ualcan. path.uab.edu) was selected for further verification of the facticity of the differential expression results above, Student's *t*-test was used to verify expression differences, and $p < 0.05$ indicates that the difference is statistically significant. Significantly differentially expressed subunits were selected for prognostic analysis.

Subsequently, we also explored whether the mRNA expressions of Arp2/3 subunits were correlated with each other in HCC tissues; Pearson's correlation was performed using the "corrplot" R package, the Pearson product–moment correlation coefficient (Pearson's R) represents the degree of correlation between the two subunits, and Pearson's $R > 0.4$ was considered as statistical correlation.

Besides, we downloaded the mRNA expression data of the Arp2/3 subunits in HCC cell lines from the Broad Institute Cancer Cell Line Encyclopedia (CCLE) (https://portals.broadinstitute.org/ccle), which provides public access to genomic data, analysis, and visualization of over 1,100 cell lines (Chen et al., 2019; Cheng et al., 2019). The ggplot2 package of R 4.0.4 software was used to explore Arp2/3 subunit expression levels in different HCC cell lines.

Differential Expression of Arp2/3 Subunits at the Protein Level in Hepatocellular Carcinoma Tissues

In addition to assessing the mRNA expression level of Arp2/3 members, protein expression analysis of Arp2/3 subunits was conducted using data from The Human Protein Atlas (HPA, https://www.proteinatlas.org/). The HPA provides typical immunohistochemistry profiling data for more than 8,000 patients and contains 26,941 antibodies targeting 17,165 unique proteins; it is also available for free download of nearly 20 common cancer types (Thul et al., 2017). In this study, the representative immunohistochemistry images of Arp2/3 members were downloaded directly from the HPA. We then compared the protein expression differences of Arp2/3 members in HCC and normal liver tissues.

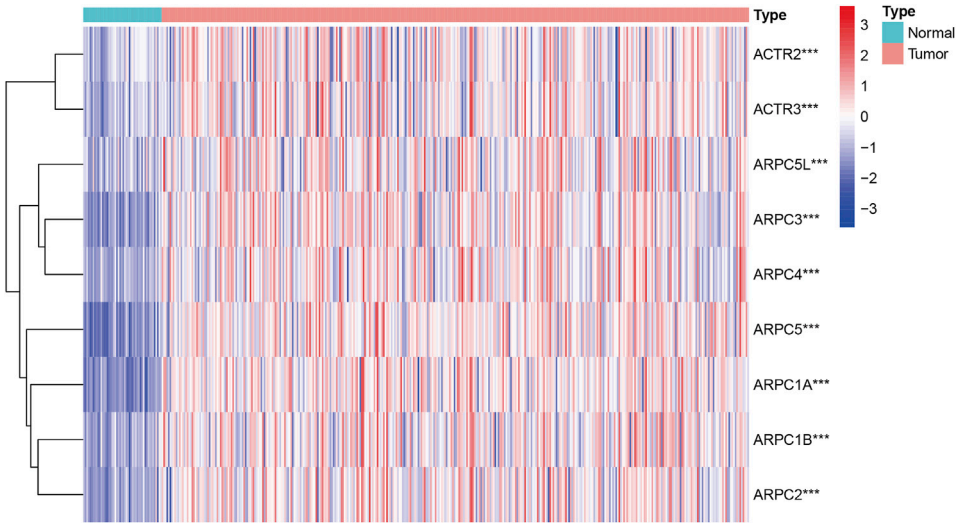


FIGURE 2 | Expression profile level of Arp2/3 complex members in HCC tissues visualized using a heatmap. Red stands for overexpression, and blue represents down-expression. ***stands for $p < 0.001$.

Prognostic Values of Arp2/3 Subunits in Hepatocellular Carcinoma

First, the mRAN expression value of Arp2/3 subunits and relevant clinical information of HCC were downloaded from the TCGA and UCSC Xena databases. After deleting some samples with incomplete follow-up information, we analyzed the remaining 370 HCC patients with complete overall survival time and 372 patients with progression-free survival information. Maximally selected rank statistics was used to determine an optimal cutoff value for each subunit, and the patients were divided into high- or low-expression groups according to the optimal cutoff value. Then Kaplan–Meier survival analysis and the log-rank test were performed to compare the overall survival (OS) or progression-free survival (PFS) difference between the two groups. Statistical significance was set at $p < 0.05$.

Furthermore, univariate and multivariate Cox regression analyses were conducted to explore whether the Arp2/3 subunits could be used as independent factors for the prognosis of HCC patients, integrating the following clinicopathological factors: age, sex, grade, and clinical stage. The results were presented with a hazard ratio (HR) and 95% confidence interval (CI), and statistical significance was set at $p < 0.05$. The subunits that significantly affected survival were chosen for further analyses.

Clinicopathological Analysis of Arp2/3 Subunits in Hepatocellular Carcinoma

UALCAN (http://ualcan.path.uab.edu) is a comprehensive interactive web server based on TCGA RNA-seq and clinical data (Chandrashekar et al., 2017). In this study, UALCAN was used to assess the association between the mRNA expression of Arp2/3 subunits and the clinicopathological parameters in HCC patients, including individual cancer stages and the nodal metastasis status. The results were obtained directly from the UALCAN website (TCGA data processed by unified standards). Student's *t*-test was used to verify expression differences, and statistical significance was set at $p < 0.05$.

Immune Infiltration Analysis of the Arp2/3 Family in Hepatocellular Carcinoma

Tumor cells and tumor-infiltrating immune cells (TIICs) interact closely in cancer progression; thus, we investigated the connections between TIICs and Arp2/3 subunits using the TIMER 2.0 platform (http://timer.comp-genomics.org/), which provides a systematic analysis of the specific gene(s) and immune cell infiltration in different cancers, including 32 types of cancers and 10,897 samples from TCGA (Li et al., 2020). In this study, the TIICs included CD4+ T cells, CD8+ T cells, B cells, neutrophils, and macrophages. We further explored the association between Arp2/3 subunits and biomarkers of subsets of TIICs using GEPIA databases (Tang et al., 2017) (http://gepia.cancer-pku.cn/). TIMER 2.0 and GEPIA correlation analysis employed the Spearman test and $p < 0.05$ was considered statistically significant, and the correlation strength was evaluated using Spearman's rank correlation Rho according to the previous studies, Rho 0.00–0.19 being "very weak," Rho 0.20–0.39 being "weak," Rho 0.40–0.59 being "moderate," Rho 0.60–0.79 being "strong," and Rho 0.80–1.0 being "very strong" (Lin et al., 2020).

Gene Set Enrichment Analysis (GSEA)

To explore the potential biological mechanism by which Arp2/3 subunits affect the carcinogenesis and progression of HCC, the transcriptome data of HCC from TCGA were selected for gene set enrichment analysis (GSEA) using GSEA 4.1.0 software. The c2. cp.kegg.v7.4. symbols.gmt downloaded from the Molecular Signatures Database (http://www.

gseamsigdb.org/gsea/msigdb/collections. jsp) was used as the reference. GSEA was executed using a random combination number of 1,000 permutations, and a false discovery rate (FDR q-value) <0.01 was used to identify the significantly enriched pathways.

Statistical Analysis

Statistical analyses were performed using R software (https://www.r-project.org/, version 4.0.4). The HTSeq FPKM mRNA data from the TCGA database were disposed using Perl 5.30.0 software (https://www.perl.org/). The "limma" R package and the Wilcox test were used to analyze the different expressions of Arp2/3 subunits in HCC. Besides, Student's *t*-test was used to verify expression differences in UALCAN. The "corrplot" R package was used for the correlation analysis of Arp2/3 members. Kaplan–Meier survival analysis and Cox proportional hazards regression analysis were conducted to assess the prognostic significance of Arp2/3 subunits. All statistical tests were two-sided, and statistical significance was set at $p < 0.05$.

RESULTS

The mRNA Expression of Arp2/3 Subunits in Pan-Cancers and Hepatocellular Carcinoma

First, the mRNA differential expression of Arp2/3 subunits in pan-cancers and the corresponding normal tissues was analyzed using the Oncomine database. As shown in **Figure 1**, overexpression of Arp2/3 subunits was observed in many kinds of cancers, including liver cancer. Based on the TCGA database, after collecting 371 HCC samples and 50 normal control samples, the mRNA differential expression of Arp2/3 subunits was obtained using the "limma" R package. We found that all of the Arp2/3 family members (ACTR2, ACTR3, ARPC1A, APRC1B, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L) were significantly upregulated in HCC tissues (**Figure 2**). The differential expression analyzed using the UALCAN databases also showed that Arp2/3 subunits were

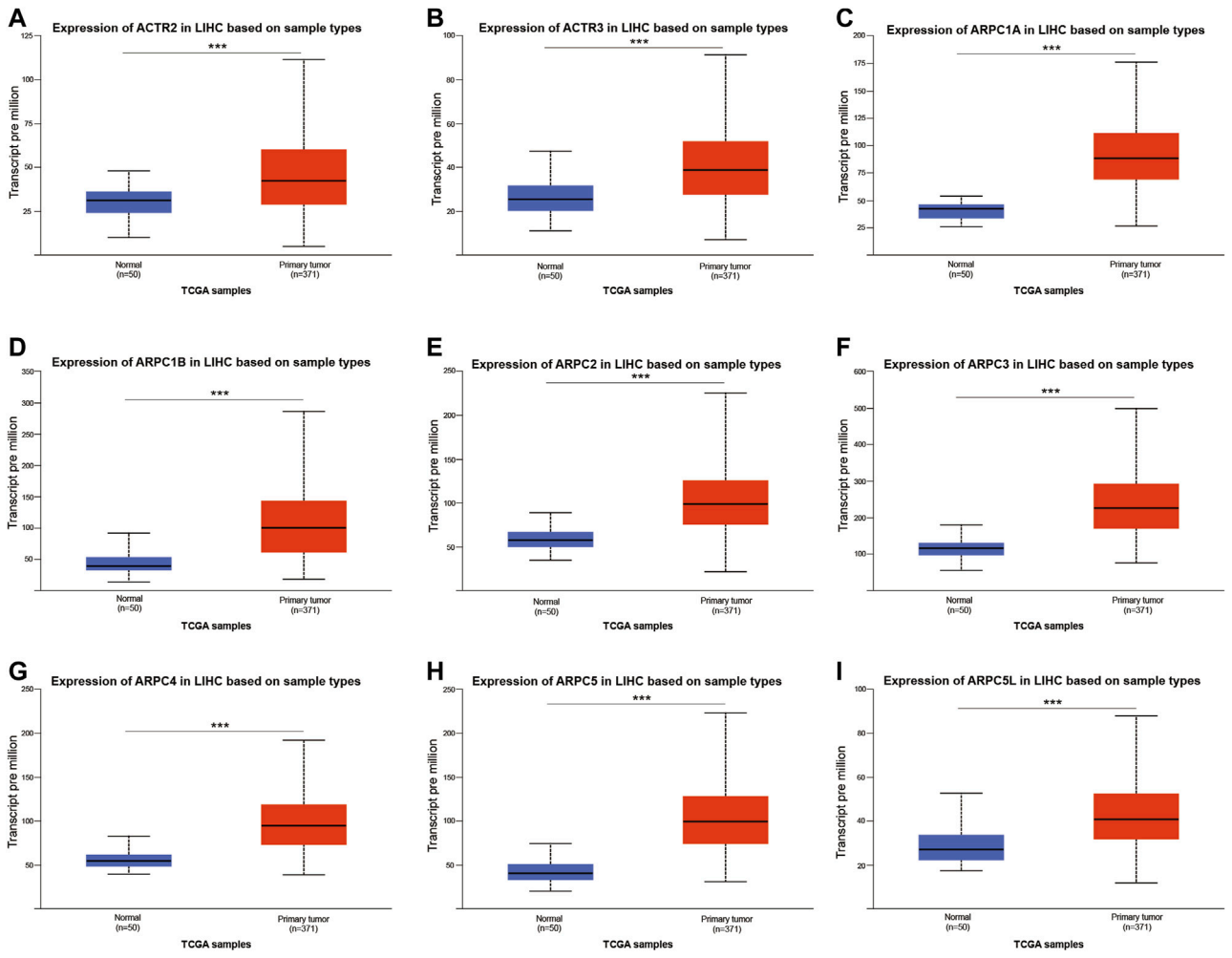


FIGURE 3 | Relative expression of Arp2/3 complex members in normal tissues and HCC tissues based on the UALCAN database **(A–I)**, ***stands for $p < 0.001$.

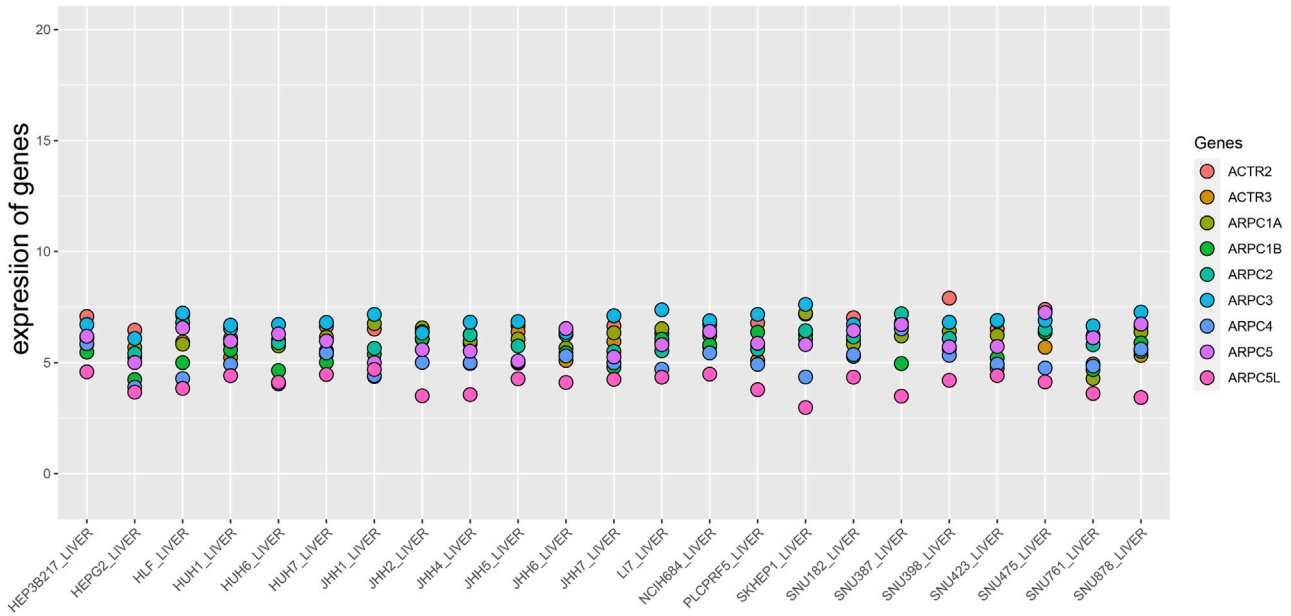


FIGURE 4 | Expression level of Arp2/3 complex members in HCC cell lines (the CCLE database). The abscissa represents different HCC cell lines, and the ordinate is the expression value of genes.

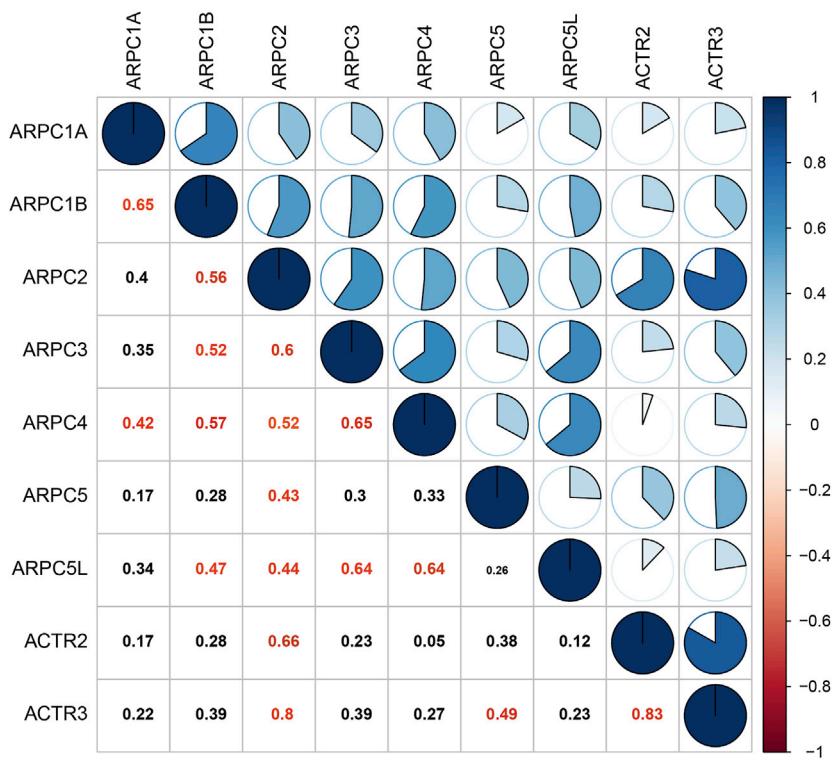


FIGURE 5 | Correlation analysis of each Arp2/3 complex member. The data were analyzed by Pearson's correlation, and Pearson's R cutoff value was 0.4. The red values represent significant correlation between the Arp2/3 members.

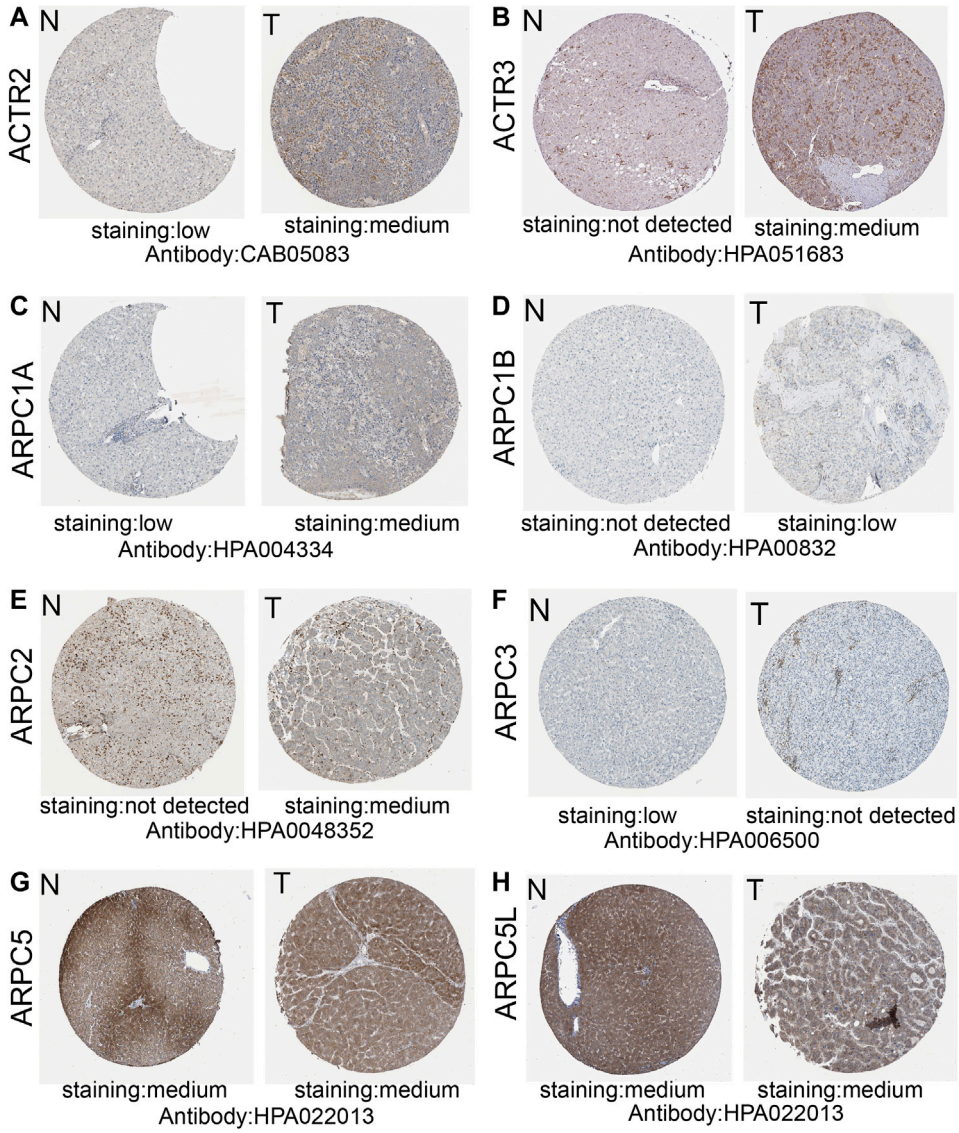


FIGURE 6 | Representative immunohistochemistry images of Arp2/3 members (except for ARPC4) of HCC tissues and normal liver tissues in the Human Protein Atlas **(A–H)**. T, HCC tissues; N, normal liver tissues.

upregulated in HCC tissues compared with normal control tissues, which was in accordance with the results above (**Figures 3A–I**). We used the CCLE databases to probe the mRNA expression of Arp2/3 subunits in HCC cell lines and found that Arp2/3 subunits were widely expressed in 23 HCC cell lines; among them, the expression levels of ACTR2 and ARPC3 were higher than those of other subunits, and the ARPC5L expression level was lowest in HCC cell lines (**Figure 4**).

Pearson's correlation was performed to explore whether the mRNA expressions of Arp2/3 subunits were correlated with each other. The results revealed that the mRNA expressions of Arp2/3 subunits were correlated to a significant degree in HCC tissues, such as ARPC1A and ARPC1B (Pearson's R = 0.65), ARPC2 and APRC3 (Pearson's R was 0.6), ARPC2 and ACTR (Pearson's R was 0.66), ARPC3 and ARPC4 (Pearson's R was 0.65), ARPC3 and APRC5L (Pearson's R was 0.64), ARPC4 and ARPC5L (Pearson's R was 0.64), and ARTC2 and ARTC3 (Pearson's R was 0.83), as shown in **Figure 5**.

Protein Expression of Arp2/3 Subunits in Hepatocellular Carcinoma Tissue

We obtained representative immunohistochemistry images from the Human Protein Atlas (HPA) to explore the protein expression conditions of Arp2/3 subunits in HCC. The results showed that higher protein expression of ACTR2, ACTR3, ARPC1A, ARPC1B, and ARPC2 was found in HCC tissues than in normal liver tissues, which have shown approximately the same results as the mRNA expression of Arp2/3 subunits. However, lower protein expression of ARPC3 was observed in

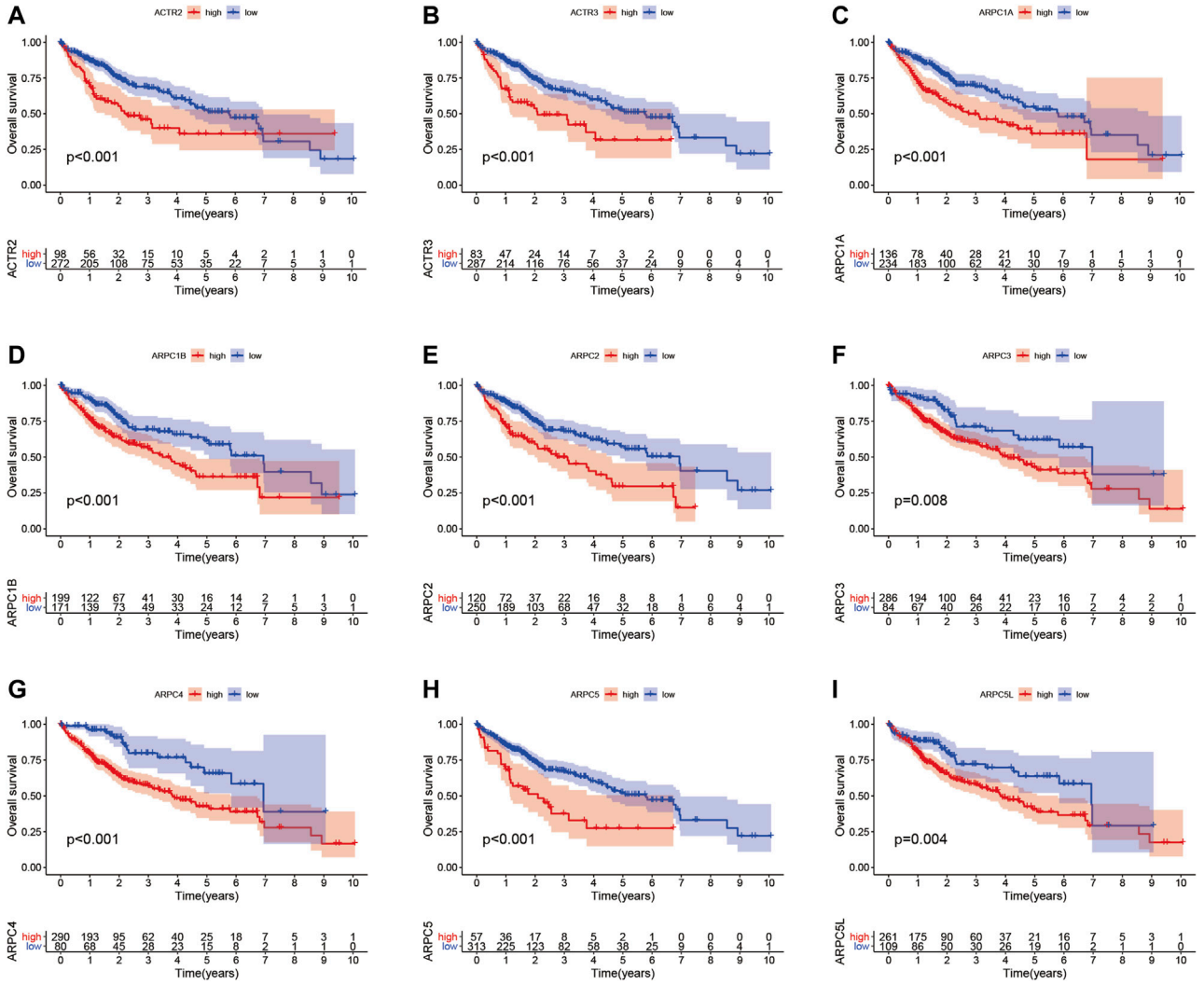


FIGURE 7 | Kaplan–Meier survival analysis of the association between mRNA expression of Arp2/3 complex members and overall survival (OS) in HCC patients **(A–I)**. The information of HCC samples was derived from the TCGA and UCSC Xena databases.

HCC tissues than in normal tissues. There were no significant differences in the expression levels of ARPC5 and ARPC5L. Currently, there is no immunohistochemical map for ARPC4 detection in the HPA. The results of Arp2/3 subunit immunohistochemistry are shown in **Figures 6A–H**.

Prognostic Values of Arp2/3 Subunits in Hepatocellular Carcinoma

To examine the prognostic values of Arp2/3 subunits in HCC patients, we performed Kaplan–Meier survival analysis and the log-rank test. The Kaplan–Meier survival curves for overall survival (OS) indicated that HCC patients with high expression of ACTR2 ($p < 0.001$), ACTR3 ($p < 0.001$), ARPC1A ($p < 0.001$), ARPC1B ($p < 0.001$), ARPC2 ($p < 0.001$), ARPC3 ($p = 0.008$), ARPC4 ($p < 0.001$), ARPC5 ($p < 0.001$), and ARPC5L ($p = 0.004$) had worse OS than those with low expression, as shown in **Figures 7A–I**. The Kaplan–Meier survival curves for PFS showed that patients with high expression of ACTR2 ($p < 0.001$), ACTR3 ($p < 0.001$), ARPC1A ($p = 0.0058$), ARPC2 ($p = 0.034$), ARPC3 ($p < 0.001$), ARPC4 ($p = 0.004$), ARPC5 ($p = 0.002$), and ARPC5L ($p = 0.005$) had shorter PFS than those with low expression (**Figures 8A–H**). These results indicate that Arp2/3 subunits lead to poor prognosis in HCC patients.

Univariate and multivariate Cox proportional hazards regression analyses were carried out to evaluate whether the Arp2/3 subunits could be independently associated with HCC survival. The results of univariate Cox regression analysis showed that the Arp2/3 subunits (ACTR2, ACTR3, ARPC1A, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L) and the clinical stage were associated with poor survival outcomes in HCC patients. The hazard ratio (HR), 95% confidence interval (CI), and p values are shown in **Table 1**. Multivariate Cox proportional hazards regression analysis revealed that the expressions of ACTR3 (HR =

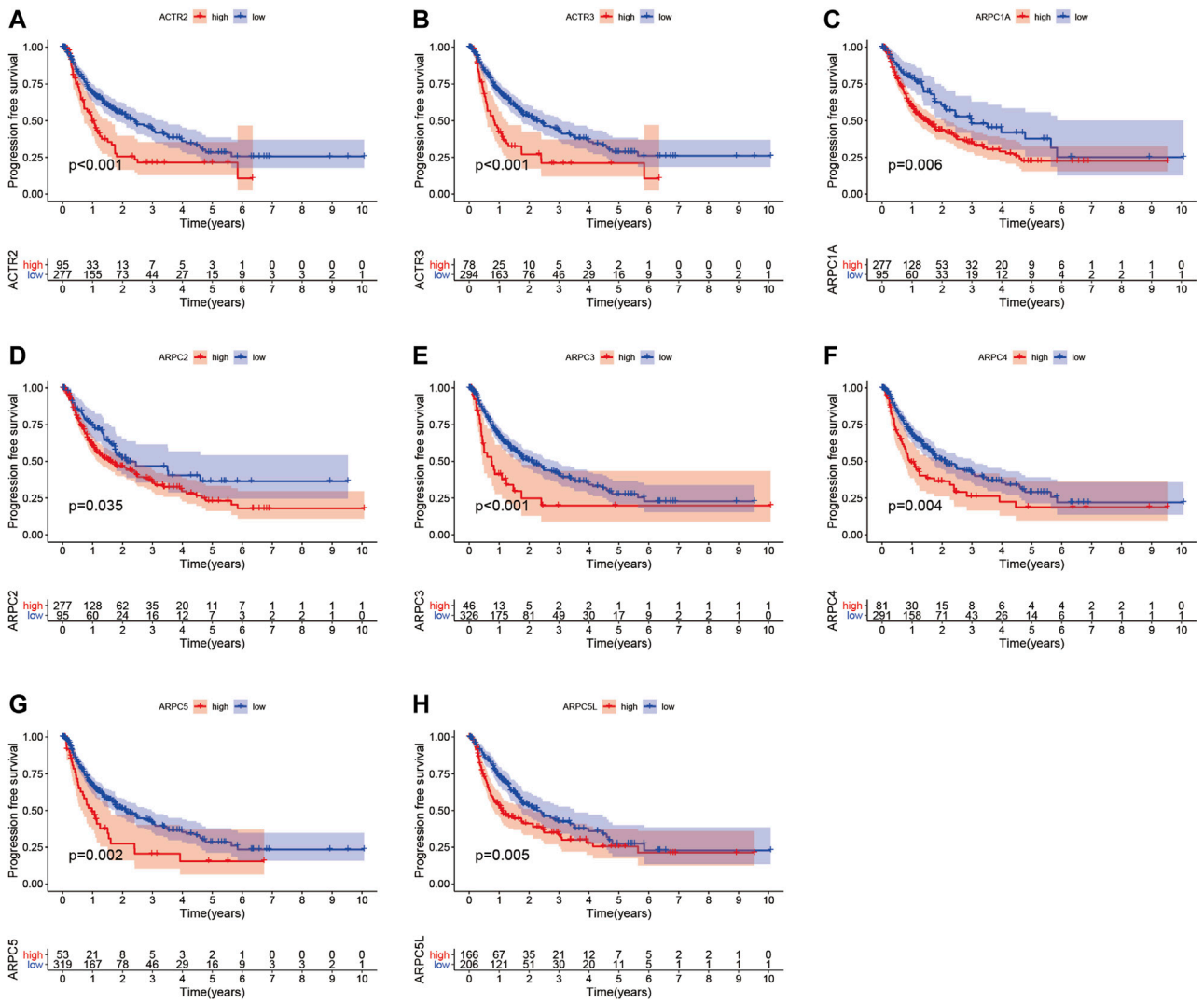


FIGURE 8 | Kaplan–Meier survival analysis of the association between mRNA expression of Arp2/3 complex members and progression-free survival (PFS) in HCC patients **(A–H)**. The information of HCC samples was derived from the TCGA and UCSC Xena databases.

TABLE 1 | Univariate Cox proportional hazards regression analyses of Arp2/3 members and clinical features in HCC.

Parameter	Univariate analysis		
	Hazard ratio	95% CI	*p* value
ARPC1A	1.011946	1.004–1.020	**0.002**
ARPC1B	1.004998	1.000–1.010	**0.053**
ARPC2	1.035206	1.016–1.055	**0.3E-03**
ARPC3	1.009771	1.002–1.017	**0.011**
ARPC4	1.014924	1.005–1.025	**0.004**
ARPC5	1.040619	1.017–1.064	**0.58E-03**
ARPC5L	1.048752	1.013–1.085	**0.006**
ACTR2	1.014755	1.004–1.026	**0.010**
ACTR3	1.064911	1.025–1.106	**0.001**
Age	1.010115	0.996–1.025	0.173
Gender	1.28922	0.883–1.882	0.188
Grade	1.133,154	0.881–1.457	0.330
Stage	1.679,735	1.369–2.062	**6.97E-07**

The bold values represent the factors that significantly associated with poor survival in HCC patients

1.0, 95%CI: 1.01–1.1, p = 0.002), ARPC2 (HR = 1.0, 95%CI: 1.00–1.0, p = 0.016), ARPC5 (HR = 1.0, 95%CI: 1.01–1.2, p = 0.002), and the clinical stage (HR = 1.6, 95%CI: 1.31–2.0, p < 0.001) were independent prognostic biomarkers of HCC survival, as shown in the forest plots in **Figures 9A–C**. Those results indicated that ACTR3, ARPC2, and ARPC5 are independently related to the prognosis of HCC patients and can be used as useful biomarkers to predict patients' survival rate.

Correlation Analysis of Arp2/3 Subunits and Clinicopathological Features in Hepatocellular Carcinoma

We used the UALCAN database to explore the relationship between the mRNA expression of Arp2/3 members and the clinicopathological parameters of HCC patients. As shown in **Figures 10A–I**, the results showed a definite association between the mRNA expression of Arp2/3 subunits (ACTR2, ACTR3,

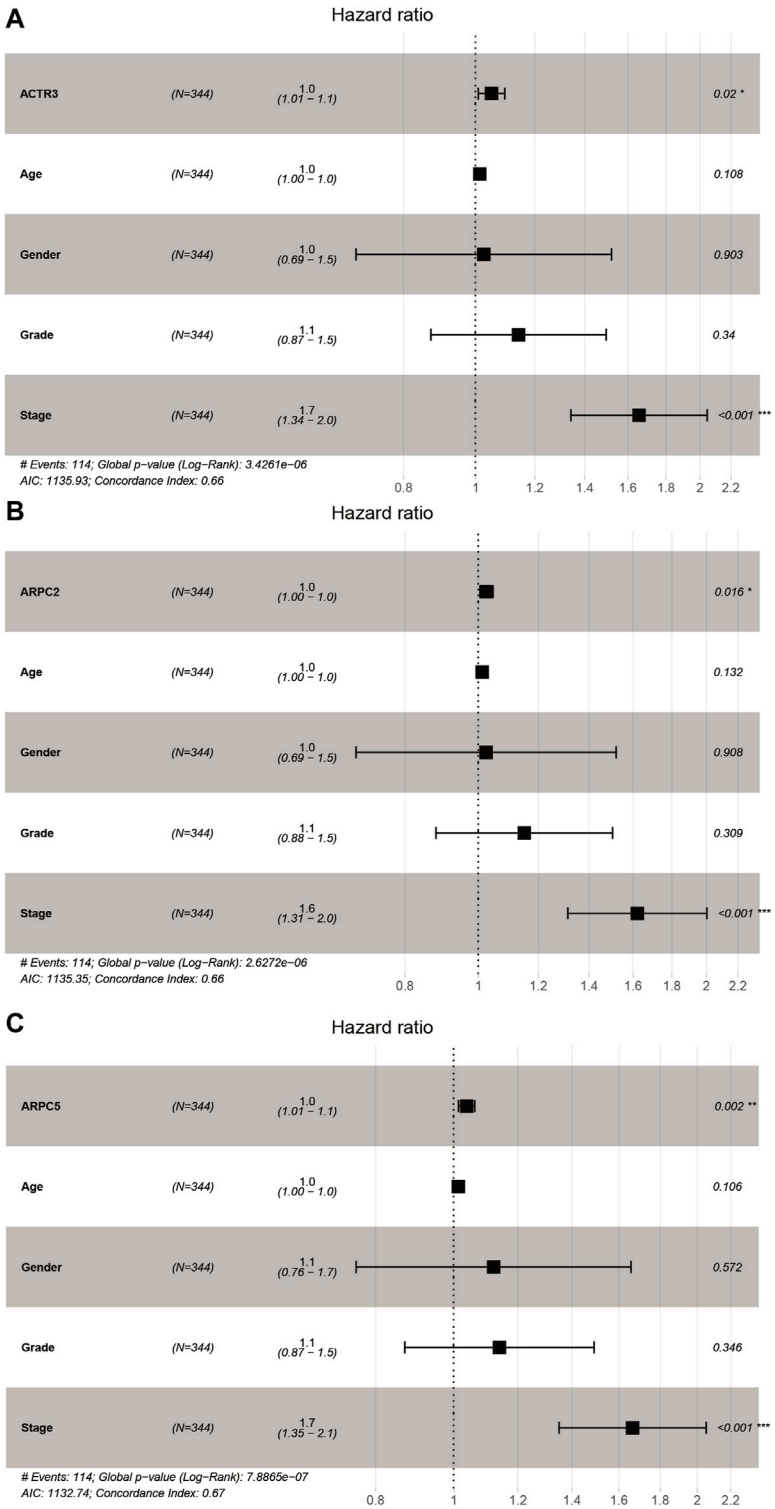


FIGURE 9 | Forest plots of multivariate Cox regression analysis of Arp2/3 members with significant prognostic significance. **(A)** ACTR3; **(B)** ARPC2; and **(C)** ARPC5. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

ARPC1A, ARPC1B, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L) and cancer stages; patients with more advanced cancer stages tended toward higher mRNA expression of Arp2/3 subunits. The mRNA expression of Arp2/3 subunits (ACTR3, ARPC1A, ARPC1B, ARPC2, ARPC3, ARPC4, and ARPC5) was remarkably higher in HCC patients than in

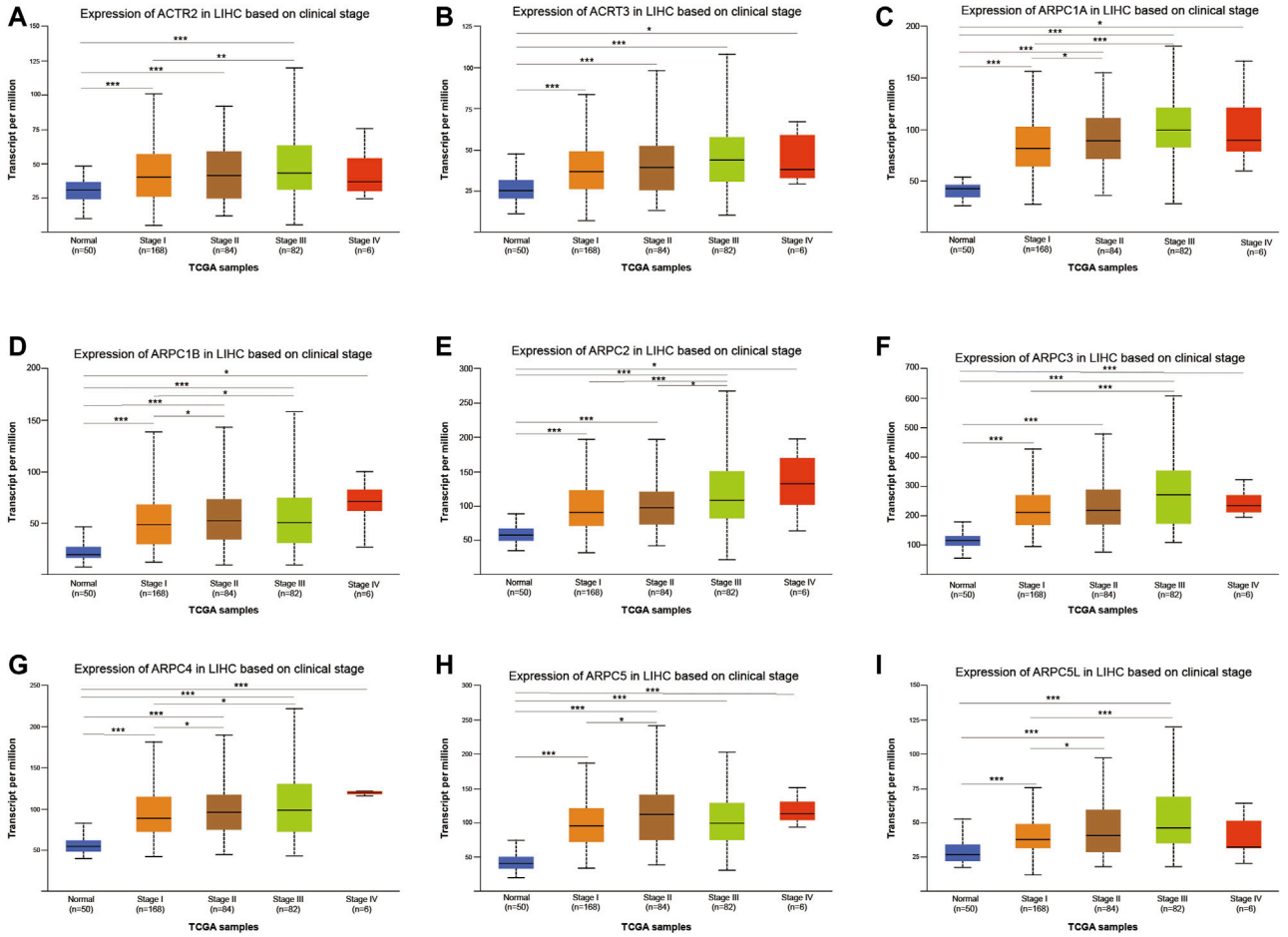


FIGURE 10 | Analysis of association between mRNA expression of Arp2/3 members and cancer stages of HCC patients **(A–I)**. The mRNA expression of Arp2/3 members in normal individuals or in HCC patients of stages I, II, III, or IV. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

normal tissues. In contrast, there was no marked difference between stage IV and normal tissues in the mRNA expression of ACTR2, ACTR3, and ARPC5L, which may be due to the small sample size in stage IV (only six samples).

We further investigated the relationship between the mRNA expression of Arp2/3 subunits and the nodal metastasis status in HCC patients. The results showed no significant relationship between the mRNA expression of Arp2/3 subunits and the nodal metastasis status (**Figures 11A–I**). This may be due to the small number of patients with lymph node metastasis in the TCGA database (n = 4). Nevertheless, the mRNA expression of Arp2/3 subunits (ACTR2, ACTR3, ARPC1A, ARPC1B, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L) was remarkably higher in N0 patients than in those in the normal group.

Association of Prognostic Arp2/3 Subunits With Immune Infiltration Level in Hepatocellular Carcinoma

Tumor-infiltrating immune cells (TIICs) in the tumor microenvironment (TME) play crucial roles in the tumorigenesis, progression, metastasis, and treatment resistance of tumors. To investigate the correlations between Arp2/3 subunits and TIICs, we first explored the associations between the independent prognostic biomarkers (ACTR3, ARPC2, and ARPC5) and the immune cells using the TIMER 2.0 platform. The results showed that the expression of ARPC2 was negatively correlated with tumor purity (Rho = -0.169, p = 1.6e-03), whereas the expression of ACTR3 and APRC5 was irrelevant to tumor purity. The expression levels of ACTR3, ARPC2, and ARPC5 were positively correlated with the immune infiltration of CD4+ T cells, CD8+ T cells, B cells, neutrophils, and macrophages, as shown in **Figures 12A–C**.

In addition, to investigate which types of TIIC subsets were correlated with prognostic Arp2/3 subunits, we analyzed the co-expression relationship of prognostic subunits and typical biomarkers of TIICs using the GEPIA database. As shown in **Table 2**, the higher expression of ACTR3 and ARPC2 was positively correlated with the expression of biomarkers of TIIC subsets, including B cells, CD8+ T cells, Th1 cells, Th2 cells, Th17 cells, Treg cells, neutrophils, M1 macrophages, and M2

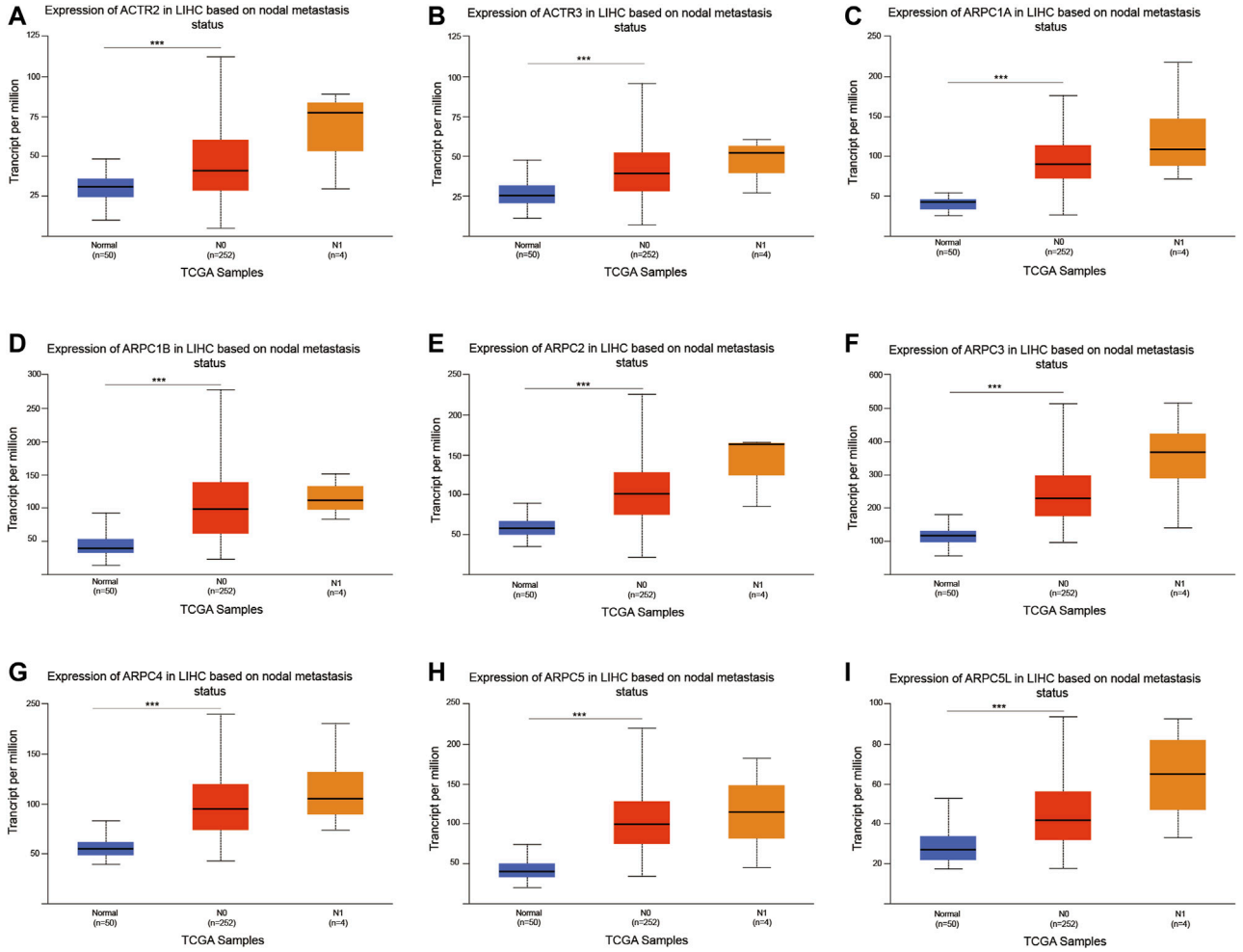


FIGURE 11 | Analysis of association between mRNA expression of Arp2/3 members and the nodal metastasis status of HCC patients **(A–I)**. The mRNA expression of Arp2/3 members in normal individuals or in HCC patients of nodal metastasis status N0 or N1. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

macrophages. ACTR3 expression in HCC was significantly correlated with STAT1 (Th1), STAT6 (Th2), STAT3 (Th17), CCR8 (Treg), CDb11(neutrophils), PTGS2, and IRF5 (M1 macrophages). There was also a significant positive correlation between the expression of ARPC2 and TIIC biomarkers, including STAT1 and TNF (Th1), GATA3 and STAT5A (Th2), CCR8 and TGFB1(Treg), CDb11(neutrophils), PTGS2 and IRF5(M1 macrophages), and VSIG4 and MS4A4A (M2 macrophages). Higher expression of ARPC5 was positively correlated with most biomarkers of TIIC subsets. Some of these genes were moderately correlated with the expression of ARPC5, including STAT1(Th1), STAT6, STAT5A (Th2), STAT3(Th17), CDb11(neutrophils), and IRF5(M1 macrophages). In addition, the correlation between ARPC2, ARPC5, and ACTR3 and the other biomarkers of TIIC subsets was weak or irrelevant. The above results indicate that Arp2/3 genes might positively modulate the infiltration and activation of TIICs in HCC.

Potential Action Mechanism of Prognostic Arp2/3 Subunits in Hepatocellular Carcinoma Carcinogenesis

We identified ACTR3, ARPC2, and ARPC5 as independent prognostic biomarkers affecting the survival of HCC, and GSEA analysis was conducted to explore the potential biological mechanism by which Arp2/3 subunits lead to poor survival. According to the GSEA results, high expression of ACTR3 was positively related to 82 gene sets at FDR <0.01, the functions of which focused on regulation of the actin cytoskeleton, protein ubiquitination, the immune system process, genesis and progression of various tumors, leukocyte migration, and the DNA metabolic process; the ACTR3 overexpression was closely relevant to the "JAK-STAT signaling pathway," the "WNT signaling pathway," the "pathway in cancer," the "VEGF signaling pathway," "non–small-cell lung cancer," "pancreatic cancer," and "renal cell carcinoma," as shown in **Figure 13A**. The GSEA results also

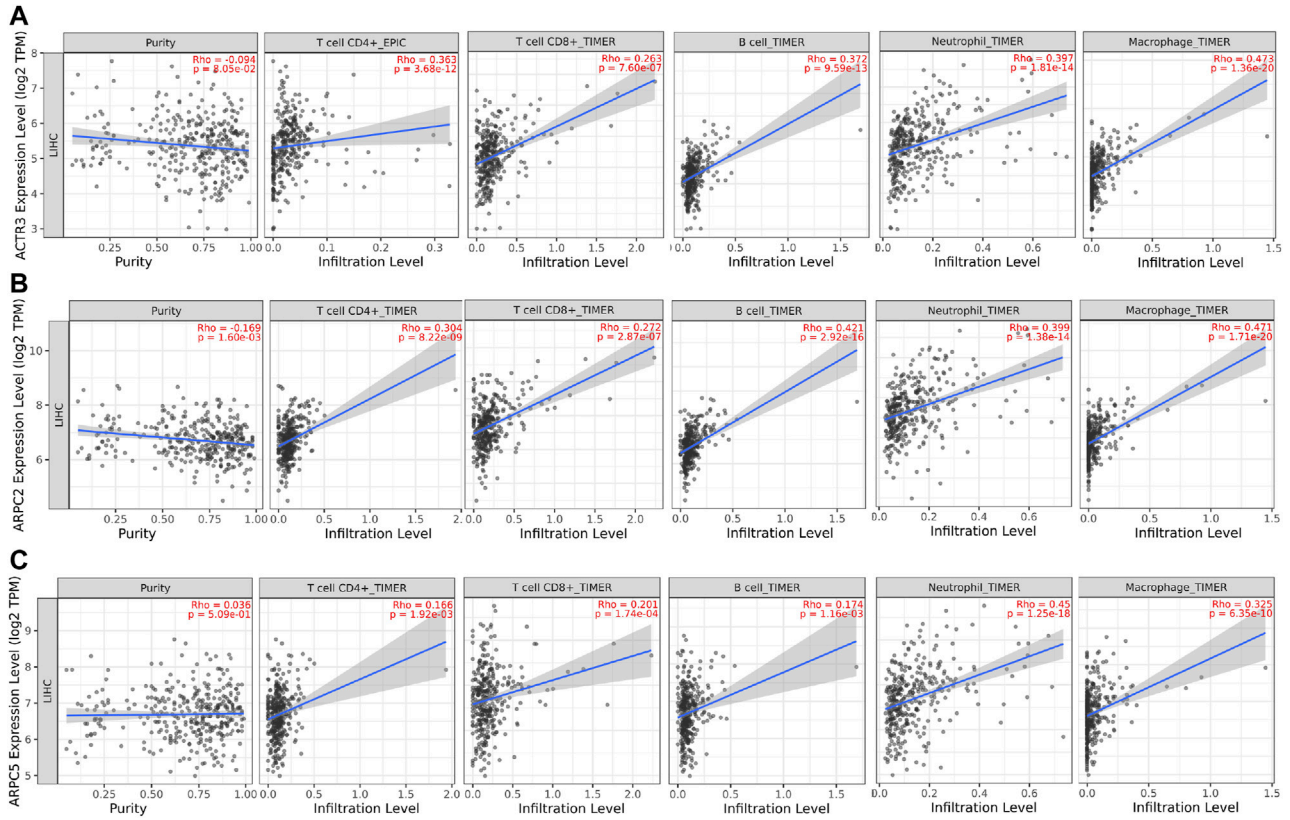


FIGURE 12 | Correlation analysis between tumor-infiltrating immune cells (TIICs: CD4+ T cells, $CD8^+$ T cells, B cells, neutrophils, and macrophages) and significant prognostic Arp2/3 members in HCC. **(A)** ACTR3; **(B)** ARPC2; and **(C)** ARPC5. Tumor purity is shown in the panels on the left.

indicated that high expression of ARPC2 was significantly positively related to 59 gene sets at FDR <0.01. Among them, the cancer-related pathways included the "WNT signaling pathway," the "cell cycle," the "pathway in cancer," "bladder cancer," "colorectal cancer," the "VEGF signaling pathway," the "MAPK signaling pathway," and the "chemokine signaling pathway." Besides, the "T cell receptor signaling pathway" and "leukocyte transendothelial migration" might be associated with immune cell infiltration (**Figure 13B**). Fourteen gene sets were significantly negatively related to the expression of ARPC2 at FDR <0.01, the functions of which focused on fatty metabolism, amino acid metabolism, and metabolism of xenobiotics by cytochrome P450 (**Figure 13B**). High expression of ARPC5 was significantly positively related to 43 gene sets at FDR <0.01, and the following pathways might be involved in tumor development and pathogenesis: the "MAPK signaling pathway," "non–small-cell lung cancer," "small-cell lung cancer," "pancreatic cancer," the "WNT signaling pathway," and the "toll-like receptor signaling pathway" (**Figure 13C**). Five gene sets were significantly negatively related to the expression of ARPC2 at FDR <0.01, including fatty metabolism and amino acid metabolism (**Figure 13C**).

DISCUSSION

The actin-related protein 2/3 complex (Arp2/3) was first isolated from *Acanthamoeba* as an affinity complex for intracellular profibrin. It plays an important role in the formation of microfilaments and is related to cell movement. Recently, increasing numbers of studies have shown that Arp2/3 subunits are upregulated in various cancer tissues or cells involved in the proliferation, invasion, and metastasis of cancer. It was reported that the expression of Arp2 was significantly higher in cases with a high histologic grade and lymph node metastasis in adenocarcinomas of the lung and breast carcinoma (Semba et al., 2006; Iwaya et al., 2007). Eeva et al. (Laurila et al., 2009) found that ARPC1A acts as a novel regulator of cell migration and invasion in pancreatic cancer and has been suggested as a potential target for cancer anti-metastasis therapy. Zhang et al. (Zhang et al., 2017a) verified that ARPC2 expression was higher in gastric cancer tissues than in normal tissues and promoted gastric cancer cell proliferation and metastasis. It was reported that ARPC2 inhibitors, such as benproperine and pimozide, inhibited tumor invasion and metastasis of cancer cells in animal models (Choi et al., 2019). Similarly, significant overexpression of ARPC4 has also been observed in pancreatic carcinoma and gastric carcinoma, indicating a close association

TABLE 2 | Correlations between significant prognostic Arp2/3 subunits' expression and biomarker expression of subsets of TIICs in HCC.

Types of TIICs	Gene markers	ARPC2		ARPC5		ACTR3	
		R	*P*	R	*P*	R	*P*
B cell	CD19	0.36	1.2E-12	0.18	5.3E-04	0.17	0.0012
	CD79A	0.34	3.6E-11	0.088	0.092	0.17	0.0012
CD8	CD8A	0.39	9.7E-15	0.031	0.55	0.23	6.3E-06
	CD8B	0.36	6.8E-13	0.11	0.038	0.12	0.027
Th1	TBX21	0.3	4.3E-09	0.11	0.032	0.19	1.7E-04
	STAT4	0.39	6.1E-15	0.19	2.9E-04	0.26	6.6E-07
	STAT1	**0.6**	**0.3E-37**	**0.46**	**5.3E-21**	**0.59**	**3.9E-36**
	TNF	**0.48**	**7E-23**	0.23	7.1E-06	0.37	1.3E-13
	IFNG	0.34	2.7E-11	0.13	0.012	0.17	0.0014
Th2	GATA3	**0.48**	**1.4E-22**	0.19	2.3E-04	0.38	2.6E-14
	STAT6	0.29	8.6E-9	**0.48**	**2.2E-22**	**0.57**	**1.6E-33**
	IL-13	0.11	0.043	0.055	0.29	0.12	0.019
	STAT5A	**0.54**	**7.4E-30**	**0.46**	**2.6E-20**	**0.48**	**3.2E-22**
Th17	STAT3	0.37	1.7E-13	**0.48**	**4.2E-23**	**0.63**	**2.1E-42**
	IL-17A	0.052	0.32	0.089	0.089	0.14	0.007
Treg	FOXP3	0.2	1E-04	0.12	0.018	0.26	4.8E-07
	CCR8	**0.55**	**6E-30**	0.36	5.1E-13	**0.56**	**2.9E-31**
	TGFB1	**0.56**	**1.2E-31**	0.23	6.4E-06	0.29	2.2E-08
Neutrophils	CD11b	**0.52**	**2.6E-27**	**0.47**	**1.8E-21**	**0.48**	**1.1e-22**
	CCR7	0.36	5E-13	0.18	5.4E-04	0.27	1.7e-07
	CD66b	0.13	0.014	–0.0051	0.92	0.1	0.054
M1 macrophages	NOS2	0.13	0.011	0.32	3.1E-10	0.35	2.4E-12
	PTGS2	**0.47**	**6.4E-22**	0.3	4.8E-09	**0.47**	**1E-21**
	IRF5	**0.44**	**2.9E-19**	**0.51**	**5.9E-26**	**0.46**	**1.8E-20**
M2 macrophages	CD163	0.33	5.7E-11	0.23	8E-06	0.15	0.003
	VSIG4	**0.47**	**3E-21**	0.27	1.7E-7	0.33	4.2E-11
	MS4A4A	**0.46**	**1.2E-20**	0.25	1.9E-6	0.36	4.6E-13

R, Spearman's rank correlation Rho; P, p *value.The bold values represent the correlation strength was above "moderate" between Arp2/3 subunits and biomarkers of subsets of TIICs.*

between APRC4 expression and tumor migration and invasion (Rauhala et al., 2013), (Kang et al., 2016). Furthermore, Xu et al. (Xu et al., 2020) have shown that ARPC4 is necessary for proliferation, migration, invasion, and pseudopodia formation in bladder cancer cells, suggesting that ARPC4 is a potential prognostic biomarker in these diseases. ARPC5 may function as an oncogene in the development of lung squamous cell carcinoma (lung SCC) and head and neck squamous cell carcinoma (HNSCC) and contributes to cancer cell migration and invasion, which is directly regulated by miRNA (Moriya et al., 2012), (Kinoshita et al., 2012). Arp2/3 complex silencing mediated by siRNA led to a reduction in the migration of pancreatic cells (Rauhala et al., 2013). In this study, we identified that ARP2/3 members were significantly overexpressed in various cancers, including HCC. Differential expression of ACTR2, ACTR3, ARPC1A, ARPC1B, and ARPC2, both at the mRNA and protein levels, was observed in patients with HCC. Moreover, there was a significant correlation between the expressions of each subunit, suggesting that ARP2/3 subunits may serve as potential biomarkers for HCC.

Furthermore, we evaluated the prognostic roles of Arp2/3 subunits in HCC using Kaplan–Meier survival analysis and Cox proportional hazards regression. We found that higher expression of ARP 2/3 subunits was associated with worse OS, and overexpression of ACTR2, ACTR3, ARPC1A, ARPC2, ARPC3, ARPC4, ARPC5, and ARPC5L was related to shorter PFS in HCC. We also identified that the expression of ACTR3, ARPC2, and ARPC5 was associated with poor survival of HCC patients as independent prognostic biomarkers. In addition, we found a specific correlation between the mRNA expression of the nine Arp2/3 members and the cancer stages of HCC. A previous study reported that ARPC2 was closely associated with the stage, nodal metastasis, and overall survival in breast cancer and that the TGF-β/EMT pathway is involved in ARPC2-mediated carcinogenesis (Cheng et al., 2019). Furthermore, ARPC2 was significantly associated with large tumor size, lymph node invasion, and high tumor stage *via* association analysis of 110 gastric cancer tissues, and ARPC2-positive patients exhibited lower RFS and OS rates than ARPC2-negative patients with gastric cancer (Zhang et al., 2017a). In addition, the ARPC5 high-expression group was associated with poor overall survival compared to that in the ARPC5 low-expression group, and multivariable analysis indicated that ARPC5 was an independent prognostic factor in patients with multiple myeloma (MM) (Xiong and Luo, 2018). Therefore, our results are consistent with those of previous studies on other tumors.

Cancer immunotherapy has caused significant breakthroughs in various malignancies. However, only a minority of patients with HCC respond to immunotherapy (Zhu et al., 2018). This is primarily due to the high heterogeneity of tumors, various

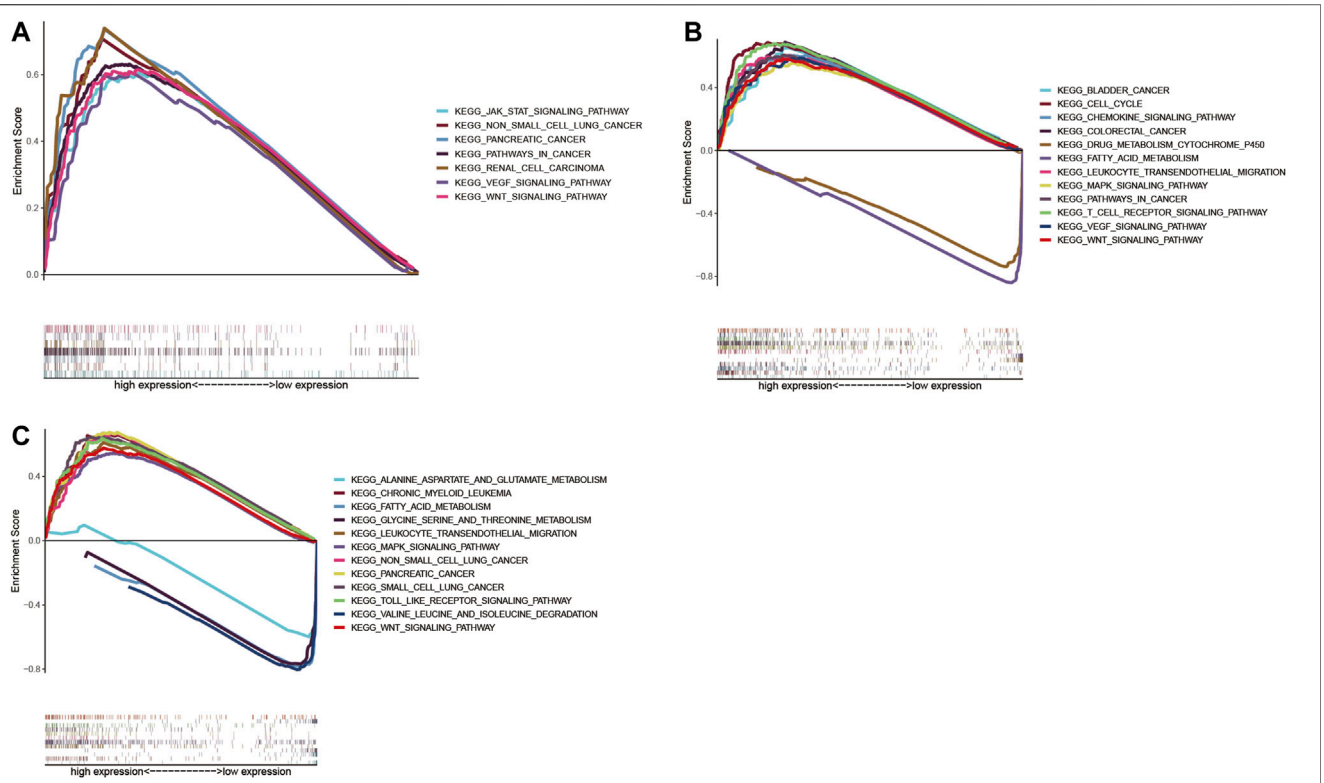


FIGURE 13 | Gene set enrichment analysis (GSEA) of significant prognostic Arp2/3 subunits in HCC based on the cancer-related Kyoto Encyclopedia of Genes and Genomes (KEGG). **(A)** ACTR3; **(B)** ARPC2; and **(C)** ARPC5.

immune microenvironments, lack of immune cell infiltration, and absence of predictive markers. Therefore, this study explored the association between the expression of prognostic genes and immune cell infiltration. The Arp2/3 complex is critical for chemotaxis and hagocytosis and is required for macrophage integrin effects and monocyte recruitment functions (Rotty et al., 2017). There is a strong relationship between the Arp2/3 complex and immune cells; for example, leukocytes need to adhere to cells to form synapses that kill infected cells, which literally squeeze their cell body during blood extravasation and efficiently migrate to the inflammatory focus (Tur-Gracia and Martinez-Quiles, 2021). Moreover, the cytoskeleton is crucially important for adhesive contact and migration in the development process of immune cells (Tur-Gracia and Martinez-Quiles, 2021). Since the Arp2/3 subunits assemble into a complex, the abnormity of one subunit is likely to affect the whole complex function. ARPC2 knockout mice have been reported to cause a dramatic decrease in peripheral T cell numbers and impaired T cell homeostasis, which was caused by a reduction in the surface TCR levels of T cells. There is a higher transcription level of ARPC2 in peripheral T cells than in thymocytes, and Arp2/3 complex–promoted actin nucleation is essential for peripheral T cell homeostasis (Zhang et al., 2017b). This study found that the expression of ACTR3, ARPC2, and ARPC5 was positively related to the immune infiltration of CD4+ T cells, CD8+ T cells, B cells, neutrophils, and macrophages in the HCC microenvironment. In addition, we discovered that ACTR3, ARPC2, and ARPC5 expression was significantly correlated with the biomarkers of $CD4^+$ T cells, CD8+ T cells, neutrophils, M1 macrophages, and M2 macrophages. The above results indicate that Arp2/3 subunits participate in the activation and recruitment of TIICs in HCC and play a dual role in tumor immunity, which promotes antitumor immune cell infiltration and recruits immunosuppressive cells. Thus, further research is needed to help us understand the role of Arp2/3 subunits and tumor-related immune cell functions and, consequently, contribute to the application of immunotherapy.

Currently, few studies have been conducted on the specific mechanism by which ARP2/3 members promote tumor development and metastasis. Zhang et al. (Zhang et al., 2017a) found that oncogenic genes, including CTNND1, EZH2, BCL2L2, CDH2, VIM, and EGFR, were upregulated by ARPC2, and tumor suppressor genes PTEN, BAK, and CDH1 were downregulated by ARPC2. In breast cancer, ARPC2 expression significantly upregulated the expression of vimentin, N-cadherin, MMP-9, ZEB1, and MMP-3, activated the TGF-β pathway, and eventually led to epithelial–mesenchymal transition (EMT) (Cheng et al., 2019). Using gene set enrichment analysis (GSEA), this study found that Arp2/3 subunits mainly participate in regulating various cancer pathways, including colorectal cancer, pancreatic cancer, bladder cancer, lung cancer, renal cell carcinoma, and the VEGF, MAPK, and Wnt signaling pathways. In addition, leukocyte transendothelial migration

and T cell receptor signaling pathways may increase immune cell infiltration. A validation study regarding the exact mechanism of ACTR2, ARPC2, and ARPC5 should be performed to confirm the above results further.

Although we discussed the important role of Arp2/3 complex members in HCC, the limitations of this study should be noted. First, the Arp2/3 subunits were analyzed and evaluated using limited data and clinical information from the genomic commons data. Moreover, protein differential expression of Arp2/3 subunits was presented with immunohistochemistry images from different samples, which may affect the results to some extent. It is preferable to collect a specific number of clinical samples and compare subunit expression levels in tumor tissues and adjacent noncancerous tissues. In addition, it is critical to verify the functional features and molecular mechanisms of Arp2/3 subunits using biological experiments and clinical research.

CONCLUSION

This study systematically analyzed the expression profile and prognostic values of Arp2/3 complex members in HCC and found that ACTR3, ARPC2, and ARPC5 could be used as independent predictors of survival and might be applied as promising molecular targets for diagnosis and therapy of HCC 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

JW conceived and designed the study. SH and DL were responsible for the collection and analysis of the research information. SH was involved in drafting the manuscript. LZ, LS, and JW critically and carefully revised the manuscript. The authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 82060435).

ACKNOWLEDGMENTS

We acknowledge and appreciate our colleagues for their valuable efforts and comments on this manuscript.

REFERENCES

Abella, J. V. G., Galloni, C., Pernier, J., Barry, D. J., Kjær, S., Carlier, M.-F., et al. (2016). Isoform Diversity in the Arp2/3 Complex Determines Actin Filament Dynamics. *Nat. Cel Biol* 18 (1), 76–86. doi:10.1038/ncb3286

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

Asia-Pacific Working Party on Prevention of Hepatocellular C (2010). Prevention of Hepatocellular Carcinoma in the Asia-Pacific Region: Consensus Statements. *J. Gastroenterol. Hepatol.* 25 (4), 657–663. doi:10.1111/j.1440-1746.2009.06167.x

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity. *Nature* 483 (7391), 603–607. doi:10.1038/nature11003

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, P., Yue, X., Xiong, H., Lu, X., and Ji, Z. (2019). RBM3 Upregulates ARPC2 by Binding the 3'UTR and Contributes to Breast Cancer Progression. *Int. J. Oncol.* 54 (4), 1387–1397. doi:10.3892/ijo.2019.4698

Cheng, Z., Wei, W., Wu, Z., Wang, J., Ding, X., Sheng, Y., et al. (2019). ARPC2 Promotes Breast Cancer Proliferation and Metastasis. *Oncol. Rep.* 41 (6), 3189–3200. doi:10.3892/or.2019.7113

Choi, J., Lee, Y. J., Yoon, Y. J., Kim, C. H., Park, S. J., Kim, S. Y., et al. (2019). Pimozide Suppresses Cancer Cell Migration and Tumor Metastasis through Binding to ARPC2, a Subunit of the Arp2/3 Complex. *Cancer Sci.* 110 (12), 3788–3801. doi:10.1111/cas.14205

Firat-Karalar, E. N., and Welch, M. D. (2011). New Mechanisms and Functions of Actin Nucleation. *Curr. Opin. Cel Biol.* 23 (1), 4–13. doi:10.1016/j.ceb.2010.10.007

García-Ponce, A., Citalán-Madrid, A. F., Velázquez-Avila, M., Vargas-Robles, H., and Schnoor, M. (2015). The Role of Actin-Binding Proteins in the Control of Endothelial Barrier Integrity. *Thromb. Haemost.* 113 (1), 20–36. doi:10.1160/th14-04-0298

Iwaya, K., Norio, K., and Mukai, K. (2007). Coexpression of Arp2 and WAVE2 Predicts Poor Outcome in Invasive Breast Carcinoma. *Mod. Pathol.* 20 (3), 339–343. doi:10.1038/modpathol.3800741

Kang, M., Lu, S., Kuan Chong, P., Guan Yeoh, K., and Pin Lim, Y. (2016). Comparative Proteomic Profiling of Extracellular Proteins between Normal and Gastric Cancer Cells. *Curr. Cancer. Drug. Target.* 16 (5), 442–454. doi:10.2174/1568009616666151209113606

Kinoshita, T., Nohata, N., Watanabe-Takano, H., Yoshino, H., Hidaka, H., Fujimura, L., et al. (2012). Actin-related Protein 2/3 Complex Subunit 5 (ARPC5) Contributes to Cell Migration and Invasion and Is Directly Regulated by Tumor-Suppressive microRNA-133a in Head and Neck Squamous Cell Carcinoma. *Int. J. Oncol.* 40 (6), 1770–1778. doi:10.3892/ijo.2012.1390

Kiuchi, T., Nagai, T., Ohashi, K., and Mizuno, K. (2011). Measurements of Spatiotemporal Changes in G-Actin Concentration Reveal its Effect on Stimulus-Induced Actin Assembly and Lamellipodium Extension. *J. Cel Biol* 193 (2), 365–380. doi:10.1083/jcb.201101035

Laurila, E., Savinainen, K., Kuuselo, R., Karhu, R., and Kallioniemi, A. (2009). Characterization of the 7q21-Q22 Amplicon Identifies ARPC1A, a Subunit of the Arp2/3 Complex, as a Regulator of Cell Migration and Invasion in Pancreatic Cancer. *Genes Chromosom. Cancer* 48 (4), 330–339. doi:10.1002/gcc.20643

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 (W1), W509–W514. doi:10.1093/nar/gkaa407

Lin, S., Zheng, L., Lu, Y., Xia, Q., Zhou, P., and Liu, Z. (2020). Comprehensive Analysis on the Expression Levels and Prognostic Values of LOX Family Genes

in Kidney Renal clear Cell Carcinoma. *Cancer Med.* 9 (22), 8624–8638. doi:10.1002/cam4.3472

Liu, Z., Yang, X., Chen, C., Liu, B., Ren, B., Wang, L., et al. (2013). Expression of the Arp2/3 Complex in Human Gliomas and its Role in the Migration and Invasion of Glioma Cells. *Oncol. Rep.* 30 (5), 2127–2136. doi:10.3892/or.2013.2669

Molinie, N., and Gautreau, A. (2018). The Arp2/3 Regulatory System and its Deregulation in Cancer. *Physiol. Rev.* 98 (1), 215–238. doi:10.1152/physrev.00006.2017

Moriya, Y., Nohata, N., Kinoshita, T., Mutallip, M., Okamoto, T., Yoshida, S., et al. (2012). Tumor Suppressive microRNA-133a Regulates Novel Molecular Networks in Lung Squamous Cell Carcinoma. *J. Hum. Genet.* 57 (1), 38–45. doi:10.1038/jhg.2011.126

Pollard, T., and Beltzner, C. C. (2002). Structure and Function of the Arp2/3 Complex. *Curr. Opin. Struct. Biol.* 12 (6), 768–774. doi:10.1016/s0959-440x(02)00396-2

Rauhala, H. E., Teppo, S., Niemelä, S., and Kallioniemi, A. (2013). Silencing of the ARP2/3 Complex Disturbs Pancreatic Cancer Cell Migration. *Anticancer Res.* 33 (1), 45–52 .

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

Rotty, J. D., Brighton, H. E., Craig, S. L., Asokan, S. B., Cheng, N., Ting, J. P., et al. (2017). Arp2/3 Complex Is Required for Macrophage Integrin Functions but Is Dispensable for FcR Phagocytosis and *In Vivo* Motility. *Develop. Cel* 42 (5), 498–513. doi:10.1016/j.devcel.2017.08.003

Semba, S., Iwaya, K., Matsubayashi, J., Serizawa, H., Kataba, H., Hirano, T., et al. (2006). Coexpression of Actin-Related Protein 2 and Wiskott-Aldrich Syndrome Family Verproline-Homologous Protein 2 in Adenocarcinoma of the Lung. *Clin. Cancer Res.* 12 (8), 2449–2454. doi:10.1158/1078-0432.ccr-05-2566

Su, X., Wang, S., Huo, Y., and Yang, C. (2018). Short Interfering RNA-Mediated Silencing of Actin-Related Protein 2/3 Complex Subunit 4 Inhibits the Migration of SW620 Human Colorectal Cancer Cells. *Oncol. Lett.* 15 (3), 2847–2854. doi:10.3892/ol.2017.7642

Tan, C. K., Law, N. M., Ng, H. S., and Machin, D. (2003). Simple Clinical Prognostic Model for Hepatocellular Carcinoma in Developing Countries and its Validation. *J Clin. Oncol.* 21 (12), 2294–2298. doi:10.1200/jco.2003.03.151

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

Thul, P. J., Akesson, 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

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *Contemp. Oncol. (Pozn)* 19 (1A), A68–A77. doi:10.5114/wo.2014.47136

Tur-Gracia, S., and Martinez-Quiles, N. (2021). Emerging Functions of Cytoskeletal Proteins in Immune Diseases. *J. Cel Sci* 134 (3). jcs253534. doi:10.1242/jcs.253534

Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380 (15), 1450–1462. doi:10.1056/nejmra1713263

Xiong, T., and Luo, Z. (2018). The Expression of Actin-Related Protein 2/3 Complex Subunit 5 (ARPC5) Expression in Multiple Myeloma and its Prognostic Significance. *Med. Sci. Monit.* 24, 6340–6348. doi:10.12659/msm.908944

Xu, N., Qu, G. Y., Wu, Y. P., Lin, Y. Z., Chen, D. N., Li, X. D., et al. (2020). ARPC4 Promotes Bladder Cancer Cell Invasion and Is Associated with Lymph Node Metastasis. *J. Cel Biochem* 121 (1), 231–243. doi:10.1002/jcb.29136

Zhang, J., Liu, Y., Yu, C. J., Dai, F., Xiong, J., Li, H. J., et al. (2017). Role of ARPC2 in Human Gastric Cancer. *Mediators Inflamm.* 2017, 5432818. doi:10.1155/2017/5432818

Zhang, Y., Shen, H., Liu, H., Feng, H., Liu, Y., Zhu, X., et al. (2017). Arp2/3 Complex Controls T Cell Homeostasis by Maintaining Surface TCR Levels via Regulating TCR(+) Endosome Trafficking. *Sci. Rep.* 7 (1), 8952. doi:10.1038/s41598-017-08357-4

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

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.

Copyright © 2021 Huang, Li, Zhuang, Sun 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 16 July 2021
doi: 10.3389/fmolb.2021.685410

Implications of Lifestyle Factors and Polygenic Risk Score for Absolute Risk Prediction of Colorectal Neoplasm and Risk-Adapted Screening

Hongda Chen[1†], Li Liu[2†], Ming Lu[1], Yuhan Zhang[1], Bin Lu[1], Ying Zhu[2], Jianbo Tian[2], Xinying Li[2], Shaofa Nie[2], Xiaoping Miao[2] and Min Dai[1]**

[1]Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Epidemiology and Biostatistics, and the Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China

Reviewed by:
Sha Tao,
Shanghai Roche Pharmaceutical Co., Ltd., China
Harry H. X. Wang,
Sun Yat-Sen University, China

***Correspondence:**
Min Dai
daimin2002@hotmail.com
Xiaoping Miao
miaoxp@hust.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 25 March 2021
Accepted: 15 June 2021
Published: 16 July 2021

Citation:
Chen H, Liu L, Lu M, Zhang Y, Lu B, Zhu Y, Tian J, Li X, Nie S, Miao X and Dai M (2021) Implications of Lifestyle Factors and Polygenic Risk Score for Absolute Risk Prediction of Colorectal Neoplasm and Risk-Adapted Screening. Front. Mol. Biosci. 8:685410. doi: 10.3389/fmolb.2021.685410

Background: Estimation of absolute risk of developing colorectal neoplasm is essential for personalized colorectal cancer (CRC) screening. We developed models to determine relative and absolute risks of colorectal neoplasm based on lifestyle and genetic variants and to validate their application in risk-adapted screening.

Methods: We prospectively collected data from 203 advanced neoplasms, 464 non-advanced adenomas, and 1,213 healthy controls from a CRC screening trial in China in 2018–2019. The risk prediction model based on four lifestyle factors and a polygenic risk score (PRS) consisted of 19 CRC-associated single-nucleotide polymorphisms. We assessed the relative and 10-year absolute risks of developing colorectal neoplasm and the yield of a risk-adapted screening approach incorporating risk models, fecal immunochemical test, and colonoscopy.

Results: Compared to the participants with favorable lifestyle and lower PRS, those with unfavorable lifestyle and higher PRS had 2.87- and 3.79-fold higher risk of colorectal neoplasm in males and females, respectively. For a 50-year-old man or a 50-year-old woman with the highest risk profile, the estimated 10-year absolute risk of developing colorectal neoplasm was 6.59% (95% CI: 6.53–6.65%) and 4.19% (95% CI: 4.11–4.28%), respectively, compared to 2.80% (95% CI: 2.78–2.81%) for men and 2.24% (95% CI: 2.21–2.27%) for women with the lowest risk profile. The positive predictive value for advanced neoplasm was 31.7%, and the number of colonoscopies needed to detect one advanced neoplasm was 3.2.

Conclusion: The risk models, absolute risk estimates, and risk-adapted screening presented in our study would contribute to developing effective personalized CRC prevention and screening strategies.

Keywords: colorectal cancer, adenoma, polygenic risk score, risk stratification, personalized screening

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and fourth leading cause of cancer-related death worldwide (Bray et al., 2018). The majority of sporadic CRC arises from normal intestinal epithelium through sequentially worsening degrees of adenomatous dysplasia (Morson, 1974; Vogelstein et al., 1988). Early detection of CRC and its precancerous lesions by means of screening has been demonstrated to be effective in reducing the mortality, even for incidence (Brenner et al., 2014; Wolf et al., 2018). To maximize the cost-effectiveness of population-based CRC screening, research studies resolving the identification of individuals at high risk and implementation of appropriate screening modalities based on risk stratification are highly valuable (Robertson and Ladabaum, 2019).

Approximately 12–35% of CRC can be attributed to genetic predisposition (Dekker et al., 2019). So far, whole-genome–wide association studies (GWASs) have identified more than 100 common genetic variants associated with the risk of CRC (Peters et al., 2013; Zhang et al., 2014; Schmit et al., 2019), among which 24 loci were validated in eastern Asian populations (Lu et al., 2019). Although individual SNPs presented modest associations with CRC, the combination of these genetic polymorphisms, known as the polygenic risk score (PRS), presented a more predominant role in CRC risk prediction (Frampton et al., 2016; Carr et al., 2020; Kastrinos et al., 2020; Thomas et al., 2020). In addition to genetic predisposition, there are several well-established modifiable lifestyle factors related to CRC (Dekker et al., 2019), including physical activity, intake of red/processed meat and dietary fruits and vegetables, lower body mass index/waist circumference, smoking, and alcohol consumption. Adherence to healthy lifestyle has been demonstrated to reduce CRC risk (Turati et al., 2017; Petimar et al., 2019; Solans et al., 2020).

Although previous studies have implied that accumulation of predisposed risk alleles and violation of healthy lifestyle are associated with increased risk of colorectal cancer (Jeon et al., 2018; Carr et al., 2020), the detailed effects of genetic predisposition and environmental exposure on colorectal adenomatous dysplasia are not completely elucidated. Moreover, the absolute risk of developing colorectal neoplasia given a specific combination of risk factors has been barely explored. In addition, from the view of translational relevance, targeting high-risk population and implementation of appropriate risk-adapted screening intervention may improve the low participation and suboptimal screening effectiveness and cost-effectiveness in population-based CRC screening programs.

Therefore, using prospectively collected samples from a large-scale population-based CRC screening trial conducted in China from 2018 through 2019, the current study was designed to estimate the relative and absolute risks of colorectal neoplasia based on the lifestyle score and the PRS and to subsequently develop and evaluate the yield of risk-adapted screening approaches incorporating risk assessment with established screening modalities including the fecal immunochemical test (FIT) and colonoscopy in detecting colorectal neoplasms.

MATERIALS AND METHODS

Study Design and Study Sample

The study was conducted in the context of the TARGET-C trial, an ongoing study comparing the effectiveness of colonoscopy, FIT, and risk-adapted screening in CRC conducted in six centers of China since May 2018. Detailed study design has been described in previous publications (Chen et al., 2019; Chen et al., 2020). Briefly, 19,582 eligible participants aged 50–74 years were randomized into three arms in a 1:2:2 ratio: 1) one-time colonoscopy; 2) annual FIT; and 3) annual risk-adapted screening. The risk-adapted screening approach used an established CRC risk scoring system, the Asia-Pacific Colorectal Screening (APCS) score (Yeoh et al., 2011), in which participants at high risk of CRC were referred for colonoscopy, while participants at low risk were referred for FIT. All participants were required to undertake an epidemiological questionnaire survey to collect information including sociodemographic factors, history of diseases and clinical treatment, living habits, and family history of cancer. In addition, participants who needed colonoscopy examination were further required to donate stool and blood samples per standardized procedures. This study was approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (18-013/1,615), and the protocol was registered in the Chinese Clinical Trial Registry (ChiCTR1800015506). All participants provided written informed consent.

For the present study, we used the data and samples collected from the baseline screening phase of the TARGET-C trial. Overall, there were a total of 3,825 participants undertaking colonoscopy examination in the baseline screening phase. After excluding participants without blood samples (n = 94), having ineligible blood DNA quality for SNP typing (n = 1,294), and having failed SNP detection in at least one sample (n = 557), we finally included 1,880 samples for the final analysis, including 24 CRCs, 179 advanced adenomas, 464 non-advanced adenomas, and 1,213 controls without any significant finding at colonoscopy. The detailed sample selection scheme is presented in **Figure 1**.

Blood Sample Handing and SNP Typing

Blood samples were collected in a BD Vacutainer K2 EDTA tube (no. 367863) prior to bowel preparation for colonoscopy. After collection, the samples were handled in the laboratory of the hospital within 4 h after withdrawal. The blood samples were centrifuged at 1,200 g for 12 min at room temperature, aliquoted, and stored at −80°C until further use. For the present study, DNA was extracted from buffy coat using a commercial DNA extraction kit (Tiangen, Beijing, China) according to the manufacturer's instructions. Based on the findings from a large-scale GWAS of colorectal cancer in eastern Asia and a GWAS of advanced colorectal adenoma in Europe (Wang et al., 2013; Lu et al., 2019), 24 SNPs were initially selected and genotyped using the Sequenom MassARRAY platform, as shown in **Supplementary Table S1**.

Stool Sample Handling and FIT

Eligible participants were asked to collect one stool specimen into a sterile container (SARSTEDT, Germany) from a single bowel

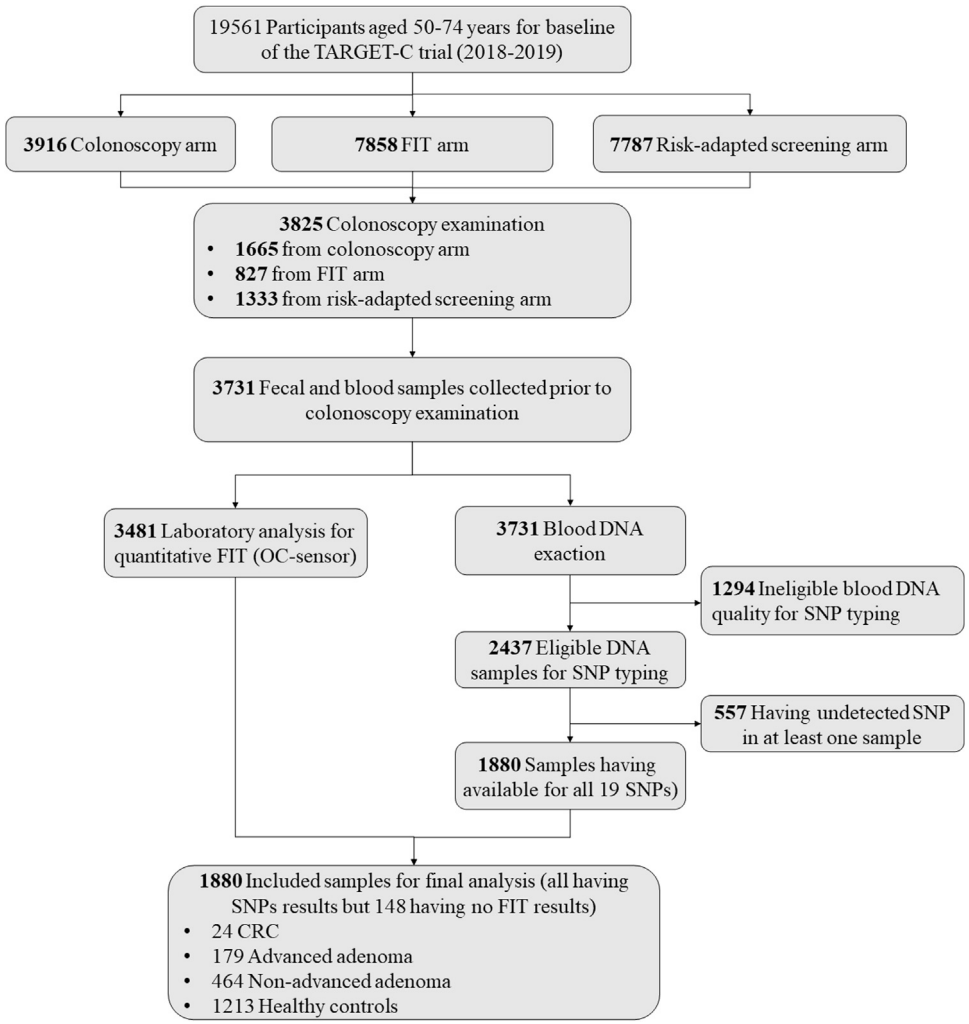


FIGURE 1 | Sample selection scheme for the present study.

movement, without any specific restrictions on diet or medicine within 24 h before colonoscopy examination. After collection, participants were asked to wrap the stool-filled container with an ice bag and to store it in a refrigerator at 4°C at home until their colonoscopy appointment. After receipt of the stool-filled containers, the samples were preserved at −20°C temporarily and were sent to the central biobank by cold chain logistics within one month, where the samples would be kept at −80°C for further analysis.

For the present study, the frozen fecal samples were tested by the quantitative FIT (OC-Sensor, Eiken Chemical, Japan) following a standard operating procedure in the central laboratory of the National Cancer Center, China. The laboratory staff was blinded to the colonoscopy results. Test values ≥100 ng Hb/ml (equivalent to 20 ug Hb/g, recommended cutoff by the manufacturer) were defined as positive in the present study.

Outcome Ascertainment and Definition

Colonoscopy examinations were performed in the designated hospital by experienced endoscopists. Standardized forms recording the colonoscopy and pathology reports were collected and verified. Moreover, to ensure the uniform standard of pathology diagnosis in different study sites, an experienced gastrointestinal pathologist from the National Cancer Center, China, independently reviewed the pathology sections of all CRCs, all advanced adenomas, and a random selection of 10% of the non-advanced adenomas. For the present study, we defined advanced adenoma as having at least one of the following features: 1) high-grade dysplasia; 2) villous or tubular–villous histologic features; and 3) adenoma of 1 cm or more in diameter.

Statistical Analysis

The distribution of demographic and lifestyle characteristics of the study population according to the screening outcomes (healthy control, non-advanced adenoma, and advanced neoplasm) was presented and compared by using the ANOVA test or chi-squared test where appropriate. Univariate logistic regression was applied to explore the associations between

individual lifestyle factors and colorectal neoplasia risk stratified by sex. The lifestyle factors potentially related to colorectal neoplasia ($p < 0.1$) were selected to construct sex-specific healthy lifestyle scores, which were created by dichotomizing the information of four lifestyle factors (waist circumference, red meat intake, fruit consumption, and smoking) for men and three lifestyle factors (waist circumference, red meat intake, and fruit consumption) for women (**Supplementary Table S2**). Then, the weighted lifestyle score (LS) was calculated as follows: $LS = \beta_1\chi_1 + \beta_2\chi_2 \ldots + \beta_k\chi_k \ldots + \beta_n\chi_n$, where β_k is the log-odds ratio (OR) for colorectal neoplasm related to the lifestyle factor k and χ_k is the value of the lifestyle factor k. After exclusion of five SNPs with statistical departure from the Hardy–Weinberg equilibrium (HWE), 19 SNPs were used to construct the PRS. The details of the SNPs are listed in **Supplementary Table S1**. Each SNP was coded as 0, 1, or 2 copies of the risk allele. The weighted PRS was calculated as follows: $PRS = \beta_1\chi_1 + \beta_2\chi_2 \ldots + \beta_k\chi_k \ldots + \beta_n\chi_n$, where β_k is the per-allele log-OR for colorectal neoplasm related to SNP k and χ_k is the allele dosage of SNP k. The weights for each SNP included in the PRS are presented in **Supplementary Table S3**.

Multiple logistic regression was used to estimate the associations of the healthy lifestyle score, the PRS, and the combination of these two scores with the risk of colorectal neoplasm, non-advanced colorectal adenoma, and advanced colorectal neoplasm by calculating ORs and 95% confidence intervals (CIs). We stratified the model by sex to allow for potential difference in the associations between men and women. In these analyses, the lifestyle score was dichotomized as unfavorable and favorable according to the sex-specific median of the score, with higher scores indicating unfavorable and lower scores representing favorable. The weighted PRS was categorized into two groups according to the 90th percentile of its distribution (higher PRS: ≥ the 90th percentile; lower PRS: < the 90th percentile). The subgroup analysis was conducted according to age and sex of the participants.

We further estimated the 10-year absolute risk and its 95% CIs for developing colorectal neoplasm for 50-year-old men and women with specific profiles of lifestyle score and PRS, based on the principles of modeling described by Freedman et al. (2009), Pfeiffer et al. (2011). Briefly, the estimation of the absolute risk of colorectal neoplasm includes estimating relative risk of colorectal neoplasm (calculated from a population-based case–control study) and attributable risk parameters and combining these estimates with baseline age-specific cancer hazard rates from the Chinese Center for Cancer Registry Data to estimate the probability of developing colorectal neoplasm during a specific time interval given a person's age, lifestyle score, and PRS. As colorectal neoplasm tends to occur at old ages, we accounted for the competing risks of non-CRC–specific mortality in the absolute risk estimation. The details of the calculation have been provided by Carr et al. (2020).

Based on risk stratification using environmental exposure and PRS described above, we designed a risk-adapted screening approach as shown in **Figure 2A**. Briefly, all subjects firstly involved in risk assessment; for subjects assessed to be at high risk, colonoscopy was recommended; for subjects assessed to be at low risk, FIT was offered; and those with positive test results were further offered colonoscopy. To determine the yield of efficiency of the risk-adapted screening scenarios and their comparison with the traditional colonoscopy-only approach, we calculated the positive predictive values (PPVs) for detecting advanced neoplasm (including CRC or advanced adenoma) and any neoplasm (including CRC, advanced adenoma, and non-advanced adenoma). The PPV was defined as the number of patients with true diagnosis of interest divided by the number of subjects deemed positive. To assess the resource load of colonoscopy, we calculated the number of colonoscopies needed to be screened (NNS) to detect one lesion.

All analyses were performed using SAS (version 9.4, SAS Institute Inc., Cary, NC) and R software (version 3.5.1) (ore Team (2018). R: A, 2018), and all statistic tests were two-sided, with the p value less than 0.05 being statistically significant.

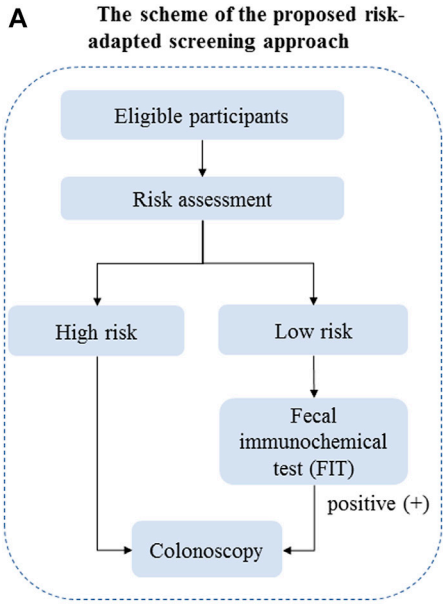
RESULTS

Basic Characteristics

A detailed sample selection scheme from the TARGET-C study is shown in **Figure 1**. Overall, 1880 participants were finally included into analysis, consisting of 203 advanced neoplasia cases (24 CRCs and 179 advanced adenomas), 464 non-advanced adenomas, and 1,213 healthy controls (**Table 1**). The mean age of the participants was 60.5 years, and 51.3% were male. Compared with healthy controls, colorectal neoplasm cases, including non-advanced adenoma and advanced neoplasm, were more likely to be male, older, and obese and to have more consumption of red meat but less intake of fruits. In addition, colorectal neoplasm cases had a higher lifestyle score and PRS compared with health controls (**Table 1**).

Association of Lifestyle Score and PRS With Colorectal Neoplasm Risk

Unfavorable lifestyle was associated with increased risk of any colorectal neoplasm, non-advanced adenoma, and advanced neoplasm, with the ORs of 1.68 (95% CI: 1.30–2.17), 1.53 (95% CI: 1.15–2.03), and 2.04 (95% CI: 1.40–3.00) for males and 1.85 (95% CI: 1.32–2.58), 1.72 (95% CI: 1.18–2.51), and 2.26 (95% CI: 1.22–4.21) for females. Higher PRSs presented positive association with the risk of colorectal neoplasm, non-advanced adenoma, and advanced neoplasm, with the ORs of 1.83 (95% CI: 1.21–2.75), 1.79 (95% CI: 1.14–2.80), and 1.87 (95% CI: 1.07–3.27) for males and 2.08 (95% CI: 1.27–3.40), 2.40 (95% CI: 1.41–4.07), and 1.30 (95% CI: 0.49–3.42) for females (**Supplementary Table S4**). Similar risk effects of unfavorable lifestyle and higher PRS on colorectal neoplasm, non-advanced adenoma, and advanced neoplasm have been found in the sensitivity analysis by adding a lifestyle–PRS interaction item in logistic regression models, although the 95% confidence intervals of the ORs became much wider due to limited samples of women (**Supplementary Table S5**). The comparisons of the risk of colorectal neoplasm, non-advanced adenoma, and advanced neoplasm with the lifestyle score and genetic score according to age increase are presented in **Supplementary Figure S1**. The



B **Comparison of the yield and efficiency of different screening scenarios**

Screening approach	Colonoscopy needed to be performed	Colonoscopy load reduced (%)	Advanced neoplasm		Any neoplasm	
			PPV	NNS	PPV	NNS
All participants undertaking screening colonoscopy	1732	Reference	10.8% (187/1732)	9.3	35.2% (610/1732)	2.8
Risk-adapted screening based on lifestyle score solely	804	53.60%	13.6% (109/804)	7.4	37.8% (304/804)	2.6
Risk-adapted screening based on PRS solely	212	87.80%	23.1%* (49/212)	4.3	55.7%* (118/212)	1.8
Risk-adapted screening based on combination of lifestyle score and PRS*	123	92.90%	31.7%* (39/123)	3.2	61.0%* (75/123)	1.6

Abbreviations: NNS, number of colonoscopy needed to be performed to detect one lesion; PPV, positive predictive value; PRS, polygenic risk score.
† Participants were deemed to be high risk only if both lifestyle and polygenic risk score were categorized as high risk.
* P-value <0.05 for the PPVs compared with the reference approach (all participants undertaking screening colonoscopy)

FIGURE 2 | (A) Scheme of the proposed risk-adapted screening approach; **(B)** comparison of the yield and efficiency of different screening scenarios.

TABLE 1 | Basic characteristics of included participants.

Characteristics	Total (n = 1880)	Healthy controls (n = 1,213)	Non-advanced adenoma (n = 464)	Advanced neoplasm (n = 203)	*p* value[a]
Sex, n (%)					<0.0001
Female	916 (48.7)	700 (57.7)	158 (34.1)	58 (28.6)	
Male	964 (51.3)	513 (42.3)	306 66.0)	145 (71.4)	
Age, mean (standard deviation), years	60.5 (6.3)	59.8 (6.3)	61.5 (6.1)	61.7 (6.3)	<0.0001
Smoking status, n (%)					
Male, ≥15 pack-years	464 (48.1)	228 (44.4)	159 (52.0)	77 (53.1)	0.05
Female, >0 pack-years	14 (1.5)	9 (1.3)	4 (2.5)	1 (1.7)	0.51
Alcohol consumption (> 250 ml alcohol/week), n (%)					
Male	181 (18.8)	91 (17.7)	54 (17.7)	36 (24.8)	0.13
Female	17 (1.9)	14 (2.0)	2 (1.3)	1 (1.7)	0.82
Waistline, mean (standard deviation), cm					
Male	86.9 (11.1)	86.1 (11.3)	87.2 (11.2)	89.5 (9.8)	0.004
Female	81.4 (11.3)	80.7 (11.5)	83.0 (11.0)	84.9 (8.8)	0.004
Red meat consumption, n (%)					
Male	901 (93.5)	475 (92.6)	285 (93.1)	141 (97.2)	0.13
Female	798 (91.4)	602 (86.0)	143 (90.5)	53 (91.4)	0.19
Fruit consumption, n (%)					
Male	585 (60.7)	298 (58.1)	196 (64.1)	91 (62.8)	0.21
Female	477 (52.1)	348 (49.7)	91 (57.6)	38 (65.5)	0.02
Physically active[b], n (%)					
Male	421 (43.7)	237 (46.2)	124 (40.5)	60 (41.4)	0.24
Female	386 (42.1)	299 (42.7)	64 (40.5)	23 (39.7)	0.81
Polygenic risk score, median (interquartile range)	0.25 (0.05–0.46)	0.22 (0.02–0.42)	0.32 (0.12–0.53)	0.27 (0.08–0.53)	<0.0001
Lifestyle score, median (interquartile range)					
Male	0.74 (0.49–0.92)	0.68 (0.49–0.92)	0.81 (0.49–0.92)	0.81 (0.58–1.00)	<0.0001
Female	0.80 (0.47–0.87)	0.78 (0.47–0.87)	0.87 (0.47–1.20)	0.87 (0.80–1.20)	0.0003

Advanced neoplasm included colorectal cancer and advanced adenoma. Favorable controls indicated no significant findings at colonoscopy examination.
[a]*p values were calculated by conducting the chi-square test for categorical variables or ANOVA test for continuous variables.*
[b]*Physically active was defined as at least 1 h moderate-to-intense leisure time physical activity per week.*

TABLE 2 | Odds ratios of risk profiles with colorectal neoplasm, non-advanced adenoma, and advanced neoplasm stratified by sex.

	Any neoplasm		Non-advanced adenoma		Advanced neoplasm	
	No. of cases/no. of controls	OR (95% CI)	No. of cases/no. of controls	OR (95% CI)	No. of cases/no. of controls	OR (95% CI)
Male						
Favorable lifestyle and lower PRS	166/264	1	118/264	1	48/264	1
Favorable lifestyle and higher PRS	28/22	2.01 (1.11–3.64)	21/22	2.12 (1.12–4.02)	7/22	1.73 (0.70–4.28)
Unfavorable lifestyle and lower PRS	219/205	1.71 (1.30–2.25)	144/205	1.59 (1.17–2.15)	75/205	2.01 (1.34–3.02)
Unfavorable lifestyle and higher PRS	38/22	2.87 (1.64–5.04)	23/22	2.41 (1.29–4.51)	15/22	3.96 (1.91–8.22)
Female						
Favorable lifestyle and lower PRS	52/270	1	40/270	1	12/270	1
Favorable lifestyle and higher PRS	9/22	2.14 (0.93–4.92)	7/22	2.18 (0.87–5.45)	2/22	2.05 (0.43–9.76)
Unfavorable lifestyle and lower PRS	135/381	1.86 (1.30–2.65)	94/381	1.69 (1.13–2.53)	41/381	2.43 (1.25–4.71)
Unfavorable lifestyle and higher PRS	20/27	3.79 (1.98–7.28)	17/27	4.26 (2.13–8.53)	3/27	2.50 (0.67–9.42)

PRS: polygenic risk score; OR: odds ratio; CI: confidence interval.
Lifestyle was binarized as favorable and unfavorable according to the median of the lifestyle score.
Genetic risk was categorized as lower and higher according to the 90th percentile of the polygenic risk score.

combined effects of lifestyle score and PRS on colorectal neoplasia were further detected. Compared to the participants with favorable lifestyle and lower PRS, those with unfavorable lifestyle and higher PRS had 2.87-, 2.41-, and 3.96-fold risk of colorectal neoplasm, non-advanced adenoma, and advanced neoplasm in males and 3.79-, 4.26-, and 2.50-fold risk of colorectal neoplasm, non-advanced adenoma, and advanced neoplasm in females (**Table 2**). Further analyses stratified by age revealed the adverse effect of unfavorable lifestyle and higher PRS was more apparent in younger age in males, but in older age in females (**Supplementary Table S6**). In the sensitivity analysis where the PRSs were categorized into the higher genetic risk group (Tertile 3) and lower genetic risk group (Tertiles 1 and 2), similar but attenuated associations between the risk profiles and colorectal neoplasm were observed (**Supplementary Table S7**).

Absolute Risk Estimates for Colorectal Neoplasm Based on Lifestyle Score and PRS

The attributable risk estimates for colorectal neoplasm stratified by age and sex are presented in **Supplementary Table S8**. The estimated colorectal adenoma incidence in China is listed in **Supplementary Table S9**. Based on the national death registry data and cancer registry information, we summarized the mortality for men and women in China in 2015 in **Supplementary Table S10**. **Table 3** presents the estimates of the 10-year absolute risk of developing colorectal neoplasm for males and females separately, aged 50 years, combining information on the lifestyle score and PRS. The 10-year absolute risk of developing colorectal neoplasm varied across risk profiles in males and females. To illustrate, for a 50-year-old man with the highest risk profile (unfavorable lifestyle and higher PRS), the estimated 10-year absolute risk of developing colorectal neoplasm was 6.59% (95% CI: 6.53–6.65%), compared to 2.80% (95% CI: 2.78–2.81%) for men with the lowest risk profile (adherence to favorable lifestyle and lower PRS). For a 50-year-old woman with the highest risk profile, the estimated 10-year absolute risk of developing colorectal neoplasm was 4.19% (95% CI: 4.11–4.28%), compared to 2.24% (95% CI: 2.21–2.27%) for females with favorable lifestyle and lower PRS (**Table 3**).

TABLE 3 | 10-year absolute risk estimates of colorectal neoplasm for 50-year-old men and women.

Subgroup	10-year absolute risk (95% CI)	
	Men	Women
Favorable lifestyle		
Lower PRS	2.80 (2.78–2.81)	2.24 (2.21–2.27)
Higher PRS	6.12 (6.04–6.20)	4.08 (4.00–4.16)
Unfavorable lifestyle		
Lower PRS	4.56 (4.54–4.58)	3.60 (3.58–3.62)
Higher PRS	6.59 (6.53–6.65)	4.19 (4.11–4.28)

Abbreviations: CI, confidence interval: PRS, polygenic risk score.
Lifestyle was binarized as favorable and unfavorable according to the median of the lifestyle score.
Genetic risk was categorized as lower and higher according to the 90th percentile of the polygenic risk score.

Yield of Risk-Adapted Screening

To evaluate the translation potential of the constructed risk assessment score in CRC screening, we proposed a risk-adapted screening approach incorporating risk assessment, FIT, and colonoscopy, and the detailed scheme is shown in **Figure 2A**. The detailed comparison of the yield and resource load of different screening scenarios is shown in **Figure 2B**. For the traditional approach in that all participants aged 50–74 years undertake screening colonoscopy, the PPV and NNS for detecting advanced neoplasm were 10.8% and 9.3, respectively; those for detecting any neoplasm were 35.2% and 2.8, respectively. Setting the yield of the colonoscopy-only screening approach as a reference, the risk-adapted screening approach showed remarkably increased PPVs, decreased colonoscopy load, and lower NNSs for detecting colorectal neoplasm. The strongest improved PPV and NNS were observed for risk-adapted screening combining the lifestyle score and PRS, yielding the PPV and NNS of 31.7% and 3.2 for detecting advanced neoplasm, respectively, and 61.0% and 1.6 for detecting any neoplasm.

DISCUSSION

Using samples from the population-based CRC screening trial conducted in China, we presented the relative risk and 10-year absolute risk estimates for developing colorectal neoplasm considering the lifestyle score and PRS. Moreover, we further empirically demonstrated that the risk-adapted CRC screening incorporating risk assessment and established screening modalities has great translational potential in terms of increasing PPV for detecting advanced neoplasm and reducing resource load of colonoscopy. Such findings of our study may therefore have strong implications in future CRC screening, which may aid the suboptimal participation and efficiency of the current one-size-fits-all screening strategies, especially for countries with relatively low disease burden and limited healthcare resources.

The relative risk estimates of the genetic or lifestyle factors with colorectal neoplasm have been reported previously (Wang et al., 2013; Bailie et al., 2017). To our knowledge, this is the first estimate of absolute risk for developing colorectal neoplasm based on genetic and lifestyle factors. The absolute risk results support our hypothesis that lifestyle factors may modify the risk of colorectal neoplasm. The 10-year absolute risk associated with unfavorable lifestyle was greatest in the group at higher genetic risk, which also emphasizes the benefit of adhering to healthy lifestyle. Within the lower polygenic risk category, unfavorable lifestyle resulted in about 60% higher 10-year absolute risk of colorectal neoplasm in both males and females when compared to favorable lifestyle, suggesting that the genetic insusceptibility of colorectal neoplasm can be offset by an unhealthy lifestyle. In addition, although one may perceive that having a higher genetic risk score means powerless against genetic predisposition, the current study implies that a healthy lifestyle and a personalized CRC screening, at least to some extent, may reduce the risk of colorectal neoplasm (Carr et al., 2020). The current study added more values of genetic and lifestyle factors in the risk prediction of colorectal precancerous lesions, which provides a validation for previous findings of the health effects of genetic and lifestyle factors on colorectal carcinogenesis.

The estimate of absolute risk of CRC has implications of personalized CRC screening. Previously, Joen J and colleagues (Jeon et al., 2018) derived the risk prediction models using 9 lifestyle and environmental factors and 63 SNPs, and the results showed that the starting ages of screening varied significantly for individuals with different risk profiles, although screening is recommended to begin at 50 years for individuals with no family history of CRC. In addition, Frampton M. J. and colleagues (Frampton et al., 2016) derived a PRS using 37 SNPs and calculated absolute risk of CRC from the United Kingdom population age structure, and they found that personalized screening using the PRS would then result in 26% fewer men and women being eligible for screening with 7 and 5% fewer screen-detected cases. For our study, due to lack of evidence of reference absolute risk of colorectal adenoma, we therefore did not provide recommendation of personalized starting age based on the risk profiles. However, the absolute risk estimates can still be useful to facilitate communication and to better inform the public about the magnitude and potentials of CRC prevention and may help to define those likely to maximally benefit from chemoprevention and screening.

In our previous study (Chen et al., 2020), we demonstrated that the risk-adapted screening approach incorporating the environmental risk assessment, FIT, and colonoscopy had satisfying participation rate and superior yield than the FIT-based screening strategy in population-based CRC screening. The current study finding expanded our previous research and indicated that the risk-adapted screening approach incorporating the PRS additionally may further improve the screening yield in terms of higher PPVs for advanced neoplasm and lower NNSs. However, such validation was only validated in a retrospective manner, and technical issues and operational details such as optimal positivity threshold of risk score, population compliance rate, and health resource allocation need to be considered carefully before such a strategy is introduced. In addition, with the development of the novel early detection biomarkers, how to incorporate the PRS with established and novel screening modalities deserves further research.

In the present study, the SNPs used for the PRS were selected based on previous GWAS findings, especially for eastern Asian population. There were several issues needed to be taken into consideration when comparing our results with others. First, the discovered SNPs associated with advanced adenoma are rather limited; therefore, the prediction accuracy of the constructed PRS may not be optimal given the majority of the samples were actually adenomas from a CRC screening setting. Second, there were some disparities regarding the CRC-associated SNPs between the Asian and European populations. For our study, we specifically selected some validated CRC-associated SNPs in the eastern Asian populations, which may not be the susceptibility genes in the other populations. Third, we adopted a targeted genotyping approach using the Sequenom MassARRAY platform rather than using the whole-genome sequencing, which only detected a limited number of SNPs simultaneously. We anticipated that the risk prediction accuracy could be further enhanced, if more colorectal neoplasm–associated SNPs were

included in the PRS. However, from the view of translational use in population-based screening, the proposed PRS should have the merit of affordable cost and convenient-to-detect nature.

Our study has specific strengths. To our knowledge, this is the first study to estimate the absolute risk of colorectal neoplasm based on genetic and lifestyle factors in Chinese population. The current study estimates the probability of developing colorectal neoplasm over a 10-year time interval using data from a nation-wide colonoscopy screening program and a multicentered CRC screening trial. Thus, it is expected that the risk prediction is to some extent representative of the general China population. Moreover, the lifestyle and genetic factors involved in the prediction model can be attained in the clinical setting, which provides more practical meaning for generalization.

Our study also has some limitations. First, the sample size was relatively small, which disenabled the combination of lifestyle and genetic information in a very detailed way. Second, all lifestyle factors were collected retrospectively by interview; therefore, recall bias might not be avoided. However, since all information was collected prior to colonoscopy, this recall bias might cause undeferential misclassification of exposures, which consequently resulted in an underestimated effect of unhealthy lifestyle. Third, the numbers of included SNPs in the PRS were limited.

To sum up, using samples from a multicentered CRC screening trial conducted in China, we presented the relative risk and 10-year absolute risk estimates for developing colorectal neoplasm considering the lifestyle score and genetic predisposition. Retrospective analysis revealed that the proposed risk-adapted screening approach incorporating the lifestyle score, PRS, FIT, and colonoscopy showed improved screening efficiency than the traditional colonoscopy-based screening approach. The proposed risk prediction models and the risk-adapted screening approach would contribute to the development of effective personalized CRC prevention and screening strategies in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available upon reasonable request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HC, LL, XM, and MD conceptualized and designed the study. ML, YuZ, BL, YiZ, JT, XL, and SN participated in acquisition of data and analysis and interpretation of data. HC and LL participated in statistical analysis and drafted the manuscript. All authors critically revised the manuscript and approved the final version.

FUNDING

This work was supported by the Natural Science Foundation of Beijing Municipality (7202169), Beijing Nova Program of Science and Technology (Z191100001119065), CAMS Innovation Fund for Medical Sciences (2017-I2M-1-006), National Natural Science Foundation of China (81925032, 81703309, and 81974491), the Technical Innovation Special Project of Hubei Province (2019ACA135), and Health Commission of Hubei Province Scientific Research Project (WJ2019Q027). The funders had no role in study design and conduct; data collection, management, analysis, or interpretation; manuscript preparation, review, or approval; or the decision to submit the manuscript for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.685410/full#supplementary-material

REFERENCES

Bailie, L., Loughrey, M. B., and Coleman, H. G. (2017). Lifestyle Risk Factors for Serrated Colorectal Polyps: A Systematic Review and Meta-Analysis. *Gastroenterology* 152 (1), 92–104. doi:10.1053/j.gastro.2016.09.003

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

Brenner, H., Stock, C., and Hoffmeister, M. (2014). Effect of Screening Sigmoidoscopy and Screening Colonoscopy on Colorectal Cancer Incidence and Mortality: Systematic Review and Meta-Analysis of Randomised Controlled Trials and Observational Studies. *BMJ* 348, g2467. doi:10.1136/bmj.g2467

Carr, P. R., Weigl, K., Edelmann, D., Jansen, L., Chang-Claude, J., Brenner, H., et al. (2020). Estimation of Absolute Risk of Colorectal Cancer Based on Healthy Lifestyle, Genetic Risk, and Colonoscopy Status in a Population-Based Study. *Gastroenterology* 159 (1), 129–138.e9. doi:10.1053/j.gastro.2020.03.016

Chen, H., Li, N., Shi, J., Ren, J., Liu, c., Zhang, y., et al. (2019). Comparative Evaluation of Novel Screening Strategies for Colorectal Cancer Screening in China (TARGET-C): a Study Protocol for a Multicentre Randomised Controlled Trial. *BMJ Open* 9 (4), e025935. doi:10.1136/bmjopen-2018-025935

Chen, H., Lu, M., Liu, C., Zou, S., Du, L., Liao, X., et al. (2020). Comparative Evaluation of Participation and Diagnostic Yield of Colonoscopy vs Fecal Immunochemical Test vs Risk-Adapted Screening in Colorectal Cancer Screening: Interim Analysis of a Multicenter Randomized Controlled Trial (TARGET-C). *Am. J. Gastroenterol.* 115 (8), 1264–1274. doi:10.14309/ajg.0000000000000624

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., and Wallace, M. B. (2019). Colorectal Cancer. *The Lancet* 394 (10207), 1467–1480. doi:10.1016/s0140-6736(19)32319-0

Frampton, M. J. E., Law, P., Litchfield, K., Morris, E. J., Kerr, D., Turnbull, C., et al. (2016). Implications of Polygenic Risk for Personalised Colorectal Cancer Screening. *Ann. Oncol.* 27 (3), 429–434. doi:10.1093/annonc/mdv540

Freedman, A. N., Slattery, M. L., Ballard-Barbash, R., Willis, G., Cann, B. J., Pee, D., et al. (2009). Colorectal Cancer Risk Prediction Tool for white Men and Women without Known Susceptibility. *Jco* 27 (5), 686–693. doi:10.1200/jco.2008.17.4797

Jeon, J., Du, M., Schoen, R. E., Hoffmeister, M., Newcomb, P. A., Berndt, S. I., et al. (2018). Determining Risk of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. *Gastroenterology* 154 (8), 2152–2164.e19. doi:10.1053/j.gastro.2018.02.021

Kastrinos, F., Samadder, N. J., and Burt, R. W. (2020). Use of Family History and Genetic Testing to Determine Risk of Colorectal Cancer. *Gastroenterology* 158 (2), 389–403. doi:10.1053/j.gastro.2019.11.029

Lu, Y., Kweon, S.-S., Tanikawa, C., Jia, W.-H., Xiang, Y.-B., Cai, Q., et al. (2019). Large-Scale Genome-wide Association Study of East Asians Identifies Loci Associated with Risk for Colorectal Cancer. *Gastroenterology* 156 (5), 1455–1466. doi:10.1053/j.gastro.2018.11.066

Morson, B. (1974). The Polyp-Cancer Sequence in the Large Bowel. *Proc. R. Soc. Med.* 67 (6 Pt 1), 451–457. doi:10.1177/00359157740676p115

Peters, U., Jiao, S., Schumacher, F. R., Hutter, C. M., Aragaki, A. K., Baron, J. A., et al. (2013). Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-wide Meta-Analysis. *Gastroenterology* 144 (4), 799–807.e24. doi:10.1053/j.gastro.2012.12.020

Petimar, J., Smith-Warner, S. A., Rosner, B., Chan, A. T., Giovannucci, E. L., and Tabung, F. K. (2019). Adherence to the World Cancer Research Fund/American Institute for Cancer Research 2018 Recommendations for Cancer Prevention and Risk of Colorectal Cancer. *Cancer Epidemiol. Biomarkers Prev.* 28 (9), 1469–1479. doi:10.1158/1055-9965.epi-19-0165

Pfeiffer, R. M., and Petracci, E. (2011). Variance Computations for Functionals of Absolute Risk Estimates. *Stat. Probab. Lett.* 81 (7), 807–812. doi:10.1016/j.spl.2011.02.002

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. . Available at: https://www.R-project.org/.

Robertson, D. J., and Ladabaum, U. (2019). Opportunities and Challenges in Moving from Current Guidelines to Personalized Colorectal Cancer Screening. *Gastroenterology* 156 (4), 904–917. doi:10.1053/j.gastro.2018.12.012

Schmit, S. L., Edlund, C. K., Schumacher, F. R., Gong, J., Harrison, T. A., Huyghe, J. R., et al. (2019). Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J. Natl. Cancer Inst.* 111 (2), 146–157. doi:10.1093/jnci/djy099

Solans, M., Chan, D. S. M., Mitrou, P., Norat, T., and Romaguera, D. (2020). A Systematic Review and Meta-Analysis of the 2007 WCRF/AICR Score in Relation to Cancer-Related Health Outcomes. *Ann. Oncol.* 31 (3), 352–368. doi:10.1016/j.annonc.2020.01.001

Thomas, M., Sakoda, L. C., Hoffmeister, M., Rosenthal, E. A., Lee, J. K., van Duijnhoven, F. J. B., et al. (2020). Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am. J. Hum. Genet.* 107 (3), 432–444. doi:10.1016/j.ajhg.2020.07.006

Turati, F., Bravi, F., Di Maso, M., Bosetti, C., Polesel, J., Serraino, D., et al. (2017). Adherence to the World Cancer Research Fund/American Institute for Cancer Research Recommendations and Colorectal Cancer Risk. *Eur. J. Cancer* 85, 86–94. doi:10.1016/j.ejca.2017.08.015

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., et al. (1988). Genetic Alterations during Colorectal-Tumor Development. *N. Engl. J. Med.* 319 (9), 525–532. doi:10.1056/nejm198809013190901

Wang, J., Carvajal-Carmona, L. G., Chu, J.-H., Zauber, A. G., Kubo, M., Matsuda, K., et al. (2013). Germline Variants and Advanced Colorectal Adenomas: Adenoma Prevention with Celecoxib Trial Genome-wide Association Study. *Clin. Cancer Res.* 19 (23), 6430–6437. doi:10.1158/1078-0432.ccr-13-0550

Wolf, A. M. D., Fontham, E. T. H., Church, T. R., Flowers, C. R., Guerra, C. E., LaMonte, S. J., et al. (2018). Colorectal Cancer Screening for Average-Risk Adults: 2018 Guideline Update from the American Cancer Society. *CA: A Cancer J. Clinicians* 68 (4), 250–281. doi:10.3322/caac.21457

Yeoh, K.-G., Ho, K.-Y., Chiu, H.-M., Zhu, F., Ching, J. Y. L., Wu, D.-C., et al. (2011). The Asia-Pacific Colorectal Screening Score: a Validated Tool that Stratifies Risk for Colorectal Advanced Neoplasia in Asymptomatic Asian Subjects. *Gut* 60 (9), 1236–1241. doi:10.1136/gut.2010.221168

Zhang, B., Jia, W. H., Jia, W.-H., Matsuda, K., Kweon, S.-S., Matsuo, K., et al. (2014). Large-scale Genetic Study in East Asians Identifies Six New Loci Associated with Colorectal Cancer Risk. *Nat. Genet.* 46 (6), 533–542. doi:10.1038/ng.2985

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.

Copyright © 2021 Chen, Liu, Lu, Zhang, Lu, Zhu, Tian, Li, Nie, Miao 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 Molecular Biosciences

REVIEW
published: 20 July 2021
doi: 10.3389/fmolb.2021.716415

Annexin A3, a Calcium-Dependent Phospholipid-Binding Protein: Implication in Cancer

*Liu Yang[1†], Pingan Lu[2†], Xiaohui Yang[1], Kaiguo Li[1] and Song Qu[1]**

[1]Key Laboratory of High-Incidence Tumor Prevention and Treatment (Guangxi Medical University), Ministry of Education, Department of Radiation Oncology, Guangxi Medical University Cancer Hospital, Nanning, China, [2]Faculty of Medicine, Amsterdam Medical Centre, University of Amsterdam, Amsterdam, Netherlands

Annexin A3 (ANXA3), also known as lipocortin III and placental anticoagulant protein III, has been reported to be dysregulated in tumor tissues and cancer cell lines, and harbors pronounced diagnostic and prognostic value for certain malignancies, such as breast, prostate, colorectal, lung and liver cancer. Aberrant expression of ANXA3 promotes tumor cell proliferation, invasion, metastasis, angiogenesis, and therapy resistance to multiple chemotherapeutic drugs including platinum-based agents, fluoropyrimidines, cyclophosphamide, doxorubicin, and docetaxel. Genetic alterations on the ANXA3 gene have also been reported to be associated with the propensity to form certain inherited, familial tumors. These diverse functions of ANXA3 in tumors collectively indicate that ANXA3 may serve as an attractive target for novel anticancer therapies and a powerful diagnostic and prognostic biomarker for early tumor detection and population risk screening. In this review, we dissect the role of ANXA3 in cancer in detail.

Keywords: annexin A3, biomarkers, drug resistance, neoplasm, signal transduction, tumorigenesis

OPEN ACCESS

Edited by:
William C Cho, QEH, China

Reviewed by:
Liwei Lang, Augusta University, United States
Dinesh Kumar, Centre of Bio-Medical Research (CBMR), India

****Correspondence:***
Song Qu
13607887386@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *28 May 2021*
Accepted: *08 July 2021*
Published: *20 July 2021*

Citation:
Yang L, Lu P, Yang X, Li K and Qu S (2021) Annexin A3, a Calcium-Dependent Phospholipid-Binding Protein: Implication in Cancer. Front. Mol. Biosci. 8:716415. doi: 10.3389/fmolb.2021.716415

INTRODUCTION

Annexin A3 (ANXA3), a water-soluble protein consisting of 323 amino acid residues, is encoded by the ANXA3 gene located on human chromosome 4q13–q22 (Mussunoor and Murray, 2008). Also known as lipocortin III and placental anticoagulant protein II, ANXA3 subordinate to the annexin family, which is a well-characterized multigene family of structurally homologous, but functionally different calcium-dependent membrane phospholipid-binding proteins ubiquitously distributed in a wide array of cell types (Moss and Morgan, 2004).

The annexin family can be divided into five classes (A-E) based on their biological origins; among them, 12 subtypes of class A annexins, annexins A1 to A11 and A13, are derived from human and vertebrate orthologues, while class B till E are originated from non-vertebrate metazoans, fungi and molds, plants and protists respectively (Gerke and Moss, 2002). ANXA3 has been demonstrated to be

Abbreviations: ANXA3, Annexin A3; Bax, Bcl-2-associated X protein; Bcl-2, B-cell lymphoma 2; CAF, Cancer associated fibroblast; CDKs, Cyclin-dependent kinases; CRC, Colorectal cancer; EMT, Epithelial-Mesenchymal Transition; HCC, Hepatocellular carcinoma; HIF-1α, Hypoxia-inducible factor 1-alpha; IκBα, Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha; LC3B, Microtubule-associated proteins 1A/1B light chain 3B; MET, Mesenchymal-epithelial transition; NF-κB, Nuclear factor kappa-light-chain-enhancer of activated B cells; NSCLC, Non-small cell lung cancer; p27, Cyclin-dependent kinase inhibitor 1B; p38, p38 mitogen-activated protein kinases; PARP, Poly (ADP-ribose) polymerase; SNP, Single-nucleotide polymorphism; TKI, Tyrosine kinase inhibitor; TNBC, Triple-negative breast cancer; TNF, Tumor necrosis factor; VEGF, Vascular endothelial growth factor; VEGFR, Vascular endothelial growth factor receptor.

virtually solely expressed in differentiated myeloid cell lines and accounts for about 1% of the cytosolic protein of human neutrophils (Sopkova et al., 2002). To date, two isoforms of ANXA3 with the molecular mass of 33 and 36 kDa have been documented. The 36 kDa ANXA3 isoform is mainly expressed in monocytes, while the 33 kDa ANXA3 isoform is more abundantly observed in neutrophils (Le Cabec and Maridonneau-Parini, 1994).

Four conserved annexin repeats structural domains (I-IV) constitute the C-terminal protein core of ANXA3 (Sopkova et al., 2002) (**Figure 1**). Each structural domain encompasses five *a*-helices (A-E) consisting of 70 amino acid residues (Favier-Perron et al., 1996; Gerke and Moss, 2002). In addition, one principal calcium-binding site is present on the convex face formed by *a*-helices A and B in each structural domain (Raynal and Pollard, 1994; Mussunoor and Murray, 2008). The highly variable N-terminus of 20 amino acids in length contains two tryptophan residues, which are W5 in the N-terminal segments and W190 at the end of the IIIA-IIIB loop (Sopkova et al., 2002). These two tryptophan residues are essential for the protein stability as well as the interaction of ANXA3 with intracellular calcium ions and negatively charged phospholipids, thereby regulating a diverse range of biological functions of ANXA3 (Hofmann et al., 2000; Sopkova et al., 2002). Due to its phospholipid-binding capacity and calmodulin-dependent nature, ANXA3 primarily participates in membrane-associated activities, such as intracellular and extracellular signal transduction, vesicular transport, membrane fusion and endocytosis, formation and transport of ion channels, and interactions of cytoskeleton proteins (Swairjo and Seaton, 1994; Perron et al., 1997; Mussunoor and Murray, 2008). The involvement of ANXA3 in cellular signal transduction facilitates its multifaceted regulatory roles in various physiological activities, including cell division, differentiation, motility and apoptosis as well as anti-inflammation, anticoagulation and angiogenesis (Moss and Morgan, 2004; Mussunoor and Murray, 2008). Meanwhile, dysregulation of ANXA3 has been reported to play a pivotal role in cancer development and progression (Mussunoor and Murray, 2008) (**Table 1** and **Figure 4**). However, data published so far about its expression in different malignancies are inconsistent. To the best of our knowledge, ANXA3 has been reported to be overexpressed in a majority cancer types including breast (Pendharkar et al., 2016; Zhou et al., 2017a; Zhou et al., 2017b; Guo et al., 2017; Aravind Kumar et al., 2018; Du et al., 2018; Kim et al., 2018; Li et al., 2018; Zhou et al., 2018), colorectal (Madoz-Gurpide et al., 2006; Marshall et al., 2010; Yip et al., 2010; Yang L. et al., 2018; Yang Q. et al., 2018; Xu et al., 2019), bladder (Tsai et al., 2018), ovarian (Jiang et al., 2019), gastric (Takahashi et al., 2015; Wang and Li, 2016) and pancreatic cancer (Baine et al., 2011a; Baine et al., 2011b; Wan et al., 2020) as well as hepatocellular (Pan et al., 2015a; Pan et al., 2015b; Tong et al., 2018) and nasopharyngeal carcinoma (Ruan et al., 2010), while downregulated in renal (Bianchi et al., 2010), prostate (Wozny et al., 2007; Köllermann et al., 2008; Peraldo-Neia et al., 2011) and papillary thyroid cancer (Jung et al., 2010). Furthermore, the expression of ANXA3 in lung cancer remains controversial, with the evidence of both upregulated (Liu et al., 2009; Győrffy et al., 2013; Wang et al., 2019; Jin et al., 2020; Liu et al., 2021) and downregulated (Rho et al., 2009; Wu et al., 2018; Lohinai et al., 2019) expression patterns documented in the literature. In an immunohistochemistry-based study of organotypic *ex vivo* human HCC clinical samples and HCC patient-derived xenografts, Tong et al. found that overexpression of ANXA3 was associated with enhanced resistance to sorafenib and led to poor survival of HCC patients receiving sorafenib treatment. Their data further indicates that targeting ANXA3 could effectively inhibit tumor growth and sensitize the response of tumor cells to sorafenib treatment (Tong et al., 2018). In addition, ANXA3 mRNAs and proteins were overexpressed in gastric cancer tissues and various gastric cancer cell lines, as detected by RT-PCR and Western blot analyses (Wang and Li, 2016). This aberrant expression was further correlated with the depth of tumor infiltration and TNM stage in both univariate and multivariate analyses of a cohort of 183 gastric cancer patients, which indicates the potential of ANXA3 as an independent prognosticator for the survival of gastric cancer patients (Wang and Li, 2016). Likewise, markedly elevated ANXA3 expression was detected in bladder cancer by multiplexed liquid chromatography multiple-reaction-monitoring mass spectrometry assay (LC-MRM-MS). Investigators from this study further suggested that ANXA3 might serve as a reliable non-invasive diagnostic biomarker for bladder cancer (Tsai et al., 2018). Similarly, ANXA3 was overexpressed in colorectal cancer (CRC) tissues compared to adjacent normal tissues, as shown from immunohistochemistry and western blot results (Yang Q. et al., 2018). Moreover, a Max Vision immunohistochemistry-based retrospective analysis of a cohort of 309 breast cancer patients demonstrated that ANXA3 expression in triple negative breast cancer (TNBC) patients was significantly higher than other breast cancer subtypes (Zhou et al., 2017b). Given that overexpression of ANXA3 has a vital impact on tumor progression, we could expect that downregulation of ANXA3 can also exert certain regulatory effects on tumorigenesis. Interestingly, ANXA3 expression level was diminished in prostate tumor tissues and was correlated with increasing pathological stages and Gleason scores (Köllermann et al., 2008). Immunohistochemistry and tissue microarray data further confirmed that ANXA3 could be used as an independent prognostic factor to predict the survival of prostate cancer patients and to support population risk stratification (Köllermann et al., 2008). Downregulation of ANXA3 was also reported in papillary thyroid cancer (PTC), and PTC patients with decreased ANXA3 expression exhibited substantially elevated lymph node metastasis scores and tumor growth (Jung et al., 2010).

Collectively, aberrant expression of ANXA3 plays a crucial role in malignant tumor development. It stimulates tumor cell proliferation, facilitates invasion, migration and metastasis, induces angiogenesis, desensitizes patient response to antitumor treatments and predisposes the emergence of certain inherited familial tumors (Gerke and Moss, 2002; Moss and Morgan, 2004; Mussunoor and Murray, 2008; Tong et al., 2018; Sarquis et al., 2020). Therefore, it is important to shedding

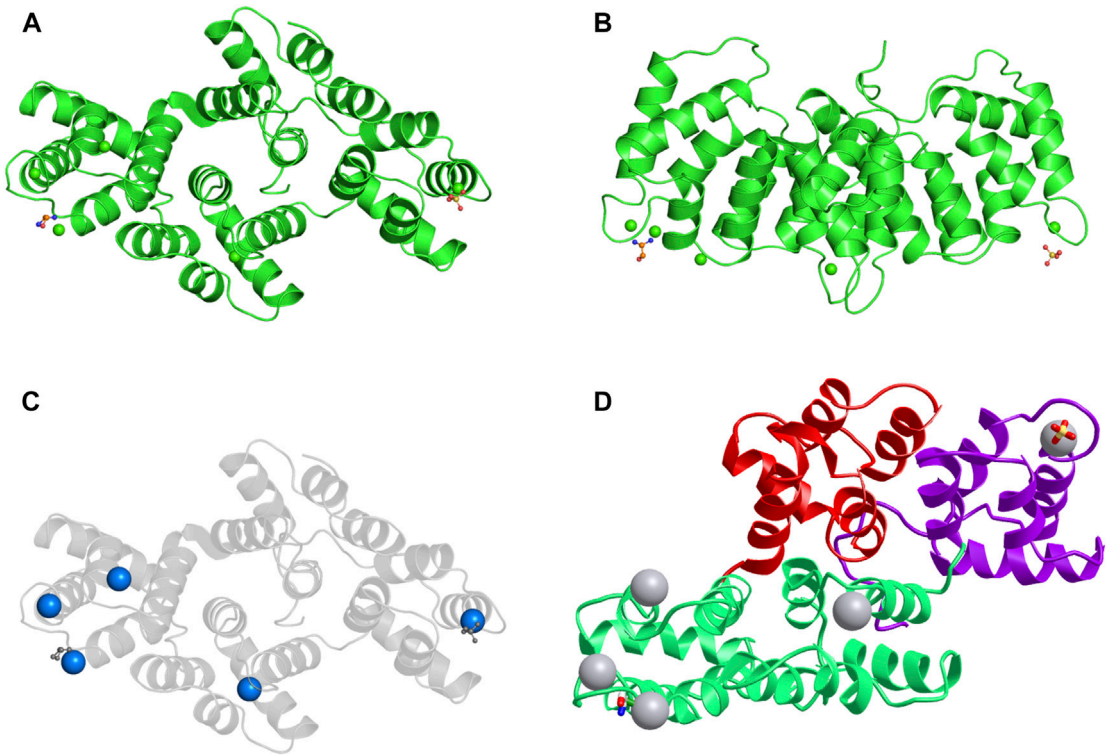


FIGURE 1 | The structure of the ANXA3 protein. **(A,B)** 3D structure of ANXA3 colored by chain and viewed from the front **(A)** and side **(B)**. **(C)** 3D structure viewed from front with calcium-binding sites indicated by calcium ions colored as blue spheres. **(D)** 3D structure colored by domain. 5 *a*-helices are clearly visible in each structural domain. The remaining two chemical structures are one sulfate ion and one ethanolamine. Data derived from Protein Data Bank and iCn3D web-based 3D structure viewer.

light on the functions of ANXA3 in tumor biology in order to improve the early detection of preneoplastic tumors, to overcome anticancer therapy resistance and to develop novel, targeted approaches to treat solid tumors. This review focusses on the roles of ANXA3 in cancer.

THE ROLE OF ANXA3 IN TUMORIGENESIS

Sustaining Proliferative Signaling

Sustaining proliferative signaling is a common strategy used by cancer cells to facilitate their progression and aggressiveness. Recent years, the pro-proliferative role of ANXA3 has been substantiated to be heavily implicated in various types of malignancies, and a diversity of evidence has been provided about the underlying mechanisms mediating this process (**Figure 2**). In hepatocellular carcinoma, ANXA3 has been shown to activate the Notch and MAPK/ERK/JNK signaling pathway, resulting in enhanced cell proliferation and promotion of stem-cell like characteristics (Pan et al., 2015a; Tong et al., 2015). Support on this finding was delivered by *in vitro* data of colorectal cancer, in which the phosphorylation of ERK and JNK was found to be significantly reduced once a depletion of ANXA3 was established using small interfering RNA (siRNA) (Xu et al., 2019). This finding was further corroborated by another study on chemoresistant non-small cell lung cancer (NSCLC) cells, which showed that high level of ANXA3 secreted by cancer associated fibroblasts (CAFs) in the tumor microenvironment activated the JNK/survivin signaling, thereby helping cancer cells escaping the cisplatin-induced apoptosis (Wang et al., 2019). Another putative mechanism contributing to the pro-proliferative effect of ANXA3 was provided by the work of Wan et al., in which PI3K/Akt signaling pathway was found to be substantially inhibited in pancreatic cancer patients with overexpression of miR-382, a miRNA that suppresses the expression of the ANXA3 gene (Wan et al., 2020). Investigators in this study further observed a decline of clone formation ability and proliferative behavior in pancreatic cancer cells overexpressing miR-382. Interestingly, a research on breast cancer has described an opposite correlation; the depletion of ANXA3 using short hairpin RNA plasmids has been shown to promote cell proliferation in both cell-line models and mouse xenograft models (Du et al., 2018). However, a series of studies on breast cancer challenged this finding. Collectively, these studies demonstrated that ANXA3 is highly expressed in luminal A, B and triple negative breast cancer subtypes, and that ANXA3 inhibition could significantly impair tumor growth *in vivo*, concomitant with a lower proliferation index and a higher apoptosis rate and G0/1 cell count *in vitro* (Zhou et al., 2017a; Li et al., 2018; Zhou et al., 2018).

TABLE 1 | The role of ANXA3 in tumor proliferation, invasion and metastasis.

Types	Expression	Models	Evidence	References
Hepatocellular carcinoma (HCC)	↑	6 cell lines and HCC tumor tissues	ANXA3 is overexpressed in sorafenib-resistant HCC cells, which inhibits PKCδ/p38-associated apoptosis and stimulates p38-mediated autophagy for cell survival	Tong et al. (2018)
	↑	7 cell lines; Tissues from 83 patients	Endogenous and secretory ANXA3 promotes tumor growth and stemness acquisition through dysregulating JNK pathway	Tong et al. (2015)
	↑	4 cell lines; 107 patients with HCC	The expression of HIF-1a, CD133, Notch1, Notch2 is significantly increased in ANXA3-overexpressing cells. Overexpression of ANXA3 enhances the proportion and tumorigenicity of CD133 + cells	Pan et al. (2015a)
	↑	Sk-hep-1 and SMMC-7721 cell line	Small interfering RNA silencing ANXA3 inhibits tumorigenesis and metastasis	Pan et al. (2015b)
Breast cancer	↑	MDA-MB 231 cell line and tissue from 30 primary breast cancer patients	ANXA3 knockdown inhibits proliferation, invasion, migration, and colony formation of tumor cells	Kim et al. (2018)
	↑	18 female nude mice inoculated with MDA-MB-231 cells	Breast cancer cells transfected with ANXA3 silencing shRNA exhibit significantly lower tumor weight, volume and tumorigenic activity	Li et al. (2018)
	↑	Samples from 158 patients; Tissue specimens	Enhanced cell proliferation indexes are positively correlated with the ANXA3 mRNA and protein expression level	Zhou et al. (2018)
	↑	16 pairs of breast cancer tissues and adjacent normal tissues; 3 cell lines	Silencing ANXA3 suppresses the NFκB pathway via upregulating IκBα, leading to mesenchymal-epithelial transition (MET) with attenuated invasion and metastasis, but promotes tumor cell proliferation	Du et al. (2018)
	↑	2 cell lines	Migration and invasion ability is lower in ANXA3 silenced cells	Zhou et al. (2017a)
	↑	309 breast cancer patients and their tissue specimens	ANXA3 is correlated with increased number of lymphatic metastases and advanced histological grading	Zhou et al. (2017b)
Lung carcinoma	↑	3 cell lines	Activation of the ANXA3/JNK pathway inhibits cisplatin-induced apoptosis	Wang et al. (2019)
	↑	Excised tissues from 21 lung AdC patients	ANXA3 expression is positively correlated with lymph node metastasis and the clinicopathological stages of lung adenocarcinoma (AdC)	Liu et al. (2009)
	↑	5 cell lines; 102 LC tissues and 102 paracancerous tissues	Knockdown of ANXA3 expression by miR-1253 inhibited proliferation, invasion and increased apoptosis rate	Liu et al. (2021)
Colorectal cancer (CRC)	↑	Tumor tissues from 107 CRC patients; 5 cell lines	miR-340-5p directly targets ANXA3, resulting in enhanced CRC cell proliferation, migration, and invasion	Yang et al. (2018a)
	↑	2 cell lines	ANXA3 depletion inhibits proliferation and facilitates apoptosis in oxaliplatin-resistant cells	Xu et al. (2019)
Pancreatic cancer	↑	Sample from 115 PC patients; 4 cell lines	ANXA3 knockdown using microRNA-382 inhibits PI3K/AKT signaling pathway, give rise to suppression of pancreatic cancer proliferation, invasion and metastasis	Wan et al. (2020)
Gastric cancer	↑	Tissues from 183 GC patients; 5 cell lines	ANXA3 depletion suppresses cell proliferation, invasion and metastasis	Wang and Li, (2016)
Ovarian cancer	↑	OC tissues; Normal ovarian cell line and OC cell lines	Overexpression ANXA3 leads to augmented proliferative and migratory behavior	Jiang et al. (2019)
Papillary thyroid carcinoma	↓	25 patients-derived tissue specimens	Reduced ANXA3 immunohistochemical staining is correlated with tumors with higher lymph node metastasis scores and larger sizes	Jung et al. (2010)

Aside from promoting the pro-proliferative pathways, aberrant expression of ANXA3 has also been shown to downregulate multiple pro-apoptotic proteins and cyclin-dependent kinases (CDKs), which facilitates the evasion of apoptosis and cell cycle arrest. For example, *in vitro* investigations on breast cancer cell lines HCC-1954 and MDA-MB-231 demonstrated that ANXA3 silencing using siRNA significantly reduced the expression of CDK4 and enhanced the expression of E2F1 and $p27^{Kip1}$ (Kim et al., 2018). Furthermore, high levels of ANXA3 in hepatocellular carcinoma has been demonstrated to attenuate the PKCδ/p38 associated apoptosis (Tong et al., 2018). Of note, p38 not only plays a role in the regulation of apoptosis, but also acts a key regulator in autophagy, which is another defining feature of tumor cells that supplies metabolic fuel sources for unlimited proliferation (Webber, 2010; Webber and Tooze, 2010). Unsurprisingly, investigators from this study further detected a substantially increased level of autophagic marker LC3B in both HCC cells *in vitro* and mouse xenografts *in vivo*, thereby confirming the positive correlation between ANXA3 expression and autophagic activity (Tong et al., 2018). The anti-apoptotic ability of ANXA3 was further substantiated by the work of Wang et al., wherein pro-apoptotic proteins caspase 3 and caspase 8 was significantly downregulated in NSCLC cells overexpressing ANXA3 (Wang et al., 2019). Conversely, miRNA-induced silencing of ANXA3 markedly upregulated caspase 3

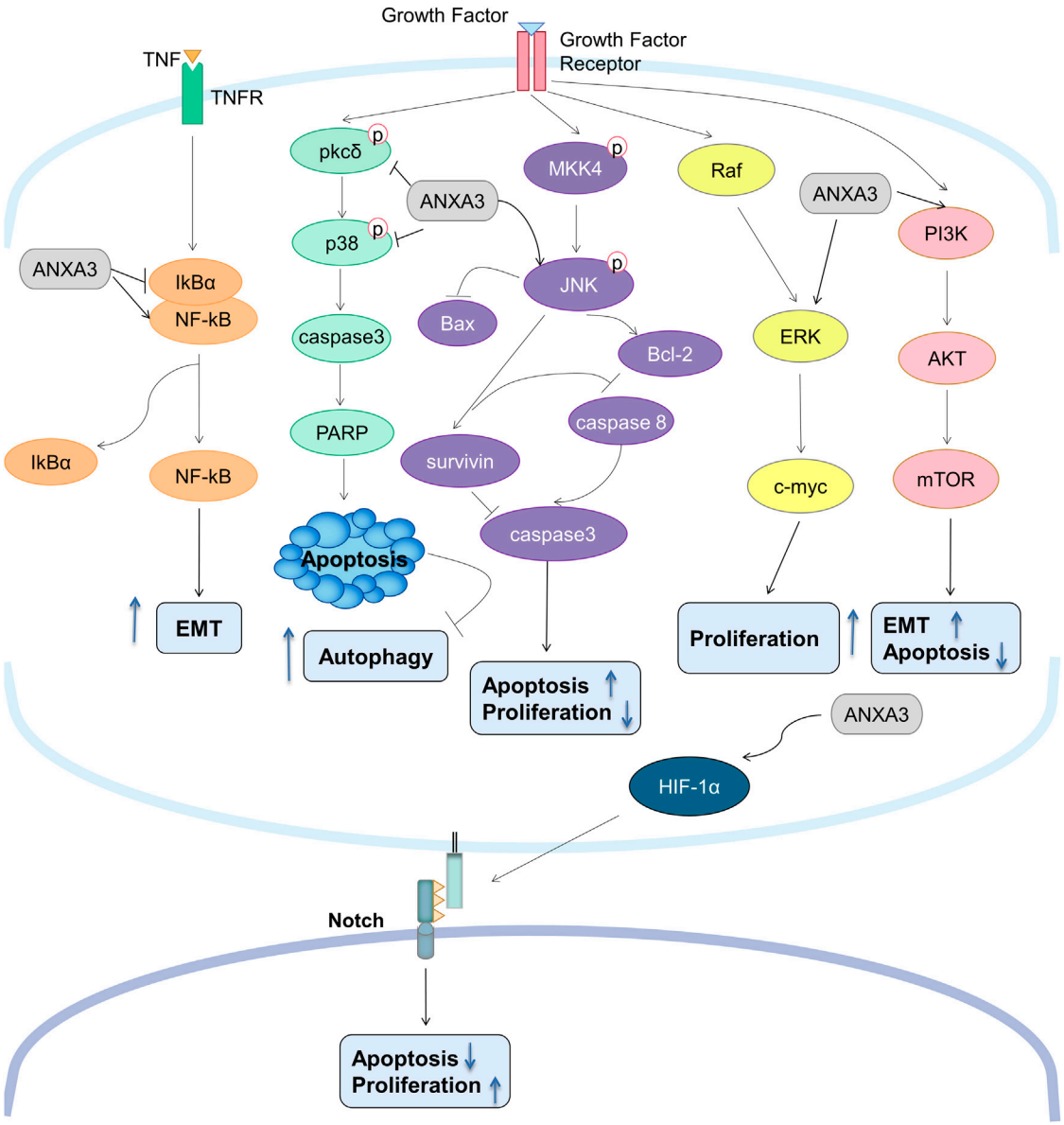


FIGURE 2 | Overview of the signal transduction network of ANXA3 in tumors. ANXA3 promotes tumorigenesis *via*: (1) activating NFkB pathway leading to increased EMT; (2) inhibiting PKCδ/p38 pathway leading to decreased apoptosis and increased autophagy; (3) activating JNK/survivin and Raf/ERK/c-myc pathways leading to increased proliferation; (4) activating PI3K/Akt/mTOR pathway leading to increased EMT and decreased apoptosis; and (5) activating HIF-1α/Notch pathway leading to increased proliferation and decreased apoptosis. p38, p38 mitogen-activated protein kinases; PARP, Poly (ADP-ribose) polymerase; bax, Bcl-2-associated X protein; bcl-2, B-cell lymphoma 2; IκBα, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; HIF-1α, Hypoxia-inducible factor 1-alpha; EMT, Epithelial-Mesenchymal Transition; TNF, Tumor necrosis factor.

expression as well as the expression of pro-apoptotic protein Bax, while suppressing the expression anti-apoptotic protein Bcl-2 (Liu et al., 2021). Consistent data were also published by Xu et al., who observed an ANXA3 knockdown-induced upregulation of c-caspase 3 and c-PARP in colorectal cancer cells (Xu et al., 2019).

Promoting Invasion and Metastasis

Invasion and metastasis are the major cause of poor clinical outcomes of malignant diseases. *In vitro* investigations have revealed that ANXA3 overexpression significantly stimulated the invasion and migration of breast cancer cells (Ibrahim et al., 2012; Guo et al., 2017; Kim et al., 2018). Clinically, ANXA3 overexpression has been demonstrated to be correlated with the occurrence of lymph node metastasis and the clinicopathological stages of breast cancer (Zhou et al., 2017b) and lung adenocarcinoma (Liu et al., 2009). The correlation was further reversely validated by the work of Zhou et al., in which knockdown of ANXA3 by shRNA impaired the invasion and

migration abilities of luminal A and triple negative breast cancer cells (Zhou et al., 2017a). Similar approaches employing miRNA and siRNA to probe the effect of ANXA3 depletion have also been applied in *in vitro* cell-line models of colorectal carcinoma (Yang L. et al., 2018; Xu et al., 2019) and ovarian carcinoma (Jiang et al., 2019), which presented consistent results. Intriguingly, an inverse relationship between ANXA3 expression and the execution of invasion-metastasis cascade in malignant tumors has also been documented. In surgical tumor specimens from 25 thyroid papillary cancer patients receiving thyroidectomy, downregulation of ANXA3 was detected in tumor tissues compared to the adjacent non-tumor tissues (Jung et al., 2010). Further immunohistochemistry results showed that reduced ANXA3 staining was correlated with thyroid papillary tumors with higher lymph node metastasis scores and larger sizes (Jung et al., 2010). Nonetheless, no data are currently available about what factors govern the downregulated expression of ANXA3 and its potentially anti-oncogenic activities in thyroid neoplasms.

A number of attempts has been made to decipher the underlying molecular changes mediating the ANXA3-induced invasion and migration. In ANXA3 overexpressing gastric cancer cells and patient-derived tumor specimens, an enhanced degree of epithelial-mesenchymal transition (EMT) was observed, which was evidenced by western blot results indicating increased expressions of mesenchymal markers vimentin and *ß*-catenin, a decreased expression of epithelial marker E-cadherin, and increased expressions of EMT-related transcription factors fibronectin, Slug and Snail (Wang and Li, 2016). Contrariwise, silencing of ANXA3 inhibited the expression of N-cadherin and vimentin in pancreatic cancer, while elevating the expression of E-cadherin (Wan et al., 2020). Additionally, ANXA3-knockdown pancreatic cancer cells exhibited decreased expressions of VEGF-C and VEGF-D (Wan et al., 2020), both of which are proteins previously shown to be positively associated with the number of lymph node metastases (Schulz et al., 2011). Similar findings were reported by research involving triple negative breast cancer cell lines MDA-MB-231 and MDA-MB-486, which found mesenchymal-epithelial transition (MET) in cancer cells receiving ANXA3-targetting shRNA plasmids, evidenced by decreased mesenchymal markers (vimentin and N-cadherin) and increased epithelial markers (E-cadherin and *γ*-cadherin) (Du et al., 2018). This reversed transition pattern could be considered as an indicator of diminished invasion and migration abilities. Interestingly, this study further indicated that IκBα knockdown could partially reverse the ANXA knockdown-induced MET state concomitant with an increased phospho-NF-κB p65 expression. These findings suggest that NF-κB signaling might be crucially involved in the ANXA3-medicated tumor cell invasion and migration (Du et al., 2018).

Inducing Angiogenesis

Angiogenesis is the essential mechanism that allows continuous nutrients supply and assists tumor cells to combat hypoxia in the tumor microenvironment. Over the last decade, a series of studies has identified novel clues shedding light on the functional link between ANXA3 and tumor angiogenesis. At histological level, less blood vessels were observed in H&E-stained tissue slides of triple negative breast cancer xenografts from mice pretreated with ANXA3-silencing shRNA plasmids (Li et al., 2018). At molecular level, ANXA3 silencing in pancreatic cancer cells resulted in a decrement of the expression of vascular endothelial growth factor receptor 3 (VEGFR3), which is a protein suggested to play a key role in lymphatic vascularization in pancreatic cancer (Wan et al., 2020). In addition, investigators from this study observed an ANXA3-induced upregulation of PI3K/Akt signaling pathway, which is a canonical pathway also capable of promoting neovascularization apart from its pro-proliferative effects (Sun et al., 2016; Wan et al., 2020). The hypoxia-inducible factor-1α (HIF-1α)/VEGF pathway is another well-established response of cancer cell to initiate angiogenesis and thereby survive hypoxia (Xu et al., 2018; Zhang et al., 2018; Zhang P.-C. et al., 2020). Employing liver cancer stem-like cells, Pan et al. revealed a positive correlation between the expressions of ANXA3 and HIF1α, which further confirmed the pro-angiogenic role of ANXA3 (Pan et al., 2015a). Similar findings were also presented by a research on bone-cancer induced pain (BCP), showing that downregulation of ANXA3 using shRNA substantially inhibited the expression of HIF1α and VEGF in the ipsilateral spinal cord and microglial cells of mice undertaken 21 days of bone cancer induction (Zhang Z. et al., 2020). Besides cancer-related investigations, fundamental studies using human umbilical vein endothelial cell (HUVEC) model provided extra confirming data (Park et al., 2005; Mohr et al., 2017). Employing the hypoxia responsive element (HRE)-luciferase reporter assay, investigators demonstrated that ANXA3 upregulated the HIF1α transactivation activity and VEGF production, give rise to enhanced tube formation and migration of HUVECs(Park et al., 2005). These findings suggest that ANXA3 might act as a key driver in angiogenic processes.

Paradoxically, a negative relationship between the expressions of ANXA3 and HIF1α has been reported in renal cell carcinoma (RCC) (Bianchi et al., 2010). Whereas RCC cultures not expressing HIF1α (HIF1α-negative) exhibited a similar level of ANXA3 expression as the matched noncancerous cortex, a markedly decreased ANXA3 expression was detected in HIF1α-positive RCC cultures compared to the matched counterparts (Bianchi et al., 2010). Moreover, the abundance of 36-kDa ANXA3 was significantly reduced in HIF1α-positive RCC cultures, while the 33-kDa ANXA3 showed a pronounced increment (Bianchi et al., 2010). Most importantly, the total expression of ANXA3 protein was significantly lower in RCC cultures *in vitro* as well as in surgical excised RCC tissues compared to their paired counterparts (Bianchi et al., 2010). It was speculated that the downregulation of the total ANXA3 protein in RCC was associated with the decrement of its 36-kDa isoform, as the N-terminus of ANXA3 capable of promoting self-expression is present in the 36-kDa isoform but absent in the 33-kDa isoform (Hofmann et al., 2000; Gerke et al., 2005; Bianchi et al., 2010). However, the mechanism underlying the diminution of 36-kDa ANXA3 in RCC remains unclear, and the differences between the two isoforms regarding their expressions and functions in other cancer models are still far from fully understood.

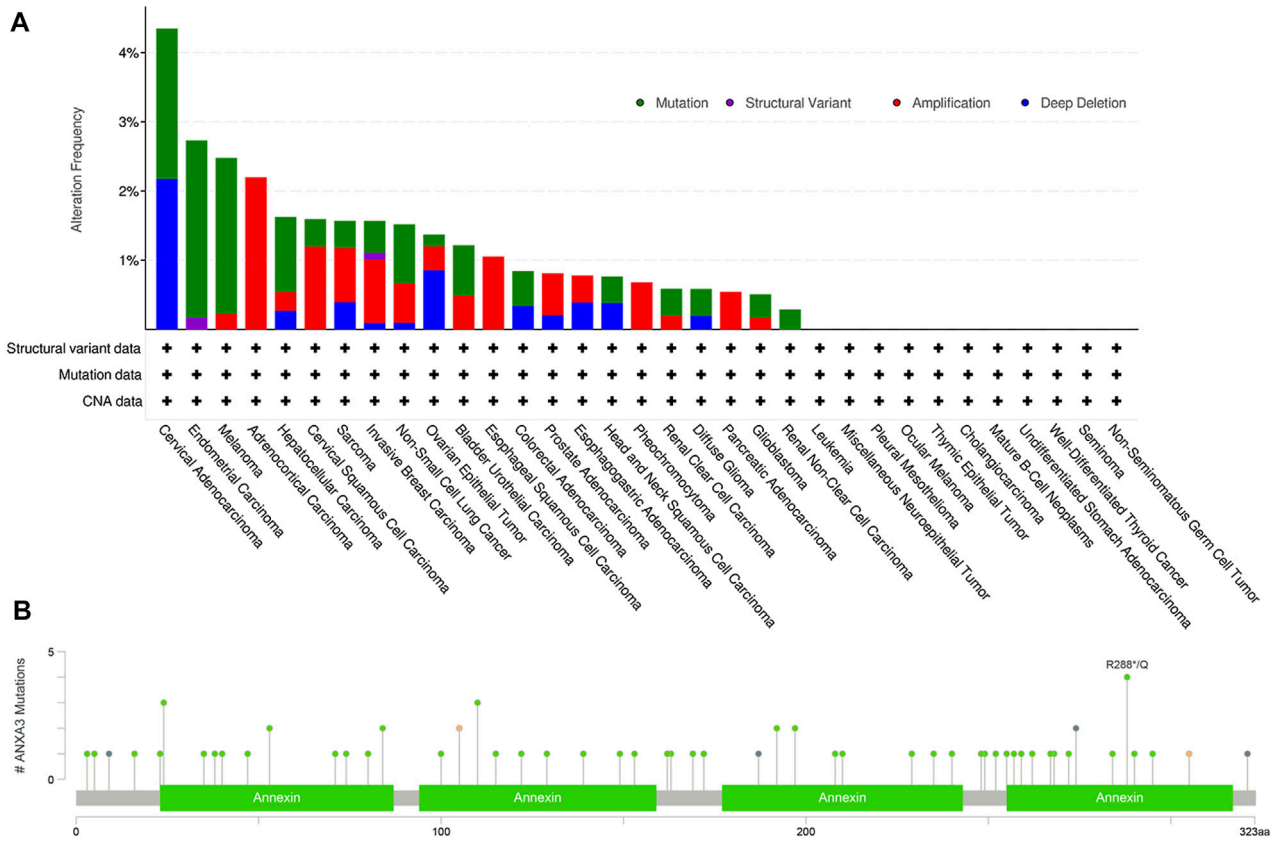


FIGURE 3 | Genetic alterations of the ANXA3 gene. **(A)** The alteration frequency of the ANXA3 gene in different cancer types. Y-axis represented the alteration frequency, and X-axis showed the cancer types with descending alteration frequencies. The colors used in the histogram represented different types of genetic alterations, which were shown with the top right legend. **(B)** The type, location and number of mutations occurred on the ANXA3 gene. Y-axis showed the number of ANXA3 mutations and the X-axis represented the amino acid sequence. Green, black, orange, and violet dots represented missense mutations, truncating mutations, splice-site mutations, and gene fusions respectively. Data has been retrieved from cBioPortal (http://www.cbioportal.org, June 2021).

Genetic Predisposition

Genetic predisposition usually serves as a prerequisite in cancer initiation and harbors pronounced clinical relevance especially in aggressive malignancies where an urgent need exists for novel risk screening methods with good predictive performance and clinical utility. Cancer-associated mutations can not only drive accelerated cancer progression but can also exhibit inherited patterns, contributing to familial hereditary tumors (AlHarthi et al., 2020). So far, very limited evidence has been published reporting the association between genetic alterations of the ANXA3 gene with cancer susceptibility. In a case-control study of 29 TNBC patients searching for risk-associated SNPs through microarray-based whole genome SNP sequencing, the non-synonymous SNP exm4087722 was detected in the ANXA3 gene (NM_005139) with a minor allele frequency (MAF) smaller than 0.05 (Aravind Kumar et al., 2018). Whole genome sequencing was also applied in another research on 3 Brazilian families with hereditary papillary thyroid cancer (PTC), which found point mutation p.D283N on the ANXA3 gene (MIM No.106490) (Sarquis et al., 2020). Yet, it is important to note that above observations are both based on small sample sizes and that more conclusive data from larger cohorts are still awaited about the role of genetic alterations of the ANXA3 gene in cancer susceptibility.

To generate a more comprehensive picture of the potentially oncogenic alterations on the ANXA3 gene, a computational analysis was performed using cBioPortal (http://www.cbioportal.org) querying 10,953 patients/10967 samples from 32 TCGA PanCancer Atlas studies (June 2021). Overall, the gene exhibited a low alteration frequency, with a detectable alteration in 121 (1.1%) of the total queried patients. The top three malignant diseases (**Figure 3A**) with the highest ANXA3 gene alterations are cervical adenocarcinoma (4.35% alterations including 2.17% mutation and 2.17% deep deletion), endometrial carcinoma (2.73% alterations including 2.56% mutation and 0.17% structural variant), and melanoma (2.48% alterations including 2.25% mutation and 0.23% amplification). Furthermore, the largest number of mutations on the ANXA3 gene (4 registered cases) occurred at amino acid 288 on its fourth annexin repeat domain, which could be either a nonsense mutation or a missense mutation replacing the arginine into glutamine (R288*/Q) (**Figure 3B**). Cancer types in which these mutations were detected are glioblastoma multiforme, uterine endometrioid carcinoma, and breast invasive ductal carcinoma.

TABLE 2 | The role of ANXA3 in therapy resistance.

Therapy resistance	Years	Tumor types	Main findings	References
Platinum resistance	2020	Non-small cell lung carcinoma	Anxa3-silencing siRNAs can eliminate oxaliplatin resistance in A549 NSCLC cells	Jin et al. (2020)
	2019	Non-small cell lung carcinoma	ANXA3 secreting CAF induces cisplatin resistance in NSCLC cells A549, H661 and SK-MES-1. The putative mechanism is the activation of JNK/survivin pathway	Wang et al. (2019)
	2019	Colorectal carcinoma	Depletion of ANXA3 in oxaliplatin resistant HCT116 and SW480 colorectal cancer cells inhibits the cell viability and BrdU incorporation, increases apoptosis and diminishes migration and invasion	Xu et al. (2019)
	2015	Hepatocellular carcinoma	Increasing resistance to cisplatin in ANXA3-overexpressing tumor cells *in vitro* and in mouse xenografts *in vivo*	Pan et al. (2015b)
	2012	Ovarian carcinoma	Lower intracellular accumulation and DNA binding of cisplatin and carboplatin in ANXA3 overexpressing ovarian cancer cells, accompanied by decreased p53 levels	Yin et al. (2012)
TKIs resistance	2018	Hepatocellular carcinoma cells	ANXA3 overexpression inhibits PKC/p38-mediated apoptosis and actives p38-mediated autophagy in sorafenib-resistant HepG2 and Huh7 HCC cells as well as in patient-derived xenografts	Tong et al. (2018)
	2013	45 cell lines	ANXA3 is associated with resistance against gefitinib, sorafenib, sunitinib and lapatinib	Pénzváltó et al. (2013)
Fluoropyrimidine resistance	2015	Hepatocellular carcinoma cells	Overexpressed ANXA3 significantly enhanced the IC50 of 5-FU in both cell-line models and mouse xenografts models	Pan et al. (2015b)
	2015	Gastric cancer	SNP rs2867461 in the ANXA3 gene is significantly correlated with the sensitivity against fluoropyrimidine	Takahashi et al. (2015)
Cyclophosphamide resistance	2013	Prostate cancer	ANXA3 expression in cyclophosphamide (CPA)-resistant PC3-D3 and PC3-D4 prostate cancer cells is higher compared to the chemo-sensitive wild type PC3 cell line	Kubisch et al. (2013)
	2010			Thoenes et al. (2010)
Doxorubicin and docetaxel	2018	Triple negative breast cancer	ANXA3 knockdown promotes the uptake of doxorubicin and sensitizes response to doxorubicin and docetaxel	Du et al. (2018)

However, none of the mutations was further annotated with documented clinical implications from the literature, indicating that our understanding of their oncogenic effects is currently still highly rudimentary.

THE ROLE OF ANXA3 IN ANTICANCER DRUG RESISTANCE

Chemotherapy is the standard-of-care treatment for a majority types of tumors, yet its efficacy is frequently plagued by the evolution of fatal drug resistance or temporary, fluctuating, or partial drug responses in patients. Based on the previously described tumorigenic characteristics of ANXA3, it is well-reasoned to hypothesize that ANXA3 also play a role in attenuating the vulnerability of cancer cells against chemotherapeutic drugs. As a result, in the past decade, investigators in this field have generated a substantial body of evidence delineating the involvement of ANXA3 in the development of anticancer drug resistance (**Table 2** and **Figure 4**).

Platinum resistance. ANXA3 has been previously described as a potential marker for platinum sensitivity in patients with hepatocellular carcinoma (Pan et al., 2015b), ovarian carcinoma (Yan et al., 2010; Yin et al., 2012), non-small cell lung carcinoma (Wang et al., 2019; Jin et al., 2020) and colorectal carcinoma (Xu et al., 2019). In an attempt to ascertain the mechanism of platinum resistance in patients with hepatocellular carcinoma, researchers observed an enhanced degree of resistance to cisplatin-induced cell death in ANXA3-overexpressing tumor cells *in vitro* and in engrafted mice *in vivo* (Pan et al., 2015b). Support on this finding was provided by Yan et al., in which a significantly lower intracellular accumulation and DNA binding of cisplatin and carboplatin was detected in ANXA3 overexpressing ovarian cancer cells, accompanied by a decrement of the intracellular p53 level (Yan et al., 2010). Likewise, a substantially increased ANXA3 expression has been found in the serum of platinum-resistant ovarian cancer patients compared to the platinum sensitive group, which underscored the value of ANXA3 as a noninvasive marker for drug response prediction (Yin et al., 2012). Mechanistically, high level of ANXA3 released by CAFs has been confirmed to activate the JNK/survivin pathway in both cell-line models (A549, H661 and SK-MES-1) and mouse xenograft models of NSCLC, resulting in a markedly attenuated IC50 of cisplatin *in vitro* and an augmented xenograft tumor growth under cisplatin exposure *in vivo* (Wang et al., 2019). Conversely, abrogation of oxaliplatin resistance was achieved with ANXA3-silencing siRNAs in A549 NSCLC cells (Jin et al., 2020). Similarly, *in vitro* observations in oxaliplatin-resistant HCT116 and SW480 colorectal cancer cells further evidenced that ANXA3-knockdown has significantly impeded the phosphorylation of ERK and JNK, which led to decreased cell viability and BrdU incorporation as well as increased apoptosis and impaired migration and invasion abilities (Xu et al., 2019).

TKIs resistance. In a microarray analysis testing the sensitivity of 45 different cancer cell lines to anticancer tyrosine kinase inhibitors (TKIs) targeting the ERBB/RAS pathway, ANXA3 was indicated to be associated with resistance against multiple TKIs

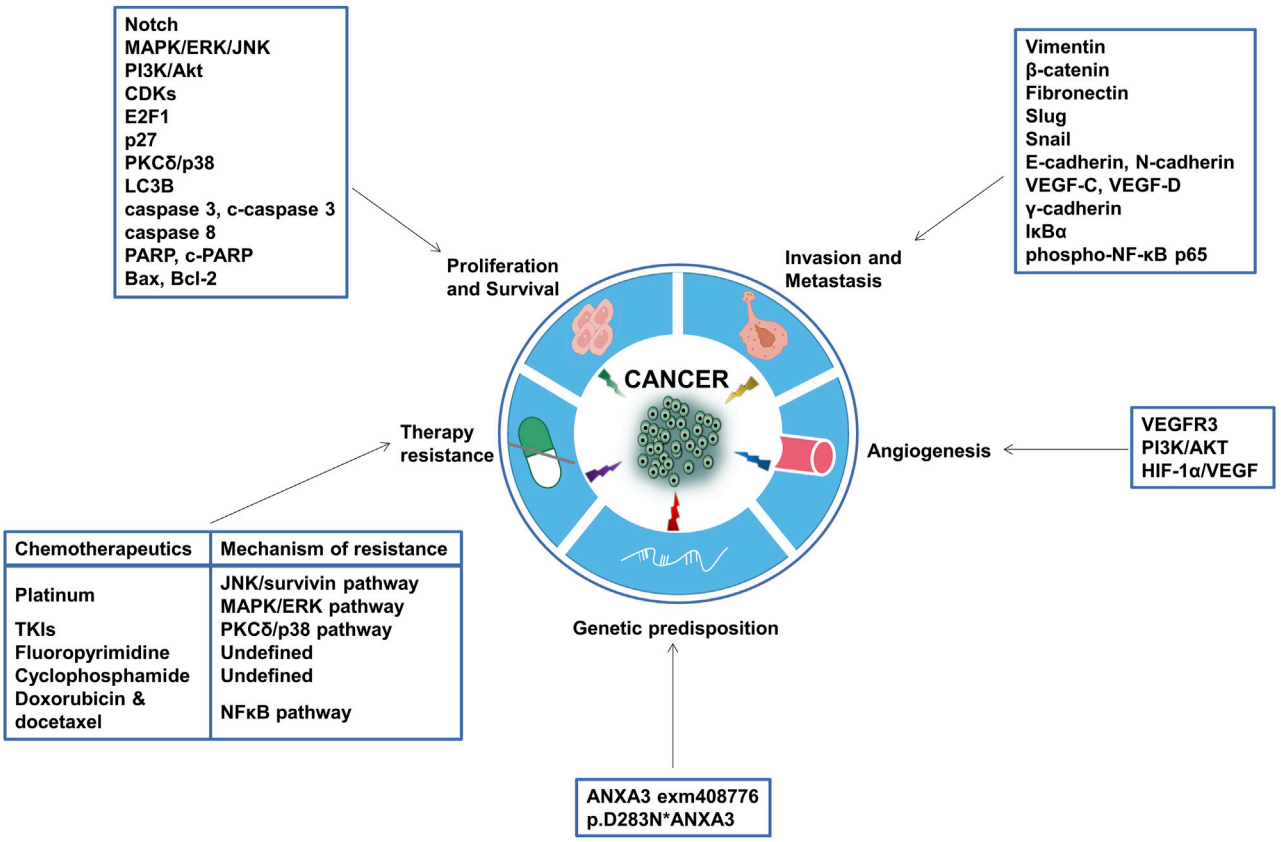


FIGURE 4 | Tumorigenic functions of ANXA3. Aberrant expression of ANXA3 can exert multifaceted tumorigenic functions, such as promoting tumor cell proliferation, invasion, metastasis and angiogenesis, enhancing the propensity to form certain inherited tumors, and inducing resistance to chemotherapeutic agents. The putative mechanisms, target genes, effector proteins and pathways through which ANXA3 modulates these processes are annotated in the square text boxes. CDKs, Cyclin-dependent kinases; p27, Cyclin-dependent kinase inhibitor 1B; p38, p38 mitogen-activated protein kinases; LC3B, Microtubule-associated proteins 1A/1B light chain 3B; PARP, Poly (ADP-ribose) polymerase; bax, Bcl-2-associated X protein; bcl-2, B-cell lymphoma 2; VEGF, Vascular endothelial growth factor; IκBα, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha; NF-κB, nuclear factor kappa-light-chain-enhancer of activated B cells; VEGFR, Vascular endothelial growth factor receptor; HIF-1α, Hypoxia-inducible factor 1-alpha.

including gefitinib, sorafenib, sunitinib, and lapatinib (Pénzváltó et al., 2013). Among these TKIs, only sorafenib resistance possesses a comprehensive *in vitro* and clinical evaluation, which is provided by the work of Tong et al. In this study, ANXA3 overexpression was detected in sorafenib-resistant HepG2 and Huh7 hepatocellular carcinoma cells as well as in patient-derived xenografts; conversely, inhibition of ANXA3 achieved with shANXA3 re-sensitized the response to sorafenib *in vitro* and limited tumor growth *in vivo* (Tong et al., 2018). The mechanistic rationale underlying this phenomenon was the inhibition of PKCδ/p38-mediated apoptosis and activation of p38-mediated autophagy, both of which are driven by a substantially elevated ANXA3 expression (Tong et al., 2018). Clinically, HCC patients with high levels of ANXA3 exhibited inferior outcomes under sorafenib treatment (Tong et al., 2018). Interestingly, investigators from this study further demonstrated that the combination of sorafenib with anti-ANXA3 monoclonal antibody effectively overcomes sorafenib resistance in both *ex vivo* organotypic cultures and patient-derived mouse xenografts (Tong et al., 2018), suggesting that targeting ANXA3 might be an actionable therapeutic approach for sorafenib-refractory HCC patients.

Fluoropyrimidine resistance. The involvement of ANXA3 in 5-fluorouracil (5-FU) resistance has been previously reported in hepatocellular carcinoma, in which enforced ANXA3 expression established by ANXA3-overexpressing lentiviruses significantly enhanced the IC50 of 5-FU in both cell-line models and mouse xenografts models (Pan et al., 2015b). Furthermore, a bioinformatics analysis on the genomic data of 119 fluoropyrimidine-treated gastric cancer patients reported that SNP rs2867461 in the ANXA3 gene exhibited a significant correlation with the sensitivity to fluoropyrimidine treatment and might therefore serve as a potential genetic biomarker in predicting the therapeutic response (Takahashi et al., 2015).

Cyclophosphamide resistance. Employing 2D-DIGE analysis, Thoenes et al. observed a significantly upregulated ANXA3 expression in cyclophosphamide (CPA)-resistant PC3-D3 and PC3-D4 prostate cancer cells compared to the chemo-sensitive

wild type PC3 cell line (Thoenes et al., 2010). Moreover, western blot analysis validated the CPA-induced upregulation of ANXA3 in the isolated tumor tissues from mice that were engrafted with PC3-wt cells and subsequently exposed to a 70-day of metronomic CPA treatment (Thoenes et al., 2010). This finding is further consistent with the data from another microarray-based study on CPA-resistant PC3 tumors (Kubisch et al., 2013).

Doxorubicin and docetaxel. The knockdown of ANXA3 has been demonstrated to promote the uptake of doxorubicin in human MDA-MB-231 breast cancer cells and 4T1 mouse mammary cancer cells (Du et al., 2018). MTT assay data from this study further indicated a substantially increased sensitivity of ANXA3-knockdown breast cancer cells to doxorubicin and docetaxel (Du et al., 2018). The reduction of the NF-κB signaling activity, which was shown to be a result of ANXA3-knockdown, was thought to be a putative mechanism underlying this improved drug response (Du et al., 2018). Despite these findings, current data is extremely limited regarding the role of ANXA3 in drug resistance against antineoplastic antibiotics and taxanes, which indicates the pressing need for further explorations in this field.

THE DIAGNOSTIC AND PROGNOSTIC VALUE OF ANXA3 IN CANCER TREATMENT AND MANAGEMENT

ANXA3 is one of the candidate biomarkers that has been extensively studied in different cancer models over the last decade. Collectively, studies have demonstrated that ANXA3 possesses potential diagnostic and prognostic value in the clinical treatment and management of an array of malignancies, including breast cancer (Pendharkar et al., 2016; Zhou et al., 2017b; Kim et al., 2018; Zhou et al., 2018), prostate cancer (Wozny et al., 2007; Köllermann et al., 2008; Schlomm et al., 2010; Guo et al., 2020), colorectal cancer (Marshall et al., 2010; Yip et al., 2010; Chang et al., 2014; Yang Q. et al., 2018) and some relatively rarer cancers such as bladder cancer (Hofmann et al., 2000), pancreatic cancer (Baine et al., 2011b) and testis cancer (Hofmann et al., 2000). In this section, we will summarize data from various studies which evaluated the clinical value of ANXA3 or provided novel clues on its diagnostic and prognostic potential.

Breast cancer. Currently, stratifying patients into their correct subtypes and discriminating between early and late stages of these subtypes remain as a prime challenge in breast cancer (Pendharkar et al., 2016). This is of considerable importance for an optimal treatment design and patient outcome, especially in those patients with aggressive phenotypes such as triple-negative and basal-like breast cancer. Moreover, multiple shortcomings of the current mammography screening method have been reported, with the main concern remains on its suboptimal sensitivity in young women with dense breasts (Arif et al., 2015). Interestingly, the expression level of ANXA3 has been reported to be different per breast cancer subtype (Pendharkar et al., 2016; Zhou et al., 2017b; Kim et al., 2018; Zhou et al., 2018). Specifically, ANXA3 expression in basal subtype of breast cancer (MDA-MB-231, HCC-70, HCC-1954) was found to be significantly higher than other subtypes (Kim et al., 2018). Other studies further confirmed that a significantly higher expression of ANXA3 was detect in TNBC cells in comparison to luminal A and B subtypes (Zhou et al., 2017b; Zhou et al., 2018). Furthermore, employing 2D-DIGE and iTRAQ approaches, Pendharkar et al. demonstrated that ANXA3 upregulation is a marker for differentiating invasive ductal carcinoma with luminal B HER2 positive (LB) and HER2 enriched (HE) subtypes as well as their early and late stages (Pendharkar et al., 2016). These findings collectively confirmed the potential of ANXA3 in improving the accuracy of subtypes differentiation and risk stratification of breast cancer.

Prostate cancer. The low explanatory power of existing clinical and histological parameters has also been reported in prostate cancer, where prognosis prediction relies solely on clinical stages, Gleason score and serum PSA and no molecular marker has been successfully translated into the routine clinical applications to date (Köllermann et al., 2008; Schlomm et al., 2010). The inferior performance of serum PSA in screening small neoplasms of initial stages (Köllermann et al., 2008) and the patient inconvenience from periodic biopsy sampling for Gleason score and pathological stage evaluation (Guo et al., 2020) further emphasize the pronounced clinical value of novel molecular biomarkers in prostate cancer. ANXA3 has been demonstrated to be downregulated in prostate cancer tissues (Wozny et al., 2007; Köllermann et al., 2008; Peraldo-Neia et al., 2011) but upregulated in the chemo-resistant PC3-D3 and PC3-D4 cell lines (Thoenes et al., 2010; Kubisch et al., 2013). Immunohistochemistry-based studies have reported a substantially less abundant ANXA3 staining in prostate cancer tissue when comparing with the surrounding epithelium and prostatic intraepithelial neoplasia (Wozny et al., 2007; Köllermann et al., 2008). In addition, the proportion of the ANXA3-negative tissue has been demonstrated to be correlated with advanced pathological stage and Gleason score as well as a reduced PSA-free survival (Köllermann et al., 2008). Using Kaplan-Meier analysis and multivariate cox regression, the study further indicated a significant association between ANXA3 staining abundance and biochemical relapse of prostate cancer (Köllermann et al., 2008). Support on this finding was provided by the work of Schlomm et al., in which ANXA3 was identified as an independent prognostic marker for the postoperative PSA recurrence (Schlomm et al., 2010). Of note, Schlomm et al. further demonstrated that incorporation of ANXA3 into the current risk stratification nomogram predicting the biochemical relapse after radical prostatectomy provided an enhancement of its predictive performance (AUC from 0.71 to 0.73) (Schlomm et al., 2010). Aside from postoperative recurrence risk, the value of ANXA3 in discriminating prostate cancer from benign tumors has also been elucidated by the work Guo et al., in which a 14-gene panel involves ANXA3 was constructed and evaluated on pre-biopsy urine and tissue specimens, displaying desirable predictive performance (AUC = 0.854) (Guo et al., 2020). Further evaluation of this gene predictor set on prospective and retrospective cohorts confirmed its ability in distinguishing low-risk patients from high-risk patients (Guo et al., 2020).

Unlike a wide spectrum of cancers overexpressing ANXA3, it is notable that the protein is downregulated in prostate cancer tissues. The putative mechanism underlying this downregulated expression pattern has been proposed in the context of autoimmune reactions in prostate cancer (Köllermann et al., 2008). Autoimmune antibodies against ANXA3 have been previously detected in the serum of prostate cancer patients (Köllermann et al., 2008), which was thought to be a result of the release of ANXA3 by prostatic epithelium cells in the form of prostasomes (Wozny et al., 2007; Köllermann et al., 2008). These granules, which are secreted by both normal and malignant prostate epithelium, have been broadly considered to harbor a strong immunogenicity (Feist et al., 2000; Minelli et al., 2005; Larsson et al., 2006). Studies have shown that, compared to patients with benign prostate hyperplasia and other noncancerous prostate disorders, prostate cancer patients carried significantly higher levels of anti-prostasome antibodies in their blood (Feist et al., 2000; Minelli et al., 2005; Larsson et al., 2006). This finding has linked anti-prostasome antibody titer to malignant transformation. It was therefore postulated that some prostate tumors could induce the formation of autoimmune antibodies against ANXA3 to neutralize its tumor-suppressive effects (Köllermann et al., 2008). This speculation also explained why decreased ANXA3 expression is associated with unfavorable clinicopathological features in prostate cancer patients (Köllermann et al., 2008). Nevertheless, the functional rationale of the antitumor activity of ANXA3 in prostate cancer, which is inconsistent to its tumorigenic roles in other malignancies, still necessitates additional experimental analyses.

Colorectal cancer. Colorectal cancer (CRC) is another well-known malignancy that can benefit most from early detection via population screening, since almost every colorectal tumor arises from a benign adenomatous polyp and is easily surgically resectable once detected (Marshall et al., 2010; Yang Q. et al., 2018). However, the unpleasant and inconvenient nature of current CRC screening procedures has led to a low compliance in screening participation (Kronborg, 2004; Marshall et al., 2010), which results in a higher incidence of advanced tumors and metastases. In an attempt searching for novel tumor-specific immunogens in CRC, mass spectrometry analyses have reproducibly detected the presence of ANXA3 in the tumor protein extracts blotted with patient's own sera (Yang Q. et al., 2018). Further verifications using immunohistochemistry and western blot assays confirmed the increased ANXA3 expression in tumor tissues compared to non-tumor tissues (Yang Q. et al., 2018). These findings collectively indicated the potential of ANXA3 as a novel serum antibody screening marker in CRC. Additionally, a series of microarray-based studies have designed novel gene predictor sets that can potentially contribute to improving the performance of the existing screening methods and overcoming the current low screening compliance (Marshall et al., 2010; Yip et al., 2010; Chang et al., 2014). A blood-based 7-gene panel (ANXA3, CLEC4D, LMNB1, PRRG4, TNFAIP6, VNN1, and IL2RB) developed by Marshall et al. exhibited good predictive performance in discriminating CRC from controls (AUC = 0.80) and in stratifying patients' current relative risk (Marshall et al., 2010). This gene predictor set was further validated in a Malaysian cohort, showing consistent results (Yip et al., 2010). In a study published in 2014, Chang et al. evaluated 17 previously described CRC-associated genes (including those from Marshall et al.) and further constructed a novel blood-based 7-gene model comprised of CpEB4, EIF2S3, MGC20553, MS4A1, ANXA3, TNFAIp6 and IL2RB, which showed a more superior performance in logistic regression analyses (AUC = 0.99) (Chang et al., 2014).

Other cancers. With the hypothesis that peripheral blood mononuclear cells (PBMCs) act as the first line of defense against early emerged neoplastic cells and thereby capable of more accurately representing the tumor biology in initial stages, Baine et al. investigated the differentially expressed genes in PBMCs of patients with pancreatic cancer (PC) and found that ANXA3 was significantly upregulated in PC patients compared to healthy controls (Baine et al., 2011b). The study further established a blood-based 7-gene predictor panel (ANXA3, SSBP2, Ube2b-rs1, CA5B, F5, TBC1D8, ARG1, and ADAMTS20) with a sensitivity of 83% and specificity of 75% in discriminating PC patients (Baine et al., 2011b). In a LC−MRM/MS analysis of urine samples from 30 bladder cancer patients and 89 non-cancer patients, ANXA3 was most frequently detected in bladder cancer samples (Tsai et al., 2018). Validation using western blot analysis further confirmed the elevated level of ANXA3 in both urine and tumor specimens from bladder cancer patients (Tsai et al., 2018). In another bioinformatics study based on previously published microarray data, a significantly augmented expression of the ANXA3 gene was detected in testicular carcinoma *in situ* (CIS) samples and further validated using RT-qPCR, suggesting the potential of ANXA3 serving as a novel clinicopathological marker for testicular CIS (Almstrup et al., 2007).

CONCLUSION

Cancer is a major public health issue and a highly lethal disease worldwide. The high mortality rate of many cancers can be attributed to the evasion of detection of benign or preneoplastic tumors, the occurrence of intrinsic or acquired therapy resistance, and the suboptimal predictive power of the routinely used screening techniques and clinicopathological parameters. Emerging evidence has suggested that ANXA3 might be a promising biomarker candidate and an attractive therapeutic target that harbors the potential to address these obstacles. According to the gathered data in this review, differential expression of ANXA3 has been demonstrated in a wide array of cancers, which is capable of sustaining cell proliferation signaling, promoting invasion and metastasis, inducing angiogenesis and resistance to various chemotherapeutic agents. Genetic alterations in the ANXA3 gene have also been reported to be associated with an enhanced tumor susceptibility. As a candidate molecular marker, ANXA3 exhibited pronounced clinical value in risk stratification, early detection, patient differentiation and active surveillance. Nonetheless, some limitations to the current findings and conclusions still need to be noted. Despite the substantial body of evidence on its tumorigenic role, the precise upstream and downstream signaling transduction of ANXA3 remains incompletely elucidated, and there is a lack of mechanistic studies on the role of ANXA3 in highly prevalent

cancers. Besides, the underlying mechanisms of ANXA3-induced chemoresistance are not fully understood, and investigations so far have not covered all chemotherapeutic drugs, nor other anticancer therapies, such as radiotherapy. Furthermore, most of the studies evaluating the diagnostic and prognostic value of ANXA3 suffer from small sample sizes, which indicates that more conclusive data from larger patient cohorts are still urgently awaited. Therefore, the eventual clinical implementation of ANXA3 as a therapeutic target or as a diagnostic or prognostic biomarker still requires further investigations to 1) elucidate the complete picture of the upstream and downstream signaling pathways of ANXA3; 2) obtain deeper insights into the mechanisms underlying ANXA3-induced chemoresistance as well as the role of ANXA3 in other types of anticancer therapy resistance such as radioresistance; and 3) evaluate the clinical significance of ANXA3 as a diagnostic or prognostic biomarker on larger patient cohorts.

AUTHOR CONTRIBUTIONS

LY and PL wrote initial article, XY and KL provided helpful comments on this article and SQ conceived the idea and supervised the project. All the authors have read and approved the final version of this article. LY and PL contributed equally to this work as co-first authors.

FUNDING

This work was supported by the Development and Promotion Project of Guangxi Medical and Health Appropriate Technology (S2018038), the Key Laboratory of High-Incidence-Tumor Prevention and Treatment (Guangxi Medical University) (GK 2018-08, GKE-ZZ202012) and the General Program of Guangxi Natural Science Foundation (2018GXNSFAA281032).

REFERENCES

AlHarthi, F. S., Qari, A., Edress, A., and Abedalthagafi, M. (2020). Familial/inherited Cancer Syndrome: a Focus on the Highly Consanguineous Arab Population. *NPJ Genom. Med.* 5, 3. doi:10.1038/s41525-019-0110-y

Almstrup, K., Leffers, H., Lothe, R. A., Skakkebæk, N. E., Sonne, S. B., Nielsen, J. E., et al. (2007). Improved Gene Expression Signature of Testicular Carcinoma *In Situ*. *Int. J. Androl.* 30 (4), 292–303. discussion 303. doi:10.1111/j.1365-2605.2007.00758.x

Aravind Kumar, M., Singh, V., Naushad, S. M., Shanker, U., and Lakshmi Narasu, M. (2018). Microarray-based SNP Genotyping to Identify Genetic Risk Factors of Triple-Negative Breast Cancer (TNBC) in South Indian Population. *Mol. Cel Biochem.* 442 (1-2), 1–10. doi:10.1007/s11010-017-3187-6

Arif, S., Qudsia, S., Urooj, S., Chaudry, N., Arshad, A., and Andleeb, S. (2015). Blueprint of Quartz crystal Microbalance Biosensor for Early Detection of Breast Cancer through Salivary Autoantibodies against ATP6AP1. *Biosens. Bioelectron.* 65, 62–70. doi:10.1016/j.bios.2014.09.088

Baine, M. J., Chakraborty, S., Smith, L. M., Mallya, K., Sasson, A. R., Brand, R. E., et al. (2011a). Transcriptional Profiling of Peripheral Blood Mononuclear Cells in Pancreatic Cancer Patients Identifies Novel Genes with Potential Diagnostic Utility. *PLoS One* 6 (2), e17014. doi:10.1371/journal.pone.0017014

Baine, M. J., Menning, M., Smith, L. M., Mallya, K., Kaur, S., Rachagani, S., et al. (2012b). Differential Gene Expression Analysis of Peripheral Blood Mononuclear Cells Reveals Novel Test for Early Detection of Pancreatic Cancer. *Cancer Biomark* 11 (1), 1–14. doi:10.3233/cbm-2012-0260

Bianchi, C., Bombelli, S., Raimondo, F., Torsello, B., Angeloni, V., Ferrero, S., et al. (2010). Primary Cell Cultures from Human Renal Cortex and Renal-Cell Carcinoma Evidence a Differential Expression of Two Spliced Isoforms of Annexin A3. *Am. J. Pathol.* 176 (4), 1660–1670. doi:10.2353/ajpath.2010.090402

Chang, Y.-T., Yao, C. T., Su, S. L., Chou, Y. C., Chu, C. M., Huang, C. S., et al. (2014). Verification of Gene Expression Profiles for Colorectal Cancer Using 12 Internet Public Microarray Datasets. *World J. Gastroenterol.* 20 (46), 17476–17482. doi:10.3748/wjg.v20.i46.17476

Du, R., Liu, B., Zhou, L., Wang, D., He, X., Xu, X., et al. (2018). Downregulation of Annexin A3 Inhibits Tumor Metastasis and Decreases Drug Resistance in Breast Cancer. *Cell Death Dis.* 9 (2), 126. doi:10.1038/s41419-017-0143-z

Favier-Perron, B., Lewit-Bentley, A., and Russo-Marie, F. (1996). The High-Resolution Crystal Structure of Human Annexin III Shows Subtle Differences with Annexin V†,‡. *Biochemistry* 35 (6), 1740–1744. doi:10.1021/bi952092o

Feist, E., Dörner, T., Kuckelkorn, U., Scheffler, S., Burmester, G.-R., and Kloetzel, P.-M. (2000). Diagnostic Importance of Anti-proteasome Antibodies. *Int. Arch. Allergy Immunol.* 123 (1), 92–97. doi:10.1159/000024427

Gerke, V., Creutz, C. E., and Moss, S. E. (2005). Annexins: Linking Ca2+ Signalling to Membrane Dynamics. *Nat. Rev. Mol. Cel Biol.* 6 (6), 449–461. doi:10.1038/nrm1661

Gerke, V., and Moss, S. E. (2002). Annexins: from Structure to Function. *Physiol. Rev.* 82 (2), 331–371. doi:10.1152/physrev.00030.2001

Guo, J., Liu, D., Zhang, X., Johnson, H., Feng, X., Zhang, H., et al. (2020). Establishing a Urine-Based Biomarker Assay for Prostate Cancer Risk Stratification. *Front. Cel Dev. Biol.* 8, 597961. doi:10.3389/fcell.2020.597961

Guo, L., Zhang, K., and Bing, Z. (2017). Application of a Co-expression Network for the Analysis of Aggressive and Non-aggressive Breast Cancer Cell Lines to Predict the Clinical Outcome of Patients. *Mol. Med. Rep.* 16 (6), 7967–7978. doi:10.3892/mmr.2017.7608

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 (12), e82241. doi:10.1371/journal.pone.0082241

Hofmann, A., Raguénès-Nicol, C., Favier-Perron, B., Mesonero, J., Huber, R., Russo-Marie, F., et al. (2000). The Annexin A3–Membrane Interaction Is Modulated by an N-Terminal Tryptophan†. *Biochemistry* 39 (26), 7712–7721. doi:10.1021/bi992359+

Ibrahim, S. A., Yip, G. W., Stock, C., Pan, J.-W., Neubauer, C., Poeter, M., et al. (2012). Targeting of Syndecan-1 by microRNA miR-10b Promotes Breast Cancer Cell Motility and Invasiveness via a Rho-GTPase- and E-cadherin-dependent Mechanism. *Int. J. Cancer* 131 (6), E884–E896. doi:10.1002/ijc.27629

Jiang, J. N., Hong, Q. Y., Gao, H. J., Lai, B. L., Lan, J. F., and Yang, Q. (2019). Lnc00908 Promotes the Development of Ovarian Cancer by Regulating microRNA-495-5p. *Eur. Rev. Med. Pharmacol. Sci.* 23 (4), 1388–1396. doi:10.26355/eurrev_201902_17095

Jin, Y. F., Huang, Y. T., and Chen, P. F. (2020). ANXA3 Deletion Inhibits the Resistance of Lung Cancer Cells to Oxaliplatin. *Eur. Rev. Med. Pharmacol. Sci.* 24 (7), 3741–3748. doi:10.26355/eurrev_202004_20838

Jung, E.-J., Moon, H.-G., Park, S.-T., Cho, B.-I., Lee, S.-M., Jeong, C.-Y., et al. (2010). Decreased Annexin A3 Expression Correlates with Tumor Progression in Papillary Thyroid Cancer. *Prot. Clin. Appl.* 4 (5), 528–537. doi:10.1002/prca.200900063

Kim, J.-Y., Jung, E. J., Park, H. J., Lee, J.-H., Song, E. J., Kwag, S.-J., et al. (2018). Tumor-Suppressing Effect of Silencing of Annexin A3 Expression in Breast Cancer. *Clin. Breast Cancer* 18 (4), e713–e719. doi:10.1016/j.clbc.2017.11.009

Köllermann, J., Schlomm, T., Bang, H., Schwall, G. P., von Eichel-Streiber, C., Simon, R., et al. (2008). Expression and Prognostic Relevance of Annexin A3 in Prostate Cancer. *Eur. Urol.* 54 (6), 1314–1323. doi:10.1016/j.eururo.2008.01.001

Kronborg, O. (2004). Colon Polyps and Cancer. *Endoscopy* 36 (1), 3–7. doi:10.1055/s-2004-814120

Kubisch, R., Meissner, L., Krebs, S., Blum, H., Günther, M., Roidl, A., et al. (2013). A Comprehensive Gene Expression Analysis of Resistance Formation upon

Metronomic Cyclophosphamide Therapy. *Translational Oncol.* 6 (1), 1-IN3. doi:10.1593/tlo.12295

Larsson, A., Ronquist, G., Wülfing, C., Eltze, E., Bettendorf, O., Carlsson, L., et al. (2006). Antiprostasome Antibodies: Possible Serum Markers for Prostate Cancer Metastasizing Liability. *Urol. Oncol. Semin. Original Invest.* 24 (3), 195–200. doi:10.1016/j.urolonc.2005.07.009

Le Cabec, V., and Maridonneau-Parini, I. (1994). Annexin 3 Is Associated with Cytoplasmic Granules in Neutrophils and Monocytes and Translocates to the Plasma Membrane in Activated Cells. *Biochem. J.* 303 (Pt 2), 481–487. doi:10.1042/bj3030481

Li, J., Zhou, T., Liu, L., Ju, Y. C., Chen, Y. T., Tan, Z. R., et al. (2018). The Regulatory Role of Annexin 3 in a Nude Mouse Bearing a Subcutaneous Xenograft of MDA-MB-231 Human Breast Carcinoma. *Pathol. - Res. Pract.* 214 (10), 1719–1725. doi:10.1016/j.prp.2018.09.009

Liu, Q., Wang, S., Pei, G., Yang, Y., Min, X., Huang, Y., et al. (2021). Impact Analysis of miR-1253 on Lung Cancer Progression through Targeted Regulation of ANXA3. *Cancer Manag. Res.* 13, 1767–1776. doi:10.2147/cmar.S251679

Liu, Y., Xiao, Z., Li, M., Li, M., Zhang, P., Li, C., et al. (2009). Quantitative Proteome Analysis Reveals Annexin A3 as a Novel Biomarker in Lung Adenocarcinoma. *J. Pathol.* 217 (1), 54–64. doi:10.1002/path.2429

Lohinai, Z., Megyesfalvi, Z., Suda, K., Harko, T., Ren, S., Moldvay, J., et al. (2019). Comparative Expression Analysis in Small Cell Lung Carcinoma Reveals Neuroendocrine Pattern Change in Primary Tumor versus Lymph Node Metastases. *Transl Lung Cancer Res.* 8 (6), 938–950. doi:10.21037/tlcr.2019.11.30

Madoz-Gúrpide, J., López-Serra, P., Martínez-Torrecuadrada, J. L., Sánchez, L., Lombardía, L., and Casal, J. I. (2006). Proteomics-based Validation of Genomic Data. *Mol. Cell Proteomics* 5 (8), 1471–1483. doi:10.1074/mcp.M600048-MCP200

Marshall, K. W., Mohr, S., Khettabi, F. E., Nossova, N., Chao, S., Bao, W., et al. (2009). A Blood-Based Biomarker Panel for Stratifying Current Risk for Colorectal Cancer. *Int. J. Cancer* 126 (5), NA. doi:10.1002/ijc.24910

Minelli, A., Ronquist, G., Carlsson, L., Mearini, E., Nilsson, O., and Larsson, A. (2005). Antiprostasome Antibody Titres in Benign and Malignant Prostate Disease. *Anticancer Res.* 25 (6c), 4399–4402.

Mohr, T., Haudek-Prinz, V., Slany, A., Grillari, J., Micksche, M., and Gerner, C. (2017). Proteome Profiling in IL-1β and VEGF-Activated Human Umbilical Vein Endothelial Cells Delineates the Interlink between Inflammation and Angiogenesis. *PLoS One* 12 (6), e0179065. doi:10.1371/journal.pone.0179065

Moss, S. E., and Morgan, R. O. (2004). The Annexins. *Genome Biol.* 5 (4), 219. doi:10.1186/gb-2004-5-4-219

Mussunoor, S., and Murray, G. (2008). The Role of Annexins in Tumour Development and Progression. *J. Pathol.* 216 (2), 131–140. doi:10.1002/path.2400

Pan, Q.-Z., Pan, K., Wang, Q.-J., Weng, D.-S., Zhao, J.-J., Zheng, H.-X., et al. (2015a). Annexin A3 as a Potential Target for Immunotherapy of Liver Cancer Stem-like Cells. *Stem Cells* 33 (2), 354–366. doi:10.1002/stem.1850

Pan, Q.-Z., Pan, K., Weng, D.-S., Zhao, J.-J., Zhang, X.-F., Wang, D.-D., et al. (2015b). Annexin A3 Promotes Tumorigenesis and Resistance to Chemotherapy in Hepatocellular Carcinoma. *Mol. Carcinog.* 54 (8), 598–607. doi:10.1002/mc.22126

Park, J. E., Lee, D. H., Lee, J. A., Park, S. G., Kim, N.-S., Park, B. C., et al. (2005). Annexin A3 Is a Potential Angiogenic Mediator. *Biochem. Biophysical Res. Commun.* 337 (4), 1283–1287. doi:10.1016/j.bbrc.2005.10.004

Pendharkar, N., Gajbhiye, A., Taunk, K., RoyChoudhury, S., Dhali, S., Seal, S., et al. (2016). Quantitative Tissue Proteomic Investigation of Invasive Ductal Carcinoma of Breast with Luminal B HER2 Positive and HER2 Enriched Subtypes towards Potential Diagnostic and Therapeutic Biomarkers. *J. Proteomics* 132, 112–130. doi:10.1016/j.jprot.2015.11.024

Pénzváltó, Z., Tegze, B., Szász, A. M., Sztupinszki, Z., Likó, I., Szendrői, A., et al. (2013). Identifying Resistance Mechanisms against Five Tyrosine Kinase Inhibitors Targeting the ERBB/RAS Pathway in 45 Cancer Cell Lines. *PLoS One* 8 (3), e59503. doi:10.1371/journal.pone.0059503

Peraldo-Neia, C., Migliardi, G., Mello-Grand, M., Montemurro, F., Segir, R., Pignochino, Y., et al. (2011). Epidermal Growth Factor Receptor (EGFR) Mutation Analysis, Gene Expression Profiling and EGFR Protein Expression in Primary Prostate Cancer. *BMC Cancer* 11, 31. doi:10.1186/1471-2407-11-31

Perron, B., Lewit-Bentley, A., Geny, B., and Russo-Marie, F. (1997). Can Enzymatic Activity, or Otherwise, Be Inferred from Structural Studies of Annexin III?. *J. Biol. Chem.* 272 (17), 11321–11326. doi:10.1074/jbc.272.17.11321

Raynal, P., and Pollard, H. B. (1994). Annexins: the Problem of Assessing the Biological Role for a Gene Family of Multifunctional Calcium- and Phospholipid-Binding Proteins. *Biochim. Biophys. Acta (Bba) - Rev. Biomembranes* 1197 (1), 63–93. doi:10.1016/0304-4157(94)90019-1

Rho, J.-H., Roehrl, M. H. A., and Wang, J. Y. (2009). Glycoproteomic Analysis of Human Lung Adenocarcinomas Using Glycoarrays and Tandem Mass Spectrometry: Differential Expression and Glycosylation Patterns of Vimentin and Fetuin A Isoforms. *Protein J.* 28 (3-4), 148–160. doi:10.1007/s10930-009-9177-0

Ruan, L., Wang, G.-L., Chen, Y., Yi, H., Tang, C.-E., Zhang, P.-F., et al. (2010). Identification of Tyrosine Phosphoproteins in Signaling Pathway Triggered TGF-A by Using Functional Proteomics Technology. *Med. Oncol.* 27 (4), 1407–1414. doi:10.1007/s12032-009-9394-6

Sarquis, M., Moraes, D. C., Bastos-Rodrigues, L., Azevedo, P. G., Ramos, A. V., Reis, F. V., et al. (2020). Germline Mutations in Familial Papillary Thyroid Cancer. *Endocr. Pathol.* 31 (1), 14–20. doi:10.1007/s12022-020-09607-4

Schlomm, T., Chun, F. K., and Erbersdobler, A. (2010). From Gene to Clinic: TMA-Based Clinical Validation of Molecular Markers in Prostate Cancer. *Methods Mol. Biol.* 664, 177–189. doi:10.1007/978-1-60761-806-5_17

Schulz, P., Fischer, C., Detjen, K. M., Rieke, S., Hilfenhaus, G., Marschall, Z., et al. (2011). Angiopoietin-2 Drives Lymphatic Metastasis of Pancreatic Cancer. *FASEB j.* 25 (10), 3325–3335. doi:10.1096/fj.11-182287

Sopkova, J., Raguenes-Nicol, C., Vincent, M., Chevalier, A., Lewit-Bentley, A., Russo-Marie, F., et al. (2002). Ca2+ and Membrane Binding to Annexin 3 Modulate the Structure and Dynamics of its N Terminus and Domain III. *Protein Sci.* 11 (7), 1613–1625. doi:10.1110/ps.4230102

Sun, Y., Wu, C., Ma, J., Yang, Y., Man, X., Wu, H., et al. (2016). Toll-like Receptor 4 Promotes Angiogenesis in Pancreatic Cancer via PI3K/AKT Signaling. *Exp. Cel Res.* 347 (2), 274–282. doi:10.1016/j.yexcr.2016.07.009

Swairjo, M. A., and Seaton, B. A. (1994). Annexin Structure and Membrane Interactions: a Molecular Perspective. *Annu. Rev. Biophys. Biomol. Struct.* 23, 193–213. doi:10.1146/annurev.bb.23.060194.001205

Takahashi, H., Kaniwa, N., Saito, Y., Sai, K., Hamaguchi, T., Shirao, K., et al. (2015). Construction of Possible Integrated Predictive index Based on EGFR and ANXA3 Polymorphisms for Chemotherapy Response in Fluoropyrimidine-Treated Japanese Gastric Cancer Patients Using a Bioinformatic Method. *BMC Cancer* 15, 718. doi:10.1186/s12885-015-1721-z

Thoenes, L., Hoehn, M., Kashirin, R., Ogris, M., Arnold, G. J., Wagner, E., et al. (2010). *In Vivo* chemoresistance of Prostate Cancer in Metronomic Cyclophosphamide Therapy. *J. Proteomics* 73 (7), 1342–1354. doi:10.1016/j.jprot.2010.02.019

Tong, M., Che, N., Zhou, L., Luk, S. T., Kau, P. W., Chai, S., et al. (2018). Efficacy of Annexin A3 Blockade in Sensitizing Hepatocellular Carcinoma to Sorafenib and Regorafenib. *J. Hepatol.* 69 (4), 826–839. doi:10.1016/j.jhep.2018.05.034

Tong, M., Fung, T.-M., Luk, S. T., Ng, K.-Y., Lee, T. K., Lin, C.-H., et al. (2015). ANXA3/JNK Signaling Promotes Self-Renewal and Tumor Growth, and its Blockade Provides a Therapeutic Target for Hepatocellular Carcinoma. *Stem Cel Rep.* 5 (1), 45–59. doi:10.1016/j.stemcr.2015.05.013

Tsai, C.-H., Chen, Y.-T., Chang, Y.-H., Hsueh, C., Liu, C.-Y., Chang, Y.-S., et al. (2018). Systematic Verification of Bladder Cancer-Associated Tissue Protein Biomarker Candidates in Clinical Urine Specimens. *Oncotarget* 9 (56), 30731–30747. doi:10.18632/oncotarget.24578

Wan, X., Guo, D., Zhu, Q., and Qu, R. (2020). microRNA-382 Suppresses the Progression of Pancreatic Cancer through the PI3K/Akt Signaling Pathway by Inhibition of Anxa3. *Am. J. Physiol. Gastrointest. Liver Physiol.* 319 (3), G309–g322. doi:10.1152/ajpgi.00322.2019

Wang, K., and Li, J. (2016). Overexpression of ANXA3 Is an Independent Prognostic Indicator in Gastric Cancer and its Depletion Suppresses Cell Proliferation and Tumor Growth. *Oncotarget* 7 (52), 86972–86984. doi:10.18632/oncotarget.13493

Wang, L., Li, X., Ren, Y., Geng, H., Zhang, Q., Cao, L., et al. (2019). Cancer-associated Fibroblasts Contribute to Cisplatin Resistance by modulatingANXA3 in Lung Cancer Cells. *Cancer Sci.* 110 (5), 1609–1620. doi:10.1111/cas.13998

Webber, J. L. (2010). Regulation of Autophagy by P38α MAPK. *Autophagy* 6 (2), 292–293. doi:10.4161/auto.6.2.11128

Webber, J. L., and Tooze, S. A. (2010). Coordinated Regulation of Autophagy by P38α MAPK through mAtg9 and p38IP. *Embo j* 29 (1), 27–40. doi:10.1038/emboj.2009.321

Wozny, W., Schroer, K., Schwall, G. P., Poznanović, S., Stegmann, W., Dietz, K., et al. (2007). Differential Radioactive Quantification of Protein Abundance Ratios between Benign and Malignant Prostate Tissues: Cancer Association of Annexin A3. *Proteomics* 7 (2), 313–322. doi:10.1002/pmic.200600646

Wu, C., Song, G., Xiong, G., Lu, L., Zhou, X., and Duan, Y. (2018). Genome-Wide Analysis of mRNA Expression Profiling Identified Lung Cancer-Related Gene in Xuanwei, China. *Clin. Lab.* 64 (9), 1517–1526. doi:10.7754/Clin.Lab.2018.180402

Xu, R., Yin, J., Zhang, Y., and Zhang, S. (2019). Annexin A3 Depletion Overcomes Resistance to Oxaliplatin in Colorectal Cancer via the MAPK Signaling Pathway. *J. Cel Biochem.* 120 (9), 14585–14593. doi:10.1002/jcb.28720

Xu, Z., Zhu, C., Chen, C., Zong, Y., Feng, H., Liu, D., et al. (2018). CCL19 Suppresses Angiogenesis through Promoting miR-206 and Inhibiting Met/ERK/Elk-1/HIF-1α/VEGF-A Pathway in Colorectal Cancer. *Cel Death Dis.* 9 (10), 974. doi:10.1038/s41419-018-1010-2

Yan, X., Yin, J., Yao, H., Mao, N., Yang, Y., and Pan, L. (2010). Increased Expression of Annexin A3 Is a Mechanism of Platinum Resistance in Ovarian Cancer. *Cancer Res.* 70 (4), 1616–1624. doi:10.1158/0008-5472.CAN-09-3215

Yang, L., Men, W. L., Yan, K. M., Tie, J., Nie, Y. Z., and Xiao, H. J. (2018a). MiR-340-5p Is a Potential Prognostic Indicator of Colorectal Cancer and Modulates ANXA3. *Eur. Rev. Med. Pharmacol. Sci.* 22 (15), 4837–4845. doi:10.26355/eurrev_201808_15619

Yang, Q., Roehrl, M. H., and Wang, J. Y. (2018b). Proteomic Profiling of Antibody-Inducing Immunogens in Tumor Tissue Identifies PSMA1, LAP3, ANXA3, and Maspin as colon Cancer Markers. *Oncotarget* 9 (3), 3996–4019. doi:10.18632/oncotarget.23583

Yin, J., Yan, X., Yao, X., Zhang, Y., Shan, Y., Mao, N., et al. (2012). Secretion of Annexin A3 from Ovarian Cancer Cells and its Association with Platinum Resistance in Ovarian Cancer Patients. *J. Cel Mol Med.* 16 (2), 337–348. doi:10.1111/j.1582-4934.2011.01316.x

Yip, K.-T., Das, P. K., Suria, D., Lim, C.-R., Ng, G.-H., and Liew, C.-C. (2010). A Case-Controlled Validation Study of a Blood-Based Seven-Gene Biomarker Panel for Colorectal Cancer in Malaysia. *J. Exp. Clin. Cancer Res.* 29, 128. doi:10.1186/1756-9966-29-128

Zhang, P.-C., Liu, X., Li, M.-M., Ma, Y.-Y., Sun, H.-T., Tian, X.-Y., et al. (2020a). AT-533, a Novel Hsp90 Inhibitor, Inhibits Breast Cancer Growth and HIF-1α/VEGF/VEGFR-2-mediated Angiogenesis *In Vitro* and *In Vivo*. *Biochem. Pharmacol.* 172, 113771. doi:10.1016/j.bcp.2019.113771

Zhang, W., Xiong, Z., Wei, T., Li, Q., Tan, Y., Ling, L., et al. (2018). Nuclear Factor 90 Promotes Angiogenesis by Regulating HIF-1α/VEGF-A Expression through the PI3K/Akt Signaling Pathway in Human Cervical Cancer. *Cel Death Dis.* 9 (3), 276. doi:10.1038/s41419-018-0334-2

Zhang, Z., Deng, M., Huang, J., Wu, J., Li, Z., Xing, M., et al. (2020b). Microglial Annexin A3 Downregulation Alleviates Bone Cancer-Induced Pain through Inhibiting the Hif-1α/vascular Endothelial Growth Factor Signaling Pathway. *Pain* 161 (12), 2750–2762. doi:10.1097/j.pain.0000000000001962

Zhou, T., Liu, S., Yang, L., Ju, Y., and Li, C. (2018). The Expression of ANXA3 and its Relationship with the Occurrence and Development of Breast Cancer. *J. Buon* 23 (3), 713–719.

Zhou, T., Li, Y., Yang, L., Liu, L., Ju, Y., and Li, C. (2017a). Silencing of ANXA3 Expression by RNA Interference Inhibits the Proliferation and Invasion of Breast Cancer Cells. *Oncol. Rep.* 37 (1), 388–398. doi:10.3892/or.2016.5251

Zhou, T., Li, Y., Yang, L., Tang, T., Zhang, L., and Shi, J. (2017b). Annexin A3 as a Prognostic Biomarker for Breast Cancer: A Retrospective Study. *Biomed. Res. Int.* 2017, 1–7. doi:10.1155/2017/2603685

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.

Copyright © 2021 Yang, Lu, Yang, Li and Qu. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 21 July 2021
doi: 10.3389/fmolb.2021.688298

Identification of a Prognostic Signature Associated With the Homeobox Gene Family for Bladder Cancer

*Bingqi Dong[1†], Jiaming Liang[2†], Ding Li[3†], Wenping Song[3], Jinbo Song[1], Mingkai Zhu[1], Shiming Zhao[1], Yongkang Ma[1] and Tiejun Yang[1]**

[1]Department of Urology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [3]Department of Pharmacy, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Florence Le Calvez-Kelm,
International Agency for Research on Cancer (IARC), France
Ajay Dixit,
University of Minnesota Twin Cities, United States

***Correspondence:**
Tiejun Yang
tiejunyang@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 30 March 2021
Accepted: 23 June 2021
Published: 21 July 2021

Citation:
Dong B, Liang J, Li D, Song W, Song J, Zhu M, Zhao S, Ma Y and Yang T (2021) Identification of a Prognostic Signature Associated With the Homeobox Gene Family for Bladder Cancer. Front. Mol. Biosci. 8:688298. doi: 10.3389/fmolb.2021.688298

Background: Bladder cancer (BLCA) is a common malignant tumor of the genitourinary system, and there is a lack of specific, reliable, and non-invasive tumor biomarker tests for diagnosis and prognosis evaluation. Homeobox genes play a vital role in BLCA tumorigenesis and development, but few studies have focused on the prognostic value of homeobox genes in BLCA. In this study, we aim to develop a prognostic signature associated with the homeobox gene family for BLCA.

Methods: The RNA sequencing data, clinical data, and probe annotation files of BLCA patients were downloaded from the Gene Expression Omnibus database and the University of California, Santa Cruz (UCSC), Xena Browser. First, differentially expressed homeobox gene screening between tumor and normal samples was performed using the "limma" and robust rank aggregation (RRA) methods. The mutation data were obtained with the "TCGAmutation" package and visualized with the "maftools" package. Kaplan–Meier curves were plotted with the "survminer" package. Then, a signature was constructed by logistic regression analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using "clusterProfiler." Furthermore, the infiltration level of each immune cell type was estimated using the single-sample gene set enrichment analysis (ssGSEA) algorithm. Finally, the performance of the signature was evaluated by receiver-operating characteristic (ROC) curve and calibration curve analyses.

Results: Six genes were selected to construct this prognostic model: TSHZ3, ZFHX4, ZEB2, MEIS1, ISL1, and HOXC4. We divided the BLCA cohort into high- and low-risk groups based on the median risk score calculated with the novel signature. The overall survival (OS) rate of the high-risk group was significantly lower than that of the low-risk group. The infiltration levels of almost all immune cells were significantly higher in the high-risk group than in the low-risk group. The average risk score for the group that responded to immunotherapy was significantly lower than that of the group that did not.

Conclusion: We constructed a risk prediction signature with six homeobox genes, which showed good accuracy and consistency in predicting the patient's prognosis and response to immunotherapy. Therefore, this signature can be a potential biomarker and treatment target for BLCA patients.

Keywords: bladder cancer, homeobox gene family, prognostic signature, immunotherapy, biomarkers

INTRODUCTION

Bladder cancer (BLCA) is a common urological tumor, and its morbidity and mortality rates are increasing year by year (Siegel et al., 2019). High recurrence and early metastasis lead to the poor prognosis of BLCA. The detection of exfoliated tumor cells in urine or bladder lavage samples has a high sensitivity (84%) for the diagnosis of high-grade BLCA but is less sensitive for low-grade BLCA (Babjuk et al., 2019). Cystoscopy, the main method for the diagnosis of BLCA, is invasive, time-consuming, and tedious. Currently, specific, reliable, and non-invasive tumor biomarker tests for the diagnosis and prognosis evaluation of BLCA are desperately needed.

The homeobox gene family is a group with a homologous segment of approximately 180 bp in length that encodes a homologous domain of 60 amino acids and is an important transcriptional regulator that plays a vital role in tumor formation and development, regulating cell proliferation, migration, and apoptosis (Laughon and Scott, 1984; Srebrow et al., 1998; Yang et al., 2015). Current studies have shown that the homeobox gene family is aberrantly expressed in different tumors, such as bladder, bile duct, endometrial, and breast cancers (Rao et al., 2002). In BLCA, ISL1 and LHX5 play important roles in multiple stages of bladder tumorigenesis (Akhir et al., 2020); ZHX3 promotes migration and invasion *in vitro* and *in vivo* (Deng et al., 2021). Therefore, the homeobox gene family plays an important role in the development and progression of BLCA. Although progress has been made in the study of individual family members, the role and prognostic value of the homeobox gene family in BLCA remain unclear.

In this study, we analyzed the mRNA expression of a large number of BLCA samples in public databases [The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO)]. We constructed a prognostic signature for BLCA based on six homeobox genes with significant differential expression between BLCA tissues and normal tissues. This signature can predict a patient's prognosis and response to immunotherapy and thus has good clinical application value. The design flow chart for the entire analysis process of this study is shown in **Figure 1**.

MATERIALS AND METHODS

Data Collection

The RNA sequencing (RNA-Seq) data, clinical data, and probe annotation files of BLCA patients (providing 18 normal tissues and 406 tumor tissues) in TCGA were downloaded from the University of California, Santa Cruz (UCSC), Xena Browser (https://xenabrowser.net/). BLCA datasets GSE7476 (3 normal tissues and 9 tumor tissues), GSE13507 (69 normal tissues and 188 tumor tissues), GSE37815 (6 normal tissues and 18 tumor tissues), GSE65635 (4 normal tissues and 8 normal tissues), and GSE19423 (48 tumor tissues) were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) using the R package "GEOquery" (Davis and Meltzer, 2007). All 344 homeobox gene family members were extracted from the Hugo Gene Nomenclature Committee (HGNC). The

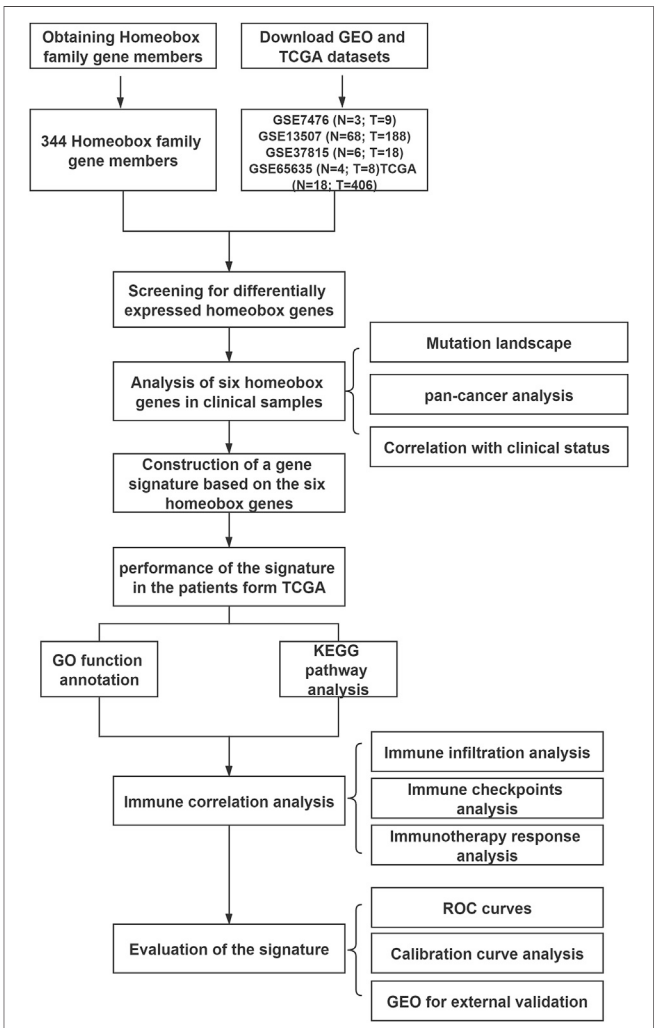


FIGURE 1 | Flow chart showing the design of the study, with GSE7476 (*N* = 3; T = 9), GSE13507 (*N* = 68; *T* = 188), GSE37815 (*N* = 6; *T* = 18), GSE65635 (*N* = 4; T = 8), and TCGA (*N* = 18; *T* = 406) datasets.

probe IDs in each BLCA dataset were transformed into gene symbols according to the annotation files.

Identification and Integration of Differentially Expressed Genes

The R package "limma" was used to identify DEGs between normal and tumor tissues in each BLCA cohort with cutoff criteria of adjusted p value <0.05 and |log fold change (FC)| > 0.5 (Ritchie et al., 2015). DEGs acquired from the five BLCA cohorts were sorted by the log fold change (logFC) value, and then the five gene lists were integrated using the RobustRankAggreg (RRA) R package (Kolde et al., 2012). The RRA method is based on the assumption that if the gene rank is high in all datasets, the probability that the gene is differentially expressed is higher and the related p value is lower.

Mutation Landscape Analysis

TCGA BLCA mutation data containing 411 tumor samples were acquired from the R package "TCGAmutations." The mutation landscape for the six signature genes in BLCA was visualized using the R package "Maftools" (Mayakonda et al., 2018).

Construction and Evaluation of the Prognosis Model

We randomly divided the TCGA BLCA cohort (n = 406) in a 7:3 ratio into a training dataset (n = 285) and a testing dataset (n = 121). Logistic regression analysis was used to integrate the prognostic value of the six homeobox family genes into a six-gene signature model for BLCA. The formula for calculating the risk score for each sample is as follows:

$$\begin{aligned} Risk\ score = {} & 0.062 \times TSHZ3 + 0.122 \times ZFHX4 - 0.031 \times ZEB2 \\ & - 0.206 \times MEIS1 + 0.012 \times ISL1 - 0.061 \times HOXC4. \end{aligned}$$

We calculated the risk score using the expression profiles of each sample based on the formula of the signature model. Then, we divided the BLCA cohort into high- and low-risk groups based on the median risk score. The R package "survival ROC" was used to establish the receiver-operating characteristic (ROC) curves for predicting one-, three-, and five-year overall survival (OS) for the two risk groups. Furthermore, we used the R package "rms" to construct calibration curves and evaluate the precision of the one-, three-, and five-year OS predictions for the BLCA cohort.

Estimation of Immune Cell Infiltration

We identified a group of 782 genes that represent 28 immune cell types involved in innate and adaptive immunity to estimate the infiltration level of different immune cell types in the tumor microenvironment (Charoentong et al., 2017). Subsequently, the single-sample gene set enrichment analysis (ssGSEA) algorithm with the R package "GSVA" was used to evaluate the infiltration level of each immune cell type based on the expression profiles of each sample in BLCA and the immune cell gene marker (Hänzelmann et al., 2013).

Functional Enrichment Analysis

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases include collections of gene sets associated with the function of cells and organisms. Functional enrichment analysis of a set of genes that are dysregulated under certain conditions revealed which GO terms or KEGG pathways are overrepresented for that gene set. The TCGA BLCA cohort was divided into high-risk and low-risk groups according to the median risk score. Then, the R package "limma" was used to identify DEGs between the two risk groups. GO and KEGG analyses of the DEGs between the two risk groups were performed using the R package "clusterProfiler" (Yu et al., 2012). A cutoff value of adjusted p value < 0.05 was used to determine the significant pathways.

Prediction of the Immunotherapy Response

The response of each sample to PD-1/PD-L1 and CTLA4 inhibitors was evaluated according to the gene expression profiles of the BLCA cohort with the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (http://tide.dfci.harvard.edu) (Jiang et al., 2018).

Survival Analysis

The samples were divided into high- and low-risk groups based on the median risk score, and the differences in OS and progression-free survival between the high-risk and low-risk groups were estimated using the Kaplan–Meier method. Survival curves were compared using the log-rank test. The significance threshold was defined as $p < 0.05$.

Statistical Analysis

Statistical analyses were performed using the log-rank test for univariate analysis. Pearson's correlation test was used to assess the relationship between the risk score and immune markers, characteristic gene expression, and the immune cell infiltration score. The relationship between the characteristic gene expression and the immune cell infiltration score was also evaluated. Student's t-tests were used to determine statistical significance of differences between variables. Statistical significance was defined as $p < 0.05$. All statistical analyses were performed in R version 4.0.2.

RESULTS

Identification of the Differentially Expressed Homeobox Gene Family Members in Bladder Cancer

To screen differentially expressed homeobox genes (DEHGs) in BLCA, four GEO datasets, GSE7476, GSE13507, GSE37815, and GSE65635, as well as TCGA gene expression dataset containing 406 BLCA samples and 18 normal samples from the UCSC Xena Browser were obtained. The R package "limma" was used to

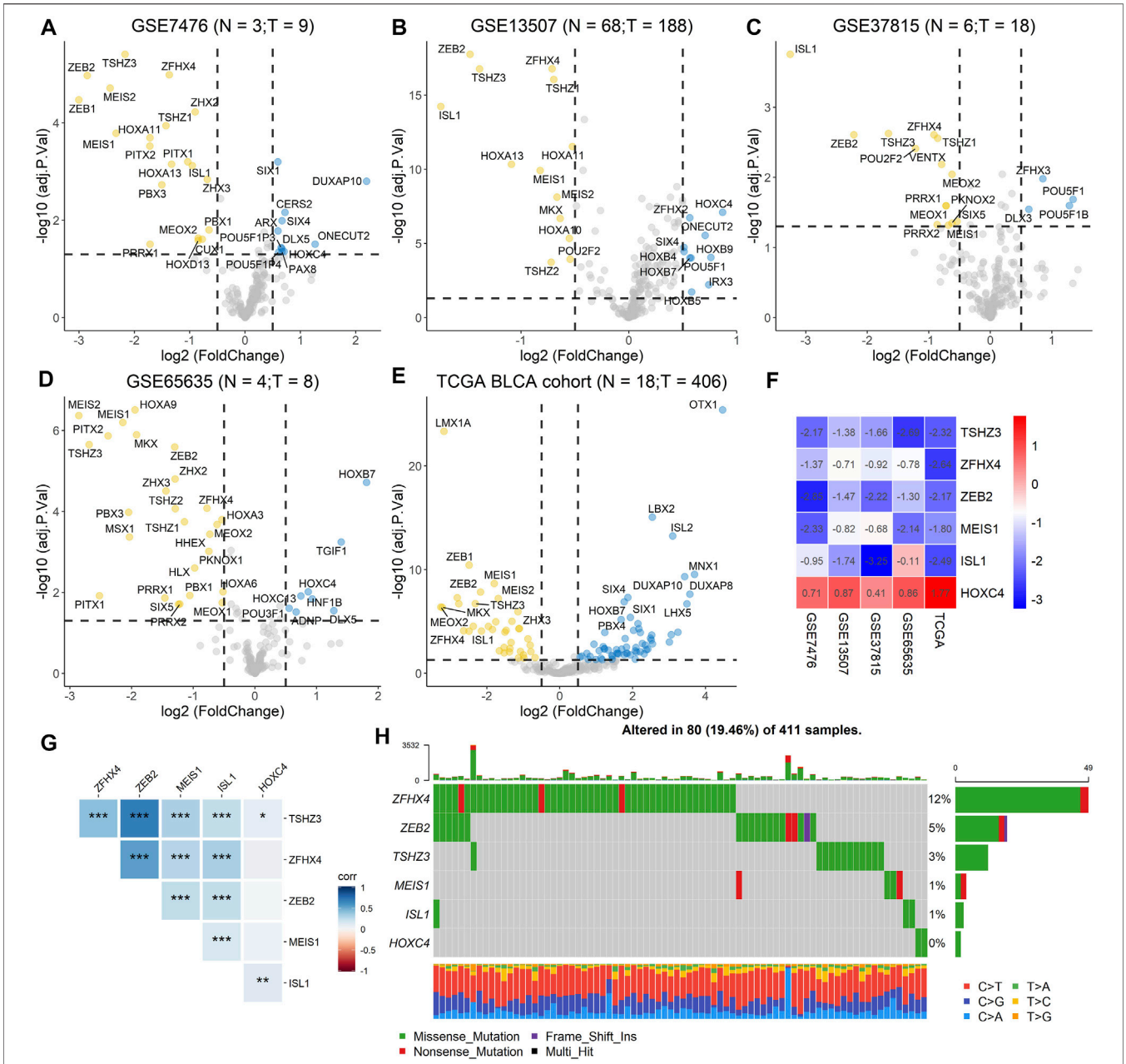


FIGURE 2 | Identification of DEHGs for BLCA, analysis of the mutation landscape, and correlation analysis of the six DEHGs. **(A)** DEHGs for BLCA in the GSE7476 dataset. **(B)** DEHGs for BLCA in the GSE13507 dataset. **(C)** DEHGs for BLCA in the GSE37815 dataset. **(D)** DEHGs for BLCA in the GSE65635 dataset. **(E)** DEHGs for BLCA in the TCGA dataset. **(F)** LogFC values of each gene in different datasets (GSE7476, GSE13507, GSE37815, GSE65635, and TCGA). **(G)** Analysis of the correlations among the six DEHGs. **(H)** Mutation landscape of the six DEHGs in TCGA BLCA patients.

determine the DEHGs of each dataset using |logFC| > 0.5 and adjusted $p < 0.05$ criteria, and the volcanoes were plotted (**Figures 2A–E**). Furthermore, the RRA method based on the expression of each gene in all datasets was used to screen out the candidate genes (score < 0.05) (**Supplementary Table 1**). As a result, six homeobox genes, TSHZ3, ZFHX4, ZEB2, MEIS1, ISL1, and HOXC4, were screened out, and then the logFC values of each gene in different datasets were calculated and are shown in **Figure 2F**. Moreover, correlation analysis of the six homeobox genes was performed, and the results showed that there were significant positive correlations between most genes (**Figure 2G**).

Correlation of the Six Homeobox Genes With Clinical Status and Mutation Landscape

To explore the clinical significance of these six genes, pancancer analysis in BLCA (**Figure 3A**) and 23 other tumors

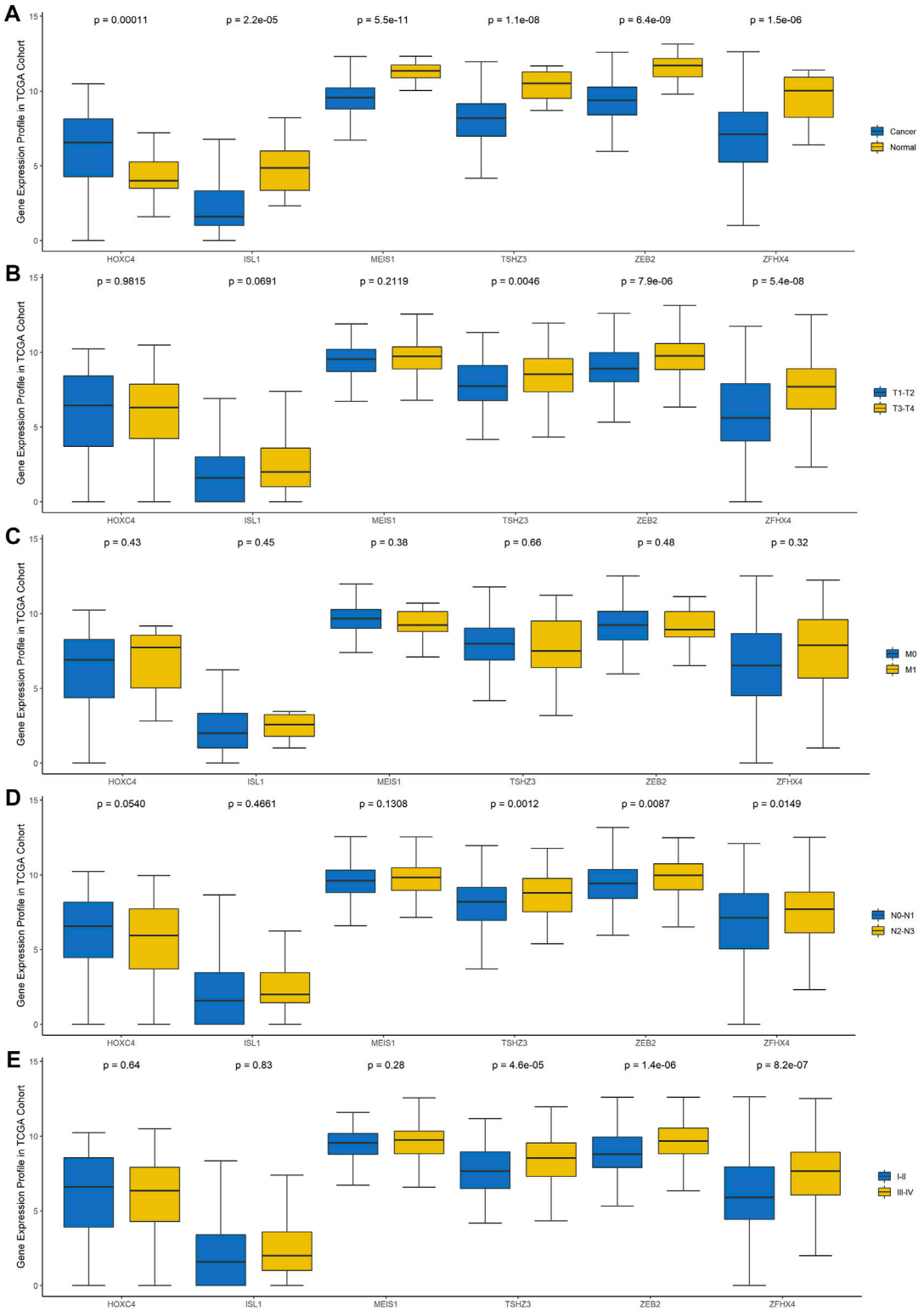


FIGURE 3 | Gene expression profile of these six genes in TCGA cohort. **(A)** Differences in the expression of the six genes between BLCA tissues and normal tissues. **(B)** Differences in the expression of the six genes in different T stages. **(C)** Differences in the expression of the six genes in different N stages. **(D)** Differences in the expression of the six genes in different M stages. **(E)** Differences in the expression of the six genes in different clinical stages.

(**Supplementary Figure 1**) was performed, and the results revealed that the expression of TSHZ3, ZFHX4, ZEB2, MEIS1, and ISL1 was significantly lower than that in normal tissues, while the expression of HOXC4 was higher than that in normal tissues, especially in BLCA, breast invasive carcinoma (BRCA), prostate adenocarcinoma (PRAD), and head and neck squamous cell carcinoma (HNSC). Furthermore, we analyzed the correlation between these six homeobox genes and tumor size, regional lymph node involvement, and distant metastases (TNM) as well as the BLCA stage and found that TSHZ3, ZFHX4, and ZEB2 were positively correlated with T stage, N stage, and BLCA stage, but there was no significant correlation with metastasis (**Figures 3B–E**). In addition, we analyzed the mutation landscape of these six DEHGs in BLCA. Among the 411 samples, 19.46% had at least one gene mutation; ZFHX4 mutation was the most common change, accounting for 12% of mutations; ZEB2, TSHZ3, MEIS1, and ISL1 mutations accounted for 5, 3, 1, and 1% of all mutations, respectively. The waterfall diagram formed according to the mutation landscape of these six DEHGs showed that most mutations were missense mutations (**Figure 2H**). The driver genes ERBB2, HDAC1, PARP1, ERBB3, FGFR3, mTOR, AXL, EZH2, FGFR1, FGFR2, CSF1R, KIT, FGFR4, RET, and ERBB4 are key targets in the treatment of BLCA. Furthermore, we assessed the correlations between these six genes and BLCA driver genes in the BLCA dataset, and it was found that these six genes have a strong correlation with these driver genes (**Supplementary Figure 2**).

A High Risk Score Was Associated With a Poor Clinical Outcome

The prognostic value of the six-homeobox-gene signature was evaluated in the training dataset and testing dataset. We calculated the risk score for each BLCA sample in the training set, ranked them according to this score, and divided them into high-risk and low-risk groups based on the median risk score. We used scatter plots to show the survival status of BLCA patients based on risk scores, and we then performed a chi-square test on the data (**Figures 4A,C**). The results demonstrated that patients in the high-risk group had a higher mortality rate than those in the low-risk group (p = 0.033). The heat map with the gene expression profile of these six homeobox genes showed that ISL1, ZFHX4, TSHZ3, and ZEB2 were more highly expressed in high-risk BLCA samples, while HOXC4 and MEIS1 were highly expressed in the low-risk group (**Figure 4E**). The results for the testing dataset were consistent with those for the training dataset (**Figures 4B,D,F**). Kaplan–Meier analysis was performed on the training dataset, the testing dataset, and all datasets (**Figures 4G–I**), and the results revealed that the survival time of the low-risk group was significantly longer than that of the high-risk group.

GO Function Annotation and KEGG Pathway Analysis Between the High-Risk and Low-Risk Groups

The DEHGs between the two risk groups were analyzed using GO functional annotation and KEGG pathway analysis with the R software package "clusterProfiler." The GO analysis of biological process (BP), molecular function (MF), and cell component (CC) terms showed that most of the enriched terms were related to immunity, including B cell-mediated immunity, immunoglobulin-mediated immune response, immunoglobulin complex, and antigen binding (**Figure 5A**). The KEGG pathway analysis showed that the DEHGs were mainly enriched in cytokine-cytokine receptor interactions, *Staphylococcus aureus* infection, cell adhesion molecules, etc., most of which are related to immunity (**Figure 5B**).

The Signature Composed of Six Homeobox Genes Was Closely Related to Immunity

Since the results of the GO functional annotation and KEGG pathway analysis showed that the signature was related to immunity, analysis of the risk score and immune cell infiltration was then performed to further confirm the conclusion. The results showed that there were differences in the infiltration of most immune cells, except for CD56dim natural killer cells, eosinophils, and monocytes, between the high- and low-risk groups, which demonstrated that the signature was significantly correlated with immune infiltration (**Figure 6A**). In addition, we analyzed the correlation of each gene with the infiltration of immune cells, and the results indicated that TSHZ3, ZFHX4, and ZEB2 were related to almost all immune cell types and that MEIS1, ISL1, and HOXC4 were related to some immune cell types (**Figure 6B**). Furthermore, we also analyzed the correlation analysis between these six genes and cytokines related to T cell function. The results showed that five out of the six genes, TSHZ3, ZFHX4, ZEB2, MEIS1, and ISL1, had a strong correlation with most cytokines, while HOXC4 had a strong correlation with IL-17A (**Supplementary Figure 3**). Similarly, the analysis of the correlations between the expression of these six homeobox genes and immune checkpoints showed that TSHZ3, ZFHX4, and ZEB2 were significantly correlated with the expression of CTLA-4, PD-L1, PD-L2, and PD-1. MEIS1 was strongly correlated with the expression of PD-1. In addition, ISL1 was significantly correlated with CTLA-4, PD-L2, and PD-1 expressions (**Figure 6C**). Then, we analyzed the relationship between the risk score and the response to immunotherapy. The samples were divided into response and no-response groups, and the difference in risk scores between the two groups was assessed. The results showed that the risk scores were higher in the no-response group than in the response group (**Figure 6D**).

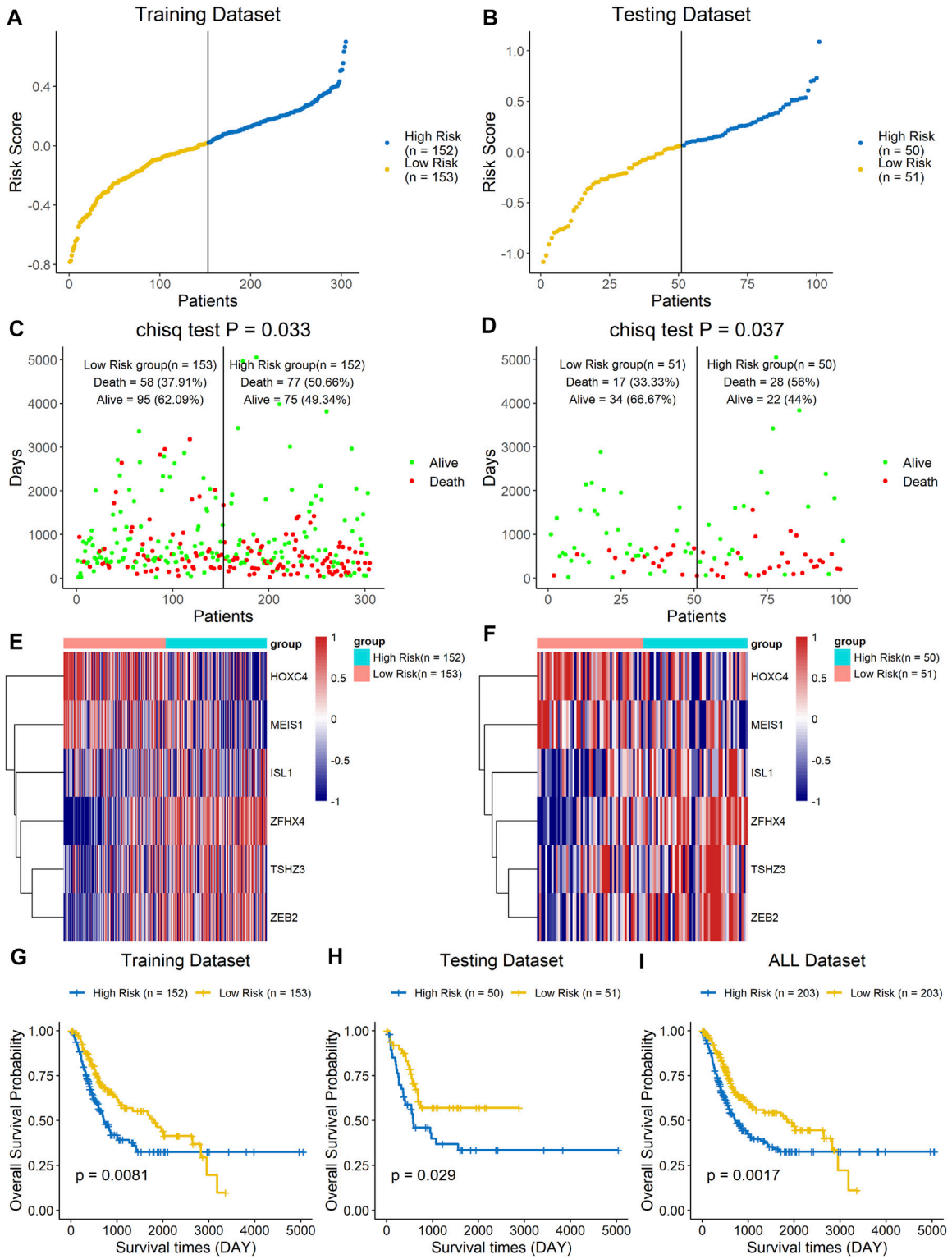


FIGURE 4 | A high risk score was associated with a poor clinical outcome. The BLCA cohort was divided into two groups based on the median estimated score, and the two groups were then compared. The ranked dot plot indicates the risk score distribution in the training dataset **(A)** and testing dataset **(B)**. Scatter plot presenting the patients' overall survival status in the training dataset **(C)** and testing dataset **(D)**. Heat map with the gene expression profiles of these six genes in the training dataset **(E)** and testing dataset **(F)**. Kaplan–Meier curve analysis of the signature in the training set **(G)**, testing set **(H)**, and entire dataset **(I)**.

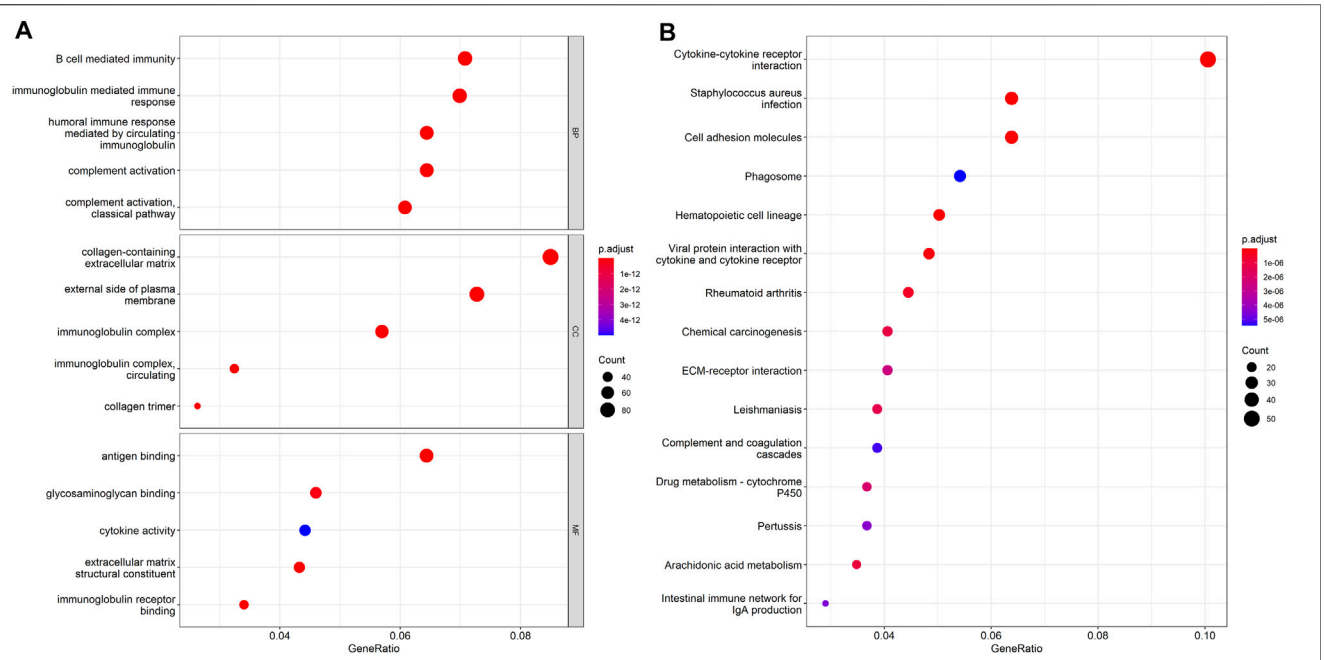


FIGURE 5 | Functional enrichment between the high-risk and low-risk groups. **(A)** GO function annotation. **(B)** KEGG pathway analysis.

Evaluation and External Validation of the Signature Model Performance

The ROC curves of the training set, testing set, and entire dataset (combination of training and testing sets) were plotted, and the area under the ROC curve (AUC) was calculated to verify the accuracy of this signature. The AUCs for one-, three-, and five-year OS were 0.631, 0.606, and 0.609 in the training set; 0.679, 0.652, and 0.671 in the testing set; and 0.647, 0.629, and 0.633, respectively, in the entire dataset (**Figures 7A–C**). To compare the consistency of the model predictions with actual clinical outcomes, calibration curves for one-, three-, and five-year OS were constructed for the training set (**Supplementary Figures 4A–C**), testing set (**Supplementary Figures 4D–F**), and entire dataset (**Supplementary Figures 4G–I**). The calibration curves showed satisfactory agreement between the predicted and observed values for one-, three-, and five-year OS. We further validated the prediction ability of this prognostic signature using the GEO datasets GSE13507, GSE19423, and GSE37815 for external validation. The risk score of each sample was calculated, and the samples were divided into high-risk and low-risk groups based on the optimal splitting point. Kaplan–Meier analysis of GSE13507 (p = 0.17), GSE19423 (p = 0.027), and GSE37815 (p = 0.012) showed that the high-risk group tended to have a shorter survival time than the low-risk group (**Figures 7D–F**).

DISCUSSION

There are many studies on biomarkers of BLCA, such as urine cytology and urine biomarkers; the detection of exfoliated tumor cells in urine or bladder lavage has a high sensitivity for the diagnosis of high-grade BLCA but is less sensitive for low-grade BLCA. There are many biomarkers with unique functions, such as radiotherapy markers, chemotherapy markers, and immunotherapy markers, but these markers have a single function (Giordano and Soria, 2020), and most of them involved single targets, which easily cause false-positive or false-negative results. The application of RNA-Seq and bioinformatic analysis of databases has provided a theoretical basis for mechanistic studies of tumorigenesis and development. Zhu et al. identified some immune-related genes as prognostic factors in BLCA (Zhu et al., 2020). Lian et al. established a signature including eight long non-coding RNAs as a candidate prognostic biomarker for BLCA (Lian et al., 2019). At present, there are few biomarkers that can predict both clinical outcomes and immunotherapy response. In this study, a clinical prediction model containing six homeobox genes was constructed through next-generation sequencing (NGS), which can not only predict the prognosis of patients but also predict the patient's immune response. With the popularity of sequencing technology, its price and convenience continue to improve, and this study has good clinical applicability. Although the homeobox gene family is closely related to BLCA (Cantile et al., 2011), few studies have focused on its prognostic value in BLCA. Therefore, we analyzed the RNA-Seq data of a large number of samples from TCGA and GEO public databases and screened out six significant DEHGs, namely, TSHZ3, ZFHX4, ZEB2, MEIS1, ISL1, and HOXC4, by the RRA method.

Some of these six homeobox genes have been reported to regulate tumor progression and were identified as potential prognostic markers in previous studies. For example, aberrant HOXC4 expression is prevalent and plays an important role in

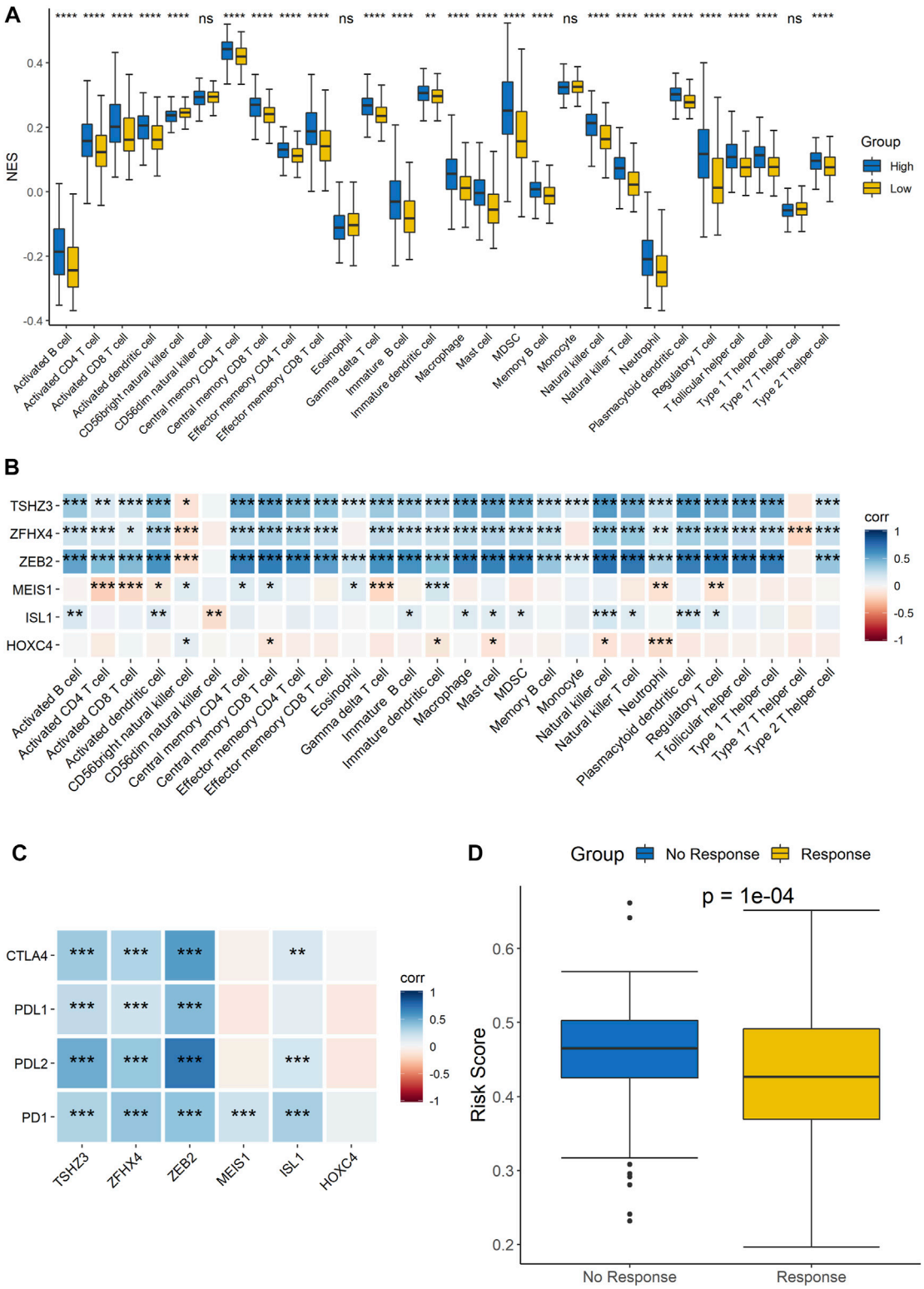


FIGURE 6 | Correlation analysis of the signature and immune characteristics. **(A)** Correlations between the signature and each immune cell infiltration score. **(B)** Correlations between each signature gene and each immune cell infiltration score. **(C)** Correlations between the expression level of immune checkpoints and the six signature genes. **(D)** Prediction of the difference in risk scores between immunotherapy responders and non-responders.

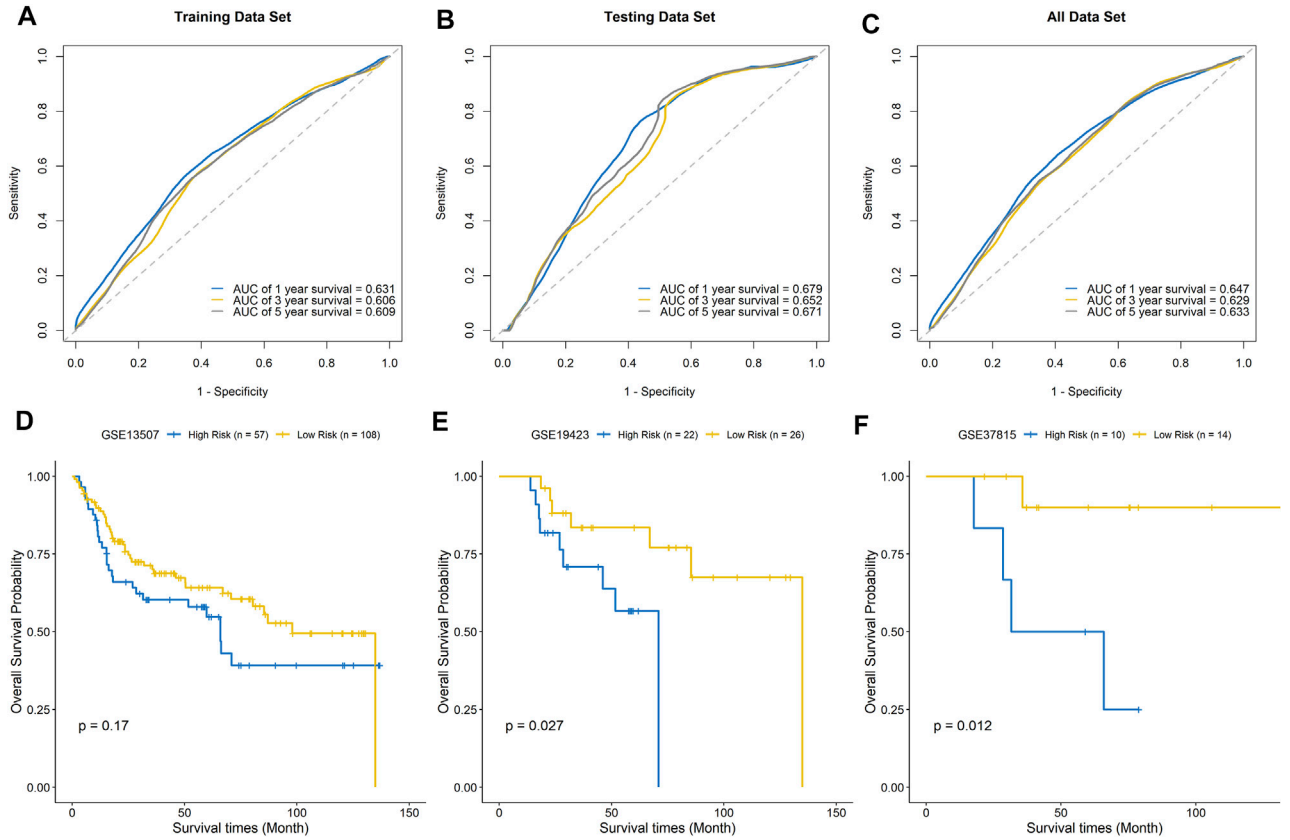


FIGURE 7 | Evaluation of the signature model. ROC curves for predicting one-, three-, and five-year survival in the training set **(A)**, testing set **(B)**, and entire dataset **(C)**. External validation of the signature model using the GEO BLCA cohorts GSE13507 **(D)**, GSE19423 **(E)**, and GSE37815 **(F)**.

the development of prostate cancer (Luo and Farnham, 2020). Moreover, HOXC4 can promote hepatocellular carcinoma progression by transactivating Snail (Yang et al., 2021). The expression of TSHZ3 is significantly downregulated in human glioma tissues and cell lines, and overexpression of TSHZ3 decreases the invasiveness of U87 and U251 glioblastoma cells (Li et al., 2018). In addition, the downregulation or deletion of TSHZ3 function is involved in the pathogenesis of ovarian cancer (McBride et al., 2012), which suggests that TSHZ3 plays an oncogenic role. ZFHX4 is required for the regulation of glioblastoma tumor–initiating cells, and its inhibition leads to reduced tumorigenesis and increased glioma-free survival time. Mutations in ZFHX4 are strongly associated with a poor prognosis, and downregulation of ZFHX4 inhibits the progression of esophageal squamous carcinoma (Qing et al., 2017). ZEB2 can promote the migration and invasion of gastric cancer cells by regulating epithelial–mesenchymal transition (EMT) and is a potential target for gene therapy of invasive gastric cancer (Dai et al., 2012). Deregulation of negative feedback between GATA3 and ZEB2 can promote breast cancer metastasis (Si et al., 2015). The expression level of MEIS1 in acute myeloid leukemia (AML) is negatively correlated with prognosis (Rozovskaia et al., 2001). ISL1 plays an important role in a variety of cellular processes, including cytoskeleton genesis, organogenesis, and tumorigenesis (Zheng and Zhao, 2007), and has been found to be a highly specific marker for pancreatic endocrine tumors and metastases (Schmitt et al., 2008). In addition, it was also significantly associated with aggressive tumor characteristics, tumor recurrence, tumor progression, and disease-specific mortality (DSM) in BLCA and plays an important role in multiple stages of bladder tumorigenesis (Akhir et al., 2020).

We constructed a predictive signature based on these six prognostic homeobox genes. The expression profiles of the signature genes showed that tumors with higher risk scores tended to exhibit elevated ISL1, ZFHX4, TSHZ3, and ZEB2 levels, while those with lower risk scores tended to exhibit elevated HOXC4 and MEIS1 levels. Patients with high risk scores according to the signature had a poor prognosis. Then, we performed survival analysis on the training dataset, the testing dataset, and all datasets. The results showed that the high-risk group had a shorter survival time than the low-risk group. Finally, we validated the performance of the signature using GEO datasets. Overall, the signature can predict the prognosis of patients accurately and has good prognostic value.

Errors in the process of DNA replication are random and universal and subject to correction and repair by the DNA

mismatch repair system. Once the dynamic balance between the two is disrupted, it will easily lead to the occurrence of gene mutations, which will affect the expression of the corresponding genes and facilitate tumorigenesis and development (Turajlic et al., 2019). We analyzed the mutation landscape of these six genes in BLCA. Among the 411 samples, 19.46% had at least one gene mutation. Driver genes are important genes associated with tumor development and play a driving role in the process of cancer development and progression (Martínez-Jiménez et al., 2020). Currently, the driver genes of BLCA include ERBB2, HDAC1, PARP1, and mTOR. These genes are important targets in BLCA treatment (Scholtes et al., 2020). We performed correlation analysis between these six genes and BLCA driver genes in the BLCA dataset and found that these six genes have a strong correlation with the driver genes. The results indicated that the six homeobox genes play an important role in the development of BLCA and that the signature could be used in the prediction of BLCA prognosis.

As a major component of the tumor microenvironment (TME), immune infiltration has been shown to contribute to tumor progression and the immunotherapeutic response (Balkwill et al., 2012), and tumor-infiltrating immune cells, particularly T cells, are the cellular basis of immunotherapy. A better understanding of immune cells in the TME is critical to deciphering the mechanisms of immunotherapy, defining predictive biomarkers, and identifying new therapeutic targets (Zhang and Zhang, 2020; Ma et al., 2021). In our GO analysis, most of the enriched functional terms were immune-related, and the same results were obtained by KEGG analysis. Then, the analysis of the risk score and immune cell infiltration showed that there were differences in the infiltration of most immune cells between the high- and low-risk groups, which demonstrated that this signature was significantly correlated with immune infiltration. Immune cells in tumors work together to control tumor growth, and the effectiveness of immunotherapy depends on the synergistic response of innate and adaptive immune cells, particularly T cells (Moynihan et al., 2016). The function of T cells is usually classified based on whether they secrete specific effector molecules or cytokines, and effector $CD4^+$ T cells include different functional subtypes (Th1 cells secrete IL-2 and IFN-γ; Th2 secretes IL-4, -13; Th17 secretes IL-17A, etc.), while effector $CD8^+$ T cells secrete cytotoxic mediators (perforin and granzymes) or proinflammatory cytokines (TNF-α, IFN-γ) (Szabo et al., 2019). Therefore, in order to analyze the correlation between these six genes and T cell function, we further analyzed the correlation of these six genes with those cytokines, and the results showed that five out of the six genes, TSHZ3, ZFHX4, ZEB2, MEIS1, and ISL1, had a strong correlation with most cytokines, while HOXC4 had a strong correlation with IL-17A. The expression of the six homeobox genes in the signature was correlated with most immune checkpoints (CTLA-4, PD-L1, PD-L2, and PD-1). At present, the most commonly used immunotherapy drugs in clinical practice are immune checkpoint inhibitors. Therefore, we analyzed the relationship between the risk score and the response to immunotherapy. The results showed that the risk scores were higher in the no-response group than in the response group. Above all, this signature was highly correlated with immunity and will be a good predictor of the patient's response to immunotherapy.

However, this study has some limitations. This study is based on TCGA and GEO databases, the reliability of its data is unknown, and this study lacks experimental evidence and is mostly based on bioinformatics prediction, which limits its immediate applicability in clinical practice. In addition, the number of non-tumor tissues assessed in this study is rather small (n = 18), which constitutes a potentially important bias influencing the results. The GEO datasets used for validation were relatively small. Further validation of model prediction accuracy with clinical data is needed.

CONCLUSION

We constructed a risk prediction signature with six homeobox genes, which showed good accuracy and consistency in predicting the patient's prognosis and the response to immunotherapy. Therefore, this signature could be a potential biomarker and treatment target for BLCA 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

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

TY made substantial contributions to the conception, design, interpretation, and preparation of the final manuscript. BD, JL, WS, JS, MZ, SZ, and YM participated in the coordination of data acquisition and data analysis and reviewed the manuscript. DL reviewed and revised the manuscript.

FUNDING

This work was supported by the Henan Provincial Medical Education Research Project (Grant No. WJLX2017057) and

Science and Technology Project of Henan Province (Grant No. 162102310040).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.688298/full#supplementary-material

Supplementary Figure 1 | Pancancer analysis of the six homeobox genes.

Supplementary Figure 2 | Correlation analysis of the six homeobox genes and BLCA driver genes.

Supplementary Figure 3 | Correlation analysis of these six genes and cytokines related to T cell function.

Supplementary Figure 4 | Calibration curves of the signature model in predicting one-, three-, and five-year OS in the training set **(A–C)**, testing set **(D–F)**, and entire dataset **(G–I)**.

Supplementary Table 1 | Results based on the RRA method.

REFERENCES

Akhir, M. K. A. M., Choy, C. S., Abdullah, M. A., Ghani, F. A., Veerakumarasivam, A., and Hussin, H. (2020). The Role of ISL1 and LHX5 LIM Homeobox Genes in Bladder Tumourigenesis. *Malays J. Med. Sci.* 27 (1), 37–45. doi:10.21315/mjms2020.27.1.4

Babjuk, M., Burger, M., Compérat, E. M., Gontero, P., Mostafid, A. H., Palou, J., et al. (2019). European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma *In Situ*) - 2019 Update. *Eur. Urol.* 76 (5), 639–657. doi:10.1016/j.eururo.2019.08.016

Balkwill, F. R., Capasso, M., and Hagemann, T. (2012). The Tumor Microenvironment at a Glance. *J. Cel Sci* 125 (Pt 23), 5591–5596. doi:10.1242/jcs.116392

Cantile, M., Franco, R., Schiavo, G., Procino, A., Cindolo, L., Botti, G., et al. (2011). The HOX Genes Network in Uro-Genital Cancers: Mechanisms and Potential Therapeutic Implications. *Cmc* 18 (32), 4872–4884. doi:10.2174/092986711797535182

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. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Dai, Y.-H., Tang, Y.-P., Zhu, H.-Y., Lv, L., Chu, Y., Zhou, Y.-Q., et al. (2012). ZEB2 Promotes the Metastasis of Gastric Cancer and Modulates Epithelial Mesenchymal Transition of Gastric Cancer Cells. *Dig. Dis. Sci.* 57 (5), 1253–1260. doi:10.1007/s10620-012-2042-6

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

Deng, M., Wei, W., Duan, J., Chen, R., Wang, N., He, L., et al. (2021). ZHX3 Promotes the Progression of Urothelial Carcinoma of the Bladder via Repressing of RGS2 and Is a Novel Substrate of TRIM21. *Cancer Sci.* 112, 1758–1771. doi:10.1111/cas.14810

Giordano, A., and Soria, F. (2020). Role and Efficacy of Current Biomarkers in Bladder Cancer. *AME Med. J.* 5, 6. doi:10.21037/amj.2020.02.06

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

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

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust Rank Aggregation for Gene List Integration and Meta-Analysis. *Bioinformatics* 28 (4), 573–580. doi:10.1093/bioinformatics/btr709

Laughon, A., and Scott, M. P. (1984). Sequence of a Drosophila Segmentation Gene: Protein Structure Homology with DNA-Binding Proteins. *Nature* 310 (5972), 25–31. doi:10.1038/310025a0

Li, Y., Huang, Y., Qi, Z., Sun, T., and Zhou, Y. (2018). MiR-338-5p Promotes Glioma Cell Invasion by Regulating TSHZ3 and MMP2. *Cell Mol Neurobiol* 38 (3), 669–677. doi:10.1007/s10571-017-0525-x

Lian, P., Wang, Q., Zhao, Y., Chen, C., Sun, X., Li, H., et al. (2019). An Eight-Long Non-coding RNA Signature as a Candidate Prognostic Biomarker for Bladder Cancer. *Aging* 11 (17), 6930–6940. doi:10.18632/aging.102225

Luo, Z., and Farnham, P. J. (2020). Genome-wide Analysis of HOXC4 and HOXC6 Regulated Genes and Binding Sites in Prostate Cancer Cells. *PLoS One* 15 (2), e0228590. doi:10.1371/journal.pone.0228590

Ma, W., Liang, J., Liu, J., Tian, D., and Chen, Z. (2021). Establishment and Validation of an Eight-Gene Metabolic-Related Prognostic Signature Model for Lung Adenocarcinoma. *Aging* 13 (6), 8688–8705. doi:10.18632/aging.202681

Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., et al. (2020). A Compendium of Mutational Cancer Driver Genes. *Nat. Rev. Cancer* 20 (10), 555–572. doi:10.1038/s41568-020-0290-x

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 (11), 1747–1756. doi:10.1101/gr.239244.118

McBride, D. J., Etemadmoghadam, D., Cooke, S. L., Alsop, K., George, J., Butler, A., et al. (2012). Tandem Duplication of Chromosomal Segments Is Common in Ovarian and Breast Cancer Genomes. *J. Pathol.* 227 (4), 446–455. doi:10.1002/path.4042

Moynihan, K. D., Opel, C. F., Szeto, G. L., Tzeng, A., Zhu, E. F., Engreitz, J. M., et al. (2016). Eradication of Large Established Tumors in Mice by Combination Immunotherapy that Engages Innate and Adaptive Immune Responses. *Nat. Med.* 22 (12), 1402–1410. doi:10.1038/nm.4200

Qing, T., Zhu, S., Suo, C., Zhang, L., Zheng, Y., and Shi, L. (2017). Somatic Mutations in ZFHX4 Gene Are Associated with Poor Overall Survival of Chinese Esophageal Squamous Cell Carcinoma Patients. *Sci. Rep.* 7 (1), 4951. doi:10.1038/s41598-017-04221-7

Rao, M. K., Maiti, S., Ananthaswamy, H. N., and Wilkinson, M. F. (2002). A Highly Active Homeobox Gene Promoter Regulated by Ets and Sp1 Family Members in normal Granulosa Cells and Diverse Tumor Cell Types. *J. Biol. Chem.* 277 (29), 26036–26045. doi:10.1074/jbc.M203374200

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

Rozovskaia, T., Feinstein, E., Mor, O., Foa, R., Blechman, J., Nakamura, T., et al. (2001). Upregulation of Meis1 and HoxA9 in Acute Lymphocytic Leukemias with the T(4 : 11) Abnormality. *Oncogene* 20 (7), 874–878. doi:10.1038/sj.onc.1204174

Schmitt, A. M., Riniker, F., Anlauf, M., Schmid, S., Soltermann, A., Moch, H., et al. (2008). Islet 1 (Isl1) Expression Is a Reliable Marker for Pancreatic Endocrine Tumors and Their Metastases. *Am. J. Surg. Pathol.* 32 (3), 420–425. doi:10.1097/PAS.0b013e318158a397

Scholtes, M., Akbarzadeh, M., Zwarthoff, E., Boormans, J., Mahmoudi, T., and Zuiverloon, T. (2020). Targeted Therapy in Metastatic Bladder Cancer: Present Status and Future Directions. *Appl. Sci.* 10 (20), 7102. doi:10.3390/app10207102

Si, W., Huang, W., Zheng, Y., Yang, Y., Liu, X., Shan, L., et al. (2015). Dysfunction of the Reciprocal Feedback Loop between GATA3- and ZEB2-Nucleated Repression Programs Contributes to Breast Cancer Metastasis. *Cancer Cell* 27 (6), 822–836. doi:10.1016/j.ccell.2015.04.011

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

Srebrow, A., Friedmann, Y., Ravanpay, A., Daniel, C. W., and Bissell, M. J. (1998). Expression ofHoxa-1 andHoxb-7 Is Regulated by Extracellular Matrix-dependent Signals in Mammary Epithelial Cells. *J. Cel. Biochem.*

69 (4), 377–391. doi:10.1002/(sici)1097-4644(19980615)69:4<377::aid-jcb1>3.0.co;2-k

Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., et al. (2019). Single-cell Transcriptomics of Human T Cells Reveals Tissue and Activation Signatures in Health and Disease. *Nat. Commun.* 10 (1), 4706. doi:10.1038/s41467-019-12464-3

Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving Genetic Heterogeneity in Cancer. *Nat. Rev. Genet.* 20 (7), 404–416. doi:10.1038/s41576-019-0114-6

Yang, H., Zhou, J., Mi, J., Ma, K., Fan, Y., Ning, J., et al. (2015). HOXD10 Acts as a Tumor-Suppressive Factor via Inhibition of the RHOC/AKT/MAPK Pathway in Human Cholangiocellular Carcinoma. *Oncol. Rep.* 34 (4), 1681–1691. doi:10.3892/or.2015.4194

Yang, T., Zhang, X.-B., Li, X.-N., Sun, M.-Z., and Gao, P.-Z. (2021). Homeobox C4 Promotes Hepatocellular Carcinoma Progression by the Transactivation of Snail. *neo* 68 (1), 23–30. doi:10.4149/neo_2020_200403N350

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 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 (8), 807–821. doi:10.1038/s41423-020-0488-6

Zheng, Q., and Zhao, Y. (2007). The Diverse Biofunctions of LIM Domain Proteins: Determined by Subcellular Localization and Protein-Protein Interaction. *Biol. Cel* 99 (9), 489–502. doi:10.1042/bc20060126

Zhu, J., Wang, H., Ma, T., He, Y., Shen, M., Song, W., et al. (2020). Identification of Immune-Related Genes as Prognostic Factors in Bladder Cancer. *Sci. Rep.* 10 (1), 19695. doi:10.1038/s41598-020-76688-w

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.

Copyright © 2021 Dong, Liang, Li, Song, Song, Zhu, Zhao, Ma 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fmolb.2021.687229

Molecular Characterization of Advanced Colorectal Cancer Using Serum Proteomics and Metabolomics

*Jun Rao[1†], Xianghui Wan[1†], Fangfang Tou[1], Qinsi He[1], Aihua Xiong[1], Xinyi Chen[2], Wenhao Cui[3] and Zhi Zheng[1]**

[1]Jiangxi Cancer Hospital, Jiangxi Cancer Hospital of Nanchang University, Nanchang, China, [2]Department of Hematology and Oncology, Beijing University of Chinese Medicine, Beijing, China, [3]Department of Pharmacology, Kyoto Prefectural University of Medicine, Kyoto, Japan

Colorectal cancer (CRC) is a growing public health concern due to its high mortality rate. Currently, there is a lack of valid diagnostic biomarkers and few therapeutic strategies are available for CRC treatment, especially for advanced CRC whose underlying pathogenic mechanisms remain poorly understood. In the present study, we investigated the serum samples from 20 patients with stage III or IV advanced CRC using data-independent acquisition (DIA)-based proteomics and ultra-performance liquid chromatography coupled to time-of-flight tandem mass spectrometry (UPLC-TOF-MS/MS) metabolomics techniques. Overall, 551 proteins and 719 metabolites were identified. Hierarchical clustering analysis revealed that the serum proteomes of advanced CRC are more diversified than the metabolomes. Ten biochemical pathways associated with cancer cell metabolism were enriched in the detected proteins and metabolites, including glycolysis/gluconeogenesis, biosynthesis of amino acids, glutathione metabolism, and arachidonic acid metabolism, etc. A protein-protein interaction network in advanced CRC serum was constructed with 80 proteins and 21 related metabolites. Correlation analysis revealed conserved roles of lipids and lipid-like molecules in a regulatory network of advanced CRC. Three metabolites (hydroquinone, leucenol and sphingomyelin) and two proteins (coagulation factor XIII A chain and plasma kallikrein) were selected to be potential biomarkers for advanced CRC, which are positively and significantly correlated with CEA and/or CA 19–9. Altogether, the results expanded our understanding of the physiopathology of advanced CRC and discovered novel potential biomarkers for further validation and application to improve the diagnosis and monitoring of advanced CRC.

Keywords: colorectal cancer, DIA-MS, UPLC/Q-TOF-MS/MS, correlation analysis, biomarker

OPEN ACCESS

Edited by:
Tianhui Chen, University of Chinese Academy of Sciences, China

Reviewed by:
Rishi Kumar Jaiswal, Loyola University Chicago, United States
Chuan-Ming Xie, Army Medical University, China

****Correspondence:***
Zhi Zheng
zhengxia_2007@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *29 March 2021*
Accepted: *16 July 2021*
Published: *27 July 2021*

Citation:
Rao J, Wan X, Tou F, He Q, Xiong A, Chen X, Cui W and Zheng Z (2021) Molecular Characterization of Advanced Colorectal Cancer Using Serum Proteomics and Metabolomics. Front. Mol. Biosci. 8:687229. doi: 10.3389/fmolb.2021.687229

INTRODUCTION

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer-related death worldwide (Bray et al., 2018; Cantor et al., 2020). It is estimated that by the year of 2030, CRC may account for one in every ten cancer cases and deaths, and increase the global health burden by 60% (Bray et al., 2018). The high mortality rate of CRC is mainly due to its late diagnosis when CRC is already in advanced stages and metastasis has already occurred. Only 9% of CRC

patients are practically diagnosed at stage I, and most (91%) are diagnosed at stage II, III or IV (Hammond et al., 2016). Well documented risk factors of CRC include cigarette smoking, physical inactivity, obesity, and high consumption of alcohol or red meat (Hissong and Pittman, 2020). Family history and certain medical conditions including inflammatory bowel disease are also associated with CRC (Xue et al., 2017).

The pathogenic mechanisms of CRC are complex and heterogeneous, and molecular changes in the tumor determine both the histologic type of premalignant lesion and the time to malignant transformation (Hissong and Pittman, 2020). Secondary to inactivation of the adenomatous polyposis coli (APC) tumor suppressor gene, chromosomal instability is a commonly characterized molecular event in CRC, which subsequently results in hyper-activation of the WNT signaling pathway (Fearon and Vogelstein, 1990). Another molecular event, the microsatellite instability, occurs in 15% of CRC (Hissong and Pittman, 2020). CRC also involves abnormalities in MLH-1, PMS-2, MSH-2, MSH-6, or POL-E gene which are all necessary for repairing DNA mismatches. In the past few years, genomic and transcriptomic landscapes of CRC have been investigated and many genomic alterations and extensive molecular heterogeneity of the disease have been identified (Cancer Genome Atlas Netwo, 2012; Vasaikar et al., 2019). For example, a genome-scale analysis of 276 CRC patients was conducted to characterize somatic alterations in CRC (Vasaikar et al., 2019). The results showed that 24 genes including APC, TP53, ARID1A, and SOX9 are significantly mutated, suggesting a number of new potential therapeutic strategies to CRC.

Recent advances in proteomics and metabolomics techniques have extended our understanding of pathways that control cell proliferation, differentiation, and death (Chen et al., 2019). Identification of changed proteins or metabolites in the development of CRC is important to the discovery of new potential biomarkers for early diagnosis (Ritchie et al., 2010). Identified proteins, metabolites and their corresponding pathways are attractive therapeutic targets for cancer treatment. Proteomics is a high-throughput large-scale approach that allows for simultaneous detection of thousands of proteins in many sample types such as cell, tissue, or body fluids. In 2016, Ward et al. employed surface-enhanced laser desorbtion/ionisation technique to characterize the serum proteomes of 62 CRC patients and 31 healthy individuals, and the study identified complement C3a des-arg, α1-antitrypsin and transferrin to have diagnostic potentials (Ward et al., 2006). Vasaikar et al. conducted a proteogenomic study on prospectively collected CRC tumor tissues and adjacent normal tissues (Vasaikar et al., 2019). An association between increased glycolysis in microsatellite instability-high (MSI-H) tumors and decreased CD8 T cell infiltration was identified, suggesting the glycolysis pathway could be a potential target to reverse the resistance of MSI-H tumors to immune check-point blockade treatment. Similar to proteomics, metabolomics is a large-scale high-throughput omics technology that enables comprehensive and semi-quantitative detection of a large number of metabolites in biological samples. Metabolomic studies in various cancers such as CRC, gastric cancer, liver cancer, and pancreatic cancer have demonstrated its great potentials in improving tumor diagnosis and therapy (Zheng et al., 2017; Fan et al., 2018; Zheng et al., 2018). For instance, Kim et al. performed urine-NMR metabolomics study on 92 patients with colorectal neoplasia and 156 healthy individuals to screen for advanced adenoma and stage 0 CRC (Kim et al., 2019), and the receiver operating characteristics curve analysis results revealed that 3-aminoisobutyrate, taurine, and alanine were good indicators of CRC.

TABLE 1 | Characteristics of 20 CRC patients.

Characteristics	CRC patients
No. of subjects	20
Race	Han
Age (median)	51 years
Gender (%men)	60% (12/20)
UICC Stage	
III	15% (3/20)
IV	85% (17/20)
CA19-9 (ng/ml)	
Average	2000.60
≤37	50% (10/20)
>37	50% (10/20)
CEA (ng/ml)	
Average	503.41
≤5	55% (11/20)
>5	45% (9/20)

For general CRC treatment, different strategies including surgery, radiation therapy, chemotherapy, targeted drug therapy, and immunotherapy have been adopted (Khiavi et al., 2019). For advanced-stage CRC, chemotherapy is commonly recommended, and targeted therapies including anti-epidermal growth factor receptor (anti-EGFR) agents are frequently used in combination with chemotherapy (Rawla et al., 2019; Wang et al., 2019). The treatment effects of invasive CRC will depend on tumor location, stage, and underlying molecular changes including genetic and metabolic alterations. Comprehensive molecular characterization studies of advanced CRC, particularly combined proteomic and metabolomic study, have been rare. In the current study, we conducted nontargeted DIA-MS proteomics and UPLC/Q-TOF-MS/MS metabolomics analyses on 20 serum samples from advanced CRC patients. The aim is to identify key regulatory elements (proteins and/or metabolites) and pathways in advanced CRC, which may serve to be potential biomarkers for early diagnosis and novel therapeutic targets of advanced CRC.

MATERIALS AND METHODS

Study Participants

The study was approved by the Ethics Committee of Jiangxi Cancer Hospital and performed in accordance with the Declaration of Helsinki. Written informed consents were obtained from all participants. In total, 20 patients (N1—N20) diagnosed with advanced CRC at stage III or IV were recruited. The average age of the 20 patients was 51, ranging from 29 to 76.

None of the participants had been diagnosed with other major chronicle diseases or cancers, and none had received any drug treatment in the previous 3 months before sampling. The demographic and clinical characteristics of the 20 patients were listed in **Table 1**.

Proteomic Analysis

The proteomic analysis of the serum samples was conducted using the combination of DIA and a data dependent acquisition (DDA)-based ion library as previously reported (Chen et al., 2019). Each sample of 2 μL serum was first diluted with a lysis buffer containing 100 mM Tris-HCL (pH 8.5, Sigma, MO, United States), 8 M Urea (Sigma, MO, United States), 1 mM EDTA, and 1 mM PMSF, and then centrifuged at 15000 *g* for 15 min at 4°C. The extracted proteins in the supernatant was quantified using a BCA protein assay kit (Bi Yuntian, Shanghai, China), digested in trypsin (Promega, Madison, WI) after reduction and alkylation using the FASP (filter aided sample preparation) method (Wisniewski et al., 2009). The concentration of digested peptides was determined by measuring the absorbance at 280 nm using a NanoDrop 2000 instrument (ThermoFisher Scientific, United States). Each 3 μg of trypsin-digested peptides was mixed with iRT peptides (Biognosys, Schlieren, Switzerland) and analyzed in the DDA mode on an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific, United States) equipped with an EASY-nLC 1000 system (ThermoFisher Scientific, United States) (Chen et al., 2019). The peptides were separated on a 150 μm I.D. × 15 cm C18 trap column (C18, 1.9 μm, 120 Å, Dr Maisch GmbH) with a mobile solution flow rate of 600 nL/min. The gradient elution program was as the following: 7–20% solvent B for 80 min, 20–32% solvent B for 25 min, 32–90% solvent B for 13 min. Data was acquired with full scans (m/z 350–1500) at a mass resolution of 60,000 at m/z 200. The top 20 precursor ions were selected for fragmentation in the HCD (high energy collision dissociation) cell at normalized collision energy of 32%, and fragment ions were scanned at a resolution of 30,000 at m/z 200. The automatic gain control (AGC) was set to 4e5 for full MS with maximum ion injection time of 50 ms, and 5e4 for MS/MS with maximum ion injection time of 54 ms. The dynamic exclusion was 30 s.

The DIA analysis was performed the same as for DDA. The full scan in DIA analysis was at a resolution of 60,000 over m/z 350—1500, DIA scan resolution was 30,000, collision energy was 32%, AGC target was $5e^5$, and maximum injection time was 74 ms. There were 45 variable DIA windows set from 350 to 1500 m/z. Protein identification and quantification were performed using Spectronaut pulsar X 12.0 (Biognosys) with default setting. For protein identification, DDA raw files were searched against the human Uniprot fasta database, and three to six fragments with the highest quality were selected for each peptide to generate a spectral library. Peptide FDR (false discovery rate), PSM FDR, and protein FDR were all set to 1%. The iRT Calibration R square was set to 0.8. For protein quantification using the DIA data, RT regression type was set as Local (Non- Linear) Regression. All results were filtered by a Q-value cutoff of 0.01 (corresponds to a FDR of 1%). *p*-value estimator was performed by Kermel Density Estimator. Area was used for protein quantification. Every peptide was validated with at least three fragment-ions.

Metabolomics Analysis

Metabolites in the serum samples were extracted with 120 μL of 50% methanol buffer (Chen et al., 2019). For global metabolomics analysis, an ultra-performance liquid chromatography (UPLC) system (SCIEX, Cheshire, United Kingdom) coupled to a high-resolution tandem mass spectrometer (Triple TOF 5600 plus; SCIEX) were used. An ACQUITY UPLC T3 column (100 mm × 2.1 mm, 1.8 μm, Waters, United Kingdom) was employed for reversed phase separation. The two mobile phase solutions were solvent A (water, 0.1% formic acid) and solvent B (Acetonitrile, 0.1% formic acid), the mobile phase solution flow rate was 0.4 ml/min. The gradient elution program was as the following: 0–0.5 min, 5% B; 0.5–7 min, 5–100% B; 7–8 min, 100% B; 8–8.1 min, 100–5% B; 8.1–10 min, 5% B. The injection volume for each sample was 4 μL. The Q-TOF was performed in both positive and negative ion modes (Chen et al., 2019). The ionspray voltage floating in positive and negative ion mode were set at 5000 and −4500 V, respectively. The XCMS software was used for MS data pretreatments including peak picking, peak grouping, retention time correction, second peak grouping, and annotation of isotopes and adducts. Online databases including Human Metabolome Database (HMDB) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were employed for metabolite annotations. An in-house fragment spectrum library of metabolites was used for compound identification by MS^2 matching. Five pooled quality control (QC) samples were prepared by combining 10 μL of each extraction and injected with the true samples for quality control purpose.

Data Analysis

All of the raw mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the iProX partner repository and the dataset identifier is PXD025041. Both the proteomic and metabolomic data were normalized by defining the median of each protein/metabolite value equal to 1.00, while missing values (if any) were filled with the observed minimum value (Chen et al., 2019). Hierarchical clustering with average linkage using Pearson correlation as a distance metric was conducted using the Mev (MultiExperiment Viewer, 4.8) software. The categories of identified proteins were determined using the online PANTHER (protein annotation through evolutionary relationship) classification system (www.pantherdb.org). A multi-omics data analysis tool, OmicsBean (http://www.omicsbean.com), was used for bioinformatics analyses including Gene Ontology (GO) analysis, KEGG pathway, and protein-protein interaction network analysis. Before correlation analysis, proteins and metabolites with the same values in more than ten samples were filtered out due to their obviously high correlations. Pearson's product-moment correlation analysis was conducted using R statistical software. The corresponding *p*-values were calculated using the cor.test function. The calculated *p*-values were accordingly adjusted to control the

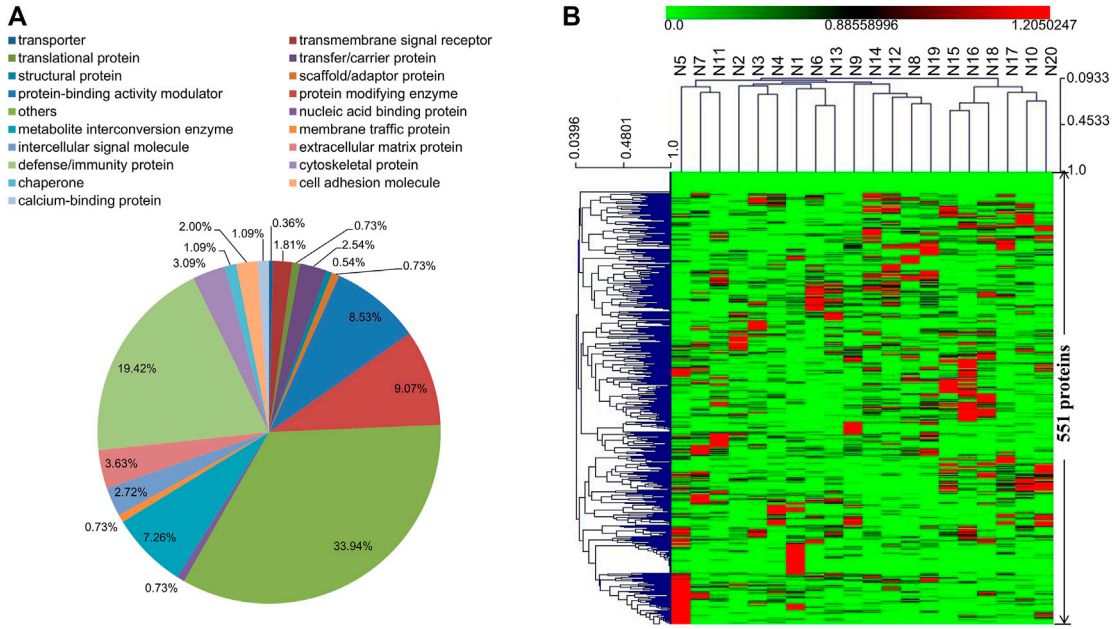


FIGURE 1 | (A) Functional classification of the identified 551 proteins in serum samples (N1–N20) using the PANTHER classification system (www.pantherdb.org). **(B)** Hierarchical clustering analysis of the identified 551 proteins in serum samples (N1–N20).

false discovery rate (FDR) (Rao et al., 2014). The graphical presentations of correlations were composed with Cytoscape version 3.4.0.

RESULTS

The Proteomics Characterization of Advanced Colorectal Cancer Serums

A total of 551 proteins were identified in the DIA proteomics analysis, and the majority of them are defense/immunity proteins, protein modifying enzymes, protein-binding activity modulators, and metabolite interconversion enzymes (**Supplementary Table S1** and **Figure 1A**). Other types of identified proteins include extracellular matrix proteins, signaling molecules, intercellular signal molecules, transmembrane signal receptors, transfer/carrier proteins, and cell adhesion molecules. Except for 25 proteins which had low abundances in all samples, the rest 526 proteins displayed remarkable changes across the 20 tested samples as illustrated in the heat-map of hierarchical clustering analysis (**Figure 1B**). On one hand, the enrichment of certain proteins seemed to be sample-specific. For example, 34 proteins at the bottom of the heat-map including collectin-10, lithostathine-1-alpha, and osteopontin were abundant only in sample N5. On the other hand, several proteins were specifically enriched in one or more samples. For instance, the levels of Immunoglobulin kappa variable 1–39 and C-C motif chemokine 18 were higher in samples N15 and N16 than in other samples.

The identified 551 proteins were further annotated according to GO database (**Supplementary Figures S1, S2, S3**). Most of the annotation terms contain cellular process, response to stimulus, and biological regulation. The main terms for cellular process include cellular anatomical entity, intracellular, and protein-containing complex. The major molecular function terms were binding, catalytic activity, and molecular function regulartor. To better understand the biological functions and/or interactions of the identified 551 proteins, we also carried out pathway annotation analysis in KEGG and mapped 251 proteins to 189 pathways. The top fifteen proteins are in complement and coagulation cascades, PI3K-Akt signaling pathway, and pathways in cancer (**Supplementary Table S2**).

The Metabolomics Characterization of the Advanced Colorectal Cancer Serum Samples

The same 20 serum samples analyzed by proteomics study were also subjected to non-targeted UHPLC-Q-TOF-MS/MS metabolomics analysis. A total of 9,193 positive-mode and 7,571 negative-mode ion features were detected. Based on MS/MS spectrum matching, 567 metabolites were determined in positive-mode and 431 in negative-mode data (**Supplementary Tables S3, S4**). Eventually, 719 non-redundant metabolites were identified in the CRC serums, and they can be classified into 15 categories according to the HMDB database (**Figure 2A**). Among these 719 metabolites, about 51% are lipids and lipid-like molecules. The hierarchical clustering analysis result of the metabolomics data indicated that CRC serum metabolites have less changes in their abundances than proteins in the 20 tested samples. In our previous study using six pooled samples from healthy blood donators for integrative proteomics and metabolomics analyses, similar results that serum metabolites

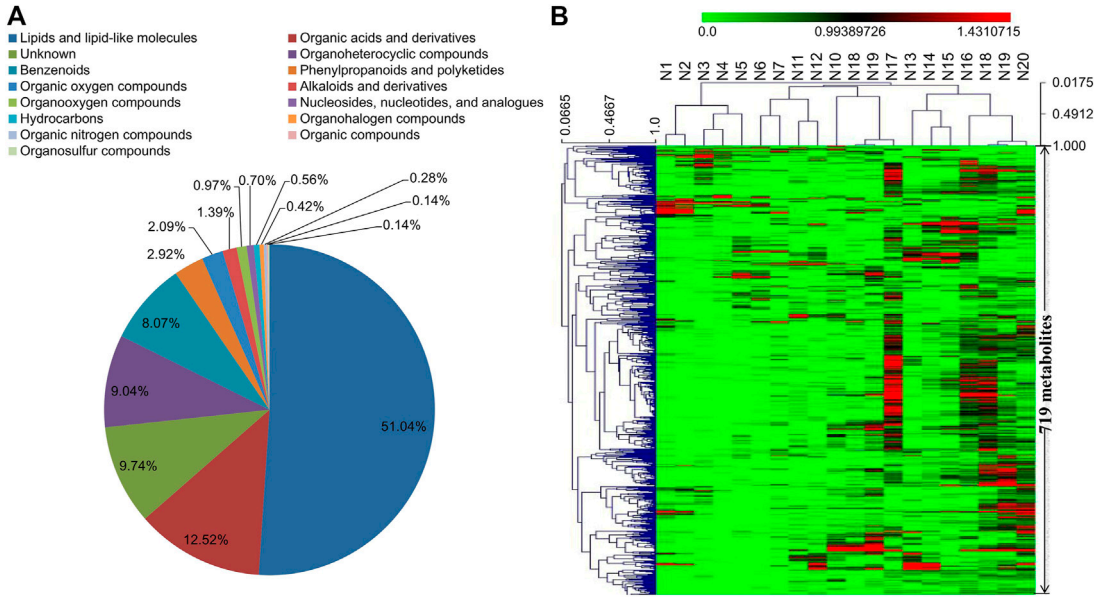


FIGURE 2 | (A) Category information of the 719 metabolites identified in serum samples (N1–N20) according to the database from HMDB. **(B)** Hierarchical clustering analysis of the identified 719 metabolites in serum samples (N1–N20).

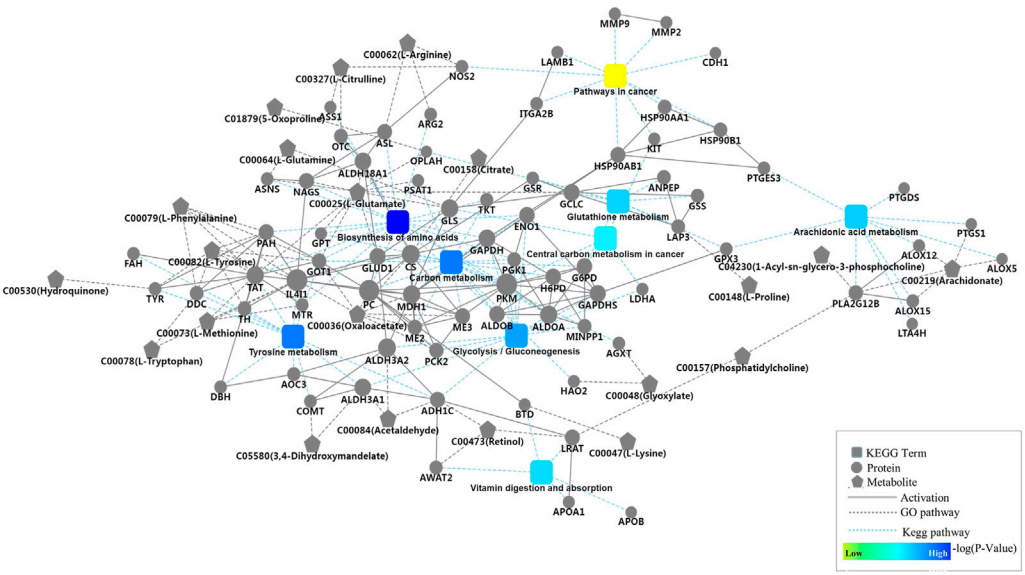


FIGURE 3 | Protein-protein interaction network analysis based on nine key pathways associated with cancer cell metabolism, which involved glycolysis/gluconeogenesis, carbon metabolism, protein digestion and absorption, biosynthesis of amino acids, glutathione metabolism, vitamin digestion and absorption, central carbon metabolism in cancer, arachidonic acid metabolism, and tyrosine metabolism, as well as 80 proteins and 21 metabolites.

exhibit less change than proteins were also observed (**Supplementary Figure S4**) (Chen et al., 2020; Deng et al., 2020). The 719 metabolites were mapped to 135 KEGG pathways, and the top three pathways with most metabolites are metabolic pathways, glycerophospholipid metabolism, and biosynthesis of amino acids (**Supplementary Tables S5, S6**).

Protein-Protein Interaction Network Analysis

The 189 mapped pathways in the proteomics analysis and the 135 mapped pathways in the metabolomics analysis have 69 pathways in common, including 238 proteins and 187 metabolites in the 69 pathways (**Supplementary Table S7**).

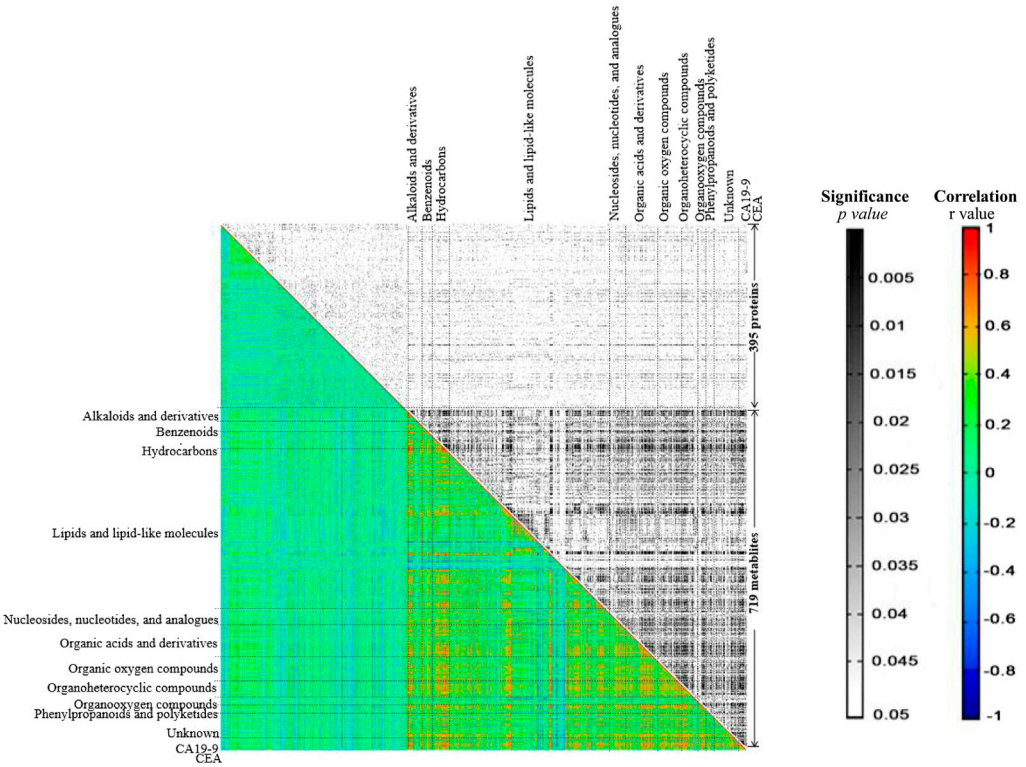


FIGURE 4 | The heat-map generated from correlation analysis. *X* and *Y*-axes were categorized into proteins/metabolites/CA19-9/CEA. In the black and white area rectangles represent *p*-values resulting from Pearson correlation coefficient, while in the colored area rectangles represent *r* values respective to Pearson correlation coefficient computation.

Ten of the 69 pathways are associated with cancer metabolisms including pathways in cancer, glycolysis/gluconeogenesis, carbon metabolism, protein digestion and absorption, biosynthesis of amino acids, glutathione metabolism, vitamin digestion and absorption, central carbon metabolism in cancer, arachidonic acid metabolism, and tyrosine metabolism. Protein-protein interaction network analysis was conducted based on the ten pathways to provide further insight into the developmental and physiological processes underlying advanced CRC. All of the ten pathways except protein digestion and absorption were covered in the network with 80 proteins and 21 metabolites (**Figure 3**). The 80 proteins were mapped to corresponding pathways via KEGG analysis, and the metabolites were connected to proteins *via* GO database annotations. Two pathways, biosynthesis of amino acids and carbon metabolism, dominate the network and include a large number of proteins such as PKM, GAPDH, ALDOA, and ALDOB that plays important roles in cellular proliferation. Other pathways such as glutathione metabolism, tyrosine metabolism, and glycolysis/gluconeogenesis in the network generate key products that promote cell survival and growth. Arachidonic acid metabolism pathway in the network consists of nine detected proteins, and this pathway play important roles in the development of various cancers (Borin et al., 2017).

Correlation Analysis Between the Detected Proteins and Metabolites

To further explore the regulatory network in advanced CRC, network-based analysis (Rao et al., 2014) was conducted to analyze the correlations among the identified proteins and metabolites (**Supplementary Table S8**), as well as their correlations with two tumor markers, CEA and CA 19–9. Pearson pair-wise correlation of 1116 elements including 395 proteins, 719 metabolites and the two tumor markers were calculated, and the results were presented in a heat-map displaying a total of 622,170 correlations with scores ranging from −0.8978 to 0.9991 (**Figure 4**). There were 23,201 significant correlations with $r^2 \geq 0.49$ and FDR ≤0.05. Among them, 22,891 were positive correlations and the rest 310 were negative ones. There were much more significant correlations between metabolites than those between proteins or with tumor markers. Lipids and lipid-like metabolites dominated the significant correlations, accounting for nearly 63% of all significant correlations. These significant correlations were not observed in a previous study on the healthy control serums (**Supplementary Figure S5**) (Chen et al., 2020; Deng et al., 2020). Two important metabolites, citric acid and glutamine, are positively and significantly correlated (**Supplementary Table S9**), and other correlations with citric acid and glutamine were shown in **Figure 5**. There were 120

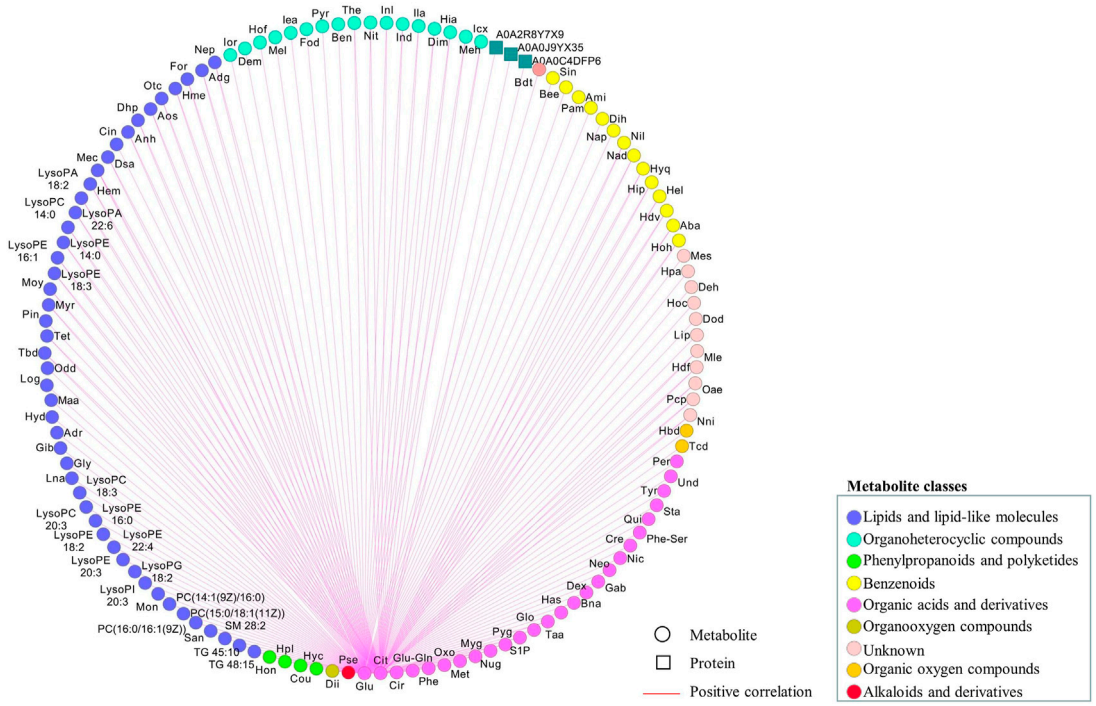


FIGURE 5 | Regulatory network associated with citric acid or glutamine based on significant correlations ($r^2 \geq 0.49$ and FDR ≤0.05). Metabolites and proteins were respectively represented as circular and rectangle, and their relations as edges. Metabolites categorized into different pathways were displayed in different node colors. The positive correlations were shown in red. Computations of the correlations were performed under the R environment. Cytoscapewas employed to generate the graphical output of the networks.

TABLE 2 | The list of significant correlations associated with CEA or CA 19–9.

Protein/metabolite 1	Protein/metabolite 2	Correlation coefficient	*p* value
CEA	Hydroquinone	0.74	1.99E-04
CEA	CA19-9	0.96	1.46E-11
CEA	Immunoglobulin heavy variable 1–69D	0.74	2.18E-04
CA19-9	Immunoglobulin lambda variable 4–60	0.78	4.22E-05
CEA	Immunoglobulin lambda variable 4–60	0.82	1.00E-05
CEA	Immunoglobulin kappa variable 2–40	0.71	4.84E-04
CA19-9	16beta-16-Hydroxy-3-oxo-1,12-oleanadien-28-oic acid	0.71	4.82E-04
CEA	16beta-16-Hydroxy-3-oxo-1,12-oleanadien-28-oic acid	0.73	2.50E-04
CA19-9	SM 28:3; SM(d14:2/14:1)	0.76	9.02E-05
CEA	SM 28:3; SM(d14:2/14:1)	0.79	3.42E-05
CA19-9	Coagulation factor XIII A chain	0.75	1.37E-04
CEA	Coagulation factor XIII A chain	0.82	1.16E-05
CA19-9	Leucenol	0.72	3.40E-04
CEA	Leucenol	0.76	9.88E-05
CEA	Plasma kallikrein (fragment)	0.70	5.39E-04
CA19-9	Plasma kallikrein (fragment)	0.74	1.73E-04
CA19-9	1-(1′,3′-Benzodioxol-5′-yl)-2-butanamine	0.81	1.34E-05
CEA	1-(1′,3′-Benzodioxol-5′-yl)-2-butanamine	0.87	8.56E-07

positive and significant correlations with glutamine and 63 positive and significant correlations with citric acid. In total, there were 182 positive and significant correlations between citric acid/glutamine and 126 molecules, and most of the 126 molecules were lipids and lipid-like metabolites. Other molecules which had positive and significant correlations with citric acid or glutamine include 24 organic acids and their derivatives, 14 benzenoids, 17 organoheterocyclic compounds, and three proteins. There were 18 positive and significant correlations between ten elements (five metabolites

and five proteins, **Table 2**) and the two cancer marks (CEA and CA 19–9). The correlation between CEA and CA 19–9 was also very strong. In the previous study using six pooled healthy blood samples, there was no significant correlation between CEA and CA 19–9, and there was no significant correlation between the above mentioned ten elements and the two cancer markers (**Supplementary Figure S5**) (Chen et al., 2020; Deng et al., 2020).

DISCUSSION

CRC is one of the most common and lethal cancers worldwide, but has poor diagnosis and few effective treatment options, especially for patients with advanced CRC. These limitations highlight the importance of gaining new understanding of the pathogenesis of advanced CRC. We employed an integrated proteomics and metabolomics strategy to investigate the serum samples from 20 CRC patients, including 17 diagnosed with stage IV CRC and three with stage III advanced CRC. The ages of the 20 patients ranged from 29 to 76. The 20 patients had various types of advanced CRC including sigmoid colon cancer, right colon cancer, rectal cancer, ascending colon cancer, and adenocarcinoma of the junction of rectum and sigmoid. The study was a multi-omics one consisting of both proteomics and metabolomics investigations on each serum sample, and this enabled us to explore both proteome and metabolome changes in the serums of advanced CRC patients and the interactions between the proteome and metabolome.

Experimental human body fluid samples include blood, breast milk, tears, urine and malignant pleural efusions, etc., among which blood is the most commonly adopted one for discovering new biomarker to predict treatment effects and prognosis of diseases including cancers (Deng et al., 2020). Blood can be noninvasively collected in large quantity through a simple procedure, and changes in blood proteome and metabolome can reflect physical or pathological disturbances to an otherwise balanced and homeostatic system (Zhou et al., 2019). In the present study, DIA-MS and UPLC/Q-TOF-MS/MS technologies were used to investigate proteomes and metabolomes of advanced CRC serum samples. The proteomics study detected a total of 551 proteins, most of which are defense/immunity proteins, protein modifying enzymes, and protein-binding activity modulators. Other types of detected proteins include metabolite interconversion enzymes, extracellular matrix proteins, and signaling molecules. A total of 719 named metabolites were determined in the metabolomics study, and 649 of them can be classified into 14 categories such as lipids and lipid-like molecules, organic acids and derivatives, and organoheterocyclic compounds, etc. These 649 metabolites cover most of the central metabolism pathways such as carbohydrate super pathway, amino acid super pathway, lipid super pathway, and nucleotide super pathway. Compared to previous reports, the current study revealed changes of more serum proteins and metabolites related to advanced CRC, which is helpful in uncovering more molecular changes and selecting novel biomarkers.

The KEGG pathway analysis revealed that the detected proteins and metabolites share 69 common pathways, and ten of them were associated with cancer cell metabolisms including glycolysis/gluconeogenesis, biosynthesis of amino acids, glutathione metabolism, and arachidonic acid metabolism. We used OmicsBean online software to construct a protein-protein interaction network in advanced CRC serum. This network covers nine cancer-associated pathways, 80 proteins and 21 metabolites. Metabolic reprogramming, as a hallmark of cancer, has become a hot topic in cancer research over the past decade (Koppenol et al., 2011). The well documented Warburg effect is characterized by an increase in glucose uptake and lactate production. There are 13 detected proteins in glycolysis/gluconeogenesis pathway (**Figure 3**), and most play important roles in the development of cancers including CRC (Gimm et al., 2001; Patel et al., 2008; Tsai et al., 2010; Duell et al., 2012; Ahmad et al., 2013; Cui et al., 2014; Leithner et al., 2015; Yun et al., 2015; Dayton et al., 2016; He et al., 2016; Dai et al., 2018). For example, phosphoglycerate kinase 1 (PGK1) is reported to be a promoter of metastasis in CRC, and high expression of ALDOA is associated with poor CRC prognosis.

Glutathione metabolism plays both beneficial and pathogenic roles in a series of malignancies (Bansal and Simon, 2018). In our study, eight proteins including G6PD, GPX3, and LAP3 which are reported to be involved in colon cancer cell growth were connected to glutathione metabolism (Pelosof et al., 2017; Zhang et al., 2017; Yang et al., 2018). Nine proteins in the constructed network are associated with arachidonic acid metabolism, suggesting that arachidonic acid pathway may play important roles in advanced CRC. Many studies have demonstrated the connection between arachidonic acid metabolism and carcinogenesis (Hong et al., 2004). Habermann et al. reported that SNPs inside PTGS1, ALOX5, ALOX12, and ALOX1 affect fatty acid metabolisms in CRC (Habermann et al., 2013). In addition, 21 metabolites including citrate, oxaloacetate, arachidonate and nine standard amino acids were connected to the nine cancer-associated pathways via related proteins.

A regulatory network to reveal key regulatory elements in advanced CRC was also constructed by correlation analysis (Rao et al., 2014). A large number of significant correlations were discovered, most of which were positive correlations. The highly positive associations between every two metabolites suggested the conserved roles of metabolome in the human serum, which was in line with the observation in hierarchical clustering analysis. Lipids and lipid-like molecules dominate the significant correlations, suggesting their essential roles in advanced CRC (Beloribi-Djefaflia et al., 2016). In the metabolomics studies of aqueous humor samples from patients with high myopia and various mature seeds including maize kernels, the identified amino acids have conserved roles, which are totally different from the findings in the current study (Toubiana et al., 2012; Rao et al., 2014; Ji

et al., 2017). Most of the significant correlations with citric acid and glutamine are also correlated to lipids and lipid-like molecules. There were 18 positive and significant correlations to CEA or CA 19–9. CEA and CA19-9 are acknowledged markers for diagnosing early stages of CRC and predicting treatment effect, but with limitations (Hissong and Pittman, 2020). In the present study, CEA and CA 19–9 are strongly correlated, but the levels of CEA and CA 19–9 exceeded the standard reference values in only 55% of the patients. Five metabolites and five proteins with strong and positive correlations with CEA or CA 19–9 have been demonstrated to be potential biomarkers involved in modulating cancer cell growth (Byeon et al., 2018). For example, hydroquinone was determined to be able to increase skin cancer risk, while the biosynthesis of sphingomyelin was reported to modulate cancer cell death and growth (Lewis et al., 2018). Kulp et al. found that mimosine can block cell cycle progression in asynchronous human breast cancer cells by chelating irons (Kulp and Vulliet, 1996). Activation peptide of the coagulation factor XIII (AP-F13A1) and plasma kallikrein (fragment) were identified to be novel biomarkers for the screening of CRC and lung cancer, respectively (Chee et al., 2008; Peltier et al., 2018). Together with CEA and CA 19–9, these three metabolites (hydroquinone, sphingomyelin and mimosine and) and two proteins (coagulation factor XIII A chain and plasma kallikrein) are potential biomarkers, to improve the accuracy of diagnosis and monitoring of CRC. These potential new biomarkers need to be validated in further studies with more patients and controls.

CONCLUSION

In summary, the present study reported an integrative proteomics and metabolomics investigation of advanced CRC serums. The constructed protein-protein interaction network and correlation analysis revealed key regulatory elements and pathways in advanced CRC, and new potential biomarkers for diagnosis and monitoring of CRC were selected. Since the number of patients in the present study was limited, future validation studies need to be conducted to validate the discoveries reported in this study.

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 local ethics committee of Jiangxi Cancer Hospita. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

All authors contributed to the study concept and design. JR and XW did the main statistical analysis. FT, QH, and AX contributed to the literature review and interpretation of the data. XC and WC designed this study and wrote the manuscript. ZZ supervised and provided administrative, technical and material support of the study. All authors have read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No. 81960865), Jiangxi provincial department of science (Nos. 20171BBG70119 and 20181 BAB205047), and National Cancer Center Climbing Fund (No. NCC201814B045); in part by grants from Health and Family Planning Commission of Jiangxi Province (No. 20203533).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.687229/full#supplementary-material

REFERENCES

Ahmad, S. S., Glatzle, J., Bajaeifer, K., Bühler, S., Lehmann, T., Königsrainer, I., et al. (2013). Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. *Int. J. Oncol.* 43, 586–590. doi:10.3892/ijo.2013.1971

Bansal, A., and Simon, M. C. (2018). Glutathione metabolism in cancer progression and treatment resistance. *J. Cel Biol* 217, 2291–2298. doi:10.1083/jcb.201804161

Beloribi-Djefaflia, S., Vasseur, S., and Guillaumond, F. (2016). Lipid metabolic reprogramming in cancer cells. *Oncogenesis* 5, e189. doi:10.1038/oncsis.2015.49

Borin, T., Angara, K., Rashid, M., Achyut, B., and Arbab, A. (2017). Arachidonic acid metabolite as a novel therapeutic target in breast cancer metastasis. *Ijms* 18, 2661. doi:10.3390/ijms18122661

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

Byeon, S. E., Yi, Y., Lee, J., Yang, W. S., Kim, J. H., Kim, J., et al. (2018). Hydroquinone exhibits *in vitro* and *in vivo* anti-cancer activity in cancer cells and mice. *Int. J. Mol. Sci.* 19, 903.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Cantor, D. I., Cheruku, H. R., Westacott, J., Shin, J.-S., Mohamedali, A., and Ahn, S. B. (2020). Proteomic investigations into resistance in colorectal cancer. *Expert Rev. Proteomics* 17, 49–65. doi:10.1080/14789450.2020.1713103

Chee, J., Naran, A., Misso, N. L., Thompson, P. J., and Bhoola, K. D. (2008). Expression of tissue and plasma kallikreins and kinin B1 and B2 receptors in lung cancer. *Biol. Chem.* 389, 1225–1233. doi:10.1515/bc.2008.139

Chen, W., Liu, Z., and You, X. (2020). Serum proteomics study on patients with CSCC by DIA/SWATH. *The Pract. J. Cancer (Chinese Version)* 35, 1928–1931.

Chen, X., Rao, J., Zheng, Z., Yu, Y., Lou, S., Liu, L., et al. (2019). Integrated tear proteome and metabolome reveal panels of inflammatory-related molecules via key regulatory pathways in dry eye syndrome. *J. Proteome Res.* 18, 2321–2330. doi:10.1021/acs.jproteome.9b00149

Cui, J., Shi, M., Xie, D., Wei, D., Jia, Z., Zheng, S., et al. (2014). FOXM1 promotes the warburg effect and pancreatic cancer progression via transactivation of LDHA expression. *Clin. Cancer Res.* 20, 2595–2606. doi:10.1158/1078-0432.ccr-13-2407

Dai, L., Pan, G., Liu, X., Huang, J., Jiang, Z., Zhu, X., et al. (2018). High expression of ALDOA and DDX5 are associated with poor prognosis in human colorectal cancer. *Cmar* Vol. 10, 1799–1806. doi:10.2147/cmar.s157925

Dayton, T. L., Jacks, T., and Vander Heiden, M. G. (2016). PKM 2, cancer metabolism, and the road ahead. *EMBO Rep.* 17, 1721–1730. doi:10.15252/embr.201643300

Deng, W., Rao, J., Chen, X., Li, D., Zhang, Z., Liu, D., et al. (2020). Metabolomics study of serum and urine samples reveals metabolic pathways and biomarkers associated with pelvic organ prolapse. *J. Chromatogr. B* 1136, 121882. doi:10.1016/j.jchromb.2019.121882

Duell, E. J., Sala, N., Travier, N., Munoz, X., Boutron-Ruault, M. C., Clavel-Chapelon, F., et al. (2012). Genetic variation in alcohol dehydrogenase (ADH1A, ADH1B, ADH1C, ADH7) and aldehyde dehydrogenase (ALDH2), alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Carcinogenesis* 33, 361–367. doi:10.1093/carcin/bgr285

Fan, X., Rao, J., Zhang, Z., Li, D., Cui, W., Zhang, J., et al. (2018). Macranthoidin B modulates key metabolic pathways to enhance ROS generation and induce cytotoxicity and apoptosis in colorectal cancer. *Cell Physiol Biochem* 46, 1317–1330. doi:10.1159/000489147

Fearon, E. R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767. doi:10.1016/0092-8674(90)90186-i

Gimm, O., Chi, H., Dahia, P. L. M., Perren, A., Hinze, R., Komminoth, P., et al. (2001). Somatic Mutation and Germline Variants of MINPP1, a Phosphatase Gene Located in Proximity to PTEN on 10q23.3, in Follicular Thyroid Carcinomas1. *J. Clin. Endocrinol. Metab.* 86, 1801–1805. doi:10.1210/jcem.86.4.7419

Habermann, N., Ulrich, C. M., Lundgreen, A., Makar, K. W., Poole, E. M., Caan, B., et al. (2013). PTGS1, PTGS2, ALOX5, ALOX12, ALOX15, and FLAP SNPs: interaction with fatty acids in colon cancer and rectal cancer. *Genes Nutr.* 8, 115–126. doi:10.1007/s12263-012-0302-x

Hammond, W. A., Swaika, A., and Mody, K. (2016). Pharmacologic resistance in colorectal cancer: a review. *Ther. Adv. Med. Oncol.* 8, 57–84. doi:10.1177/1758834015614530

He, J., Jin, Y., Chen, Y., Yao, H.-B., Xia, Y.-J., Ma, Y.-Y., et al. (2016). Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. *Ott* Vol. 9, 6099–6109. doi:10.2147/ott.s110203

Hissong, E., and Pittman, M. E. (2020). Colorectal carcinoma screening: Established methods and emerging technology. *Crit. Rev. Clin. Lab. Sci.* 57, 22–36. doi:10.1080/10408363.2019.1670614

Hong, J., Bose, M., Ju, J., Ryu, J., Chen, X., Sang, S., et al. (2004). Modulation of arachidonic acid metabolism by curcumin and related -diketone derivatives: effects on cytosolic phospholipase A2, cyclooxygenases and 5-lipoxygenase. *Carcinogenesis* 25, 1671–1679. doi:10.1093/carcin/bgh165

Ji, Y., Rao, J., Rong, X., Lou, S., Zheng, Z., and Lu, Y. (2017). Metabolic characterization of human aqueous humor in relation to high myopia. *Exp. Eye Res.* 159, 147–155. doi:10.1016/j.exer.2017.03.004

Khiavi, M. A., Safary, A., and Somi, M. H. (2019). Recent advances in targeted therapy of colorectal cancer: impacts of monoclonal antibodies nanoconjugates. *Bioimpacts* 9, 123.

Kim, E. R., Kwon, H. N., Nam, H., and Kim, Jae. J. (2019). Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal. *Sci. Rep.* 9, 4786. doi:10.1038/s41598-019-41216-y

Koppenol, W. H., Bounds, P. L., and Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* 11, 325–337. doi:10.1038/nrc3038

Kulp, K. S., and Vulliet, P. R. (1996). Mimosine blocks cell cycle progression by chelating iron in asynchronous human breast cancer cells. *Toxicol. Appl. Pharmacol.* 139, 356–364. doi:10.1006/taap.1996.0176

Leithner, K., Hrzenjak, A., Trötzmüller, M., Moustafa, T., Köfeler, H. C., Wohlkoenig, C., et al. (2015). PCK2 activation mediates an adaptive response to glucose depletion in lung cancer. *Oncogene* 34, 1044–1050. doi:10.1038/onc.2014.47

Lewis, A. C., Wallington-Beddoe, C. T., Powell, J. A., and Pitson, S. M. (2018). Targeting sphingolipid metabolism as an approach for combination therapies in haematological malignancies. *Cell Death Discov* 4, 1–11. doi:10.1038/s41420-018-0075-0

Patel, M., Lu, L., Zander, D. S., Sreerama, L., Coco, D., and Moreb, J. S. (2008). ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors. *Lung cancer* 59, 340–349. doi:10.1016/j.lungcan.2007.08.033

Pelosof, L., Yerram, S., Armstrong, T., Chu, N., Danilova, L., Yanagisawa, B., et al. (2017). GPX3 promoter methylation predicts platinum sensitivity in colorectal cancer. *Epigenetics* 12, 540–550. doi:10.1080/15592294.2016.1265711

Peltier, J., Roperch, J. P., Audebert, S., Borg, J. P., and Camoin, L. (2018). Activation peptide of the coagulation factor XIII (AP-F13A1) as a new biomarker for the screening of colorectal cancer. *Clin. Proteomics* 15, 15. doi:10.1186/s12014-018-9191-3

Rao, J., Cheng, F., Hu, C., Quan, S., Lin, H., Wang, J., et al. (2014). Metabolic map of mature maize kernels. *Metabolomics* 10, 775–787. doi:10.1007/s11306-014-0624-3

Rawla, P., Barsouk, A., Hadjinicolaou, A. V., and Barsouk, A. (2019). Immunotherapies and targeted therapies in the treatment of metastatic colorectal cancer. *Med. Sci.* 7, 83. doi:10.3390/medsci7080083

Ritchie, S. A., Ahiahonu, P. W. K., Jayasinghe, D., Heath, D., Liu, J., Lu, Y., et al. (2010). Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection. *BMC Med.* 8, 13. doi:10.1186/1741-7015-8-13

Toubiana, D., Semel, Y., Tohge, T., Beleggia, R., Cattivelli, L., Rosental, L., et al. (2012). Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. *Plos Genet.* 8, e1002612. doi:10.1371/journal.pgen.1002612

Tsai, S.-T., Chien, I.-H., Shen, W.-H., Kuo, Y.-Z., Jin, Y.-T., Wong, T.-Y., et al. (2010). ENO1, a potential prognostic head and neck cancer marker, promotes transformation partly via chemokine CCL20 induction. *Eur. J. Cancer* 46, 1712–1723. doi:10.1016/j.ejca.2010.03.018

Vasaikar, S., Huang, C., Wang, X., Petyuk, V. A., Savage, S. R., Wen, B., et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 177, 1035–e19. doi:10.1016/j.cell.2019.03.030

Wang, B., Li, F., Zhou, X., Ma, Y., and Fu, W. (2019). Is microsatellite instability-high really a favorable prognostic factor for advanced colorectal cancer? A meta-analysis. *World J. Surg. Oncol.* 17, 169. doi:10.1186/s12957-019-1706-5

Ward, D. G., Suggett, N., Cheng, Y., Wei, W., Johnson, H., Billingham, L. J., et al. (2006). Identification of serum biomarkers for colon cancer by proteomic analysis. *Br. J. Cancer* 94, 1898–1905. doi:10.1038/sj.bjc.6603188

Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359.

Xue, X., Bredell, B. X., Anderson, E. R., Martin, A., Mays, C., Nagao-Kitamoto, H., et al. (2017). Quantitative proteomics identifies STEAP4 as a critical regulator of mitochondrial dysfunction linking inflammation and colon cancer. *Proc. Natl. Acad. Sci. USA* 114, E9608–E9617. doi:10.1073/pnas.1712946114

Yang, Q., Roehrl, M. H., and Wang, J. Y. (2018). Proteomic profiling of antibody-inducing immunogens in tumor tissue identifies PSMA1, LAP3, ANXA3, and maspin as colon cancer markers. *Oncotarget* 9, 3996–4019. doi:10.18632/oncotarget.23583

Yun, J., Mullarky, E., Lu, C., Bosch, K. N., Kavalier, A., Rivera, K., et al. (2015). Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells

by targeting GAPDH. *Science* 350, 1391–1396. doi:10.1126/science.aaa5004

Zhang, X., Zhang, X., Li, Y., Shao, Y., Xiao, J., Zhu, G., et al. (2017). PAK4 regulates G6PD activity by p53 degradation involving colon cancer cell growth. *Cell Death Dis* 8–e2820. e2820. doi:10.1038/cddis.2017.85

Zheng, Z., Liu, P., Xu, L., Peng, Z., Zhang, Y., Chen, X., et al. (2018). Metabolomics analysis of salvage chemotherapy on refractory acute myeloid leukemia patients. *RSC Adv.* 8, 14445–14453. doi:10.1039/c7ra13298k

Zheng, Z., Xu, L., Zhang, S., Li, W., Tou, F., He, Q., et al. (2017). Peiminine inhibits colorectal cancer cell proliferation by inducing apoptosis and autophagy and modulating key metabolic pathways. *Oncotarget* 8, 47619–47631. doi:10.18632/oncotarget.17411

Zhou, W., Yang, F., Peng, J., Wang, F., Lin, Y., Jiang, W., et al. (2019). High pretreatment serum CA19-9 level predicts a poor prognosis for patients with stage III colon cancer after curative resection and adjuvant chemotherapy. *J. Cancer* 10, 3810–3818. doi:10.7150/jca.31375

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 Rao, Wan, Tou, He, Xiong, Chen, Cui 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 in Molecular Biosciences

REVIEW
published: 30 July 2021
doi: 10.3389/fmolb.2021.689139

Overview on the Role of E-Cadherin in Gastric Cancer: Dysregulation and Clinical Implications

Huichen Zhao[1], Huihui Hu[1], Beibei Chen[1,2], Weifeng Xu[1], Jing Zhao[1], Chen Huang[1], Yishu Xing[1], Huifang Lv[1], Caiyun Nie[1], Jianzheng Wang[1], Yunduan He[1], Sai-Qi Wang[1,2] and Xiao-Bing Chen[1,2]**

[1]Department of Medical Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2]State Key Laboratory of Esophageal Cancer Prevention and Treatment, Zhengzhou University, Zhengzhou, China

Gastric cancer is the fifth most common cancer and the third most common cause of cancer death all over the world. E-cadherin encoded by human *CDH1* gene plays important roles in tumorigenesis as well as in tumor progression, invasion and metastasis. Full-length E-cadhrin tethered on the cell membrane mainly mediates adherens junctions between cells and is involved in maintaining the normal structure of epithelial tissues. After proteolysis, the extracellular fragment of the full-length E-cadhein is released into the extracellular environment and the blood, which is called soluble E-cadherin (sE-cadherin). sE-cadherin promotes invasion and metastasis as a paracrine/autocrine signaling molecule in the progression of various types of cancer including gastric cancer. This review mainly summarizes the dysregulation of E-cadherin and the regulatory roles in the progression, invasion, metastasis, and drug-resistance, as well as its clinical applications in diagnosis, prognosis, and therapeutics of gastric cancer.

Keywords: gastric cancer, e-cadherin, CDH1, EMT, precision therapy

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China

Reviewed by:
Sayan Chatterjee,
Guru Gobind Singh Indraprastha University, India
Jinliang Xing,
Fourth Military Medical University, China

****Correspondence:***
Xiao-Bing Chen
zlyychenxb0807@zzu.edu.cn
Sai-Qi Wang
zlyywangsq3925@zzu.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *31 March 2021*
Accepted: *19 July 2021*
Published: *30 July 2021*

Citation:
Zhao H, Hu H, Chen B, Xu W, Zhao J, Huang C, Xing Y, Lv H, Nie C, Wang J, He Y, Wang S-Q and Chen X-B (2021) Overview on the Role of E-Cadherin in Gastric Cancer: Dysregulation and Clinical Implications. Front. Mol. Biosci. 8:689139. doi: 10.3389/fmolb.2021.689139

INTRODUCTION

Gastric cancer (GC) is one of the common malignant tumors of the digestive tract. The incidence and mortality of gastric cancer are much higher than the world average, which seriously affects our health (Sung et al., 2021). And it is a multifactorial disease. Both genetic and environmental factors are important to its pathogenesis. Gastric cancer is mostly sporadic, but about 10% of gastric cancers have family clustering characteristics, of which 1/3 are considered to have genetic background. This part of gastric cancer is called hereditary diffuse gastric cancer (HDGC) (Blair et al., 2020). In addition, smoking (Shikata et al., 2008) and alcohol consumption (Deng et al., 2021), eating habits (Takezaki et al., 1999; Kobayashi et al., 2002), chronic atrophic gastritis, and EBV or *Helicobacter pylori* infection (Arif and Syed, 2007; Holleczek et al., 2020) have been considered the most important risk factors for gastric cancer (FORMAN, 1991).

Studies have shown that the initiation and metastasis of gastric cancer is largely related to the loss of E-cadherin expression. E-cadherin encoded by gene *CDH1* has genetic and epigenetic abnormalities in both germline and sporadic gastric cancer, mainly including gene expression level changes, germline and somatic mutations, 16q22.1 allele deletion, promoter methylation andnoncoding RNA (noncoding RNA, ncRNAs) regulated epigenetic gene silencing, etc., leading to abnormal expression of E-cadherin. This paper aims to summarize the relationship between E-cadherin and the tumorigenesis, development, metastasis and drug resistance of gastric cancer

from the aspects of biological function, inactivation mechanism and clinical significance, discussing the clinical application of E-cadherin in the early diagnosis, prognosis, and therapy of gastric cancer as well as development status of E-cadherin activator, which could provide new ideas for precise therapy of gastric cancer.

E-CADHERIN

The Function of E-Cadherin

Cadherin which is a type of cell surface transmembrane glycoprotein has an important influence on the cell-cell adhesion function. E-cadherin belongs to class I classical cadherin, and it is an important part of the intercellular adhesion connection in epithelial tissue (Mendonsa et al., 2018).

Human E-cadherin is encoded by the *CDH1* gene and located on chromosome 16q22.1. The *CDH1* mutation is the only germline molecular mutation related to hereditary diffuse gastric cancer and lobular breast cancer (Gall and Frampton, 2013; Biswas, 2020; Blair et al., 2020). E-cadherin is a transmembrane glycoprotein consisting of three domains: extracellular domains (ECD), transmembrane domains and intracellular domains (ICD). ECD is composed of five cadherin repeats and contains 4 calcium ion binding sites, which mediate the adhesion function of E-cadherin; ICD interacts with α-, β-catenin and other catenin family members. It binds and connects to the cytoskeleton of actin to maintain the stability of cell structure, inhibiting the movement of individual cells, and participating in cell signal transduction (Gall and Frampton, 2013; Biswas, 2020). The abnormal expression of E-cadherin has a significant impact on the interaction between cells, leading to the destruction of the dynamic balance of epithelial tissues, making it easier for cells to gain mobility and invasiveness, making tumors more prone to infiltration and metastasis (Bruner and Derksen, 2018; Na et al., 2020).

Roles of E-Cadherin in Gastric Cancer

The Expression of E-Cadherin in Gastric Cancer

Lots of studies have showed that E-cadherin is a critical tumor suppressor in several carcinomas, including GC (Birchmeier, 1995; Shimada et al., 2012). The decreased expression of E-cadherin was mainly found in diffuse type gastric cancer (Gamboa-Dominguez et al., 2005). 33–50% of sporadic diffuse gastric cancers have somatically inactive E-cadherin mutations. One of the articles depicted 63.6% of signet ring cell carcinoma patients were confirmed E-cadherin mutations (Machado et al., 2001). E-cadherin dysfunction may be contributed by lots of molecular mechanisms, including *CDH1* mutations (Oliveira et al., 2004), DNA hypermethylation (Grady et al., 2000; Machado et al., 2001), loss of heterozygosity (LOH), and non-coding RNAs that regulate E-cadherin expression. Among them, *CDH1* germline mutation is the main cause of diffuse gastric cancer.

And the expression of E-cadherin is an independent factor for the prognosis of gastric cancer. The previous study found that E-cadherin positive patients have better prognosis than that of negative patients (Wijnhoven et al., 2002). In addition, it has been reported in the literature that the prognosis of young patients is worse than that of the elderly (Schildberg et al., 2014).

Role of E-Cadherin in Tumorigenesis

E-cadherin plays an important role in the process of tumorigenesis. It can modulate intracellular signaling to promote tumor growth. Cell adhesion which E-cadherin mediateted, plays a central role in the Wnt signaling pathway. Under physiological conditions, β-catenin in the cytoplasm remains inactive by binding to the APC/GSK3β/Axin/CK1 degradation complex. Wnt signaling inhibits the degradation process by inhibiting the GSK3β complex through phosphorylation (Kourtidis et al., 2013; Zhan et al., 2017; Bruner and Derksen, 2018).The complex of the combination of E-cadherin and β-catenin has a negative regulatory effect on the β-catenin/Wnt pathway. Bruner HC et al. found that the increased density of extracellular matrix could destroy the E-cadherin/β-catenin complex of gastric cancer cells, thereby regulating the proliferation and the response to chemotherapy of gastric cancer (Jang et al., 2018). Moreover, Notch signaling is a critical way in tumorigenesis. It can induce the cyclooxygenase-2 (COX-2) expression through the binding of the Notch1 receptor intracellular domain to the COX-2 promoter, leading to tumor progression (Yeh et al., 2009). And inhibiting Notch signaling could induce G2/M cell cycle arrest through activating nuclear PTEN in gastric cancer (Kim et al., 2016).

In addition, changes of E-cadherin could cause the destruction of cell adhesion, leading to changes in loss of contact inhibition, and changes in cell migration and matrix interactions, thereby resulting in tumorigenesis (Berx and van Roy, 2009). E-cadherin deletion is regulated by a variety of factors, including changes in gene and mRNA levels. For example, *CDH1* germline mutations are closely related to the occurrence of hereditary diffuse gastric cancer (Guilford et al., 1998; Oliveira et al., 2004). In addition, the promoter DNA hypermethylation region and E-cadherin repressors (Machado et al., 2001), such as Snail, ZEB2/SIP1, and Slug, can combine with the E-box region of E-cadherin to induce E-cadherin inactivation, thereby promoting tumorigenesis (Batlle et al., 2000; Cano et al., 2000; Comijn et al., 2001; Conacci-Sorrell et al., 2003). Therefore, we will introduce these four inactivating mechanisms in E-Cadherin clearly. A better understanding of E-cadherin inactivation would provide an opportunity for future therapeutic intervention.

Gene Mutation

Loss of E-cadherin expression is closely related to the sporadic and genetic forms of gastric cancer, and *CDH1* mutation is one of the important reasons for the loss of E-cadherin expression (Corso et al., 2013). *CDH1* mutations include germline mutations and somatic mutations. Germline mutations can affect the entire coding sequence, including small frame divisions, splice sites, meaningless, missense mutations, and large rearrangements. According to reports, compared with missense mutations, non-missense mutations may be pathogenic, and the frequency of missense mutations is higher than that of nonsense mutations (Corso et al., 2011;

Simões-Correia et al., 2012; Choi et al., 2020). High-frequency mutations at multiple sites in *CDH1* are one of the risk factors and signs of hereditary diffuse gastric cancer. And about 50–80% of diffuse gastric cancers have decreased or missing E-cadherin expression. Diffuse gastric cancer with *CDH1* mutation has more aggressive phenotypic characteristics, especially with a higher Ki67 marker index, p53 mutation and Her-2 positive gastric cancer (Muzashvili et al., 2020). And the clinical trail which is called "Hereditary Gastric Cancer Syndromes: An Integrated Genomic and Clinicopathologic Study of the Predisposition to Gastric Cancer" is recruiting.

CDH1 gene mutation or transcription silencing is related to familial diffuse gastric cancer. Research confirms that *CDH1* mutations are more likely to be found in countries with a low incidence of gastric cancer.Thus preventive genetic screening is very important (Corso et al., 2021). And due to the poor sensitivity of gastroscopy for the detection of signet ring cell carcinoma, the monitoring of *CDH1* mutation patients is very limited. Therefore, prophylactic total gastrectomy may be the most desirable option for individuals with *CDH1* mutations and a family history of diffuse gastric cancer (Shenoy, 2019; Castro et al., 2020).

In addition to these inactivating *CDH1* mutation, the occurrence of gastric cancer is related to the polymorphisms of several single nucleic acid glycosides in the *CDH1* gene. Among them, the relation of *CDH1*-160C >a polymorphism to GC is currently the most extensively studied. In the present study, it found that the *CDH1*-160 AA genotype could increase the risk of gastric cancer (Zhang et al., 2008; Al-Moundhri et al., 2010).

DNA Methylation

DNA methylation is the addition of methyl groups at C5 of the cytosine ring to produce 5-methylcytosine. DNA methylation can control gene expression by changing chromatin structure, DNA conformation, DNA stability, and the way that DNA interacts with proteins. For example, the methylation of CpG islands in tumor suppressor gene promoters causes them to prevent transcription factors from entering the binding site in the promoter through steric hindrance, thereby inhibiting gene transcription. Promoter hypermethylation is considered to be another major mechanism for *CDH1* inactivation in the development of various cancers, including gastric cancer (Qu et al., 2013). And 79% of diffuse gastric cancer has found the promoter CpG hypermethylation, but only 55% of intestinal gastric cancer found it (Yoshiura et al., 1995; Ushijima and Sasako, 2004). Thus, according to the well-known two-hit inactivation mechanism proposed by Knudson, hypermethylation of the promoter CpG could be the second hit in abrogating E-cadherin expression (Carbone and Minna, 1993). IL-1b is demonstrated to be an important step in mediating E-cadherin methylation in *Helicobacter pylori*-related gastric cancer (Qian et al., 2008).

DNA methyltransferase 3A isoform B (DNMT3Ab) has been found to play a crucial role in the EMT process of gastric cancer. DNMT3Ab mediates epigenetic inactivation of E-cadherin gene through DNA hypermethylation and histone modification of H3K9me2 and H3K27me3. DNMT3Ab deletion effectively restored the expression of E-cadherin by reducing the methylation, H3K9me2, and H3K27me3 levels on the e-cadherin promoter. These results confirm that the DNMT3Ab targeting the DNMT3Ab/Snail/E-cadherin axis may provide a promising therapeutic strategy for the treatment of metastatic gastric cancer with high expression of DNMT3Ab (Cui et al., 2018).

Non-coding RNA

At present, more and more evidences show that non-coding RNA, including micro RAN (microRNA, miRNA) and long non-coding RNA (long non-coding RNA, lncRNA), play a vital role in carcinogenesis. So far, people have found a total of 2654 miRNAs described in the human genome (Kozomara et al., 2019). It is reported that miRNA can regulate the signal pathways related to E-cadherin, and inhibit the occurrence and development of gastric cancer by inhibiting the EMT process. Therefore, it is considered as a potential clinical biomarker for the treatment of gastric cancer (Lazar et al., 2008). For example, the increase of miR-1275 expression regulates the expression of vimentin/E-cadherin by directly inhibiting the expression of JAZF1, thereby inhibiting the metastasis of gastric cancer (Mei et al., 2019). It is reported that the expression of miRNA-203 is reduced in gastric cancer. Inhibition of miR-203 expression will increase the expression of phospho-ERK1/2 (pERK1/2) and Slug, as well as reduce the expression of E-cadherin, thereby promoting tumorigenesis (Gao et al., 2017a). Studies have shown that miR-21 inhibitors can inhibit tumor growth by up-regulating the expression of E-cadherin and PTEN, and down-regulating the expression of N-cadherin, β-catenin, vimentin, and Slug, suggesting that miR-21 may increase the expression of β-catenin. Promote endoderm transformation of gastric cancer cells induced by transforming growth factor PTEN1 (Li et al., 2016). MiR-340 targets SPP1 to inhibit the PI3K/AKT pathway, and activates the Snail/Slug signaling pathway, further inhibiting the proliferation, migration, invasion and EMT of gastric cancer cells (Song et al., 2019a; Chen et al., 2019). MiR-19a/miR-96-mediated low expression of KIF26A suppresses metastasis by regulating FAK pathway in gastric cancer (Ma et al., 2021). In intestinal gastric cancer, the loss of microRNA-101 can lead to E-cadherin functional downregulation through EZH2 up-regulation (Carvalho et al., 2012).

LncRNA is the main non-coding RNA complex with chromatin-modifying proteins. Previous studies indicated that lncRNA participates in the occurrence and progression of gastric cancer through regulating gene expressions at epigenetic, transcriptional or post-transcriptional levels. Therefore, they are considered as potential diagnostic hallmarkers (Yang et al., 2015; Yao et al., 2017). HOX antisense intergenic RNA (HOTAIR) is one of the earlistlncRNAs, and it is still one of the most widely studied. HOTAIR has been found to be upregulated in gastric cancer, and can promote tumor metastasis by binding to EZH2 with the E-cadherin promoter (Chen et al., 2017; Song et al., 2019b). Studies have found that HOTAIR and its combination with PRC2 inhibited the expression of miR34a, which controls the target genes C-Met (HGF/C-Met/Snail pathway) and Snail, thereby helping gastric

cancer cells to develop EMT and promote tumor metastasis (Liu et al., 2015). Thus, the expression of HOTAIR was seemed to be an independent predictor of overall survival (Chen et al., 2017).The role of lncRNA CHRF promotes cell invasion and migration through EMT, which may provide a potential target for the biological diagnosis and therapy of gastric cancer (Gong et al., 2020). DCLK1, miR-15b, and lncRNA SNHG1 play a potential role in the occurrence of gastric cancer. Overexpression of lncRNA SNHG1 can promote the expression of DCLK1 and Nothc1 in gastric cancer cells. MiR-15b targets DCLK1 to regulate the expression of Notch1 and inhibit the EMT process of gastric cancer cells. LncRNA SNHG1 enhances the role of DCLK1/Notch1 in the EMT process by regulating the expression of miR-15b (Liu et al., 2020). LncRNA CCAT2 promotes gastric cancer proliferation and invasion through regulating the E-cadherin and LATS2 (Wang et al., 2016).And LINC00240 promotes GC tumorigenesis *via* a LINC00240/miR-124-3p/DNMT3B axis as an oncogene, so it may be a potential diagnostic biomarker for GC (Lazar et al., 2008).

Role of E-Cadherin in Metasitasis

Metastasis is one of the key molecular steps affecting the prognosis of advanced GC patients. And Epithelial-mesenchymal transition (EMT) is one of the important steps in the process related to metastasis in gastric cancer (Valastyan and Weinberg, 2011).EMT refers to the phenomenon that epithelial cells transform into mesenchymal cells under specific physiological and pathological condition (Aban et al., 2021). And it can down-regulate the epithelial markers E-cadherin, ZO-1, etc., which can inhibite the characteristics and behavior of epithelial cells. At the same time, the mesenchymal markers N-cadherin and vimentin are up-regulated, and then activating the mesenchymal characteristics of the cell to relax the cells from the basement membrane, which is more spindle-shaped and motility in phenotype, and at the same time obtains apoptosis resistance. Therefore, EMT is considered to be one of the main mechanisms that determine the spread of invasive and metastatic cancer cells (Nieto et al., 2016).

The decrease of E-cadherin expression on the cell membrane leads to weakening or disappearance of the interaction between cells and inhibit the activation of transcription factors (Snail, Slug, Twist and ZEB-1), resulting in EMT (Kalluri and Weinberg, 2009; Aban et al., 2021). Studies have found that Snail family transcription factors are strong repressors of E-cadherin gene transcription. Epithelial cells that ectopicly express Snail have a fibroblast phenotype, and are tumorigenic and invasive. Therefore, the loss of E-cadherin protein may be due to germline mutations observed in diffuse gastric cancer, or it may be due to the over-expression of transcription inhibitors (Snail, Slug, Twist, ZEB-1) in advanced cancers.This phenomenon is transcription silence (Cano et al., 2000). And some studies found that Snail, Slug and ZEB1 expression were related to tumor differentiation, lymph node metastasis and pathological staging (Uchikado et al., 2011; Chen et al., 2016; Okubo et al., 2017; Xue et al., 2019). And EMT can activate β-catenin signaling pathway to promote tumor metastasis in gastric cancer (Park et al., 2015). It is demonstrated that abnormal activation of the hedgehog (Hh) signaling pathway may be involved in inducing EMT in gastric cancer.And the levels of the Hh pathway marker Gli-1 were associated with levels of the Snail and E-cadherin.All three markers were related to depth of invasion, lymph node metastasis, and pTNM stage in gastric cancer (Wang et al., 2014).

In addition, there are some signaling pathways that promote tumor invasion and metastasis by regulating the expression of E-cadherin. The Rho A, Rac1, and Cdc42. play a critical role in mataining the cytoskeleton, increasing cell fluidity, and taking part in the phenotype of cell mesenchyme. Increase of the RhoA aactivity resulted in the cancer cell proliferation and cell cycle disorders. Previous studies have shown that E-cadherin missense mutations associated with diffuse gastric cancer can induce an increase in RhoA activity, resulting in the proliferation and movement of gastric cancer cells (Suriano et al., 2003). In addition, E-cadherin can directly activate the Rho GTPase pathway, and it can also activate Rho GTPase through EGFR (Bremm et al., 2008). A recent study found that RhoA Y42 mutation is associated with poor prognosis of gastric cancer. Y42C mutation of RhoA has a higher protein level, which can promote the proliferation and movement of gastric cancer cells.As for EGFR, its activation is associated with the E-cadherin expression. And E-cadherin's downstream effector molecules related to RAS and RAF/MEK pathways and other tumor-forming pathways such as FAK/C-SRC and PI3K/AKT/mTOR pathways, thus contributing to tumor cell proliferation and metasitasis. In addition, the loss of E-cadherin can increase β-catenin into the nucleus, thereby inhibiting PTEN expression, and then participating in the promotion of tumorigenesis through the AKT/mTOR pathway (Bremm et al., 2008). MMPs alterations are also important to tumor invasion and metastasis in gastric cancer. For example, in gastric cancer, the MMP1 and MMP2 increased expression is related to the loss of E-cadherin expression, thereby, leading cancer proferliation and metastasis (Zhou et al., 2010).MMP-3 and MMP-7 have association with the development of *Helicobacter pylori*-related gastric cancer (Yang et al., 2018).

Role of E-Cadherin in Drug Resistance

Previous studies indicated that EMT played a critical role in drug resistance (Singh and Settleman, 2010; Chen et al., 2014; Lee et al., 2014). And the loss of E-cadherin expression is a key step in the EMT process. E-cadherin is significantly down-regulated or absent in tumor drug-resistant cells, which can increase the sensitivity of tumor cells to anti-tumor drugs (Selga et al., 2008; Gao et al., 2017b). E-cadherin can reverse the resistance of transformation therapy drugs by negatively regulating the expression of BCL-2, and up-regulate the expression of cell cycle inhibitor P27, exerting a proliferation inhibitory effect. It can also inhibit tumor cells by increasing the expression of tumor suppressor gene PTEN protein (Yang et al., 2008). The latest research data show that microRNA-421 regulated by hypoxia-inducible factor-1α promotes gastric cancer metastasis by targeting E-cadherin and caspase-3, inhibits cell apoptosis, and induces cisplatin resistance (Ge et al., 2016). Eukaryotic translation initiation factor 5A2 (EIF5A2) (by up-regulating

epithelial markers E-cadherin and β-cadherin, down-regulating mesenchymal markers Vimentin and N-cadherin mediate the EMT process to regulate the resistance of gastric cancer cells to cisplatin (Sun et al., 2018).

E-Cadherin and Helicobacter Pylori-Related Gastric Cancer

Helicobacter pylori (*Helicobacter pylori*, *H.pylori*) is one of the most common pathogens in humans. *H.pylori* infection can dysregulate Wnt signaling pathway and cause EMT in gastric tissue, then, increase the risk of gastric cancer. It induces a variety of inflammatory reactions by infiltrating macrophages, neutrophils, regulatory T cells and natural killer cells, which significantly affects the gastric microenvironment (Baj et al., 2020). *H. pylori* infection Cytoxin-associated gene A (CagA) is a HP virulence factor. Researc5h data found that *Helicobacter pylori* can induce epithelial-mesenchymal transition of gastric epithelial cells through CagA, thus showing characteristics similar to tumor stem cells (Bessède et al., 2014). VacA, another *H. pylori* virulence factor, could activate PI3K/Akt pathway and then induce Wnt/β-catenin signaling, leading to phosphorylation of GSK3β and translocation of the β-catenin to the nucleus to activate CCND1 gene (Nakayama et al., 2009). *H.pylori* infection can reduce the function of E-cadherin by activating the matrix MMP-3 and -7, and induce the migration and invasion of gastric cancer cells. It can also regulate miRNA, such as miR-128/miR-148a, and affect the MMPs/E-cadherin signaling pathway, thereby promoting the occurrence of tumors (Yang et al., 2018).In addition, *H.pylori* could initiate E-cadherin methylation which may subsequently progress to intestinal metaplasia and invasive cancer.But it need to be further investigated (Chan et al., 2003a; Qian et al., 2008).

Nuclear factor kappa B (NF-κB) signaling way is the most important way in *H.pylori* -related gastric cancer (DiDonato et al., 2012). NF-κB is a group of transcription factors (RelA, RelB, c-Rel, NF-κB1/p50, and NF-κB2/p52), which form homo- and heterodimers and upregulate or suppress expression of many genes (Neumann and Naumann, 2007). The NF-κB-driven gene products include cytokines/chemokines IL-1, IL-8, TNF, IL-6, MCP-1, pro- ,and anti-apoptotic factors cIAPs, c-FLIP, A20, Bcl-XL, angiogenesis regulator vascular endothelial growth factor (VEGF), matrix metalloproteinases (MMP)-2, MMP-9 in non-transformed or tumor cells in response to a variety of stimuli, including growth factors, cytokines, hormones, microbial and chemical compounds (Maeda and Omata, 2008). In a *H. pylori* infection gastric cell line, the infection can promote NF-κB- and AKT-mediated MMP-9 production, which lead to cell migration and invasion (Maubach et al., 2013). And previous study has demonstrated that NF-κB decreases the expression of E-cadherin in gastric cancer through regulation of other transcription factors, promoting EMT, tumor angiogenesis and metastatic dissemination (Nam et al., 2011; Hu et al., 2013).

CLINICAL SIGNIFICANCE OF E-CADHERIN

E-cadherin plays an important role in cell connection, and loss of E-cadherin is crucial in the occurrence and development of gastric cancer. Data show that E-cadherin deletion is often associated with poorly differentiated cancer, lymph node metastasis, and tumor staging (Gao et al., 2017b). Therefore, evaluating the protein level of E-cadherin and changes in the *CDH1* gene may provide promising prospects for the diagnosis, prognosis, and treatment of gastric cancer.

E-Cadherin Can be Used as a Potential Marker for the Diagnosis of Gastric Cancer

Kahtan Al-Bayaty, Mustafa and others tested the serum E-cadherin status of 30 gastritis patients, 20 gastric ulcer patients, and 20 gastric cancer patients. They found that the E-cadherin level of all patients increased, and the gastric cancer group was significantly higher than the other groups. Increase, so E-cadherin can be used as a potential marker for the diagnosis of gastric cancer (Carneiro et al., 2012). The extracellular domain of E-cadherin can be broken down into 80 kDa fragments by proteases under certain pathological stimuli. These fragments can be soluble E-cadherin (sE-cadherin). In patients with intestinal and diffuse gastric cancer, serum soluble E-cadherin also presents a completely different pattern. The level of serum E-cadherin increases in intestinal gastric cancer, while in diffuse gastric cancer, especially in the advanced stage, its level decreases, so it may be a biomarker for the diagnosis of intestinal gastric cancer (Juhasz et al., 2003).

E-Cadherin Can be Used as a Predictor of Gastric Cancer Prognosis

The abnormal expression of E-cadherin is related to the aggressiveness of gastric cancer, suggesting that this marker may serve as a negative prognostic factor for gastric cancer (Lazar et al., 2008; Xia et al., 2017; Kumar et al., 2021). Zhila Torabizadeh et al. detected the protein expression level of E-cadherin in the tumor tissues and adjacent tissues of 70 patients with gastric cancer. And they found among these 70 patients, 48.6% showed abnormal E-cadherin expression.the abnormal expression of E-cadherin has a strong correlation with tumor stage, tumor grade, depth of invasion and local lymph node involvement, this marker can be used as a predictor of tumor aggressiveness in gastric cancer (Torabizadeh et al., 2017; Wool Eom et al., 2020).

It can be seen that detecting the expression of E-cadherin or the alteration of the *CDH1* gene encoded by it may provide promising applications for the diagnosis, prognosis or treatment targets of gastric cancer. E-cadherin and β-cadherin are often more related to advanced disease and poor prognosis. Chan AO et al. measured the expression of soluble E-cadherin in 116 patients with gastric adenocarcinoma included in the trial, and found that soluble E-cadherin can be used as a long-term predictor of cancer patients. It is an independent factor of survival, and it can be a potentially valuable prognostic factor

TABLE 1 | Compounds that restore E-cadherin protein expression in gastric cancer.

Drug	References	Characteristic	Function and mechanism
Astragaloside	Zhu and Wen, (2018)	A lanolin alcohol type tetracyclic triterpene saponins and one of the main active ingredients of Astragalus	Inhibit PI3K/Akt/NF-κB pathway and inhibit the conversion of E-cadherin to N-cadherin induced by TGF-β1
Celecoxib	Zhou et al. (2007)	Selective COX-2 inhibitor	Induce apoptosis a, inhibit angiogenesis, and increase the expression of
Cariin mesylate eribulin	Kurata et al. (2018)	Flavonoid glycosides derived from plants of the genus epimedium	Down-regulate the TGF-β/Smad signaling pathway
Curcumin	Liu et al. (2019)	A diketone compound extracted from plant rhizomes	Downregulate the expression of N-cadherin, snail1, Wnt3a, p-β-catenin, p-LRP6 and Bcl-2, and upregulate the expression of E-cadherin and bax, and increase the activity of caspase-3, caspase-8, caspase-9
Decitabine	Li et al. (2015)	DNA methylation inhibitor	Upregulation of NEDD4-1
Dihydroartemisinin	Li et al. (2019)	Artemisinin derivatives	Down-regulate the activities of PI3K/AKT and snail, and inhibit the epithelial-mesenchymal transition of gastric cancer cells
Dihydromyricetin (DHM)	Wang et al. (2019)	Traditional Chinese medicine with extensive anti-tumor effects	Up-regulate E-cadherin and down-regulates vimentin through the JNK/MMP-2 pathway
Estrogen derivative megestrol and other estrogen receptor modulators	Shimada et al. (2018)	Estrogen receptor modulator	Induce apoptosis before DNA damage
Luteolin	Zang et al. (2017)	A flavonoid found in fruits and green plants	Inhibite Notch1 signal transduction
Metformin	Valaee et al. (2017)	An anti-diabetic drug for treating type 2 diabetes	Decrease mesenchymal markers (including vimentin and β-catenin), and induce epithelial markers (E-cadherin)
MLN4924	Lan et al. (2016)	NEDD8-activating enzyme inhibitor	Activate E-cadherin and inhibit MMP-9
Oxamflatin	Faghihloo et al. (2016)	HDAC inhibitor	Induce E-cadherin expression and reduce cell viability
Triallyl trisulfide (DATS)	Jiang et al. (2017)	Garlic extraction	Increase the phosphorylation of cyclin A2, cyclin B1, JNK, ERK and p38, activate the MAPK pathway, up-regulate the expression of E-cadherin and down-regulate the expression of MMP-9

for patients with gastric cancer before treatment (Chan et al., 2003b).

DEVELOPMENT OF E-CADHERIN ACTIVATOR

Although EMT is a complex process that is regulated by multiple genes, the loss of E-cadherin is one of the most important characteristics. Therefore, compounds that target to restore E-cadherin expression may become potential EMT reversal agents and provide new ideas for inhibiting tumor metastasis.

So far, many compounds have been found to restore E-cadherin expression. Here, as shown in **Table 1**, we summarize the compounds that can restore E-cadherin protein expression and inhibit EMT in gastric cancer. In the HDGC, hypermethylation of the promoter CpG could be the second hit in abrogating E-cadherin expression, therefore, the histone deacetylase (HDAC) inhibitors and DNA-demethylating agents has become potential drugs that can regulate expression of E-cadherin (Grady et al., 2000). HDAC inhibitors are able to block substrate recognition and inducing gene expression (Sharma et al., 2010). Oxamflatin, the HDAC inhibitors can induce E-cadherin expression and reduce cell viability in gastric cancer. Thus, it can be further considered for the prevention of tumor metastasis (Faghihloo et al., 2016). DNA methylation inhibitors are nucleoside analogs that can inhibit DNA methylation by trapping DNA methyltransferases, leading to their depletion inside the cell (Egger et al., 2004). Decitabine (DAC), a DNA methylation inhibitor, has been demonstrated to promote gastric cancer cell migration and invasion *via* the upregulation of NEDD4-1 (Li et al., 2015). Cyclooxygenase-2 (COX-2) participates in cancer invasion and metastasis by regulating E-cadherin expression through the NF-κB/Snail signaling pathway in gastric cancers (Chen et al., 2013; Liu et al., 2013). Thus COX-2 inhibitor have been shown to be chemopreventive against gastric cancer (Ohno et al., 2001; Hu et al., 2004). Celecoxib, a COX-2 selective inhibitor, has been demonstrated to induce apoptosis and inhibit angiogenesis of gastric cancer and then has an inhibitory impact on E-cadherin, resulting in suppressing the invasion of advanced gastric cancer (Zhou et al., 2007). And the potential application of Allium genus to GC chemoprevention and treatment support through *CDH1* restoration and COX2 downregulation. In addition, an estrogen derivative megestrol and other estrogen receptor modulators can specifically inhibit the viability of gastric cancer cells by inducing apoptosis before DNA damage (Shimada et al., 2018). The latest evidence shows that luteolin affects cell proliferation, migration, apoptosis and reverses EMT by inhibiting Notch1 pathway, thereby inhibiting the progression of gastric cancer (Zang et al., 2017). Metformin can inhibit EMT in a glucose-independent manner, thereby inhibiting the invasion and migration ability of gastric cancer cells (Valaee et al., 2017). The recently research discovered that, NEDD8-Activating

Enzyme (NEDD8-Activating Enzyme, NAE) inhibitor MLN4924 can significantly inhibit the migration of gastric cancer cells through transcriptional activation of E-cadherin and inhibition of MMP-9 (Lan et al., 2016). In addition, the natural compound triallyl trisulfide (DATS) extracted from garlic can inhibit gastric cancer cell metastasis by up-regulating E-cadherin and down-regulating MMP-9 (Jiang et al., 2017). Dihydromyricetin (DHM) up-regulates E-cadherin and down-regulates Vimentin through the JNK/MMP-2 pathway, inhibiting the migration and invasion of gastric cancer cells (Wang et al., 2019). Icariin mesylate (Eribulin) can inhibit the EMT changes of triple-negative breast cancer. Kurata T et al. further found that in gastric cancer cells, it can inhibit EMT by regulating the TGF-β/Smad signaling pathway. Icariin sulfonate combined with 5-FU may be a promising treatment option for the treatment of gastric cancer peritoneal metastasis (Kurata et al., 2018). Astragaloside inhibits the conversion of E-cadherin to N-cadherin induced by TGF-β1 by inhibiting the PI3K/Akt/NF-κB pathway to inhibit the survival, invasion and metastasis of gastric cancer cells (Zhu and Wen, 2018). Li Nan et al. found that dihydroartemisinin can effectively inhibit the malignant proliferation of gastric cancer cells, down-regulate the activities of PI3K/AKT and Snail, and inhibit the epithelial-mesenchymal transition of gastric cancer cells (Li et al., 2019). And the research reveals that, the curcumin could inhibit the migration and invasion of gatric cancer cells, downregulate the expression of N-cadherin, snail1, Wnt3a, p-β-catenin, p-LRP6, and Bcl-2, and upregulate the expression of E-cadherin and Bax. Therefore, it might provide potential strategies for gastric cancer treatment (Liu et al., 2019).

SUMMARY AND OUTLOOK

Although the current clinical therapy of gastric cancer has made great progress, its effect is far from satisfactory. In China, owing to lack of early symptoms obvious and biomarkers for early diagnosis, about 75% of gastric cancer have been at advanced stage when diagnosed, and usually accompanied by metastasis. Therefore, early diagnosis of gastric cancer is important to reduce mortality. E-cadherin maintains the integrity of the epithelial cell layer through cell adhesion. The decreased expression of E-cadherin in gastric cancer is related to the malignancy and poor prognosis of gastric cancer patients. The study found that soluble E-cadherin could be used as an independent factor to predict the long-term survival of tumor patients, and it could be used as a biomarker for the early diagnosis of gastric cancer, with broad application prospect. Because patients with advanced gastric cancer have poor therapeutic effect and poor sensitivity to radiotherapy, chemotherapy, immunotherapy, etc., precise targeted therapy is more important for the treatment of gastric cancer. In gastric cancer, diffuse gastric cancer is more closely related to E-cadherin expression and *CDH1* mutations. Diffuse gastric cancer with *CDH1* mutation has more aggressive phenotypic characteristics, and targeted E-cadherin therapy may provide new ideas for inhibiting tumor metastasis, which can be used as a potential target for future gene and genetic therapy and has important value.

AUTHOR CONTRIBUTIONS

H-CZ took responsibility for the topic selection and design, thesis writing and revision. H-H, B-BC, JZ and F-WX performed writing and modifying the manuscript. X-BC and S-QW designed the topic, supervised the entire study, and revised the manuscript. CH, Y-SX, H-FL, CY-N, JZ-W and Y-DH revised the manuscript for intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

We would like to thank the financial support from the National Natural Science Foundation of China (No. 81472714), 1,000 Talents Program of Central plains (No. 204200510023), Project of international scientific and technological cooperation research of Henan Province (No. 182102410023), Science and Technique Foundation of Henan Province (No. 202102310413), Medical Science and Technique Foundation of Henan Province (Nos. 2018020486 and SB201901101), Young and middle-aged Health and Technology Innovation Leading talent Project of Henan Province (YXKC2020008), Natural Science Foundation of Henan Province (No. 212300410270), State Key Laboratory of Esophageal Cancer Prevention and Treatment (No. Z2020000X) and Science and Technique Foundation of Henan Province (No. 202102310121).

REFERENCES

Aban, C. E., Lombardi, A., Neiman, G., Biani, M. C., La Greca, A., Waisman, A., et al. (2021). Downregulation of E-Cadherin in Pluripotent Stem Cells Triggers Partial EMT. *Sci. Rep.* 11 (1), 2048. doi:10.1038/s41598-021-81735-1

Al-Moundhri, M. S., Al-Khanbashi, M., Al-Kindi, M., Al-Nabhani, M., Burney, I. A., Al-Farsi, A., et al. (2010). Association ofE-cadherin(CDH1) Gene Polymorphisms and Gastric Cancer Risk. *Wjg* 16 (27), 3432–3436. doi:10.3748/wjg.v16.i27.3432

Arif, M., and Syed, S. (2007). Association of *Helicobacter pylori* with Carcinoma of Stomach. *J. Pak Med. Assoc.* 57 (7), 337–341.

Baj, J., Brzozowska, K., Forma, A., Maani, A., Sitarz, E., and Portincasa, P. (2020). Immunological Aspects of the Tumor Microenvironment and Epithelial-Mesenchymal Transition in Gastric Carcinogenesis. *Ijms* 21 (7), 2544. doi:10.3390/ijms21072544

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., et al. (2000). The Transcription Factor Snail Is a Repressor of E-Cadherin Gene Expression in Epithelial Tumour Cells. *Nat. Cell Biol* 2 (2), 84–89. doi:10.1038/35000034

Berx, G., and van Roy, F. (2009). Involvement of Members of the Cadherin Superfamily in Cancer. *Cold Spring Harbor Perspect. Biol.* 1 (6), a003129. doi:10.1101/cshperspect.a003129

Bessède, E., Staedel, C., Acuña Amador, L. A., Nguyen, P. H., Chambonnier, L., Hatakeyama, M., et al. (2014). *Helicobacter pylori* Generates Cells with Cancer Stem Cell Properties via Epithelial-Mesenchymal Transition-like Changes. *Oncogene* 33 (32), 4123–4131. doi:10.1038/onc.2013.380

Birchmeier, W. (1995). E-cadherin as a Tumor (Invasion) Suppressor Gene. *Bioessays* 17 (2), 97–99. doi:10.1002/bies.950170203

Biswas, K. H. (2020). Molecular Mobility-Mediated Regulation of E-Cadherin Adhesion. *Trends Biochem. Sci.* 45 (2), 163–173. doi:10.1016/j.tibs.2019.10.012

Blair, V. R., McLeod, M., Carneiro, F., Coit, D. G., D'Addario, J. L., van Dieren, J. M., et al. (2020). Hereditary Diffuse Gastric Cancer: Updated Clinical Practice Guidelines. *Lancet Oncol.* 21 (8), e386–e397. doi:10.1016/S1470-2045(20)30219-9

Bremm, A., Walch, A., Fuchs, M., Mages, J., Duyster, J., Keller, G., et al. (2008). Enhanced Activation of Epidermal Growth Factor Receptor Caused by Tumor-Derived E-Cadherin Mutations. *Cancer Res.* 68 (3), 707–714. doi:10.1158/0008-5472.CAN-07-1588

Bruner, H. C., and Derksen, P. W. B. (2018). Loss of E-cadherin-dependent Cell-Cell Adhesion and the Development and Progression of Cancer. *Cold Spring Harb Perspect. Biol.* 10 (3), a029330. doi:10.1101/cshperspect.a029330

Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., et al. (2000). The Transcription Factor Snail Controls Epithelial-Mesenchymal Transitions by Repressing E-Cadherin Expression. *Nat. Cell Biol* 2 (2), 76–83. doi:10.1038/35000025

Carbone, D. P., and Minna, J. D. (1993). Antioncogenes and Human Cancer. *Annu. Rev. Med.* 44, 451–464. doi:10.1146/annurev.me.44.020193.002315

Carneiro, P., Fernandes, M. S., Figueiredo, J., Caldeira, J., Carvalho, J., Pinheiro, H., et al. (2012). E-cadherin Dysfunction in Gastric Cancer - Cellular Consequences, Clinical Applications and Open Questions. *FEBS Lett.* 586 (18), 2981–2989. doi:10.1016/j.febslet.2012.07.045

Carvalho, J., van Grieken, N. C., Pereira, P. M., Sousa, S., Tijssen, M., Buffart, T. E., et al. (2012). Lack of microRNA-101 Causes E-Cadherin Functional Deregulation through EZH2 Up-Regulation in Intestinal Gastric Cancer. *J. Pathol.* 228 (1), a–n. doi:10.1002/path.4032

Castro, R., Lobo, J. o., Pita, I. s., Videira, F., Pedro-Afonso, L., Dinis-Ribeiro, M., et al. (2020). Random Biopsies in Patients Harboring a CDH1 Mutation: Time to Change the Approach? *Rev. Esp Enferm Dig.* 112 (5), 367–372. doi:10.17235/reed.2020.6720/2019

Chan, A. O.-O., Chu, K.-M., Lam, S.-K., Wong, B. C.-Y., Kwok, K.-F., Law, S., et al. (2003b). Soluble E-Cadherin Is an Independent Pretherapeutic Factor for Long-Term Survival in Gastric Cancer. *Jco* 21 (12), 2288–2293. doi:10.1200/JCO.2003.08.078

Chan, A. O.-O., Lam, S., Wong, B. C., Wong, W., Yuen, M., Yeung, Y., et al. (2003a). Promoter Methylation of E-Cadherin Gene in Gastric Mucosa Associated with *Helicobacter pylori* Infection and in Gastric Cancer. *Gut* 52 (4), 502–506. doi:10.1136/gut.52.4.502

Chen, D.-D., Cheng, J.-T., Chandoo, A., Sun, X.-W., Zhang, L., Lu, M.-D., et al. (2019). microRNA-33a Prevents Epithelial-Mesenchymal Transition, Invasion, and Metastasis of Gastric Cancer Cells through the Snail/Slug Pathway. *Am. J. Physiology-Gastrointestinal Liver Physiol.* 317 (2), G147–G160. doi:10.1152/ajpgi.00284.2018

Chen, D., Huang, J., Zhang, K., Pan, B., Chen, J., De, W., et al. (2014). MicroRNA-451 Induces Epithelial-Mesenchymal Transition in Docetaxel-Resistant Lung Adenocarcinoma Cells by Targeting Proto-Oncogene C-Myc. *Eur. J. Cancer* 50 (17), 3050–3067. doi:10.1016/j.ejca.2014.09.008

Chen, W.-M., Chen, W.-D., Jiang, X.-M., Jia, X.-F., Wang, H.-M., Zhang, Q.-J., et al. (2017). HOX Transcript Antisense Intergenic RNA Represses E-Cadherin Expression by Binding to EZH2 in Gastric Cancer. *Wjg* 23 (33), 6100–6110. doi:10.3748/wjg.v23.i33.6100

Chen, X., Li, J., Hu, L., Yang, W., Lu, L., Jin, H., et al. (2016). The Clinical Significance of Snail Protein Expression in Gastric Cancer: a Meta-Analysis. *Hum. Genomics* 10 (S2), 22. doi:10.1186/s40246-016-0070-6

Chen, Z., Liu, M., Liu, X., Huang, S., Li, L., Song, B., et al. (2013). COX-2 Regulates E-Cadherin Expression through the NF-κB/Snail Signaling Pathway in Gastric Cancer. *Int. J. Mol. Med.* 32 (1), 93–100. doi:10.3892/ijmm.2013.1376

Choi, S., Jang, J., Heo, Y. J., Kang, S. Y., Kim, S. T., Lee, J., et al. (2020). CDH1 Mutations in Gastric Cancers Are Not Associated with Family History. *Pathol. - Res. Pract.* 216 (5), 152941. doi:10.1016/j.prp.2020.152941

Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., et al. (2001). The Two-Handed E Box Binding Zinc finger Protein SIP1 Downregulates E-Cadherin and Induces Invasion. *Mol. Cell* 7 (6), 1267–1278. doi:10.1016/s1097-2765(01)00260-x

Conacci-Sorrell, M., Simcha, I., Ben-Yedidia, T., Blechman, J., Savagner, P., and Ben-Ze'Ev, A. (2003). Autoregulation of E-Cadherin Expression by Cadherin-Cadherin Interactions. *J. Cell Biol* 163 (4), 847–857. doi:10.1083/jcb.200308162

Corso, G., Carvalho, J., Marrelli, D., Vindigni, C., Carvalho, B., Seruca, R., et al. (2013). Somatic Mutations and Deletions of the E-Cadherin Gene Predict Poor Survival of Patients with Gastric Cancer. *Jco* 31 (7), 868–875. doi:10.1200/JCO.2012.44.4612

Corso, G., Corso, F., Bellerba, F., Carneiro, P., Seixas, S., Cioffi, A., et al. (2021). Geographical Distribution of E-Cadherin Germline Mutations in the Context of Diffuse Gastric Cancer: A Systematic Review. *Cancers* 13 (6), 1269. doi:10.3390/cancers13061269

Corso, G., Pedrazzani, C., Pinheiro, H., Fernandes, E., Marrelli, D., Rinnovati, A., et al. (2011). E-cadherin Genetic Screening and Clinico-Pathologic Characteristics of Early Onset Gastric Cancer. *Eur. J. Cancer* 47 (4), 631–639. doi:10.1016/j.ejca.2010.10.011

Cui, H., Hu, Y., Guo, D., Zhang, A., Gu, Y., Zhang, S., et al. (2018). DNA Methyltransferase 3A Isoform B Contributes to Repressing E-Cadherin through Cooperation of DNA Methylation and H3K27/H3K9 Methylation in EMT-Related Metastasis of Gastric Cancer. *Oncogene* 37 (32), 4358–4371. doi:10.1038/s41388-018-0285-1

Deng, W., Jin, L., Zhuo, H., Vasiliou, V., and Zhang, Y. (2021). Alcohol Consumption and Risk of Stomach Cancer: A Meta-Analysis. *Chemico-Biological Interactions* 336, 109365. doi:10.1016/j.cbi.2021.109365

DiDonato, J. A., Mercurio, F., and Karin, M. (2012). NF-κB and the Link between Inflammation and Cancer. *Immunol. Rev.* 246 (1), 379–400. doi:10.1111/j.1600-065X.2012.01099.x

Egger, G., Liang, G., Aparicio, A., and Jones, P. A. (2004). Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature* 429 (6990), 457–463. doi:10.1038/nature02625

Faghihloo, E., Araei, Y., Mohammadi, M., Mirzaei, H., Mohammadi, H. R., and Mokhtari-Azad, T. (2016). The Effect of Oxamflatin on the E-Cadherin Expression in Gastric Cancer Cell Line. *Cancer Gene Ther.* 23 (11), 396–399. doi:10.1038/cgt.2016.52

Forman, D. (1991). *Helicobacter pylori* Infection: A Novel Risk Factor in the Etiology of Gastric Cancer. *JNCI J. Natl. Cancer Inst.* 83 (23), 1702–1703. doi:10.1093/jnci/83.23.1702

Gall, T. M. H., and Frampton, A. E. (2013). Gene of the Month: E-Cadherin (CDH1). *J. Clin. Pathol.* 66 (11), 928–932. doi:10.1136/jclinpath-2013-201768

Gamboa-Dominguez, A., Dominguez-Fonseca, C., Chavarri-Guerra, Y., Vargas, R., Reyes-Gutierrez, E., Green, D., et al. (2005). E-cadherin Expression in Sporadic Gastric Cancer from Mexico: Exon 8 and 9 Deletions Are Infrequent Events Associated with Poor Survival. *Hum. Pathol.* 36 (1), 29–35. doi:10.1016/j.humpath.2004.09.020

Gao, H., Lan, X., Li, S., and Xue, Y. (2017b). Relationships of MMP-9, E-Cadherin, and VEGF Expression with Clinicopathological Features and Response to Chemosensitivity in Gastric Cancer. *Tumour Biol.* 39 (5), 101042831769836. doi:10.1177/1010428317698368

Gao, P., Wang, S., Jing, F., Zhan, J., and Wang, Y. (2017a). microRNA-203 Suppresses Invasion of Gastric Cancer Cells by Targeting ERK1/2/Slug/E-Cadherin Signaling. *Cbm* 19 (1), 11–20. doi:10.3233/CBM-160167

Ge, X., Liu, X., Lin, F., Li, P., Liu, K., Geng, R., et al. (2016). MicroRNA-421 Regulated by HIF-1α Promotes Metastasis, Inhibits Apoptosis, and Induces Cisplatin Resistance by Targeting E-Cadherin and Caspase-3 in Gastric Cancer. *Oncotarget* 7 (17), 24466–24482. doi:10.18632/oncotarget.8228

Gong, J., Wang, Y., and Shu, C. (2020). LncRNA CHRF Promotes Cell Invasion and Migration via EMT in Gastric Cancer. *Eur. Rev. Med. Pharmacol. Sci.* 24 (3), 1168–1176. doi:10.26355/eurrev_202002_20168

Grady, W. M., Willis, J., Guilford, P. J., Dunbier, A. K., Toro, T. T., Lynch, H., et al. (2000). Methylation of the CDH1 Promoter as the Second Genetic Hit in Hereditary Diffuse Gastric Cancer. *Nat. Genet.* 26 (1), 16–17. doi:10.1038/79120

Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., et al. (1998). E-cadherin Germline Mutations in Familial Gastric Cancer. *Nature* 392 (6674), 402–405. doi:10.1038/32918

Holleczek, B., Schöttker, B., and Brenner, H. (2020). *Helicobacter pylori* Infection, Chronic Atrophic Gastritis and Risk of Stomach and Esophagus Cancer: Results from the Prospective Population-based ESTHER Cohort Study. *Int. J. Cancer* 146 (10), 2773–2783. doi:10.1002/ijc.32610

Hu, P. J., Yu, J., Zeng, Z. R., Leung, W. K., Lin, H. L., Tang, B. D., et al. (2004). Chemoprevention of Gastric Cancer by Celecoxib in Rats. *Gut* 53 (2), 195–200. doi:10.1136/gut.2003.021477

Hu, Z., Liu, X., Tang, Z., Zhou, Y., and Qiao, L. (2013). Possible Regulatory Role of Snail in NF-Kb-Mediated Changes in E-Cadherin in Gastric Cancer. *Oncol. Rep.* 29 (3), 993–1000. doi:10.3892/or.2012.2200

Jang, M., Koh, I., Lee, J. E., Lim, J. Y., Cheong, J.-H., and Kim, P. (2018). Increased Extracellular Matrix Density Disrupts E-Cadherin/β-Catenin Complex in Gastric Cancer Cells. *Biomater. Sci.* 6 (10), 2704–2713. doi:10.1039/c8bm00843d

Juhasz, M., Ebert, M. P., Schulz, H. U., Röcken, C., Molnar, B., Tulassay, Z., et al. (2003). Dual Role of Serum Soluble E-Cadherin as a Biological Marker of Metastatic Development in Gastric Cancer. *Scand. J. Gastroenterol.* 38 (8), 850–855. doi:10.1080/00365520310003985

Jiang, X., Zhu, X., Huang, W., Xu, H., Zhao, Z., Li, S., et al. (2017). Garlic-derived Organosulfur Compound Exerts Antitumor Efficacy via Activation of MAPK Pathway and Modulation of Cytokines in SGC-7901 Tumor-Bearing Mice. *Int. Immunopharmacology* 48, 135–145. doi:10.1016/j.intimp.2017.05.004

Kalluri, R., and Weinberg, R. A. (2009). The Basics of Epithelial-Mesenchymal Transition. *J. Clin. Invest.* 119 (6), 1420–1428. doi:10.1172/JCI39104

Kim, S.-J., Lee, H.-W., Baek, J.-H., Cho, Y.-H., Kang, H. G., Jeong, J. S., et al. (2016). Activation of Nuclear PTEN by Inhibition of Notch Signaling Induces G2/M Cell Cycle Arrest in Gastric Cancer. *Oncogene* 35 (2), 251–260. doi:10.1038/onc.2015.80

Kobayashi, M., Tsubono, Y., Sasazuki, S., Sasaki, S., and Tsugane, S. (2002). Vegetables, Fruit and Risk of Gastric Cancer in Japan: a 10-year Follow-Up of the JPHC Study Cohort I. *Int. J. Cancer* 102 (1), 39–44. doi:10.1002/ijc.10659

Kourtidis, A., Ngok, S. P., and Anastasiadis, P. Z. (2013). p120 Catenin. *Prog. Mol. Biol. Transl Sci.* 116, 409–432. doi:10.1016/B978-0-12-394311-8.00018-2

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA Sequences to Function. *Nucleic Acids Res.* 47 (D1), D155–D162. doi:10.1093/nar/gky1141

Kumar, P., Sebastian, A., Verma, K., Dixit, R., Kumari, S., Singh, J., et al. (2021). mRNA Expression Analysis of E-Cadherin, VEGF, and MMPs in Gastric Cancer: a Pilot Study. *Indian J. Surg. Oncol.* 12 (Suppl. 1), 85–92. doi:10.1007/s13193-020-01096-5

Kurata, T., Fushida, S., Kinoshita, J., Oyama, K., Yamaguchi, T., Okazaki, M., et al. (2018). Low-dose Eribulin Mesylate Exerts Antitumor Effects in Gastric Cancer by Inhibiting Fibrosis via the Suppression of Epithelial–Mesenchymal Transition and Acts Synergistically with 5-fluorouracil. *Cmar* 10, 2729–2742. doi:10.2147/CMAR.S167846

Lan, H., Tang, Z., Jin, H., and Sun, Y. (2016). Neddylation Inhibitor MLN4924 Suppresses Growth and Migration of Human Gastric Cancer Cells. *Sci. Rep.* 6, 24218. doi:10.1038/srep24218

Lazar, D., Taban, S., Ardeleanu, C., Dema, A., Sporea, I., Cornianu, M., et al. (2008). The Immunohistochemical Expression of E-Cadherin in Gastric Cancer; Correlations with Clinicopathological Factors and Patients' Survival. *Rom. J. Morphol. Embryol.* 49 (4), 459–467.

Lee, H.-J., Zhuang, G., Cao, Y., Du, P., Kim, H.-J., and Settleman, J. (2014). Drug Resistance via Feedback Activation of Stat3 in Oncogene-Addicted Cancer Cells. *Cancer Cell* 26 (2), 207–221. doi:10.1016/j.ccr.2014.05.019

Li, C., Song, L., Zhang, Z., Bai, X.-X., Cui, M.-F., and Ma, L.-J. (2016). MicroRNA-21 Promotes TGF-B1-Induced Epithelial-Mesenchymal Transition in Gastric Cancer through Up-Regulating PTEN Expression. *Oncotarget* 7 (41), 66989–67003. doi:10.18632/oncotarget.11888

Li, D., Xu, C.-Y., Cui, R.-J., Tang, J.-B., Sun, H., Yang, Z.-K., et al. (2015). DNA Methylation Inhibitor, Decitabine, Promotes MGC803 Gastric Cancer Cell Migration and Invasion via the Upregulation of NEDD4-1. *Mol. Med. Rep.* 12 (6), 8201–8208. doi:10.3892/mmr.2015.4424

Li, N., Zhang, S., Luo, Q., Yuan, F., Feng, R., Chen, X., et al. (2019). The Effect of Dihydroartemisinin on the Malignancy and Epithelial-Mesenchymal Transition of Gastric Cancer Cells. *Cpb* 20 (9), 719–726. doi:10.2174/1389201020666190611124644

Liu, W. H., Yuan, J. B., Zhang, F., and Chang, J. X. (2019). Curcumin Inhibits Proliferation,migration and Invasion of Gastric Cancer Cells via Wnt3a/β-Catenin/EMT Signaling Pathway. *Zhongguo Zhong Yao Za Zhi* 44 (14), 3107–3115. doi:10.19540/j.cnki.cjcmm.20190304.002

Liu, X.-J., Chen, Z. F., Li, H. L., Hu, Z. N., Liu, M., Tian, A. P., et al. (2013). Interaction between Cyclooxygenase-2, Snail, and E-Cadherin in Gastric Cancer Cells. *Wjg* 19 (37), 6265–6271. doi:10.3748/wjg.v19.i37.6265

Liu, Y.-w., Sun, M., Xia, R., Zhang, E.-b., Liu, X.-h., Zhang, Z.-h., et al. (2015). LincHOTAIR Epigenetically Silences miR34a by Binding to PRC2 to Promote the Epithelial-To-Mesenchymal Transition in Human Gastric Cancer. *Cell Death Dis* 6, e1802. doi:10.1038/cddis.2015.150

Liu, Z.-Q., He, W.-F., Wu, Y.-J., Zhao, S.-L., Wang, L., Ouyang, Y.-Y., et al. (2020). LncRNA SNHG1 Promotes EMT Process in Gastric Cancer Cells through Regulation of the miR-15b/DCLK1/Notch1 axis. *BMC Gastroenterol.* 20 (1), 156. doi:10.1186/s12876-020-01272-5

Ma, R.-R., Zhang, H., Chen, H.-F., Zhang, G.-H., Tian, Y.-R., and Gao, P. (2021). MiR-19a/miR-96-mediated Low Expression of KIF26A Suppresses Metastasis by Regulating FAK Pathway in Gastric Cancer. *Oncogene* 40 (14), 2524–2538. doi:10.1038/s41388-020-01610-7

Machado, J. C., Oliveira, C., Carvalho, R., Soares, P., Berx, G., Caldas, C., et al. (2001). E-cadherin Gene (CDH1) Promoter Methylation as the Second Hit in Sporadic Diffuse Gastric Carcinoma. *Oncogene* 20 (12), 1525–1528. doi:10.1038/sj.onc.1204234

Maeda, S., and Omata, M. (2008). Inflammation and Cancer: Role of Nuclear Factor-kappaB Activation. *Cancer Sci.* 99 (5), 836–842. doi:10.1111/j.1349-7006.2008.00763.x

Maubach, G., Sokolova, O., Wolfien, M., Rothkötter, H.-J., and Naumann, M. (2013). Ca2+/calmodulin-dependent Kinase II Contributes to Inhibitor of Nuclear Factor-Kappa B Kinase Complex Activation inHelicobacter Pyloriinfection. *Int. J. Cancer* 133 (6), 1507–1512. doi:10.1002/ijc.28148

Mei, J.-W., Yang, Z.-Y., Xiang, H.-G., Bao, R., Ye, Y.-Y., Ren, T., et al. (2019). MicroRNA-1275 Inhibits Cell Migration and Invasion in Gastric Cancer by Regulating Vimentin and E-Cadherin via JAZF1. *BMC Cancer* 19 (1), 740. doi:10.1186/s12885-019-5929-1

Mendonsa, A. M., Na, T.-Y., and Gumbiner, B. M. (2018). E-cadherin in Contact Inhibition and Cancer. *Oncogene* 37 (35), 4769–4780. doi:10.1038/s41388-018-0304-2

Muzashvili, T., Tutisani, A., Chabradze, G., Beridze, N., and Museridze, N. (2020). The Study of the Expression of Cdh1, Ki67, P53 and Her2 in Diffuse Gastric Carcinoma. *Georgian Med. News* 299, 147–150.

Na, T.-Y., Schecterson, L., Mendonsa, A. M., and Gumbiner, B. M. (2020). The Functional Activity of E-Cadherin Controls Tumor Cell Metastasis at Multiple Steps. *Proc. Natl. Acad. Sci. USA* 117 (11), 5931–5937. doi:10.1073/pnas.1918167117

Nakayama, M., Hisatsune, J., Yamasaki, E., Isomoto, H., Kurazono, H., Hatakeyama, M., et al. (2009). *Helicobacter pylori* VacA-Induced Inhibition of GSK3 through the PI3K/Akt Signaling Pathway. *J. Biol. Chem.* 284 (3), 1612–1619. doi:10.1074/jbc.M806981200

Nam, S. Y., Ko, Y. S., Jung, J., Yoon, J., Kim, Y. H., Choi, Y. J., et al. (2011). A Hypoxia-dependent Upregulation of Hypoxia-Inducible Factor-1 by Nuclear Factor-Kb Promotes Gastric Tumour Growth and Angiogenesis. *Br. J. Cancer* 104 (1), 166–174. doi:10.1038/sj.bjc.6606020

Neumann, M., and Naumann, M. (2007). Beyond IκBs: Alternative Regulation of NF-KB Activity. *FASEB j.* 21 (11), 2642–2654. doi:10.1096/fj.06-7615rev

Nieto, M. A., Huang, R. Y.-J., Jackson, R. A., and Thiery, J. P. (2016). EMT: 2016. *Cell* 166 (1), 21–45. doi:10.1016/j.cell.2016.06.028

Ohno, R., Yoshinaga, K., Fujita, T., Hasegawa, K., Iseki, H., Tsunozaki, H., et al. (2001). Depth of Invasion Parallels Increased Cyclooxygenase-2 Levels in Patients with Gastric Carcinoma. *Cancer* 91 (10), 1876–1881. doi:10.1002/1097-0142(20010515)91:10<1876::aid-cncr1209>3.0.co;2-h

Okubo, K., Uenosono, Y., Arigami, T., Yanagita, S., Matsushita, D., Kijima, T., et al. (2017). Clinical Significance of Altering Epithelial-Mesenchymal Transition in Metastatic Lymph Nodes of Gastric Cancer. *Gastric Cancer* 20 (5), 802–810. doi:10.1007/s10120-017-0705-x

Oliveira, C., de Bruin, J., Nabais, S., Ligtenberg, M., Moutinho, C., Nagengast, F. M., et al. (2004). Intragenic Deletion of CDH1 as the Inactivating Mechanism of the Wild-type Allele in an HDGC Tumour. *Oncogene* 23 (12), 2236–2240. doi:10.1038/sj.onc.1207335

Park, J. W., Park, D. M., Choi, B. K., Kwon, B. S., Seong, J. K., Green, J. E., et al. (2015). Establishment and Characterization of Metastatic Gastric Cancer Cell Lines from Murine Gastric Adenocarcinoma Lacking Smad4, P53, and E-Cadherin. *Mol. Carcinog.* 54 (11), 1521–1527. doi:10.1002/mc.22226

Qian, X., Huang, C., Cho, C. H., Hui, W. M., Rashid, A., and Chan, A. O. O. (2008). E-cadherin Promoter Hypermethylation Induced by Interleukin-1β Treatment or *H. pylori* Infection in Human Gastric Cancer Cell Lines. *Cancer Lett.* 263 (1), 107–113. doi:10.1016/j.canlet.2007.12.023

Qu, Y., Dang, S., and Hou, P. (2013). Gene Methylation in Gastric Cancer. *Clinica Chim. Acta* 424, 53–65. doi:10.1016/j.cca.2013.05.002

Schildberg, C. W., Abba, M., Merkel, S., Agaimy, A., Dimmler, A., Schlabrakowski, A., et al. (2014). Gastric Cancer Patients Less Than 50 Years of Age Exhibit Significant Downregulation of E-Cadherin and CDX2 Compared to Older Reference Populations. *Adv. Med. Sci.* 59 (1), 142–146. doi:10.1016/j.advms.2014.03.002

Selga, E., Morales, C., Noé, V., Peinado, M. A., and Ciudad, C. J. (2008). Role of Caveolin 1, E-Cadherin, Enolase 2 and PKCalpha on Resistance to Methotrexate in Human HT29 colon Cancer Cells. *BMC Med. Genomics* 1, 35. doi:10.1186/1755-8794-1-35

Sharma, S., Kelly, T. K., and Jones, P. A. (2010). Epigenetics in Cancer. *Carcinogenesis* 31 (1), 27–36. doi:10.1093/carcin/bgp220

Shenoy, S. (2019). CDH1 (E-Cadherin) Mutation and Gastric Cancer: Genetics, Molecular Mechanisms and Guidelines for Management. *Cmar* 11, 10477–10486. doi:10.2147/CMAR.S208818

Shikata, K., Doi, Y., Yonemoto, K., Arima, H., Ninomiya, T., Kubo, M., et al. (2008). Population-based Prospective Study of the Combined Influence of Cigarette Smoking and *Helicobacter pylori* Infection on Gastric Cancer Incidence: the Hisayama Study. *Am. J. Epidemiol.* 168 (12), 1409–1415. doi:10.1093/aje/kwn276

Shimada, S., Akiyama, Y., Mogushi, K., Ishigami-Yuasa, M., Kagechika, H., Nagasaki, H., et al. (2018). Identification of Selective Inhibitors for Diffuse-type Gastric Cancer Cells by Screening of Annotated Compounds in Preclinical Models. *Br. J. Cancer* 118 (7), 972–984. doi:10.1038/s41416-018-0008-y

Shimada, S., Mimata, A., Sekine, M., Mogushi, K., Akiyama, Y., Fukamachi, H., et al. (2012). Synergistic Tumour Suppressor Activity of E-Cadherin and P53 in a Conditional Mouse Model for Metastatic Diffuse-type Gastric Cancer. *Gut* 61 (3), 344–353. doi:10.1136/gutjnl-2011-300050

Simões-Correia, J., Figueiredo, J., Lopes, R., Stricher, F., Oliveira, C., Serrano, L., et al. (2012). E-cadherin Destabilization Accounts for the Pathogenicity of Missense Mutations in Hereditary Diffuse Gastric Cancer. *PLoS One* 7 (3), e33783. doi:10.1371/journal.pone.0033783

Singh, A., and Settleman, J. (2010). EMT, Cancer Stem Cells and Drug Resistance: an Emerging axis of Evil in the War on Cancer. *Oncogene* 29 (34), 4741–4751. doi:10.1038/onc.2010.215

Song, S. Z., Lin, S., Liu, J. N., Zhang, M. B., Du, Y. T., Zhang, D. D., et al. (2019a). Targeting of SPP1 by microRNA-340 Inhibits Gastric Cancer Cell Epithelial-Mesenchymal Transition through Inhibition of the PI3K/AKT Signaling Pathway. *J. Cell Physiol* 234 (10), 18587–18601. doi:10.1002/jcp.28497

Song, Y., Wang, R., Li, L.-W., Liu, X., Wang, Y.-F., Wang, Q.-X., et al. (2019b). Long Non-coding RNA HOTAIR Mediates the Switching of Histone H3 Lysine 27 Acetylation to Methylation to Promote Epithelial-To-Mesenchymal Transition in Gastric Cancer. *Int. J. Oncol.* 54 (1), 77–86. doi:10.3892/ijo.2018.4625

Sun, J., Xu, Z., Lv, H., Wang, Y., Wang, L., Ni, Y., et al. (2018). eIF5A2 Regulates the Resistance of Gastric Cancer Cells to Cisplatin via Induction of EMT. *Am. J. Transl Res.* 10 (12), 4269–4279.

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

Suriano, G., Oliveira, M. J., Huntsman, D., Mateus, A. R., Ferreira, P., Casares, F., et al. (2003). E-cadherin Germline Missense Mutations and Cell Phenotype: Evidence for the independence of Cell Invasion on the Motile Capabilities of the Cells. *Hum. Mol. Genet.* 12 (22), 3007–3016. doi:10.1093/hmg/ddg316

Takezaki, T., Gao, C.-M., Ding, J.-H., Liu, T.-K., Li, M.-S., and Tajima, K. (1999). Comparative Study of Lifestyles of Residents in High and Low Risk Areas for Gastric Cancer in Jiangsu Province, China; with Special Reference to Allium Vegetables. *J. Epidemiol.* 9 (5), 297–305. doi:10.2188/jea.9.297

Torabizadeh, Z., Nosrati, A., Sajadi Saravi, S. N., Yazdani Charati, J., and Janbabai, G. (2017). Evaluation of E-Cadherin Expression in Gastric Cancer and its Correlation with Clinicopathologic Parameters. *Int. J. Hematol. Oncol. Stem Cell Res* 11 (2), 158–164.

Uchikado, Y., Okumura, H., Ishigami, S., Setoyama, T., Matsumoto, M., Owaki, T., et al. (2011). Increased Slug and Decreased E-Cadherin Expression Is Related to Poor Prognosis in Patients with Gastric Cancer. *Gastric Cancer* 14 (1), 41–49. doi:10.1007/s10120-011-0004-x

Ushijima, T., and Sasako, M. (2004). Focus on Gastric Cancer. *Cancer Cell* 5 (2), 121–125. doi:10.1016/s1535-6108(04)00033-9

Valaee, S., Yaghoobi, M. M., and Shamsara, M. (2017). Metformin Inhibits Gastric Cancer Cells Metastatic Traits through Suppression of Epithelial-Mesenchymal Transition in a Glucose-independent Manner. *PLoS One* 12 (3), e0174486. doi:10.1371/journal.pone.0174486

Valastyan, S., and Weinberg, R. A. (2011). Tumor Metastasis: Molecular Insights and Evolving Paradigms. *Cell* 147 (2), 275–292. doi:10.1016/j.cell.2011.09.024

Wang, F. J., Zong, X. Y., Du, J. L., Wang, W. S., Yuan, D. P., and Chen, X. B. (2019). Effects of Dihydromyricetin on the Migration and Invasion of Human Gastric Cancer MKN45 Cells and its Mechanism. *Zhongguo Ying Yong Sheng Li Xue Za Zhi* 35 (5), 428–432. doi:10.12047/j.cjap.5809.2019.093

Wang, Y. J., Liu, J. Z., Lv, P., Dang, Y., Gao, J. Y., and Wang, Y. (2016). Long Non-coding RNA CCAT2 Promotes Gastric Cancer Proliferation and Invasion by Regulating the E-Cadherin and LATS2. *Am. J. Cancer Res.* 6 (11), 2651–2660.

Wang, Z.-s., Shen, Y., Li, X., Zhou, C.-z., Wen, Y.-g., Jin, Y.-b., et al. (2014). Significance and Prognostic Value of Gli-1 and Snail/E-Cadherin Expression in Progressive Gastric Cancer. *Tumor Biol.* 35 (2), 1357–1363. doi:10.1007/s13277-013-1185-1

Wijnhoven, B. P. L., Dinjens, W. N. M., and Pignatelli, M. (2002). E-cadherin-catenin Cell-Cell Adhesion Complex and Human Cancer. *Br. J. Surg.* 87 (8), 992–1005. doi:10.1046/j.1365-2168.2000.01513.x

Wool Eom, B., Won, R. K., Won Ryu, K., Man Yoon, H., and Kook, M.-C. (2020). Predictive Value of E-Cadherin and EpCAM for Detection of Metastatic Lymph Node in Early Gastric Cancer. *Chin. J. Cancer Res.* 32 (5), 614–620. doi:10.21147/j.issn.1000-9604.2020.05.06

Xia, M., Xie, Y., Zan, L., Reddy, S., Tan, C., Li, J., et al. (2017). Membranous Staining of β-catenin and E-Cadherin Expression in Patients with Gastric Cancer. *Int. J. Clin. Exp. Pathol.* 10 (8), 8980–8990.

Xue, Y., Zhang, L., Zhu, Y., Ke, X., Wang, Q., and Min, H. (2019). Regulation of Proliferation and Epithelial-To-Mesenchymal Transition (EMT) of Gastric Cancer by ZEB1 via Modulating Wnt5a and Related Mechanisms. *Med. Sci. Monit.* 25, 1663–1670. doi:10.12659/MSM.912338

Yang, H., Kong, W., He, L., Zhao, J.-J., O'Donnell, J. D., Wang, J., et al. (2008). MicroRNA Expression Profiling in Human Ovarian Cancer: miR-214 Induces Cell Survival and Cisplatin Resistance by Targeting PTEN. *Cancer Res.* 68 (2), 425–433. doi:10.1158/0008-5472.CAN-07-2488

Yang, Y., Li, X., Du, J., Yin, Y., and Li, Y. (2018). Involvement of microRNAs-MMPs-E-Cadherin in the Migration and Invasion of Gastric Cancer Cells Infected with *Helicobacter pylori*. *Exp. Cell Res.* 367 (2), 196–204. doi:10.1016/j.yexcr.2018.03.036

Yang, Z.-G., Gao, L., Guo, X. B., and Shi, Y. L. (2015). Roles of Long Non-coding RNAs in Gastric Cancer Metastasis. *Wjg* 21 (17), 5220–5230. doi:10.3748/wjg.v21.i17.5220

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 (24), 5661–5667. doi:10.26355/eurrev_201712_14010

Yeh, T.-S., Wu, C.-W., Hsu, K.-W., Liao, W.-J., Yang, M.-C., Li, A. F.-Y., et al. (2009). The Activated Notch1 Signal Pathway Is Associated with Gastric Cancer Progression through Cyclooxygenase-2. *Cancer Res.* 69 (12), 5039–5048. doi:10.1158/0008-5472.CAN-08-4021

Yoshiura, K., Kanai, Y., Ochiai, A., Shimoyama, Y., Sugimura, T., and Hirohashi, S. (1995). Silencing of the E-Cadherin Invasion-Suppressor Gene by CpG Methylation in Human Carcinomas. *Proc. Natl. Acad. Sci.* 92 (16), 7416–7419. doi:10.1073/pnas.92.16.7416

Zang, M.-d., Hu, L., Fan, Z.-y., Wang, H.-x., Zhu, Z.-l., Cao, S., et al. (2017). Luteolin Suppresses Gastric Cancer Progression by Reversing Epithelial-Mesenchymal Transition via Suppression of the Notch Signaling Pathway. *J. Transl Med.* 15 (1), 52. doi:10.1186/s12967-017-1151-6

Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt Signaling in Cancer. *Oncogene* 36 (11), 1461–1473. doi:10.1038/onc.2016.304

Zhang, X.-F., Wang, Y.-M., Ge, H., Cao, Y.-Y., Chen, Z.-F., Wen, D.-G., et al. (2008). Association of CDH1 Single Nucleotide Polymorphisms with Susceptibility to Esophageal Squamous Cell Carcinomas and Gastric Cardia Carcinomas. *Dis. Esophagus* 21 (1), 21–29. doi:10.1111/j.1442-2050.2007.00724.x

Zhou, Y., Li, G., Wu, J., Zhang, Z., Wu, Z., Fan, P., et al. (2010). Clinicopathological Significance of E-Cadherin, VEGF, and MMPs in Gastric Cancer. *Tumor Biol.* 31 (6), 549–558. doi:10.1007/s13277-010-0068-y

Zhou, Y., Ran, J., Tang, C., Wu, J., Honghua, L., Xingwen, L., et al. (2007). Effect of Celecoxib on E-Cadherin, VEGF, Microvessel Density and Apoptosis in Gastric Cancer. *Cancer Biol. Ther.* 6 (2), 269–275. doi:10.4161/cbt.6.2.3629

Zhu, J., and Wen, K. (2018). Astragaloside IV Inhibits TGF-B1-Induced Epithelial-Mesenchymal Transition through Inhibition of the PI3K/Akt/NF-Kb Pathway in Gastric Cancer Cells. *Phytotherapy Res.* 32 (7), 1289–1296. doi:10.1002/ptr.6057

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 Zhao, Hu, Chen, Xu, Zhao, Huang, Xing, Lv, Nie, Wang, He, Wang 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 in Molecular Biosciences

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fmolb.2021.692120

Prognostic Prediction, Immune Microenvironment, and Drug Resistance Value of Collagen Type I Alpha 1 Chain: From Gastrointestinal Cancers to Pan-Cancer Analysis

Yi Liu[1,2,3], *Jinmin Xue*[1,2,3], *Maoxi Zhong*[1,2,3], *Zhi Wang*[1,2,3], *Jie Li*[1,2,3] *and Yuxi Zhu*[1,2,3]*

[1]*Department of Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China,* [2]*Department of Oncology, Jinshan Hospital of the First Affiliated Hospital of Chongqing Medical University, Chongqing, China,* [3]*Chongqing Clinical Cancer Research Center, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China*

OPEN ACCESS

Edited by:
Xin Wang,
City University of Hong Kong, SAR China

Reviewed by:
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

****Correspondence:***
Yuxi Zhu
zhuyuxi@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *07 April 2021*
Accepted: *28 June 2021*
Published: *30 July 2021*

Citation:
Liu Y, Xue J, Zhong M, Wang Z, Li J and Zhu Y (2021) Prognostic Prediction, Immune Microenvironment, and Drug Resistance Value of Collagen Type I Alpha 1 Chain: From Gastrointestinal Cancers to Pan-Cancer Analysis. Front. Mol. Biosci. 8:692120. doi: 10.3389/fmolb.2021.692120

Background: Gastrointestinal cancers patients might experience multiple primary tumors in the digestive tract. Therefore, identifying potential biomarkers can help us better understand the underlying mechanism. From the GEO database, four profiles of gastrointestinal cancers were gathered for the screening process, and six hub genes were found by bioinformatics analysis. *Collagen type I alpha 1 chain* (*COL1A1*), one of the hub genes, is a component of the extracellular matrix and is critical for tumor microenvironment. However, the expression level, signaling pathway, prognostic prediction, and immunological value of *COL1A1* in different cancers remain unclear.

Methods: We comprehensively analyzed gene expression and genetic alteration patterns of *COL1A1* among 33 types of malignancies from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression projects. Besides, we explored the correlation of *COL1A1* with cancer prognosis, immune infiltrates, PD-L1, tumor mutational burden (TMB)/microsatellite instability status (MSI), and the pathway and drug sensitivity of co-expressed genes.

Results: The results showed that *COL1A1* was highly expressed and associated with poor prognosis in the majority of cancers. The most common alteration type of *COL1A1* was missense mutation, and *COL1A1* was associated with poor prognosis in KIRP, LGG, MESO, SKCM, and STAD. For the immunologic significance, *COL1A1* expression was closely related to high TMB in THYM, LAML, ACC, KICH, PRAD, and LGG, and high MSI in TGCT, MESO, PRAD, COAD, SARC, and CESC. In addition, *COL1A1* was positively correlated with the abundance of CAFs, macrophages, and tumor-infiltrating lymphocytes. However, it was negatively correlated with $CD8^+$ T cells mainly in CESC, HNSC-HPV+, and SKCM. Besides, as a component of the extracellular matrix, *COL1A1* was involved in the activation of epithelial-mesenchymal transition (EMT), and high expression of *HTRA1* was resistant to multiple drugs.

Conclusion: *COL1A1* can serve as a prognostic and immunological biomarker in different cancers.

Keywords: *COL1A1*, pan-cancer, prognosis, tumor immune microenvironment, gastrointestinal cancers

INTRODUCTION

Esophageal, gastric, and colorectal cancer are the most common types of gastrointestinal cancers. According to Global Cancer Statistics 2018, colon, gastric, rectal, and esophageal cancer have been ranked the fourth, sixth, eighth, and ninth, respectively, in the incidence of human malignant tumors, and their mortality rates were 5.8, 8.2, 3.2, and 5.3%, respectively (Bray et al., 2018). However, it has been found that gastrointestinal cancers patients might experience multiple primary tumors in the digestive tract (Yoshida et al., 2020; van de Ven et al., 2020). Furthermore, it was reported that several cases were related to synchronous or metachronous primary gastrointestinal tract malignancies (Bratislav et al., 2015; Yoshikawa et al., 2016; Kim et al., 2017; Arakawa et al., 2018). Hence, we identified some hub genes including Collagen *Type I Alpha 1 Chain* (*COL1A1*) of multiple primary tumors in the gastrointestinal tract based on bioinformatics technology, which might become potential diagnostic biomarkers of gastrointestinal cancers.

COL1A1 is the gene which encodes the pro-alpha 1 chains of type I collagen whose triple helix comprises two alpha 1 chains and one alpha 2 chain (Prockop, 1990); the protein encoded by this gene is an important component of the extracellular matrix (ECM). *COL1A1* is known to be overexpressed in several cancers other than gastrointestinal cancers (Shi et al., 2015; Zhang et al., 2018; Yin et al., 2019), such as thyroid cancer (Huang et al., 2019), lung cancer (Grigoroiu et al., 2015), breast cancer (Liu et al., 2018b), and renal cancer (Boguslawska et al., 2016). However, there is no pan-cancer analysis to comprehensively elucidate the potential role of *COL1A1* in various tumor types. Thus we expand our research from gastrointestinal cancers to the pan-cancer analysis of *COL1A1*.

MATERIALS AND METHODS

Microarray Data of Gastrointestinal Cancers

We downloaded four gene expression profiles from the GEO database (http://www.ncbi.nlm.nih.gov/geo) (Barrett et al., 2013): the GSE20347 of esophageal squamous cell carcinoma (ESCC), the GSE92396 of esophageal adenocarcinoma (EAC), the GSE103236 of stomach adenocarcinoma (STAD), and the GSE110224 of colorectal adenocarcinoma (CRAC). GSE203475 (Hu et al., 2010) was based on GPL571 ([HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array) and contained 17 pairs of esophageal squamous cell carcinoma tissues and matched normal adjacent tissues. GSE92396 was based on GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]) and contained 12 esophageal adenocarcinoma tissues and 10 normal esophageal tissues. GSE1032366 (Chivu Economescu et al., 2010) was based on GPL4133 [Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F (Feature Number version)], which contained 10 pairs of cancerous and normal adjacent tissue from gastric adenocarcinoma patients. GSE110224 (Vlachavas et al., 2019) was based on GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). It included 17 pairs of histologically confirmed colorectal adenocarcinoma tissues and normal adjacent tissues.

Identification of the DEGs in Profile

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) is a tool for analyzing differentially expressed genes in the GEO database, which can compare the expression of genes in tumor and normal samples. Adj. $p < 0.05$ and $|\log FC| > 1$ were set as the cutoff criteria to select DEGs for these microarray, respectively. Then the overlapping DEGs among the four datasets were identified by the online Venn diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Gene Ontology and Pathway Enrichment Analysis of DEGs

Database for Annotation, Visualization, and Integrated Discovery (DAVIDv6.8 https://david.ncifcrf.gov/) provides a comprehensive set of functional annotation tools to understand the biological meaning behind a large number of genes (Huang da et al., 2009). DAVID was employed to carry out gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of DEGs. The GO and KEGG pathways were plotted by http://www.bioinformatics.com.cn, an online platform for data analysis and visualization.

Hub Genes Screening From the PPI Network

The protein–protein interaction (PPI) network of differentially expressed genes was constructed based on the online website STRING (STRING; http://string-db.org) (version 11.0) (Szklarczyk et al., 2019) and was further illustrated by the Cytoscape software (Shannon et al., 2003). The MCODE plug-in in Cytoscape was utilized to identify key modules and hub genes. The preferred cutoff values were determined as degree cutoff values = 2, max. depth = 100, the node score = 0.2, and the k-score = 2.

Gene Expression Analysis in Pan-Cancer

The Oncomine (https://www.oncomine.org) database is currently the world's largest cancer gene-chip database and an integrated data-mining platform, which can analyze differential gene expressions in normal and tumor tissues (Rhodes et al., 2004).

Firstly, we used Oncomine to analyze the differential expression of *COL1A1* between tumor tissues and normal tissues. Next, the "Gene_DE" module of TIMER2 (tumor immune estimation resource, version 2) (http://timer.cistrome.org/) was employed to analyze the differential expression of *COL1A1* in different tumors and normal tissues (Li et al., 2020). For those tumors that lack normal or have a highly limited number of normal tissues, the "Expression analysis-Box Plots" module of GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2) (http://gepia2.cancer-pku.cn/#analysis) (Tang et al., 2019) was used to obtain the expression difference between these tumor tissues and the corresponding normal tissues. In addition, we explored the *COL1A1* expression difference in different stages by the "Pathological Stage Plot" module of GEPIA2.

Genetic Alteration Analysis

CBioportal (https://www.cbioportal.org) is an online database which provides visualization, analysis, and the ability to download large-scale cancer genomics data (Cerami et al., 2012). Herein, cBioportal was employed to obtain the alteration frequency and mutation type of *COL1A1* across all TCGA tumors. Then we explored the overall, disease-specific, progression-free, and disease-free survival differences with or without *COL1A1* genetic alteration of the tumor with the highest alteration frequency.

Methylation Profile of *COL1A1*

UALCAN is a comprehensive and interactive web resource for analyzing cancer OMICS data (Chandrashekar et al., 2017). We investigated the *COL1A1* promoter DNA methylation level in gastrointestinal cancers and some certain types of cancer by UALCAN. MEXPRESS is a web tool which can visualize DNA methylation, expression, and clinical data (Koch et al., 2015). MEXPRESS was employed to determine the association between *COL1A1* DNA methylation and clinical data.

Survival Prognosis Analysis

We used the "Survival Map" module of GEPIA2 to analyze the overall (OS) and disease-free survival (DFS) of *COL1A1* among all tumors. Furthermore, the "Survival Analysis" module was used to further analyze the survival outcome of the specific type of tumor.

Immune Infiltration Analysis

The occurrence and development of tumors are closely related to the tumor immune microenvironment. We used the "Immune-Gene" module of TIMER2 to analyze the relationship between *COL1A1* expression and cancer-associated fibroblasts, $CD8^+$ T cells, and macrophages. TISIDB (http://cis.hku.hk/TISIDB/index.php) is an integrated repository portal for tumor and immune system interaction, which integrates multiple heterogeneous data types (Ru et al., 2019). We used TISIDB to analyze *COL1A1* expression and tumor-infiltrating lymphocytes (TILs), immunoinhibitors, immunostimulators, and the MHC molecule.

PD-L1, TMB, and MSI in Pan-Cancer

PD-L1, TMB, and MSI are important predictive markers of immunotherapy. We used GEPIA to analysis the association of *COL1A1* expression and PD-L1 (CD274) among different cancers. RNA-seq data of 33 types of tumor were downloaded from the Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) data portal website (https://portal.gdc.cancer.gov/). TMB is derived from the article "The Immune Landscape of Cancer" published by Thorsson et al. (2019); MSI is derived from the "Landscape of Microsatellite Instability Across 39 Cancer Types" article published by Bonneville et al. (2017). R software v4.0.3 was utilized for statistical analysis. If not otherwise stated, the rank-sum test detected two sets of data, and p-value <0.05 was considered statistically significant.

Genes Co-expressed With *COL1A1* in Pan-Cancer: Pathway and Drug Sensitivity Analysis

We used the STRING database (https://string-db.org/) (Szklarczyk et al., 2019) to obtain the top 50 proteins most relevant to *COL1A1* protein expression. The parameters were set as follows: a minimum required interaction score ["Low confidence (0.150)"], meaning of network edges ("evidence"), max number of interactors to show ("no more than 50 interactors" in first shell), and active interaction sources ("Experiments"). Moreover, we used the GEPIA2 database to obtain the top 100 genes most correlated to *COL1A1*. Then we drew the Venn diagram to obtain the overlapping genes by the online Venn diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Next, we used GSCALite (http://bioinfo.life.hust.edu.cn/web/GSCALite/) (Liu et al., 2018a) to conduct a pathway and drug sensitivity analysis on *COL1A1* and the obtained genes.

RESULTS

Identification of DEGs in Gastrointestinal Cancers

Based on the cutoff criteria, four mRNA expression profiles of esophageal squamous cell carcinoma, esophageal adenocarcinoma, gastric adenocarcinoma, and colorectal adenocarcinoma were gathered from the GEO database in order to screen out potential biomarkers of gastrointestinal cancers, including GSE20347, GSE92396, GSE103236, and GSE110224, respectively. After analysis using the GEO2R tool and Venn diagram, a total of 21 DEGs were identified: 19 genes were upregulated and 2 genes were downregulated. The Venn diagram is shown in **Figure 1A** and **Figure 1B**.

GO and KEGG Analysis for the DEGs in Gastrointestinal Cancers

For all DEGs, extracellular regions, cell surfaces, the proteinaceous extracellular matrix, and the collagen trimer were mostly enriched in cellular components (CC). The biological process (BP) mainly included the collagen catabolic process, extracellular matrix organization and disassembly, cell adhesion, cartilage development, and skeletal system development. As for molecular function (MF), the DEGs were related to extracellular matrix structural constituent, serine-type endopeptidase activity, metalloendopeptidase activity, protein binding, and extracellular matrix binding (**Figure 1C**). Furthermore, the KEGG pathway included focal adhesion, ECM-receptor interaction, PI3K-Akt

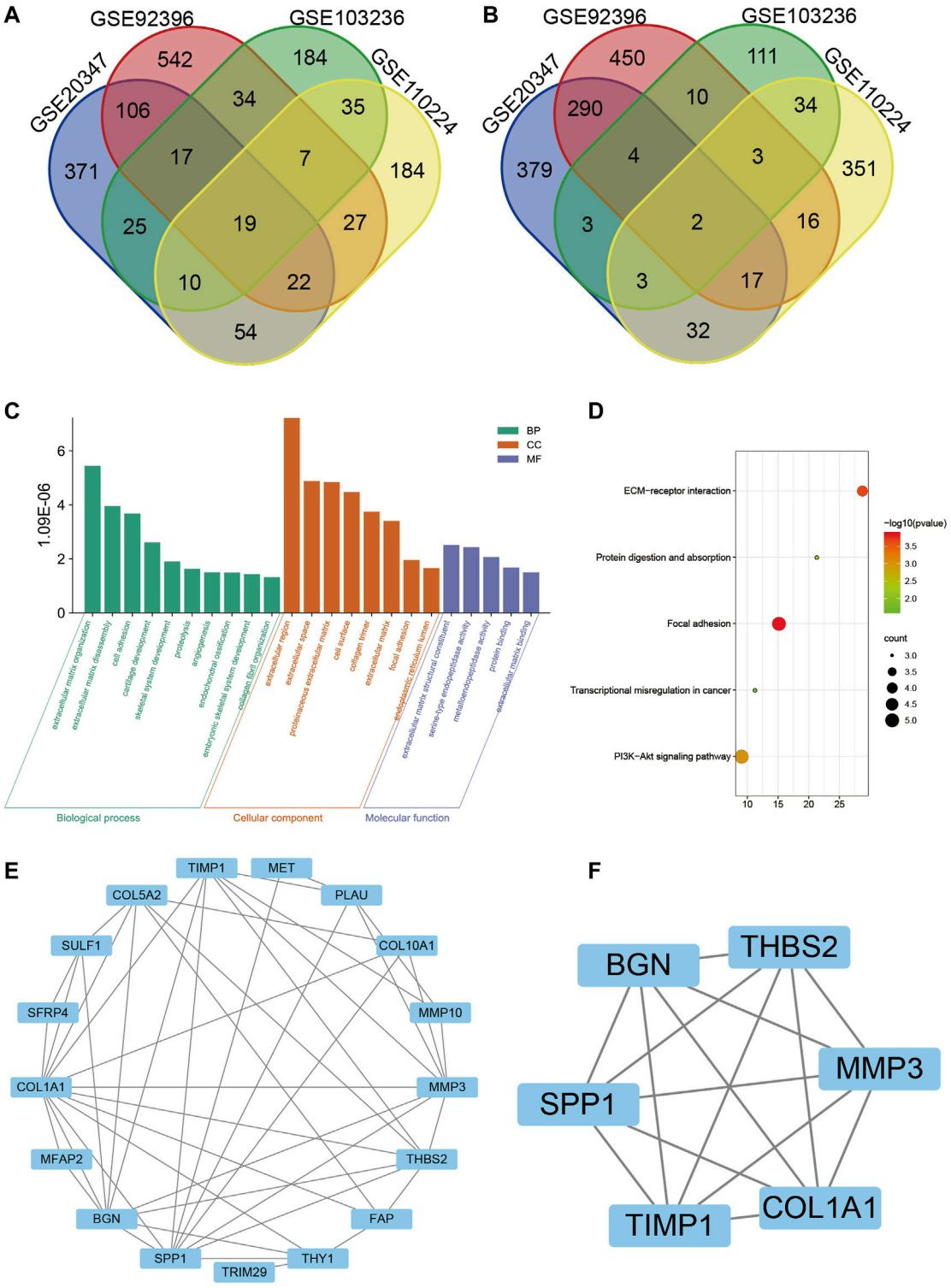


FIGURE 1 | Identification of 21 differentially expressed genes (DEGs) from four microarrays (GSE20347, GSE92396, GSE103236, and GSE110224) of gastrointestinal cancers, among which 19 were upregulated and 2 were downregulated. **(A)** Upregulated genes. **(B)** Downregulated genes. **(C)** Gene Ontology (GO) analysis of DEGs, including biological process (BP), cellular component (CC), and molecular function (MF), respectively. **(D)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs. **(E)** Protein–protein interaction (PPI) network of DEGs by Cytoscape. **(F)** Interaction network of hub genes, including *SPP1*, *BGN*, *THBS2*, *MMP3*, *COL1A1*, and *TIMP1*.

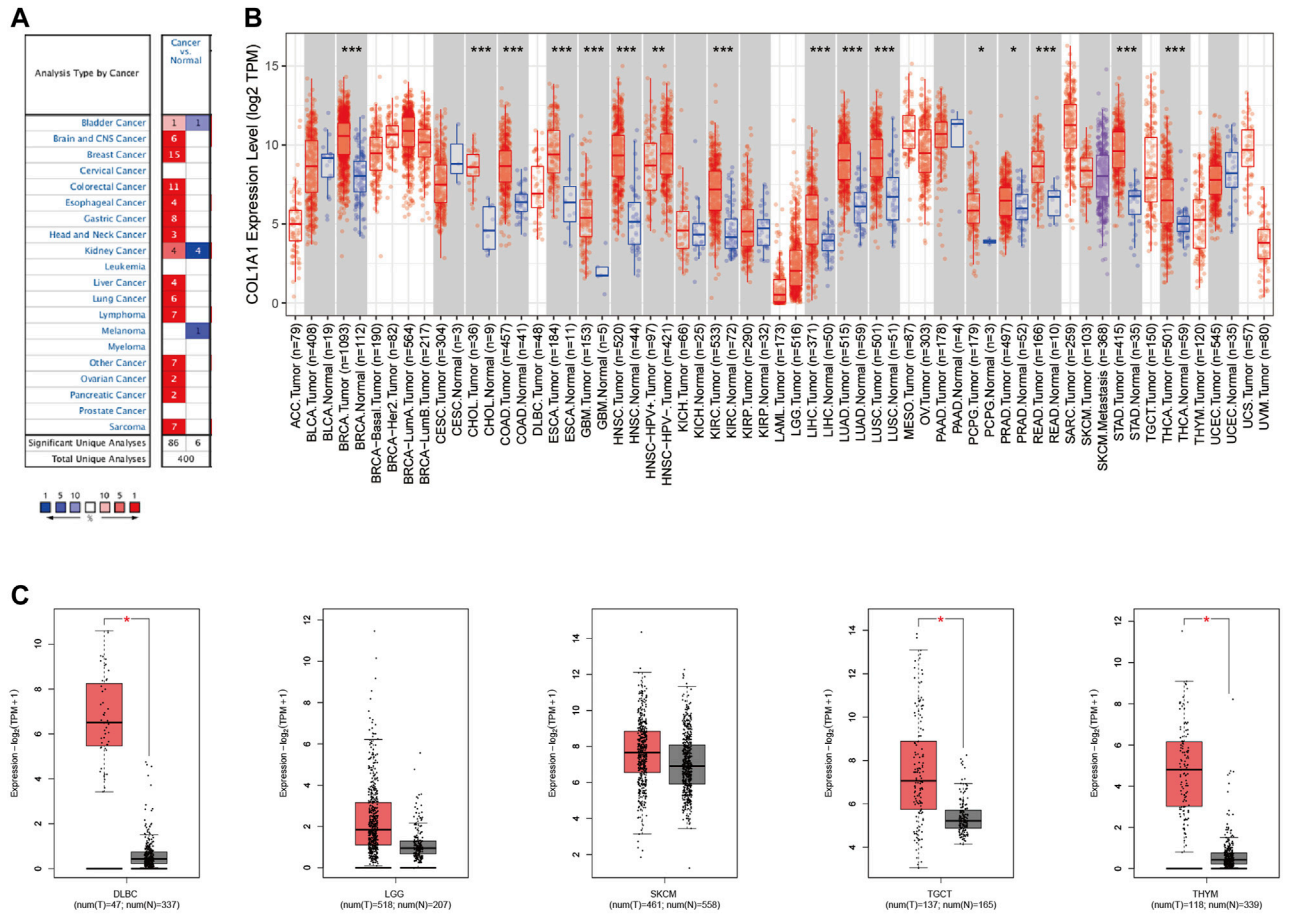


FIGURE 2 | (A) *COL1A1* expression levels in diverse cancer types and normal tissues by Oncomine. The red color represents increased expression and the blue color represents decreased expression of *COL1A1* in different cancers compared with normal tissues. **(B)** *COL1A1* expression in diverse cancers relative to the non-carcinoma tissue samples based on the TCGA database through TIMER2. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. **(C)** *COL1A1* expression in DLBC, LGG, SKCM, TGCT, and THYM by GEPIA2. $^{*}p < 0.05$.

signaling pathway, protein digestion and absorption, and transcriptional mis-regulation in cancer (**Figure 1D**).

PPI Network Construction and Modules Analysis

The PPI network of the DEGs was constructed through the STRING online website (**Figure 1E**). Thus the data were imported into Cytoscape software for visualization and the MCODE plug-in was used to further screen the hub gene. Finally, a total of six hub genes were identified, namely *SPP1, BGN, THBS2, MMP3, COL1A1*, and *TIMP1* (**Figure 1F**). All the hub genes were upregulated in tumor tissues.

The above results showed that the hub genes were mainly enriched in the extracellular matrix by GO and KEGG, and the protein of *COL1A1* is an important component of the extracellular matrix. Moreover, we found that *COL1A1* was highly expressed in other cancers in addition to gastrointestinal cancers (Grigoroiu et al., 2015; Boguslawska et al., 2016; Liu et al., 2018b; Huang et al., 2019). Therefore, we aimed to explore *COL1A1* in pan-cancer.

Gene Expression Analysis of *COL1A1* in Pan-Cancer

Through the Oncomine database, we found that *COL1A1* was highly expressed in a variety of tumors (**Figure 2A**). Moreover, through the TIMER2 database and GEPIA2 database, we found that *COL1A1* was overexpressed mainly in breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal adenocarcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), renal clear cell carcinoma (KIRC), hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid adenocarcinoma (THCA), diffuse large B lymphoma (DLBC), testicular germ cell tumors (TGCT), and thymoma (THYM) (**Figures 2B,C**).

We further analyzed the expression of *COL1A1* in different tumor stages. The results showed that *COL1A1* expression is closely related to the late stage of ACC, BCLA, ESCA, KICH, KIRP, STAD, and THCA (**Supplementary Figure 1**).

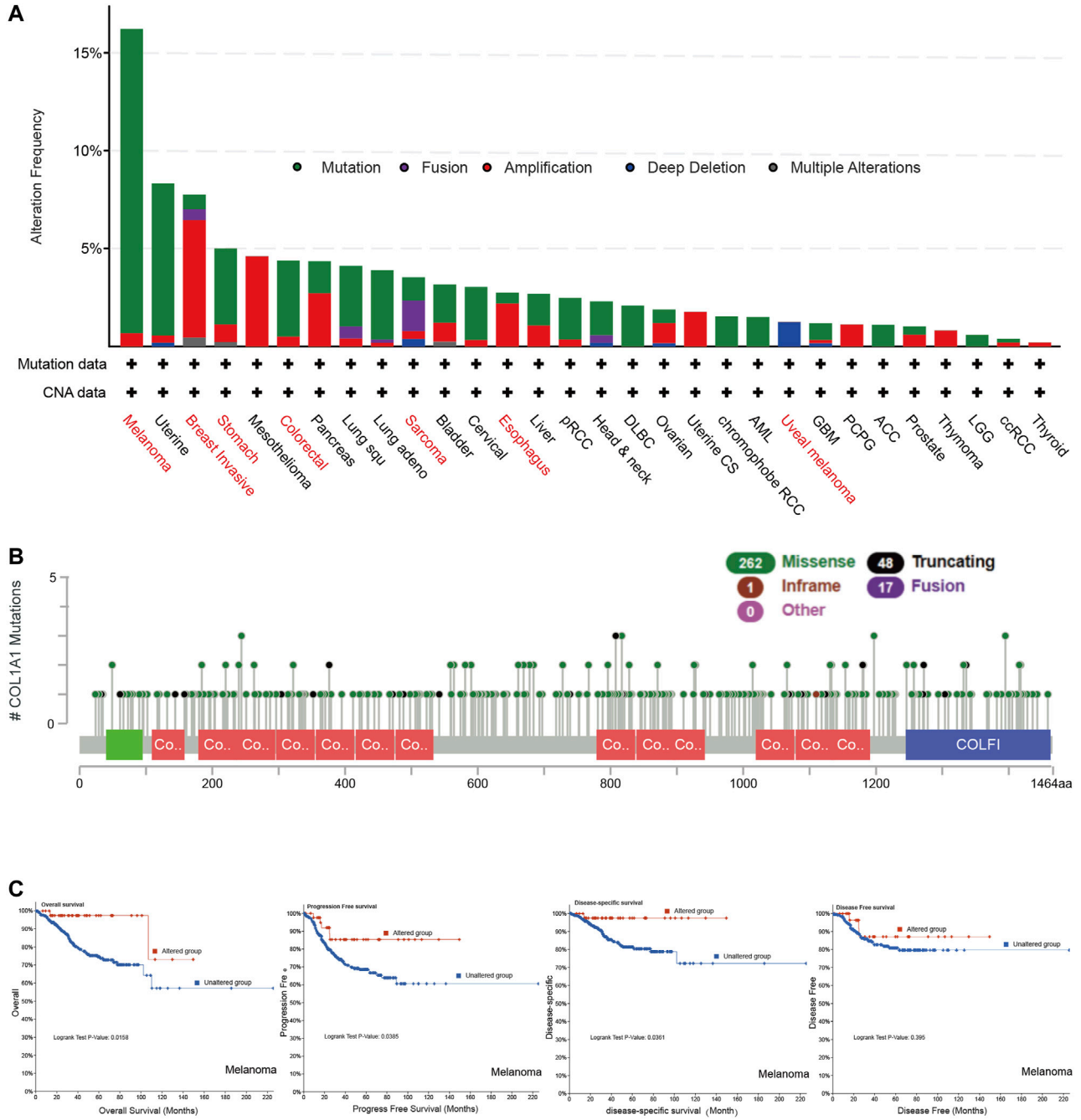


FIGURE 3 | (A) Genetic alteration type and frequency of *COL1A1* in different tumors of TCGA by cBioportal. **(B)** Mutation sites of *COL1A1*, which contain 262 missense, 48 truncating, 17 fusion, and 1 inframe mutations. **(C)** *COL1A1* alteration status and overall, progression-free, disease-specific, and disease-free survival of melanoma.

Genetic Alteration Analysis of *COL1A1*

We found *COL1A1* mutation in various tumors by cBioportal. Among them, the highest mutation frequency was in melanoma (16.22%). Breast invasive carcinoma had the highest amplification frequency (6%) and sarcoma had the highest fusion frequency of *COL1A1*. It is noteworthy that all mesothelioma cases had copy number amplification (4.6%) and all uveal melanoma had deletion of *COL1A1* (1.25%) (**Figure 3A**). The main alteration types of esophageal, gastric, and colorectal cancer were mutation and amplification. Furthermore, **Figure 3B** presented the types, sites, and case number of the *COL1A1* genetic alteration. The most common alteration type was missense mutation (**Figure 3B**). In addition, we analyzed the relationship between genetic changes and the

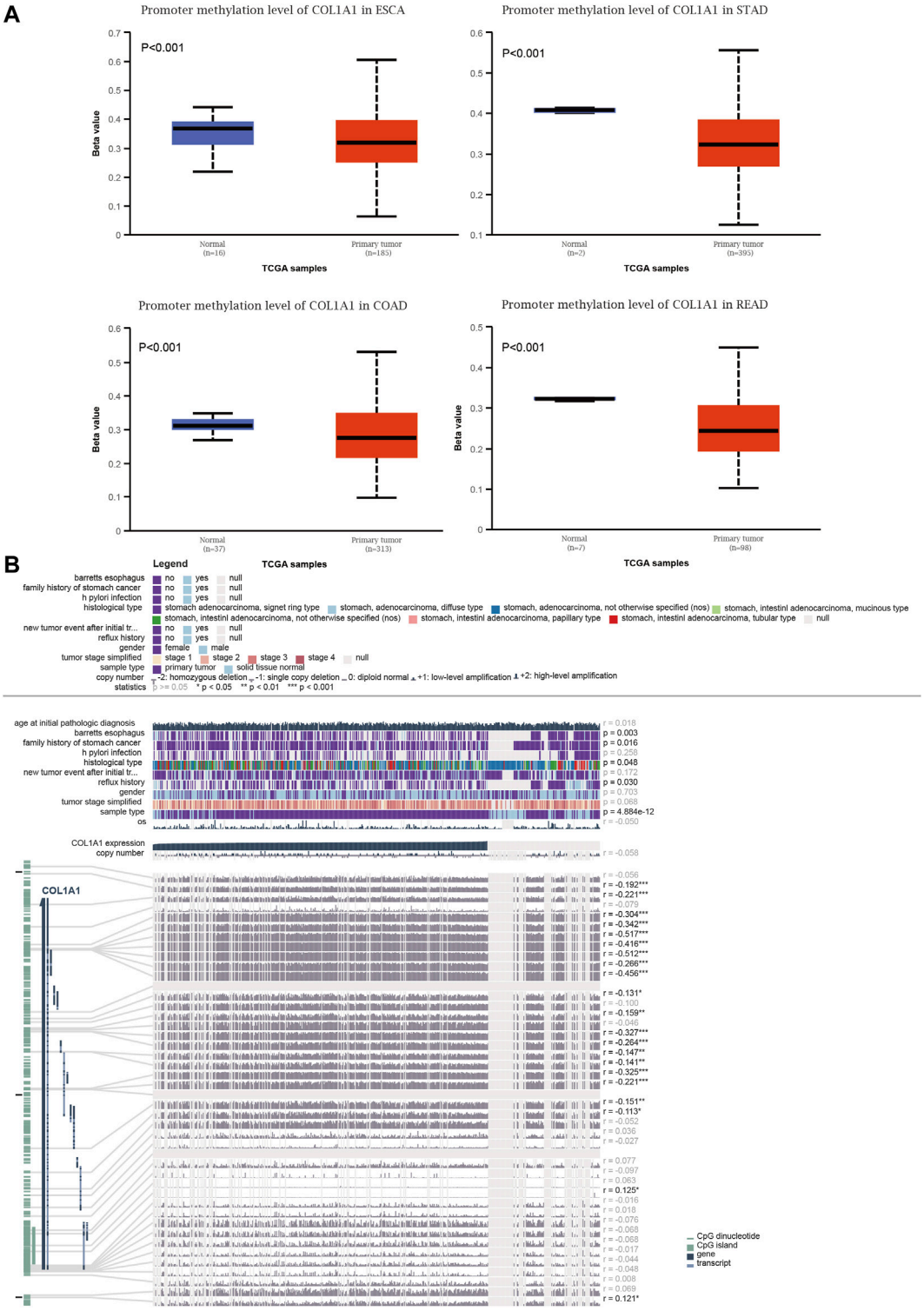


FIGURE 4 | (A) Box plots of the *COL1A1* methylation level in normal and gastrointestinal cancers tissues by UALCAN. **(B)** The detailed information of correlation between *COL1A1* expression and DNA methylation, copy number, and clinical data in STAD by MEXPRESS. Copy number of *COL1A1* was examined using Wilcoxon's rank-sum test and Pearson's correlation analysis. **p* < 0.05, ***p* < 0.01, ****p* < 0.001.

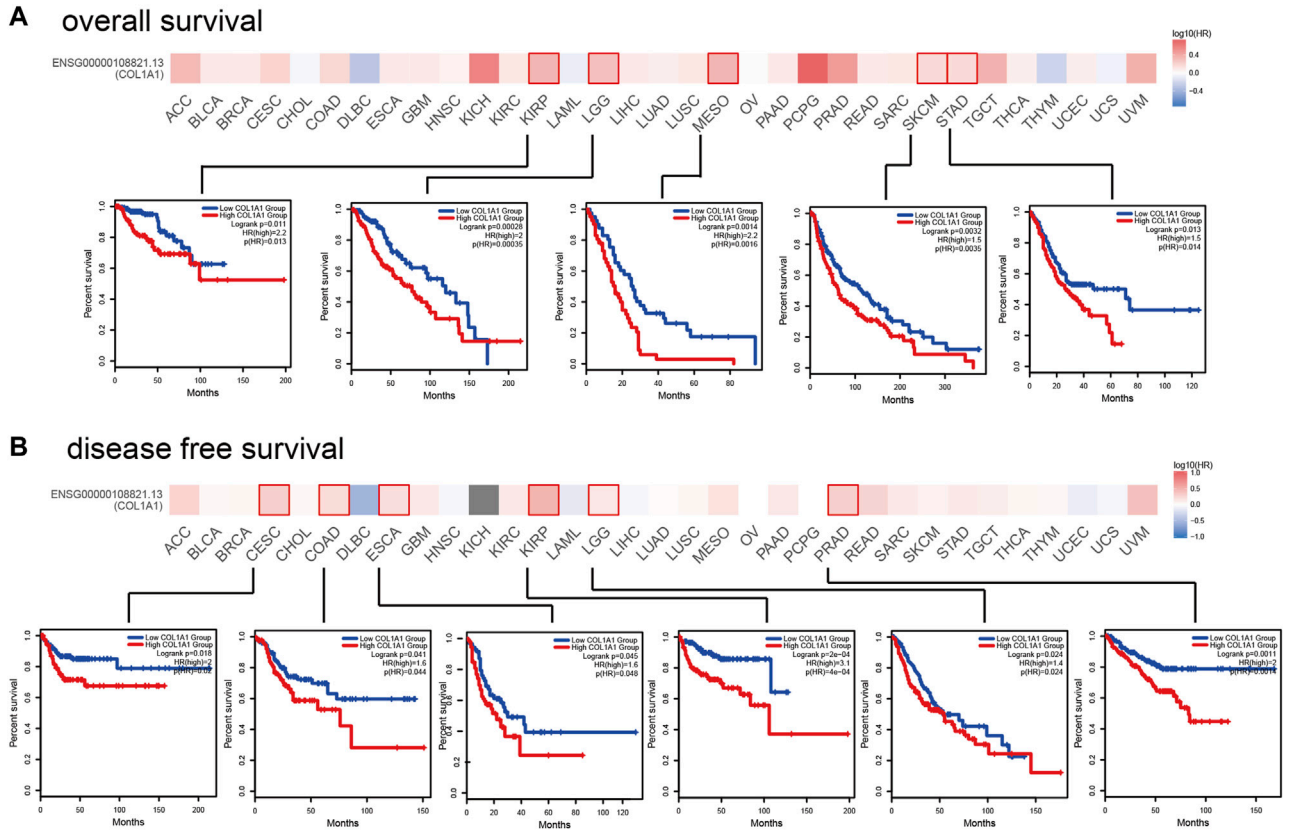


FIGURE 5 | The prognostic impact of *COL1A1* in 33 types of human tumors were examined using GEPIA2. **(A)** High expression of *COL1A1* was associated with poor overall survival in KIRP (HR = 2.2, p = 0.011), LGG (HR = 2, p = 0.00028), MESO (HR = 2.2, p = 0.0014), SKCM (HR = 1.5, p = 0.0032), and STAD (HR = 1.5, p = 0.013). **(B)** High expression of *COL1A1* was associated with poor disease-free survival in CESC (HR = 2, p = 0.018), COAD (HR = 1.6, p = 0.041), ESCA (HR = 1.6, p = 0.045), KIRP (HR = 3.1, p < 0.001), LGG (HR = 1.4, p = 0.024), and PRAD (HR = 2, p = 0.0011).

prognosis of melanoma. The results showed that the alternations of *COL1A1* are related to the better prognosis of melanoma with overall survival (p = 0.0158), progression-free survival (p = 0.0385), and disease-specific survival (p = 0.0361). Although these alternations showed a tendency of better disease-free survival, it was not statistically significant (p = 0.395) (**Figure 3C**).

Methylation Profile of *COL1A1*

Promoter methylation levels were significantly lower in primary tumors than in normal tissues in gastrointestinal cancers (**Figure 4A**). The promoter methylation levels of other tumors are shown in **Supplementary Figure 2**. Moreover, we investigated the detailed information of the correlation between the *COL1A1* expression level and DNA methylation, copy number, and clinical data in STAD (**Figure 4B**). The results showed that *COL1A1* expression was negatively correlated with DNA methylation of CpG islands (p < 0.05, r < 0). Furthermore, *COL1A1* expression was closely associated with Barrett's esophagus (p = 0.003), family history (p = 0.016), histology type (p = 0.048), and reflux history (p = 0.030). Copy number variation (CNV) showed no significance in STAD (r = −0.058, p > 0.05).

The Relationship Between Survival and Gene Expression of *COL1A1* in Pan-Cancer

In order to analyze the correlation between *COL1A1* gene expression and prognosis, we used GEPIA2 to analyze the OS and DFS of *COL1A1* among different cancers. The results indicated that high expression of *COL1A1* is related to poor overall survival of KIRP (HR = 2.2, p = 0.011), LGG (HR = 2, p = 0.00028), MESO (HR = 2.2, p = 0.0014), SKCM (HR = 1.5, p = 0.0032), and STAD (HR = 1.5, p = 0.013) (**Figure 5A**). Moreover, high expression of *COL1A1* is related to poor disease-free survival of CESC (HR = 2, p = 0.018), COAD (HR = 1.6, p = 0.041), ESCA (HR = 1.6, p = 0.045), KIRP (HR = 3.1, p < 0.001), LGG (HR = 1.4, p = 0.024), and PRAD (HR = 2, p = 0.0011) (**Figure 5B**).

Immune Infiltration Analysis of *COL1A1* in Pan-Cancer

In addition, we observed a strong positive correlation between *COL1A1* expression and cancer-associated fibroblasts in most tumors. Moreover, BLCA, ESCA, HNSC, COAD, READ, and STAD showed a statistically positive correlation of *COL1A1* expression and macrophages. However, CESC, HNSC-HPV+, and SKCM showed a negative relationship between *COL1A1* expression

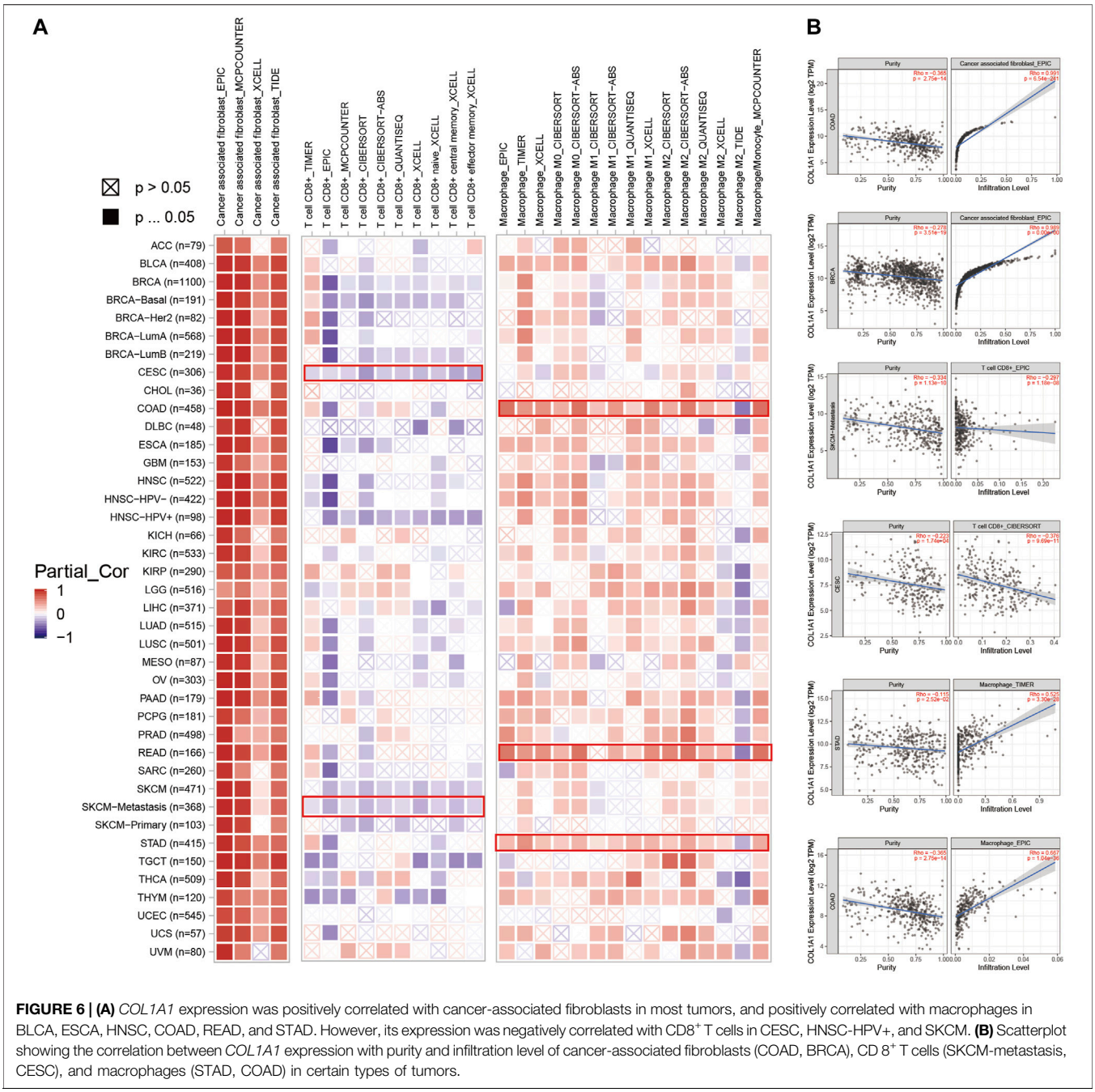


FIGURE 6 | (A) *COL1A1* expression was positively correlated with cancer-associated fibroblasts in most tumors, and positively correlated with macrophages in BLCA, ESCA, HNSC, COAD, READ, and STAD. However, its expression was negatively correlated with CD8$^+$ T cells in CESC, HNSC-HPV+, and SKCM. **(B)** Scatterplot showing the correlation between *COL1A1* expression with purity and infiltration level of cancer-associated fibroblasts (COAD, BRCA), CD 8$^+$ T cells (SKCM-metastasis, CESC), and macrophages (STAD, COAD) in certain types of tumors.

and CD8$^+$ T cells (**Figure 6A**). Furthermore, we presented the scatterplot of the correlation between *COL1A1* expression with purity and the infiltration level of cancer associated fibroblast, CD 8$^+$ T cells, and macrophages in certain types of tumors (**Figure 6B**), and the scatterplot showed consistent results with the heatmap.

The results of TISIDB showed that *COL1A1* expression is positively related to TILs in most tumors, especially THCA (**Supplementary Figure 3A**). Whereas the association of *COL1A1* expression with the MHC molecule (**Supplementary Figure 3B**), immunoinhibitors (**Supplementary Figure 3C**), and immunostimulators (**Supplementary Figure 3D**) is diverse. LGG and THCA showed a positive relation in the MHC molecule. THCA showed a positive relation and TGCT showed a negative relation in most kinds of immunoinhibitors, and TGCT showed a negative relation in most kinds of immunostimulators.

The Relationship of *COL1A1* Expression and PD-L1 and TMB/MSI in Pan-Cancer

We found that *COL1A1* expression was positively correlated to PD-L1 in COAD, DLBC, GBM, LAML, LGG, LIHC, LUAD, PAAD, PRAD, and THCA (**Figure 7A**). Moreover, through analysis of *COL1A1* expression and TMB/MSI, we found that

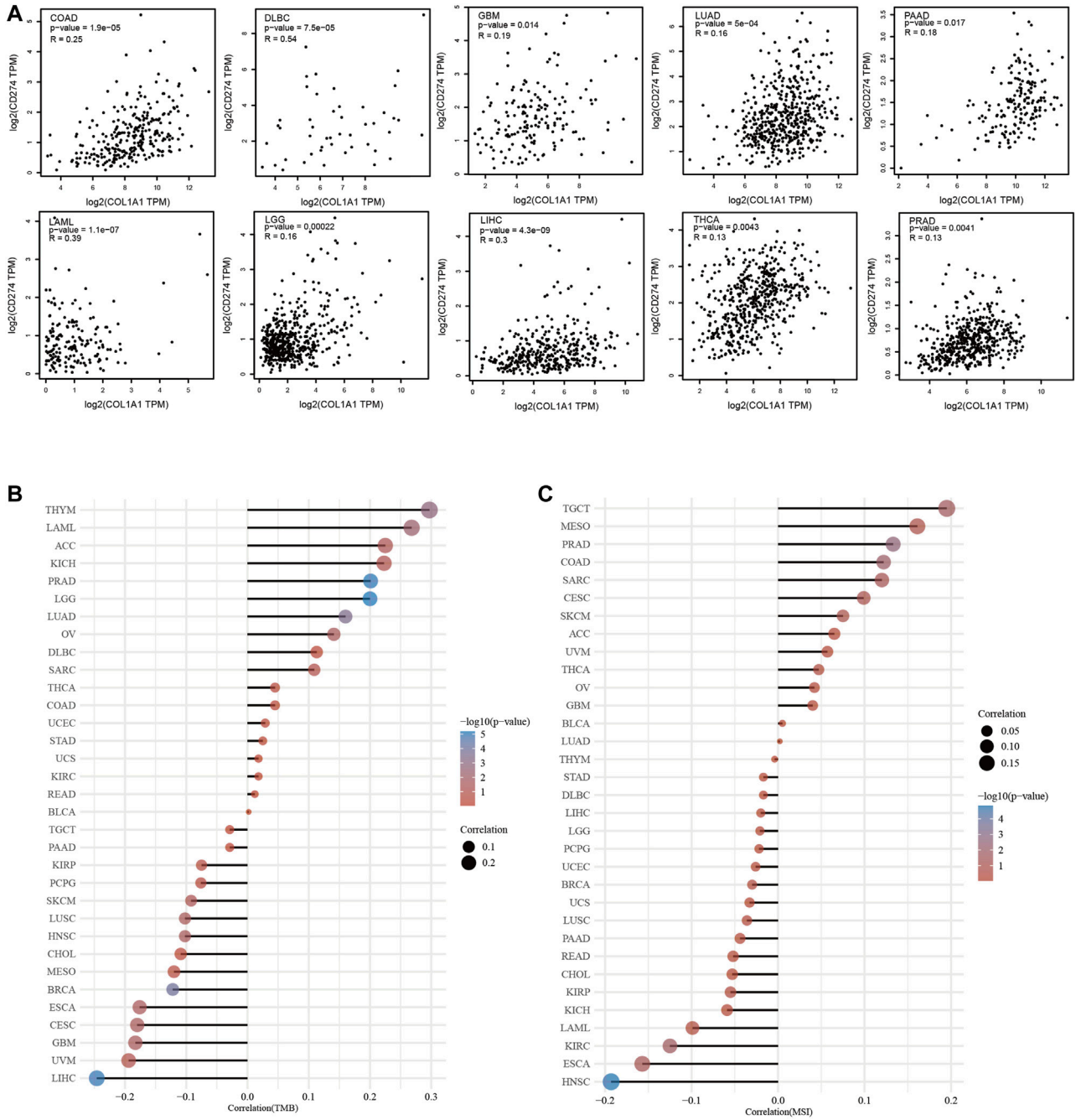


FIGURE 7 | (A) The association of *COL1A1* expression and PD-L1 in COAD, DLBC, GBM, LAML, LGG, LIHC, LUAD, PAAD, PRAD, and THCA. **(B)** Spearman correlation analysis of TMB and *COL1A1* gene expression. **(C)** Spearman correlation analysis of MSI and *COL1A1* gene expression. The horizontal axis in the figure represents the correlation coefficient between genes and TMB/MSI, the ordinate is different tumors, the size of the dots in the figure represents the size of the correlation coefficient, and the different colors represent the significance of the *p* value. The bluer the color, the smaller the *p* value.

COL1A1 is closely related to TMB and MSI in a variety of tumors and can be used as a predictor of immunotherapy. Among them, *COL1A1* expression was associated with high TMB in THYM, LAML, ACC, KICH, PRAD, and LGG (**Figure 7B**). What is more, high correlation of *COL1A1* and MSI was observed in TGCT, MESO, PRAD, COAD, SARC, and CESC (**Figure 7C**).

Genes Co-Expressed With *COL1A1* in Pan-Cancer: Pathway and Drug Sensitivity Analysis

We obtained 50 experimentally verified proteins related to *COL1A1* based on STRING (**Figure 8A**). Besides, we obtained 100 genes related to *COL1A1* expression from GEPIA2. Then the Venn

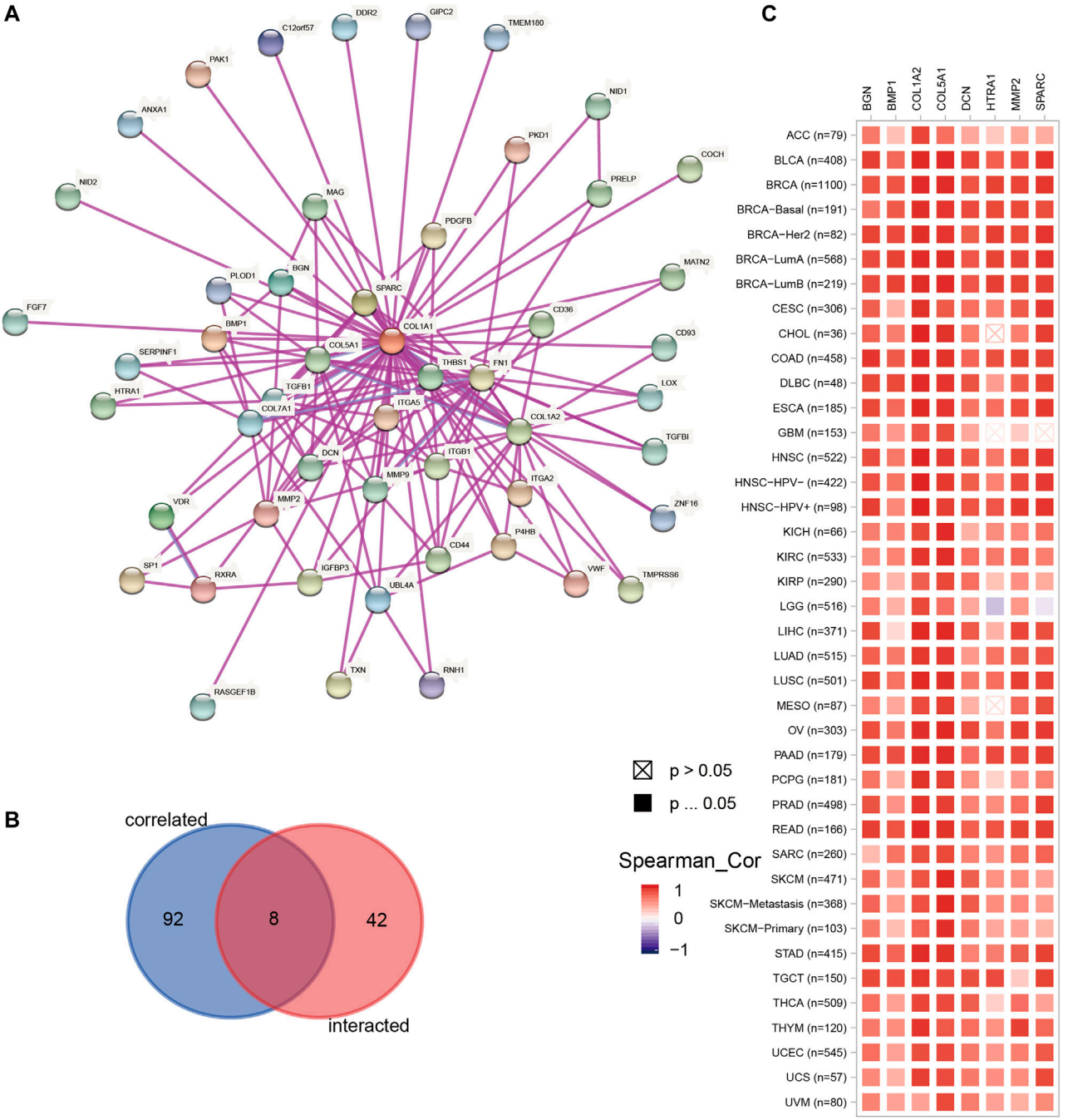


FIGURE 8 | (A) The top 50 experimentally determined *COL1A1*-binding proteins using STRING. **(B)** The Venn diagram showing the overlap of the top 50 *COL1A1*-binding proteins and top 100 *COL1A1*-correlated genes in GEPIA2. The overlapping genes were *MMP2*, *SPARC*, *COL5A1*, *DCN*, *BMP1*, *BGN*, *COL1A2*, and *HTRA1*. **(C)** The heatmap confirmed the association of *COL1A1* and *MMP2*, *SPARC*, *COL5A1*, *DCN*, *BMP1*, *BGN*, *COL1A2*, and *HTRA1* in pan-cancer.

diagram showed the overlapping genes (**Figure 8B**), of which eight genes were closely related to the expression of *COL1A1*, namely, *MMP2*, *SPARC*, *COL5A1*, *DCN*, *BMP1*, *BGN*, *COL1A2*, and *HTRA1*. Heatmaps (**Figure 8C**) confirmed that these gene expressions were closely related to *COL1A1* among tumors. Then, we performed pathway enrichment and drug sensitivity analysis of *COL1A1* and these eight genes in 33 tumors by GSCALite. The results showed that the pathway was mainly activated in epithelial-mesenchymal transition (EMT) (**Figure 9A**), and high expression of *HTRA1* was resistant to multiple drugs (**Figure 9B**).

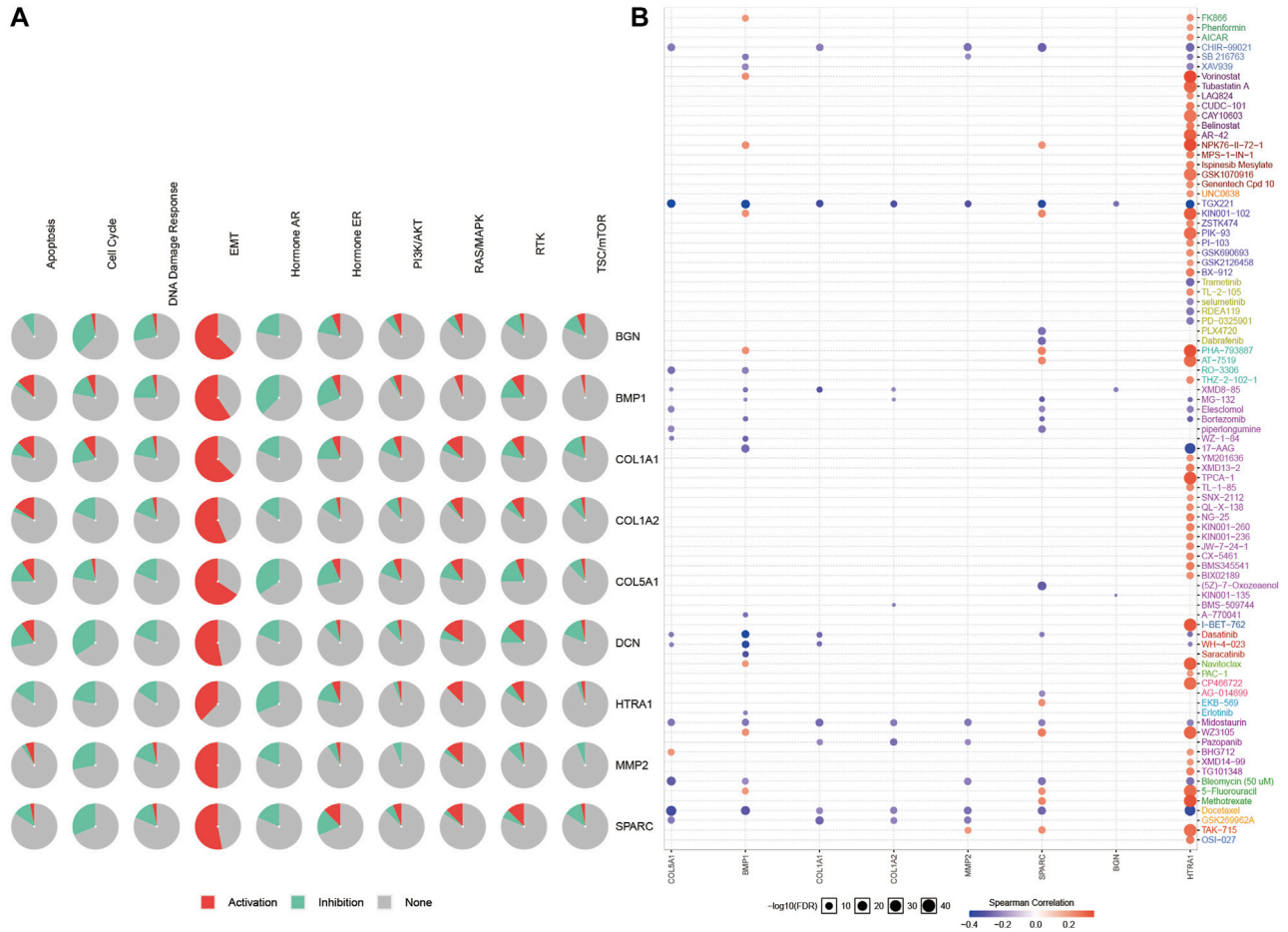


FIGURE 9 | Pathway enrichment **(A)** and drug sensitivity analysis **(B)** of *COL1A1*, *MMP2*, *SPARC*, *COL5A1*, *DCN*, *BMP1*, *BGN*, and *COL1A2* based on GSCALite. The Spearman correlation represented the gene expression that correlated with the drug. Positive correlation means that the high expression of the gene is resistant to the drug, and vice versa.

DISCUSSION

Recently, several trials reported on the multiple primary tumors of the digestive tract. However, it is difficult to distinguish them from metastatic cancer, thus the diagnosis is relatively complicated. In this study, we screened out four chips of esophageal, gastric, and colorectal cancer from the GEO database and identified 21 DEGs by using the bioinformatics analysis. Furthermore, the PPI network analysis identified six hub genes. All of these genes were upregulated in gastrointestinal cancers. We utilized the methodology of bioinformatics in our previously published studies and successfully screened out some hub genes of certain types of cancer. In this study, we continue to use these same methods and referred to other methods based on some studies.

Moreover, we analyzed one of these six genes, *COL1A1*. It has been reported that the *COL1A1* is highly expressed not only in gastrointestinal cancers but also in other cancers, involving tumorigenesis, metastasis, and immune infiltration. However, there is no pan-cancer research to comprehensively analyze *COL1A1*. Therefore, we expanded the research scope to pan-cancer and analyzed its gene expression level, mutation, DNA methylation, tumor prognosis, tumor immune microenvironment, and the association with predictive markers of immunotherapy, the pathway, and drug sensitivity in 33 kinds of tumors.

COL1A1 is the main member of the type I collagen family, and the protein is a component of the extracellular matrix. We found that *COL1A1* was highly expressed in various cancers (Ma et al., 2019; Dong et al., 2020; Zhang et al., 2021). The expression of *COL1A1* was negatively correlated with the level of the methylation of its promoter. Hypomethylation of the *COL1A1*'s promoter region might increase its expression level, thereby promoting tumorigenesis. On the other hand, EMT is a dynamic process where inactive polarized epithelial cells transit into active mesenchymal cells (Nieto et al., 2016). *COL1A1* DNA methylation is closely related to the occurrence of EMT (Casalino and Verde, 2020). Researchers have found that the knockdown of *COL1A1* can inhibit the EMT process through the TGF-β signaling pathway, thereby inhibiting the invasion and

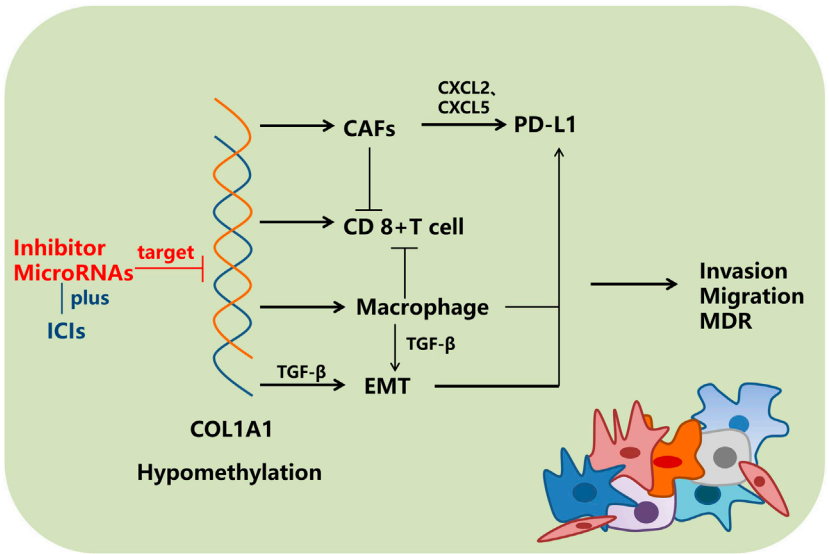


FIGURE 10 | Overview of the regulatory mechanism of *COL1A1*. MDR: multidrug resistance.

metastasis of liver cancer and bladder cancer (Ma et al., 2019; Zhu et al., 2019). Thus the methylation of the *COL1A1*'s promoter region might be closely related to the occurrence of EMT. Therefore, *COL1A1* can be viewed as not only the prognostic marker for multiple primary tumors of the digestive tract but also the prognostic marker for pan-cancer.

At present, the efficacy of immunotherapy is still not satisfied in many cancers, and patients might experience immunoresistance during the treatment. The results of immune checkpoint inhibitors (ICIs) are closely related to the tumor immune microenvironment and the expression levels of PD-L1, TMB, and MSI. An impressive finding of this study was that the expression of *COL1A1* was positively correlated with the abundance of CAFs in almost all kinds of cancers, and macrophage infiltration in gastrointestinal cancers, BLCA, and HNSC, whereas it was negatively correlated with the quantity of $CD8^+$ T cells in CESC, HNSC-HPV+, and SKCM, respectively. Moreover, its expression is also related to the levels of PD-L1, TMB, and MSI. It is reported that the relationship between immune cells and PD-L1 is complicated. On the one hand, CAFs and macrophages induced $CD8^+$ T cell deletion (Lakins et al., 2018; Dong et al., 2021). On the other hand, CAFs promote the expression of PD-L1 in tumor cells by secreting CXCL2 and CXCL5 (Inoue et al., 2019; Li et al., 2019), and macrophages upregulate PD-L1 expression through TGF-β–induced EMT (Vakili-Ghartavol et al., 2018; Jiang and Zhan, 2020), which plays a critical role in tumor immunosuppression and immune evasion. This might be helpful to understand the mechanisms of immunoresistance. Furthermore, Peng et al. (2020) found that collagen promotes ICIs resistance through LAIR1-dependent $CD8^+$ T cell exhaustion. Thus, it is worth studying whether *COL1A1*, a member of the collagen family, causes resistance of ICIs through LAIR1-dependent $CD8^+$ T cell exhaustion or other mechanisms. In addition, we hypothesized that if the expression of *COL1A1* or relative pathways were inhibited, $CD8^+$ T cells might be enriched in the tumor microenvironment, and the number of CAFs, macrophages, and the PD-L1 expression level might decrease, resulting in attenuated immunoresistance.

The FDA pooled analysis showed that the efficacy of immunotherapy combined with chemotherapy was better than that of immunotherapy alone (mPFS 7.7 vs. 4.2 months) for the first-line treatment of advanced non-small cell lung cancer with 1–49% of the PD-L1 expression level (Akinboro et al., 2021). If high expression of *COL1A1* was detected in these patients, we hypothesize that adding the *COL1A1* inhibitor to ICIs plus chemotherapy would further raise the efficacy, and relevant clinical trials could be carried out to study its efficacy and toxicity in the future. Another phase II study compared the sequence of chemotherapy and anti–PD-1 antibody in neoadjuvant therapy for locally advanced esophageal squamous cell carcinoma. The results showed that immunotherapy given 2 days after chemotherapy had a higher pathological complete response rate (36.4 vs. 7.7%) compared with simultaneous administration (Zhao et al., 2021). This result might be explained by the killing of activated T cells caused by chemotherapy. For cancers with high *COL1A1* expression, we found a low infiltration of $CD\ 8^+$ T cells by TIMER2. Future experiments could explore whether the efficacy might be better if the *COL1A1* inhibitor was given before ICIs.

The combination of radiotherapy and immunotherapy might be another research highlight. It is believed that radiation could activate the immune system and reverse cold tumors into hot tumors, and thus improve the efficacy of immunotherapy. However, the results of the FORCE study showed that nivolumab combined with radiotherapy (5×4 Gy) did not increase the objective response rate of advanced non-small cell lung cancer compared to nivolumab alone (Bozorgmehr et al., 2020). For these patients, whether the efficacy of combination therapy can be enhanced if the *COL1A1* inhibitor were administrated might be worthy of further discussion.

It was found that the efficacy of EGFR-TKI was not satisfied for NSCLC patients with a high level of PD-L1, and drug resistance may appear during the treatment (Su et al., 2018). Yamazaki et al. (2018) found that type I collagen induces EGFR-TKI resistance by activating mTOR. Under this condition, we can further study whether EGFR-TKI combined with an mTOR inhibitor or *COL1A1* inhibitor can overcome EGFR-TKI resistance.

In the era of precision medicine, it would be exciting to carry out a biomarker-guided basket trial to study the potential role of *COL1A1* inhibitors in different cancers, and it might be an optional treatment for patients with advanced tumors, just like pembrolizumab could be used in MSI-H patients.

Nevertheless, although we employed multiple bioinformatics databases to analyze the role of *COL1A1* across 33 tumors, this study still has some limitations. Firstly, the current study merely focused on *COL1A1* at the transcriptional level, but lacked the description of its protein transcriptome and the post-translational level. Secondly, the results of our study lack external validation in other public datasets. Finally, the databases used in this study were mainly based on the gene-chip and sequencing data of tumor tissue from TCGA, so the cell-level analysis of immune cell markers could have introduced systematic bias. Future studies should focus more on single-cell sequencing technology, *in vitro* experiments, and translational research.

CONCLUSION

In conclusion, we found that *COL1A1* could be a marker for different cancers and is critical for tumor immune microenvironments (**Figure 10**). Further exploration of the function of *COL1A1* might provide additional information on precision oncology.

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

YL wrote the initial manuscript. JX, MZ, ZW, JL and YZ edited and approved the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Natural Science Foundation Project of Chongqing Science and Technology Commission (CSTC), China (Grant No. cstc2018jcyjAX0012).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.692120/full#supplementary-material

REFERENCES

Akinboro, O., Vallejo, J., J., Mishra-Kalyani, P., S., Larkins, E., A., Drezner, N., L., Tang, S., et al. (2021). Outcomes of Anti-PD-(l1) Therapy in Combination with Chemotherapy versus Immunotherapy (IO) Alone for First-Line (1L) Treatment of Advanced Non-small Cell Lung Cancer (NSCLC) with PD-L1 Score 1-49%: FDA Pooled Analysis. *J. Clin. Oncol.* 39 (Suppl. 15; abstr 9001). Abstract retrieved from Abstracts in American Society of Clinical Oncology. (Accession No. 9001).

Arakawa, K., Hata, K., Yamamoto, Y., Nishikawa, T., Tanaka, T., Kiyomatsu, T., et al. (2018). Nine Primary Malignant Neoplasms-Involving the Esophagus, Stomach, colon, Rectum, Prostate, and External Ear Canal-Without Microsatellite Instability: a Case Report. *BMC Cancer* 18, 24. doi:10.1186/s12885-017-3973-2

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-Update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Boguslawska, J., Kedzierska, H., Poplawski, P., Rybicka, B., Tanski, Z., and Piekielko-Witkowska, A. (2016). Expression of Genes Involved in Cellular Adhesion and Extracellular Matrix Remodeling Correlates with Poor Survival of Patients with Renal Cancer. *J. Urol.* 195, 1892–1902. doi:10.1016/j.juro.2015.11.050

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H.-Z., et al. (2017). Landscape of Microsatellite Instability across 39 Cancer Types. *JCO Precision Oncol.*, 1, 15 2017. doi:10.1200/po.17.00073

Bozorgmehr, F., Fischer, J. R., Bischof, M., Atmaca, A., Wetzel, S., Faehling, M., et al. (2020). LBA58 ORR in Patients Receiving Nivolumab Plus Radiotherapy in Advanced Non-small Cell Lung Cancer: First Results from the FORCE Trial. *Ann. Oncol.* 31, S1187. doi:10.1016/j.annonc.2020.08.2291

Bratislav, T., Branimir, N., Mihailo, B., Jovan, K., Veljović, M., and Dejan, Z. (2015). Synchronous Gastric and Colonic Cancer--A Case Report. *Vojnosanit Pregl* 72, 642–645.

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

Casalino, L., and Verde, P. (2020). Multifaceted Roles of DNA Methylation in Neoplastic Transformation, from Tumor Suppressors to EMT and Metastasis. *Genes* 11, 922. doi:10.3390/genes11080922

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

Chivu Economescu, M., Necula, L. G., Dragu, D., Badea, L., Dima, S. O., Tudor, S., et al. (2010). Identification of Potential Biomarkers for Early and Advanced Gastric Adenocarcinoma Detection. *Hepatogastroenterology* 57, 1453–1464.

Dong, L., Chen, C., Zhang, Y., Guo, P., Wang, Z., Li, J., et al. (2021). The Loss of RNA N6-Adenosine Methyltransferase Mettl14 in Tumor-Associated Macrophages Promotes CD8+ T Cell Dysfunction and Tumor Growth. *Cancer Cell.* doi:10.1016/j.ccell.2021.04.016

Dong, S., Zhu, P., and Zhang, S. (2020). Expression of Collagen Type 1 Alpha 1 Indicates Lymph Node Metastasis and Poor Outcomes in Squamous Cell Carcinomas of the Lung. *PeerJ* 8, e10089. doi:10.7717/peerj.10089

Grigoroiu, M., Tagett, R., Draghici, S., Dima, S., Nastase, A., Florea, R., et al. (2015). Gene-expression Profiling in Non-small Cell Lung Cancer with Invasion of Mediastinal Lymph Nodes for Prognosis Evaluation. *Cancer Genomics Proteomics* 12, 231–242.

Hu, N., Clifford, R. J., Yang, H. H., Wang, C., Goldstein, A. M., Ding, T., et al. (2010). Genome Wide Analysis of DNA Copy Number Neutral Loss of Heterozygosity (CNNLOH) and its Relation to Gene Expression in Esophageal Squamous Cell Carcinoma. *BMC Genomics.* 11, 576. doi:10.1186/1471-2164-11-576

Huang, C., Yang, X., Han, L., Fan, Z., Liu, B., Zhang, C., et al. (2019). The Prognostic Potential of Alpha-1 Type I Collagen Expression in Papillary Thyroid Cancer. *Biochem. Biophysical Res. Commun.* 515, 125–132. doi:10.1016/j.bbrc.2019.04.119

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

Inoue, C., Miki, Y., Saito, R., Hata, S., Abe, J., Sato, I., et al. (2019). PD-L1 Induction by Cancer-Associated Fibroblast-Derived Factors in Lung Adenocarcinoma Cells. *Cancers* 11, 1257. doi:10.3390/cancers11091257

Jiang, Y., and Zhan, H. (2020). Communication between EMT and PD-L1 Signaling: New Insights into Tumor Immune Evasion. *Cancer Lett.* 468, 72–81. doi:10.1016/j.canlet.2019.10.013

Kim, S.-H., Park, B.-S., Kim, H. S., and Kim, J. H. (2017). Synchronous Quintuple Primary Gastrointestinal Tract Malignancies: Case Report. *Wjg* 23, 173–177. doi:10.3748/wjg.v23.i1.173

Koch, A., De Meyer, T., Jeschke, J., and Van Criekinge, W. (2015). MEXPRESS: Visualizing Expression, DNA Methylation and Clinical TCGA Data. *BMC Genomics* 16, 636. doi:10.1186/s12864-015-1847-z

Lakins, M. A., Ghorani, E., Munir, H., Martins, C. P., and Shields, J. D. (2018). Cancer-associated Fibroblasts Induce Antigen-specific Deletion of CD8 + T Cells to Protect Tumour Cells. *Nat. Commun.* 9, 948. doi:10.1038/s41467-018-03347-0

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

Li, Z., Zhou, J., Zhang, J., Li, S., Wang, H., and Du, J. (2019). Cancer-associated Fibroblasts Promote PD-L1 Expression in Mice Cancer Cells via Secreting CXCL5. *Int. J. Cancer* 145, 1946–1957. doi:10.1002/ijc.32278

Liu, C.-J., Hu, F.-F., Xia, M.-X., Han, L., Zhang, Q., and Guo, A.-Y. (2018a). GSCALite: a Web Server for Gene Set Cancer Analysis. *Bioinformatics* 34, 3771–3772. doi:10.1093/bioinformatics/bty411

Liu, J., Shen, J. X., Wu, H. T., Li, X. L., Wen, X. F., Du, C. W., et al. (2018b). Collagen 1A1 (COL1A1) Promotes Metastasis of Breast Cancer and Is a Potential Therapeutic Target. *Discov. Med.* 25, 211–223.

Ma, H.-P., Chang, H.-L., Bamodu, O. A., Yadav, V. K., Huang, T.-Y., Wu, A. T. H., et al. (2019). Collagen 1A1 (COL1A1) Is a Reliable Biomarker and Putative Therapeutic Target for Hepatocellular Carcinogenesis and Metastasis. *Cancers* 11, 786. doi:10.3390/cancers11060786

Nieto, M. A., Huang, R. Y.-J., Jackson, R. A., and Thiery, J. P. (2016). EMT: 2016. *Cell* 166, 21–45. doi:10.1016/j.cell.2016.06.028

Peng, D. H., Rodriguez, B. L., Diao, L., Chen, L., Wang, J., Byers, L. A., et al. (2020). Collagen Promotes Anti-PD-1/pd-L1 Resistance in Cancer through LAIR1-dependent CD8+ T Cell Exhaustion. *Nat. Commun.* 11, 4520. doi:10.1038/s41467-020-18298-8

Prockop, D. J. (1990). Mutations that Alter the Primary Structure of Type I Collagen. The Perils of a System for Generating Large Structures by the Principle of Nucleated Growth. *J. Biol. Chem.* 265, 15349–15352. doi:10.1016/s0021-9258(18)55398-9

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

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, 4200–4202. doi:10.1093/bioinformatics/btz210

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

Shi, Y., Wang, J., Xin, Z., Duan, Z., Wang, G., and Li, F. (2015). Transcription Factors and microRNA-Co-Regulated Genes in Gastric Cancer Invasion in *Ex Vivo*. *PLoS One* 10, e0122882. doi:10.1371/journal.pone.0122882

Su, S., Dong, Z.-Y., Xie, Z., Yan, L.-X., Li, Y.-F., Su, J., et al. (2018). Strong Programmed Death Ligand 1 Expression Predicts Poor Response and De Novo Resistance to EGFR Tyrosine Kinase Inhibitors Among NSCLC Patients with EGFR Mutation. *J. Thorac. Oncol.* 13, 1668–1675. doi:10.1016/j.jtho.2018.07.016

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., 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

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2019). The Immune Landscape of Cancer. *Immunity* 51, 411–412. doi:10.1016/j.immuni.2019.08.004

Vakili-Ghartavol, R., Mombeiny, R., Salmaninejad, A., Sorkhabadi, S. M. R., Faridi-Majidi, R., Jaafari, M. R., et al. (2018). Tumor-associated Macrophages and Epithelial-Mesenchymal Transition in Cancer: Nanotechnology Comes into View. *J. Cel Physiol.* 233, 9223–9236. doi:10.1002/jcp.27027

Van De Ven, S. E. M., Falger, J. M., Verhoeven, R. H. A., Baatenburg De Jong, R. J., Spaander, M. C. W., Bruno, M. J., et al. (2021). Increased Risk of Second Primary Tumours in Patients with Oesophageal Squamous Cell Carcinoma: A Nationwide Study in a Western Population. *United Eur. Gastroenterol. j.*, 9. 497–506. doi:10.1177/2050640620977129

Vlachavas, E. I., Pilalis, E., Papadodima, O., Koczan, D., Willis, S., Klippel, S., et al. (2019). Radiogenomic Analysis of F-18-Fluorodeoxyglucose Positron Emission Tomography and Gene Expression Data Elucidates the Epidemiological Complexity of Colorectal Cancer Landscape. *Comput. Struct. Biotechnol. J.* 17, 177–185. doi:10.1016/j.csbj.2019.01.007

Yamazaki, S., Higuchi, Y., Ishibashi, M., Hashimoto, H., Yasunaga, M., Matsumura, Y., et al.(2018). Collagen Type I Induces EGFR-TKI Resistance in EGFR-Mutated Cancer Cells by mTOR Activation through Akt-independent Pathway. *Cancer Sci.* 109, 2063–2073. doi:10.1111/cas.13624

Yin, Y., Du, L., Li, X., Zhang, X., and Gao, Y. (2019). miR-133a-3p Suppresses Cell Proliferation, Migration, and Invasion and Promotes Apoptosis in Esophageal Squamous Cell Carcinoma. *J. Cel Physiol.* 234, 12757–12770. doi:10.1002/jcp.27896

Yoshida, N., Eto, K., Kurashige, J., Izumi, D., Sawayama, H., Horinouchi, T., et al. (2020). Comprehensive Analysis of Multiple Primary Cancers in Patients with Esophageal Squamous Cell Carcinoma Undergoing Esophagectomy. *Ann. Surg.* Publish Ahead of Print. doi:10.1097/sla.0000000000004490

Yoshikawa, T., Aoki, K., Mitsuhashi, Y., Tomiura, S., Suto, A., Miura, T., et al. (2016). A Case of Double Cancer of Initially Unresectable Sigmoid Colon Cancer and Advanced Gastric Cancer Treated with Curative Resection after mFOLFOX6 Therapy. *Gan To Kagaku Ryoho* 43, 365–368.

Zhang, C., Liu, S., Wang, X., Liu, H., Zhou, X., and Liu, H. (2021). COL1A1 Is a Potential Prognostic Biomarker and Correlated with Immune Infiltration in Mesothelioma. *Biomed. Res. Int.* 2021, 1–13. doi:10.1155/2021/5320941

Zhang, Z., Fang, C., Wang, Y., Zhang, J., Yu, J., Zhang, Y., et al. (2018). COL1A1: A Potential Therapeutic Target for Colorectal Cancer Expressing Wild-type or Mutant KRAS. *Int. J. Oncol.* 53, 1869–1880. doi:10.3892/ijo.2018.4536

Zhao, L., Xing, W., Yang, Y., Zhang, Y., Ma, B., Fu, X., et al. (2021). The Sequence of Chemotherapy and Anti-PD-1 Antibody Influence the Efficacy of Neoadjuvant Immunochemotherapy in Locally Advanced Esophageal Squamous Cell Cancer: A Phase II Study. *J. Clin. Oncol.* 39 (Suppl. 15;abstr 4051). Abstract retrieved from Abstracts in American Society of Clinical Oncology(Accession No. 4051). doi:10.1200/jco.2021.39.15_suppl.4051

Zhu, H., Chen, H., Wang, J., Zhou, L., and Liu, S. (2019). Collagen Stiffness Promoted Non-muscle-invasive Bladder Cancer Progression to Muscle-Invasive Bladder Cancer. *Ott*. 12, 3441–3457. doi:10.2147/ott.S194568

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 Liu, Xue, Zhong, Wang, Li 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.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fmolb.2021.689466

Genomic Analysis Reveals Heterogeneity Between Lesions in Synchronous Primary Right-Sided and Left-Sided Colon Cancer

*Hanqing Hu[1†], Qian Zhang[2†], Rui Huang[1†], Zhifeng Gao[3], Ziming Yuan[1], Qingchao Tang[1], Feng Gao[1], Meng Wang[2], Weiyuan Zhang[1], Tianyi Ma[1], Tianyu Qiao[1], Yinghu Jin[1] and Guiyu Wang[2]**

[1]Colorectal Cancer Surgery Department, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Colorectal Cancer Surgery Department, Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China, [3]Department of Surgery, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

OPEN ACCESS

Edited by:
Xin Wang,
City University of Hong Kong, SAR China

Reviewed by:
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

***Correspondence:**
Guiyu Wang
guiywang@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 31 March 2021
Accepted: 07 June 2021
Published: 04 August 2021

Citation:
Hu H, Zhang Q, Huang R, Gao Z, Yuan Z, Tang Q, Gao F, Wang M, Zhang W, Ma T, Qiao T, Jin Y and Wang G (2021) Genomic Analysis Reveals Heterogeneity Between Lesions in Synchronous Primary Right-Sided and Left-Sided Colon Cancer. Front. Mol. Biosci. 8:689466. doi: 10.3389/fmolb.2021.689466

Background: The synchronous primary right-sided and left-sided colon cancer (sRL-CC) is a peculiar subtype of colorectal cancer. However, the genomic landscape of sRL-CC remains elusive.

Methods: Twenty-eight paired tumor samples and their corresponding normal mucosa samples from 14 patients were collected from the Second Affiliated Hospital of Harbin Medical University from 2011 to 2018. The clinical–pathological data were obtained, and whole-exome sequencing was performed based on formalin-fixed and paraffin-embedded samples of these patients, and then, comprehensive bioinformatic analyses were conducted.

Results: Both the lesions of sRL-CC presented dissimilar histological grade and differentiation. Based on sequencing data, few overlapping SNV signatures, onco-driver gene mutations, and SMGs were identified. Moreover, the paired lesions harbored a different distribution of copy number variants (CNVs) and loss of heterozygosity. The clonal architecture analysis demonstrated the polyclonal origin of sRL-CC and inter-cancerous heterogeneity between two lesions.

Conclusion: Our work provides evidence that lesions of sRL-CC share few overlapping mutational signatures and CNVs, and may originate from different clones.

Keywords: synchronous multiple primary cancer, genome, right-sided colon cancer, left-sided colon cancer, heterogeneity

INTRODUCTION

Synchronous multiple primary colorectal cancer (SM-CRC) refers to at least two primary lesions identified in a single patient. Synchronous primary right-sided and left-sided colon cancer (sRL-CC) is a peculiar subtype of SM-CRC. Previous studies have shown that genomic molecular aberration is a contributor to synchronous neoplasia (Ogino et al., 2006). However, the genomic landscape of sRL-CC remains to be elucidated.

Recently, many works have shown that solitary left-sided colon cancer and right-sided colon cancer have different biological behaviors. Clinical studies revealed that patients with right colon

cancer (RCC) have a worse overall survival (Ishihara et al., 2018). Patients with metastatic left colon cancer (LCC) respond more effectively to cetuximab-based target therapy compared to those with metastatic RCC (Elez et al., 2015; Arnold et al., 2017; Tejpar et al., 2017). In addition to clinical trials, molecular studies have shown different genomic phenotypes in LCC and RCC. *TP53* and *APC* are mutated more frequently in LCC, whereas *PIK3CA*, *CTNNB1*, *ATM*, *PTEN*, and *BRCA1* are mutated more frequently in RCC (Yamauchi et al., 2012; Takahashi et al., 2016; Salem et al., 2017). The consensus molecular subtype (CMS) according to the transcriptome has divided colorectal cancer into four subtypes: CMS1 (immune activation and JAK-STAT activation), CMS2 (WNT activation, MYC activation, EGFR or SRC activation, and VEGF or VEGFR activation), CMS3 (DNA damage repair, glutaminolysis, and lipidogenesis), and CMS4 (mesenchymal transition and complement activation immunosuppression) (Guinney et al., 2015). CMS1 and CMS3 are frequent in RCC, whereas CMS2 and CMS4 are frequent in LCC (Dienstmann et al., 2017). Moreover, LCC and RCC also present different immune landscapes. For RCC, decreased infiltration of $CD8^+$ T cells and Th1 cells were identified. For LCC, infiltration of $CD56^{high}$ natural killer cells and activation of IFN-α signaling were identified (Zhang et al., 2018). Despite continuous anatomy, left-sided and right-sided colon cancer could be two sides of a coin in solitary colorectal cancer, which provokes us to explore the molecular phenotype of sRL-CC.

Previous studies had reported the heterogeneity and independent genetic origin of synchronous colorectal cancer (Cereda et al., 2016; Wang et al., 2018), but these studies have not compared the genetic phenotype according to the tumor location. In our present work, we analyzed 14 cases of sRL-CC to investigate single nucleotide variation, somatic mutation, and copy number alteration in sRL-CC patients based on whole-exome sequencing data.

MATERIALS AND METHODS

Sample Collection

Nineteen sRL-CC patients were enrolled from March 2011 to October 2018 and five patients were excluded due to the small sample volume of the tumor tissue. The diagnosis was confirmed by two experts in the Department of Pathology from the Second Affiliated Hospital of Harbin Medical University. Fourteen patients denied hereditary history of CRC and were diagnosed with sporadic CRC. All the samples were formalin-fixed and paraffin-embedded. The splenic flexure, descending colon, and sigmoid colon were classified as the left-sided colon, while the caecum, hepatic flexure, and ascending colon were classified as the right-sided colon (Missiaglia et al., 2014). The clinicopathological data of the 14 sRL-CC were all available. All the samples were acquired with the approval of the ethics committee from the Second Affiliated Hospital of Harbin Medical University and written informed consent was obtained from all the participants.

DNA Extraction and Whole-Exome Sequencing

The DNA was extracted from FFPE samples using a QIAamp DNA FFPE Tissue Kit (Qiagen, China), according to the manufacturer's protocol, after each slide was reviewed by two pathological experts to ensure tumor purity was greater than 50%. The quality of DNA and contamination was evaluated on 1% agarose gels and the concentration of DNA was calculated by a Qubit DNA Assay Kit in a Qubit 2.0 Fluorometer (Invitrogen, China). Two micrograms of genomic DNA was used to prepare the captured libraries by an Agilent SureSelect Human All Exon V5 kit (Agilent Technologies, China), following the manufacturer's recommendations. The reads library was sequenced on the Illumina Novaseq 6000 platform.

Raw Data Processing for Calling Single Nucleotide Variants and Indels

The clean data were obtained after discarding the adapter and paired reads of the raw data from the Novaseq 6000 platform. Next, quality control was performed including reads number, error rate, and percentage of reads with average quality (>Q20 and >Q30). Burrows–Wheeler Aligner (BWA) software was used to map the paired-end clean reads to the reference genome (UCSC Human Genome Reference hg38) (Li and Durbin, 2009). Aligned reads were processed in terms of marked duplicates, realignment of indels, and base recalibration by the Genome Analysis Toolkit (GATK) (McKenna et al., 2010). Variants were identified in accordance to dbSNP (Sherry et al., 2001) and the 1000 Genomes database (Genomes Project et al., 2012), which was annotated by ANNOVAR (Wang et al., 2010). Next, SNVs and indels were identified by MuTech (Cibulskis et al., 2013) and Strelka (Saunders et al., 2012).

Based on these data, cluster analysis was conducted on 96 somatic mutational nucleotide types through nonnegative matrix factorization (Nik-Zainal et al., 2012), and three signatures were identified (Alexandrov et al., 2013a). These signatures were clustered based on 30 known signatures according to the COSMIC database to reveal the biological process of the signatures (Alexandrov et al., 2013b). The degree of similarity was assessed by the cosine similarity coefficient index.

Mutation Signature Analysis

To investigate the predisposing genes, we analyzed germline mutation in the normal tissue. Compared with the Cancer Gene Census (CGC) database (Gerlinger et al., 2014) using an in-house algorithm, the predisposing genes were identified. To determine driver mutation genes in carcinogenesis, the mutation spectrum was aligned with published driver mutations *via* OncodriveCLUSTL (Arnedo-Pac et al., 2019) and OncodriveFM (Gonzalez-Perez and Lopez-Bigas, 2012) software referring to four databases, the CGC513 database (Futreal et al., 2004), 125 mutation genes reported by Bert Vogelstein (Vogelstein et al., 2013), SMG127 (Kandoth et al.,

TABLE 1 | Demographic characteristic of the patients in our study.

	Age	Gender	CEA (ng/ml)	CA199 (U/ml)	Location	Histology	Differentiation	Stage	Perineural invasion	Venous invasion	Lymphatic invasion
1	60	F	6.58	146.26	Right colon	TA/MA	Moderately	$T_3N_{1c}M_0$	Yes	Yes	Yes
					Sigmoid colon	TA	Well/ moderately	$T_1N_0M_0$	No	No	No
2	36	F	18.71	32.57	Right colon	TA	Moderately	$T_3N_0M_0$	No	No	No
					Sigmoid colon	TA	Moderately	$T_2N_0M_0$	No	No	No
3	84	M	3.15	17.26	Right colon	TA/MA	Poorly	$T_3N_{1a}M_0$	No	Yes	Yes
					Left colon	TA	Well	$T_2N_0M_0$	No	No	No
4	68	M	6.85	3.22	Right colon	TA	Moderately	$T_3N_0M_0$	No	Yes	No
					Sigmoid colon	TA	Moderately	$T_3N_0M_0$	Yes	Yes	No
5	70	M	10.34	16.39	Right colon	VA	Moderately	$T_3N_0M_0$	Yes	No	Yes
					Sigmoid colon	VA	Moderately	$T_1N_0M_0$	No	No	No
6	74	F	3.65	4.06	Right colon	TA	Moderately	$T_3N_0M_0$	No	No	No
					Sigmoid colon	TA	Moderately	$T_2N_{1a}M_0$	No	No	Yes
7	83	M	1.81	34.54	Right colon	TA	Well	$T_1N_0M_0$	No	No	No
					Sigmoid colon	TA	Well	$T_1N_0M_0$	No	No	No
8	45	F	1.37	11.06	Right colon	TA/MA	Well/ moderately	$T_2N_0M_0$	No	No	No
					Left colon	TA	Well/ moderately	$T_1N_0M_0$	No	No	No
9	75	M	49.83	31.53	Right colon	TA	Moderately	$T_3N_0M_0$	No	No	No
					Sigmoid colon	TA	Moderately	$T_3N_0M_0$	No	No	No
10	84	M	0.95	1.92	Right colon	VA	Well/ moderately	$T_2N_0M_0$	No	No	No
					Sigmoid colon	TA	Moderately	$T_1N_0M_0$	No	No	No
11	58	M	28.36	79.83	Right colon	TA	Moderately	$T_3N_0M_0$	Yes	Yes	Yes
					Sigmoid colon	TA	Moderately	$T_2N_0M_0$	No	No	No
12	75	M	1.66	78.52	Right colon	TA/MA	Moderately	$T_3N_0M_0$	No	No	No
					Sigmoid colon	TA	Moderately	$T_1N_0M_0$	No	No	No
13	60	M	3.7	43.47	Right colon	TA	Moderately	$T_3N_0M_0$	Yes	Yes	Yes
					Sigmoid colon	TA	Moderately	$T_3N_0M_0$	Yes	No	No
14	50	M	24.08	8.98	Right colon	MA	Poorly	$T_3N_0M_0$	Yes	No	No
					Sigmoid colon	TA	Moderately	$T_1N_0M_0$	Yes	No	Yes

F, female; M, male; TA, tubular adenocarcinoma; VA, villous adenocarcinoma; MA, mucinous adenocarcinoma.

2013), and Comprehensive435 (Tamborero et al., 2013). A significantly mutated genes test was applied to define the SMGs in the tumor, and a mutation relation rest was applied to explore the relationship between SMGs (Dees et al., 2012).

Phylogenetic Tress and Clonal Architecture Analysis

Phylogenetic trees analysis was performed based on the WES data. The branch and trunk lengths reflected the number of nonsynonymous mutations, which were also marked beside the branch and trunk (Gerlinger et al., 2014). In accordance with the Pyclone algorithm, clonal architecture analysis was performed (Roth et al., 2014).

Copy Number Analysis

The copy number variants (CNVs) were evaluated by control-FREEC software based on the WES data (Boeva et al., 2012). The profiles of CNVs were calculated with alignment to the BAM data. Next, the CNVs were normalized to obtain the number of CNVs in different regions on the chromosome.

Statistical Analyses

Statistical analyses were conducted by SPSS software (version 25.0). The comparison between the left-sided and right-sided lesions with SNP signature and TMB was performed by a two-tailed paired t test. The differences were considered to be significant at $p < 0.05$. The p values are described in the corresponding figure legend.

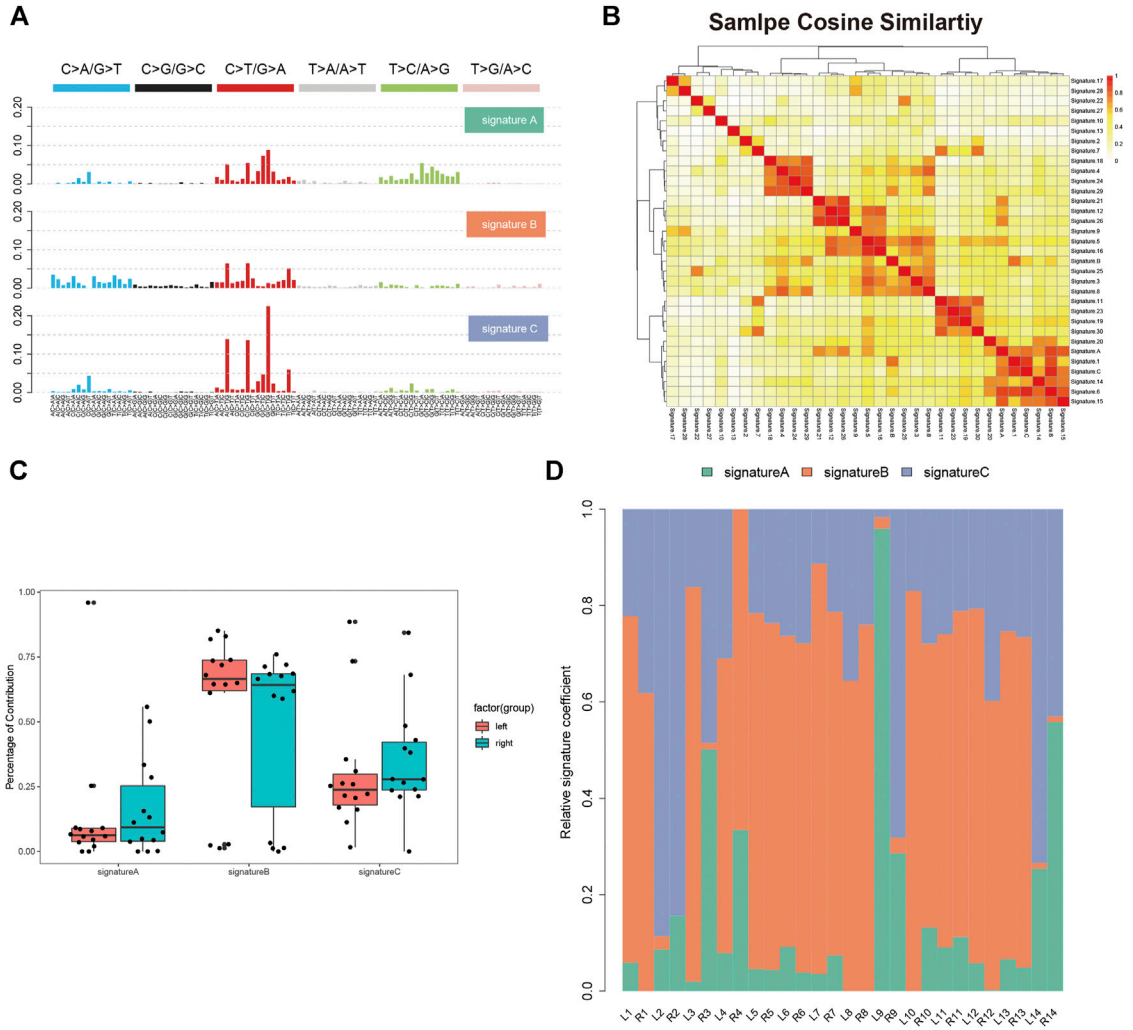


FIGURE 1 | Single nucleotide variants analysis of sRL-CC. **(A)** Three SNV signatures are obtained from WES data in 14 sRL-CC patients. **(B)** Correlation analysis of three SNV signatures with known 30 signatures. The bar represents the coefficient index. **(C)** The box plot shows the percentage of the three signatures in left-sided and right-sided lesions when analyzed as a whole. Signature A: $p = 0.722$, two-tailed paired t test; signature B: $p = 0.539$, two-tailed paired t test; and signature C: $p = 0.589$, two-tailed paired t test. **(D)** The distribution of the three signatures in two lesions of each patient is presented by the heatmap.

RESULTS

Clinical Characteristics of Synchronous Primary Right-Sided and Left-Sided Colon Cancer

We collected 9,876 colorectal cancer patients, who were confirmed by pathological examination, in the Second Affiliated Hospital of Harbin Medical University from 2011 to 2018. A total of 7,369 patients received radical colectomy, and the samples were preserved by the formalin-fixed paraffin-embedded method. Among these patients, 67 were diagnosed with SM-CRC and 19 SM-CRC patients were identified as sRL-CC. The tumor location was evaluated by colonoscopy or computed tomography. During the surgery, the tumor location was confirmed by two experienced surgeons and recorded in the pathological reports. At last, 14 patients were enrolled for further analysis and 5 patients were excluded due to the small size of the tumor. The demographic characteristics of the 14 patients are presented in **Table 1**. As shown in **Table 1**, the median age of these patients was 69 years. Of the 14 patients, 4 were women and 10 were men. The CEA and CA-199 level was elevated (cutoff value, CEA: 5 ng/ml, CA-199: 30 U/ml) in most cases, suggesting a heavy tumor burden. As for the two lesions in most patients, the left lesions were located at the sigmoid colon and the right lesions were located at the right colon. However, in two patients, the left lesions were located at the left colon and right lesions were located at the right colon. Mucinous adenocarcinoma

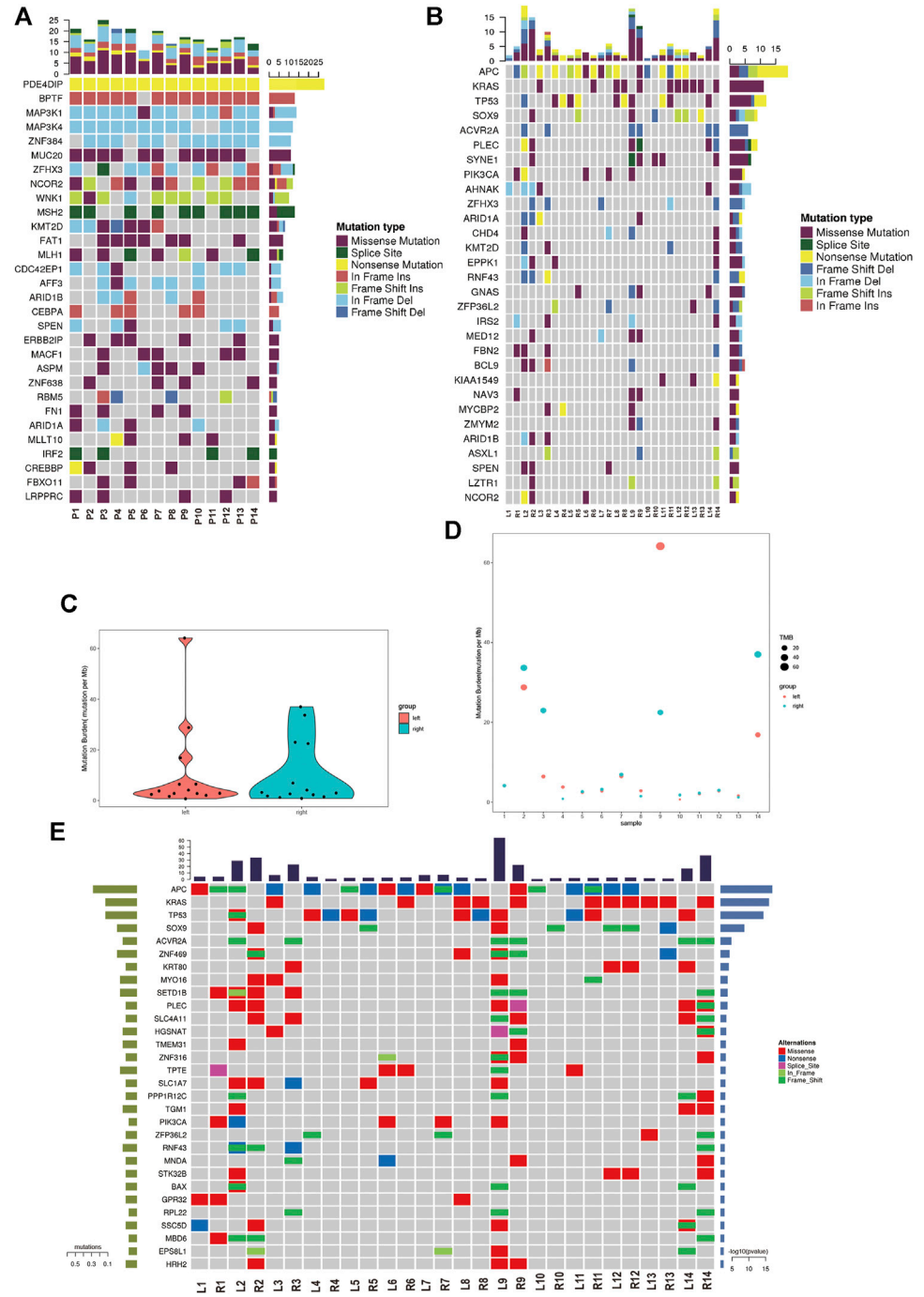


FIGURE 2 | The mutational analysis of sRL-CC. **(A)** The frequent predisposing genes are presented by the heatmap. The tumor mutation burden is marked on the top of the heatmap, the name of predisposing genes is marked on the left, and the number of mutated samples is marked on the right. **(B)** The oncodriver gene mutations are presented by the heatmap. The tumor mutation burden is marked on the top of the heatmap, the name of predisposing genes is marked on the left, and the number of mutated samples is marked on the right. **(C)** The tumor mutation burden is calculated in the left-sided and right-sided lesions. **(D)** The tumor mutation burden is calculated in both lesions of each patient. **(E)** The significant mutated genes are presented by the heatmap. The tumor mutation burden is marked on the top of the heatmap.

was identified in right-sided lesions but not in left-sided lesions. The right-sided lesions showed an advanced disease compared to the left-sided lesions according to the pathological stages and indicators, which suggested that left-sided and right-sided lesions may be at different stages of CRC development.

Single Nucleotide Variants Analysis of Synchronous Primary Right-Sided and Left-Sided Colon Cancer

We analyzed the single nucleotide variants (SNVs) based on the WES data of 14 paired tumor specimen and their corresponding normal mucosa. Six types of SNVs were identified in sRL-CC patients, and C>T/G>A transition was preponderant nucleotide substitution in all the samples (**Supplementary Figure S1A**). Then, we evaluated the percentage of each type, which were similar in left and right lesions (**Supplemental Figure S1B**). To explore the signatures of the SNV spectrum, we clustered SNVs to obtain three signatures (**Figure 1A**). In signature A, C>T/G>A and T>C/A>G transition was abundant, signature B featured C>A/G>T and C>T/G>A transition, and signature C was defined as a high level of C>T/G>A transition (**Figure 1A**). To investigate the function of the three signatures, we performed cosine similarity analysis according to the COSMIC database (Blokzijl et al., 2018). It revealed that signature A was mostly related to signature 6 (correlation coefficient: 0.837), which was associated with defective DNA mismatch repair. Signature B was mostly related to signature 1 (correlation coefficient: 0.725), which was associated with the biological process initiated by spontaneous deamination of 5-methylcytosine. Signature C was also mostly related to signature 6 (correlation coefficient: 0.920) (**Figure 1B**). To evaluate the composition of signatures in paired samples, we analyzed the left-sided and right-sided lesions as a whole. In spite of no statistical significance resulting from the small number of patients, the average percentage of signature A and C contribution was higher in right lesions, but the average percentage of signature B contribution was higher in left lesions (**Figure 1C**). Then, we evaluated the signatures in a single patient, respectively. For instance, signature B was dominated in L2, but signatures A and C were dominated in R2. In patient 9, signature A was dominated in the left lesion and signature C was dominated in the right lesions. There were still some special cases, such as patients 5–8 and 13, in whom the dominated signature was similar (**Figure 1D**). This finding suggests that distinctive mutational processes happen in the right or left lesion during carcinogenesis.

Mutation Signature Between Lesions of Synchronous Primary Right-Sided and Left-Sided Colon Cancer

We analyzed the germline mutation in the normal tissues of these patients to define predisposing genes using the in-house method and several frequent predisposing genes were identified. The nonsense mutation of *PDE4DIP* was found in all of 14 patients. The in-frame insertion of *BPTF* was found in 13 patients except patient 6. The in-frame deletion of *MAP3K1/4* and *ZNF384* was found in the majority of the patients. The missense mutation of *MUC20* was found in 11 out of 14 patients (**Figure 2A**). Cancer driver gene mutation plays a significant role in carcinogenesis, and we explored the oncodriver mutation in sRL-CC. The completely matched lesions failed to be identified in all 14 patients. Although similar in a small part, the features of cancer driver gene mutation between lesions were diversified in most patients. For example, frameshift insertion of *APC* and frameshift deletion of *TP53* was detected in L2 but not in R2, whereas missense mutation of *SOX9*, frameshift deletion of *ZFHX3*, and missense mutation of *SYNE1* were detected in R2 but not in L2 (**Figure 2B**). The significant mutation genes (SMGs) reflect the mutational phenotype of diseases; and then, we wanted to investigate the SMGs in all the lesions. When treated as a whole, the tumor mutation burden (TMB) was alike between left- and right-sided lesions (**Figure 2C**). When analyzed separately, TMB was different in patients 2, 3, 4, 8, 9, 10, and 14 and was similar in patients 5, 6, 7, 11, 12, and 13. Particularly, the TMB was almost the same in patient 1 (L: 4.1537674 vs. R: 4.1569451) (**Figure 2D**). Among these significant mutation genes, the *APC* mutation tended to happen in the left lesions, while *BAX*, *KRAS*, and *SOX9* mutations tended to happen in right-sided lesions. The *PI3KCA* mutation was also distributed unevenly, which was found in L2, L6, L9, R1, and R7 but not in the corresponding lesions on the other side (**Figure 2E**). The frequent mutation genes identified in sRL-CCs were similar to those identified in solitary colon cancer. Two studies identified eight frequent mutation genes (*APC*, *TP53*, *SMAD4*, *PIK3CA*, *KRAS*, *ARID1A*, *SOX9*, and *FAM123B*). *APC*, *TP53*, *KRAS*, and *SOX9* were the top four mutation genes in sRL-CCs (Cancer Genome Atlas, 2012; Vasaikar et al., 2019). The *ACVR2A* mutation was found in hypermutation colon cancers (Vasaikar et al., 2019), which was also identified in sRL-CCs. Mutation genes in a sample may have a synergistic effect or a mutually exclusive effect. Through mutation relation rest analysis (MRT), a synergistic and mutually exclusive relationship was identified. We found that *ACVR2A* had a synergistic effect with *PLEC*, *SLC4A11*, and *CEL* (**Supplementary Figure S2A**). Besides, we found that the *APC* mutation was mutually exclusive with *EPS8L1*, *PPP1R12C*, *SETD1B*, *RNF43*, and *SLC4A11*. The *KRAS* mutation was mutually exclusive with *TP53*, *SLC1A7*, and *PIK3CA* (**Supplementary Figure S2B**). These works showed the different mutational landscape in two lesions of a single sRL-CC patient.

Copy Number Variant Analysis of Synchronous Primary Right-Sided and Left-Sided Colon Cancer

Subsequently, we assessed the CNVs in both lesions of sRL-CCs. The GISTIC curve showed that amplification of the chromosome was predominant and more deletion events were identified in left-sided lesions (**Figure 3A**). The distribution of somatic CNVs in left-sided lesions was also different from those in right-sided lesions (**Figure 3B**). Loss of heterozygosity (LOH) was evaluated by β-allelic frequency (BAF). The right-sided and left-sided lesions shared different BAF (**Figure 3B**). When analyzed separately, the heterogeneity stood out more clearly. As shown in the Circos plot (**Figure 3C**), the copy number was almost normal in L2 and R4 lesions, whereas the copy number was amplified in R2 and L4 lesions. Similar phenomena were also found in other patients of our cohort (**Supplementary Figure S3**). These data imply that the lesions share different patterns

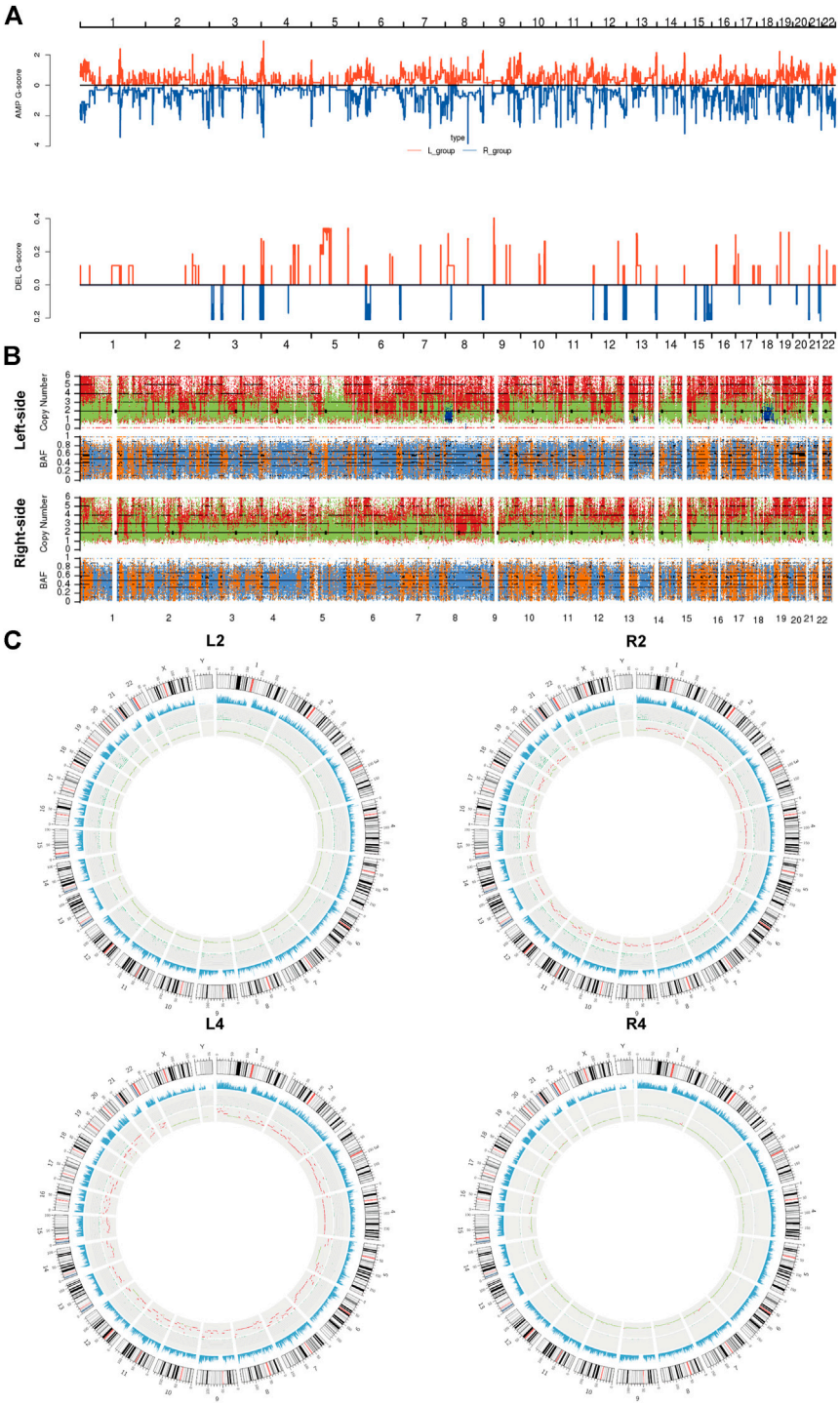


FIGURE 3 | The copy number variants analysis of sRL-CC. **(A)** GISTIC analysis of left-sided and right-sided lesions. The G-score represents the degree of CNV amplification **(upper panel)** and deletion **(lower panel)**. L-group means left-sided lesions and R-group means right-sided lesions. **(B)** The upper chart illustrates the distribution of copy number variants. Red presents the CNV gain, green presents normal CNV, and blue presents CNV loss. The lower chart illustrates the distribution of β allelic frequency (BAF). Orange means the normal distribution of allele and blue means loss of heterozygosity. **(C)** Circos plots illustrate the molecular landscape of genome in patients 2 and 4 in terms of copy number variants in each patient. The first circle means the depth of sequencing, the second circle means the density of SNP insert and deletion, and the third circle means the distribution of CNVs. In the third circle, red means the CNV gains, green means normal copy number, and blue means CNV loss.

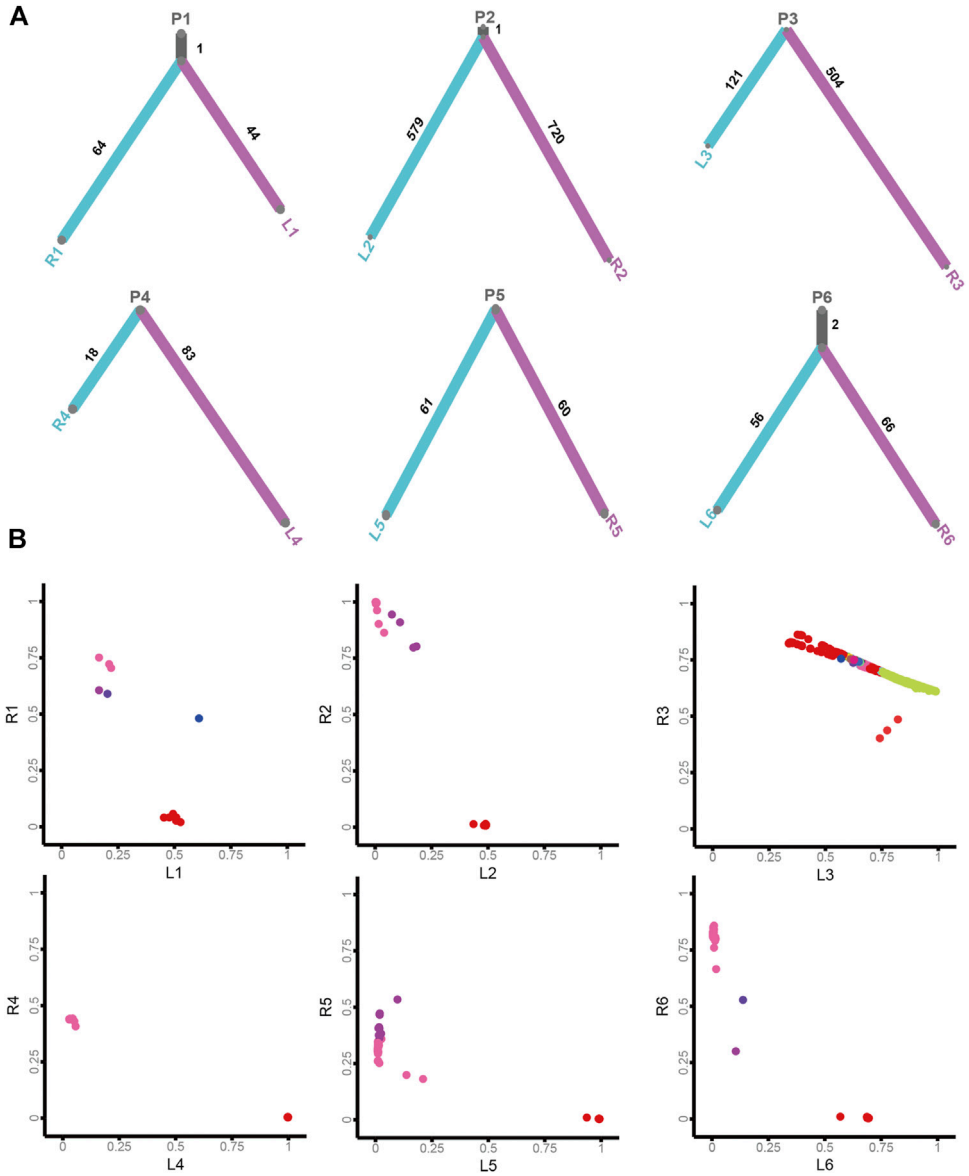


FIGURE 4 | The clonal architecture and evolution analysis of sRL-CC. **(A)** The phylogenetic tress analysis of patients 1–6. The trunk represents the common mutations and the branch represents unique mutations in each lesion. The number of mutations is marked. **(B)** The scatter plot demonstrates cancer cell fraction (CCF) of the mutations in both lesions of patients 1–6. Different colors represent different clusters. The clusters located in the middle area of the plot mean shared clusters by both lesions, and those located near the X-axis and Y-axis mean unique clusters in each lesion.

with respect to copy number and different stages during carcinogenesis.

Clonal Architecture and Evolution Analysis of Synchronous Primary Right-Sided and Left-Sided Colon Cancer

To acquire insights into the origin of both lesions shaping sRL-CC tumorigenesis, phylogenetic trees analyses were constructed to evaluate the ancestral relationship of individual lesions, significantly overlapping variant sets failed to be identified in most cases in our cohort. None of the overlapping variant sets were identified in patients 3, 4, 5, 7, 8, 9, 10, 11, 13, and 14, and few overlapping variant sets were identified in patients 1, 2, and 6. But as an exception, both the lesions in patient 12 shared 53 overlapping variant sets (**Figure 4A** and **Supplementary Figure S4A**). Clonal architecture analysis revealed different clusters between two lesions. Most of the patients shared few overlapping clusters. In patients 3 and 12, relatively more sharing clusters were identified with some independent clusters. However, in patient 9, a single shared cluster was identified (**Figure 4B** and **Supplementary Figure S4B**). These data may suggest different origins and clonal architectures in the lesion of sRL-CCs.

DISCUSSION

In this study, we performed an unprecedented molecular characterization of both lesions and corresponding normal adjacent tissue in sRL-CCs with comprehensive data from WES. Our work revealed that heterogeneity occurred in both lesions in the same patient and each lesion was in a different stage during carcinogenesis. Our study is in line with the multi-omics analysis in single primary colorectal cancer. Vasaikar et al. reported some significant mutations such as *APC*, *TP53*, *KRAS*, *SOX9*, and *ACVR2A*, which were also found in our cohort (Vasaikar et al., 2019). The mutation of mismatch repair genes is a key molecular event in colorectal cancer (Sahin et al., 2019; Schrock et al., 2019). Our work has also identified that SNV signature is associated with defective DNA mismatch repair. All this evidence confirms that our data are accurate enough to reflect the biological process in sRL-CCs.

Recently, more and more attention has been paid to the multiple primary tumors. Grolleman et al. has reported the mutation of the base excision repair gene *NTHL1* that can trigger the development of malignancy in many organs (Grolleman et al., 2019). Our work also showed the mutational signature involved in the dysfunction of DNA mismatch repair in sRL-CCs. Our study has also highlighted predisposing genes in sRL-CC. The nonsense mutation of *PDE4DIP* was identified in all the cases in our cohort. The *PDE4DIP* gene encodes protein myomegalin and exerts its function as an anchor to sequester components to Golgi and/or centrosomes (Wang et al., 2014; Wu et al., 2016; Bouguenina et al., 2017; Yang et al., 2017). However, the role of *PDE4DIP* in cancers is poorly understood. Thus, prospective studies are needed to explore the role of nonsense mutation in *PDE4DIP* of sRL-CCs.

Above all, our study reveals heterogeneity of both lesions in sRL-CCs. We have identified different copy numbers in paired right-sided and left-sided lesions, which indicates that the lesions are in different stages of carcinogenesis. LOH is an important mechanism for the disability of tumor suppressor genes. Through the control-FREEC method (Boeva et al., 2012), the regions of LOH have been identified. The paired lesions share unmatched distribution of LOH on the chromosomes. This evidence suggests that inconsistent degrees of genome instability happen in the both lesions. Combined with the phylogenetic trees and clonal architecture analysis, we have proved the multi-origins of the paired lesions and different clonal fractions in the paired lesions. Studies based on multiple primary cancer in other organs have drawn a similar conclusion. Ma et al. reported that multicentric lesions harbor distinct oncogenic alterations and genomic heterogeneity (Ma et al., 2017). In synchronous bilateral renal cancer, the lesions in the kidneys also originate from separated clones (Linehan, 2009). In line with these previous studies, our work has shown the polyclonal origin of sRL-CC.

Limitations do exist in our study. The sample size is small and only 14 patients are included. Besides, our study is a single center, retrospective study and all the samples have to be preserved by FFPE due to the long time span. Thus, a multi-center, prospective study could collect a larger number of sRL-CC patients and fresh tumor tissue in a shorter time. Moreover, genomic analysis has trouble in reflecting the whole scenarios of sRL-CC. Systematic studies including the transcriptome, epi-transcriptome, proteome, and metabolome can provide a wider horizon for sRL-CC.

In conclusion, we performed whole-exome sequencing analyses, which are suggestive of heterogeneity between lesions and the polyclonal origin of sRL-CC. Moreover, we illustrated the genomic landscape of sRL-CC and provided an insight into the molecular pattern of sRL-CC, which could make treatment more precise and effective.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the Chinese laws forbid the publication the genomic data based on the Chinese population. Requests to access the datasets should be directed to the corresponding author GW, guiywang@163.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee from the Second Affiliated Hospital of Harbin Medical 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

HH, RH, and GW: study design, data interpretation, preparation, and submission of the manuscript. ZY, QZ, QT, and FG: clinical and pathological data collection, DNA extraction, and DNA quality control. MW, WZ, TM, TQ, and YJ: whole-exsome sequencing and raw data collection.

FUNDING

This work was supported by the China Postdoctoral Science Foundation (Grant number: 2018M641856), Heilongjiang Postdoctoral Fund (Grant number: LBH-Z18200), Heilongjiang Applied and Development Program (Grant number: GA19C003), Natural Science Foundation of Heilongjiang Province (Grant number: H2017016), Postgraduate Research & Practice Innovation Program of Harbin Medical University (Grant number: YJSKYCX 2019-51HYD), and National Natural Science Foundation of China (Grant number: 81872034).

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.689466/full#supplementary-material

REFERENCES

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013a). Signatures of Mutational Processes in Human Cancer. *Nature* 500, 415–421. doi:10.1038/nature12477

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J., and Stratton, M. R. (2013b). Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep.* 3, 246–259. doi:10.1016/j.celrep.2012.12.008

Arnedo-Pac, C., Mularoni, L., Muiños, F., Gonzalez-Perez, A., and Lopez-Bigas, N. (2019). OncodriveCLUSTL: A Sequence-Based Clustering Method to Identify Cancer Drivers. *Bioinformatics* 35, 4788–4790. doi:10.1093/bioinformatics/btz501

Arnold, D., Lueza, B., Douillard, J.-Y., Peeters, M., Lenz, H.-J., Venook, A., et al. (2017). Prognostic and Predictive Value of Primary Tumour Side in Patients With RAS Wild-Type Metastatic Colorectal Cancer Treated with Chemotherapy and EGFR Directed Antibodies in Six Randomized Trials. *Ann. Oncol.* 28, 1713–1729. doi:10.1093/annonc/mdx175

Blokzijl, F., Janssen, R., van Boxtel, R., and Cuppen, E. (2018). Mutational Patterns: Comprehensive Genome-Wide Analysis of Mutational Processes. *Genome Med.* 10, 33. doi:10.1186/s13073-018-0539-0

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., et al. (2012). Control-FREEC: A Tool for Assessing Copy Number and Allelic Content Using Next-Generation Sequencing Data. *Bioinformatics* 28, 423–425. doi:10.1093/bioinformatics/btr670

Bouguenina, H., Salaun, D., Mangon, A., Muller, L., Baudelet, E., Camoin, L., et al. (2017). EB1-Binding-Myomegalin Protein Complex Promotes Centrosomal Microtubules Functions. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10687–E10696. doi:10.1073/pnas.1705682114

Cancer Genome Atlas (2012). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Cereda, M., Gambardella, G., Benedetti, L., Iannelli, F., Patel, D., Basso, G., et al. (2016). Patients with Genetically Heterogeneous Synchronous Colorectal Cancer Carry Rare Damaging Germline Mutations in Immune-Related Genes. *Nat. Commun.* 7, 12072. doi:10.1038/ncomms12072

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., et al. (2013). Sensitive Detection of Somatic Point Mutations in Impure and Heterogeneous Cancer Samples. *Nat. Biotechnol.* 31, 213–219. doi:10.1038/nbt.2514

Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., et al. (2012). MuSiC: Identifying Mutational Significance in Cancer Genomes. *Genome Res.* 22, 1589–1598. doi:10.1101/gr.134635.111

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus Molecular Subtypes and the Evolution of Precision Medicine in Colorectal Cancer. *Nat. Rev. Cancer* 17, 79–92. doi:10.1038/nrc.2016.126

Elez, E., Argiles, G., and Tabernero, J. (2015). First-Line Treatment of Metastatic Colorectal Cancer: Interpreting FIRE-3, PEAK, and CALGB/SWOG 80405. *Curr. Treat Options Oncol.* 16, 52. doi:10.1007/s11864-015-0369-x

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., et al. (2004). A Census of Human Cancer Genes. *Nat. Rev. Cancer*. 4, 177–183. doi:10.1038/nrc1299

Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An Integrated Map of Genetic Variation From 1,092 Human Genomes. *Nature* 491, 56–65. doi:10.1038/nature11632

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., Varela, I., et al. (2014). Genomic Architecture and Evolution of Clear Cell Renal Cell Carcinomas Defined by Multiregion Sequencing. *Nat. Genet.* 46, 225–233. doi:10.1038/ng.2891

Gonzalez-Perez, A., and Lopez-Bigas, N. (2012). Functional Impact Bias Reveals Cancer Drivers. *Nucleic Acids Res.* 40, e169. doi:10.1093/nar/gks743

Grolleman, J. E., de Voer, R. M., Elsayed, F. A., Nielsen, M., Weren, R. D. A., Palles, C., et al. (2019). Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-Tumor Phenotype. *Cancer Cell* 35, 256–266. doi:10.1016/j.ccell.2018.12.011

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

Ishihara, S., Murono, K., Sasaki, K., Yasuda, K., Otani, K., Nishikawa, T., et al. (2018). Impact of Primary Tumor Location on Postoperative Recurrence and Subsequent Prognosis in Nonmetastatic Colon Cancers. *Ann. Surg.* 267, 917–921. doi:10.1097/sla.0000000000002206

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational Landscape and Significance Across 12 Major Cancer Types. *Nature* 502, 333–339. doi:10.1038/nature12634

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Linehan, W. M. (2009). Genetic Basis of Bilateral Renal Cancer: Implications for Evaluation and Management. *J. Clin. Oncol.* 27, 3731–3733. doi:10.1200/jco.2009.23.0045

Ma, P., Fu, Y., Cai, M. C., Yan, Y., Jing, Y., Zhang, S., et al. (2017). Simultaneous Evolutionary Expansion and Constraint of Genomic Heterogeneity in Multifocal Lung Cancer. *Nat. Commun.* 8, 823. doi:10.1038/s41467-017-00963-0

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Missiaglia, E., Jacobs, B., D'Ario, G., Di Narzo, A. F., Soneson, C., Budinska, E., et al. (2014). Distal and Proximal colon Cancers Differ in Terms of Molecular, Pathological, and Clinical Features. *Ann. Oncol.* 25, 1995–2001. doi:10.1093/annonc/mdu275

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., et al. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* 149, 979–993. doi:10.1016/j.cell.2012.04.024

Ogino, S., Brahmandam, M., Kawasaki, T., Kirkner, G. J., Loda, M., and Fuchs, C. S. (2006). Epigenetic Profiling of Synchronous Colorectal Neoplasias by Quantitative DNA Methylation Analysis. *Mod. Pathol.* 19, 1083–1090. doi:10.1038/modpathol.3800618

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., et al. (2014). PyClone: Statistical Inference of Clonal Population Structure in Cancer. *Nat. Methods* 11, 396–398. doi:10.1038/nmeth.2883

Sahin, I. H., Akce, M., Alese, O., Shaib, W., Lesinski, G. B., El-Rayes, B., et al. (2019). Immune Checkpoint Inhibitors for the Treatment of MSI-H/MMR-D Colorectal Cancer and a Perspective on Resistance Mechanisms. *Br. J. Cancer* 121, 809–818. doi:10.1038/s41416-019-0599-y

Salem, M. E., Weinberg, B. A., Xiu, J., El-Deiry, W. S., Hwang, J. J., Gatalica, Z., et al. (2017). Comparative Molecular Analyses of Left-Sided Colon, Right-Sided colon, and Rectal Cancers. *Oncotarget* 8, 86356–86368. doi:10.18632/oncotarget.21169

Saunders, C. T., Wong, W. S. W., Swamy, S., Becq, J., Murray, L. J., and Cheetham, R. K. (2012). Strelka: Accurate Somatic Small-Variant Calling From Sequenced Tumor-normal Sample Pairs. *Bioinformatics* 28, 1811–1817. doi:10.1093/bioinformatics/bts271

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

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: The NCBI Database of Genetic Variation. *Nucleic. Acids Res.* 29, 308–311. doi:10.1093/nar/29.1.308

Takahashi, Y., Sugai, T., Habano, W., Ishida, K., Eizuka, M., Otsuka, K., et al. (2016). Molecular Differences in the Microsatellite Stable Phenotype Between Left-Sided and Right-Sided Colorectal Cancer. *Int. J. Cancer*. 139, 2493–2501. doi:10.1002/ijc.30377

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., et al. (2013). Comprehensive Identification of Mutational Cancer Driver Genes Across 12 Tumor Types. *Sci. Rep.* 3, 2650. doi:10.1038/srep02952

Tejpar, S., Stintzing, S., Ciardiello, F., Tabernero, J., Van Cutsem, E., Beier, F., et al. (2017). Prognostic and Predictive Relevance of Primary Tumor Location in Patients With RAS Wild-Type Metastatic Colorectal Cancer. *JAMA Oncol.* 3, 194–201. doi:10.1001/jamaoncol.2016.3797

Vasaikar, S., Huang, C., Wang, X., Petyuk, V. A., Savage, S. R., Wen, B., et al. (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* 177, 1035–1049. doi:10.1016/j.cell.2019.03.030

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., and Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science* 339, 1546–1558. doi:10.1126/science.1235122

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants From High-Throughput Sequencing Data. *Nucleic. Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wang, X., Fang, H., Cheng, Y., Li, L., Sun, X., Fu, T., et al. (2018). The Molecular Landscape of Synchronous Colorectal Cancer Reveals Genetic Heterogeneity. *Carcinogenesis* 39, 708–718. doi:10.1093/carcin/bgy040

Wang, Z., Zhang, C., and Qi, R. Z. (2014). A Newly Identified Myomegalin Isoform Functions in Golgi Microtubule Organization and ER-Golgi Transport. *J. Cell Sci.* 127, 4904–4917. doi:10.1242/jcs.155408

Wu, J., de Heus, C., Liu, Q., Bouchet, B. P., Noordstra, I., Jiang, K., et al. (2016). Molecular Pathway of Microtubule Organization at the Golgi Apparatus. *Dev. Cell.* 39, 44–60. doi:10.1016/j.devcel.2016.08.009

Yamauchi, M., Morikawa, T., Kuchiba, A., Imamura, Y., Qian, Z. R., Nishihara, R., et al. (2012). Assessment of Colorectal Cancer Molecular Features Along Bowel Subsites Challenges the Conception of Distinct Dichotomy of Proximal Versus Distal Colorectum. *Gut* 61, 847–854. doi:10.1136/gutjnl-2011-300865

Yang, C., Wu, J., de Heus, C., Grigoriev, I., Liv, N., Yao, Y., et al. (2017). EB1 and EB3 Regulate Microtubule Minus End Organization and Golgi Morphology. *J. Cell Biol.* 216, 3179–3198. doi:10.1083/jcb.201701024

Zhang, L., Zhao, Y., Dai, Y., Cheng, J. N., Gong, Z., Feng, Y., et al. (2018). Immune Landscape of Colorectal Cancer Tumor Microenvironment From Different Primary Tumor Location. *Front. Immunol.* 9, 1578. doi:10.3389/fimmu.2018.01578

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 Hu, Zhang, Huang, Gao, Yuan, Tang, Gao, Wang, Zhang, Ma, Qiao, Jin 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 06 August 2021
doi: 10.3389/fmolb.2021.705008

CUX1 Facilitates the Development of Oncogenic Properties *Via* Activating Wnt/β-Catenin Signaling Pathway in Glioma

Fan Feng[1,2,3], *Zongqing Zhao*[2,3], *Yunfei Zhou*[3], *Yanhao Cheng*[2,3], *Xiujie Wu*[2,3] *and Xueyuan Heng*[2,3]*

[1]*Institute of Clinical Medicine College, Guangzhou University of Chinese Medicine, Guangzhou, China,* [2]*Institute of Brain Science and Brain-Like Intelligence, Linyi People's Hospital, Linyi, China,* [3]*Department of Neurosurgery, Linyi People's Hospital, Linyi, China*

Background: Homeobox cut like 1 (CUX1), which often presents aberrated expression in many cancer cells, exerts a crucial role in tumorigenesis. Evidence describing CUX1 in gliomagenesis is scarce, and the effects of CUX1 on the Wnt/β-catenin pathway have not been reported. Our study aimed to explore the biological functions and molecular mechanisms involved in CUX1 activity in glioma.

Methods: Datasets for bioinformatics analysis were obtained from the GEO, TCGA, CGGA, GTEX and CCLE databases. qRT-PCR, western blotting (WB), and immunohistochemistry (IHC) assays were used to investigate the expression patterns of CUX1 among glioma and brain tissues. CUX1 knockdown and overexpression vectors were transfected into glioma cell lines, the CCK-8, clone formation assay, wound healing, Transwell assay, and flow cytometry were performed to detect changes in cell viability, invasiveness, and the cell cycle. WB and immunofluorescence (IF) assays were used to explore changes in cell cycle-related and Wnt/β-catenin signaling protein levels.

Results: Overexpression of CUX1 was identified in glioma tissues, and especially in glioblastoma (GBM), when compared to normal controls and correlated with poor prognosis. In comparison with untreated cells, TJ905 glioma cells overexpressing CUX1 showed higher proliferation and invasion abilities and S phase cell-cycle arrest, while the knockdown of CUX1 suppressed cell invasive ability and induced G1 phase arrest. Active Wnt/β-catenin signaling was enriched and clustered in a CUX1-associated GSEA/GSVA analysis. IF and WB assays indicated that CUX1 regulated the distribution of Axin2/β-catenin in glioma cells and regulated the expression of proteins downstream of the Wnt/β-catenin signaling pathway, suggesting that CUX1 served as an upstream positive regulator of the Wnt/β-catenin pathway. Finally, the knockdown of Axin2 or β-catenin could reverse the tumor-promoting effects caused by CUX1 overexpression, suggesting that CUX1 induced gliomagenesis and malignant phenotype by activating the Wnt/β-catenin signaling pathway.

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Rishi Kumar Jaiswal,
Loyola University Chicago, United States
Tabish H. Khan,
University of Virginia, United States

***Correspondence:**
Xueyuan Heng
hengxuey001@163.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *04 May 2021*
Accepted: *26 July 2021*
Published: *06 August 2021*

Citation:
Feng F, Zhao Z, Zhou Y, Cheng Y, Wu X and Heng X (2021) CUX1 Facilitates the Development of Oncogenic Properties Via Activating Wnt/β-Catenin Signaling Pathway in Glioma. Front. Mol. Biosci. 8:705008. doi: 10.3389/fmolb.2021.705008

Conclusion: Our data suggested that the transcription factor CUX1 could be a novel therapeutic target for glioma with gene therapy.

Keywords: CUX1, glioma, proliferation, cell cycle, wnt/β-catenin signaling, bioinformatics analysis

INTRODUCTION

Glioma is the most devastating primary intracranial tumor, characterized by diffuse invasion, induction of neuro degeneration, and resistance to various chemotherapeutic agents (Kang and Adamson, 2015; Babu et al., 2016). Despite the availability of multimodal treatments, the diagnosis and outcomes for glioma patients are still poor, with a median survival time of approximately 14.6 months (Li et al., 2019; Pinel et al., 2019). Hence, elucidating the molecular pathogenic process in gliomagenesis is vital for identify novel therapeutic targets.

The cut-like homeobox 1 (CUX1), also known as CULT1 or CDP, belongs to the family of homeodomain (HD) transcription factors (Liu et al., 1991; Sansregret and Nepveu, 2008; Hulea and Nepveu, 2012). CUX1 shares four similar DNA-binding domain architectures, including a CUT homeodomain and three CUT repeats (CR1, CR2, CR3), which are evolutionarily and functionally conserved from metazoans to humans (Ludlow et al., 1996; Vanden Heuvel et al., 2005; Kedinger et al., 2009). As a major member of the CUX family, it is involved in multiple biological processes including cell differentiation, proliferation, cell cycle regulation, tissue development, and double strand breaks (DSBs) repair response, besides, CUX1 expression was abnormally elevated in many malignant tumors and was implicated in tumorigenesis (Ramdzan and Nepveu, 2014; Wong et al., 2014; Jo et al., 2017). Many lines of evidence have indicated that overexpression of CUX1 was present in various types of cancers, such as melanoma, pancreatic cancer, multiple myeloma, and breast cancer (Fan et al., 2014; Vadnais et al., 2014; Krug et al., 2020). In our previous study, we also found that the CUX1 level was associated with tumor WHO grade and the malignant proliferation index in glioma, which provides additional evidence for the investigation of the CUX1-drived molecular mechanism in gliomagenesis (Wu et al., 2019).

The canonical Wnt/β-catenin signaling pathway was activated in gliomagenesis and involved in proliferation, apoptosis, cell invasiveness and angiogenesis (Wan et al., 2011; Nager et al., 2012). The upregulation of WNT3a and WNT1 in glioma stem cells (GSCs) has been reported in the progression of malignant transformation and gliomagenesis (Liu et al., 2011; Zhang et al., 2011). Besides, the expression of β-catenin, WNT3a and TCF4 were prevalent in astrocytic tumors and the nuclear accumulation of β-catenin was correlated with histological malignancy grade (Sareddy et al., 2009; Kaur et al., 2013). Griesmann et al. 2013 identified WNT5a as a crucial downstream effector of transcription factor CUX1 and CUX1/WNT5a/NFAT axis exerted crucial role in tumorigenesis and drug-resistance in pancreatic cancer (Griesmann et al., 2013). EMT is a complex biological process in which epithelial cells undergo varieties of biochemical changes and finally transformed into mesenchymal phenotype, which exerted a pivotal role in tumor progress, invasion and metastasis (Pastushenko and Blanpain, 2019). There was also evidence to show that the activation of CUX1/WNT signaling regulated epithelial-mesenchymal transition (EMT) in EBV infected epithelial cells (Malizia et al., 2009). Recently, a vitro study has also indicated that by cooperating with GLIS1, CUX1 promoted tumor cell migration and invasion by stimulating TCF/β-catenin transcriptional activity and epithelial-mesenchymal transition (EMT) in breast cancer (Vadnais et al., 2014). Yet, the biological function of CUX1 and it-induced potential molecular mechanism in gliomagenesis were still elusive.

The objective of this study was to investigate the effects of CUX1 expression on physiological functions and the regulatory mechanism of CUX1 gene in the Wnt/β-catenin pathway. Our data provide insight into the pathogenesis, diagnosis, and novel targets of glioma in gene therapy.

MATERIALS AND METHODS

Patients and Specimens

This study included 80 glioma tissue samples and 15 normal brain tissues collected between August 2015 and August 2019 at Linyi People's Hospital, Shandong, China (**Supplementary Tables S1, S2**). Patients who had undergone radiotherapy and chemotherapy were excluded from the study. All fresh glioma tumor tissues were immediately transferred to liquid nitrogen and stored at −80°C for subsequent western blotting (WB) and qRT-PCR assays. The remaining tissues were fixed in 4% paraformaldehyde (PFA) for immunohistochemical staining. Each patient signed informed consent to participate in the study. The use of patients' samples was approved by the Ethics Committee of Linyi People's Hospital and this study protocol was performed in accordance with the Declaration of Helsinki.

Reagents and Antibodies

Antibodies for WB, immunohistochemistry (IHC), and immunofluorescence (IF) included CUX1 rabbit mAb (No.11733-1-AP; WB 1:1500; IHC 1:150; IF 1:300), c-Myc rabbit mAb (No.24072-1-AP; WB 1:1000), β-catenin mouse mAb (No.66379-1-Ig; WB 1:4000; IF 1:200), MMP-2 rabbit mAb (No.10373-2-AP; WB 1:3000), MMP-7 rabbit mAb (No.10374-2-AP; WB 1:2000), MMP-9 rabbit mAb (No.27306-2-AP; WB 1:2000), RHOA mouse mAb (No.66733-1-Ig; WB 1:3000), RHOB rabbit mAb (No.14326-1-AP; WB 1:2500), RHOC rabbit mAb (No.10632-1-AP; WB 1:1500), ROCK1 rabbit mAb (No.21850-1- AP; WB 1:2000), CDK2 rabbit mAb (No.10122-1-AP; WB 1:800), CDK4 mAb (No.11026-1-AP; WB 1:3000), CDK6 mAb (No.14052-1-AP; WB 1:1500), CyclinD1 rabbit

mAb (No.26939-1-AP; WB 1:1000), CyclinE1 rabbit mAb (No.11554-1-AP; WB 1:800), GAPDH mouse mAb (No.60004-1-Ig; WB 1:3000); all antibodies were purchased from Proteintech (Wuhan, China). Antibodies against Axin2 (No. ET1703-96; WB 1:3000; IF 1:300) were obtained from Huabio Technology (Hangzhou, China). Dulbecco's Modified Eagle's Medium (DMEM) and fetal bovine serum (FBS) were purchased from Gibco (Carsbad, CA, United States). The cell cycle kit and CCK-8 kit were obtained from Beyotime (Jiangsu, China).

Vector Construction and Transfection Assays

To overexpress CUX1, the coding sequence (NM_181552.4) was obtained from the NCBI and was synthesized to facilitate cloning. The synthesized sequence was ligated and subcloned into pBABE lentivirus vector and an empty vector was used as the control group. Three siRNA sequences targeting CUX1 were used: shRNA#1: forward 5′-GCAUAAGCUCAGUCUGAAATT-3′, reverse 3′-UUUCAGACUGAGUUAUGCTT-5'; shRNA#2: 5′-GCAAGGAGCCAUUUCACAATT-3′, reverse 3′-UUGUGA AAUGGCUCCUUGCTT-5'. Lentiviral particles were purchased from Genechem. Co., Ltd. The siRNAs and lentiviral particles were transfected into cells using Lipofectamine 3000 (Invitrogen, United States) according to the manufacturer's protocol. Puromycin was used to screen the stable cells, which were collected for subsequent experiments. As a positive control for inhibition of Wnt/β-catenin signal transduction by siRNA, we used siRNA targeting Axin2 with the sequence (forward 5′-GCATAGATT GTTACTGCTA-3′, reverse 3′-GCACAGATTATTACTGCTA-5′) and siRNA targeting CTNNB1 with the sequence (5′-CCU UCACUCAAGAACAAGUTT-3′, reverse 3′-TTGGAAGUG AGUUCUUCUUCA-5′).

Cell Culture

The cell lines HA 1800, U251, A172, SF295, CRT, TJ905, and PT2 were cultured in DMEM containing penicillin/streptomycin and 10% FBS in a 5% CO_2 incubator at 37°C. The cells were sub-cultured when cultures reached a confluence of 70–80%; DMEM was replaced every 2–3 days. All glioma cell lines and Normal human astrocyte cells (NHAs) were purchased from Cell Bank, Shanghai Academy of Science, Shanghai, China.

RNA Extraction and Real-time Quantitative Reverse Transcription PCR

RNA was harvested using the TRIzol reagent (Invitrogen), according to the manufacturer's guidelines. The RNA was transcribed into cDNA by miscriptreverse transcription kit (Takara, Japan). The level of mRNA was quantified by q-PCR with the QuantiTect SYBR Green PCR kit (Takara, Japan). The reaction conditions were 95°C for 10 min, followed by 95°C for 15 s for 40 cycles, and 60°C for 60 s. Primers used for PCR were as follows: CUX1-forward: 5′-AGCCGAAACCATAGCTCTTGA-3′; CUX1-reverse: 5′-GCCCTTTCGAGGTCCGTCAT-3′; GAPDH-forward: 5′- TGCACCACCAACTGCTTAGC-3′; GAPDH-reverse: 5′-GGCA TGCACTGTGGTCATGAG-3′. The mRNA level of target genes was compared to GAPDH by qPCR using the comparative cycle threshold ($2^{-\Delta\Delta CT}$) method. All assays were independently performed in triplicate.

Western Blotting Analysis

Expression of target genes was determined by WB as previously described with the following modifications (17). Briefly, samples were lysed by radioimmunoprecipitation assay (RIPA) buffer containing a phosphatase and protease inhibitor cocktail (Beyotime, Jiangsu, China). The supernatants were examined for protein content using the BCA method. Equal amounts of protein were separated by SDS-PAGE gels, then transferred onto PVDF membranes (Millipore, Billerica, MA, United States). After blocking for 1–3 h, the membranes were probed with proper primary and secondary antibodies. Bands were visualized by enhanced chemiluminescence and captured with ultrasensitive chemiluminescence imaging system. Protein bands were quantified by using the ImageJ software version 1.8.0 (NIH, US) and all experiments were independently performed in triplicate.

Immunofluorescence Analysis and Immunohistochemical Staining

For IF assays, cells were grown on coverslips and were fixed in 4% PFA on ice for 20 min. After washing twice with phosphate-buffered saline (PBS), cells were permeabilized with TritonX-100 (Sigma) for 15 min. Cells were blocked with 3% goat serum and 2% bovine serum albumin (BSA), then incubated with β-catenin antibody and Axin2 antibody at 4°C overnight. Subsequently, coverslips were incubated with TRITC-conjugated anti-rabbit secondary antibody after washing twice with PBS at room temperature for 40 min. Then, coverslips were washed again with PBS and incubated with DAPI at 37°C for 30 min. Fluorescence was monitored by Olympus inverted confocal microscope. The semi-quantitative fluorescence intensity analysis of different cell compartments was assessed by ImageJ software. Immunohistochemistry assay was performed with avidin-biotin immunoperoxidase technique as described previously (Wu et al., 2019). CUX1 antibody was used at a dilution of 1:3000. Staining results were scored according to the following standards: The stained sample was assigned a combined IHC score (0–6), which was obtained by adding the intensity score (0–3) and the percentage score (0–100%). Each combined score is shown (0 [negative], score 2 [++, <10%], score 3 [+, >50%], score 4 [++, <50%], score 5 [++, >50%], score 6 [+++,100%]). We defined samples with scores <4 as the low CUX1 expression group and scores ≥4 as the high CUX1 expression group.

Cell Proliferation Assays

For the colony formation assay, transfected cells were seeded in 6-well plates (500 cells/well) and incubated with 10% FBS for 2 weeks at 37°C for 2 weeks, then fixed with 4% PFA. After washing with PBS, cells were strained with 0.1% crystal violet (Beyotime) for 30 min. The colony forming efficiency was

monitored under a microscope. This assay was performed in triplicate. Cell viability was also assessed using the cell counting kit 8 (CCK-8, Beyotime) method at 0, 24, 48, and 72 h. According to the manufacturer's instructions, cells were seeded in 96-well plate (2000 cells/well). Each well contained 100 μl DMEM and 10 μl CCK-8. The culture plate was incubated at room temperature for 3 h. Absorbance of each well at 450 nm was measured using the Synergy Microplate Reader (Biotek). All assays were independently performed in triplicate.

Flow Cytometry Assay

For cell cycle analysis, glioma cells were harvested at 48 h post-transfection by trypsinization and fixed with pre-cooled 70% ethanol at −20 °C overnight. After being centrifuged at 1000 × g for 5 min, the cells were stained with 400 μl propidium iodide (PI) and 100 μl RNase A (Beyotime, Shanghai, China) at 37°C for 30 min. A total of 10,000 cells per sample were analyzed by flow cytometry (BD, Bioscience, CA, United States), and the cell-cycle populations were determined by ModFit software. All assays were independently performed in triplicate.

Transwell Assay and Wound Healing Assay

For the assessment of cell invasion, approximately 5×10^4 cells in 100 μl DMEM medium without FBS were placed in the upper Matrigel chamber (BD, Bioscience, United States). A volume of 500 μl DMEM with 10% FBS was added to the lower chamber. Next, 10 μg/ml mitomycin was added to the DMEM to repress cell proliferation to greatest extent. After incubation for 24 h at 37°C in 5% CO_2 atmosphere, the cells on the surface of the insert were removed using a cotton swab. Then, glioma cells that passed through the membrane were stained with 1% crystal violet and counted under a microscope.

For the wound healing assay, 5×10^5 transfected glioma cells were cultured in 6-well plates and incubated until the cells grew to 90% confluence. The confluence plates were scratched with a P-10 pipette tip, washed with PBS to remove the detached cells and cultured in serum-free medium supplemented with 10 μg/ml mitomycin. The wound-closing procedure was observed under a microscope at 0, 24, and 48 h, respectively. All assays were independently performed in triplicate.

Gene Expression Profile Data Acquisition

Microarray RNA-seq data of glioma patients were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE16011, GSE51006, GSE67089, GSE45921, GSE7696, and GSE4290. The GPL8542 platform (Affymetrix Gene Chip Human Genome U133 Plus 2.0 Array) was used. The clinical data and related RNA-seq data from patients with glioma were obtained from TCGA database (https://xenabrowser.net/datapages/). Data for co-expression analysis were obtained from the CGGA (http://www.cgga.org.cn/), GTEx and CCLE database (https://xenabrowser.net/datapages/).

Bioinformatics Analysis

Gene set enrichment analysis (GSEA) was performed to identify the essential biological pathways that were significantly enriched in the $CUX1^{high}$ group. The enrichment status estimates of CUX1-associated signaling pathways were also obtained using the gene set variation analysis (GSVA) package. Other R packages, "ggpubr", "ggplot2", "limma", "dplyr", and "tidyr" were also applied to visualize the results of bioinformatics analysis.

Statistical Analysis

Statistical analysis was performed using Origin2018 and Sigmaplot version 14.0 for Windows. All data were obtained from three independent assays and presented in the form of Mean ± SD. The Student's *t*-test was used to analyze differences between two groups. Comparisons between more than three groups were determined using one-way ANOVA analysis of variance followed by the Turkey post-hoc test. The χ^2test examined the relationship between CUX1 expression and clinicopathological characteristics. Kaplan-Meier survival analysis and the log-rank test were used to investigate the overall survival (OS). A p-value <0.05 was considered statistically significant.

RESULTS

Expression of Homeobox Cut Like 1 in Glioma Was Associated With Poor Prognosis

We initially analyzed the mRNA levels of CUX1 in several types of intracranial tumors and found that the expression of CUX1 mRNA was higher in glioma (GSE50161, **Figure 1A**). To identify the CUX1 expression pattern in clinical glioma, we analyzed the RNA-seq data using the microarray GSE7696, GSE4290, GSE45921, and GSE16011 from GEO database. The statistical results of the four datasets indicated that the level of CUX1mRNA was significantly upregulated in glioma compared to normal brain tissues (**Figures 1B–E**). The astrocyte and glioma cells were differentiated from neuro stem cells (NSCs) and glioma stem cells (GSCs), respectively. NSCs, as the most active cells in central nervous system (CNS), are in a state of continuous proliferation and division and are prone to mutation. There are growing evidence to support the possibility that GSCs are derived from the accumulation of mutations in NSCs and astrocyte progentiors (Guelfi et al., 2016). Zong et al. 2015 also reported that both astrocyte/oligodendrocytes and their progenitors (NSCs) could serve as the original cells of glioma (Zong et al., 2015). In the microarray GSE67089, CUX1mRNA was significantly overexpressed in GSCs and GBM cells compared with NHA and NSCs, which suggested that CUX1 may exert a crucial role in gliomagenesis (**Figure 1F**). Consistent with the findings in the RNA-seq datasets, the results of qRT-PCR and WB assays showed that CUX1 mRNA and proteins were significantly increased in glioma tissues compared to normal brain tissues (**Figures 1G–I**). Moreover, the expression of CUX1 mRNA and protein in the normal human astrocyte cell line (HA 1800) and several glioma cell lines (U251, TJ905, PT2, A172 and

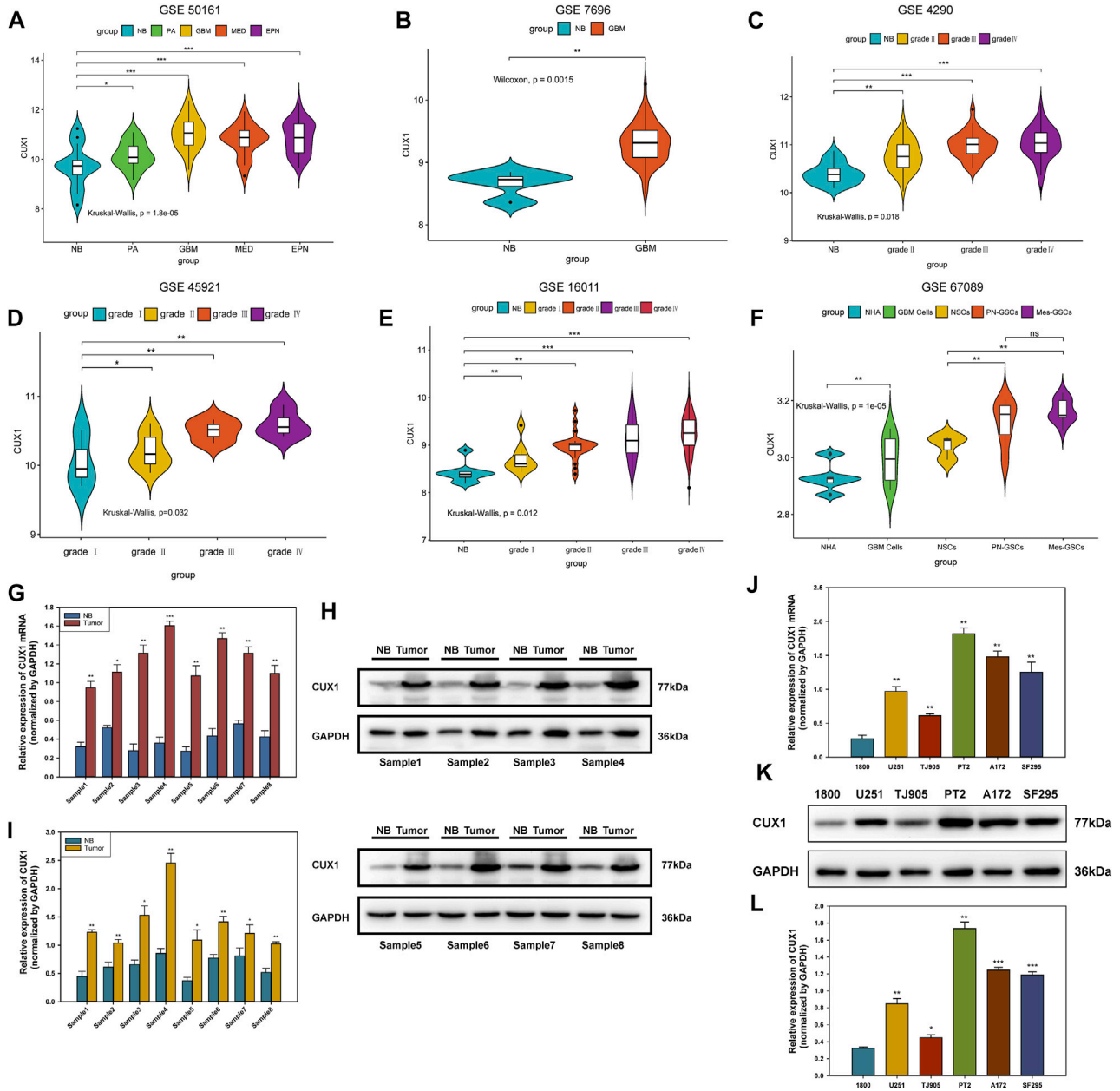


FIGURE 1 | CUX1 expression was upregulated in glioma. **(A)** Association between the expression of CUX1mRNA with various subtypes of intracranical tumors available on the GEO database (GSE501161). **(B–E)** Correlation between the CUX1mRNA level and WHO grades of glioma. Increased expression of CUX1mRNA was detected with pathological grade rising, based on analysis of RNA-seq data from GSE7696, GSE4290, GSE45921, and GSE16011. **(F)** Expression patterns of CUX1mRNA in normal human astrocyte (NHA) and glioblastoma (GBM) cells as well as their progenitor cells (NSCs and GSCs). **(G)** The expression levels of CUX1mRNA in fresh glioma tissues and adjacent non-tumor tissues were analyzed by RT-qPCR. **(H,I)** Western blotting analysis showing the levels of CUX1 protein and mRNA in fresh glioma tissues and adjacent noncancerous tissues among eight groups of samples. **(J–L)** qRT-PCR and Western blot analysis showed the expression of CUX1 mRNA and protein in normal astrocyte (HA 1800) and five glioma cell lines (U251, TJ905, PT2, A172 and SF295). (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

SF295) were detected by qRT-PCR and WB assays. The results showed that CUX1 expression was higher in glioma cell lines (especially PT2 and A172) compared to HA1800 cells ($p < 0.05$, **Figures 1J–L**). In addition, we reviewed the clinical characteristics of 80 patients with glioma. The IHC assay revealed that CUX1 was mainly located in the cytoplasm/cytomembrane of glioma cells and CUX1 levels were upregulated with increasing World Health Organization (WHO) grade in the glioma samples. (**Figures 2A–C**, **Supplementary Tables S1, S2**). To evaluate the prognostic

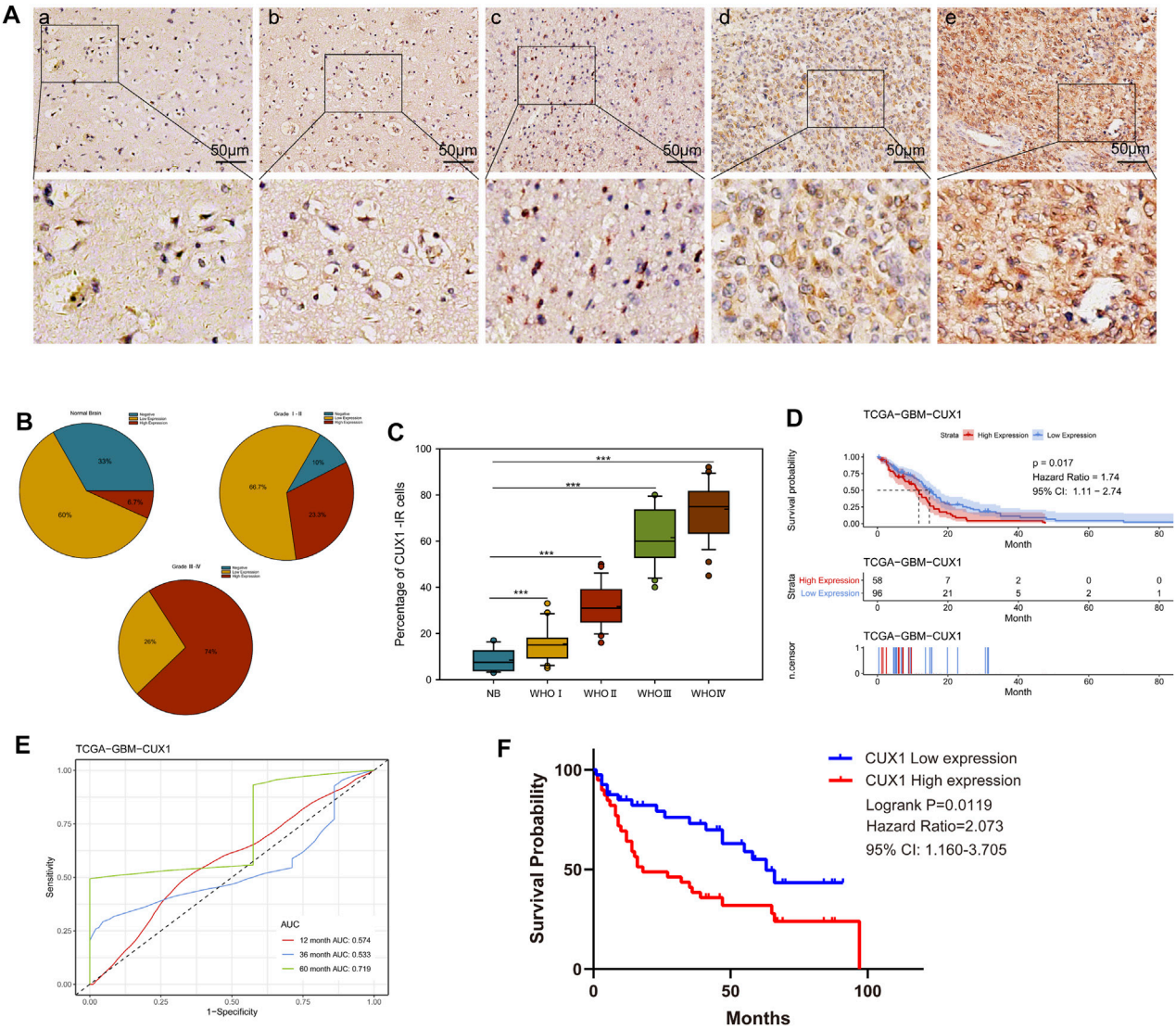


FIGURE 2 | High CUX1 expression predicted poor prognosis in patients with glioma. **(A–C)** The expression levels of CUX1 in 80 glioma tissues and 15 normal brain tissues, measured by immunohistochemistry (IHC). **(A)** Representative images for CUX1 staining in nontumor **(a)** brain samples and WHO grade I-IV glioma tissues **(b–e)**. IHC-stained sections, the scale bar corresponds to 50 μm. **(B)** Statistical analysis of CUX1 expression in normal brain tissue and different clinical grade glioma specimens. **(C)** Quantitative analysis of cell counts showing that the percentage of CUX1-immunoreactivity (IR) cells was significantly increased in glioma tissues compared with nontumor brain tissues. **(D)** Kaplan Meier (KM) curve comparing overall of glioblastoma patients according to expression of CUX1 in TCGA database. Log-rank test. **(E)** Receiver operator characteristic curve analysis of CUX1. AUC, Area under the curve. **(F)** KM-plot of CUX1 expression in patients with glioma based on clinical information data from our study. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

value of CUX1 in glioblastoma patients, the clinical information of 577 patients was obtained from the TCGA RNA-seq database. After dividing the cases into two equal strata according to CUX1 expression, Kaplan-Meier survival analysis revealed that CUX1 overexpression was correlated with shorter OS compared to patients with negative or low CUX1 expression, which was also in line with the results of the survival analysis in our study (**Figures 2D,F**). Meanwhile, in the Receiver operator characteristic (ROC) curve analysis, the area under curve (AUC) for CUX1 expression in predicting the 60 months survival was 71.9% (**Figure 2E**). The findings mentioned above revealed that CUX1 served as a negative prognostic factor in patients with glioma.

Effects of Knockdown and Overexpression of Homeobox Cut Like 1 on Cell Proliferation and Cell Cycle Progression in Glioma Cells

The gain-of-function and loss-of-function assays were performed to investigate the role of CUX1 in PT2 and TJ905 cell lines. Transfected cells were detected for CUX1 at the protein by WB analysis. The results demonstrated that the expression of CUX1

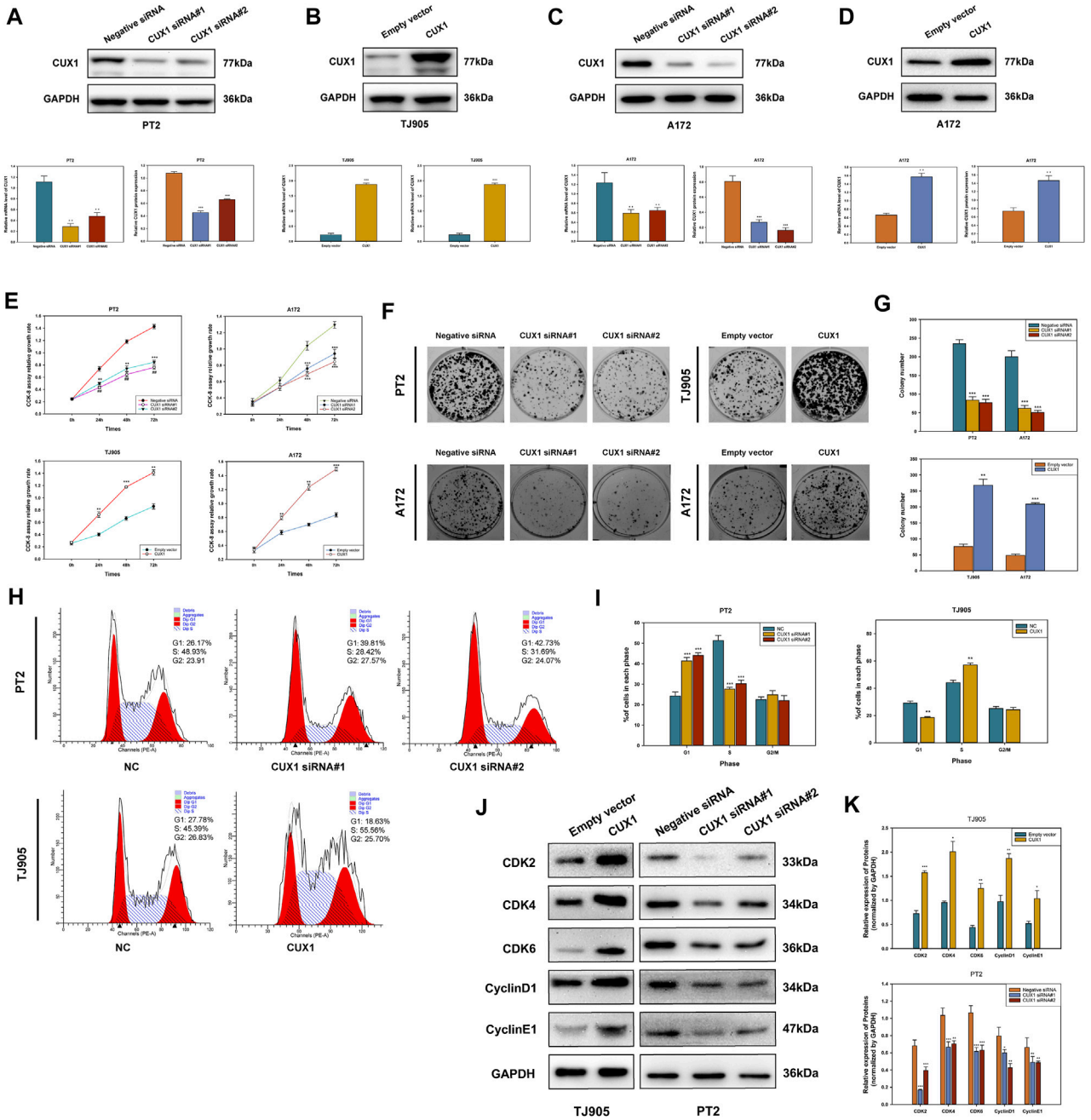


FIGURE 3 | Effects of knockdown and overexpression of CUX1 on cell proliferation and cell cycle progression in glioma cells. **(A)** Transfected with CUX1 siRNA#1 and siRNA#2 decreased the CUX1 protein and mRNA level compared to negative siRNA in PT2 cells. **(B)** Transfected with CUX1 plasmid upregulated the expression of CUX1 protein compared to empty vector in TJ905 cells. **(C,D)** The interference effects of TROAP siRNA and overexpression plasmid in A172 cells. **(E)** Proliferation in PT2, TJ905 and A172 cells was detected by CCK8 assay. **(F,G)** The effect of CUX1 on PT2, TJ905 and A172 cells proliferation was measured by the colony formation assay. **(H,I)** Flow cytometry presented CUX1 induced cell cycle arrest in PT2 and TJ905 cell lines. **(J,K)** Western blot showed the expression of relevant cell cycle proteins in PT2 and TJ905 cells. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

was significantly lower following transfection with two CUX1-siRNA vectors in PT2 cells, and specifically following transfection with CUX1-siRNA#1 ($p < 0.05$, **Figure 3A**). Conversely, transfection with the CUX1 plasmid markedly increased the level of CUX1 protein and mRNA compared to transfection with empty vector in TJ905 cells ($p < 0.05$, **Figure 3B**). Different glioma cell lines exhibit different phenotypes and properties. To eliminate potential interference by different cell

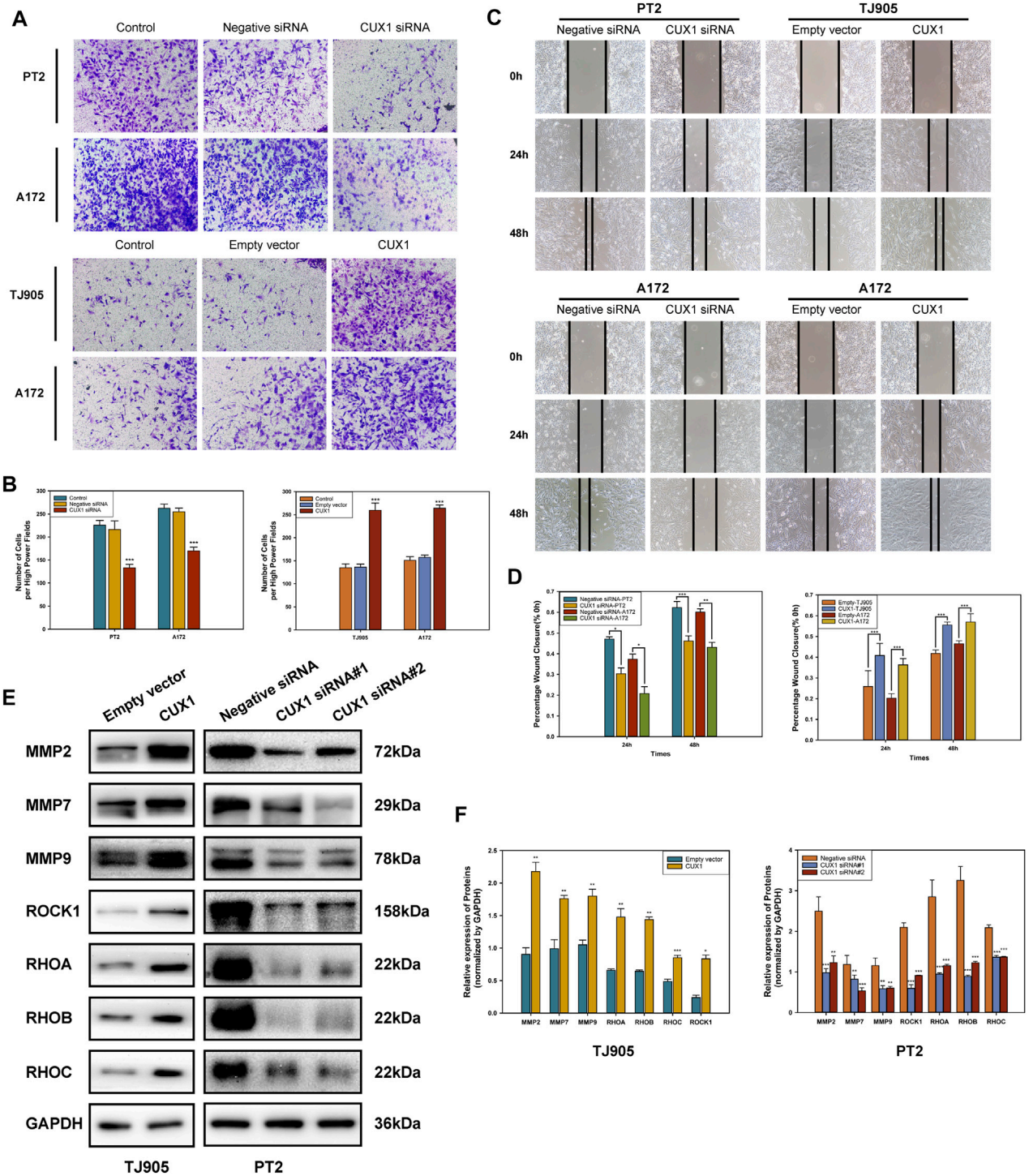


FIGURE 4 | Effects of overexpression and knockdown of CUX1 on the migration and invasion of glioma cells. **(A,B)** Transwell assay showed the invasion of TJ905, PT2 and A172 cells after transfection with CUX1 siRNA or CUX1 overexpression plasmid (magnification ×200). **(C,D)** Wound healing assay was used to detect the cell migration ability after overexpression or knockdown of CUX1 after 24 and 48 h in TJ905, PT2 and A172 cells (magnification ×200). **(E,F)** Western blot analysis showed the change of migration-related proteins in PT2 and TJ905 cells.

line phenotypes, we also selected the A172 cell line for knockdown and overexpression transfection assays, followed by a series of cellular function assays ($p < 0.05$, **Figures 3C,D**). The effects of CUX1 on glioma cell proliferation were examined by preforming CCK-8 and colony formation assays. The results indicated that CUX1 overexpression significantly

enhanced viability and colony formation in TJ905-trasfected cells compared to control cells, while CUX1 knockdown significantly decreased proliferation of PT2 cells. These observations were also consistent with the results of parallel CCK8 and colony formation assays in the A172 cell line ($p < 0.05$, **Figures 3E–G**). Next, the role of CUX1 in cell cycle progression was further explored *via* flow cytometry analysis. The findings showed that there were more TJ905 cells in the S phase and less in the G1 phase following transfection with the CUX1 overexpression vector; whereas, transfection of CUX1-siRNA resulted in a higher proportion of PT2 cells in the G1 stage ($p < 0.05$, **Figures 3H,I**). Similarly, the effects of CUX1 expression on cell cycle proteins were also detected by WB analysis. A higher expressions of Cyclin D1, Cyclin E1, CDK2, CDK4 and CDK6 were observed in TJ905 cells presenting CUX1 overexpression compared to controls, and conversely, these proteins were all down-regulated following CUX1-siRNA treatment ($p < 0.05$, **Figures 3J,K**).

Homeobox Cut Like 1 Promoted Invasion and Migration in Glioma Cells

The roles of CUX1 in regulating cell invasion and migration ability were investigated using Transwell invasion and wound healing assays. Our findings indicated that the invasive and migrative abilities of TJ905 cells had increased following transfection with CUX1 overexpression plasmid, and this ability was inhibited following transfection with CUX1-siRNA. Similarly, A172 glioma cells also exhibited enhanced migration and invasion capability when treated with the CUX1 overexpression plasmid compared to the empty vector, while attenuated migration was observed following transfection with CUX1 siRNA ($p < 0.05$, **Figures 4A–D**). WB analysis was then performed to assess the effects of CUX1 overexpression on migration-associated genes. Knockdown of CUX1 significantly down-regulated the expression of MMP2, MMP7, MMP9, ROCK1, RHOA, RHOB, and RHOC protein in PT2 cells compared to control, while the respective gene levels were dramatically elevated in TJ905 cells (**Figures 4E,F**). These data suggested that CUX1 promoted glioma infiltration and spread by regulating the expression of migration-related proteins.

The Downregulation of Homeobox Cut Like 1 Reversed the Epithelial-Mesenchymal Transition

The epithelial-mesenchymal transition (EMT) is a complex process that enable epithelial cells lose cell adhesiveness/polarity and then transform into the mesenchymal cells, which have an increased invasive and metastatic ability and gain the malignant phenotype. Moreover, EMT is further correlated with increased activities of matrix metalloproteins (MMPs), which facilities cell invasion *via* degrading extracellular matrix proteins. Hence, we next used WB analysis to measure the effects of CUX1 downregulation/upregulation on the activities of EMT markers in glioma cell lines. The results of WB assay showed that levels of mesenchymal markers (N-cadherin, Snail and Vimentin) decreased in PT2 and A172 cells following CUX1-siRNA transfection, while, the expression of epithelial adhesion molecular E-cadherin was upregulated ($p < 0.05$). After transfecting with CUX1 overexpression vector in TJ905 and A172cells, WB assay demonstrated a decline of E-cadherin protein levels, but the expression levels of N-cadherin, snail and vimentin were obviously upregulated ($p < 0.05$, **Figures 5A–D**). Immunofluorescence staining assays provided further confirmation that the effect of CUX1 on the epithelial-mesenchymal transition in PT2, TJ905 and A172 glioma cell lines ($p < 0.05$, **Figures 5E,F**). Taken together, these results suggested that the CUX1 overexpression promoted the epithelial-mesenchymal transition, while, downregulation of CUX1 could reverse the malignant phenotype of glioma.

Homeobox Cut Like 1 Promoted the Malignant Phenotype by Activating Wnt/β-Catenin Signaling

Based on GSEA enrichment analysis, we identified 30 CUX1-associated significantly enriched pathways (adj.*p*.value < 0.05), including 11 activated pathways (normalized enrichment score [NES] >0) and 19 inactivated pathways (NES<0). The results, as shown in **Figures 6A–C**, indicated that the Wnt/β-Catenin pathway was activated by CUX1, which was also confirmed by GSVA analysis ($p < 0.05$, **Figure 6D**). Moreover, the transcription factor CUX1, by targeting key subunits of the Wnt/β-Catenin pathway, showed a positive correlation with mRNA expression of Axin2, CTNNB1, and TCF4 in most normal and cancer tissues or cell lines (**Supplementary Figure S1A–E**). The interaction network of CUX1 protein was analyzed using the GeneMANIA database, which suggested that CUX1 and several key regulators of the Wnt/β-catenin signaling were co-expressed (**Supplementary Figure S1F–G**). To confirm the bioinformatics results, a series of functional assays were performed to explore the potential molecular mechanism involved in CUX1-induced gliomagenesis. Axin2 and β-catenin (CTNNB1) serve as crucial regulatory factors involved in Wnt/β-catenin signaling. The IF assays revealed that CUX1 overexpression significantly induced substantial accumulation of Axin2 in cytoplasm and β-catenin in the nucleus and cytoplasm of TJ905 cells compared to controls ($p < 0.05$, **Figures 6E,F**). In addition, the WB analysis showed that when compared with control group, cells transfected with the CUX1 overexpression vector significantly increased the expression of downstream genes involved in Wnt/β-catenin signaling, such as CyclinD1, C-Myc, MMP7, and TCF4 ($p < 0.05$, **Figures 7A,B**). While the expression of the respective proteins and the CUX1-induced tumor-promoting effects (tumor cell proliferation and invasion) were significantly reversed by silencing Axin2 or β-Catenin expression ($p < 0.05$, **Figures 7C–G**). Importantly, Western blot assay also found the expression levels of EMT markers (N-cadherin, E-cadherin, Vimentin and Snail) were reversed after inhibiting Wnt/β-catenin pathway (knocking-down Axin2 or β-catenin) in TJ905 cells ectopically overexpressing CUX1 ($p < 0.05$, **Figures 7H,I**). Taken together, our findings indicated that CUX1 induced a

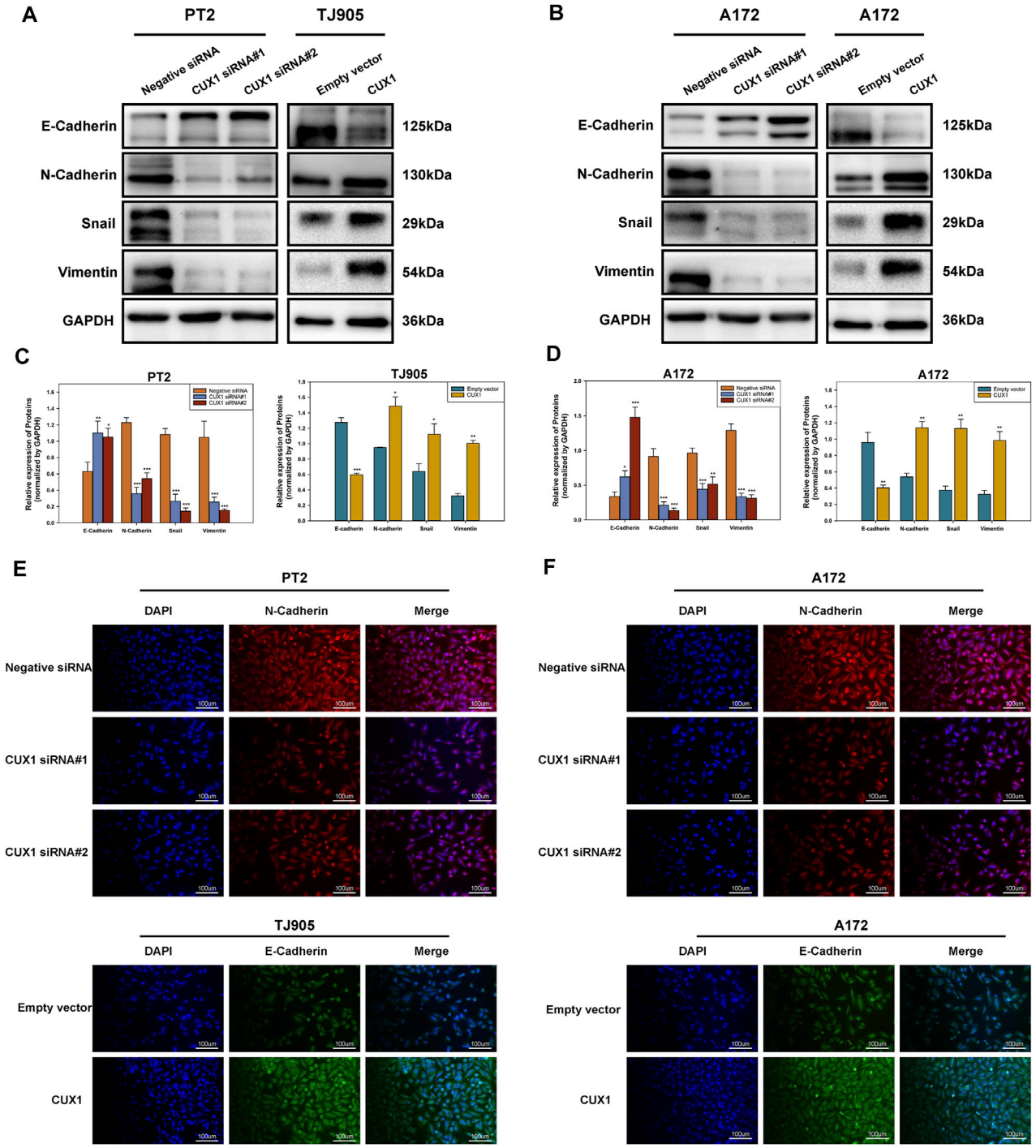


FIGURE 5 | The downregulation of CUX1 reversed the epithelial-mesenchymal transition (EMT). **(A–D)** Western blotting showing the decreased protein levels of mesenchymal marker (N-cadherin, vimentin and snail) and the increased protein levels of epithelial marker (E-cadherin) after knockdown of CUX1 in PT2 cells, while conversed phenomenon was presented in TJ905 following CUX1 overexpression treatment. The same effect was observed after treatment with CUX1-siRNA/CUX1 plasmid in A172 cell line. Representative images were presented in **A** and **B**, and quantitative analysis were shown in **C** and **D**. **(E,F)** Immunofluorescence staining showing the same outcomes as for CUX1 in PT2, TJ905 and A172 cells after siRNA treatment and overexpression plasmid treatment. The scalebar corresponds to 100 μm. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

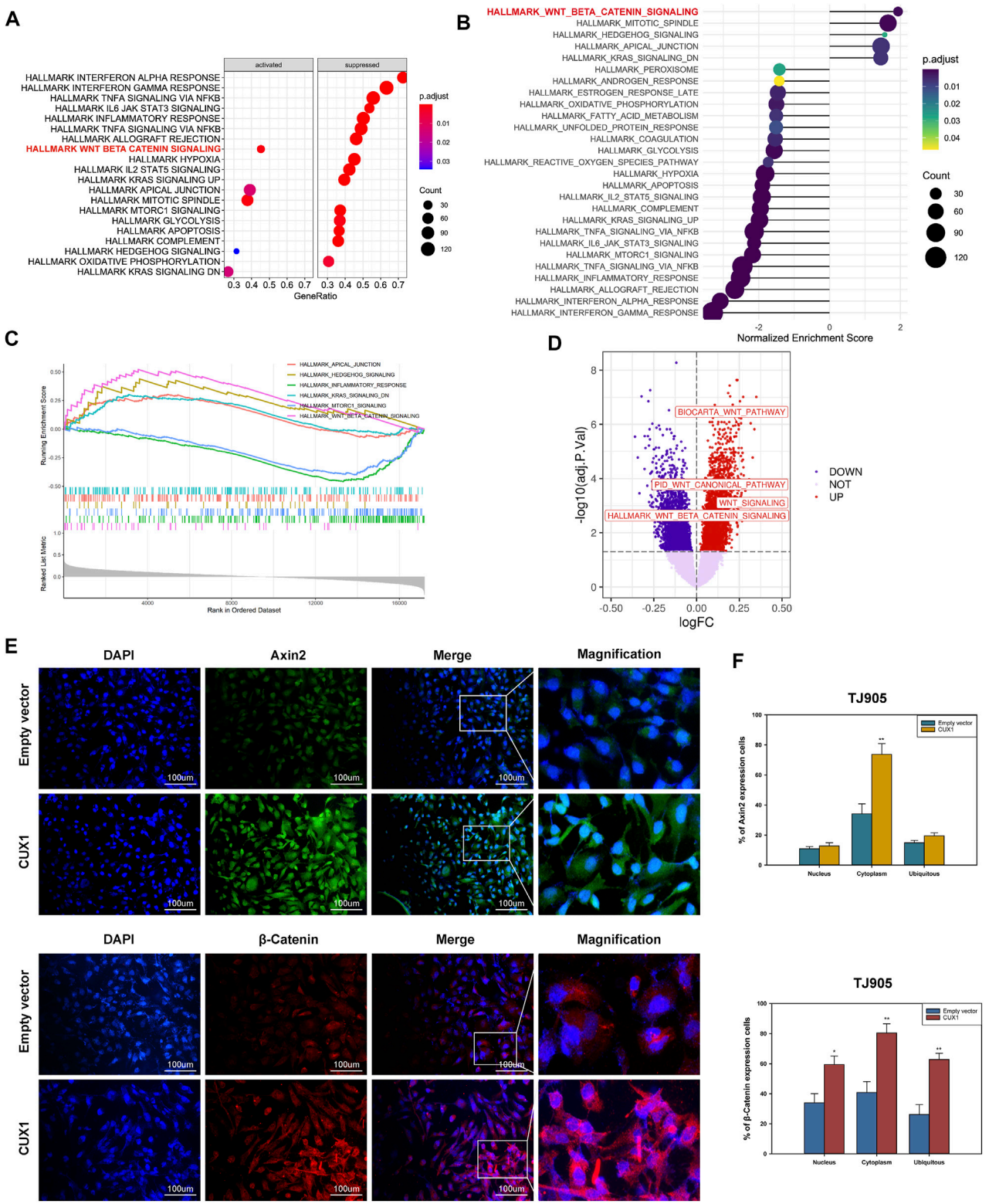


FIGURE 6 | Overexpression of CUX1 enhanced the activity of Wnt/β catenin signaling pathway in glioma cells. **(A–C)** Gene set enrichment analysis (GSEA) pathways enriched by CUX1. **(A)** The dotplot of partially enriched pathway. **(B)** The dotplot of the up- and down regulated pathways. **(C)** The up-regulated pathway of Wnt/β-catenin. **(D)** Wnt/β-catenin pathway was enriched and upregulated in glioma by gene set variation analysis (GSVA) analysis. **(E,F)** Immunoflurescence showed that overexpression of CUX1 increased expression of Axin2 proteins in cytoplasm of TJ905 cells compared to empty group. Meanwhile, CUX1 overexpression also increased the expression of β-catenin in nucleus and cytoplasm of TJ905 cells compared to control.

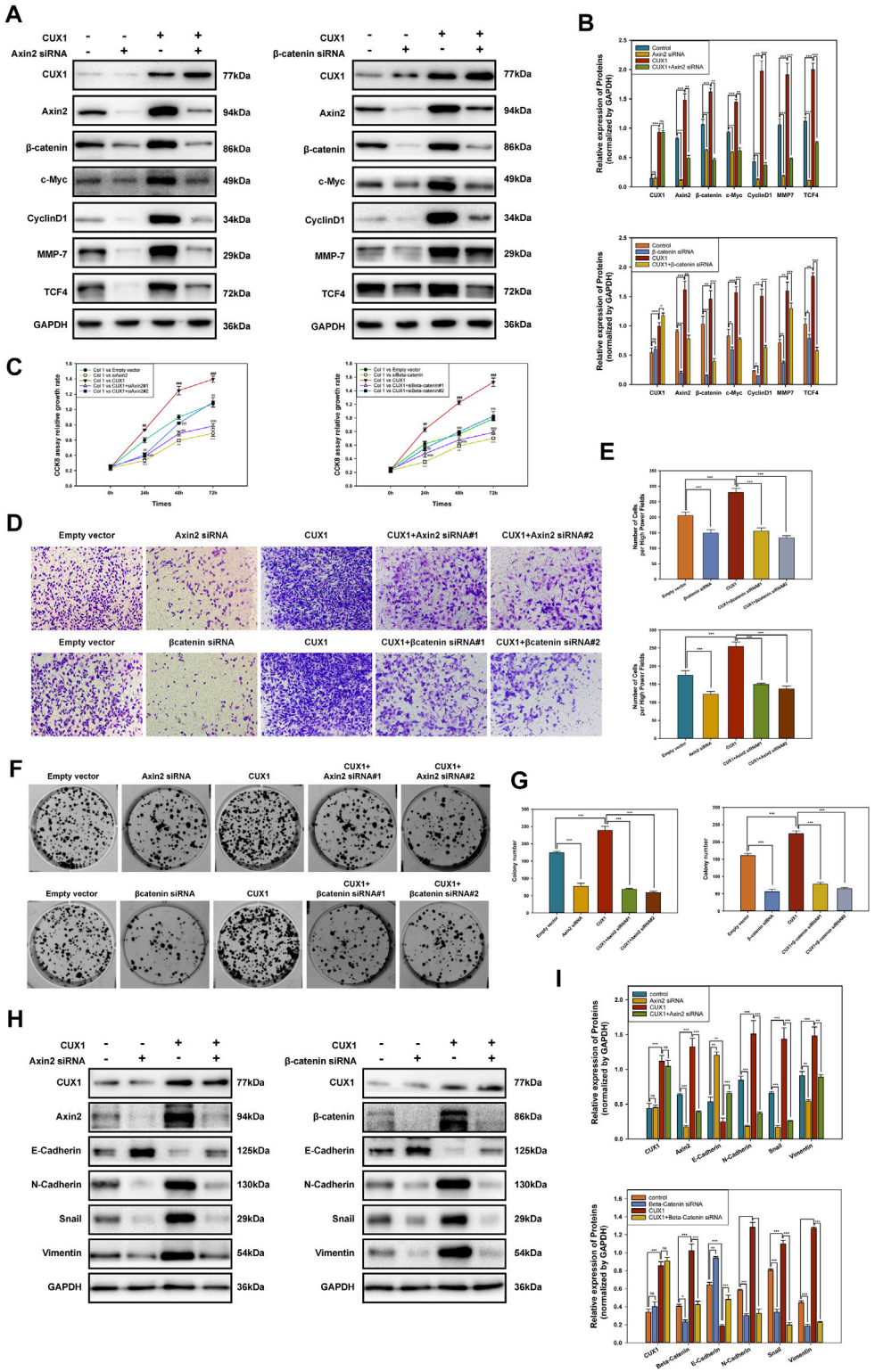


FIGURE 7 | Inhibiting Axin2 or β-catenin could partially reverse the malignant phenotype caused by CUX1 overexpression in gliomagenesis. **(A,B)** Western blot analysis presented the levels of subunits in Wnt/β-catenin signaling pathway after knocking-down Axin2 or β-catenin in TJ905 cells ectopically overexpressing CUX1. **(C)** CCK8 assay was used to detect the change of proliferation capability after inhibiting Wnt/β-catenin in TJ905 cell lines. *, compared with the empty vector group; #, compared with the empty vector group; α, compared with the empty vector group; γ, compared with the empty vector group. **(D,E)** Transwell assay to evaluate the reversal effect of silencing Wnt/β-catenin pathway on the migration activity in TJ905 cell lines. **(F,G)** The change of proliferation capability after inhibiting Wnt/β-catenin was detected by colony formation assays in TJ905 cell lines. **(H,I)** Western blot analysis presented the levels of EMT markers (N-cadherin, E-cadherin, Vimentin and Snail) after inhibiting Wnt/β-catenin pathway (knocking-down Axin2 or β-catenin) in TJ905 cells ectopically overexpressing CUX1.

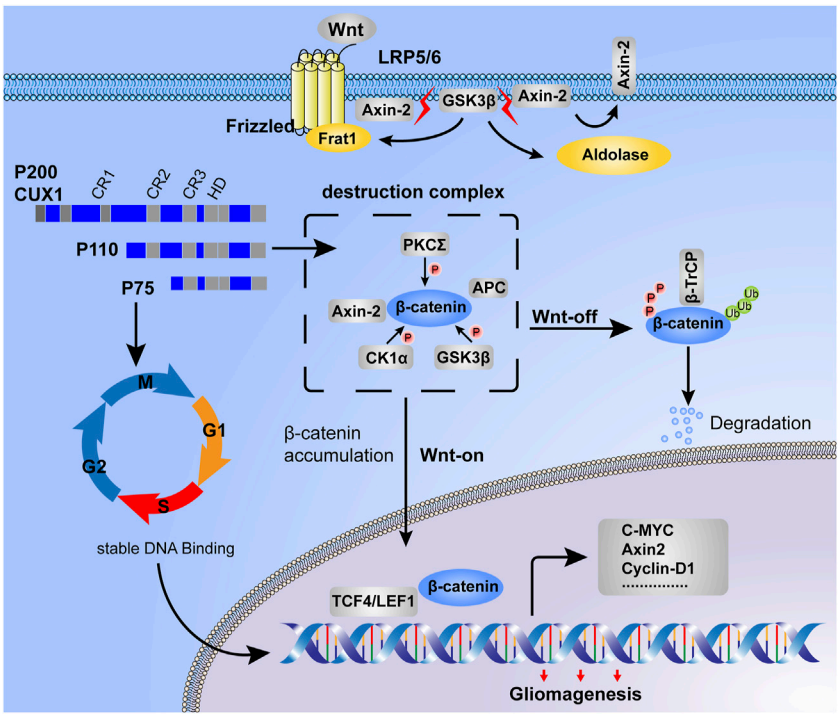


FIGURE 8 | Working model for CUX1-induced malignant phenotype in gliomagenesis. Destruction complex was the essential step for β-catenin phosphorylation, meanwhile, the phosphorylated status of β-catenin was determined for its stability. In the Wnt-off state, β-catenin was recruited by the destruction complex and induced degradation in proteasome. However, in the state of Wnt-on, due to the disassembly of the destruction complex, β-catenin was accumulated in cytoplasm and then translocated into the nucleus. On the one hand, DNA-binding dynamics of CUX1 showed a cell cycle-dependent manner, our study found that glioma cells with CUX1 overexpression represented the accelerated entry into S phase and cell proliferation. On the other hand, we also found that transcription factor CUX1 could upregulate the expression of Axin2/β-catenin and activated Wnt/β-catenin signaling pathway to promoted the transcription of downstreamed cell motility-related genes (such as CyclinD1, C-Myc and MMP7) and progress of epithelial-mesenchymal transition (EMT), inducing the malignant phenotype during gliomagenesis.

malignant phenotype and tumorigenesis, at least in part *via* activation of the Wnt/β-Catenin signaling pathway (**Figure 8**).

DISCUSSION

Gliomas, a CNS tumor, comprise approximately 80% of all primary malignant brain tumors (Shaver et al., 2019). Due to the multifocal tumor infiltration and spreading, the outcomes for these patients remain poor and the 5 years OS of glioblastoma patients is less than 5% (Falkenstein et al., 2020; Stylli, 2020). It is widely accepted that gliomagenesis is likely to involve cell immortalization, abnormal activation of signaling pathways, and multiple gene mutations. Hence, gene therapy has become a novel therapeutic option for glioma.

The present study examined the molecular mechanisms and physiological functions of CUX1 in glioma. When analyzing datasets from the GEO and TCGA RNA-seq databases, the findings indicated that CUX1 was overexpressed in glioma tissues and exhibited a significant association with shorter OS, which was also confirmed with our clinical analysis results. Furthermore, our previous study also indicated that CUX1 overexpression was significantly correlated with several clinicopathological features, such as P53 mutation, Ki67 expression, and WHO tumor grade (Wu et al., 2019). These results indicated that CUX1 had prognostic value in patients with glioma. CUX1 is widely expressed in all metazoans and is functionally conserved in human tissues (Liu et al., 2020). In accordance with our results, recent studies have reported that CUX1 overexpression was significantly associated with tumorigenesis and poor prognosis in patients with multiple myeloma, uterine leiomyomas, colorectal cancer, and high-grade breast cancer (Moon et al., 2002; Cancer Genome Atlas, 2012; Vadnais et al., 2014). Kühnemuth et al. reported that CUX1 promoted tumor-associated macrophages (TAM) to present M2 polarization-dependent angiogenesis and tumorigenesis *via* inhibiting binding of NF-κB P65 to the promoter region of the chemokine CXCL10 in pancreatic ductal adenocarcinoma (Kuhnemuth et al., 2015). Kojima et al. 2011 also found a complex regulatory cascade involving CUX1, which activated miR-24 to repress AFP gene expression *via* inhibition of the ZBTB20 transcription factor, suggesting that aberrantly elevated CUX1 expression was correlated with tumorigenesis in hepatocarcinoma (Kojima et al., 2011). A genome-wide location analysis revealed that the biochemical functions and expression of CUX1 were regulated in a cell cycle-dependent

manner. The level of histone nuclear factor D (HiNF-D), which is a DNA-binding partner of CUX1, was elevated and accompanied Cdc25A phosphorylation, whereas, DNA-binding activity was attenuated in early G1 and G2 phases in normal cells (Sansregret et al., 2010; Sansregret and Nepveu, 2011). Moreover, mouse embryonic fibroblasts (MEFs) from CUX1$^{-/-}$ mice proliferated more slowly and presented cells having a longer G1 phase, whereas, cells with CUX1 overexpression showed accelerated entry into S phase, and stimulated proliferation in many cell types (Sansregret et al., 2006). In line with these findings, the present study revealed that CUX1 overexpression induced cell accumulation in S phase and increase the expression of cell cycle-related proteins, including CDK2, CDK4, CDK6, and Cyclin E and Cyclin D1. To the authors' knowledge, this was the first *in vitro* study to report that CUX1 promoted proliferation by activating G1/S transition in glioma cells.

Tumorigenesis is a multistep and complex process that involves adhesion and degradation of extracellular matrix components and the active migration from the primary tumor (Panciera et al., 2020). Our study found that CUX1 overexpression significantly increased the invasion and migration of glioma cells. Conversely, silencing CUX1 showed significantly retarded capability of infiltration and spread in glioma cells, while, decreasing the expression of migration-related genes. Recent studies indicated that even non-pure epithelial cells can switch from non-polar epithelial phenotype into mesenchymal phenotype, and the existence of this phenomenon was also confirmed in different experimental models, referred to as EMT-like process (Tam and Weinberg, 2013; Iser et al., 2017). Iser et al. 2017 reported that reactive astrocytes might undergo EMT-like under stimuli of GBM cells. As a result, the positive feedback of astrocytes and glioma cells could induce cell migration and invasion to support gliomagenesis (Iser et al., 2019). Our data showed that knockdown of CUX1 suppressed the expression levels of mesenchymal markers (N-cadherin, vimentin and snail), and enhanced the expression of epithelial marker E-cadherin. CUX1 overexpression presented the opposite effects, which indicated that CUX1 participated in the regulation of EMT-like process in glioma. These findings above implied that CUX1 exerted invasion, migration and malignant phenotype-promoting effects on glioma cells.

The Wnt/β-catenin pathway plays a crucial role in various physiological activities, meanwhile, aberrant mutation and activation of Wnt/β-catenin signaling is linked with the occurrence and development of glioma (He et al., 2019; Ma et al., 2019; Chen et al., 2020). β-catenin, as the major subunit of the canonical Wnt pathway, accumulates in cytoplasm and then is transported into the nucleus upon genetic mutations or activation of Wnt, where it binds with T-cell factor/lymphoid enhancer-binding factor (TCF/LEF) to promote the transcription of downstream cell motility-related genes, such as Cyclin D1, C-Myc, MMP7, and TCF4 (Gekas et al., 2016; Mao et al., 2017; Shang et al., 2017). To elucidate the molecular mechanisms underlying CUX1 in gliomagenesis, we performed single gene GSEA analysis using the KEGG curated geneset; the "Hallmark WNT Beta Catenin signaling" was among the top positively enriched pathways in the CUX1 high-expression group. A positive correlation of CUX1 with key subunits of the Wnt/β-catenin pathway was also identified in GTEx, TCGA, CGGA and CCLE datasets. It is well-known that the canonical Wnt pathway is mediated by β-catenin, hence, we speculated that the CUX1-induced gliomagenesis was correlated with the activation of Wnt/β-catenin signaling pathway.

In the present of Wnt ligands (On State), Wnt binds both LRP5/6 and Fzd receptor to initiate phosphorylation of the Axin2/APC/GSK3β complex (destruction complex). Then, β-catenin phosphorylation is inhibited, which prevents its degradation in the proteasome. β-catenin accumulates in the cytosol and then translocates to start the Wnt response gene transcription (Zhang et al., 2012; Stamos and Weis, 2013; He et al., 2019). Our IF assays revealed that CUX1 overexpression induced accumulation of Axin2 in cytoplasm and β-catenin in nucleus and cytoplasm, which was in accordance with the notion that the degradation of β-catenin was controlled by Axin2/APC/GSK3β complex in cytoplasm. We hypothesized that aberrantly upregulated CUX1 might prevent β-Catenin from being degraded by Axin2 complex, leading to the substantial accumulation of β-Catenin in cytoplasm and subsequent transportation into the nucleus, where it increased the transcription of downstream genes involved in the activation of Wnt/β-Catenin signaling, ultimately inducing tumorigenesis. As expected, our WB results indicated that CUX1 overexpression increased the level of a series of downstream genes belonging the Wnt/β-catenin pathway such as Cyclin D1, C-Myc, MMP7, and TCF4. The expression of these proteins and the CUX1-induced tumor-promoting effect could be reversed by silencing Axin2 or β-catenin expression. Importantly, WB assay also found the levels of EMT markers (N-cadherin, E-cadherin, vimentin and snail) were reversed after inhibiting Wnt/β-catenin pathway (knocking-down Axin2 or β-catenin) in TJ905 cells ectopically overexpressing CUX1 ($p < 0.05$, **Figures 7H,I**). It is evidence that Wnt/β-catenin pathway participated the CUX1-induced malignant phenotype in glioma. The findings above were consistent with our hypothesis, confirming that CUX1 was a positive upstream regulator of the Wnt/β-catenin pathway in gliomagenesis.

CONCLUSIONS

These data suggested that CUX1 was up-regulated in glioma and correlated with poor prognosis. Moreover, CUX1 could behave as an oncogene that stimulated the malignant phenotype (EMT) by activating the Wnt/β-catenin signaling pathway in glioma cells and as a potential target for the development of anti-glioma drugs.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Linyi People's Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FF, XH designed the study, ZZ, YZ and YC researched the literature. FF wrote the manuscript and performed bioinformatics analysis. XH, FF contributed to figures. ZZ and FF performed the invitro experiments and data analysis. XW, XH supervised the study. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Nature Science Foundation of Shandong Province (ZR2014HM077).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.705008/full#supplementary-material

REFERENCES

Babu, R., Komisarow, J. M., Agarwal, V. J., Rahimpour, S., Iyer, A., Britt, D., et al. (2016). Glioblastoma in the Elderly: the Effect of Aggressive and Modern Therapies on Survival. *J. Neurosurg.* 124, 998–1007. doi:10.3171/2015.4.JNS142200

Cancer Genome Atlas, N (2012). Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature.* 487, 330–337. doi:10.1038/nature11252

Chen, Y., Fang, R., Yue, C., Chang, G., Li, P., Guo, Q., et al. (2020). Wnt-Induced Stabilization of KDM4C Is Required for Wnt/β-Catenin Target Gene Expression and Glioblastoma Tumorigenesis. *Cancer Res.* 80, 1049–1063. doi:10.1158/0008-5472.CAN-19-1229

Falkenstein, F., Gessi, M., Kandels, D., Ng, H. K., Schmidt, R., Warmuth-Metz, M., et al. (2020). Prognostic Impact of Distinct Genetic Entities in Pediatric Diffuse Glioma WHO -Grade II -Report From the German/Swiss SIOP-LGG 2004 Cohort. *Int. J. Cancer.* 147, 2159–2175. doi:10.1002/ijc.32995

Fan, X., Wang, H., Zhou, J., Wang, S., Zhang, X., Li, T., et al. (2014). The Transcription Factor CUTL1 Is Associated With Proliferation and Prognosis in Malignant Melanoma. *Melanoma Res.* 24, 198–206. doi:10.1097/CMR.0000000000000064

Griesmann, H., Ripka, S., Pralle, M., Ellenrieder, V., Baumgart, S., Buchholz, M., et al. (2013). WNT5A-NFAT Signaling Mediates Resistance to Apoptosis in Pancreatic Cancer. *Neoplasia.* 15, 11–IN9. doi:10.1593/neo.121312

Guelfi, S., Duffau, H., Bauchet, L., Rothhut, B., and Hugnot, J.-P. (2016). Vascular Transdifferentiation in the CNS: A Focus on Neural and Glioblastoma Stem-like Cells. *Stem Cell Int.* 2016, 1–13. doi:10.1155/2016/2759403

He, L., Zhou, H., Zeng, Z., Yao, H., Jiang, W., and Qu, H. (2019). Wnt/β-catenin Signaling cascade: A Promising Target for Glioma Therapy. *J. Cel Physiol.* 234, 2217–2228. doi:10.1002/jcp.27186

Hulea, L., and Nepveu, A. (2012). CUX1 Transcription Factors: From Biochemical Activities and Cell-Based Assays to Mouse Models and Human Diseases. *Gene.* 497, 18–26. doi:10.1016/j.gene.2012.01.039

Iser, I. C., Lenz, G., and Wink, M. R. (2019). EMT-Like Process in Glioblastomas and Reactive Astrocytes. *Neurochem. Int.* 122, 139–143. doi:10.1016/j.neuint.2018.11.016

Iser, I. C., Pereira, M. B., Lenz, G., and Wink, M. R. (2017). The Epithelial-To-Mesenchymal Transition-like Process in Glioblastoma: An Updated Systematic Review and In Silico Investigation. *Med. Res. Rev.* 37, 271–313. doi:10.1002/med.21408

Jo, Y. S., Kim, M. S., Yoo, N. J., and Lee, S. H. (2017). Intratumoral Heterogeneity for Inactivating Frameshift Mutation of CUX1 and SIRT1 Genes in Gastric and Colorectal Cancers. *Pol. J. Pathol.* 68, 258–260. doi:10.5114/pjp.2017.71534

Kang, J. H., and Adamson, C. (2015). Novel Chemotherapeutics and Other Therapies for Treating High-Grade Glioma. *Expert Opin. Investig. Drugs.* 24, 1361–1379. doi:10.1517/13543784.2015.1048332

Kaur, N., Chettiar, S., Rathod, S., Rath, P., Muzumdar, D., Shaikh, M. L., et al. (2013). Wnt3a Mediated Activation of Wnt/β-Catenin Signaling Promotes Tumor Progression in Glioblastoma. *Mol. Cell Neurosci.* 54, 44–57. doi:10.1016/j.mcn.2013.01.001

Kedinger, V., Sansregret, L., Harada, R., Vadnais, C., Cadieux, C., Fathers, K., et al. (2009). p110 CUX1 Homeodomain Protein Stimulates Cell Migration and Invasion in Part Through a Regulatory cascade Culminating in the Repression of E-Cadherin and Occludin. *J. Biol. Chem.* 284, 27701–27711. doi:10.1074/jbc.M109.031849

Kojima, K., Takata, A., Vadnais, C., Otsuka, M., Yoshikawa, T., Akanuma, M., et al. (2011). MicroRNA122 Is a Key Regulator of α-fetoprotein Expression and Influences the Aggressiveness of Hepatocellular Carcinoma. *Nat. Commun.* 2, 338. doi:10.1038/ncomms1345

Krug, S., Weissbach, J., Blank, A., Perren, A., Haybaeck, J., Fendrich, V., et al. (2020). CUX1-Transcriptional Master Regulator of Tumor Progression in Pancreatic Neuroendocrine Tumors. *Cancers.* 12, 1957. doi:10.3390/cancers12071957

Kühnemuth, B., Mühlberg, L., Schipper, M., Griesmann, H., Neesse, A., Milosevic, N., et al. (2015). CUX1 Modulates Polarization of Tumor-Associated Macrophages by Antagonizing NF-Kb Signaling. *Oncogene.* 34, 177–187. doi:10.1038/onc.2013.530

Li, C., Wang, S., Yan, J.-L., Torheim, T., Boonzaier, N. R., Sinha, R., et al. (2020). Characterizing Tumor Invasiveness of Glioblastoma Using Multiparametric Magnetic Resonance Imaging. *J. Neurosurg.* 132, 1465–1472. doi:10.3171/2018.12.JNS182926

Liu, C., Tu, Y., Sun, X., Jiang, J., Jin, X., Bo, X., et al. (2011). Wnt/Beta-Catenin Pathway in Human Glioma: Expression Pattern and Clinical/prognostic Correlations. *Clin. Exp. Med.* 11, 105–112. doi:10.1007/s10238-010-0110-9

Liu, N., Sun, Q., Wan, L., Wang, X., Feng, Y., Luo, J., et al. (2020). CUX1, A Controversial Player in Tumor Development. *Front. Oncol.* 10, 738. doi:10.3389/fonc.2020.00738

Liu, S., McLeod, E., and Jack, J. (1991). Four Distinct Regulatory Regions of the Cut Locus and Their Effect on Cell Type Specification in Drosophila. *Genetics.* 127, 151–159. doi:10.1093/genetics/127.1.151

Ludlow, C., Choy, R., and Blochlinger, K. (1996). Functional Analysis of Drosophila and Mammalian Cut Proteins in Flies. *Dev. Biol.* 178, 149–159. doi:10.1006/dbio.1996.0205

Ma, L., Lin, K., Chang, G., Chen, Y., Yue, C., Guo, Q., et al. (2019). Aberrant Activation of β-Catenin Signaling Drives Glioma Tumorigenesis via USP1-Mediated Stabilization of EZH2. *Cancer Res.* 79, 2018–85. doi:10.1158/0008-5472.CAN-18-1304

Malizia, A. P., Lacey, N., Walls, D., Egan, J. J., and Doran, P. P. (2009). CUX1/Wnt Signaling Regulates Epithelial Mesenchymal Transition in EBV Infected Epithelial Cells. *Exp. Cel Res.* 315, 1819–1831. doi:10.1016/j.yexcr.2009.04.001

Mao, X.-W., Xiao, J.-Q., Xu, G., Li, Z.-Y., Wu, H.-F., Li, Y., et al. (2017). CUL4B Promotes Bladder Cancer Metastasis and Induces Epithelial-To-Mesenchymal Transition by Activating the Wnt/β-Catenin Signaling Pathway. *Oncotarget.* 8, 77241–77253. doi:10.18632/oncotarget.20455

Moon, N. S., Rong Zeng, W., Premdas, P., Santaguida, M., Bérubé, G., and Nepveu, A. (2002). Expression of N-Terminally Truncated Isoforms of CDP/CUX Is

Increased in Human Uterine Leiomyomas. *Int. J. Cancer.* 100, 429–432. doi:10.1002/ijc.10510

Nager, M., Bhardwaj, D., Cantí, C., Medina, L., Nogués, P., and Herreros, J. (2012). β-Catenin Signalling in Glioblastoma Multiforme and Glioma-Initiating Cells. *Chemother. Res. Pract.* 2012, 1–7. doi:10.1155/2012/192362

Panciera, T., Citron, A., Di Biagio, D., Battilana, G., Gandin, A., Giulitti, S., et al. (2020). Reprogramming Normal Cells into Tumour Precursors Requires ECM Stiffness and Oncogene-Mediated Changes of Cell Mechanical Properties. *Nat. Mater.* 19, 797–806. doi:10.1038/s41563-020-0615-x

Pastushenko, I., and Blanpain, C. (2019). EMT Transition States During Tumor Progression and Metastasis. *Trends Cel Biol.* 29, 212–226. doi:10.1016/j.tcb.2018.12.001

Pinel, S., Thomas, N., Boura, C., and Barberi-Heyob, M. (2019). Approaches to Physical Stimulation of Metallic Nanoparticles for Glioblastoma Treatment. *Adv. Drug Deliv. Rev.* 138, 344–357. doi:10.1016/j.addr.2018.10.013

Ramdzan, Z. M., and Nepveu, A. (2014). CUX1, a Haploinsufficient Tumour Suppressor Gene Overexpressed in Advanced Cancers. *Nat. Rev. Cancer.* 14, 673–682. doi:10.1038/nrc3805

Sansregret, L., Gallo, D., Santaguida, M., Leduy, L., Harada, R., and Nepveu, A. (2010). Hyperphosphorylation by Cyclin B/CDK1 in Mitosis Resets CUX1 DNA Binding Clock at Each Cell Cycle. *J. Biol. Chem.* 285, 32834–32843. doi:10.1074/jbc.M110.156406

Sansregret, L., Goulet, B., Harada, R., Wilson, B., Leduy, L., Bertoglio, J., et al. (2006). The P110 Isoform of the CDP/Cux Transcription Factor Accelerates Entry into S Phase. *Mol. Cel Biol.* 26, 2441–2455. doi:10.1128/MCB.26.6.2441-2455.2006

Sansregret, L., and Nepveu, A. (2011). Gene Signatures of Genomic Instability as Prognostic Tools for Breast Cancer. *Future Oncol.* 7, 591–594. doi:10.2217/fon.11.34

Sansregret, L., and Nepveu, A. (2008). The Multiple Roles of CUX1: Insights From Mouse Models and Cell-Based Assays. *Gene.* 412, 84–94. doi:10.1016/j.gene.2008.01.017

Sareddy, G. R., Panigrahi, M., Challa, S., Mahadevan, A., and Babu, P. P. (2009). Activation of Wnt/β-catenin/Tcf Signaling Pathway in Human Astrocytomas. *Neurochem. Int.* 55, 307–317. doi:10.1016/j.neuint.2009.03.016

Shang, S., Hua, F., and Hu, Z.-W. (2017). The Regulation of β-catenin Activity and Function in Cancer: Therapeutic Opportunities. *Oncotarget.* 8, 33972–33989. doi:10.18632/oncotarget.15687

Shaver, M., Kohanteb, P., Chiou, C., Bardis, M., Chantaduly, C., Bota, D., et al. (2019). Optimizing Neuro-Oncology Imaging: A Review of Deep Learning Approaches for Glioma Imaging. *Cancers.* 11, 829. doi:10.3390/cancers11060829

Stamos, J. L., and Weis, W. I. (2013). The -Catenin Destruction Complex. *Cold Spring Harbor Perspect. Biol.* 5, a007898. doi:10.1101/cshperspect.a007898

Stylli, S. S. (2020). Novel Treatment Strategies for Glioblastoma. *Cancers.* 12, 2883. doi:10.3390/cancers12102883

Tam, W. L., and Weinberg, R. A. (2013). The Epigenetics of Epithelial-Mesenchymal Plasticity in Cancer. *Nat. Med.* 19, 1438–1449. doi:10.1038/nm.3336

Vadnais, C., Shooshtarizadeh, P., Rajadurai, C. V., Lesurf, R., Hulea, L., Davoudi, S., et al. (2014). Autocrine Activation of the Wnt/β-Catenin Pathway by CUX1 and GLIS1 in Breast Cancers. *Biol. Open.* 3, 937–946. doi:10.1242/bio.20148193

Vanden Heuvel, G. B., Brantley, J. G., Alcalay, N. I., Sharma, M., Kemeny, G., Warolin, J., et al. (2005). Hepatomegaly in Transgenic Mice Expressing the Homeobox geneCux-1. *Mol. Carcinog.* 43, 18–30. doi:10.1002/mc.20091

Wan, Z., Shi, W., Shao, B., Shi, J., Shen, A., Ma, Y., et al. (2011). Peroxisome Proliferator-Activated Receptor γ Agonist Pioglitazone Inhibits β-catenin-mediated Glioma Cell Growth and Invasion. *Mol. Cel Biochem.* 349, 1–10. doi:10.1007/s11010-010-0637-9

Wong, C. C., Martincorena, I., Martincorena, I., Rust, A. G., Rashid, M., Alifrangis, C., et al. (2014). Inactivating CUX1 Mutations Promote Tumorigenesis. *Nat. Genet.* 46, 33–38. doi:10.1038/ng.2846

Wu, X., Feng, F., Yang, C., Zhang, M., Cheng, Y., Zhao, Y., et al. (2019). Upregulated Expression of CUX1 Correlates with Poor Prognosis in Glioma Patients: a Bioinformatic Analysis. *J. Mol. Neurosci.* 69, 527–537. doi:10.1007/s12031-019-01355-3

Zhang, J., Huang, K., Shi, Z., Zou, J., Wang, Y., Jia, Z., et al. (2011). High -Catenin/Tcf-4 Activity Confers Glioma Progression via Direct Regulation of AKT2 Gene Expression. *Neuro-Oncology.* 13, 600–609. doi:10.1093/neuonc/nor034

Zhang, K., Zhang, J., Han, L., Pu, P., and Kang, C. (2012). Wnt/Beta-catenin Signaling in Glioma. *J. Neuroimmune Pharmacol.* 7, 740–749. doi:10.1007/s11481-012-9359-y

Zong, H., Parada, L. F., and Baker, S. J. (2015). Cell of Origin for Malignant Gliomas and its Implication in Therapeutic Development. *Cold Spring Harb Perspect. Biol.* 7, a020610. doi:10.1101/cshperspect.a020610

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 Feng, Zhao, Zhou, Cheng, Wu and Heng. 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 Molecular Biosciences

REVIEW
published: 11 August 2021
doi: 10.3389/fmolb.2021.683844

Meta-Analysis of the Diagnostic Value of Cell-free DNA for Renal Cancer

Yipeng Xu[1†], Yingjun Jiang[2†], Mingke Yu[3†], Jianmin Lou[3], Mei Song[4], Han Xu[5], Yingying Cui[6], Xiaowei Zeng[1,2], Qibo Wang[1], Hanyun Ma[7], Zongping Wang[1], Shaoxing Zhu[1], Guorong Li[8] and An Zhao[9]**

[1]Department of Urology, Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital, Hangzho, China, [2]Hangzhou Traditional Chinese Medicine Hospital, Zhejiang Chinese Medical University, Hangzhou, China, [3]The Second Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou, China, [4]Department of Ultrasound, Cancer Hospital of the University of Chinese Academy of Sciences, Hangzhou, China, [5]Central Research Laboratory, Children's Hospital of Nanchang University, Hangzhou, China, [6]Translational Radiation Oncology Research Laboratory, Department of Radiooncology and Radiotherapy, Charite University Hospital, Berlin, Germany, [7]Comprehensive Cancer Center, Charité Universitätsmedizin, Berlin, Germany, [8]Department of Urology, North Hospital, University of Jean-Monnet, Saint-Etienne, France, [9]Experimental Research Center, Zhejiang Cancer Hospital, Institute of Cancer Research and Basic Medicine, Cancer Hospital of the University of Chinese Academy of Sciences, Hangzhou, China*

OPEN ACCESS

Edited by:
Ismail Hosen, University of Dhaka, Bangladesh

Reviewed by:
Liping Sun, The First Affiliated Hospital of China Medical University, China
Maria Zvereva, Lomonosov Moscow State University, Russia

***Correspondence:**
An Zhao
zhaoan@zjcc.org.cn
Shaoxing Zhu
zhusx@zjcc.org.cn

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *22 March 2021*
Accepted: *27 July 2021*
Published: *11 August 2021*

Citation:
Xu Y, Jiang Y, Yu M, Lou J, Song M, Xu H, Cui Y, Zeng X, Wang Q, Ma H, Wang Z, Zhu S, Li G and Zhao A (2021) Meta-Analysis of the Diagnostic Value of Cell-free DNA for Renal Cancer. Front. Mol. Biosci. 8:683844. doi: 10.3389/fmolb.2021.683844

Cell-free DNA (cf-DNA) has been reported to represent a suitable material for liquid biopsy in the diagnosis and prognosis of various cancers. We performed a meta-analysis of published data to investigate the diagnostic value of cf-DNA for renal cancer (RCa). Systematic searches were conducted using Pubmed, Embase databases, Web of Science, Medline and Cochrane Library to identify relevant publications until the 31st March 2021. For all patients, we evaluated the true diagnostic value of cf-DNA by calculating the number of true positive, false positive, true negative, and false negative, diagnoses by extracting specificity and sensitivity data from the selected literature. In total, 8 studies, featuring 754 RCa patients, and 355 healthy controls, met our inclusion criteria. The overall diagnostic sensitivity and specificity for cf-DNA was 0.71 (95% confidence interval (CI), 0.55–0.83) and 0.79 (95% CI, 0.66–0.88), respectively. The pooled positive likelihood ratio and pooled negative likelihood ratio were 3.42 (95% CI, 2.04–5.72) and 0.36 (95% CI, 0.23–0.58), respectively. The area under the summary receiver operating characteristic curve was 0.82 (95% CI, 0.79–0.85), and the diagnostic odds ratio was 7.80 (95% CI, 4.40–13.85). Collectively, our data demonstrate that cf-DNA has high specificity and sensitivity for diagnosing RCa. Therefore, cf-DNA is a useful biomarker for the diagnosis of RCa.

Keywords: cf-DNA, renal cancer, diagnosis, liquid biopsy, meta-analysis

INTRODUCTION

Renal cancer (RCa) is the 13th most common cancer in the world and accounts for 2.4% of all cancers; the highest incidence of RCa has been reported in developed countries (Capitanio et al., 2019). More than 73,000 new cases of RCa are diagnosed in the United States every year, with 330,000 new cases globally (Capitanio et al., 2019; Siegel et al., 2019). Although developments in computed tomography and magnetic resonance imaging have increased the proportion of diagnoses for the early stages of RCa, only 47% of RCa patients with locoregional disease can survive more than 5 years (Choueiri and Motzer, 2017). Moreover, metastases occur in 4.2–7.1% of RCa patients when

the diameter of the tumor is <4 cm, and this relates to an 8% 5-years survival rate for these patients (Lughezzani et al., 2009). To date, there is no biomarker for RCa like PSA for prostate cancer or EGFR for lung cancer. The discovery and application of novel biomarkers for RCa are still expected in clinical.

A number of recent publications have been reported that the potential for using cell-free DNA (cf-DNA) for the diagnosis of certain diseases (Jiang and Lo, 2016; Bianchi and Chiu, 2018). During tumorigenesis and the progression of cancer, it is likely that cf-DNA will be released into a patient's blood by cells undergoing apoptosis, or by exosomes (Bardelli and Pantel, 2017). Consequently, cf-DNA, consisting of nucleic acid chains from various cell types, could be detected in the blood, stools, urine or saliva (Stewart et al., 2018). A variety of strategies could therefore be used to analyze cf-DNA, including real-time polymerase chain reaction (RT-PCR), digital PCR or next generation sequencing (van Ginkel et al., 2017; Cohen et al., 2021; Yang et al., 2021). cf-DNA has also been found to be able to serve as the prognostic indicator for tumor progression and drug resistance in cancer patients (Adalsteinsson et al., 2017; Remon et al., 2017; Alix-Panabières and Pantel, 2021). Since then, several studies assessing the value of cf-DNA in RCa have been published. However, the diagnostic performance of this novel biomarker has not been evaluated systematically. Therefore, the purpose of our study was to assess the diagnostic performance of cf-DNA for the detection of RCa.

MATERIALS AND METHODS

Search Strategy

We carried out systematic literature searches to identify relevant publications in the PubMed, embase databases, Web of Science, Medline and Cochrane Library up to the March 31, 2021, without language or date restrictions.

The search strategy included the following terms: ("kidney neoplasms" OR "kidney cancer" OR "renal cancer") AND ("diagnosis" OR "biomarker") AND ("overall survival (OS)" OR "disease-free survival (DFS)" OR "progression-free survival (PFS)" OR "prognosis" OR "survival" AND "circulating tumor DNA" OR "cell-free nucleic acids" OR "ct-DNA" OR "cf-DNA". Three researchers (Yipeng Xu, Yingjun Jiang and Mingke Yu) independently assessed the eligibility of each potentially relevant study by screening the titles and abstracts. Disagreements between the two researchers were resolved by discussion with two additional researchers (An Zhao and Shaoxing Zhu). Additional publications were identified by searching the reference lists of the selected papers.

Inclusion Criteria

The inclusion criteria were as follows (Capitanio et al., 2019): at least one diagnostic or prognostic parameters for cf-DNA detection was reported in RCa patients, or could be calculated from the published data (Siegel et al., 2019); samples were collected from the peripheral blood (Choueiri and Motzer, 2017); the techniques were clearly stated in the articles; and Lughezzani et al. (2009) studies must feature negative controls.

Exclusion Criteria

The exclusion criteria were as follows (Capitanio et al., 2019): repeated or overlapped publications which included the same study population and genes (Siegel et al., 2019); experiments based exclusively on cell lines or tumor tissue rather than clinical samples; and Choueiri and Motzer (2017) studies with a poor sample size (≤10).

Data Extraction

All eligible studies were independently reviewed by two investigators (Y.J. Y.X.). The following items were extracted from each article: first author's name, year of publication, number of patients, TNM stage, sample origin, methods of DNA detection, detection markers, and information relating the article's quality. A range of diagnostic data were also extracted, including specificity, sensitivity, true positive (TP) rate, false positive (FP) rate, true negative (TN) rate, and false negative (FN) rate. We also acquired a range of survival data, including OS, PFS, hazard ratio (HR), p value, Kaplan-Meier survival curves, and 95% confidence intervals (95% CIs). Engauge Digitizer 4.1 was used to read the Kaplan-Meier curves in order to identify articles with accurate HRs.

Risk of Bias in Individual Studies and Synthesis of the Results

Deek's funnel plot and Quality Assessment of Diagnostic Accuracy Studies (QUADAS) 2 tool were adopted to analyze qualitative publication bias, and a P-value of <0.05 was considered statistically significant. Risk-of-bias assessment was performed independently by two authors (Y.J. Y.X.) according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) 2 tool (Whiting et al., 2011). Disagreement was solved by a third party (M.Y.). This tool provides a measure of the risk of bias and applicability over four domains of interest (**Figure 2A**). No publication bias in the pooled diagnostic effects was determined by Deek's funnel plot (p = 0.43) (**Figure 2B**).

Statistical Methods

The systematic review and meta-analysis were performed using RevMan Version 5.3 and STATA 11.0 (Stata Crop). Diagnostic variables, including positive likelihood ratios (PLR), negative likelihood ratios (NLR), and summary receiver operating characteristic curves (SROC), were analyzed by STATA 11.0 (Stata Crop), and the diagnostic odds ratio (DOR) was analyzed by Meta-DiSc software, version 1.4. Specificity was defined as the proportion of patients with no cf-DNA detection in the blood samples when compared with all negative control volunteers without RCa. Sensitivity was defined as the proportion of RCa patients containing cf-DNA in their blood samples. TP, FP, TN, and FN, were calculated by analyzing the specificity, sensitivity, and the number of people enrolled in each group (experimental group and control group). Significant heterogeneity was defined as when $p < 0.05$ or $I^2 > 50\%$, and a random-effect model was used for heterogeneity analysis.

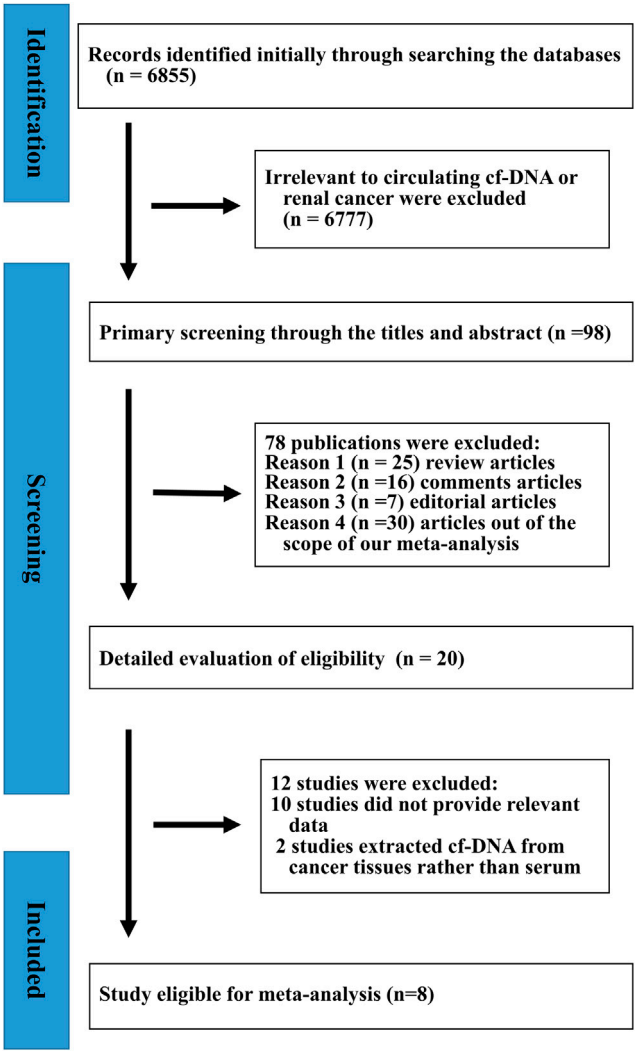


FIGURE 1 | Flowchart describing the selection of publications for meta-analysis.

RESULTS

Study Selection

In total, we retrieved 6,855 articles. Of these, 6,777 articles were excluded because they did not specifically refer to cf-DNA and RCa. By reviewing each title and abstract, we identified 25 review articles, 16 comments, seven editorials, and 30 articles, that were outside of the scope of our meta-analysis. Twenty studies were recognized as potentially relevant publications, and a full-text review was performed to identify data relating to diagnoses and prognoses. As shown in **Figure 1**, careful screening and verification identified eight studies that were eligible for meta-analysis. The main characteristics and details of these eligible studies Salinas-Sanchez et al. (2021), Lasseter et al. (2020), Yamamoto et al. (2018), Lu et al. (2016), Wan et al. (2013), De Martino et al. (2012), Ellinger et al. (2012), Hauser et al. (2010) are summarized in **Table 1**. These eight eligible studies featured a total of 754 patients, with a median sample size of 87 (range: 33–229, mean: 94). In total, 355 controls were enrolled by the eight eligible studies, of which 312 were healthy individuals and 43 were patients with benign renal tumors (De Martino et al., 2012). Four studies included patients at stages I-IV, while the remaining three studies featured patients in stages I-III (De Martino et al., 2012; Ellinger et al., 2012; Wan et al., 2013). The number of TP, FP, FN, and TN, cases in these studies are shown in **Table 1**. One of the eligible articles were performed in Spain, two studies were performed in East Asia (China and Japan), three studies were performed in Germany, and the other was performed in the United States.

Detection of Cf-DNA

Cf-DNA was primarily detected by next generation sequencing or PCR-based method, and it could be characterized by composition (size, fragment or integrity), concentration (total, panel or specified gene) or genetic characteristics (methylation or nucleotide variants) (**Table 1**). Of these, three studies extracted DNA from serum Hauser et al. (2010), De Martino et al. (2012), Ellinger et al. (2012), while the other five studies extracted DNA from plasma (Wan et al., 2013; Lu et al., 2016; Yamamoto et al., 2018; Lasseter et al., 2020; Salinas-Sanchez et al., 2021). The blood volume required for the detection of cf-DNA varied from 0.8 to 3 ml. Notably, in the three enrolled studies, 1 ml of serum was used for cf-DNA detection (Hauser et al., 2010; De Martino et al., 2012; Ellinger et al., 2012). All of the studies collected blood samples prior to initial treatment.

Risk of Bias Within Studies

The quality of the selected studies was evaluated in accordance with the QUADAS-2 criteria; the results of these evaluations are shown in **Figure 2**. Two studies were considered to be low-risk with regards to bias and applicability, and the other six studies were estimated as suboptimal for unclear risk in several areas, including patient selection, reference standards, and index testing. With DOR as the effect variable, the heterogeneity test gave a p value of 0.015, and an I^2 value of 59.6%, suggesting that the heterogeneity was existed between these studies. In addition, meta-regression analysis was performed to analysis the heterogeneity. Among several potential variables, including the source of cf-DNA (serum or plasma), proportion of patients with RCa and region (Asia/USA/Europe), were not significant factors ($p > 0.05$, data not shown).

Meta-Analysis of Diagnostic Value

All eight eligible studies were used to evaluate the diagnostic accuracy between cf-DNA expression and RCa. As shown in **Figure 3**, the overall diagnostic sensitivity and specificity were 0.71 (95% CI, 0.55–0.83) and 0.79 (95% CI, 0.66–0.88), respectively. The level of cf-DNA was significantly correlated with specificity ($p < 0.001$, $I^2 = 88.80\%$) and sensitivity ($p < 0.001$, $I^2 = 89.25\%$) (**Figure 3**).

The pooled PLR and NLR were 3.42 (95% CI, 2.04–5.72) and 0.36 (95% CI, 0.23–0.58) (**Figure 4**). The SROC was 0.82 (95% CI,

TABLE 1 | Characteristics of Studies Evaluating the cf-DNA Levels of Patients with Renal Cancer

Study ID	Region/ Year	Sample size (case/ Control)	Mean age (case/ Control)	Sample/Method	Cf-DNA characterize	TNM (i/ II/III/IV)	TP	FP	FN	TN	Sensitivity/ Specificity
Salinas-Sánchez et al (2021)	Spain/ 2021	82/20	59.7/59.5	Plasma cf-DNA/ qPCR	Fragments of cf-DNA	58(I + II)/ 24(III + IV)	32	2	50	18	39.1%/90.0%
Lasseter et al (2020)	American/ 2020	34/34	NM	Plasma cf-DNA/ cfMeDIP–seq	Methylation score of cf-DNA	20/3/6/9	34	4	0	30	100%/88%
Yamamoto et al (2018)	Japan/ 2018	92/41	68/57	Plasma cf-DNA/ qPCR	Fragments of cf-DNA	58/4/ 15/15	58	9	34	32	63.0%/78.1%
Lu et al (2016)	Germany/ 2016	229/40	NM	Plasma mitochondrial cf-DNA/qPCR	Fragments of cf-DNA	108/19/ 94/2	160	5	69	35	70.0%/88.0%
Wan et al (2013)	China/ 2013	92/44	NM	Plasma cf-DNA/ qPCR	Fragment of cf-DNA	59(I + II)/ 33/0	65	13	27	31	70.6%/71.2%
De Martino et al (2012)	American/ 2012	157/43	64.7/62.5	Serum cf-DNA/ qPCR	Total cf-DNA	92(I + II)/ 65/0	80	3	77	40	51.0%/93.0%
Ellinger et al (2012)	Germany/ 2012	33/79	64.8/31.5	Serum mitochondrial cf-DNA/qPCR	Integrity of mitochondrial cf-DNA	21/1/ 11/0	27	45	6	34	81.8%/43.0%
Hauser et al (2010)	Germany/ 2010	35/54	66/28.5	Serum cf-DNA/ qPCR	Integrity of cf-DNA	21/1/ 11/2	26	20	9	34	74.2%/62.9%

qPCR, Quantitative real-time PCR; NM, Not Mentioned; TP, True Positive; FP, False Positive; FN, False Negative; TN, True Negative.

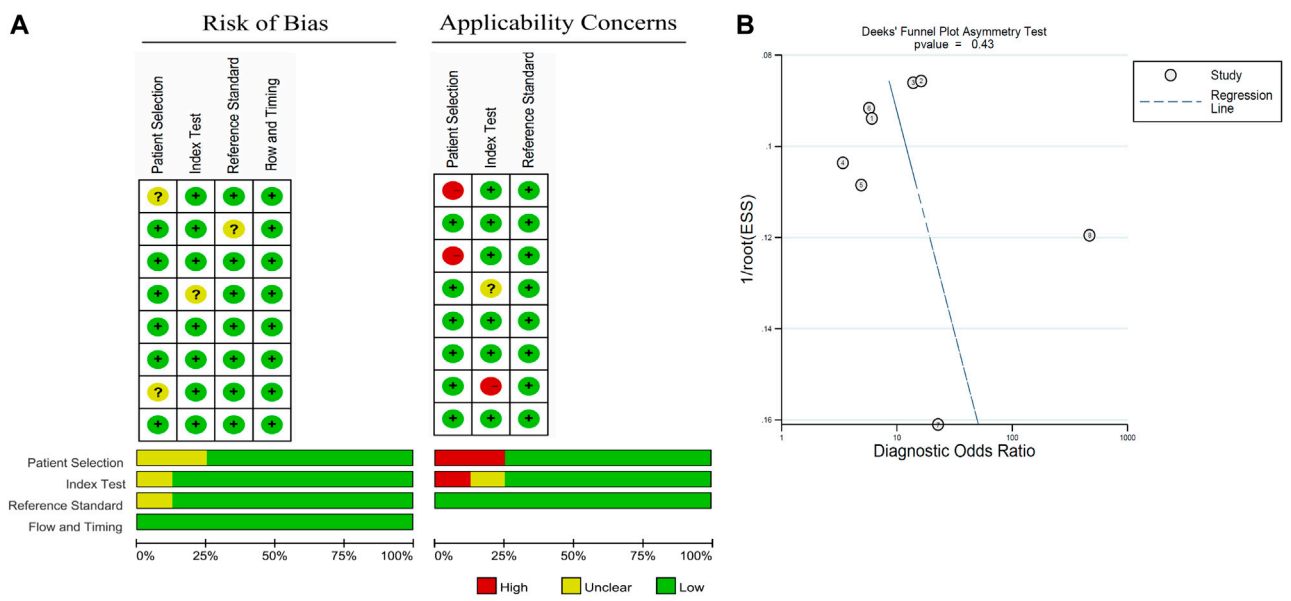


FIGURE 2 | Graphical overview of the overall risk of bias and applicability judgements for the 8 studies included according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) 2 tool and Deek's funnel plot.

0.79–0.85) (**Figure 5A**) and the DOR was 7.80 (95% CI, 4.40–13.85) (**Figure 5B**).

Prognoses

A total of six eligible studies De Martino et al. (2012), Wan et al. (2013), Lu et al. (2016), Yamamoto et al. (2018), Lasseter et al. (2020), Salinas-Sanchez et al. (2021) showed an association between cf-DNA and prognosis for patients with RCa. Two study investigated the association between cf-DNA expression and OS Lasseter et al. (2020), Salinas-Sanchez et al. (2021), while another study investigated the association between cf-DNA and DFS De Martino et al. (2012), these data could not be merged. As for the other three studies, one investigated the association between cf-DNA expression and PFS Yamamoto et al. (2018), while the other investigated the association between cf-DNA expression and RFS (recurrence free survival) (Wan et al., 2013; Lu et al., 2016); these data could not be merged. The specific details of these four studies are shown in **Supplementary Table S1**.

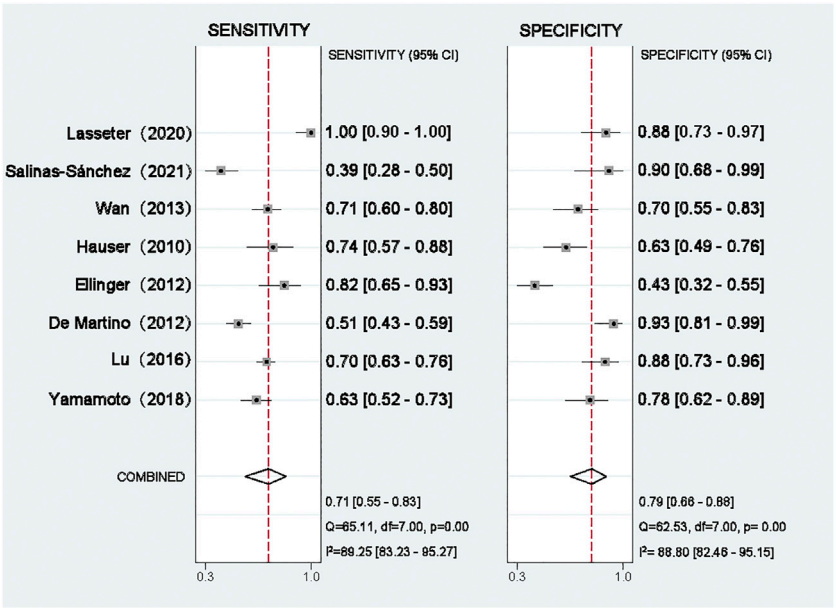


FIGURE 3 | Forest plot of the pooled sensitivity and specificity.

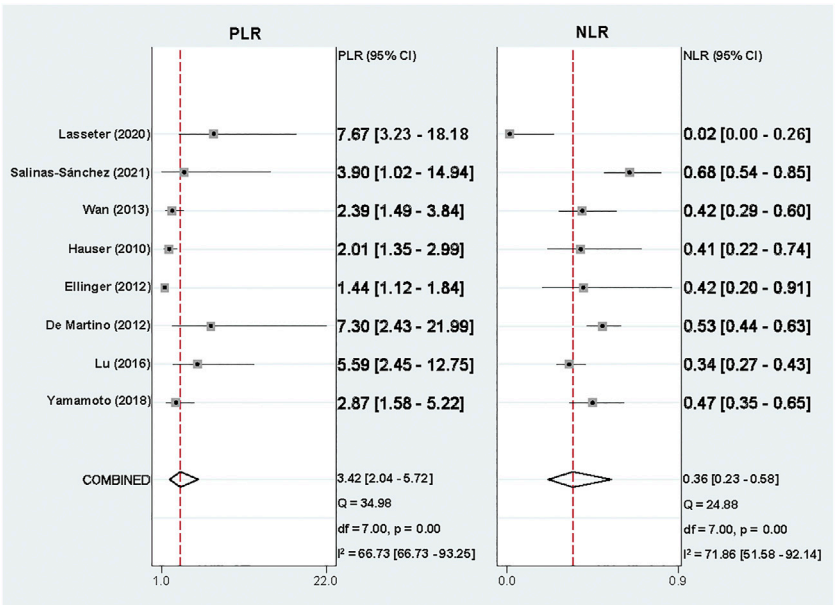


FIGURE 4 | Forest plot of positive likelihood ratio and negative likelihood ratio.

DISCUSSION

Early stage RCa is usually asymptomatic and is therefore usually discovered by chance (Ljungberg et al., 2019). The diagnosis of RCa still predominantly depends on radiological and histopathological examinations; however, these techniques are associated with exposure to radiation, and are in invasive. Unlike the case of prostate-specific antigen (PSA) for prostate cancer, there is no specific biomarker for diagnosing the early stages of RCa, or predicting disease progression in such patients. This reduces patient compliance and means that it is difficult for us to screen patients, and monitor patients over long periods of follow-up. The identification of a diagnostic indicator in peripheral blood samples of RCa patients, that could be used for diagnosis and screening, would be of significant clinical value.

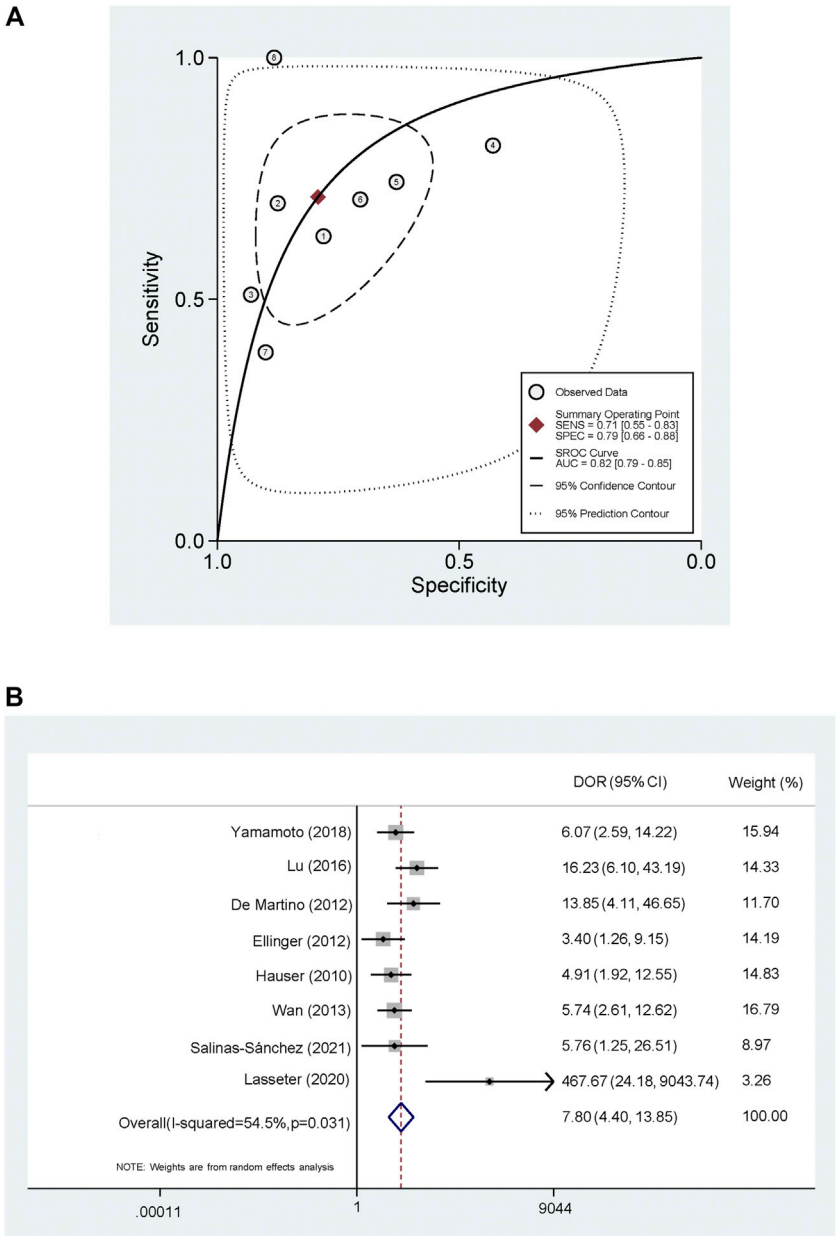


FIGURE 5 | Summary receiver operating characteristic curve and Forest plot of diagnostic odds ratio **(A)** Summary receiver operating characteristic curve **(B)** Forest plot of diagnostic odds ratio.

Liquid biopsy is widely regarded as a new diagnostic technique for cancer (Husain and Velculescu, 2017). Recent research, involving patients with urogenital cancer Di Meo et al. (2017), indicates that circulating tumor cells, cell-free nucleic acids, circulating tumor DNA, circulating cell-free RNA, and extracellular vesicles and their cargo, extracted from blood and urine, have significant potential for monitoring disease status (Zhao et al., 2015; Li et al., 2017; Zhang et al., 2018). Compared with radiological and histopathological examinations, these new methods are minimally invasive and carry minimal risk, such methods may also provide us with the possibility to test patients continuously for disease recurrence and response to treatment.

The presence of fragmented DNA in the blood was first reported by Mandel and Metais Mandel (1948) in 1948. In recent decades, the detection of cf-DNA has been applied to various different types of cancer. For example, identifying the EGFR T790M mutation in plasma samples is already known to be an effective method for determining EGFR status in patients with non-small cell lung cancer (NSCLC) (Qiu et al., 2015). Furthermore, the loss of the EGFR T790M mutation in plasma is associated with early progression to advanced NSCLC patients receiving osimertinib

(Cohen et al., 2021). The genomic profiles of ct-DNA have also been shown to closely match those of the corresponding tumors, with important implications for both molecular pathology and clinical oncology (Siravegna et al., 2017). Although the experience of cf-DNA detection in RCa patients is very limited, our ability to diagnose disease by liquid biopsy is highly likely to become increasingly powerful in the future.

To investigate the clinical utility of cf-DNA in the diagnosis of RCa, we performed a meta-analysis and found that the detection of cf-DNA conveyed an obvious advantage to the specificity of RCa diagnosis (specificity: 0.79; 95% CI: 0.66–0.88). Furthermore, the sensitivity of cf-DNA for the diagnosis of RCC was also high (sensitivity: 0.71; 95% CI: 0.55–0.83). In these analyses, higher PLR values indicated that the test results were more likely to be disease-related, while lower NLR values indicated that the test results were more likely to be disease-independent. The area under the ROC curve (AUC) was used to further evaluate the accuracy of these diagnostic tests, cf-DNA showed a high diagnostic ability (AUC = 0.82) to diagnose RCa (**Figure 5**). There are few studies on the prognosis of cf-DNA in RCa, it is not clear whether levels of cf-DNA expression could be used to evaluate patient prognosis. Interestingly, Yamamoto et al. (2018) have reported that patients with longer cf-DNA fragments (>160 bp) had a longer PFS than those with shorter fragments (<160 bp), suggesting that cf-DNA may have different release mechanisms between the normal cells and the tumor cells. In addition, multiple aspects can be assessed in circulating cf-DNA, including expression levels, integrity, methylation and mutations (Di Meo et al., 2017). As advances in laboratory technology and knowledge, the meta-analysis of diagnostic value of subtype of cf-DNA is needed in the future.

There are several limitations associated with our meta-analysis that should be taken into consideration. First, the lack of an appropriate cf-DNA gene target in RCa patients might contribute to the presence of bias. Like many other types of cancer, RCa is also considered as a malignancy with high histological and etiological heterogeneity. Further studies of specific target genes would help us to fully understand the use of cf-DNA detection for patients with RCa. Second, due to an enrichment of studies reporting positive results, it is impossible to exclude the possibility of selection bias. Other sources of bias may have arisen due to differences in detection equipment and materials, such as PCR primers and amplification systems.

CONCLUSIONS

Based on the studies selected for meta-analysis, our data indicate that cf-DNA could serves as a liquid biopsy that is effective for the diagnosis of RCa.

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

Three researchers (YX, YJ and MY) independently assessed the eligibility of each potential study by screening the titles and abstracts. Any disagreements between the two researchers were resolved by discussion with two additional researchers (AZ and SZ). We increased the number of relevant studies by screening the references of each selected paper. The manuscript was written by YX, YJ, HX and MS, and was revised by YC, HM and XZ. AZ, QW and GL revised the paper for English. QW, JL and ZW was responsible for the statistical analysis.

FUNDING

This work was supported by the National Natural Science Foundation of China (Reference number: 81402117), Natural Science Foundation of Zhejiang Province (Reference number: LY17H160043; Reference number: LY16H160034) and Medical and Health Project of Zhejiang Province (Reference number: 2016KYA038).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.683844/full#supplementary-material

REFERENCES

Adalsteinsson, V. A., Ha, G., Freeman, S. S., Choudhury, A. D., Stover, D. G., Parsons, H. A., et al. (2017). Scalable Whole-Exome Sequencing of Cell-free DNA Reveals High Concordance with Metastatic Tumors. *Nat. Commun.* 8 (1), 1324. doi:10.1038/s41467-017-00965-y

Alix-Panabières, C., and Pantel, K. (2021). Liquid Biopsy: From Discovery to Clinical Application. *Cancer Discov.* 11 (4), 858–873. doi:10.1158/2159-8290.CD-20-1311

Bardelli, A., and Pantel, K. (2017). Liquid Biopsies, what We Do Not Know (Yet). *Cancer cell* 31 (2), 172–179. doi:10.1016/j.ccell.2017.01.002

Bianchi, D. W., and Chiu, R. W. K. (2018). Sequencing of Circulating Cell-free DNA during Pregnancy. *New Engl. J. Med.* 379 (5), 464–473. doi:10.1056/NEJMra1705345

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

Choueiri, T. K., and Motzer, R. J. (2017). Systemic Therapy for Metastatic Renal-Cell Carcinoma. *New Engl. J. Med.* 376 (4), 354–366. doi:10.1056/NEJMra1601333

Cohen, J. D., Douville, C., Dudley, J. C., Mog, B. J., Popoli, M., Ptak, J., et al. (2021). Detection of Low-Frequency DNA Variants by Targeted Sequencing of the Watson and Crick Strands. *Nat. Biotechnol.* doi:10.1038/s41587-021-00900-z

De Martino, M., Klatte, T., Haitel, A., and Marberger, M. (2012). Serum Cell-free DNA in Renal Cell Carcinoma: a Diagnostic and Prognostic Marker. *Cancer* 118 (1), 82–90. doi:10.1002/cncr.26254

Di Meo, A., Bartlett, J., Cheng, Y., Pasic, M. D., and Yousef, G. M. (2017). Liquid Biopsy: a Step Forward towards Precision Medicine in Urologic Malignancies. *Mol. Cancer* 16 (1), 80. doi:10.1186/s12943-017-0644-5

Ellinger, J., Muller, D. C., Muller, S. C., Hauser, S., Heukamp, L. C., von Ruecker, A., et al. (2012). Circulating Mitochondrial DNA in Serum: a Universal Diagnostic Biomarker for Patients with Urological Malignancies. *Urol. Oncol.* 30 (4), 509–515. doi:10.1016/j.urolonc.2010.03.004

Hauser, S., Zahalka, T., Ellinger, J., Fechner, G., Heukamp, L. C., Ruecker, A. V., et al. (2010). Cell-free Circulating DNA: Diagnostic Value in Patients with Renal Cell Cancer. *Anticancer Res.* 30 (7), 2785–2789.

Husain, H., and Velculescu, V. E. (2017). Cancer DNA in the Circulation: The Liquid Biopsy. *Jama* 318 (13), 1272–1274. doi:10.1001/jama.2017.12131

Jiang, P., and Lo, Y. M. D. (2016). The Long and Short of Circulating Cell-free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genetics: TIG* 32 (6), 360–371. doi:10.1016/j.tig.2016.03.009

Lasseter, K., Nassar, A. H., Hamieh, L., Berchuck, J. E., Nuzzo, P. V., and Korthauer, K. (2020). Plasma Cell-free DNA Variant Analysis Compared with Methylated DNA Analysis in Renal Cell Carcinoma. *Genet. Med.* 22 (8), 1366–1373. doi:10.1038/s41436-020-0801-x

Li, G., Zhao, A., Peoch, M., Cottier, M., and Mottet, N. (2017). Detection of Urinary Cell-free miR-210 as a Potential Tool of Liquid Biopsy for clear Cell Renal Cell Carcinoma. *Urol. Oncol.* 35 (5), 294–299. doi:10.1016/j.urolonc.2016.12.007

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernandez-Pello, S., et al. (2019). European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur. Urol.* 75 (5), 799–810. doi:10.1016/j.eururo.2019.02.011

Lu, H., Busch, J., Jung, M., Rabenhorst, S., Ralla, B., Kilic, E., et al. (2016). Diagnostic and Prognostic Potential of Circulating Cell-free Genomic and Mitochondrial DNA Fragments in clear Cell Renal Cell Carcinoma Patients. *Clinica Chim. Acta Int. J. Clin. Chem.* 452, 109–119. doi:10.1016/j.cca.2015.11.009

Lughezzani, G., Jeldres, C., Isbarn, H., Perrotte, P., Shariat, S. F., Sun, M., et al. (2009). Tumor Size Is a Determinant of the Rate of Stage T1 Renal Cell Cancer Synchronous Metastasis. *J. Urol.* 182 (4), 1287–1293. doi:10.1016/j.juro.2009.06.018

Mandel, P. M. P. (1948). Les acides nucléiques du plasma sanguin chez l'homme. *C R. Seances Soc. Biol. Fil* 142, 241–243.

Qiu, M., Wang, J., Xu, Y., Ding, X., Li, M., Jiang, F., et al. (2015). Circulating Tumor DNA Is Effective for the Detection of EGFR Mutation in Non-small Cell Lung Cancer: a Meta-Analysis. *Cancer Epidemiol. Biomarkers Prev.* 24. 206–212. doi:10.1158/1055-9965.EPI-14-0895

Remon, J., Caramella, C., Jovelet, C., Lacroix, L., Lawson, A., Smalley, S., et al. (2017). Osimertinib Benefit in EGFR-Mutant NSCLC Patients with T790M-Mutation Detected by Circulating Tumour DNA. *Ann. Oncol. official J. Eur. Soc. Med. Oncol.* 28 (4), 784–790. doi:10.1093/annonc/mdx017

Salinas-Sanchez, A. S., Garcia-Olmo, D. C., Martinez-Sanchiz, C., Picazo-Martínez, M. G., Giménez-Bachs, J. M., Flores-Bautista, A. B., et al. (2021). *Transl Oncol.* 14 (2). doi:10.1016/j.tranon.2020.100999

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics. *CA: a Cancer J. clinicians* 69 (1), 7–34. doi:10.3322/caac.21551

Siravegna, G., Marsoni, S., Siena, S., and Bardelli, A. (2017). Integrating Liquid Biopsies into the Management of Cancer. *Nat. Rev. Clin. Oncol.* 14 (9), 531–548. doi:10.1038/nrclinonc.2017.14

Stewart, C. M., Kothari, P. D., Mouliere, F., Mair, R., Somnay, S., Benayed, R., et al. (2018). The Value of Cell-free DNA for Molecular Pathology. *J. Pathol.* 244 (5), 616–627. doi:10.1002/path.5048

van Ginkel, J. H., van den Broek, D. A., van Kuik, J., Linders, D., de Weger, R., Willems, S. M., et al. (2017). Preanalytical Blood Sample Workup for Cell-free DNA Analysis Using Droplet Digital PCR for Future Molecular Cancer Diagnostics. *Cancer Med.* 6 (10), 2297–2307. doi:10.1002/cam4.1184

Wan, J., Zhu, L., Jiang, Z., and Cheng, K. (2013). Monitoring of Plasma Cell-free DNA in Predicting Postoperative Recurrence of clear Cell Renal Cell Carcinoma. *Urologia internationalis* 91 (3), 273–278. doi:10.1159/000351409

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., et al. (2011). QUADAS-2: a Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* 155 (8), 529–536. doi:10.7326/0003-4819-155-8-201110180

Yamamoto, Y., Uemura, M., Nakano, K., Hayashi, Y., Wang, C., Ishizuya, Y., et al. (2018). Increased Level and Fragmentation of Plasma Circulating Cell-free DNA Are Diagnostic and Prognostic Markers for Renal Cell Carcinoma. *Oncotarget* 9 (29), 20467–20475. doi:10.18632/oncotarget.24943

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 (5), e001942. doi:10.1136/jitc-2020-001942

Zhang, W., Ni, M., Su, Y., Wang, H., Zhu, S., Zhao, A., et al. (2018). MicroRNAs in Serum Exosomes as Potential Biomarkers in Clear-cell Renal Cell Carcinoma. *Eur. Urol. focus* 4 (3), 412–419. doi:10.1016/j.euf.2016.09.007

Zhao, A., Peoc'h, M., Cottier, M., Genin, C., Mottet, N., and Li, G. (2015). Cell-free RNA Content in Urine as a Possible Molecular Diagnostic Tool for clear Cell Renal Cell Carcinoma. *Int. J. Cancer* 136 (11), 2610–2615. doi:10.1002/ijc.29313

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 Xu, Jiang, Yu, Lou, Song, Xu, Cui, Zeng, Wang, Ma, Wang, Zhu, Li and Zhao. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/fmolb.2021.687444

The Oncogenic Roles of JC Virus T Antigen in Breast Carcinogenesis

Hua-chuan Zheng[1], Ying E[2], Zheng-guo Cui[3], Shuang Zhao[1] and Yong Zhang[4]*

[1]Department of Oncology and Experimental Center, The Affiliated Hospital of Chengde Medical University, Chengde, China, [2]Department of Oncology, Liaoning Cancer Hospital, Shenyang, China, [3]Department of Environmental Health, University of Fukui School of Medical Science, Fukui, Japan, [4]Department of Pathology, Liaoning Cancer Hospital, Shenyang, China

Purpose: JC virus (JCV) infects 80–90% of the population and results in progressive multifocal leukoencephalopathy upon immunodeficiency. The study aimed to pathologically clarify the oncogenic roles of T antigen in human breast cancers.

Methods: Breast cancer, dysplasia, and normal tissues were examined for *T antigen* of JCV by nested and real-time PCR. The positive rate or copy number of T antigen was compared with clinicopathological parameters of breast cancer. JCV existence was morphologically detected by immunohistochemistry and *in situ* PCR. T antigen was examined by Western blot using frozen samples of breast cancer and paired normal tissues.

Results: According to nested PCR, the positive rate of breast ductal or lobular carcinoma was lower than that of normal tissue ($p < 0.05$). T antigen existence was negatively correlated with E-cadherin expression and triple-negative breast cancer ($p < 0.05$), but positively correlated with lymph node metastasis and estrogen receptor and progestogen receptor expression ($p < 0.05$). Quantitative PCR showed that JCV copies were gradually decreased from normal, dysplasia to cancer tissues ($p < 0.05$). JCV T antigen copy number was lower in ductal adenocarcinoma than in normal tissue ($p < 0.05$), in line with in situ PCR and immunohistochemistry. JCV copies were negatively correlated with tumor size and E-cadherin expression ($p < 0.05$), but positively correlated with G grading of breast cancer ($p < 0.05$). Western blot also indicated weaker T antigen expression in breast cancer than normal tissues ($p < 0.05$).

Conclusion: JCV T antigen might play an important role in breast carcinogenesis. It can be employed as a molecular marker for the differentiation and aggressive behaviors of breast cancer.

Keywords: JC virus T antigen, oncogenesis, breast cancer, dysplasia, pathological behaviors

OPEN ACCESS

Edited by:
Tianhui Chen,
Zhejiang Cancer Hospital (CAS), China

Reviewed by:
Vikram Srivastava,
Iowa State University, United States
Peng Huang,
Okayama University, Japan

****Correspondence:***
Hua-chuan Zheng
zheng_huachuan@hotmail.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *29 March 2021*
Accepted: *13 July 2021*
Published: *12 August 2021*

Citation:
Zheng H, E Y, Cui Z, Zhao S and Zhang Y (2021) The Oncogenic Roles of JC Virus T Antigen in Breast Carcinogenesis. Front. Mol. Biosci. 8:687444. doi: 10.3389/fmolb.2021.687444

INTRODUCTION

Breast cancer is the leading cause of cancer death among women, even in less developed countries (Coughlin, 2019). Environmental, hereditary, and genetic backgrounds are considered as the most important factors for carcinogenesis and the subsequent progression of breast cancer. Among environmental factors, viral infection is associated with about 15–20% of all cancers (FahadUllah, 2019).

JC virus (JCV) is a member of the polyomavirus family (Assetta and Atwood, 2017) and has a circular and double-stranded DNA genome. The early region encodes T antigen (Reiss and Khalili, 2003), a large phosphoprotein that binds to the viral replication region to promote double helix unwinding and recruitment of DNA synthesis proteins. The late region encodes the capsid structural proteins VP1, VP2 and, VP3 and agnoprotein (Reiss and Khalili, 2003; Saxena et al., 2021). Serologically, there is asymptomatic JCV infection in 80–90% of the adult population (Ahye et al., 2020). JCV enters the human body through both digestive and respiratory tracts and persists quiescent in the kidney and lymphoid tissues (White and Khalili, 2005; Delbue et al., 2017; Dwyer et al., 2021). However, it may be activated under immunosuppressive conditions, resulting in progressive multifocal leukoencephalopathy (PML) (Reiss and Khalili, 2003; Delbue et al., 2017; Ahye et al., 2020; Dwyer et al., 2021; Saxena et al., 2021). During permissive infection, replication of viral DNA can cause lytic infection, but in non-permissive cells, either abortive infection or cell transformation is the outcome (Delbue et al., 2017; Ahye et al., 2020; Saxena et al., 2021). JCV can transform cells, which display rapid growth and division, prolongation of lifespan, anchorage-dependent growth, chromosomal instability, and dense foci formation in culture (Eash et al., 2004; White and Khalili, 2005).

Intravenous or intracranial inoculation of JCV has been found to cause astrocytomas, glioblastomas, neuroblastomas, and medulloblastomas (Miller et al., 1984; Hayashi et al., 2001). In addition, transgenic mice expressing T antigen developed pituitary adenomas or malignant peripheral nerve sheath tumors (London et al., 1978; Del Valle and Khalili, 2021). In previous studies, we established transgenic mice and found that T antigen induced lens tumors and lung cancer (Gou et al., 2015; Noguchi et al., 2013). Recently, JCV was demonstrated to correlate with colorectal, gastric, prostatic, and esophageal cancers, brain tumors, lung cancer, and B cell lymphomas (Del Valle et al., 2005; Zheng et al., 2007a; Murai et al., 2007; Kutsuna et al., 2008; Ramamoorthy et al., 2011; Shen et al., 2011; Wang et al., 2012; Noguchi et al., 2013; Anzivino et al., 2015; MalekpourAfshar et al., 2016). The oncogenesis of JCV mainly centers on T antigen. T antigen can inactivate p53 and Rb and disrupt the Wnt signaling pathway (Piña-Oviedo et al., 2006). The encoding genes of JCV are known to be preferentially expressed in astrocytes and oligodendrocytes of the human brain because these cells contain specific transcriptional factors (NF-1, Sp1, Sμbp-2, and Purα) (Piña-Oviedo et al., 2006). Reportedly, investigators have identified JCV in the liver, kidney, spleen, bone marrow, bladder, prostate, and tonsils by PCR (Zheng et al., 2009). Here, we examined the existence and copy number of JCV in breast cancer, dysplasia, and normal tissues by nested PCR, real-time PCR, immunohistochemistry, and *in situ* PCR and compared JCV existence with clinicopathological parameters of breast cancer.

MATERIALS AND METHODS

Subjects

Formalin-fixed and paraffin-embedded breast cancer (n = 112), breast dysplasia (n = 18), and normal breast (n = 48) tissues were sampled from surgical materials in The Affiliated Hospital of Chengde Medical University and Liaoning Cancer Hospital. Fresh samples of breast cancer and paired normal tissues (n = 10) were collected from both hospitals. The cancer patients did not receive a neoadjuvant before surgical operation. They signed informed consent. The ethics committees of these hospitals approved the study.

DNA Extraction and Checking

Paraffin-embedded blocks were incised into 10-μm-thick sections, microdissected under the guidance of HE slides, and subjected to deparaffinization and rehydration. DNA was extracted by the traditional proteinase K/phenol/chloroform method. To check DNA integrity, we amplified tissue DNA by targeting β-globin: Forward; 5′-ACACAACTGTGTTCACTAGC-3′; backward; 5′-GTCTCCTTAAACCTGTCTTG-3′. PCR condition was described as follows: 30 cycles of denaturation at 95°C for 20s, annealing at 55°C for 35 s, and extension at 72°C for 20 s. DNA-free amplification was employed as a negative control.

Nested PCR

PCR was carried out by targeting *T antigen*. T1 (5′-TGGCCTG TAAAGTTCTAGG CA -3′ and T2 (5′-GCAGAGTCAAGGGAT TTACCTTC-3′) primers were used for the 1st PCR, whereas T1 and T3 (5′-AGCAACCTTGATTGCTTAAGAGA-3′) were used for the 2nd PCR. The 20 μl of reaction mixtures contained 0.1 μl of Ex Taq HS (TaKaRa) with 2.0 mM of $MgCl_2$, 2.0 μl × 10 PCR buffer, 2.0 μl of dNTP mixture, 1 μM of each primer, and 200 ng of DNA. The PCR process was 32 cycles of denaturation at 95°C for 20 s, annealing at 56°C for 20 s, and extension at 72°C for 20 s. Nested PCR was carried out using 1% (volume) of the first amplicon. DNA-free amplification was employed as a negative control.

Real-Time PCR

SYBR fluorescence PCR was used to quantify JCV copies using the Bio-Rad PCR system. A plasmid of PBS-JCV T antigen was serially diluted for standard reference. These standard and sample DNAs were amplified by targeting *T antigen*: Forward: 5′-GCC ACCCCAGCCATATATTG-3′ and backward: 5′-GTTGACAGT ATCCATATGACCAGAGAA-3′. In total, the 20 μl reaction mixture contained 10.0 μl of TaqMan® (× 2) with 1.8 μl (10 μM) of each primer and 80 ng of DNA. The PCR protocol was 55 cycles of denaturation at 95°C for 25 s, annealing at 55°C for 50 s, and extension at 72°C or 25 s.

Immunohistochemistry

Serial sections were deparaffinized with xylene, rehydrated with alcohol, and subjected to antigen retrieval by irradiation in target retrieval solution (TRS, DAKO, CA, United States) for 5 min with a microwave oven. Five percent BSA was then applied to incubation for 5 min to prevent non-specific binding. The sections were incubated with anti-SV 40 T antigen antibody (Calbiochem, United States; 1:100) for 20 min and then treated with the anti-mouse Envision-PO (1:100, DAKO, CA, United States) antibody for 20 min. Incubation was performed in a microwave oven for intermittent irradiation as described previously (Gou et al., 2015). After each treatment, the slides were

washed with TBST (10 mM of Tris-HCl, 150 mM of NaCl, 0.1% Tween 20) three times for 5 min. All slides were colored with 3, 3′-diaminobenzidine and counterstained with Mayer's hematoxylin. Omission of the primary antibody was used as a negative control.

In Situ PCR

A 10-μm-thick section was prepared with proteinase K for 15 min. After TBS washing, the tissue slide was fixed with 4% paraformaldehyde and then washed with 2 × SSC. After that, 100 μl of PCR mixture (0.2 μM of primers, 0.125 nM of digoxigenin-11-dUTP, 2.5 mM of $MgCl_2$, 1 × PCR buffer, 6.25 U of Taq polymerase) was put into a membrane, sealing on the tissue. PCR amplification was performed at the condition: Denaturation at 94°C for 3 min, followed by 20 cycles of 92°C for 15 s, 55°C for 20 s, and 72°C for 30 s and final extension at 72°C for 7 min. The primers were forward: 5′-AGGTAGGCCTTTGGTCTAA-3′ and backward: 5′-TGCCTAGAACTTTACAGG-3′. After that, the tissue was washed with 2 × SSC and incubated with blocking solution (100 μg/ml of Salmon testis DNA, 100 μg/ml of yeast tRNA, and 5% BSA) for 1 h. Subsequently, the sections were reacted with anti-digoxigenin and AP (alkaline phosphatase)-conjugated antibody (Roche, 1:500) for 90 min. After being washed for 5 min and immersed in solution II (100 mM of Tris-HCl, pH 9.5, 100 mM of NaCl, and 50 mM of $MgCl_2$) for 15 min, the positive signal was colored using NBP/BCIP. Finally, methyl green was used for counterstaining.

Western Blot

We homogenized breast cancer and normal tissues in RIPA lysis buffer, and a protein assay was performed using Kaumas brilliant blue. A 50 μg protein/sample was subjected to 10% SDS-PAGE electrophoresis and electrically transferred to a PVDF membrane, which was incubated with 5% bovine serum albumin (BSA) in TBST, and then with anti-SV40 T antigen (1:200; Santa Cruz) or rabbit anti-GAPDH (1:2,000, CST) antibody. The membranes were washed with TBST and incubated with anti-rabbit or anti-mouse HRP-conjugated secondary antibody (DAKO, 1:1,000). Bands were visualized with Azure Biosystem C300 by an ECL detection kit.

Statistical Analysis

SPSS v. 26.0 software was used for statistical analysis. Statistical analysis was carried out using Fisher's test for the comparison of positive rates and Mann–Whitney *U* for the comparison of means. A p value < 0.05 was statistically regarded as significant.

RESULTS

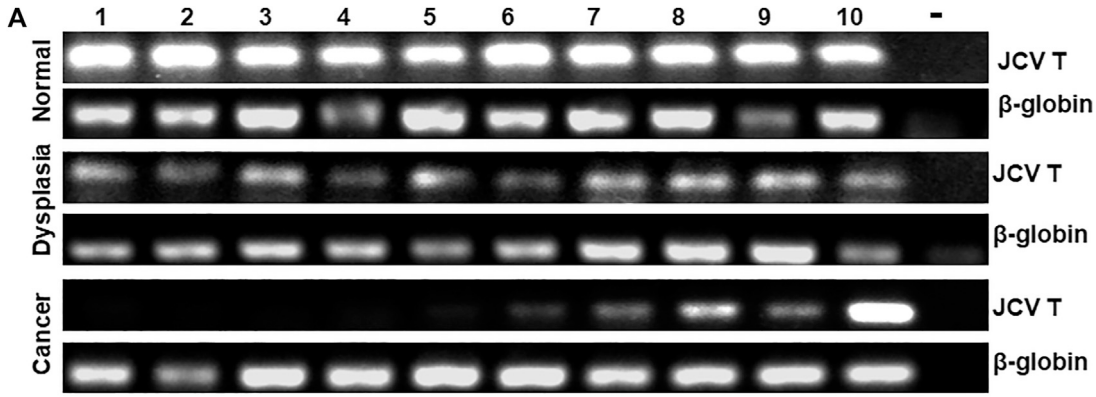
As indicated in **Figure 1A**, we used general PCR of β-globin as quality control. We observed positive bands in all samples. Nested PCR indicated that the positive rate of T antigen was lower in breast ductal or lobular carcinoma than normal tissues (**Figure 1B**, $p < 0.05$). As shown in **Table 1**, T antigen existence was negatively correlated with E-cadherin expression (64.3 vs. 30.3%, $p < 0.05$) and triple-negative breast cancer (TNBC, 52.8 vs. 18.2%, $p < 0.05$), but positively correlated with lymph node metastasis (42.9 vs. 64.7%, $p < 0.05$), estrogen receptor (ER) expression (25.0 vs. 55.0%, $p < 0.05$), and progestogen receptor (PR) expression (26.1 vs. 55.8%, $p < 0.05$).

Real-time PCR showed that T antigen copies were gradually decreased from normal, dysplasia to cancer tissues (**Figure 2A**, $p < 0.05$). They were lower in ductal adenocarcinoma than in normal tissues (**Figure 2B**, $p < 0.05$). They were negatively correlated with tumor size (**Figure 2C**, $p < 0.05$) and E-cadherin expression (**Figure 2E**) and positively correlated with G grading of breast cancer (**Figure 2D**, $p < 0.05$). *In situ* PCR demonstrated that positive cells were detectable in breast ductal and lobular epithelium, but no or weak signal was seen in breast cancer, in line with immunohistochemistry (**Figure 3**). Meanwhile, a weaker T antigen expression was seen in breast cancer than that in paired normal tissues by Western blot (**Figure 4**).

DISCUSSION

Cellular malignant transformation needs a series of genetic or epigenetic accumulation, including oncogene activation or overexpression (Del Valle and Khalili, 2021). Although JCV is a highly neurotropic virus and induces brain tumors, JCV DNA was discovered in the respiratory and the upper and lower gastrointestinal tracts (Ahye et al., 2020). Furthermore, detection of JCV in renal tubules and tonsil lymphocytes indicated JCV persistence in a quiescent state of these cells during latency and infection of other cells upon immune suppression (Zheng et al., 2007b). Upon its entrance into cells, T antigen can inactivate nuclear p53 and pRb proteins to disrupt the cell cycle (Khalili et al., 2006). It can also suppress the Wnt pathway *via* β-catenin degradation and disrupt cellular IGF-IR signaling pathways (Reviriego-Mendoza and Frisque, 2011). Additionally, T antigen decreased Bag-3 expression for apoptotic suppression by blocking the interaction of AP2 with the Bag3 promoter (Sariyer et al., 2012).

Guillory et al. (2010) established a breast cancer model of the polyoma middle T antigen transgenic mouse. Tzeng et al. (1993) injected whey acidic protein (WAP)-SV-T antigen DNA into fertilized mouse eggs and found that female mice developed breast cancer with high frequency. It was very interesting that the inactivation of PTEN in CK19-positive cells caused triple-negative breast lobular carcinoma (Zhao et al., 2017). Reddi et al. (2019) reported that diffuse large B cell lymphoma appeared secondary to JCV in PML. Querido et al. (2020)observed high-grade urothelial carcinoma in a kidney transplant recipient after JCV-related nephropathy. Malhotra et al. (2016) found that JCV seropositivity was positively associated with a high lung cancer risk in the non-smoking population. Many investigators did not detect JCV in breast cancer tissue samples (Hogan et al., 1983; Antonsson et al., 2012; Dowran et al., 2019). However, Hachana et al. (2012) detected that JCV T antigen DNA was in invasive ductal carcinomas (28/112, 25.0%) but not in invasive lobular and medullary carcinomas. In this study, we found that JCV T antigen existence was gradually decreased during breast carcinogenesis at both the DNA and protein level and negatively correlated with T staging of breast cancer. No differences in T antigen DNA and protein were found in ductal adenocarcinomas or lobular adenocarcinomas. This indicated that it might be difficult



B

	n	Positive	Negative	PR (%)	p value
Normal breast	48	38	10	79.2	
Dysplasia	18	14	4	77.8	
Ductal cancer	68	31	37	45.6	<0.001
Lobular cancer	44	21	23	47.7	0.002

FIGURE 1 | Detection of JCV T antigen in breast carcinogenesis. β-globin was positive in all cases of normal tissue, dysplasia, and cancer of the breast. T antigen was positive in some cancer cases by nested PCR **(A)**. The positive rates of JCV T antigen were compared between cancer, dysplasia, and normal tissue **(B)**. -, negative control; PR, positive rate.

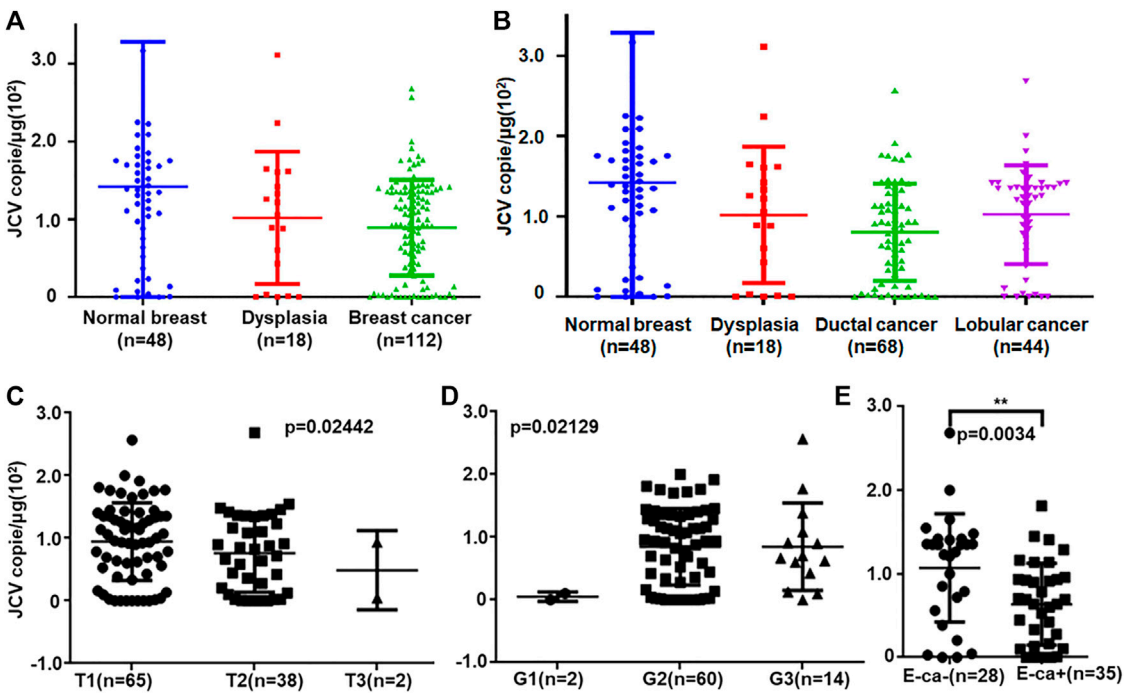


FIGURE 2 | JCV T antigen loads in breast cancers. T antigen copies were measured and calculated in normal, dysplasia, and cancer tissues by real-time PCR **(A)**. The copies were analyzed between ductal and lobular cancers **(B)**. They were also compared with T staging **(C)**, G grading **(D),** and E-cadherin expression **(E)**. E-cad, E-cadherin expression.

TABLE 1 | The correlation between JCV T antigen existence and clinicopathological features of breast cancer.

Parameters	n	JCV T antigen		χ2	*p* value
		-	+		
N staging	97			4.217	**0.040**
-		36 (57.1%)	27 (42.9%)		
+		12 (35.3%)	22 (64.7%)		
E-cadherin expression	61			7.044	**0.008**
-		10 (35.7%)	18 (64.3%)		
+		23 (69.7%)	10 (30.3%)		
ER expression	100			5.762	**0.016**
-		15 (75.0%)	5 (25.0%)		
+		36 (45.0%)	44 (55.0%)		
PR expression	100			6.275	**0.012**
-		17 (73.9%)	6 (26.1%)		
+		34 (44.2%)	43 (55.8%)		
TNBC	101			4.697	**0.030**
-		42 (47.2%)	47 (52.8%)		
+		9 (81.8%)	2 (18.2%)		

ER, estrogen receptor; PR, progesterone receptor; TNBC, triple-negative breast cancer.

for JCV infection, genomic insertion, and subsequent translation of T antigen during breast carcinogenesis and subsequent growth. However, it also cannot deny the oncogenic role of T antigen in breast cancer because it can be detected in breast cancer and its overexpression can induce breast cancer. In line with Hachana's report (Hachana et al., 2012), we also found that JCV DNA presence correlated with TNBC. Thereby, it was believed that T antigen might be closely linked to the tumorigenesis of TNBC.

JCV T antigen was involved in colorectal carcinogenesis and liver metastasis (Sinagra et al., 2014; Shoraka et al., 2020; Vilkin and Niv, 2011). Reportedly, JCV T antigen stabilized β-catenin for its nuclear translocation to initiate cancer proliferation and development (Nosho et al., 2009). Our previous study has demonstrated that lung cancers with higher JCV copy numbers displayed high proliferation and downregulation of cell adhesion, mediated by membrane β-catenin (Zheng et al., 2007a). Donadoni et al. (2018) found that mouse T antigen-overexpressing medulloblastoma cells had survival capacity, radiation resistance, a high colony formation, and a strong double-strand DNA break repair. Noch et al. (2012) demonstrated that T antigen promoted the expression of hexokinase 2 and the pentose phosphate enzyme, transaldolase-1 for glycolysis, and pentose catabolism in medulloblastoma cells. Ksiaa et al. (2010) found that JCV presence was correlated with the patient's age and differentiation and abnormal methylation of tumor suppressor genes of gastric cancer. Here, we found that JCV T antigen was positively correlated with G grading, N staging, E-cadherin hypoexpression, and non-TNBC, suggesting that T antigen existence might be employed to indicate poor differentiation and low E-cadherin-mediated metastasis of breast cancer.

If a virus plays an oncogenic role, it must infect the cells and then encode the oncogenic proteins to disrupt the cell function. According to our findings, we found that JCV copies were different according to tissue types (stomach < lung < breast) because the distinct distribution of its receptors (α 2, 6-linked sialic acid and serotonin) determined its different infection (Eash et al., 2004; Elphick et al., 2004; Gee et al., 2006). Geoghegan et al. (2017) showed that non-sialylated glycosaminoglycans served as alternative attachment receptors for the infection of JCV. Nukuzuma et al. (2016) found the suppressive effect of topoisomerase I inhibitors topotecan and β-lapachone on JCV propagation in human neuroblastoma cells. Adipocyte plasma membrane protein, PI3Kγ, and its regulatory subunit PIK3R5 promoted JCV infection in human glial cells (Clark et al., 2020; Haley et al.,

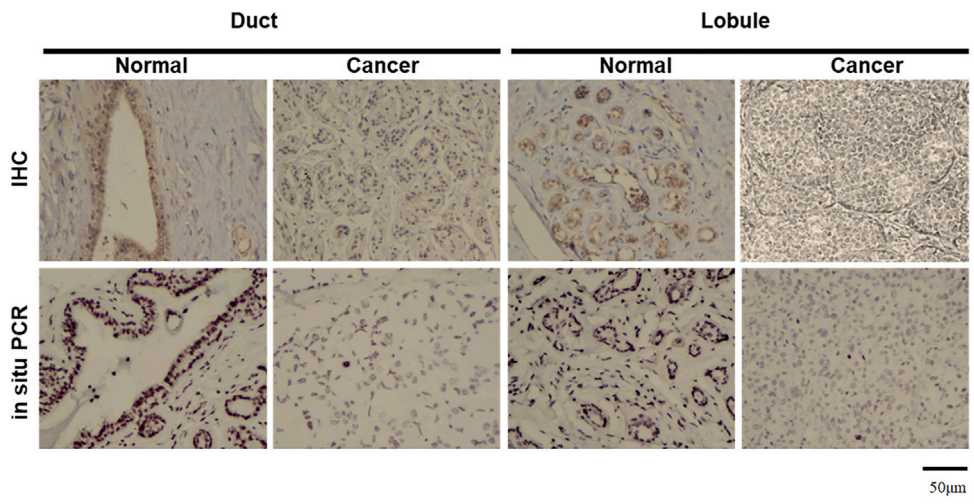


FIGURE 3 | Morphological examination of JCV T antigen in breast cancers. According to immunohistochemistry (IHC), T antigen was strongly expressed in ductal and lobular epithelial cells (brown) but positively or weakly expressed in breast ductal and lobular cancer cells. NBT/BCIP coloring is displayed as black to show positive signals in breast ductal and lobular epithelial cells by *in situ* PCR, while methyl green is denoted as a green color for counterstaining.

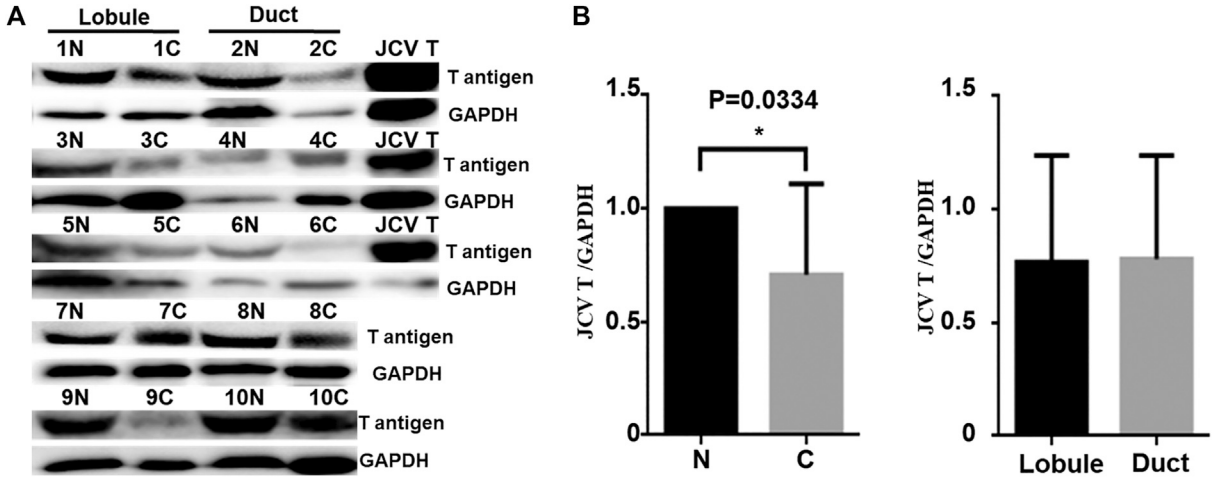


FIGURE 4 | JCV T antigen expression in human breast cancers. T antigen was screened by Western blot **(A)** and subsequently analyzed by densitometric analysis **(B)**. N, normal; C, cancer. Positive control, spontaneous breast cancer from JCV T antigen transgenic mouse.

2020). Uleri et al. (2011) showed that alternative splicing factor, SF2/ASF, negatively regulated transcription and splicing of JCV genes in glial cells. SF2/ASF hyperexpression induced growth and proliferation of JCV-transformed tumor cells. Either endogenous or ectopic LIP expression mediated the degradation of T antigen in a JCV-transgenic mouse tumor cell line (Bellizzi et al., 2015). According to our knowledge, the partner proteins of T antigen might be involved in the cell specificity of JCV T antigen, which will be confirmed in the future. Finally, estrogen and proliferation of lobular glands during breeding might increase the risk of genetic breast cancer.

In conclusion, the JCV T antigen might play an important role in breast carcinogenesis. JCV infection, insertion, transcription, translation, degradation, target, and partner chaperons of its T antigen might underlie the molecular mechanisms of its tissue-specific carcinogenesis.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Affiliated Hospital of Chengde Medical University and Liaoning Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HZ contributed to the conception of the study, performed the data analyses, and wrote the manuscript; SZ and YZ performed the experiment; YE and ZC contributed significantly to analysis and manuscript preparation.

FUNDING

This study was supported by the Award for Liaoning Distinguished Professor, Natural Science Foundation of Hebei Province (202130821010436) and the National Natural Scientific Foundation of China (grant no. 81672700).

REFERENCES

Ahye, N., Bellizzi, A., May, D., and Wollebo, H. S. (2020). The Role of the JC Virus in central Nervous System Tumorigenesis. *Int. J. Mol. Sci.* 21, 6236. doi:10.3390/ijms21176236

Antonsson, A., Bialasiewicz, S., Rockett, R. J., Jacob, K., Bennett, I. C., and Sloots, T. P. (2012). Exploring the Prevalence of Ten Polyomaviruses and Two Herpes Viruses in Breast Cancer. *PLoS One* 7, e39842. doi:10.1371/journal.pone.0039842

Anzivino, E., Rodio, D. M., Mischitelli, M., Bellizzi, A., Sciarra, A., Salciccia, S., et al. (2015). High Frequency of JCV DNA Detection in Prostate Cancer Tissues. *Cancer Genomics Proteomics* 12, 189–200.

Assetta, B., and Atwood, W. J. (2017). The Biology of JC Polyomavirus. *Biol. Chem.* 398, 839–855. doi:10.1515/hsz-2016-0345

Bellizzi, A., White, M. K., and Wollebo, H. S. (2015). Degradation of Polyomavirus JC T-Antigen by Stress Involves the LIP Isoform of C/EBP. *Cell Cycle* 14, 2075–2079. doi:10.1080/15384101.2015.1042631

Clark, P., Gee, G. V., Albright, B. S., Assetta, B., Han, Y., Atwood, W. J., et al. (2020). Phosphoinositide 3′-Kinase γ Facilitates Polyomavirus Infection. *Viruses* 12, 1190. doi:10.3390/v12101190

Coughlin, S. S. (2019). Epidemiology of Breast Cancer in Women. *Adv. Exp. Med. Biol.* 1152, 9–29. doi:10.1007/978-3-030-20301-6_2

Del Valle, L., and Khalili, K. (2021). Induction of Brain Tumors by the Archetype Strain of Human Neurotropic JCPyV in a Transgenic Mouse Model. *Viruses* 13, 162. doi:10.3390/v13020162

Del Valle, L., White, M. K., Enam, S., Oviedo, S. P. a., Bromer, M. Q., Thomas, R. M., et al. (2005). Detection of JC Virus DNA Sequences and Expression of Viral T Antigen and Agnoprotein in Esophageal Carcinoma. *Cancer* 103, 516–527. doi:10.1002/cncr.20806

Delbue, S., Comar, M., and Ferrante, P. (2017). Review on the Role of the Human Polyomavirus JC in the Development of Tumors. *Infect. Agent Cancer* 12, 10. doi:10.1186/s13027-017-0122-0

Donadoni, M., Sariyer, R., Wollebo, H., Bellizzi, A., and Kudret Sariyer, I. (2018). Viral Tumor Antigen Expression Is No Longer Required in Radiation-Resistant Subpopulation of JCV Induced Mouse Medulloblastoma Cells. *Genes Cancer* 9, 130–141. doi:10.18632/genesandcancer.174

Dowran, R., Joharinia, N., Safaei, A., Bakhtiyarizadeh, S., Alidadi Soleimani, A., Alizadeh, R., et al. (2019). No Detection of EBV, BKV and JCV in Breast Cancer Tissue Samples in Iran. *BMC Res. Notes* 12, 171. doi:10.1186/s13104-019-4178-3

Dwyer, C. M., Jokubaitis, V. G., Stankovich, J., Baker, J., Haartsen, J., Butzkueven, H., et al. (2021). High Rates of JCV Seroconversion in a Large International Cohort of Natalizumab-Treated Patients. *Ther. Adv. Neurol. Disord.* 14, 1756286421998915. doi:10.1177/1756286421998915

Eash, S., Tavares, R., Stopa, E. G., Robbins, S. H., Brossay, L., and Atwood, W. J. (2004). Differential Distribution of the JC Virus Receptor-type Sialic Acid in normal Human Tissues. *Am. J. Pathol.* 164, 419–428. doi:10.1016/s0002-9440(10)63132-x

Elphick, G. F., Querbes, W., Jordan, J. A., Gee, G. V., Eash, S., Manley, K., et al. (2004). The Human Polyomavirus, JCV, Uses Serotonin Receptors to Infect Cells. *Science* 306, 1380–1383. doi:10.1126/science.1103492

Fahad Ullah, M. (2019). Breast Cancer: Current Perspectives on the Disease Status. *Adv. Exp. Med. Biol.* 1152, 51–64. doi:10.1007/978-3-030-20301-6_4

Gee, G. V., Dugan, A. S., Tsomaia, N., Mierke, D. F., and Atwood, W. J. (2006). The Role of Sialic Acid in Human Polyomavirus Infections. *Glycoconj J.* 23, 19–26. doi:10.1007/s10719-006-5434-z

Geoghegan, E. M., Pastrana, D. V., Schowalter, R. M., Ray, U., Gao, W., Ho, M., et al. (2017). Infectious Entry and Neutralization of Pathogenic JC Polyomaviruses. *Cel Rep.* 21, 1169–1179. doi:10.1016/j.celrep.2017.10.027

Gou, W.-F., Zhao, S., Shen, D.-F., Yang, X.-F., Liu, Y.-P., Sun, H.-Z., et al. (2015). The Oncogenic Role of JC Virus T Antigen in Lens Tumors without Cell Specificity of Alternative Splicing of its Intron. *Oncotarget* 6, 8036–8045. doi:10.18632/oncotarget.3507

Guillory, B., Sakwe, A. M., Saria, M., Thompson, P., Adhiambo, C., Koumangoye, R., et al. (2010). Lack of Fetuin-A (α2-HS-Glycoprotein) Reduces Mammary Tumor Incidence and Prolongs Tumor Latency via the Transforming Growth Factor-β Signaling Pathway in a Mouse Model of Breast Cancer. *Am. J. Pathol.* 177, 2635–2644. doi:10.2353/ajpath.2010.100177

Hachana, M., Amara, K., Ziadi, S., Gacem, R. B., Korbi, S., and Trimeche, M. (2012). Investigation of Human JC and BK Polyomaviruses in Breast Carcinomas. *Breast Cancer Res. Treat.* 133, 969–977. doi:10.1007/s10549-011-1876-5

Haley, S. A., O'Hara, B. A., and Atwood, W. J. (2020). Adipocyte Plasma Membrane Protein (APMAP) Promotes JC Virus (JCPyV) Infection in Human Glial Cells. *Virology* 548, 17–24. doi:10.1016/j.virol.2020.06.002

Hayashi, H., Endo, S., Suzuki, S., Tanaka, S., Sawa, H., Ozaki, Y., et al. (2001). JC Virus Large T Protein Transforms Rodent Cells but Is Not Involved in Human Medulloblastoma. *Neuropathology* 21, 129–137. doi:10.1046/j.1440-1789.2001.00384.x

Hogan, T. F., Padgett, B. L., Walker, D. L., Borden, E. C., and Frias, Z. (1983). Survey of Human Polyomavirus (JCV, BKV) Infections in 139 Patients with Lung Cancer, Breast Cancer, Melanoma, or Lymphoma. *Prog. Clin. Biol. Res.* 105, 311–324.

Khalili, K., Gordon, J., and White, M. K. (2006). The Polyomavirus, JCV and its Involvement in Human Disease. *Adv. Exp. Med. Biol.* 577, 274–287. doi:10.1007/0-387-32957-9_20

Ksiaa, F., Ziadi, S., Mokni, M., Korbi, S., and Trimeche, M. (2010). The Presence of JC Virus in Gastric Carcinomas Correlates with Patient's Age, Intestinal Histological Type and Aberrant Methylation of Tumor Suppressor Genes. *Mod. Pathol.* 23, 522–530. doi:10.1038/modpathol.2009.184

Kutsuna, T., Zheng, H., Abdel-Aziz, H. O., Murai, Y., Tsuneyama, K., Furuta, I., et al. (2008). High JC Virus Load in Tongue Carcinomas May Be a Risk Factor for Tongue Tumorigenesis. *Virchows Arch.* 452, 405–410. doi:10.1007/s00428-007-0534-0

London, W., Houff, S., Madden, D., Fuccillo, D., Gravell, M., Wallen, W., et al. (1978). Brain Tumors in Owl Monkeys Inoculated with a Human Polyomavirus (JC Virus). *Science* 201, 1246–1249. doi:10.1126/science.211583

Malekpour Afshar, R., Mollaei, H. R., Zandi, B., and Iranpour, M. (2016). Evaluation of JC and Cytomegalo Viruses in Glioblastoma Tissue. *Asian Pac. J. Cancer Prev.* 17, 4907–4911. doi:10.22034/APJCP.2016.17.11.4907

Malhotra, J., Waterboer, T., Pawlita, M., Michel, A., Cai, Q., Zheng, W., et al. (2016). Serum Biomarkers of Polyomavirus Infection and Risk of Lung Cancer in Never Smokers. *Br. J. Cancer* 115, 1131–1139. doi:10.1038/bjc.2016.285

Miller, N. R., McKeever, P. E., London, W., Padgett, B. L., Walker, D. L., and Wallen, W. C. (1984). Brain Tumors of Owl Monkeys Inoculated with JC Virus Contain the JC Virus Genome. *J. Virol.* 49, 848–856. doi:10.1128/jvi.49.3.848-856.1984

Murai, Y., Zheng, H.-C., Aziz, H. O. A., Mei, H., Kutsuna, T., Nakanishi, Y., et al. (2007). High JC Virus Load in Gastric Cancer and Adjacent Non-cancerous Mucosa. *Cancer Sci.* 98, 25–31. doi:10.1111/j.1349-7006.2006.00354.x

Noch, E., Sariyer, I. K., Gordon, J., and Khalili, K. (2012). JC Virus T-Antigen Regulates Glucose Metabolic Pathways in Brain Tumor Cells. *PLoS One* 7, e35054. doi:10.1371/journal.pone.0035054

Noguchi, A., Kikuchi, K., Ohtsu, T., Yoshiwara, M., Nakamura, Y., Miyagi, Y., et al. (2013). Pulmonary Tumors Associated with the JC Virus T-Antigen in a Transgenic Mouse Model. *Oncol. Rep.* 30, 2603–2608. doi:10.3892/or.2013.2782

Nosho, K., Shima, K., Kure, S., Irahara, N., Baba, Y., Chen, L., et al. (2009). JC Virus T-Antigen in Colorectal Cancer Is Associated with P53 Expression and Chromosomal Instability, Independent of CpG Island Methylator Phenotype. *Neoplasia* 11, 87–95. doi:10.1593/neo.81188

Nukuzuma, S., Nakamichi, K., Kameoka, M., Sugiura, S., Nukuzuma, C., Tasaki, T., et al. (2016). Suppressive Effect of Topoisomerase Inhibitors on JC Polyomavirus Propagation in Human Neuroblastoma Cells. *Microbiol. Immunol.* 60, 253–260. doi:10.1111/1348-0421.12372

Piña-Oviedo, S., De León-Bojorge, B., Cuesta-Mejías, T., White, M. K., Ortiz-Hidalgo, C., Khalili, K., et al. (2006). Glioblastoma Multiforme with Small Cell Neuronal-like Component: Association with Human Neurotropic JC Virus. *Acta Neuropathol.* 111, 388–396. doi:10.1007/s00401-006-0050-3

Querido, S., Fernandes, I., Weigert, A., Casimiro, S., Albuquerque, C., Ramos, S., et al. (2020). High-grade Urothelial Carcinoma in a Kidney Transplant Recipient after JC Virus Nephropathy: The First Evidence of JC Virus as a Potential Oncovirus in Bladder Cancer. *Am. J. Transpl.* 20, 1188–1191. doi:10.1111/ajt.15663

Ramamoorthy, S., Devaraj, B., Miyai, K., Luo, L., Liu, Y.-T., Boland, C. R., et al. (2011). John Cunningham Virus T-Antigen Expression in Anal Carcinoma. *Cancer* 117, 2379–2385. doi:10.1002/cncr.25793

Reddi, A., Patel, N., and Morris, N. A. (2019). Diffuse Large B Cell Lymphoma Secondary to JC Virus in Progressive Multifocal Leukoencephalopathy. *J. Neurovirol.* 25, 883–886. doi:10.1007/s13365-019-00760-z

Reiss, K., and Khalili, K. (2003). Viruses and Cancer: Lessons from the Human Polyomavirus, JCV. *Oncogene* 22, 6517–6523. doi:10.1038/sj.onc.1206959

Reviriego-Mendoza, M. M., and Frisque, R. J. (2011). Interaction and Co-localization of JC Virus Large T Antigen and the F-Box Protein β-transducin-repeat Containing Protein. *Virology* 410, 119–128. doi:10.1016/j.virol.2010.10.038

Sariyer, I. K., Merabova, N., Patel, P. K., Knezevic, T., Rosati, A., Turco, M. C., et al. (2012). Bag3- Induced Autophagy Is Associated with Degradation of JCV Oncoprotein, T-Ag. *PLoS One* 7, e45000. doi:10.1371/journal.pone.0045000

Saxena, R., Saribas, S., Jadiya, P., Tomar, D., Kaminski, R., Elrod, J. W., et al. (2021). Human Neurotropic Polyomavirus, JC Virus, Agnoprotein Targets Mitochondrion and Modulates its Functions. *Virology* 553, 135–153. doi:10.1016/j.virol.2020.11.004

Shen, C.-H., Wu, J.-D., Hsu, C.-D., Jou, Y.-C., Lin, C.-T., Wang, M., et al. (2011). The High Incidence of JC Virus Infection in Urothelial Carcinoma Tissue in Taiwan. *J. Med. Virol.* 83, 2191–2199. doi:10.1002/jmv.22240

Shoraka, H. R., Abobakri, O., Naghibzade Tahami, A., Mollaei, H. R., Bagherinajad, Z., Malekpour Afshar, R., et al. (2020). Prevalence of JC and BK Viruses in Patients with Colorectal Cancer: A Systematic Review and Meta- Analysis. *Asian Pac. J. Cancer Prev.* 21, 1499–1509. doi:10.31557/apjcp.2020.21.6.1499

Sinagra, E., Raimondo, D., Gallo, E., Stella, M., Cottone, M., Orlando, A., et al. (2014). Could JC Virus Provoke Metastasis in colon Cancer?. *World J. Gastroenterol.* 20, 15745–15749. doi:10.3748/wjg.v20.i42.15745

Tzeng, Y. J., Guhl, E., Graessmann, M., and Graessmann, A. (1993). Breast Cancer Formation in Transgenic Animals Induced by the Whey Acidic Protein SV40 T Antigen (WAP-SV-T) Hybrid Gene. *Oncogene* 8, 1965–1971.

Uleri, E., Beltrami, S., Gordon, J., Dolei, A., and Sariyer, I. K. (2011). Extinction of Tumor Antigen Expression by SF2/ASF in JCV-Transformed Cells. *Genes & Cancer* 2, 728–736. doi:10.1177/1947601911424578

Vilkin, A., and Niv, Y. (2011). Association between hMLH1 Hypermethylation and JC Virus (JCV) Infection in Human Colorectal Cancer (CRC). *Clin. Epigenet* 2, 1–5. doi:10.1007/s13148-010-0013-3

Wang, J.-P., Wang, Z.-Z., Zheng, Y.-S., Xia, P., Yang, X.-H., Liu, Y.-P., et al. (2012). JC Virus Existence in Chinese Gastrointestinal Carcinomas. *Oncol. Lett.* 3, 1073–1078. doi:10.3892/ol.2012.627

White, M. K., and Khalili, K. (2005). Expression of JC Virus Regulatory Proteins in Human Cancer: Potential Mechanisms for Tumourigenesis. *Eur. J. Cancer* 41, 2537–2548. doi:10.1016/j.ejca.2005.08.019

Zhao, G.-F., Zhao, S., Liu, J.-J., Wu, J.-C., He, H.-Y., Ding, X.-Q., et al. (2017). Cytokeratin 19 Promoter Directs the Expression of Cre Recombinase in Various Epithelia of Transgenic Mice. *Oncotarget* 8, 18303–18311. doi:10.18632/oncotarget.15435

Zheng, H., Murai, Y., Hong, M., Nakanishi, Y., Nomoto, K., Masuda, S., et al. (2007). JC [corrected] Virus Detection in Human Tissue Specimens. *J. Clin. Pathol.* 60, 787–793. doi:10.1136/jcp.2006.040915

Zheng, H., Aziz, H. A., Nakanishi, Y., Masuda, S., Saito, H., Tsuneyama, K., et al. (2007). Oncogenic Role of JC Virus in Lung Cancer. *J. Pathol.* 212, 306–315. doi:10.1002/path.2188

Zheng, H. C., Yan, L., Cui, L., Guan, Y. F., and Takano, Y. (2009). Mapping the History and Current Situation of Research on John Cunningham Virus - a Bibliometric Analysis. *BMC Infect. Dis.* 9, 28. doi:10.1186/1471-2334-9-28

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 Zheng, E, Cui, Zhao 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fmolb.2021.707151

Identification of Potential Genomic Alterations and the circRNA-miRNA-mRNA Regulatory Network in Primary and Recurrent Synovial Sarcomas

Qing Yao[1†], Yong-Lai He[2†], Ning Wang[1†], Shuang-Shuang Dong[1], Mei Er Tu He Ta Mi Shi[1], Xiao Feng[1], Hao Chen[1], Li-Juan Pang[1], Hong Zou[1], Wen-Hu Zhou[3], Feng Li[1,4] and Yan Qi[1,2]**

[1]Department of Pathology, Shihezi University School of Medicine and the First Affiliated Hospital to Shihezi University School of Medicine, Shihezi, China, [2]Department of Pathology, Certral People's Hospital of Zhanjiang and Zhanjiang Central Hospital, Guangdong Medical University, Zhanjiang, China, [3]Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, China, [4]Department of Pathology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

Introduction: Synovial sarcoma (SS) is one of the most invasive soft tissue sarcomas, prone to recurrence and metastasis, and the efficacy of surgical treatment and chemotherapy for SS remains poor. Therefore, the diagnosis and treatment of SS remain a significant challenge. This study aimed to analyze the mutated genes of primary SS (PSS) and recurrent SS (RSS), discover whether these sarcomas exhibit some potential mutated genes, and then predict associated microRNAs (miRNA) and circular RNAs (circRNA) by analyzing the mutated genes. We focused on the regulation mechanism of the circRNA-miRNA-mutated hub gene in PSS and RSS.

Methods: We performed a comprehensive genomic analysis of four pairs of formalin-fixed paraffin-embedded samples of PSS and RSS, using Illumina human exon microarrays. The gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) function, and pathway enrichment of the mutated genes were analyzed, and the protein-protein interaction (PPI) network was forecast using String software 11.0. The hub genes were then obtained using the Molecular Complex Detection (MCODE) plug-in for Cytoscape 3.7.2 and were used to analyze overall survival (OS) using the Gene Expression Profiling Interactive Analysis (GEPIA) database. The corresponding miRNAs were obtained from the miRDB 5.0 and TargetScan 7.2 databases. The corresponding circRNAs of the hub genes were found through the miRNAs from these databases: Circbank, CircInteractome, and StarBase v2.0. Thereafter we set up a competing endogenous RNA (ceRNA) network with circRNA-miRNA and miRNA-messenger RNA (mRNA) pairs.

Results: Using the chi-squared test, 391 mutated genes were screened using a significance level of p-values < 0.01 from the four pairs of PSS and RSS samples. A GO pathway analysis of 391 mutated genes demonstrated that differential expression mRNAs (DEmRNAs) might be bound up with the "positive regulation of neurogenesis," "cell growth," "axon part," "cell–substrate junction," or "protein phosphatase binding" of

OPEN ACCESS

Edited by:
Xin Zhang,
Jiangmen Central Hospital, China

Reviewed by:
Xiaoyang Ye,
Institute for Systems Biology (ISB), United States
Han Shi,
Northwestern University, United States
Rossella Rota,
Bambino Gesù Children Hospital (IRCCS), Italy

***Correspondence:**
Yan Qi
qiyanyan-1998@163.com
Feng Li
lifeng7855@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 09 May 2021
Accepted: 28 July 2021
Published: 13 August 2021

Citation:
Yao Q, He Y-L, Wang N, Dong S-S, Tu He Ta Mi Shi ME, Feng X, Chen H, Pang L-J, Zou H, Zhou W-H, Li F and Qi Y (2021) Identification of Potential Genomic Alterations and the circRNA-miRNA-mRNA Regulatory Network in Primary and Recurrent Synovial Sarcomas. Front. Mol. Biosci. 8:707151. doi: 10.3389/fmolb.2021.707151

SS. The PPI network was constructed using 391 mutated genes, and 53 hub genes were identified ($p < 0.05$). Eight variant hub genes were discovered to be statistically significant using the OS analysis ($p < 0.05$). The circRNA-miRNA-mRNA (ceRNA) network was constructed, and it identified two circRNAs (hsa_circ_0070557 and hsa_circ_0070558), 10 miRNAs (hsa-let-7a-3p, hsa-let-7b-3p, hsa-let-7f-1-3p, hsa-let-7f-2-3p, hsa-mir-1244, hsa-mir-1197, hsa-mir-124-3p, hsa-mir-1249-5p, hsa-mir-1253, and hsa-mir-1271-5p) and five hub genes (CENPE, ENPP3, GPR18, MDC1, and PLOD2).

Conclusion: This study screened novel biological markers and investigated the differentiated circRNA-miRNA-mutated hub gene axis, which may play a pivotal role in the nosogenesis of PSS and RSS. Some circRNAs may be deemed new diagnostic or therapeutic targets that could be conducive to the future clinical treatment of SS.

Keywords: genome-wide SNP analysis, circRNA, ceRNA, gene, primary synovial sarcoma, recurrent synovial sarcoma

INTRODUCTION

Synovial sarcoma (SS) is a highly invasive soft tissue sarcoma (STS) that predominantly occurs in the limbs of adults in their 30s and accounts for about 5–10% of all STSs (Riedel et al., 2018; Hale et al., 2019). Its pathogenic characteristics are related to the unique chromosomal translocation t (X; 18) (p11.2; q11.2), leading to the formation and expression of the oncogenic fusion gene SS18-SSX (Riggi et al., 2018). Accepted standard treatment methods for SS include extensive surgical resection and chemotherapy. However, treatment has not been shown to significantly improve SS outcomes (Yoshimatsu et al., 2020), which were closely related to the occurrence of metastasis or local recurrence in more than half of patients, and spread easily to sites, such as the lungs and bone (El Beaino et al., 2017). In recent years, some studies have shown that different genomic alterations were found in primary and recurrent tumors, including chromatin mutation, RNA splicing, and epigenomic regulation. However, information on the underlying molecular mechanisms of aggressive SS is still lacking. Much work remains to be done to explore the different genomic alterations and investigate the natural history of primary SS (PSS) and recurrent SS (RSS). This could help guide the development of good clinical therapies for SS.

Qi et al. (2015) reported that genome-wide single nucleotide polymorphisms (SNPs) are used to assess gene functions because SNPs can use limited sample material to study DNA sequence variations occurring in the genome and transcriptome and identify their pathogenic genes (Guo and Rotter, 2019). Therefore, therapy can be targeted based on mutated genes in PSS and RSS.

In recent years, an increasing number of studies have shown that genes and microRNA (miRNA) can be regulated by circular RNA (circRNA) (Zeng et al., 2020) and that the aberrant expression of circRNA and miRNA contributes to human diseases, including cancer (Lin et al., 2020; Sarkar et al., 2020). With the emergence of high-throughput RNA sequencing and next-generation sequencing technology, circRNA has been increasingly studied (Lu and Thum, 2019). Liu et al. (2015)(Kristensen et al., 2019) found that, compared to their linear counterparts, circRNAs are highly stable because of their covalently closed structure, are less susceptible to degradation by enzymes, and can be discovered in exosomes, cell-free saliva, and plasma. Moreover, circRNAs can act as miRNA sponges to regulate gene expression (Zheng et al., 2016). Therefore, because of some features of circRNA stability, some studies have suggested that circRNAs can be used as markers for disease diagnosis and prognosis (Kristensen et al., 2019).

Thus, to provide insights into the pathogenesis of PSS and RSS, we aimed to obtain differentially expressed genes and extract hub genes from PSS and RSS, and circRNA-miRNA-messenger RNA (mRNA) competing endogenous RNA (ceRNA) networks using genome-wide SNP analysis and data from existing open-source resources.

MATERIALS AND METHODS

Patients and Tissue Specimens

Four formalin-fixed paraffin-embedded (FFPE) samples of PSS and RSS, in which SNP expression was proven to differ between primary and recurrent samples, were acquired from patients treated at the Department of Pathology, First Affiliated Hospital, Shihezi University, School of Medicine. The study included one male and three female patients with biphasic SS. The patients' ages were 64, 22, 10, and 37 years (mean age 34 years). Histological and immunohistochemical analyses confirmed the diagnosis of SS, and the presence of a fusion gene was detected using a one-step reverse transcription-polymerase chain reaction (QIAGEN, Venlo, The Netherlands). The characteristics of the four paired PSS and RSS patients enrolled in this study are shown in **Table 1**. Clinical staging was performed according to the National Comprehensive Cancer Network 2016 guidelines for soft-tissue tumors (von Mehren et al., 2016).

In this study, all patients provided informed consent, and the patients'records and information were anonymized before analysis. The study was organized following the instructions

TABLE 1 | The basic information of the patients.

Case	Sex/ age	Classification	Molecular pathological classification	Diagnosis	Location	Size (cm)	Metastases	Grade	Stages
P1	F/22	primary synovial sarcoma	SYT-SSX2	BSS	Left heel	4.5 × 3×1.5	None	III	IIA
R1	F/22	recurrent synovial sarcoma	SYT-SSX2	BSS	Left heel	3 × 3×2.5	None		
P2	F/10	primary synovial sarcoma	SYT-SSX1	BSS	Posterior to right elbow	5 × 4.8×3	None	III	IV
R2	F/10	recurrent synovial sarcoma	SYT-SSX	BSS		NA	Bone marrow cavity		
P3	M/64	primary synovial sarcoma	SYT-SSX	SS	Anterior tibialis of right	5.6 × 1.2×1.1	None	III	I
R3	M/70	recurrent synovial sarcoma	SYT-SSX	SS	Left hip	11.5 × 10.5×7	None		
P4	F/37	primary synovial sarcoma	SYT-SSX	SS	Left pubic ramus	3.5 × 3.3×1	None	III	NA
R4	F/37	recurrent synovial sarcoma	SYT-SSX	BSS	Left ischial tuberosity bursa	NA	None	NA	NA

P: primary; R: recurrent; BSS: Biphasic synovial sarcoma; Grade: Histopathological grade; Stages: clinical stages; NA: not available.

approved by the Clinical Research Ethics Committee of the First Affiliated Hospital of Shihezi University School of Medicine.

DNA Extraction and SNP Array

Genomic DNA was isolated from four FFPE tumor samples using the QIAamp DNA Micro Kit (Qiagen Inc., Valencia, CA, United States) according to the manufacturer's instructions. The minimum amount of genomic DNA successfully subjected to genomic SNP microarray analysis was 1 ug, with quality metrics of A260/280 = 1.7–2.0 and A260/230 > 1.6. According to the manufacturer's recommendation, the patient's DNA target was prepared and hybridized using an Infinium HD analysis Superchip. CEL files were imported into Illumina Human Exome-12v1.1 (Beijing Compass Biological Technology Co. Ltd.; http://www.kangpusen.com/index.html) for quality analysis according to the "Quality Control Assessment in Genotyping Console." An association analysis of controls versus cases was performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) using Fisher's exact test, chi-squared test, and estimated odds ratios. Pearson's chi-square values and *p*-values (chi-squared and Fisher exact tests) were estimated using Daploview 4.2 PLINK data.

Illumina Exon Microarray Detection

A custom microarray with exon-level resolution was investigated using Illumina Human Exome-12v1.1 to identify gross deletions and duplications. The microarray contains approximately 60,000 integrated oligonucleotide probes that have been annotated against the human genome. The density of probes increased across exons and 300 bp of flanking intronic sequence. After the validation run, only the best-performing probes were selected for use in the final array design.

The specific method used is as follows: first, a DNA standard plate was prepared, followed by a QNT standard plate with diluted Pico Green and a QNT sample plate with PicoGreen and DNA. Subsequently, DNA samples were moved to the MSA1 plate to denature and neutralize them and were then used for amplification by overnight incubation. The following day, an endpoint fragmentation was used to enzymatically fragment the DNA to avoid excessive fragmentation. Therefter, the DNA samples were treated with 2-propanol and PM1 to precipitate MSA1, and RA1 was used to resuspend the precipitated DNA. Next, the resuspended DNA samples were dispensed on BeadChips, which were incubated in an Illumina Hybridization Oven for approximately 16–24 h to hybridize the samples onto them. The following day, a stained BeadChip was prepared, and nonhybridized and nonspecifically hybridized DNA was washed away. There are eight categories of internal controls per sample in an Illumina SNP microarray experiment: staining, extension, hybridization, stringency, target removal, recovery, nonspecific binding, and nonpolymorphic controls. Quality analysis of the Illumina Human Exome-12v1.1 reporter samples indicated a call rate of 0.7526.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Functional Enrichment Analysis

GO (Dutkowski et al., 2013) is a key tool for annotating genes and analyzing their biological processes (BPs), molecular functions (MFs), and cellular components (CCs). To further understand the enrichment of SS mutated genes and their main functional annotations, the mutated genes of SNPs were evaluated by the GO annotation and KEGG (Kanehisa et al., 2021) pathways, using cluster spectral packages in R 3.51 (geoSeq). *p*-values < 0.05 were considered statistically significant.

Constructing a Protein-Protein Interaction Network and Finding Hub Genes of SNPs

The String online database (11.0; https://string-db.org/) provides reliable information on PPIs (Szklarczyk et al., 2011). In this study, PPIs with a combined score >0.9 were considered statistically significant. The PPI network was constructed

through mutant genes of SNPs in the String online database. Cytoscape is an open-source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles, and other state data (Shannon et al., 2003) and was used to analyze the obtained gene or protein networks (Cline et al., 2007). The Molecular Complex Detection (MCODE) is an application plug-in for Cytoscape 3.7.2, whose algorithm can detect densely connected regions in large PPI networks that may represent molecular complexes. Therefore, the MCODE plug-in should be used to identify the hub genes of SNPs obtained from the PPI network (Bandettini et al., 2012). The hub gene criteria were as follows: degree cut-off = 2, node score cut-off = 0.2, K-core = 2, Max depth = 100.

Survival Analysis

The total survival rate of hub genes acquired from the PPI network was analyzed using the Gene Expression Profiling Interactive Analysis (GEPIA) online database. GEPIA (http://gepia.cancer-pku.cn/detail.php) (Tang et al., 2017) is a web-based tool that provides fast and customizable capabilities based on The Cancer Genome Atlas and Genotype-Tissue Expression data. GEPIA offers key interactive and customizable features, including differential expression, correlation, patient survival analyses. The log-rank test was used for statistical analysis. In this study, overall survival (OS) was analyzed by hub genes differentially expressed between SS tissues and normal tissues. The threshold for prognostic survival significance was a p-value < 0.05. A subnetwork was constructed based on hub genes of SNPs that were meaningful for OS.

Predicting Binding Sites of miRNAs and circRNAs and Construction of the ceRNA Network

The TargetScan 7.2 (http://www.targetscan.org/vert_72/) and miRDB 5.0 (Chen and Wang, 2020) (http://mirdb.org/) online databases were used to predict the association between mutant genes and miRNAs. Only the miRNAs present in these two databases were regarded as candidate miRNAs. The Circbank (Liu et al., 2019) (http://www.circbank.cn/downloads.html), CircInteractome (Dudekula et al., 2016) (https://circinteractome.nia.nih.gov/index.html), and StarBase V2.0 (Li et al., 2014) (https://www.ncbi.nlm.nih.gov/geo/) online databases were used to find differentially expressed circRNAs (DEcircRNAs) targeted by miRNAs. The ceRNA regulatory network was constructed using circRNA-miRNA pairs and miRNA-mRNA pairs. Finally, the ceRNA network was acquired through Cytoscape 3.7.2.

RESULTS

Exome-Chip Data and Functional Annotation of SS-Associated Variants

We identified approximately 2024 single nucleotide variants, small insertions, and deletion alterations using Illumina Human Exome Microarrays in the primary and recurrent cases across the genomes of the four patients (**Figure 1**) (**Supplementary Material 1**). The PLINK analysis of the SNP-chip data revealed that 1828 SNPs were significant ($p < 0.05$). Further, GO analysis showed that SNPs of 1828 are mainly concentrated in "metabolic process," "biological regulation," "multicellular organismal process" (categorized as BPs), "membranes," "the nucleus," "macromolecular complexes" (categorized as CCs), "protein binding," "ion binding," and "hydrolase activity" (categorized as MFs) (**Figure 2A**). Of note, the KEGG analysis only focused on the extracellular matrix-receptor interaction ($p < 0.05$). To make our data meaningful, we selected the significant data ($p < 0.01$) for further analysis, including 391 SNPs (**Supplementary Material 2**). Subsequently, the result of the GO analysis indicated that 391 variant genes of SNPs were mainly concentrated on "positive regulation of neurogenesis," "cell growth," "positive regulation of neuron differentiation" (categorized as BPs), "axon part," "cell–substrate junction" (categorized as CCs), "protein phosphatase binding," and "phosphatase binding" (categorized as MFs) (**Figure 2B**). However, no statistically significant pathway was found by the KEGG analysis for 391 mutated genes of SNPs ($p > 0.05$). We, therefore, conclude that the genes from which we obtained mutations were associated with neuronal differentiation, and studies suggest that SSs have a possible neural origin (Ishibe et al., 2008).

Construction of the Protein-Protein Interaction Network and Determination of Hub Genes by Mutated Genes

The 391 mutated genes of SNPs in PSS and RSS were shown in the PPI network using String software (**Figure 3A**), and 53 hub mutated genes of SNPs in PSS and RSS were found by MCODE (Cytoscape plug-in) (**Figure 3B**). The GO analysis showed that 53 variant genes of SNPs in PSS and RSS were mainly concentrated in "positive regulation of neurogenesis," "vesicle–mediated transport in synapse" (categorized as BPs), "axon part" (categorized as CCs), and "ubiquitin–protein transferase activity" (categorized as MFs) (**Figures 3C,D**). The KEGG analysis found that 53 hub mutated genes were not statistically significant (p-value > 0.05). We, therefore, concluded that the genes from which we obtained the mutations were of great relevance for neural differentiation and could help mediate transport.

Survival Analysis of Hub Genes

Next, we assessed OS of 53 variant hub genes of SNPs in PRSSs by the GEPIA database. There were eight hub variant genes significant for OS analysis (*CENPE, CUL7, ENPP3, GPR18, IKBKG, MDC1, MUC16,* and *PLOD2*). The result showed that all eight variant hub genes were highly expressed in SS tissues ($p < 0.05$; **Figure 4**). It is suggested that these eight most significant mutant hub genes may be potential markers for PSS and RSS. Basic information regarding these eight hub genes is listed in **Table 2**.

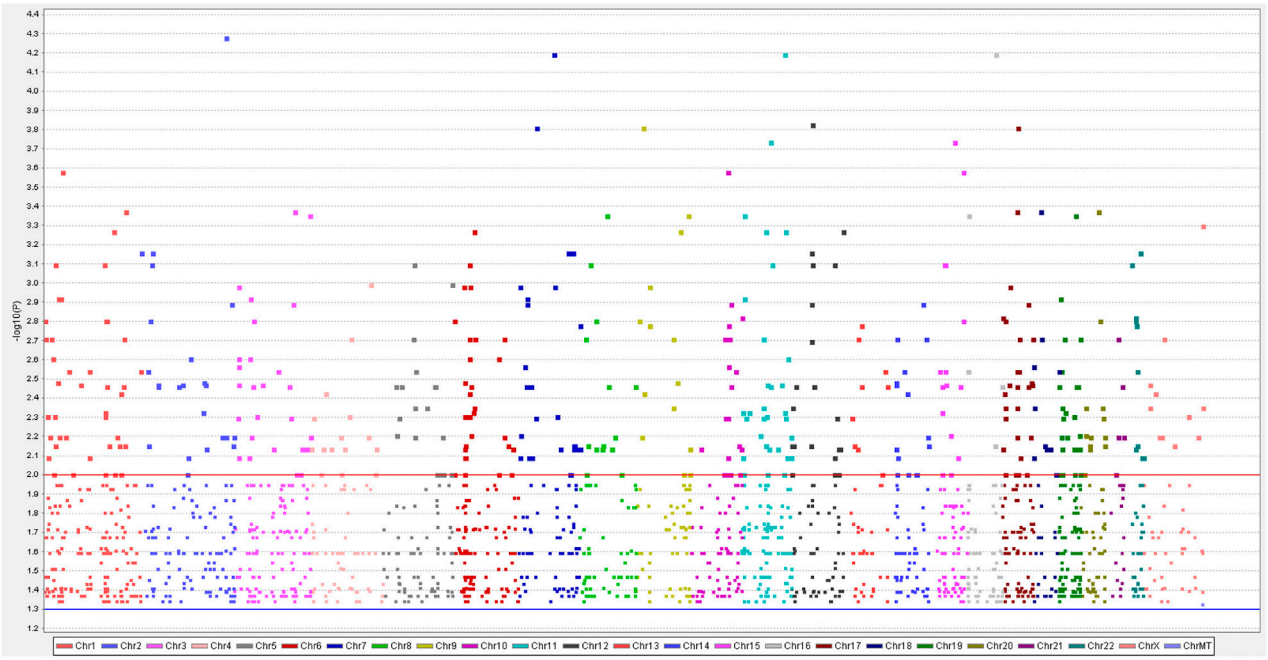


FIGURE 1 | In the aggregate to 2024 single nucleotide polymorphisms (SNPs) were obtained by a genome-wide analysis of four samples of synovial sarcomas. A Manhattan plot indicates negative log-transformed *p*-values of the case-control allele frequency significance on the Y-axis. The color-coding on the X-axis represents the number of chromosomes. Gene names indicated by single dots demonstrate SNPs of greatest significance or with potential disease relevance.

Construction of the ceRNA Network Through Mutated Hub Genes

In order to better understand the relationship among circRNA, miRNA, and mRNA, a ceRNA network was constructed with the SNPs of eight mutated genes (CENPE, CUL7, ENPP3, GPR18, IKBKG, MDC1, MUC16, and PLOD2). Then the corresponding miRNAs were found in two databases (miRBD and TargetScan) through the mutated hub genes. Five mutated hub genes (*CENPE, ENPP3, GPR18, MDC1,* and *PLOD2*) were found to target miRNAs (hsa-let-7a-3p, hsa-let-7b-3p, hsa-let-7f-1-3p, hsa-let-7f-2-3p, hsa-mir-1244, hsa-mir-1197, hsa-mir-124-3p, hsa-mir-1249-5p, hsa-mir-1253, and hsa-mir-1271-5p) in these two databases. The targeted circRNAs (hsa_circ_0070557 and hsa_circ_0070558) were found in three databases (Circbank, CircInteractome, StarBase) with miRNA. Finally, we used two circRNA nodes, 10 miRNA nodes, and five mutated hub genes nodes in Cytoscape 3.7.2 to form the ceRNA network (**Figure 5A**). Therefore, we predicted that five hub mutated genes (CENPE, ENPP3, GPR18, MDC1, and PLOD2) and two circRNAs (hsa_circ_0070557 and hsa_circ_0070558) might be suitable as potential markers for SS. The basic information regarding the circRNAs is shown in **Figure 5B**.

DISCUSSION

SS is an aggressive malignancy with biphasic differentiated sarcoma, prone to local recurrence and distant metastasis in the advanced stage (Przybyl et al., 2014). Local recurrence occurs after an average of 3.6 years and metastasis after an average of 5.7 years, and the lung is the most common site of metastasis (Krieg et al., 2011); therefore, the prognosis is poor. The origin of SS is uncertain, and studies have shown that it is derived from mesenchymal stem cells (De Logu et al., 2020). However, recent studies have suggested that neural, myogenic, or multipotent mesenchymal stem cells are considered to be the likely cellular origin (De Logu et al., 2020). It has been shown that the gene expression profile of SS is closely related to malignant peripheral nerve sheath tumors and that neural tissue-associated genes are expressed in SS tissues (Ishibe et al., 2008). Our results also showed that SS may be closely associated with neural differentiation according to the GO analysis. Histologically, SSs can be divided into monophasic, biphasic, and epithelioid types, which can present necrosis and bizarre mitoses (Soule, 1986; El Beaino et al., 2017). The morphology and clinical symptoms of SS overlap with those of various other tumors, so its diagnosis and treatment remain challenging (Thway and Fisher, 2014). Therefore, novel prognostic and predictive biomarkers, as well as novel therapeutic targets for SSs, are urgently needed.

In the era of targeted therapy, it is increasingly important to identify clinically actionable genetic alterations in cancer, and the number of these alterations is continuously increasing (Corless and Spellman, 2012). Therefore, high-throughput sequencing technologies are emerging as an essential resource for deciphering the genotype underlying a given phenotype in clinical diagnostics, cancer research, and molecularly targeted

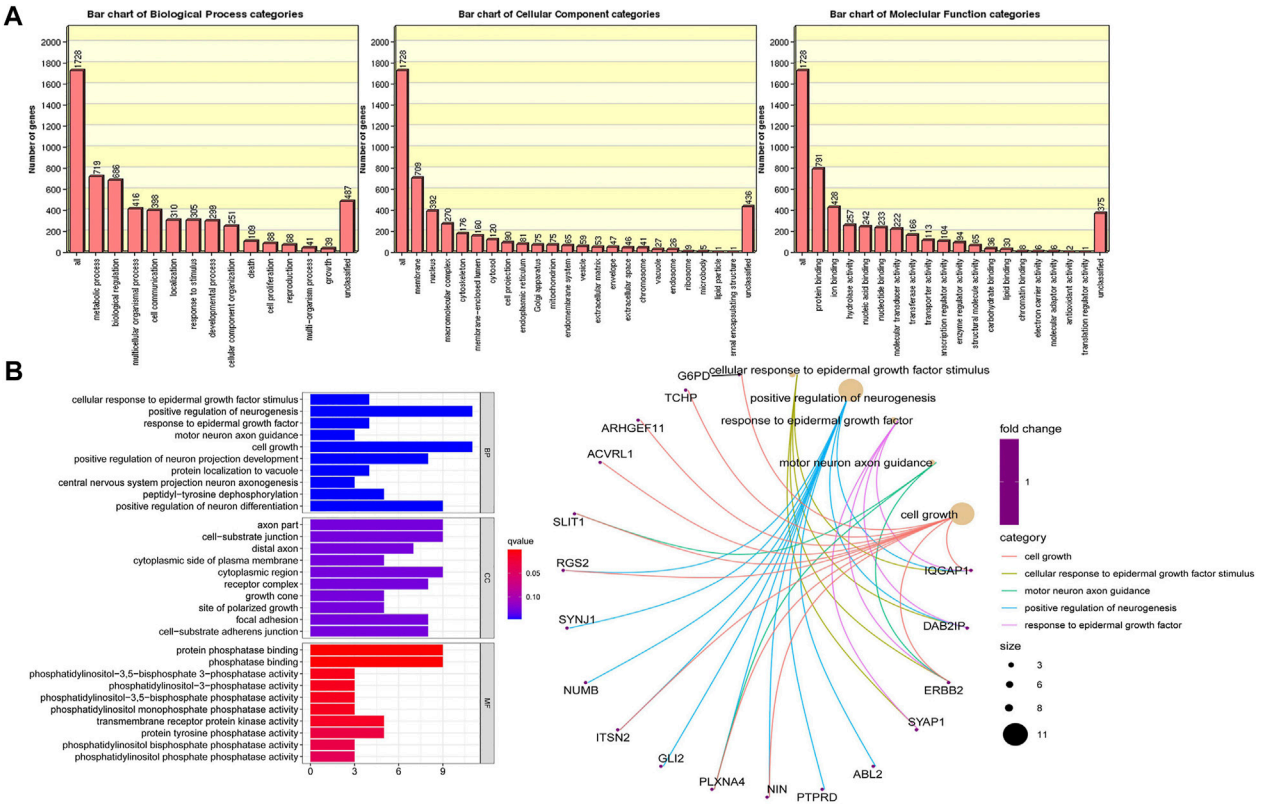


FIGURE 2 | The differential single nucleotide polymorphisms (SNPs) of target genes were found to be functionally enriched by a gene ontology enrichment analysis, which mainly included three aspects: biological processes (BPs), molecular functions (MFs) and cellular components (CCs). **(A)** Functional enrichment analysis of 1828 SNP variant genes ($p < 0.05$); **(B)** Functional enrichment analysis of 391 SNP variant genes ($p < 0.01$).

therapies for various cancers (Song et al., 2014). Many studies have found that DNA or RNA extraction is mainly performed through fresh freezing (FF), particularly for FFPE (Morris et al., 2021). FFPE tissue is the gold standard for pathological tissue preservation and represents the largest collection of patient material; its most significant advantage is its ability to be preserved long term. FFPE tissue samples provide a valuable sample resource and reduce or eliminate the need for painstaking collection and storage of cryopreserved clinical samples (Flores Bueso et al., 2020). However, the main drawback of FFPE tissue is that nucleic acid extraction is difficult owing to the need for paraffin removal with the associated loss of DNA content. Schweiger et al. (Schweiger et al., 2009) found that the decline in sequence quality was negligible in FFPE samples collected over a decade by comparing FF in whole-genome and exome sequencing. Therefore, the results suggested that reliable DNA/RNA can be obtained from FFPE tissues (Van Allen et al., 2014).

In the past few years, an increasing number of studies have shown that circRNA is much more stable than linear transcripts, mainly owing to its inability to be degraded by exonucleases, and thus this feature is thought to contribute to its effectiveness as an RNA sponge (Hansen et al., 2013). Although circRNA's use as a sponge remains controversial, the role of circRNA as an RNA sponge is increasingly confirmed and recognized. Hansen et al. have experimentally demonstrated that Sry-derived circRNA has been found to act as a sponge for miR-138 (Hansen et al., 2013). Studies have shown that circRNA and miRNA can also serve as promising diagnostic markers and therapeutic targets and that circRNAs and miRNAs are closely related to diseases, such as cancer. Moreover, the circRNA-miRNA-mutated gene axis can be analyzed using bioinformatics to study disease pathogenesis and predict meaningful DEcircRNAs (Liang et al., 2020; Ruan et al., 2020). Zhang et al. (2018) found that obtaining circRNAs microarray analysis from FFPE samples shows comparable results to reports based on fresh frozen samples. Therefore, FFPE specimens are a good alternative to fresh frozen tissue, especially when fresh frozen specimens are limited. Bai et al. (2020) have shown that circRNA and genes were identified by the regulatory network of ceRNA, and the high expression of mutant genes was negatively correlated with OS in clear cell renal cell carcinoma (ccRCC). The potential pathogenesis of the regulatory network among circRNA/miRNA/mRNA provides some potential therapeutic options for ccRCC. Therefore, it is necessary to study the regulatory mechanisms of circRNA and hub genes, to improve the diagnosis and treatment of SS.

In investigating the potential hub genes, miRNA, and circRNA associated with PRSSs, we found eight variant hub genes (*CENPE, CUL7, ENPP3, GPR18, IKBKG, MDC1, MUC16,* and *PLOD2*),

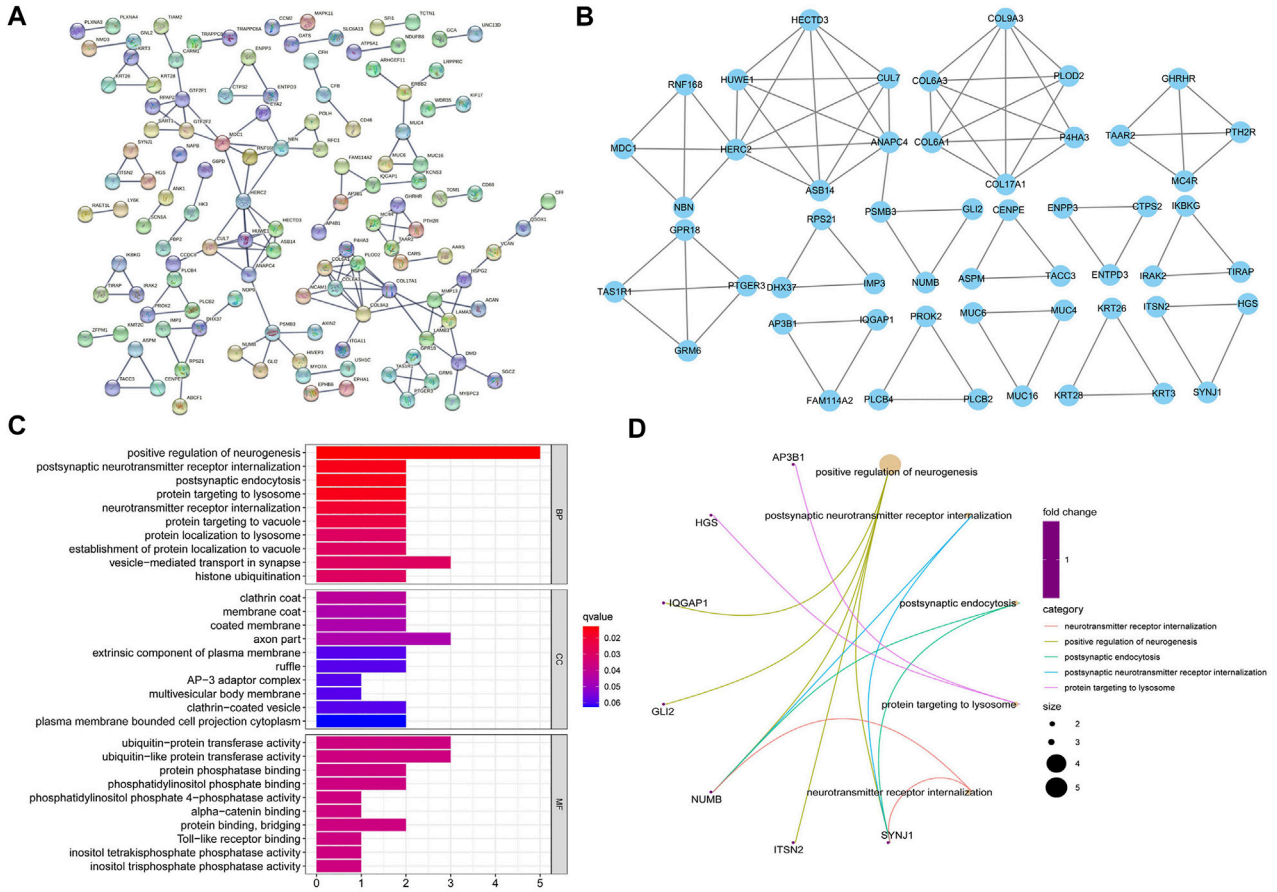


FIGURE 3 | Using the protein interaction network, we could determine a protein interaction relationship between different genes which participate in biological signal transmission, gene expression regulation, and other processes. To determine which genes were in the hub position in the regulatory network, we screened the hub genes using the Molecular Complex Detection (MCODE) plug-in. The obtained hub genes were further subjected to a gene ontology (GO) enrichment analysis to observe the hub genes' primary enriched functions. **(A)** The protein-protein interactions (PPIs) of mutated single nucleotide polymorphism (SNP) genes were analyzed using the String database, and the criteria for construction were a combined score >0.9. **(B)** The most significant module in the PPI network was obtained using the MCODE plug-in for the Cytoscape 3.7.2 online software, and the criteria were obtained as follows: degree cut-off = 2, node score cut-off = 0.2, K-core = 2, Max depth = 100. **(C/D)** GO function of the 53 hub genes is mainly enriched in these three parts: biological processes (BPs), molecular functions (MFs) and cellular components (CCs) ($p < 0.05$).

which were found to be significant on the OS analysis in SS. Some studies have shown that these mutated genes may serve as biomarkers and therapeutic targets and are related to tumor metastasis, recurrence, and prognosis. Li and Shan et al. found that CENPE can serve as a new target for the diagnosis and prognosis of rhabdomyosarcoma (Li et al., 2019) and ccRCC (Shan et al., 2019), and that it can promote the metastasis of lung cancer, and Horning et al. found that CENPE was up-regulated in prostate cancer and related to recurrent prostate cancer through RNA-sequencing data analysis.

There is increasing evidence that *CUL7* is a carcinogenic gene that can promote tumorigenesis by activating caspase-8 ubiquitination (Kong et al., 2019) and the nuclear factor-κB (Xu et al., 2020) pathway. Studies have found that *CUL7* is highly expressed in breast, lung, hepatocellular, pancreatic, ovarian, and other malignancies, thus potentially making it a novel anticancer target (Li et al., 2021). Further, some studies have shown that *ENPP3* is highly expressed in ccRCC (Moek et al., 2017). A phase I trial has been conducted with an anti-*ENPP3* antibody-drug conjugate to treat advanced refractory renal cancer (Doñate et al., 2016). Qin et al. showed that G protein-coupled receptors (GPCRs) are closely related to the occurrence and metastasis of human tumors and that *GPR18* GPCRs may be considered potential novel anticancer targets in metastatic melanoma (Qin et al., 2011). Eisinger-Mathason et al. (2013) found that the inhibition of *PLOD* enzyme activity can inhibit sarcoma metastasis, so *PLOD* is a new therapeutic target for sarcoma. Some studies have shown that IKBKG (Kang et al., 2020), MUC16 (Akahane et al., 2020) and MDC1 (Wang et al., 2018) are all related to the recurrence of breast cancer, and studies have found that MUC16 is highly expressed in most epithelial ovarian cancers and is related to the recurrence of ovarian cancer

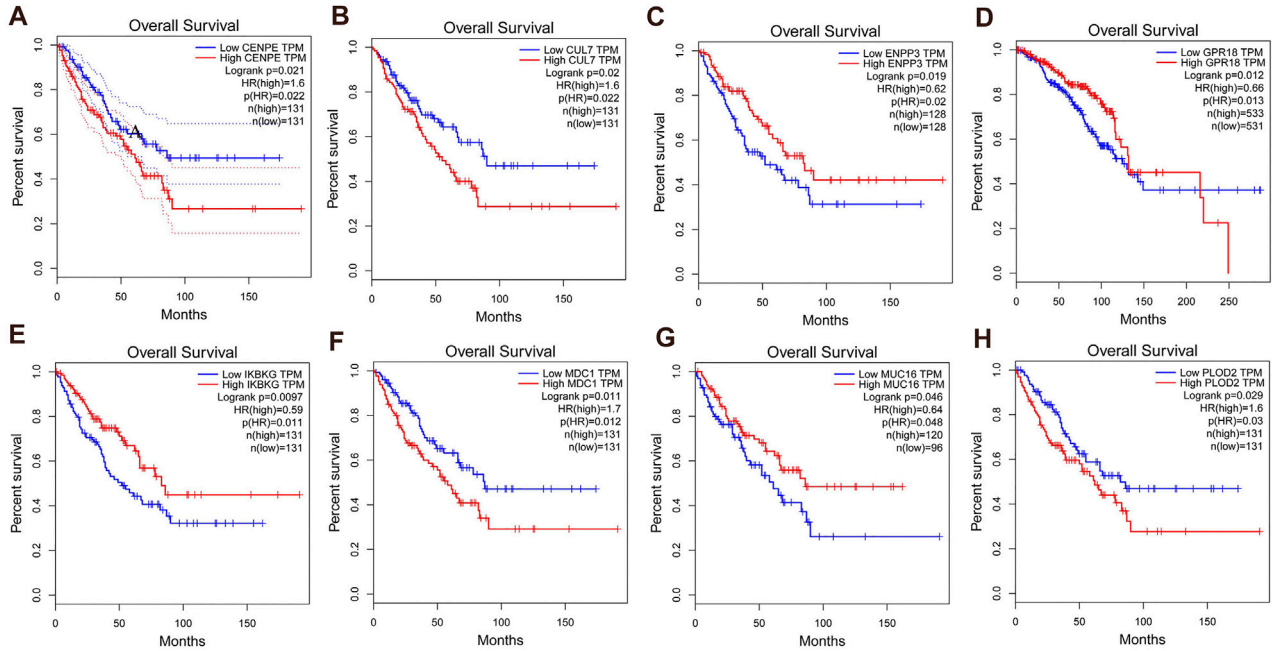


FIGURE 4 | The obtained 53 hub genes were further screened according to overall survival (OS), and eight hub genes were found to be highly expressed in synovial sarcoma tissues. The threshold for prognostic survival significance was a *p*-value < 0.05. **(A)** CENPE; **(B)** CUL7; **(C)** ENPP3; **(D)** GPR18; **(E)** IKBKG; **(F)** MDC1; **(G)** MUC16; **(H)** PLOD2.

TABLE 2 | The basic information of hub genes of SNPs.

SNPs Name	*p* value	Chr	MapInfo	Alleles	Transcripts	Genes	Location	Mutations	miRNA	circRNA
exm416631	0.001946	4	104066461	[A/G]	NM_001813	CENPE	EXON	Missense_F1535L	hsa-let-7a-3p hsa-let-7b-3p hsa-let-7f-1-3p	hsa_circ_0070557 hsa_circ_0070558
exm578521	0.006285	6	131962646	[A/G]	NM_005021	ENPP3	EXON	Missense_G44R	hsa-let-7f-2-3p	None
exm1076404	0.009823	13	99907619	[T/C]	NM_001144072	GPR18	EXON	Silent	hsa-mir-1244 hsa-let-7f-2-3p	None
exm2248707	0.004427	X	153774337	[A/C]	NM_001042351	IKBKG	EXON	Missense_V12L Missense_V42L	None	None
exm547999	0.002469	6	43008764	[T/C]	NM_014780 NM_001168370	CUL7	EXON	Missense_R1232Q Missense_R1316Q	None	None
exm528596	0.008041	6	30680111	[T/C]	NM_014641	MDC1	EXON	Missense_I536M	hsa-mir-1197 hsa-mir-124-3p hsa-mir-1249-5p hsa-mir-1253	None
exm1421643	0.009823	19	9073888	[T/C]	NM_024690	MUC16	EXON	Missense_A4520T	None	None
exm356896	0.005012	3	145809622	[A/G]	NM_182943 NM_000935	PLOD2	EXON	Missense_C282R	hsa-mir-1271-5p	None

(Giannakouros et al., 2015). Therefore, MUC16 is a novel and promising tumor-associated antigen, providing an excellent target for immunotherapy, and a phase I clinical trial has been carried out (Koneru et al., 2015).

Genes can be targeted to identify upstream miRNAs. A total of 10 miRNAs were identified in this study, with studies finding hsa-let-7f-2-3p to be associated with bladder cancer progression (He et al., 2018), hsa-mir-1244 (Li et al., 2017) and hsa-mir-1197 (Yang et al., 2020) to be associated with lung cancer, and hsa-mir-1253 (Wu et al., 2020) to be associated with prostate cancer, and other miRNAs not found to be associated with cancer. It has been established that circRNA can function as "sponges" to recruit the corresponding miRNAs by interacting with miRNA binding sites, thereby playing an indirect role in regulating gene expression (Bai et al., 2020).

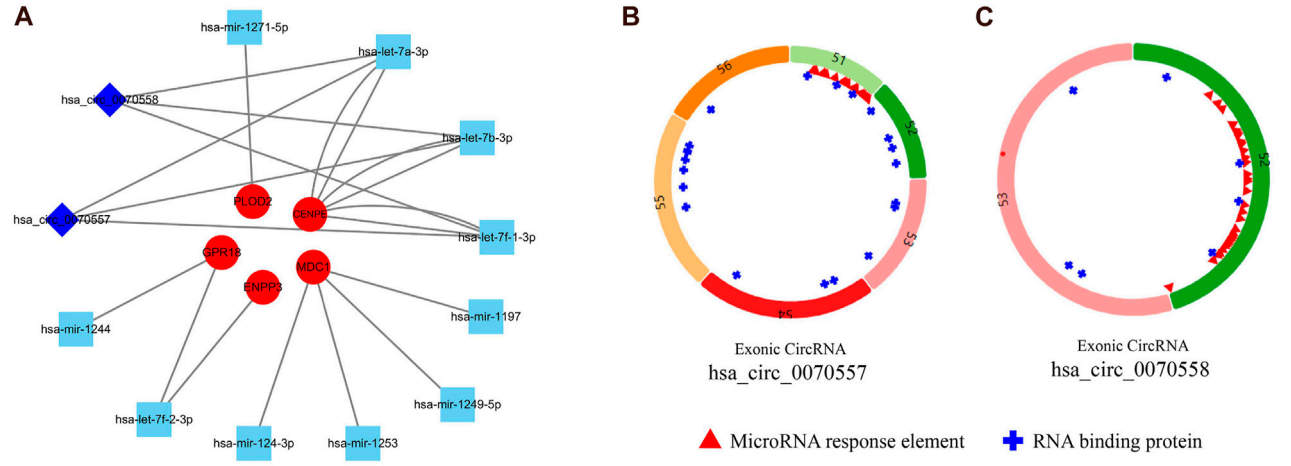


FIGURE 5 | To investigate the regulatory mechanism of synovial sarcoma, we constructed a circulatoryRNA-microRNA-messengerRNA (circRNA-miRNA-mRNA) (competing endogenous RNA) network using circRNA-miRNA pairs and miRNA-mRNA pairs with Cytoscape 3.7.2. **(A)** Squares represent circRNA, diamonds represent miRNA, and circles represent mRNA. **(B)** Basic information regarding two circRNAs (has_circ_0070557 and has_circ_0070558).

In this study, two circRNAs (hsa_circ_0070557 and hsa_circ_0070558) were identified, and both have yet to be reported. Therefore, the regulatory mechanism of these hub genes-miRNAs-circRNAs in SS remains unclear. Sun et al. (Zhengwang et al., 2021) analyzed transcriptomic changes of SS in differential expression, alternative splicing, gene fusion and circRNA by whole RNA sequencing in SS tissues compared with normal tissues, and found that these DEcircRNAs might act as miRNA sponges to regulate the expression of a large number of differentially expressed and alternatively spliced genes. Interestingly, we screened potential the genes that corresponded to the different genetic and SNPs altertions, by comparing PSS with RSS samples in our study. All of these studies may contribute to further understanding of the molecular genetic changes and construct gene-miRNAs-circRNAs network of SS.

Owing to the low incidence of SSs, paired PSS and RSS are particularly valuable for genome-wide association research, even though the small sample size is a limitation of this study. Therefore, our results need further validation, and the regulation mechanism of ceRNA in PSS and RSS should be further investigated regarding its significance in the diagnosis and prognosis of SS. We did not compare them with normal tissues, but Sun et al. reported the relationship between the transcriptome maps of synovial sarcomas and normal tissues, this is the difference between our study and Sun et al. Although there is some loss of DNA/RNA from the sample owing to formalin fixation, the results of this study are still reliable, and FFPE tissues are suitable for the relevant microarray analysis (He et al., 2016).

CONCLUSION

The clinical application of genomics is becoming highly extensive. Herein, we analyzed genome-wide SNPs in both primary and recurrent SSs. In our study, we found that the functional annotation of SNP gene mutations is related to biological adhesion, plasma membranes, and purine nucleoside binding. These eight hub genes we screened could serve as potential targets for cancer therapy. Perticurlarly for CENPE, it can be used as a novel target for sarcoma diagnosis and prognosis. And we obtained through database that CENPE interacts with hsa-let-7a/b-3p and hsa-let-7f-1-3p family and can target hsa_circ_0070557 and hsa_circ_0070558. Therefore, this may be a potential circRNA-miRNA-mRNA regulatory network, which provides a new idea for the pathogenesis of SS and is worthy of further study on potential therapeutic targets.

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: NCBI [accession: GSE178828].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Clinical Research Ethics board of the First Affiliated Hospital, Shihezi University School of Medicine. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

QY and NW performed the experiments and analyzed the data; YQ, FL and W-HZ designed or supervised the study; SSD, MT,

XF and HC provided crucial input for the project; QY, Y-LH and NW wrote 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 (grant no. 81860471, 82002846), the International Cooperation Projects of Shihezi University (grant no. GJHZ201710).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.707151/full#supplementary-material.

REFERENCES

Akahane, T., Kanomata, N., Harada, O., Yamashita, T., Kurebayashi, J., Tanimoto, A., et al. (2020). Targeted Next-Generation Sequencing Assays Using Triplet Samples of normal Breast Tissue, Primary Breast Cancer, and Recurrent/metastatic Lesions. *BMC cancer* 20 (1), 944. doi:10.1186/s12885-020-07432-w

Bai, S., Wu, Y., Yan, Y., Shao, S., Zhang, J., Liu, J., et al. (2020). Construct a circRNA/miRNA/mRNA Regulatory Network to Explore Potential Pathogenesis and Therapy Options of clear Cell Renal Cell Carcinoma. *Sci. Rep.* 10 (1), 13659. doi:10.1038/s41598-020-70484-2

Bandettini, W. P., Kellman, P., Mancini, C., Booker, O., Vasu, S., Leung, S. W., et al. (2012). MultiContrast Delayed Enhancement (MCODE) Improves Detection of Subendocardial Myocardial Infarction by Late Gadolinium Enhancement Cardiovascular Magnetic Resonance: a Clinical Validation Study. *J. Cardiovasc. Magn. Reson.* 14, 83. doi:10.1186/1532-429x-14-83

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48, D127–D131. doi:10.1093/nar/gkz757

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., et al. (2007). Integration of Biological Networks and Gene Expression Data Using Cytoscape. *Nat. Protoc.* 2 (10), 2366–2382. doi:10.1038/nprot.2007.324

Corless, C. L., and Spellman, P. T. (2012). Tackling Formalin-Fixed, Paraffin-Embedded Tumor Tissue with Next-Generation Sequencing. *Cancer Discov.* 2 (1), 23–24. doi:10.1158/2159-8290.cd-11-0319

De Logu, F., Ugolini, F., Caporalini, C., Palomba, A., Simi, S., Portelli, F., et al. (2020). TRPA1 Expression in Synovial Sarcoma May Support Neural Origin. *Biomolecules* 10 (10). doi:10.3390/biom10101446

Doñate, F., Raitano, A., Morrison, K., An, Z., Capo, L., Aviña, H., et al. (2016). AGS16F Is a Novel Antibody Drug Conjugate Directed against ENPP3 for the Treatment of Renal Cell Carcinoma. *Clin. Cancer Res.* 22 (8), 1989–1999. doi:10.1158/1078-0432.ccr-15-1542

Dudekula, D. B., Panda, A. C., Grammatikakis, I., De, S., Abdelmohsen, K., and Gorospe, M. (2016). CircInteractome: A Web Tool for Exploring Circular RNAs and Their Interacting Proteins and microRNAs. *RNA Biol.* 13 (1), 34–42. doi:10.1080/15476286.2015.1128065

Dutkowski, J., Kramer, M., Surma, M. A., Balakrishnan, R., Cherry, J. M., Krogan, N. J., et al. (2013). A Gene Ontology Inferred from Molecular Networks. *Nat. Biotechnol.* 31 (1), 38–45. doi:10.1038/nbt.2463

Eisinger-Mathason, T. S. K., Zhang, M., Qiu, Q., Skuli, N., Nakazawa, M. S., Karakasheva, T., et al. (2013). Hypoxia-dependent Modification of Collagen Networks Promotes Sarcoma Metastasis. *Cancer Discov.* 3 (10), 1190–1205. doi:10.1158/2159-8290.cd-13-0118

El Beaino, M., Araujo, D. M., Lazar, A. J., and Lin, P. P. (2017). Synovial Sarcoma: Advances in Diagnosis and Treatment Identification of New Biologic Targets to Improve Multimodal Therapy. *Ann. Surg. Oncol.* 24 (8), 2145–2154. doi:10.1245/s10434-017-5855-x

Flores Bueso, Y., Walker, S. P., Hogan, G., Claesson, M. J., and Tangney, M. (2020). Protoblock - A Biological Standard for Formalin Fixed Samples. *Microbiome* 8 (1), 122. doi:10.1186/s40168-020-00901-1

Giannakouros, P., Matte, I., Rancourt, C., and Piché, A. (2015). Transformation of NIH3T3 Mouse Fibroblast Cells by MUC16 Mucin (CA125) Is Driven by its Cytoplasmic Tail. *Int. J. Oncol.* 46 (1), 91–98. doi:10.3892/ijo.2014.2707

Guo, X., and Rotter, J. I. (2019). Genome-Wide Association Studies. *Jama* 322 (17), 1705–1706. doi:10.1001/jama.2019.16479

Hale, R., Sandakly, S., Shipley, J., and Walters, Z. (2019). Epigenetic Targets in Synovial Sarcoma: A Mini-Review. *Front. Oncol.* 9, 1078. doi:10.3389/fonc.2019.01078

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA Circles Function as Efficient microRNA Sponges. *Nature* 495 (7441), 384–388. doi:10.1038/nature11993

He, J., Abdel-Wahab, O., Nahas, M. K., Wang, K., Rampal, R. K., Intlekofer, A. M., et al. (2016). Integrated Genomic DNA/RNA Profiling of Hematologic Malignancies in the Clinical Setting. *Blood* 127 (24), 3004–3014. doi:10.1182/blood-2015-08-664649

He, Y., Deng, Y., Peng, P., Wang, N., Wang, J., Ding, Z., et al. (2018). A Novel Messenger RNA and Long Noncoding RNA Signature Associated with the Progression of Nonmuscle Invasive Bladder Cancer. *J. Cell Biochem.* 120, 1. doi:10.1002/jcb.28089

Ishibe, T., Nakayama, T., Aoyama, T., Nakamura, T., and Toguchida, J. (2008). Neuronal Differentiation of Synovial Sarcoma and its Therapeutic Application. *Clin. Orthop. Relat. Res.* 466 (9), 2147–2155. doi:10.1007/s11999-008-0343-z

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: Integrating Viruses and Cellular Organisms. *Nucleic Acids Res.* 49, D545–D551. doi:10.1093/nar/gkaa970

Kang, T., Yau, C., Wong, C. K., Sanborn, J. Z., Newton, Y., Vaske, C., et al. (2020). A Risk-Associated Active Transcriptome Phenotype Expressed by Histologically normal Human Breast Tissue and Linked to a Pro-tumorigenic Adipocyte Population. *Breast Cancer Res.* 22 (1), 81. doi:10.1186/s13058-020-01322-6

Koneru, M., O'Cearbhaill, R., Pendharkar, S., Spriggs, D. R., and Brentjens, R. J. (2015). A Phase I Clinical Trial of Adoptive T Cell Therapy Using IL-12 Secreting MUC-16ecto Directed Chimeric Antigen Receptors for Recurrent Ovarian Cancer. *J. Transl Med.* 13, 102. doi:10.1186/s12967-015-0460-x

Kong, Y., Wang, Z., Huang, M., Zhou, Z., Li, Y., Miao, H., et al. (2019). CUL7 Promotes Cancer Cell Survival through Promoting Caspase-8 Ubiquitination. *Int. J. Cancer* 145 (5), 1371–1381. doi:10.1002/ijc.32239

Krieg, A. H., Hefti, F., Speth, B. M., Jundt, G., Guillou, L., Exner, U. G., et al. (2011). Synovial Sarcomas Usually Metastasize after >5 Years: a Multicenter Retrospective Analysis with Minimum Follow-Up of 10 Years for Survivors. *Ann. Oncol.* 22 (2), 458–467. doi:10.1093/annonc/mdq394

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The Biogenesis, Biology and Characterization of Circular RNAs. *Nat. Rev. Genet.* 20 (11), 675–691. doi:10.1038/s41576-019-0158-7

Li, G.-J., Zhao, G.-Q., Yang, J.-P., Zhou, Y.-C., Yang, K.-Y., Lei, Y.-J., et al. (2017). Effect of miR-1244 on Cisplatin-Treated Non-small Cell Lung Cancer via MEF2D Expression. *Oncol. Rep.* 37 (6), 3475–3483. doi:10.3892/or.2017.5624

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. *Nucl. Acids Res.* 42, D92–D97. doi:10.1093/nar/gkt1248

Li, Q., Zhang, L., Jiang, J., Zhang, Y., Wang, X., Zhang, Q., et al. (2019). CDK1 and CCNB1 as Potential Diagnostic Markers of Rhabdomyosarcoma: Validation Following Bioinformatics Analysis. *BMC Med. Genomics* 12 (1), 198. doi:10.1186/s12920-019-0645-x

Li, Y., Zang, A., Fu, J., and Jia, Y. J. (2021). Cullin7 in Tumor Development: a Novel Potential Anti-cancer Target. *Neoplasma* 68 (3), 572–579. doi:10.4149/neo_2021_201207n1324

Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., et al. (2015). Circular RNA Is Enriched and Stable in Exosomes: a Promising Biomarker for Cancer Diagnosis. *Cell Res.* 25 (8), 981–984. doi:10.1038/cr.2015.82

Liang, Z.-Z., Guo, C., Zou, M.-M., Meng, P., and Zhang, T.-T. (2020). circRNA-miRNA-mRNA Regulatory Network in Human Lung Cancer: an Update. *Cancer Cell Int* 20, 173. doi:10.1186/s12935-020-01245-4

Lin, X., Huang, C., Chen, Z., Wang, H., and Zeng, Y. (2020). CircRNA_100876 Is Upregulated in Gastric Cancer (GC) and Promotes the GC Cells' Growth, Migration and Invasion via miR-665/YAP1 Signaling. *Front. Genet.* 11, 546275. doi:10.3389/fgene.2020.546275

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

Lu, D., and Thum, T. (2019). RNA-based Diagnostic and Therapeutic Strategies for Cardiovascular Disease. *Nat. Rev. Cardiol.* 16 (11), 661–674. doi:10.1038/s41569-019-0218-x

Moek, K. L., de Groot, D. J. A., de Vries, E. G. E., and Fehrmann, R. S. N. (2017). The Antibody-Drug Conjugate Target Landscape across a Broad Range of Tumour Types. *Ann. Oncol.* 28 (12), 3083–3091. doi:10.1093/annonc/mdx541

Morris, J. S., Luthra, R., Liu, Y., Duose, D. Y., Lee, W., Reddy, N. G., et al. (2021). Development and Validation of a Gene Signature Classifier for Consensus Molecular Subtyping of Colorectal Carcinoma in a CLIA-Certified Setting. *Clin. Cancer Res.* 27 (1), 120–130. doi:10.1158/1078-0432.ccr-20-2403

Przybyl, J., Sciot, R., Wozniak, A., Schöffski, P., Vanspauwen, V., Samson, I., et al. (2014). Metastatic Potential Is Determined Early in Synovial Sarcoma Development and Reflected by Tumor Molecular Features. *Int. J. Biochem. Cell Biol.* 53, 505–513. doi:10.1016/j.biocel.2014.05.006

Qi, Y., Wang, N., Pang, L.-J., Zou, H., Hu, J.-M., Zhao, J., et al. (2015). Identification of Potential Mutations and Genomic Alterations in the Epithelial and Spindle Cell Components of Biphasic Synovial Sarcomas Using a Human Exome SNP Chip. *BMC Med. Genomics* 8, 69. doi:10.1186/s12920-015-0144-7

Qin, Y., Verdegaal, E. M. E., Siderius, M., Bebelman, J. P., Smit, M. J., Leurs, R., et al. (2011). Quantitative Expression Profiling of G-Protein-Coupled Receptors (GPCRs) in Metastatic Melanoma: the Constitutively Active Orphan GPCR GPR18 as Novel Drug Target. *Pigment Cell Melanoma Res.* 24 (1), 207–218. doi:10.1111/j.1755-148x.2010.00781.x

Riedel, R., Jones, R., Italiano, A., Bohac, C., Thompson, J., Mueller, K., et al. (2018). Systemic Anti-cancer Therapy in Synovial Sarcoma: A Systematic Review. *Cancers* 10 (11), 417. doi:10.3390/cancers10110417

Riggi, N., Cironi, L., and Stamenkovic, I. (2018). Synovial Sarcoma: when Epigenetic Changes Dictate Tumour Development. *Swiss Med. Wkly* 148, w14667. doi:10.4414/smw.2018.14667

Ruan, Y., Li, Z., Shen, Y., Li, T., Zhang, H., and Guo, J. (2020). Functions of Circular RNAs and Their Potential Applications in Gastric Cancer. *Expert Rev. Gastroenterol. Hepatol.* 14 (2), 85–92. doi:10.1080/17474124.2020.1715211

Sarkar, J. P., Saha, I., Lancucki, A., Ghosh, N., Wlasnowolski, M., Bokota, G., et al. (2020). Identification of miRNA Biomarkers for Diverse Cancer Types Using Statistical Learning Methods at the Whole-Genome Scale. *Front. Genet.* 11, 982. doi:10.3389/fgene.2020.00982

Schweiger, M. R., Kerick, M., Timmermann, B., Albrecht, M. W., Borodina, T., Parkhomchuk, D., et al. (2009). Genome-wide Massively Parallel Sequencing of Formaldehyde Fixed-Paraffin Embedded (FFPE) Tumor Tissues for Copy-Number- and Mutation-Analysis. *PLoS One* 4 (5), e5548. doi:10.1371/journal.pone.0005548

Shan, L., Zhao, M., Lu, Y., Ning, H., Yang, S., Song, Y., et al. (2019). CENPE Promotes Lung Adenocarcinoma Proliferation and Is Directly Regulated by FOXM1. *Int. J. Oncol.* 55 (1), 257–266. doi:10.3892/ijo.2019.4805

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., 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

Song, Y., Li, L., Ou, Y., Gao, Z., Li, E., Li, X., et al. (2014). Identification of Genomic Alterations in Oesophageal Squamous Cell Cancer. *Nature* 509 (7498), 91–95. doi:10.1038/nature13176

Soule, E. H. (1986). Synovial Sarcoma. *Am. J. Surg. Pathol.* 1, 78–82. doi:10.1002/mpo.2950090107

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., et al. (2011). The STRING Database in 2011: Functional Interaction Networks of Proteins, Globally Integrated and Scored. *Nucleic Acids Res.* 39, D561–D568. doi:10.1093/nar/gkq973

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

Thway, K., and Fisher, C. (2014). Synovial Sarcoma: Defining Features and Diagnostic Evolution. *Ann. Diagn. Pathol.* 18 (6), 369–380. doi:10.1016/j.anndiagpath.2014.09.002

Van Allen, E. M., Wagle, N., Stojanov, P., Perrin, D. L., Cibulskis, K., Marlow, S., et al. (2014). Whole-exome Sequencing and Clinical Interpretation of Formalin-Fixed, Paraffin-Embedded Tumor Samples to Guide Precision Cancer Medicine. *Nat. Med.* 20 (6), 682–688. doi:10.1038/nm.3559

von Mehren, M., Randall, R. L., Benjamin, R. S., Boles, S., Bui, M. M., Conrad, E. U., et al. (2016). Soft Tissue Sarcoma, Version 2.2016, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 14 (6), 758–786. doi:10.6004/jnccn.2016.0078

Wang, S., Zou, Z., Luo, X., Mi, Y., Chang, H., and Xing, D. (2018). LRH1 Enhances Cell Resistance to Chemotherapy by Transcriptionally Activating MDC1 Expression and Attenuating DNA Damage in Human Breast Cancer. *Oncogene* 37 (24), 3243–3259. doi:10.1038/s41388-018-0193-4

Wu, Y.-P., Lin, X.-D., Chen, S.-H., Ke, Z.-B., Lin, F., Chen, D.-N., et al. (2020). Identification of Prostate Cancer-Related Circular RNA through Bioinformatics Analysis. *Front. Genet.* 11, 892. doi:10.3389/fgene.2020.00892

Xu, J., Zhang, Z., Qian, M., Wang, S., Qiu, W., Chen, Z., et al. (2020). Cullin-7 (CUL7) Is Overexpressed in Glioma Cells and Promotes Tumorigenesis via NF-Kb Activation. *J. Exp. Clin. Cancer Res.* 39 (1), 59. doi:10.1186/s13046-020-01553-7

Yang, S., Yu, F., Lin, M., Sun, L., Wei, J., Lai, C., et al. (2020). Single-nucleotide Polymorphism Rs17548629 in RIPK1 Gene May Be Associated with Lung Cancer in a Young and Middle-Aged Han Chinese Population. *Cancer Cell Int.* 20, 143. doi:10.1186/s12935-020-01215-w

Yoshimatsu, Y., Noguchi, R., Tsuchiya, R., Sei, A., Sugaya, J., Iwata, S., et al. (2020). Establishment and Characterization of NCC-SS3-C1: a Novel Patient-Derived Cell Line of Synovial Sarcoma. *Hum. Cell* 33 (3), 877–885. doi:10.1007/s13577-020-00354-6

Zeng, Z., Xia, L., Fan, S., Zheng, J., Qin, J., Fan, X., et al. (2020). Circular RNA CircMAP3K5 Acts as a MicroRNA-22-3p Sponge to Promote Resolution of Intimal Hyperplasia via TET2-Mediated SMC Differentiation. *Circulation* 143 (4), 354–371. doi:10.1161/CIRCULATIONAHA.120.049715

Zhang, F., Zhao, X., Dong, H., and Xu, J. (2018). circRNA Expression Analysis in Lung Adenocarcinoma: Comparison of Paired Fresh Frozen and Formalin-Fixed Paraffin-Embedded Specimens. *Biochem. Biophysical Res. Commun.* 500 (3), 738–743. doi:10.1016/j.bbrc.2018.04.145

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., et al. (2016). Circular RNA Profiling Reveals an Abundant circHIPK3 that Regulates Cell Growth by Sponging Multiple miRNAs. *Nat. Commun.* 7, 11215. doi:10.1038/ncomms11215

Zhengwang, S., Qiyue, X., Quan, H., Mengchen, Y., Hongqiang, Z., Geng, C., et al. (2021). *Transcriptomic Profiling of Synovial Sarcoma Reveals Novel Transcriptional and Post-Transcriptional Abnormalities*. Shanghai: Research Square. doi:10.21203/rs.3.rs-205642/v1 CrossRef Full Text

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 Yao, He, Wang, Dong, Tu He Ta Mi Shi, Feng, Chen, Pang, Zou, Zhou, Li and Qi. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fmolb.2021.693651

Pan-Cancer Analysis Identified *C1ORF112* as a Potential Biomarker for Multiple Tumor Types

Jiaxuan Chen†, *Haoming Mai*†, *Haitao Chen*, *Bin Zhou*, *Jinlin Hou** and *De-Ke Jiang**

State Key Laboratory of Organ Failure Research, Guangdong Key Laboratory of Viral Hepatitis Research, Guangdong Institute of Liver Diseases, Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, China

C1ORF112 is an evolutionarily conserved gene across vertebrates. Over the last decade, studies have suggested that *C1ORF112* may play a role in tumorigenesis. Using The Cancer Genome Atlas datasets, we explored the role of *C1ORF112* across various tumor types in this study. In most tumor types, *C1ORF112* expression was increased in tumor tissues compared to corresponding non-tumor tissues. In patients with certain tumor types, higher *C1ORF112* expression was correlated with shorter overall survival, disease-free survival, and progression-free survival. Further analyses of *C1ORF112* genetic alteration data showed that *C1ORF112* amplification and mutations may have an impact on liver hepatocellular carcinoma and uterine corpus endometrial carcinoma prognosis. In cancers including lower grade glioma and adrenocortical carcinoma, *C1ORF112* expression was linked to cancer-associated fibroblast infiltration. Gene Ontology analysis showed that *C1ORF112* was co-expressed with genes involved in biological processes such as cell cycle and mitotic regulation. The protein interaction network demonstrated that *C1ORF112* physically interacted with *RAD51*, *DMC1*, and *FIGNL1*, which have well characterized functions in DNA repair and cell cycle regulation. This pan-cancer study revealed the prognostic value and oncogenic role of *C1ORF112* across multiple tumor types.

Keywords: C1ORF112, prognosis, pan-cancer, biomarker, cell cycle

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Alessandra Bettiol,
University of Florence, Italy
Yanjie Zhao,
Qingdao University, China

****Correspondence:***
De-Ke Jiang
dekejiang17@smu.edu.cn
Jinlin Hou
jlhousmu@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *11 April 2021*
Accepted: *26 July 2021*
Published: *19 August 2021*

Citation:
Chen J, Mai H, Chen H, Zhou B, Hou J and Jiang D-K (2021) Pan-Cancer Analysis Identified C1ORF112 *as a Potential Biomarker for Multiple Tumor Types. Front. Mol. Biosci. 8:693651. doi: 10.3389/fmolb.2021.693651*

INTRODUCTION

Cancer is the most common cause of death in the world (Sung et al., 2021). To widen the potential therapy options for malignancies, a better knowledge of carcinogenesis and tumor progression through the identification of oncogenes is crucial. Large-scale and multi-omics cancer datasets, such as The Cancer Genome Atlas (TCGA) (ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020; Ye et al., 2021; Lv et al., 2021; Cheng et al., 2021; Cui et al., 2020), have made pan-cancer analysis possible in the last decade.

Chromosome 1 Open Reading Frame 112 (*C1ORF112*), also known as *FLJ10706* (Howe et al., 2021), is evolutionarily conserved, especially in primates (Edogbanya et al., 2021). Mice lacking *BC055324* (the mouse ortholog for *C1ORF112*) are embryonic lethal (https://www.mousephenotype.org/data/genes/MGI:3590554), indicating that *BC055324* is required for embryonic development. According to van Dam et al., genes enriched in the BRCA–Fanconi anemia-related DNA damage response pathway, such as *BRCA1* and *FANCI*, were shown to be co-expressed with *C1ORF112* (van Dam et al., 2012). And the

dysregulation of this pathway is linked to a higher risk of cancer (Nalepa and Clapp, 2018). Furthermore, *C1ORF112* knockdown inhibited cell proliferation of HeLa cells, implying *C1ORF112* may have a role in cancer (van Dam et al., 2012). Moreover, according to previous studies, *C1ORF112* was overexpressed in tumor tissues of stomach cancer and head and neck squamous cell carcinoma (Chen et al., 2020; Huang et al., 2020; Edogbanya et al., 2021). Besides, *C1ORF112* expression was found to be higher in tumor samples with mutant *TP53*, a well-known tumor suppressor, in a study of bladder cancer progression (Sanchez-Carbayo et al., 2007). To the best of our knowledge, no comprehensive analysis of *C1ORF112*'s function and clinical importance at the pan-cancer level has been done.

In this study, we systematically analyzed the expression status, prognostic value, genetic alteration, and molecular function of *C1ORF112* as well as the association with cancer-associated fibroblast infiltration in multiple tumor types.

MATERIALS AND METHODS

Gene Expression Analysis of C1ORF112

We constructed a *C1ORF112* mRNA expression plot using the Human Protein Atlas (HPA) database (version: 20.1) (https://www.proteinatlas.org/).

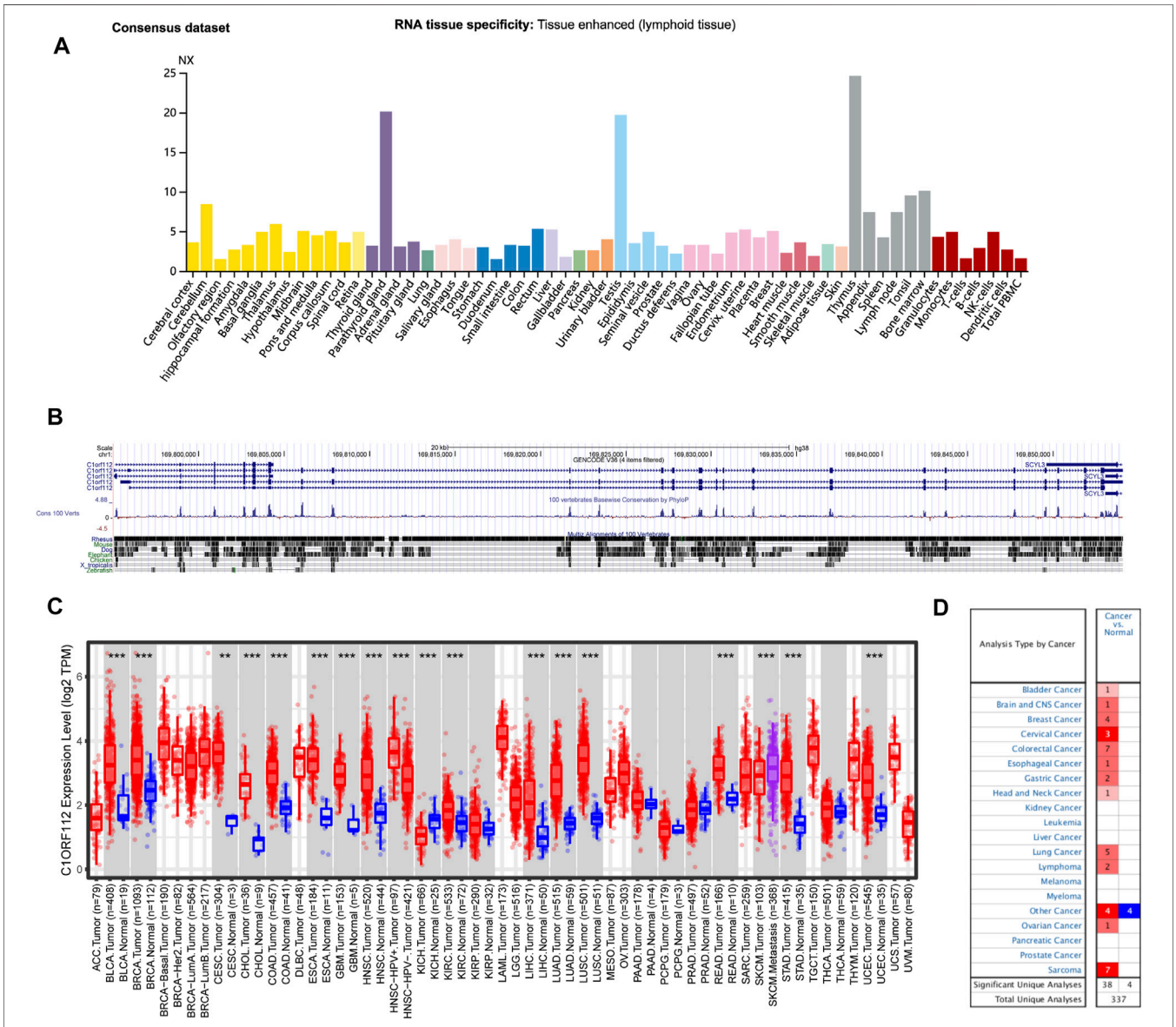


FIGURE 1 | *C1ORF112* expression status in different tumors and normal tissues and *C1ORF112* gene conservation among vertebrates. **(A)** Consensus *C1ORF112* tissue expression based on datasets of HPA (Human Protein Atlas), GTEx, and FANTOM5 (function annotation of the mammalian genome). **(B)** *C1ORF112* gene conservation analysis among vertebrates was visualized using the UCSC genome browser. **(C)** The expression status of *C1ORF112* in different tumor types was visualized by TIMER2. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. **(D)** Oncomine pooling analysis of C1ORF112 expression in various tumor types.

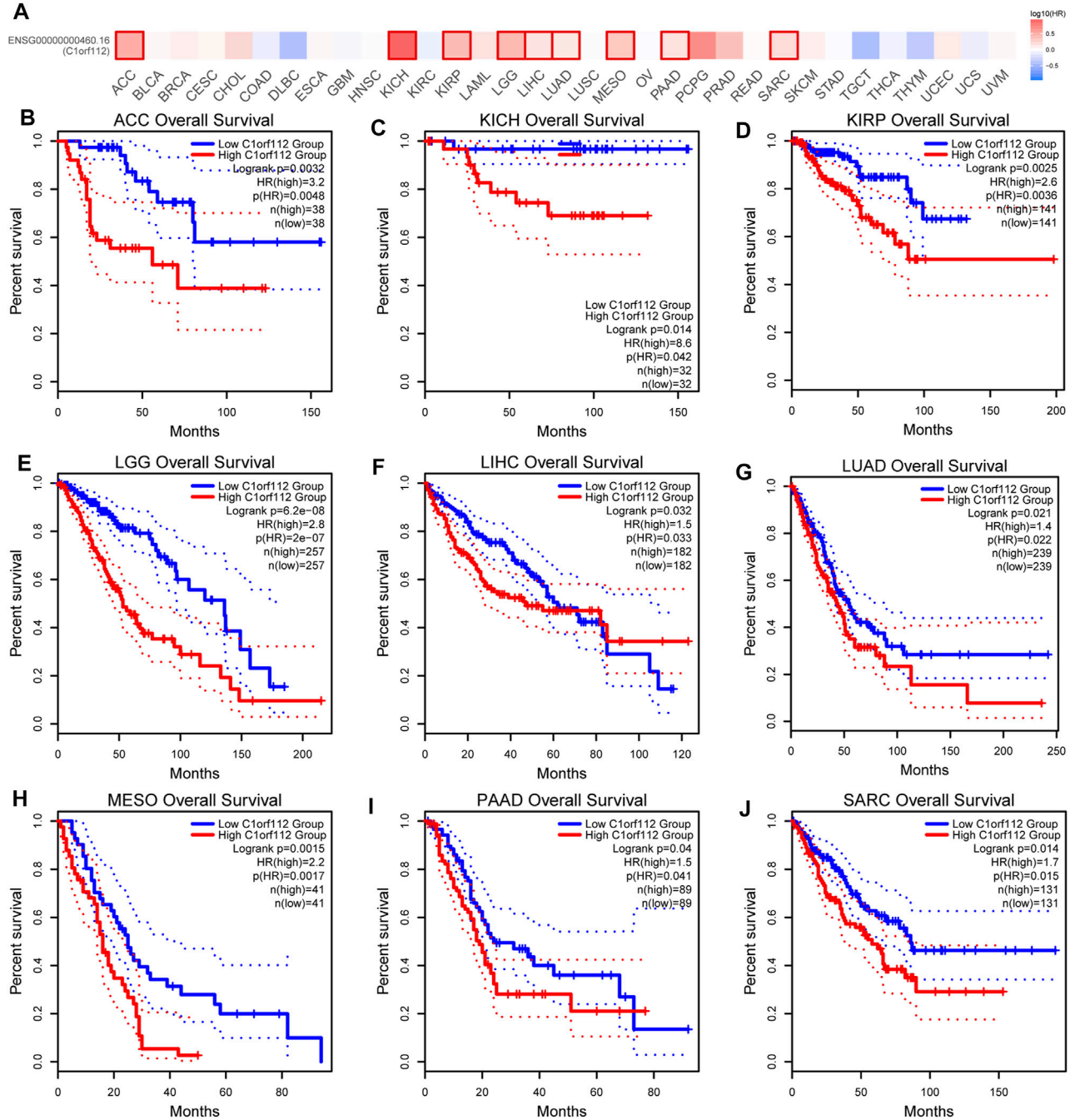


FIGURE 2 | Correlation between *C1ORF112* expression and overall survival in patients with different TCGA tumor types. GEPIA2 was used to build a survival map **(A)** and conduct overall survival analyses **(B–J)**. The survival map and Kaplan–Meier plots with significant results are displayed. The 95% confidence intervals of overall survival are indicated by red and blue dotted lines for high and low *C1ORF112* groups, respectively.

The "Gene DE" module of Tumor Immune Estimation Resource version 2 (TIMER2) (http://timer.cistrome.org/) was used to investigate *C1ORF112* expression differences in tumor and non-tumor tissues in various tumor types. In this module, *C1ORF112* expression was also evaluated between distinct breast cancer molecular subgroups, between HPV-positive and HPV-negative head and neck squamous cell carcinoma (HNSC), and between primary and metastatic skin cutaneous melanoma (SKCM).

We searched *C1ORF112* in the Oncomine database (Research Edition) (https://www.oncomine.org) to generate a pooled analysis of *C1ORF112* expression.

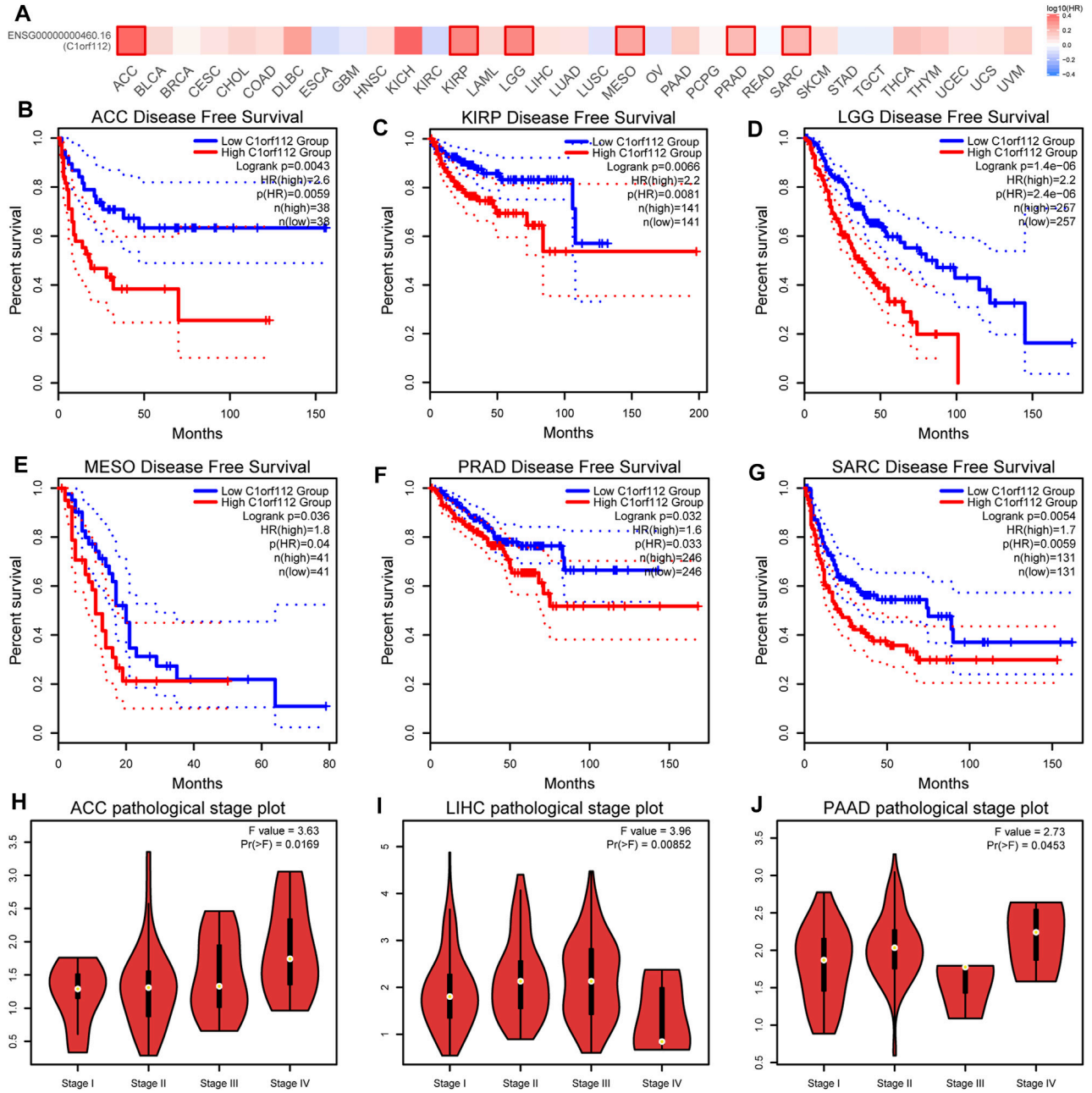


FIGURE 3 | Correlation between *C1ORF112* expression and disease-free survival in patients with different TCGA tumor types. GEPIA2 was used to build a survival map **(A)** and conduct disease-free survival **(B–G)** analyses. The survival map and Kaplan–Meier plots with significant results are displayed. The 95% confidence intervals of disease-free survival are indicated by red and blue dotted lines for high and low *C1ORF112* groups, respectively. **(H–J)** Correlation between *C1ORF112* expression and pathological stages of ACC, LIHC, and PAAD from TCGA datasets. Log2 (TPM + 1) was applied for log-scale.

Survival Prognosis Analysis

Overall survival (OS) and disease-free survival (DFS) Kaplan–Meier (K-M) plots, as well as a survival significance map of *C1ORF112* in all TCGA tumor types, were generated using the Gene Expression Profiling Interactive Analysis version 2 (GEPIA2) (http://gepia2.cancer-pku.cn/) "Survival Analysis" module. In addition, UCSC Xena Browser (https://xenabrowser.net/) was used to perform progression-free survival (PFS) analysis of *C1ORF112* using TCGA Pan-Cancer datasets (version: 2018–09–13). The expression threshold was set at 50% for high *C1ORF112* expression and low *C1ORF112* expression.

Genetic Alteration Analysis

C1ORF112 genetic alterations were analyzed using cBioPortal (version: 3.6.20) (https://www.cbioportal.org/). Based on datasets

of TCGA Pan-Cancer Atlas Studies, we calculated the frequency of *C1ORF112* gene mutation and copy number alteration in the "Cancer Types Summary" module. A mutation site plot of *C1ORF112* was created using the "Mutations" module.

To analyze the correlation between *C1ORF112* amplification status and liver hepatocellular carcinoma (LIHC) prognosis, the molecular profile was selected as copy number alterations based on "liver hepatocellular carcinoma (TCGA Pan-Cancer)" and the survival plot was generated by dividing cases based on the presence of copy number alterations (altered and unaltered groups).

To analyze the correlation between *C1ORF112* mutation status and uterine corpus endometrial carcinoma (UCEC) prognosis, the molecular profile was selected as mutations based on "uterine corpus endometrial carcinoma (TCGA Pan-Cancer)" and the survival plot was generated by dividing cases into altered and unaltered groups.

Immune Cells Infiltration Analysis

Using the "Immune" module of Tumor Immune Estimation Resource version 2 (TIMER2) (http://timer.cistrome.org/), Extended Polydimensional Immunome Characterization (EPIC) and the Tumor Immune Dysfunction and Exclusion (TIDE) algorithms were used to investigate the correlation between *C1ORF112* expression and cancer-associated fibroblast infiltration.

C1ORF112-Related Gene Enrichment Analysis

The STRING tool (version: 11.0b) (https://string-db.org/) was used to create a *Homo* Sapiens *C1ORF112* co-expression network using the following main parameters: 1) Active interaction sources: co-expression; 2) meaning of network edges: evidence; 3) maximum number of interactors: 50; and 4) minimum required interaction score: low confidence (0.150).

The GEPIA2 "Similar Gene Detection" module was used to extract 100 *C1ORF112*-correlated genes from the TCGA datasets that had the most similar expression pattern to *C1ORF112*. Gene Ontology pathway enrichment analysis was performed using the gene symbols of these 100 genes as input gene symbols in the "clusterProfiler" R package (version: 3.13). In addition, pairwise gene correlation analysis was performed using the GEPIA2 "Correlation Analysis" module.

C1ORF112-Protein Interaction Analysis

The "Network" module of BioGRID (version: 4.3) (https://thebiogrid.org/) was used to create a *C1ORF112*-protein interaction network, with the layout set to "Concentric Circles."

Conservation Analysis of C1ORF112

The UCSC genome browser (version: 2021 update) (http://www.genome.ucsc.edu/cgi-bin/hgTracks) was used to visualize gene conversation of *C1ORF112* among vertebrates.

RESULTS

Gene Expression Analysis of C1ORF112

Based on datasets of the HPA, GTEx, and FANTOM5 (function annotation of the mammalian genome), we found that *C1ORF112* was highly expressed in lymphoid tissue, such as thymus and bone, and enriched in testis and the parathyroid gland (**Figure 1A**; **Supplementary Figure S1**). Moreover, based on single-cell RNA-seq, high expression of *C1ORF112* was also observed in spermatocytes and spermatogonia (**Figure 1A**; **Supplementary Figure S1**). *C1ORF112* expression has a low tissue specificity, according to these findings. We also found that *C1ORF112* is relatively conserved among vertebrates (**Figure 1B**).

The expression pattern of *C1ORF112* in tumor tissues was then investigated. Similar to its distribution in normal tissue, *C1ORF112* mRNA distribution also showed low tumor specificity. *C1ORF112* mRNA expression was increased in various tumor tissues when compared to corresponding normal tissue (**Figure 1C**). Tumor tissues of bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), HNSC, kidney renal papillary cell carcinoma (KIRP), LIHC, lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), UCEC (all $p < 0.001$), cervical squamous cell carcinoma, and endocervical adenocarcinoma (CESC) ($p < 0.01$) had significantly higher *C1ORF112* expression when compared to corresponding normal tissue (**Figure 1C**). Significantly higher *C1ORF112* expression was observed in HPV-positive head and neck squamous cell tumor tissues compared with HPV-negative tissues ($p < 0.001$) (**Figure 1C**). Furthermore, when compared to primary SKCM tumor tissues, *C1ORF112* expression was also significantly elevated in metastatic SKCM tissues ($p < 0.001$) (**Figure 1C**). Meanwhile, significantly decreased *C1ORF112* expression was observed in kidney chromophobe (KICH) tumor tissues ($p < 0.001$) (**Figure 1C**).

We next used the Oncomine database to validate the differential *C1ORF112* expression between tumor and normal tissues. As shown in **Figure 1D**, significantly elevated *C1ORF112* expression was observed in most cancer types, including bladder cancer, brain and central nervous system cancer, breast cancer, cervical cancer, colorectal cancer, esophageal cancer, gastric cancer, head and neck cancer, lung cancer, lymphoma, ovarian cancer, and sarcoma.

Together these findings suggested that C1ORF112 may promote carcinogenesis in a variety of tumor types, and its clinical significance requires further investigation.

The Association Between C1ORF112 Expression and Prognosis of Patients With Cancer

To explore the potential prognostic value of *C1ORF112* based on TCGA datasets, we investigated the correlation between *C1ORF112* expression and prognosis of patients with different tumors by using GEPIA2. Higher *C1ORF112* expression was associated with shorter OS in cases of adrenocortical carcinoma (ACC) ($p = 3.2 \times 10^{-3}$), KICH ($p = 1.4 \times 10^{-2}$), KIRP ($p = 2.5 \times 10^{-3}$), LGG ($p = 6.2 \times 10^{-8}$), LIHC ($p = 3.2 \times 10^{-2}$), LUAD ($p = 2.1 \times 10^{-2}$), mesothelioma (MESO) ($p = 1.5 \times 10^{-3}$), pancreatic adenocarcinoma (PAAD) ($p = 4.0 \times 10^{-2}$), and sarcoma (SARC) ($p = 1.4 \times 10^{-2}$) (**Figure 2**). Furthermore, DFS

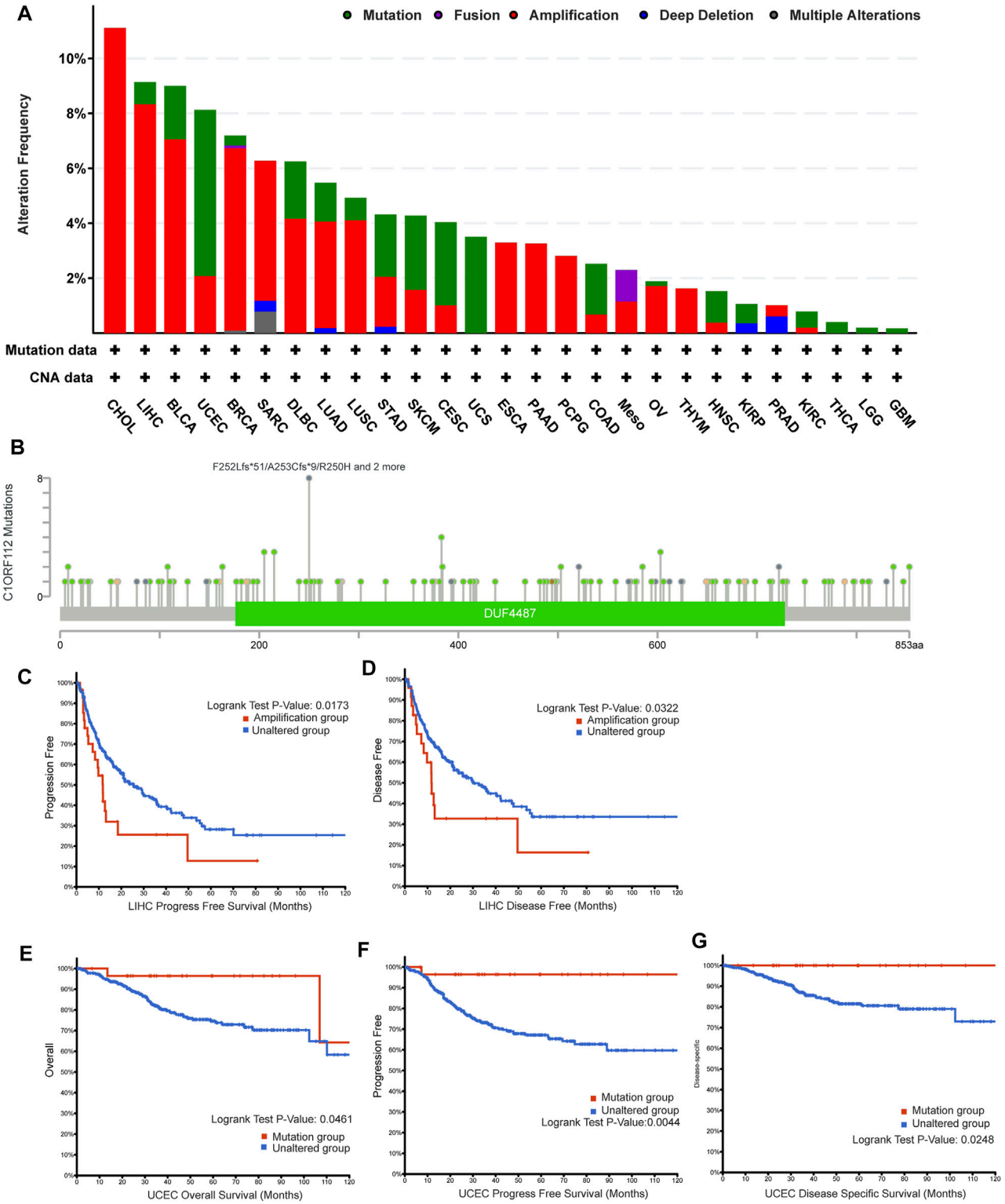


FIGURE 4 | *C1ORF112* genetic alteration in various tumor types of TCGA. The alteration frequency with *C1ORF112* genetic alteration type **(A)** and *C1ORF112* mutation site **(B)** were generated by cBioPortal. The correlations between *C1ORF112* amplification status and progression-free survival and disease-free survival of LIHC **(C,D)** were analyzed by cBioPortal. The correlations between mutation status and overall survival, progression-free survival, and disease-specific survival of UCEC **(E–G)** were analyzed by cBioPortal.

analysis showed that high *C1ORF112* expression was a marker for poor outcome for patients with ACC ($p = 4.3 \times 10^{-3}$), KIRP ($p = 6.6 \times 10^{-3}$), LGG ($p = 1.4 \times 10^{-6}$), MESO ($p = 3.6 \times 10^{-2}$), PRAD ($p = 3.2 \times 10^{-2}$), and SARC ($p = 1.4 \times 10^{-2}$) (**Figures 3A–G**). Using UCSC Xena Browser, a significant association was also noted between *C1ORF112* expression and PFS in several human cancers. In patients of ACC ($p = 9.9 \times 10^{-4}$), CHOL ($p = 4.1 \times 10^{-2}$), KICH ($p = 2.6 \times 10^{-2}$), KIRP ($p = 3.0 \times 10^{-5}$), LGG ($p = 7.9 \times 10^{-8}$), LUAD ($p = 7.7 \times 10^{-3}$), MESO ($p = 9.4 \times 10^{-4}$), PRAD ($p = 1.5 \times 10^{-2}$), SARC ($p = 4.3 \times 10^{-2}$), SKCM ($p = 1.9 \times 10^{-2}$), and UCEC ($p = 5.5 \times 10^{-4}$), higher C1ORF112 expression is associated with a worse prognosis (**Supplementary Figure S2**). These results indicated that increased *C1ORF112* expression was associated with poor prognosis in a variety of tumor types.

Moreover, we also investigated the correlation between *C1ORF112* expression and pathological stages of tumors by GEPIA2. The high expression of *C1ORF112* was significantly correlated with the advanced stage of ACC, LIHC, and PAAD (**Figures 3H–J**).

The Genetic Alteration Landscape of C1ORF112 in Different Tumors

The genetic alteration of *C1ORF112* in various tumor types in TCGA datasets was then investigated using cBioPortal. We found that CHOL tumor samples had the highest *C1ORF112* genetic alteration frequency (>10%). All of the genetic alterations occurring in CHOL tumor samples were copy number amplification (**Figure 4A**; **Supplementary Table S1**), which was the major type of genetic alteration in all TCGA tumor samples. In addition to CHOL cancer, more than 8% of LIHC, BRCA, and UCEC samples showed genetic alteration of *C1ORF112* (**Figure 4A**; **Supplementary Table S1**). As shown in **Figure 4B**, a total of 137 *C1ORF112* mutations, including 111 missense mutations, 21 truncating mutations, 4 fusion mutations, and 1 in-frame mutation, were detected in TCGA tumor samples (**Supplementary Table S2**). The residues 250–253 of the protein encoded by *C1ORF112* had nine mutations, making it the most frequently mutated region in the *C1ORF112* protein (**Figure 4B**).

Following that, we explored the link between *C1ORF112* genetic alterations and clinical outcomes of cancer patients. *C1ORF112* amplification was associated with poor prognosis in LIHC patients in terms of PFS ($p = 1.7 \times 10^{-2}$) and DFS ($p = 3.2 \times 10^{-2}$) (**Figures 4C,D**; **Supplementary Table S3**). In addition, UCEC patients with *C1ORF112* mutations (6.05%; 32 cases) showed a better prognosis in terms of OS ($p = 4.6 \times 10^{-2}$), PFS ($p = 4.4 \times 10^{-3}$), and DFS ($p = 2.5 \times 10^{-2}$) (**Figures 4E–G**; **Supplementary Table S4**).

Cancer-Associated Fibroblast Infiltration Analysis

Previous studies have found that cancer-associated fibroblasts in the stroma are involved in the regulation of different tumor-infiltrating immune cells (Chen and Song, 2019). We therefore employed the EPIC and TIDE algorithms to investigate the correlation between cancer-associated fibroblast infiltration and C1ORF112 expression in different malignancies. *C1ORF112* expression was positively correlated with cancer-associated fibroblast infiltration in ACC, CESC, KIRC, KIRP, LGG, MESO, and thyroid carcinoma (THCA) (**Figure 5**).

C1ORF112-Related Gene Enrichment Analysis

To investigate the functional mechanism of *C1ORF112* in carcinogenesis, we used GEPIA2 to extract the top 100 genes with expression patterns similar to *C1ORF112* from all tumor types in the TCGA datasets (**Supplementary Table S5**). Gene Ontology enrichment analysis indicated that these genes were closely linked to cell cycle or mitosis regulation (**Figure 6A**). Following that, 50 genes co-expressed with *C1ORF112* were obtained by the STRING tool to validate the result of Gene Ontology enrichment analysis. As shown in **Figure 6B**, the correlations of these 50 genes were mutually close; also, the genes were also enriched in cell cycle and mitotic regulation (**Supplementary Table S6**). These findings prompted us to wonder whether *C1ORF112* plays a role in these biological processes by interacting with key proteins involved in cell cycle and mitotic regulation. According to the BioGRID4.3 database, *C1ORF112* physically interacts with *RAD51*, *DMC1*, and *FIGNL1* (**Figure 6C**), which have well-characterized functions in the cell cycle, mitotic regulation, and tumorigenesis (**Supplementary Figure S3**) (Yuan and Chen, 2013; Laurini et al., 2020). Moreover, *C1ORF112* expression is strongly correlated with expression levels of RAD51 and FIGNL1 (**Figures 6D,E**). Based on these results, we speculate that *C1ORF112* may play a tumor-promoting role in cancers by driving the cell cycle and facilitating cell division.

DISCUSSION

TCGA project has profiled 33 prevalent tumor types with multi-omics data and provided an unprecedented opportunity to discover molecular aberrations at the pan-cancer level (Malta et al., 2018; Way et al., 2018; Schaub et al., 2018; Weinstein et al., 2013). Thanks to the development of bioinformatics algorithms and databases (Cerami et al., 2012; Carithers and Moore, 2013; Tang et al., 2019), numerous studies have been undertaken in recent years to identify pan-cancer molecular biomarkers and their functional roles (Ye et al., 2021; Lv et al., 2021; Hong et al., 2020). In this study, we analyzed the prognostic value and oncogenic role of *C1ORF112* in a variety of tumor types.

C1ORF112 is located at chromosome 1q24.2 and its gene synonym is *FLJ10706*. *C1ORF112* encodes nine transcripts, five of which are protein-coding (Howe et al., 2021). The *C1ORF112* protein with the longest amino acid sequence is translated from two distinct transcripts, ENST00000286031 and ENST00000359326 (Howe et al., 2021). According to the Ensembl database, *C1ORF112* orthologues have been found in 194 species. The protein-coding sequence of *C1ORF112* has a relatively high level of conservation among vertebrates, and a

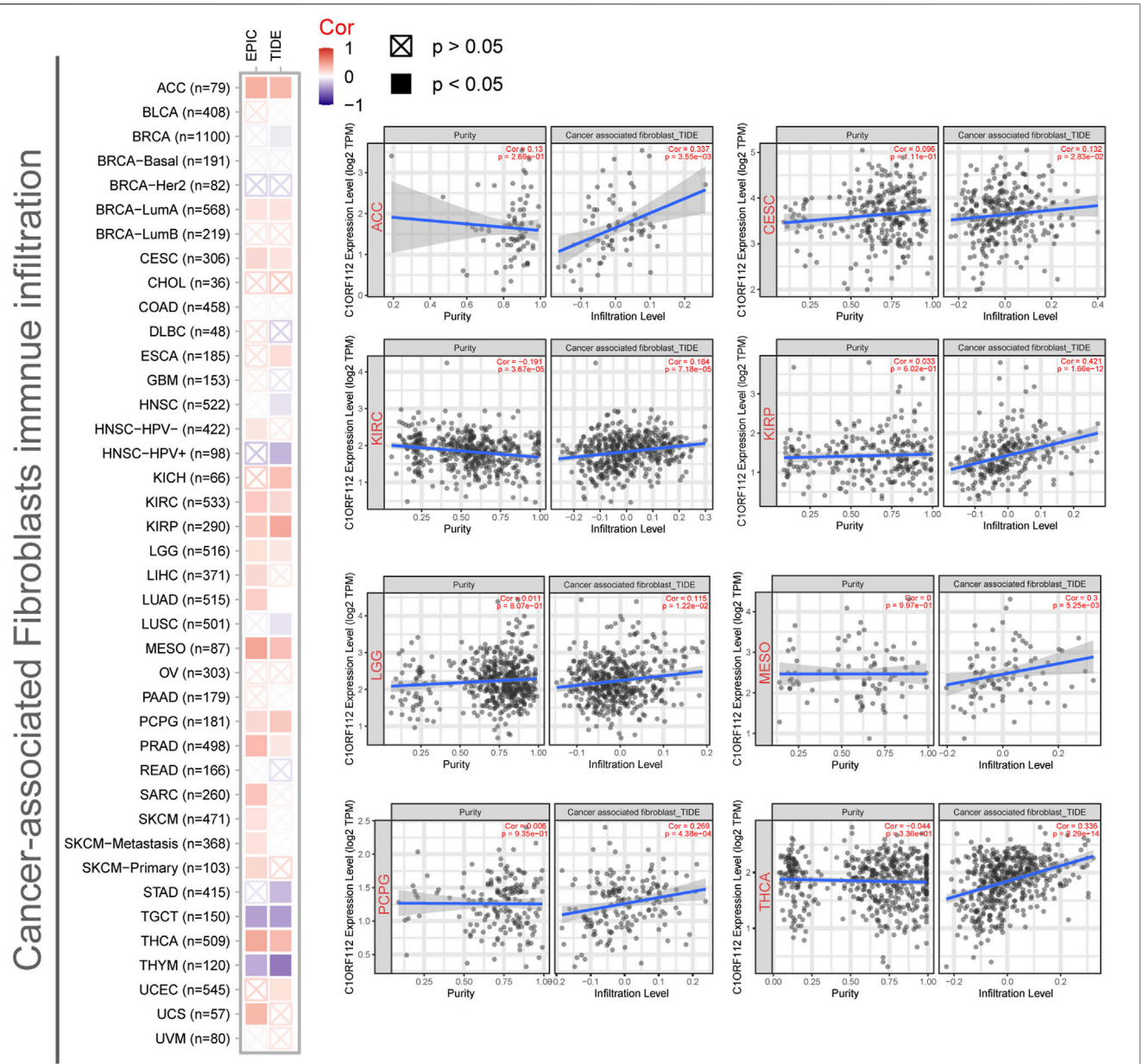


FIGURE 5 | Correlation between *C1ORF112* expression and cancer-associated fibroblast immune infiltration. EPIC and TIDE algorithms were used to calculate the correlation between *C1ORF112* expression and cancer-associated fibroblast immune infiltration in all tumor types from TCGA.

previous study revealed that *C1ORF112* might have evolved from the ancestors of eukaryotes (Edogbanya et al., 2021). Mice lacking *BC055324* (the mouse ortholog of *C1ORF112*) are embryonic fatal, and heterozygous mice have lower bone mineral density, circulating glucose level, and cardiac output (https://www.mousephenotype.org/data/genes/MGI:3590554). In the last decade, *C1ORF112* has drawn attention for its potential role in tumorigenesis. Van Dam et al. found that *C1ORF112* knockdown in HeLa cancer cells significantly lowered cell growth rate (van Dam et al., 2012). A recent study revealed that *C1ORF112* was co-expressed with stem cell-related genes, and these genes had elevated expression in gastric cancer tissues (Huang et al., 2020). Another study identified *C1ORF112* as one of the genes in a nine-gene risk model for predicting prognosis in patients with gastric cancer (Chen et al., 2020). However, the significance of *C1ORF112* in various tumor types has not been comprehensively explored. Therefore, we systemically characterized *C1ORF112* in 33 TCGA tumor types by analyzing features such as gene expression, genetic alteration, and immune infiltration.

In the present study, we found that *C1ORF112* is widely expressed in a variety of tissues and *C1ORF112* expression is upregulated in the majority of tumors. We further explored the relationship between *C1ORF112* overexpression and clinical

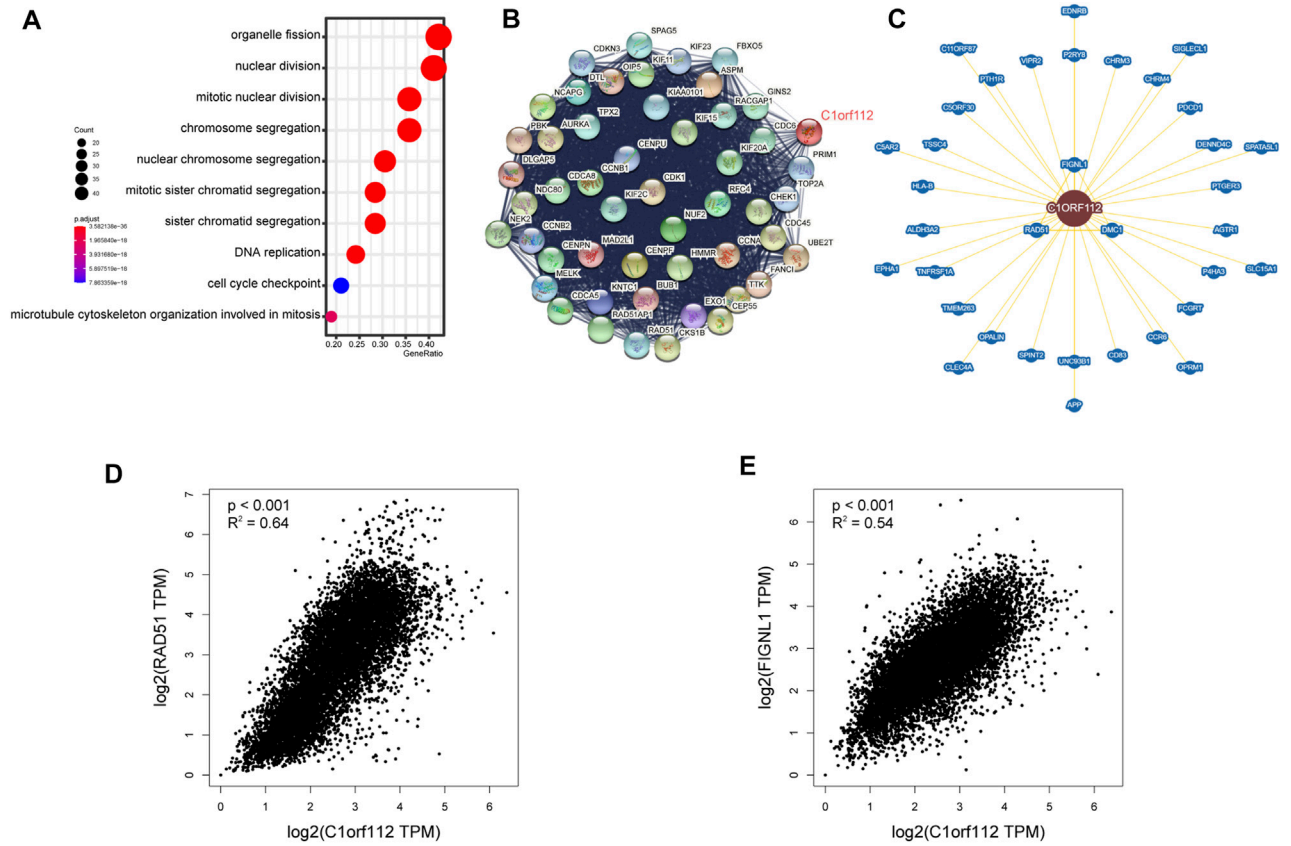


FIGURE 6 | *C1ORF112*-related gene enrichment analysis. **(A)** Gene Ontology (GO) analysis of the top 100 genes co-expressed with *C1ORF112* obtained by the GEPIA2. **(B)** Co-expression network of 50 genes co-expressed with *C1ORF112* obtained by the STRING tool. **(C)** *C1ORF112*-protein interactions obtained by BioGRID. **(D,E)** Correlation analysis between *C1ORF112* and *RAD51* and *FIGNL1* conducted by GEPIA2 across all tumor samples from TCGA.

parameters or prognosis. Survival analysis revealed that *C1ORF112* overexpression was associated with poor OS, DFS, and PFS. High *C1ORF112* expression was associated with poor prognosis in different tumor types involving ACC, CHOL, KIHC, KIRP, LGG, LUAD, MESO, PAAD, and SARC. Furthermore, upregulated *C1ORF112* expression is also significantly associated with the advanced cancer stage suggesting malignant progression. Growing evidence indicates that genomic mutations involved tumor progression and chemotherapy response (Yang et al., 2011; Cha et al., 2021). For example, Yang et al. reported that *BRCA1* and *BRCA2* mutations are significantly associated with patient survival, which may be a result of distinct response to platinum-based treatment (Yang et al., 2011). A large-scale study identified that mutations in four genes (ESR1, CDH1, RICTOR, and TP53) tended to occur in specific metastatic sites, which could be biomarkers or therapeutic targets of metastatic breast cancer patients (Cha et al., 2021). In this study, we revealed that mutations of *C1ORF112* were most common in CHOL (>10%), followed by LIHC, BRCA, and UCEC. To analyze whether *C1ORF112* genetic alterations have an impact on clinical outcomes of cancer patients, we discovered that *C1ORF112* amplification could be a risk factor for patients with liver cancer, while *C1ORF112* mutation may be protective in UCEC patients. Collectively, these findings indicate that *C1ORF112* acts as an oncogene in the progression of a variety of cancers and is a promising predictor for practical application in cancer prognosis.

Immune cells extensively intertwine with cancer cells and exert an essential effect on cancer migration and metastasis in various tumor types (Angelova et al., 2015). Recent studies have also reported that tumor immune microenvironment was associated with the expression level of various genes (Ju et al., 2020a; Ju et al, 2020b). In this study, we found *C1ORF112* expression was positively correlated with CAFs infiltration in several tumor types. CAFs are prominent components of stromal cells and have been reported to be associated with worse prognosis, chemotherapy resistance, and disease recurrence in various cancers (Calon et al., 2015; Ryner et al., 2015; Liu et al., 2016; Fiori et al., 2019). Taken together, our work elucidates the underlying effect of *C1ORF112* in tumor immunity and its prognostic values for multiple cancers.

Using STRING and GEPIA2, we identified a number of genes that were co-expressed with *C1ORF112* across different tumors and other tissues. Gene enrichment analysis revealed that these genes were strongly correlated with cell cycle or mitosis regulation, which was consistent with previous studies (van Dam et al., 2012;

Edogbanya et al., 2021). Moreover, our results showed that *C1ORF112* physically interacts with *RAD51*, *FIGNL1*, and *DMC1*. And the expressions of *RAD51* and *FIGNL*1 were strongly correlated with *C1ORF112* expression. *RAD51*, *FIGNL1*, and *DMC1* are well-characterized genes that encode proteins involved in DNA repair and cell cycle regulation (Yuan and Chen, 2013; Laurini et al., 2020). These findings validated the results of our gene enrichment analysis and paved the way for further exploration of *C1ORF112* molecular function.

This study had several limitations as well. First, the sample sizes for some uncommon tumor types were relatively small, which may cause batch effects or inaccurate results. Second, this study only provides preliminary findings linking *C1ORF112* to cancer progression in various tumors, and more experimental work is needed to determine the precise molecular function of *C1ORF112* in tumorigenesis.

In conclusion, *C1ORF112* is widely overexpressed in diverse cancers and its expression and genetic alteration are statistically associated with clinical outcomes in patients with certain tumors. Furthermore, immune infiltration analysis and *C1ORF112*-related gene enrichment analysis offer potential mechanisms by which *C1ORF112* may regulate tumor immunity, cell cycle, or DNA repair in cancers. Hence, further experimental and clinical studies are warranted to investigate *C1ORF112's* practical application in cancer therapy and prognosis prediction.

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

JC and HM performed the data analysis and prepared the draft of this manuscript. HC and BZ revised the draft. JH and D-KJ designed the research and supervised all the work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program (No. 2017BT01S131), the General Program from the Natural Science Foundation of Guangdong Province (No. 2019A1515011423), the General Programs from the National Natural Science Foundation of China (Nos. 81472618 and 81670535), the National Science and Technology Major Project (Nos. 2017ZX10202202 and 2018ZX10301202), the Key-Area Research and Development Program of Guangdong Province (No. 2019B020227004), the Innovative Research Team Project of Guangxi Province (No. 2017GXNSFGA198002), the Dean Fund of Nanfang Hospital, Southern Medical University (No. 2018Z005), the Grant for Recruited Talents to Start Scientific Research from Nanfang Hospital, and the Outstanding Youth Development Scheme of Nanfang Hospital, Southern Medical University (No. 2017J001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.693651/full#supplementary-material

REFERENCES

Angelova, M., Charoentong, P., Hackl, H., Fischer, M., Snajder, R., Krogsdam, A., et al. (2015). Characterization of the Immunophenotypes and Antigenomes of Colorectal Cancers Reveals Distinct Tumor Escape Mechanisms and Novel Targets for Immunotherapy. *Genome Biol.* 16, 64. doi:10.1186/s13059-015-0620-6

Calon, A., Lonardo, E., Berenguer-Llergo, A., Espinet, E., Hernando-Momblona, X., Iglesias, M., et al. (2015). Stromal Gene Expression Defines Poor-Prognosis Subtypes in Colorectal Cancer. *Nat. Genet.* 47 (4), 320–329. doi:10.1038/ng.3225

Carithers, L. J., and Moore, H. M. (2013). The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 45 (6), 580–585.doi:10.1038/ng.2653

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

Cha, S., Lee, E., and Won, H. (2021). Comprehensive Characterization of Distinct Genetic Alterations in Metastatic Breast Cancer across Various Metastatic Sites. *NPJ breast cancer* 7 (1), 93. doi:10.1038/s41523-021-00303-y

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

Chen, X., Zhang, D., Jiang, F., Shen, Y., Li, X., Hu, X., et al. (2020). Prognostic Prediction Using a Stemness Index-Related Signature in a Cohort of Gastric Cancer. *Front. Mol. Biosci.* 7, 570702. doi:10.3389/fmolb.2020.570702

Cheng, X., Wang, X., Nie, K., Cheng, L., Zhang, Z., Hu, Y., et al. (2021). Systematic Pan-Cancer Analysis Identifies TREM2 as an Immunological and Prognostic Biomarker. *Front. Immunol.* 12, 646523. doi:10.3389/fimmu.2021.646523

Cui, X., Zhang, X., Liu, M., Zhao, C., Zhang, N., Ren, Y., et al. (2020). A Pan-Cancer Analysis of the Oncogenic Role of Staphylococcal Nuclease Domain-Containing Protein 1 (SND1) in Human Tumors. *Genomics* 112 (6), 3958–3967. doi:10.1016/j.ygeno.2020.06.044

Edogbanya, J., Tejada-Martinez, D., Jones, N. J., Jaiswal, A., Bell, S., and Cordeiro, R. (2021). Evolution, Structure and Emerging Roles of C1ORF112 in DNA Replication, DNA Damage Responses, and Cancer. *Cell Mol Life Sci.* 78 (9), 4365–4376. doi:10.1007/s00018-021-03789-8

Fiori, M., Di Franco, S., Villanova, L., Bianca, P., Stassi, G., and De Maria, R. (2019). Cancer-associated Fibroblasts as Abettors of Tumor Progression at the Crossroads of EMT and Therapy Resistance. *Mol. Cancer* 18 (1), 70. doi:10.1186/s12943-019-0994-2

Hong, W., Gu, Y., Guan, R., Xie, D., Zhou, H., and Yu, M. (2020). Pan-cancer Analysis of the CASP Gene Family in Relation to Survival, Tumor-Infiltrating Immune Cells and Therapeutic Targets. *Genomics* 112 (6), 4304–4315. doi:10.1016/j.ygeno.2020.07.026

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* 49, D884–D891. doi:10.1093/nar/gkaa942

Huang, C., Hu, C.-G., Ning, Z.-K., Huang, J., and Zhu, Z.-M. (2020). Identification of Key Genes Controlling Cancer Stem Cell Characteristics in Gastric Cancer. *Wjgs* 12 (11), 442–459. doi:10.4240/wjgs.v12.i11.442

ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium(2020). Pan-cancer Analysis of Whole Genomes. *nature* 578 (7793), 82–93. https://pubmed.ncbi.nlm.nih.gov/?term=ICGC%2FTCGA+Pan-Cancer+Analysis+of+Whole+Genomes+Consortium%5BCorporate+Author%5D

Ju, Q., Li, X., Zhang, H., Yan, S., Li, Y., and Zhao, Y. (2020a). NFE2L2 Is a Potential Prognostic Biomarker and Is Correlated with Immune Infiltration in Brain Lower Grade Glioma: A Pan-Cancer Analysis. *Oxid. Med. Cell. longev.* 2020, 3580719. doi:10.1155/2020/3580719

Ju, Q., Li, X., Zhang, H., and Zhao, Y. (2020b). BRCA1-Associated Protein Is a Potential Prognostic Biomarker and Is Correlated with Immune Infiltration in Liver Hepatocellular Carcinoma: A Pan-Cancer Analysis. *Front. Mol. Biosci.* 7, 573619. doi:10.3389/fmolb.2020.573619

Laurini, E., Marson, D., Fermeglia, A., Aulic, S., Fermeglia, M., Pricl, S., et al. (2020). Role of Rad51 and DNA Repair in Cancer: A Molecular Perspective. *Pharmacol. Ther.* 208, 107492. doi:10.1016/j.pharmthera.2020.107492

Liu, L., Liu, L., Yao, H., Zhu, Z., Ning, Z., and Huang, Q. (2016). Stromal Myofibroblasts Are Associated with Poor Prognosis in Solid Cancers: A Meta-Analysis of Published Studies. *PloS one* 11 (7), e0159947. doi:10.1371/journal.pone.0159947

Lv, Z., Qi, L., Hu, X., Mo, M., Jiang, H., Fan, B., et al. (2021). Zic Family Member 2 (ZIC2): a Potential Diagnostic and Prognostic Biomarker for Pan-Cancer. *Front. Mol. Biosci.* 8, 631067. doi:10.3389/fmolb.2021.631067

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–e15. doi:10.1016/j.cell.2018.03.034

Nalepa, G., and Clapp, D. W. (2018). Fanconi Anaemia and Cancer: an Intricate Relationship. *Nat. Rev. Cancer* 18 (3), 168–185. doi:10.1038/nrc.2017.116

Ryner, L., Guan, Y., Firestein, R., Xiao, Y., Choi, Y., Rabe, C., et al. (2015). Upregulation of Periostin and Reactive Stroma Is Associated with Primary Chemoresistance and Predicts Clinical Outcomes in Epithelial Ovarian Cancer. *Clin. Cancer Res.* 21 (13), 2941–2951. doi:10.1158/1078-0432.ccr-14-3111

Sanchez-Carbayo, M., Socci, N. D., Richstone, L., Corton, M., Behrendt, N., Wulkfuhle, J., et al. (2007). Genomic and Proteomic Profiles Reveal the Association of Gelsolin to TP53 Status and Bladder Cancer Progression. *Am. J. Pathol.* 171 (5), 1650–1658. doi:10.2353/ajpath.2007.070338

Schaub, F. X., Dhankani, V., Berger, A. C., Trivedi, M., Richardson, A. B., Shaw, R., et al. (2018). Pan-cancer Alterations of the MYC Oncogene and its Proximal Network across the Cancer Genome Atlas. *Cell Syst.* 6 (3), 282–e2. doi:10.1016/j.cels.2018.03.003

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

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

van Dam, S., Cordeiro, R., Craig, T., van Dam, J., Wood, S. H., and de Magalhães, J. (2012). GeneFriends: an Online Co-expression Analysis Tool to Identify Novel Gene Targets for Aging and Complex Diseases. *BMC Genomics* 13, 535. doi:10.1186/1471-2164-13-535

Way, G. P., Sanchez-Vega, F., La, K., Armenia, J., Chatila, W. K., Luna, A., et al. (2018). Machine Learning Detects Pan-Cancer Ras Pathway Activation in the Cancer Genome Atlas. *Cell Rep.* 23 (1), 172–180.e3. doi:10.1016/j.celrep.2018.03.046

Weinstein, J. N., Collisson, E., Collisson, E. A., Mills, G. B., Shaw, K. R. M., 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

Yang, D., Khan, S., Sun, Y., Hess, K., Shmulevich, I., Sood, A. K., et al. (2011). Association of BRCA1 and BRCA2 Mutations with Survival, Chemotherapy Sensitivity, and Gene Mutator Phenotype in Patients with Ovarian Cancer. *JAMA* 306 (14), 1557–1565. doi:10.1001/jama.2011.1456

Ye, W., Luo, C., Liu, F., Liu, Z., and Chen, F. (2021). CD96 Correlates with Immune Infiltration and Impacts Patient Prognosis: A Pan-Cancer Analysis. *Front. Oncol.* 11, 634617. doi:10.3389/fonc.2021.634617

Yuan, J., and Chen, J. (2013). FIGNL1-containing Protein Complex Is Required for Efficient Homologous Recombination Repair. *Proc. Natl. Acad. Sci.* 110 (26), 10640–10645. doi:10.1073/pnas.1220662110

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 Chen, Mai, Chen, Zhou, Hou 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 in Molecular Biosciences

ORIGINAL RESEARCH
published: 24 August 2021
doi: 10.3389/fmolb.2021.695601

SMPDL3B Predicts Poor Prognosis and Contributes to Development of Acute Myeloid Leukemia

Huiqing Qu[1]* *and Ye Zhu*[2]

[1]Department of Blood Transfusion, Binzhou Medical University Hospital, Binzhou, China, [2]Department of Internal Medicine, People's Fifth Hospital of Jinan City Affiliated to Weifang Medical University, Jinan, China

Background: Acute myeloid leukemia (AML), characterized by the low cure rate and high relapse, urgently needs novel diagnostic or prognostic biomarkers and potential therapeutic targets. Sphingomyelin Phosphodiesterase Acid Like 3B (SMPDL3B) is a negative regulator of Toll-like receptor signaling that plays important roles in the interface of membrane biology and innate immunity. However, the potential role of SMPDL3B in human cancer, especially in AML, is still unknown.

Methods: The expression of SMPDL3B in AML samples was investigated through data collected from Gene Expression Omnibus (GEO). Association between SMPDL3B expression and clinicopathologic characteristics was analyzed with the chi-square test. Survival curves were calculated by the Kaplan–Meier method. Cox univariate and multivariate analyses were used to detect risk factors for overall survival. The biological functions of SMPDL3B in human AML were investigated both *in vitro* and *in vivo*.

Results: Expression of SMPDL3B mRNA was significantly upregulated in human AML samples and closely correlated to cytogenetics risk and karyotypes. Elevated expression of SMPDL3B was associated with poor overall survival and emerged as an independent predictor for poor overall survival in human AML. Blocked SMPDL3B expression inhibited AML cells growth both *in vitro* and *in vivo* via promoting cell apoptosis.

Conclusion: Taken together, our results demonstrate that SMPDL3B could be used as an efficient prognostic biomarker and represent a potential therapeutic target for human AML.

Keywords: acute myeloid leukemia, biomarker, SMPDL3B, prognosis, apoptosis

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Deepanshi Dhar,
Johnson & Johnson, United States
Md. Abu Musa,
Augusta University, United States

***Correspondence:**
Huiqing Qu
sdbzquhuiqing@126.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *15 April 2021*
Accepted: *26 July 2021*
Published: *24 August 2021*

Citation:
Qu H and Zhu Y (2021) SMPDL3B Predicts Poor Prognosis and Contributes to Development of Acute Myeloid Leukemia. Front. Mol. Biosci. 8:695601. doi: 10.3389/fmolb.2021.695601

INTRODUCTION

Acute myeloid leukemia (AML), characterized by the uncontrolled proliferation and accumulation of granulocyte or monocyte precursors in the bone marrow and peripheral blood, is the most common type of leukemia in adults (El Omri et al., 2020). It is estimated that there are 21,450 new cases of AML in the United States in 2019, with an annual incidence of 4.2 per 100,000 persons (Lai et al.,

Abbreviations: AML: acute myeloid leukemia; SMPDL3B: Sphingomyelin Phosphodiesterase Acid Like 3B; GEPIA: Gene Expression Profiling Interactive Analysis; GO: gene ontology; TUNEL: TdT-mediated dUTP Nick-End Labeling; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: biological processes; CC: cell component; MF: molecular function.

2019). Although new therapeutic approaches have improved outcomes in the treatment of AML in the last decades, AML has the lowest survival rate of all leukemia due to the high rate of relapse. The 5-year overall survival of patients with AML is still unsatisfactory (only 38% for younger patients, and <10% for older patients) (Kantarjian et al., 2021). Thus, the effective treatment of AML and novel personalized therapies for AML patients are urgently needed.

An increasing number of studies have revealed that AML displays a complex variety of genetic changes, which results in the malignant proliferation of AML cells and variable clinical prognosis of AML patients (Deng et al., 2018; Lai et al., 2019; Ma et al., 2019; Martin et al., 2019). These mutation and abnormal expression genes associated with AML provide significant prognostic information for determining the response to chemotherapy and survival outcome (DiNardo et al., 2021; Xuan et al., 2020; Yang et al., 2020). Consequently, a better understanding of these changes is essential for the effective treatment of AML and the design of novel personalized therapies.

The SMPDL3B (Sphingomyelin Phosphodiesterase Acid Like 3B) protein, a phosphodiesterase, plays important roles in cell membrane lipid-modulation and membrane fluidity (Mitrofanova et al., 2019; Yoo et al., 2015). SMPDL3B acts as a negative regulator of Toll-like receptor signaling and changes the cellular lipid composition and membrane fluidity in macrophages (Heinz et al., 2015). Moreover, excess SMPDL3B was reported to impair insulin receptor isoform B-dependent signaling by interfering with insulin receptor isoforms binding to caveolin-1 in podocytes in diabetic kidney disease (Mitrofanova et al., 2019). In addition, SMPDL3B modulated radiation-induced damage of human glomerular endothelial cells (Abou Daher et al., 2020) and renal podocytes (Ahmad et al., 2017). Recently, Frank W. and colleagues suggested that the elevated expression of SMPDL3B significantly correlated with poor survival of prostate cancer patients (Waldbillig et al., 2020). Moreover, the knockdown of SMPDL3B impaired the migration of PC3 cells (Waldbillig et al., 2020). However, the clinical significance and biological function of SMPDL3B in human AML have not been explored. In the present study, we aimed to investigate the expression of SMPDL3B in AML patients. Moreover, the association of SMPDL3B expression with clinical outcomes of AML patients was also explored. Furthermore, the roles of SMPDL3B in supporting AML cells growth were investigated both *in vitro* and *in vivo*.

MATERIALS AND METHODS

Cell Culture and Cell Growth/Apoptosis Assays

The human AML cell lines Kasumi-1, NB4, HL-60, THP-1, U937, MV4-11, and HEL cell lines were purchased from the Cell Bank of the Shanghai Institute for Biological Sciences (Chinese Academy of Sciences, Shanghai, China). The leukemia cell lines were grown in RPMI-1640 and HEK-293T cells were grown in high glucose DMEM medium supplemented with 10% fetal bovine serum. All the cells were maintained at cell culture incubator at 37°C and 5% CO_2. For the cell growth assays, the sorted GFP + AML cells (infected with Scramble or SMPDL3B shRNA lentivirus) were seeded in 24-well plate at the indicated numbers. The cell growth was evaluated by calculating the living cell number with TC20™ automated cell counter (Bio-Rad, Hercules, CA) by trypan blue dyeing at the indicated days. For the apoptosis assay, GFP + AML cells (infected with sgRNA lentivirus) were stained by Annexin V-PE/7-AAD apoptosis detection kit (#A213-01, Vazyme) as the instructions. At least 10,000 cells were collected by FACS to determine the percentage of the apoptotic cells. All the experiments were repeated three times.

In Silico Data Collection

The gene expression profile of SMPDL3B included 2096 blood or bone marrow samples of acute and chronic leukemia patients based on the platform of Affymetrix HG-U133 Plus 2.0 GeneChips was downloaded from the GEO (Gene Expression Omnibus) database (GSE13159) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13159). The cBio Cancer Genomics Portal (c-BioPortal) was used to download SMPDL3B mRNA expression and clinicopathological data in 200 AML patients (http://cbioportal.org).

GO and KEGG Pathway Enrichment Analysis

Candidate genes correlated with SMPDL3B in human AML patients were downloaded from the Gene Expression Profiling Interactive Analysis (GEPIA) website (http://gepia.cancer-pku.cn/). A total of 66 genes with the $|$Pearson correlation coefficient$|$ > 0.6, p < 0.05, were included and listed in Sup. **Table 1**. The gene ontology resource (http://geneontology.org/) was employed to perform GO (gene ontology) functional annotation and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis for these SMPDL3B correlated genes. The GO annotation analysis contained three categories, including cellular component (CC), biological process (BP), and molecular function (MF).

RNA Extraction and Quantitative Real-Time PCR (qRT-PCR) Analysis

Cells were lysed and the total RNAs were extracted by using TRIzol Reagent (Invitrogen, Carlsbad, CA). cDNAs were reverse transcribed with the iScript™ cDNA Synthesis Kit (Bio-Rad). SMPDL3B and reference GAPDH were amplified by qRT-PCRs performed in the QuantStudio 5 Real-Time PCR machine using the iTaq Universal SYBR Green Supermix (Bio-Rad). Relative SMPDL3B mRNA expression levels were calculated by using the ΔΔCt method, normalized to GAPDH. The following PCR primers were used: SMPDL3B forward: 5′-TGGTCAATGGGG CCAACAAT -3′ SMPDL3B reverse: 5′-GGTGGAAGGAGC TCAACCTT-3′ GAPDH forward: 5′-GAATGGG CAGCCGTTAGGAA-3′, GAPDH reverse: 5′-AAAAGCATC ACCCGGAGGAG-3’. All the primers used in this study were synthesized by Sangon Biotech (Shanghai, China).

TABLE 1 | Association between SMPDL3B expression and clinicopathological characteristics in AML patients.

Patient's parameters	SMPDL3B ^{low}, 64	SMPDL3B ^{high}, 64	*p*
Sex, male/female	35/29	33/31	0.7232
Age, <60/≥60	33/31	38/26	0.3739
WBC (median, range) $\times 10^9$/L	29 (0.4–224)	12.5 (1–203)	0.4222
Hemoglobin (median, range) g/dL	9 (6–13)	10 (7–13)	0.7582
Platelets (median, range) $\times 10^9$/L	47.5 (9–174)	42.5 (9–232)	0.2833
BM blasts (median range)%	22 (1–97)	45 (1–97)	0.0627
FAB classifications			0.0549
M0	5	8	
M1	11	18	
M2	14	17	
M3	6	9	
M4	17	6	
M5	11	4	
M6	0	1	
M7	0	1	
Cytogenetics risk			0.0014*
Favorable	15	14	
Intermediate	42	26	
Poor	7	24	
Karyotypes			<0.0001
Normal	40	25	
inv (16)	8	0	
t (8; 21)	1	5	
t (15; 17)	6	7	
Complex	2	16	
Others	7	11	
Gene mutations			
FLT3 mutation, P/N	19/45	18/46	0.8454
Activated RAS, P/N	5/59	3/61	0.4652
NPM1, P/N	17/47	13/51	0.4039
IDH1, P/N	11/53	12/52	0.8179

TABLE 2 | Univariate and multivariate Cox regression analysis of overall survival in AML patients.

Characteristics	HR	Univariate 95% CI	*p*	HR	Multivariate 95% CI	*P*
Sex (female vs. male)	1.047	0.6751–1.626	0.8373			
Age (≥60 vs. < 60)	2.719	1.979–5.061	< **0.0001**	1.883	1.130–3.140	**0.0152**
WBC (≥median vs. < median)	1.206	0.7756–1.913	0.3990			
Hemog (≥median vs. < median)	1.350	0.8809–2.145	0.1694			
Platelets (≥median vs. < median)	1.425	0.9245–2.222	0.1116			
BM blasts (≥median vs. < median)	1.262	0.8178–1.982	0.2912			
FAB classifications	///	///	**0.0146**	1.265	0.784–1.372	0.463
(M0 vs. M3)	3.562	1.334–12.43	**0.0183**			
(M1 vs. M3)	3.624	1.370–7.132	**0.0092**			
(M2 vs. M3)	2.787	1.037–5.792	**0.0467**			
(M4 vs. M3)	4.287	1.679–8.939	**0.0025**			
(M5-7 vs. M3)	3.643	1.371–9.821	**0.0127**			
Cytogenetics risk	///	///	**0.0001**	1.559	1.086–2.237	**0.0161**
(Poor vs. favorable)	3.313	1.764–5.754	**0.0002**			
(Intermediate vs. favorable)	2.481	1.249–4.050	**0.0077**			
Karyotypes			0.0662			
FLT3 (mutation vs. normal)	0.8857	0.5549–1.413	0.6145			
NPM1 (positive vs. negative)	1.121	0.6755–1.882	0.6493			
IDH1 (positive vs. negative)	2.423	1.739–8.236	**0.0011**	1.821	1.016–3.257	**0.0439**
RAS (positive vs. negative)	0.6608	0.3022–1.616	0.4091			
SMPDL3B (high vs. low)	2.139	1.367–3.287	**0.0009**	1.927	1.186–3.125	**0.0079**

Western Blot Assay

The cells were collected and lysed with RIPA lysate buffer containing 1 mmol/L PMSF (#ST505, Beyotime) and 0.1 g/L leupeptin (#SG 2012, Beyotime, Shanghai). Protein samples were quantified with BCA. For the Western blotting, 30 μg of total protein was loaded into SDS-PAGE gel and transferred to the nitrocellulose membrane. The membrane then was blocked and incubated with SMPDL3B antibody (#16552-1-AP, Proteintech) and HRP-conjugated secondary antibody. The housekeeping gene GAPDH was used as an internal control.

CRISPR–Cas9-Mediated Gene Knockout in AML Cells

SMPDL3B-knockout (SMPDL3B-KO) cell lines were performed as described previously (Yamauchi et al., 2018). Briefly, AML cells were infected with pCW-Cas9 (#50661, Addgene, Watertown, MA, United States) lentivirus and selected with 1 μg/ml puromycin treatment. Then, the cells were infected with the sgRNA lentivirus. Scramble control sgRNA (sgRNA: 5′-CCACACCTGTCTAGC ATGAC-3′) or SMPDL3B targeting sgRNA (sgRNA1: 5′-ATG GACTCATTACTAAGCCA-3′; sgRNA2: 5′- ATGGACTCATTA CTAAGCCA -3′) were cloned into the sgRNA plasmid pLenti-sgRNA (#89638, Addgene), individually. At 7 days after treatment with 1 μg/ml doxycycline, GFP + cells were seeded into a 96-well plate as a single cell per well. After cell expansion, knockout clones were verified by Western blotting.

Xenograft Tumor Model Assay

The animal assays were performed according to the guidelines and approval of the Ethical Committee of Binzhou Medical University. To establish xenograft models, 5×10^6 SMPDL3B-WT (left) or SMPDL3B-KO (right) THP-1 cells were subcutaneously injected into the flanks of 4-week-old female athymic BALB/c nude mice (Vital River, Beijing). At 10 days after injection, the size of the xenografted tumor was measured every 3 days by using a vernier caliper. At 25 days after injection, the mice were sacrificed, and the xenograft tumors were stripped. The xenograft tissues were subjected to TdT-mediated dUTP Nick-End Labeling (TUNEL) analysis.

TUNEL Assay

In order to assess the apoptosis of the xenograft tumors, the slices of the tumor tissue were baked and rehydrated, then antigen repair was performed. The DNA fragmentation was determined by the TdT-mediated dUTP nick end-labeling (TUNEL) Kit (#C1086, Beyotime, Shanghai, China.) according to the manufacturer's instructions. Briefly, the slices were incubated in H_2O_2 solution and then washed by PBS. Next, the slices were incubated with immunostaining washing solution and TUNEL staining solution. The fluorescent images were observed under a fluorescence microscope (Nikon Corp., Tokyo, Japan). The results were analyzed by IMAGE J software.

Statistical Analysis

Statistical analysis was performed using SPSS 24.0 statistical software (SPSS Inc., Chicago, IL, United States). Data with normal distribution were expressed as mean ± S.D. *t*-test (2 groups), one-way ANOVA test (3 or more groups), Pearson's chi-square test, Spearman's correlation analysis, and Fisher's exact test were used to compare variables. The Kaplan–Meier method, log-rank test, and Cox's proportional hazards model were used for survival analysis.

RESULTS

SMPDL3B Is Highly Expressed in AML Patients and Correlates With Clinical Characteristics

To evaluate whether the expression of SMPDL3B is connected to AML evolution and advancement, we initially examined its expression by *in silico* analysis. Herein, the microarray data derived from the GEO database (GSE13159) were analyzed. SMPDL3B was expressed at significantly higher levels in almost all types of myeloid leukemia cells tested as compared to healthy bone marrow samples, whereas SMPDL3B was expressed at lower levels in lymphoid leukemia cells (**Figure 1A**). Consistently, the qRT-PCR analysis showed that SMPDL3B mRNA was highly expressed in several human AML cell lines, including Kasumi-1, THP-1, HL-60, MV4-11, and so forth, but not in U937 cells (**Figure 1B**). Furthermore, the SMPDL3B mRNA expression was tested in a cohort of AML patients according to its cytogenetics risk. The results indicated that the expression of SMPDL3B mRNA was significantly higher in the patients with poor cytogenetics, whereas it was relatively lower in the patients with good or intermediate cytogenetics (**Figure 1C**). Together, these results indicated that SMPDL3B mRNA was remarkably increased in AML cells. To further explore the correlation between SMPDL3B mRNA expression and clinical characteristics in human AML, patients were divided into two groups based on the expression level of SMPDL3B (low: below the 50th percentile; high: above the 50th percentile). Statistical analysis revealed that expression of SMPDL3B was closely correlated to cytogenetics risk ($p = 0.0014$) and karyotypes ($p < 0.0001$), and it was a trend toward FAB classifications ($p = 0.0549$). However, SMPDL3B mRNA expression was not related to other clinical characteristics ($p > 0.05$) (**Table 1**).

High Expression of SMPDL3B Predicts Poor Survival in Human AML Patients

To further detect the role of SMPDL3B in AML propagation, the overall survival rate was performed by the Kaplan–Meier analysis based on SMPDL3B expression. The Kaplan–Meier analysis revealed that the AML patients with high levels of SMPDL3B mRNA had a significantly poorer overall survival in the TCGA and GEPIA groups, respectively (**Figures 2A,B**). To test whether this finding was independent from the well-established prognostic indicators, such as age, FAB classification, cytogenetics risk, and karyotypes. Cox regression analyses (univariate and multivariate) of each of the clinicopathological

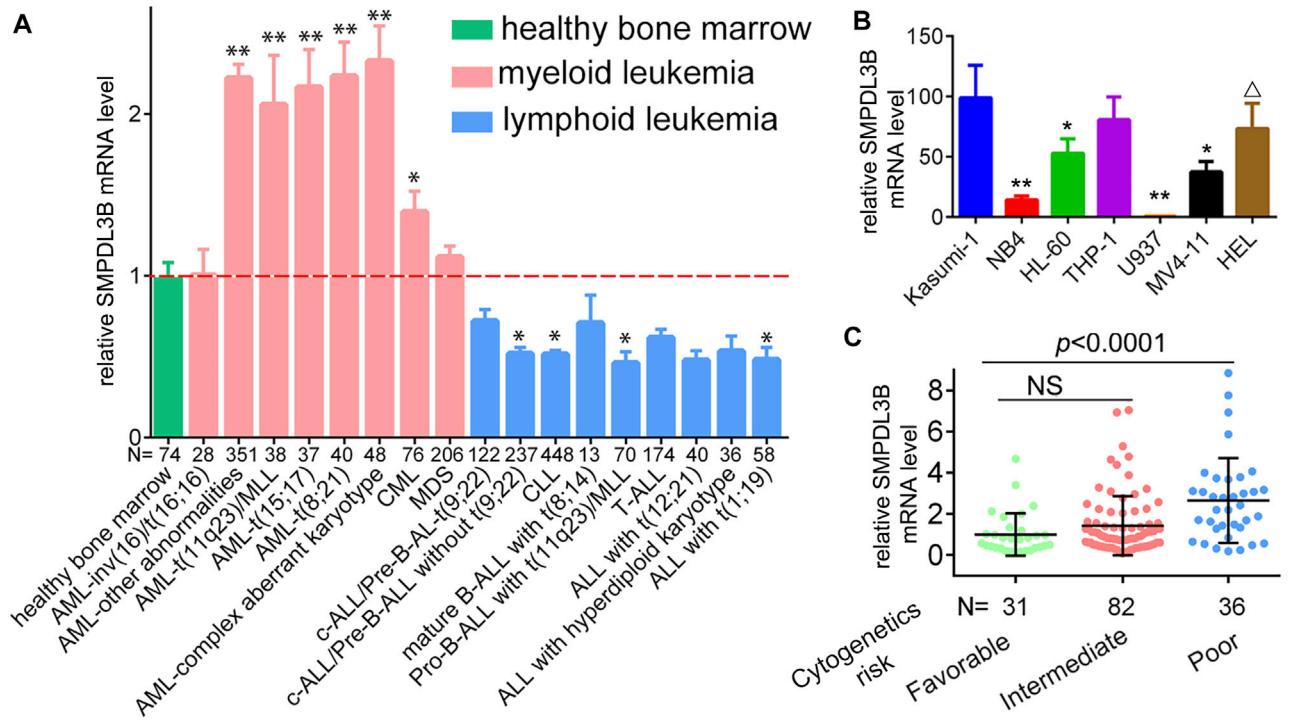


FIGURE 1 | SMPDL3B is highly expressed in AML patients and cell lines. **(A)** Relative mRNA expression of SMPDL3B in different karyotypes of AML vs. healthy samples. *p < 0.005, **p < 0.001 compared to healthy control. (Data were obtained from GSE13159.) **(B)** qRT-PCR analysis of SMPDL3B mRNA expression in the indicated cell lines. All the levels of SMPDL3B mRNA expression were normalized by GAPDH. **(C)** Relative SMPDL3B mRNA expression in human AML patients with indicated cytogenetics risk. The data were obtained from the TCGA AML database.

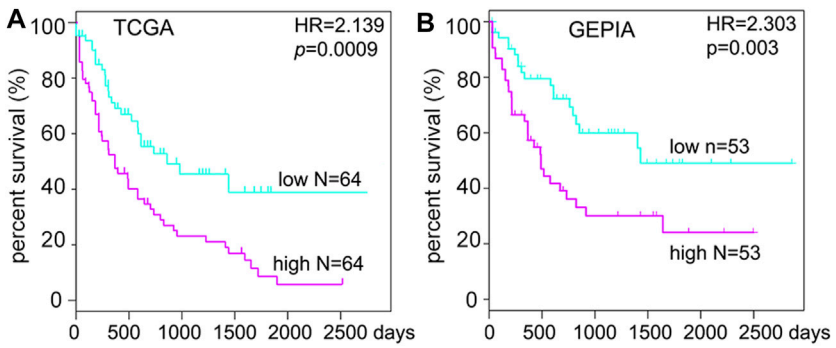


FIGURE 2 | SMPDL3B mRNA expression levels negatively correlate with the overall survival of human AML patients. Kaplan–Meier analysis of overall survival in AML patients relative to SMPDL3B mRNA expression levels. (High: above the 50th percentile; low: below the 50th percentile.) Data were obtained from the TCGA AML database **(A)** and GEPIA **(B),** respectively. p value was calculated by the log-rank test.

variables with SMPDL3B mRNA expression were conducted. Univariate analysis revealed that the overall survival of AML patients significantly correlated with age, FAB classifications, cytogenetics risk, IDH1, and SMPDL3B expression level (all p < 0.05). Further, multivariate analysis was used to analyze all the statistically significant variables revealed by univariate analysis. SMPDL3B mRNA expression level (HR = 1.927, p = 0.0079), together with age, cytogenetics risk, and IDH1, was a significant independent prognostic factor for overall survival of AML patients. Consistent with these results, the overall survival rates were significantly different between SMPDL3B low and SMPDL3B high AML patients in preplanned age <60, age ≥60, cytogenetics risk favorable, cytogenetics intermediate, and IDH1 negative subgroups (**Figures 3A–D,G**). However, there was no difference of the overall survival in cytogenetics poor and IDH1 positive subgroups; such paradox might be due to the smaller sample size (**Figures 3E,F**). Together, these results suggested that high SMPDL3B mRNA expression served as an independent poor

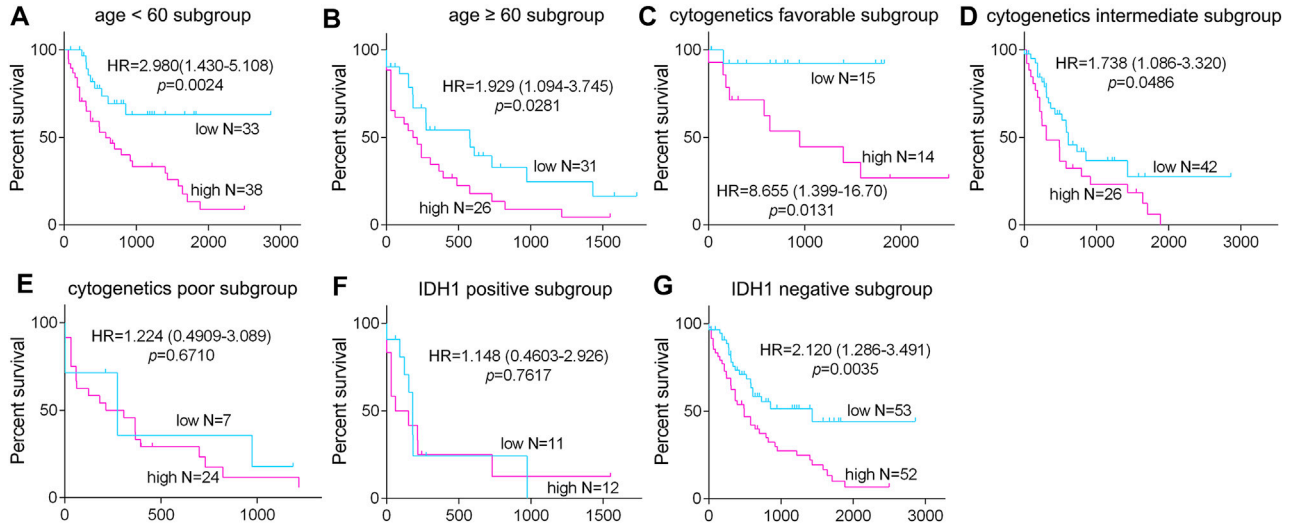


FIGURE 3 | The expression of SMPDL3B in AML subtypes and correlation with overall survival of the patients. The AML patients were divided into the indicated subgroups (**(A)**: age <60, **(B)**: age ≥60; **(C)**: cytogenetic favorable, **(D)**: cytogenetic intermediate, **(E)**: cytogenetic poor; **(F)**: IDH1 positive, and **(G)**: IDH1 negative) according to the age, cytogenetics risk, and IDH1 expression status, respectively. The Kaplan–Meier overall survival in the indicated subgroups of AML patients with SMPDL3B low and SMPDL3B high expression is plotted respectively. *p* value was calculated by log-rank test.

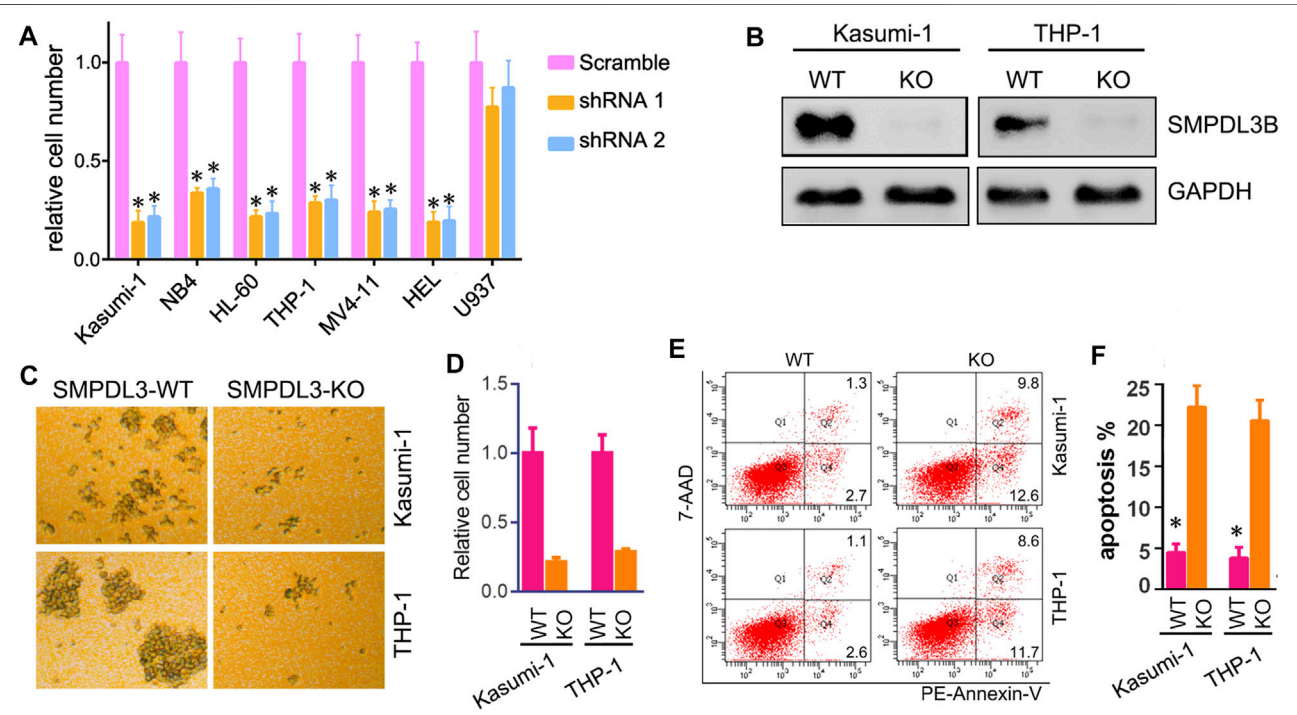


FIGURE 4 | Inhibition of SMPDL3B suppressed growth and promoted apoptosis of AML cells. **(A)** The indicated AML cells were infected with Scramble or SMPDL3B shRNA lentivirus. The cell numbers were counted at 4 days after infection. The relative cell number is plotted. $n = 5$, $^*p < 0.001$ compared to the Scramble group. **(B)** Western blot analysis of the expression of SMPDL3B in CRISPR/Cas9-edited AML cells. **(C)** Representative images of Kasumi-1 and THP-1 cells with wild type (WT) or knockout (KO) SMPDL3B. **(D)** The indicated SMPDL3B-WT or KO cells were seeded in 24-well plate. Cell numbers were counted at 4 days after being seeded. The relative cell number is plotted. $n = 5$, $p < 0.001$ compared to WT group. **(E)** Representative scatter plots showing apoptosis in SMPDL3B-WT and SMPDL3B-KO AML cells by annexin V-PE and 7-AAD staining and flow cytometry. **(F)** Quantification of the apoptotic cells. $n = 3$, $^*p < 0.001$ SMPDL3B-WT vs. SMPDL3B-KO.

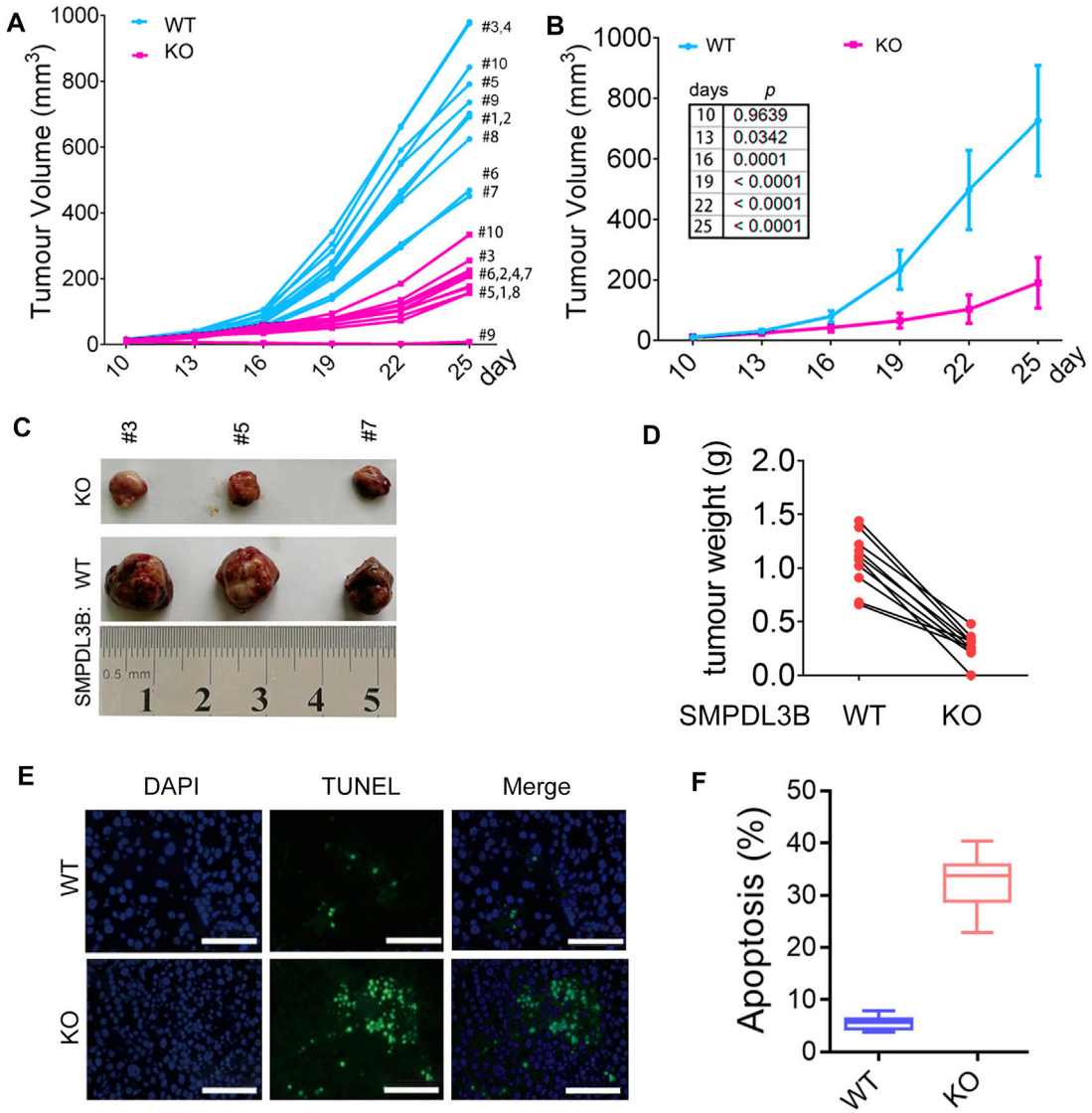


FIGURE 5 | Knockdown of SMPDL3B expression suppressed the growth of AML xenograft tumors *in vivo*. The nude mice received a subcutaneous injection of SMPDL3B-WT or SMPDL3B-KO Kasumi-1 cells were sacrificed on day 25 after injection. **(A)** The tumor volume of each xenograft at the indicated days after injection was shown. **(B)** The average volume of the tumor xenografts was shown. Data were presented as mean ± S.D. (n = 10). **(C)** Representative images of the isolated tumors. **(D)** Each paired weight of the tumor xenografts was shown (n = 10, paired t-test, $p < 0.001$). **(E)** Representative images of the TUNEL assays for the tumors were shown (bars, 100 μM). **(F)** Percent apoptosis of the SMPDL3B-WT and KO tumors. n = 5, mean ± S.D. $p < 0.01$.

prognostic biomarker associated with decreased overall survival in human AML patients.

Downregulation of SMPDL3B Inhibited Growth of Leukemia Cells Both *In Vitro* and *In Vivo*

To further study the biological function of SMPDL3B in AML cells, we knocked down the expression of SMPDL3B via lentivirus-encoded shRNAs. The qRT-PCR results showed that the shRNA could efficiently decrease the expression of SMPDL3B (data not shown). Remarkably, the CCK-8 results indicated that knockdown of SMPDL3B expression could decrease the growth of each of those leukemia cell lines that express SMPDL3B (**Figure 4A**). In contrast, the tested shRNA did not influence the growth of U937 cells (**Figure 4A**). Since Kasumi-1 and THP-1 cells had the highest expression level of SMPDL3B of the cultured cells evaluated, we constructed SMPDL3B knockout Kasumi-1 and THP-1 leukemia cell lines via the CRISPR/Cas9 system and used these two cell lines in subsequent experiments (**Figure 4B**). As expected, knockout of SMPDL3B induced visible cell growth inhibition both in Kasumi-1 and THP-1 cells (**Figures 4C,D**). Notably, the knockout of SMPDL3B significantly increased the rate of apoptosis of AML cells (**Figures 4E,F**).

To further confirm whether SMPDL3B knockout had suppressive effects on AML tumor growth *in vivo*, the

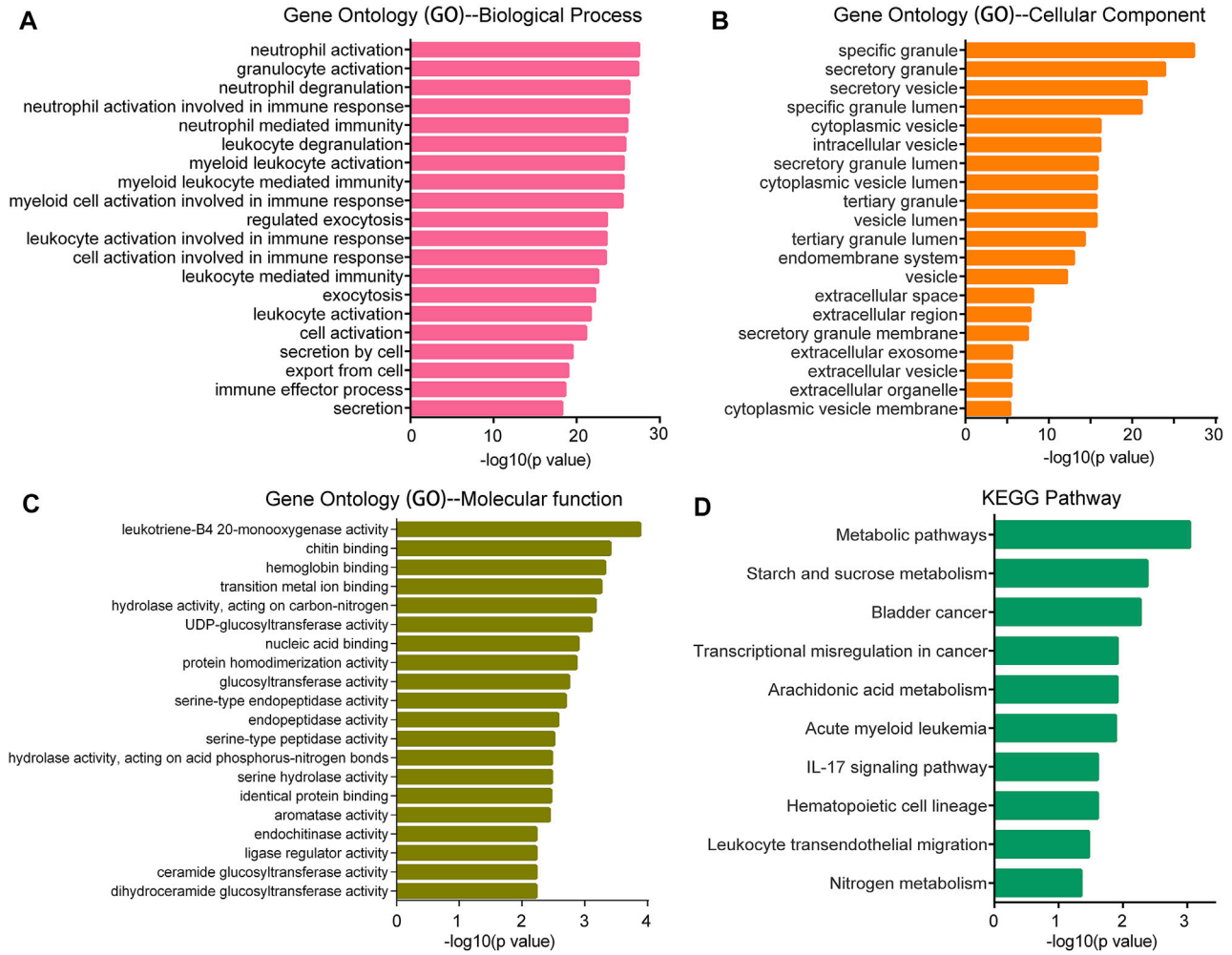


FIGURE 6 | GO and KEGG pathway enrichment analysis for the SMPDL3B correlated genes. The top 20 most enriched GO biological process **(A)**, GO cellular component **(B)**, GO molecular function **(C)** categories, and the top 10 most enriched KEGG pathways **(D)** are plotted.

Kasumi-1 cells with wild type SMPDL3B (SMDPL3B-WT) or knockout SMPDL3B (SMPDL3B-KO) were subcutaneously injected into nude mice for xenograft. Herein, each of the xenograft tumors grown from the SMPDL3B-KO cells had a smaller volume than that grown from SMPDL3B-WT cells (**Figures 5A,B**). On day 25 after injection, the mean tumor volume was 726.6 mm^3 in the WT group compared with 190.88 mm^3 in the KO group, with the inhibitory rate being 73.73% (**Figure 5B**). Furthermore, all tumors were removed, photographed (**Figure 5C**), and weighed (**Figure 5D**) on day 25 after injection. The average tumor weights in WT and KO groups were 1.066 g and 0.3033 g, respectively. Statistical analysis showed significant suppression from the SMPDL3B-KO (**Figure 5D**). Moreover, the DNA fragmentation assay in tumor tissue showed that the SMPDL3B-KO tumor had a visible increase in cell apoptosis compared to the SMPDL3B-WT tumor (**Figures 5E,F**). Together, our results indicate that SMPDL3B contributed to AML cells growth both *in vitro* and *in vivo* and might be via inhibiting cell apoptosis.

GO and KEGG Analyses for SMPDL3B Related Genes

For further interpretation of the mechanism of SMPDL3B contributing to human AML cells growth, the gene expression profile datasets were downloaded from the GEO database (GSE13159). The Sanger Box analysis tool was applied to detect the SMPDL3B correlated genes, using adjust p value <0.05 and $|R| \geq 0.6$ as cut-off criteria. After integrated bioinformatical analysis, a total of 66 genes were identified from the database (**Supplementary Table S1**). Furthermore, the gene ontology (GO) functional annotation was performed using the gene ontology resource online. GO analysis results showed that SMPDL3B correlated genes were particularly enriched in biological processes (BP), including neutrophil activation, granulocyte activation, neutrophil degranulation, and neutrophil mediated immunity (**Figure 6A**). For GO cell component (CC), the SMPDL3B correlated genes also were enriched in specific granule, secretory granule, secretory vesicle, and specific granule lumen (**Figure 6B**). In addition,

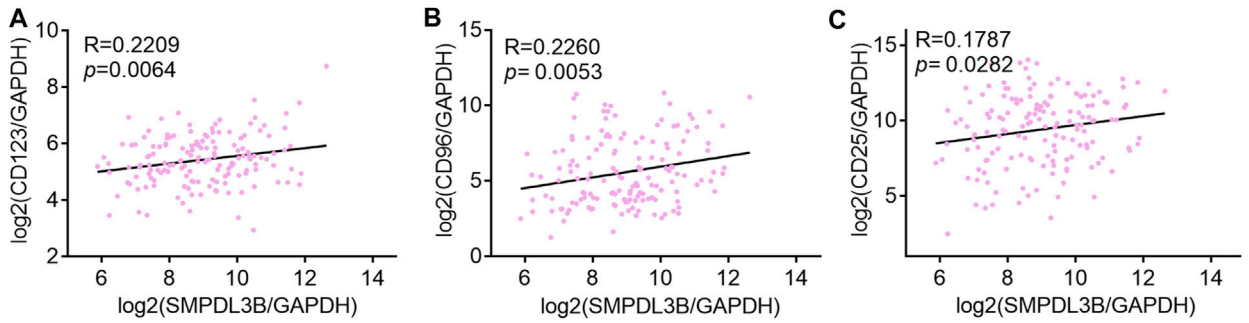


FIGURE 7 | SMPDL3B expression positively correlates with CD123, CD96, and CD25 AML patients. **(A–C)** The log2 transcription level of SMPDL3B, CD123 **(A)**, CD96 **(B)**, and CD25 **(C)** in AML patients was shown after normalized to GAPDH. n = 151, Spearman analysis $p < 0.05$.

molecular function (MF) analysis displayed that the SMPDL3B correlated genes were significantly enriched in leukotriene-B4 20-monooxygenase activity, chitin binding, hemoglobin binding, and transition metal ion binding (**Figure 6C**). In addition, we found that the most significantly enriched KEGG pathways of the SMPDL3B related genes were metabolic pathways and starch and sucrose metabolism (**Figure 6D**). Particularly, the SMPDL3B correlated genes were also enriched in acute myeloid leukemia, hematopoietic cell lineage, and leukocyte *trans*-endothelial migration (**Figure 6D**). Moreover, leukemia stem cells play prominent roles in leukemogenesis and propagation due to their capacities of self-renewal, differentiation, and proliferation. Thus, the established markers of leukemia stem cells such as CD123 and CD96 were determined. Spearman's rank tests showed that the mRNA expression of SMPDL3B was positively correlated with the mRNA expression of CD123, CD96, and CD25 in 151 AML patients, indicating that SMPDL3B may regulate myeloid leukemia development via promoting self-renewal of leukemia stem cells (**Figures 7A–C**).

DISCUSSION

AML, characterized by the low cure rate and high relapse, is the most severe adult acute leukemia. Therefore, exploring novel biomarkers and potential therapeutic targets to improve the diagnosis and therapy for human AML was urgently needed. Herein, we suggested for the first time the prognostic and biological significance of the phosphodiesterase SMPDL3B in AML patients. The transcription level of SMPDL3B was significantly elevated in AML patients compared to healthy control or lymphoid leukemia samples. This finding agreed with the results reported in other human cancers including prostate cancer (Waldbillig et al., 2020) and hepatocellular carcinoma (Liu et al., 2020). Curiously, the overexpression of SMPDL3B was associated with some clinicopathological characteristics, including cytogenetics risk (p = 0.0014) and karyotypes ($p < 0.0001$). Interestingly, cox multivariate analysis results suggested that SMPDL3B was an independent prognostic factor for the overall survival of AML. Accordantly, Frank W. et al. demonstrated that high expression of SMPLD3B was inversely associated with localized prostate cancer prognosis (Waldbillig et al., 2020). Notably, blocked expression of SMPDL3B significantly inhibited the growth of AML cells *in vivo*. These results suggest that SMPDL3B may be a good prognostic indicator and therapeutic target in human cancers. Further investigations are needed to uncover the prognostic significance of SMPDL3B in other types of cancers.

After the discovery of SMPDL3B and identification of it as a glycosylphosphatidylinositol- (GPI-) anchored protein (Masuishi et al., 2013), several seminal findings dominate thinking about the biological function of SMPDL3B. These findings include the following SMPDL3B modulates podocyte injury phenotypes in glomerular disease by shifting suPAR-mediated podocyte injury from a migratory phenotype to an apoptotic phenotype (Yoo et al., 2015); SMPDL3B functions at the interface of membrane biology and innate immunity via negative regulating Toll-like receptor signaling (Heinz et al., 2015); SMPDL3B impairs insulin signaling by interfering with insulin receptor isoforms binding to caveolin-1 in the plasma membrane (Mallela et al., 2019; Mitrofanova et al., 2019); SMPDL3B is an off-target biomarker of rituximab in focal segmental glomerulosclerosis (Fornoni et al., 2011); and the crystal structure revealed that the active site of SMPDL3B was located in a narrow boot-shaped cavity (Gorelik et al., 2016). Recently, it was reported that SMPDL3B promoted HCC cell growth, invasion, and migration via inducing ceramide hydrolysis and ceramide-1-phosphate production (Liu et al., 2020). Moreover, high expression of SMPLD3B is inversely associated with prognosis in localized prostate Cancer (Waldbillig et al., 2020). These studies provide clues as to the potential function of SMPDL3B, however, both the biological function of SMPDL3B in malignant tumors and its main substrates remain largely unclear. In this study, we firstly found that expression of SMPDL3B was significantly upregulated in human AML samples. In addition, blocked SMPDL3B expression inhibited AML cells growth both *in vitro* and *in vivo* via promoting cell apoptosis. For further interpretation of the mechanism of SMPDL3B contributing to apoptosis of human AML cells, SMPDL3B correlated genes were identified. According to the KEGG pathway analysis, the SMPDL3B correlated genes may affect AML cell apoptosis by regulating starch and sucrose metabolism. Consistently, a

previous study showed that SMPDL3B modulates insulin receptor signaling and thereby contributes to the production of ceramide-1-phosphate (Fornoni et al., 2014; Mallela et al., 2019; Mitrofanova et al., 2019). Moreover, macrophage-derived thrombospondin 1 promotes obesity-associated non-alcoholic fatty liver disease through suppressing the expression of SMPDL3B (Gwag et al., 2021). Together, these results indicated that SMPDL3B may regulate starch and sucrose metabolism via modulating insulin signaling in AML cells. Interestingly, SMPDL3B blocks the Toll-like receptor signaling pathway and negatively regulates innate immunity (Heinz et al., 2015), which may partially explain that SMPDL3B correlated genes enriched to IL-17 signaling pathway and AML. Taken together, our results reveal that SMPDL3B promotes that the survival of AML cells may be via regulating glucose metabolism or innate immunity. Further detection of the detailed mechanism of SMPDL3B support development of AML is needed.

However, several limitations to the function and mechanism of SMPDL3B in AML should be noted. Firstly, the prognostic indicator effect of SMPDL3B in AML patients was only examined by using the TCGA data. Future research that enrolls larger AML patients is necessary to further explore the association between SMPDL3B expression and overall survival or disease-free survival in AML patients. Secondly, the mechanism of SMPDL3B supporting AML growth is still elusive. In this study, the mechanism of SMPDL3B function in AML was detected by GO enrichment and KEGG pathway analysis. Although these results are consistent with previous reports, elaborate experimental designs, such as RNA sequencing or protein interaction analysis, are necessary to further explore how SMPDL3B regulates AML development.

Together, in the present study, we showed for the first time that high expression of SMPDL3B was significantly associated with unfavorable outcomes in human AML. Moreover, SMPDL3B might be identified as an independent prognostic biomarker for AML. In addition, deficiency of SMPDL3B significantly inhibited the growth of AML both *in vitro* and *in vivo*. These results indicated that SMPDL3B could serve as a promising indicator and potential therapeutic target for AML patients. Collectively, these findings call for further laboratory experiments and clinical trials to validate SMPDL3B in cancer progression.

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 Binzhou Medical University (2018-389).

AUTHOR CONTRIBUTIONS

QH designed the study. QH and ZY performed experiments (cell culture analysis, flow cytometric analysis, cell cycle, and xenograft experiments) and analyzed the data. QH wrote and edited the manuscript. All authors have read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fmolb.2021.695601/full#supplementary-material

REFERENCES

Abou Daher, A., Francis, M., Azzam, P., Ahmad, A., Eid, A. A., Fornoni, A., et al. (2020). Modulation of Radiation-induced Damage of Human Glomerular Endothelial Cells by SMPDL3B. *FASEB j.* 34, 7915–7926. doi:10.1096/fj.201902179r

Ahmad, A., Mitrofanova, A., Bielawski, J., Yang, Y., Marples, B., Fornoni, A., et al. (2017). Sphingomyelinase-like Phosphodiesterase 3b Mediates Radiation-induced Damage of Renal Podocytes. *FASEB j.* 31, 771–780. doi:10.1096/fj.201600618r

Deng, M., Gui, X., Kim, J., Xie, L., Chen, W., Li, Z., et al. (2018). LILRB4 Signalling in Leukaemia Cells Mediates T Cell Suppression and Tumour Infiltration. *Nature* 562, 605–609. doi:10.1038/s41586-018-0615-z

DiNardo, C. D., Stein, E. M., Pigneux, A., Altman, J. K., Collins, R., Erba, H. P., et al. (2021). Outcomes of Patients with IDH1-Mutant Relapsed or Refractory Acute Myeloid Leukemia Receiving Ivosidenib Who Proceeded to Hematopoietic Stem Cell Transplant. *Leukemia*. doi:10.1038/s41375-021-01229-x

El Omri, H., Taha, R. Y., Elomri, A., Kacem, N., Elsabah, H., Ellahie, A. Y., et al. (2020). Acute Myeloid Leukemia in Qatar (2010-2016): Clinical, Biological, and Prognostic Factors and Treatment Outcomes. *Front. Genet.* 11, 553. doi:10.3389/fgene.2020.00553

Fornoni, A., Sageshima, J., Wei, C., Merscher-Gomez, S., Aguillon-Prada, R., Jauregui, A. N., et al. (2011). Rituximab Targets Podocytes in Recurrent Focal Segmental Glomerulosclerosis. *Sci. translational Med.* 3, 85ra46. doi:10.1126/scitranslmed.3002231

Fornoni, A., Merscher, S., and Kopp, J. B. (2014). Lipid Biology of the Podocyte-New Perspectives Offer New Opportunities. *Nat. Rev. Nephrol.* 10, 379–388. doi:10.1038/nrneph.2014.87

Gorelik, A., Heinz, L. X., Illes, K., Superti-Furga, G., and Nagar, B. (2016). Crystal Structure of the Acid Sphingomyelinase-like Phosphodiesterase SMPDL3B Provides Insights into Determinants of Substrate Specificity. *J. Biol. Chem.* 291, 24054–24064. doi:10.1074/jbc.m116.755801

Gwag, T., Reddy Mooli, R. G., Li, D., Lee, S., Lee, E. Y., and Wang, S. (2021). Macrophage-derived Thrombospondin 1 Promotes Obesity-Associated Non-alcoholic Fatty Liver Disease. *JHEP Rep.* 3, 100193. doi:10.1016/j.jhepr.2020.100193

Heinz, L. X., Baumann, C. L., Köberlin, M. S., Snijder, B., Gawish, R., Shui, G., et al. (2015). The Lipid-Modifying Enzyme SMPDL3B Negatively Regulates Innate Immunity. *Cell Rep.* 11, 1919–1928. doi:10.1016/j.celrep.2015.05.006

Kantarjian, H., Kadia, T., DiNardo, C., Daver, N., Borthakur, G., Jabbour, E., et al. (2021). Acute Myeloid Leukemia: Current Progress and Future Directions. *Blood Cancer J.* 11, 41. doi:10.1038/s41408-021-00425-3

Lai, C., Doucette, K., and Norsworthy, K. (2019). Recent Drug Approvals for Acute Myeloid Leukemia. *J. Hematol. Oncol.* 12, 100. doi:10.1186/s13045-019-0774-x

Liu, B., Xiao, J., Dong, M., Qiu, Z., and Jin, J. (2020). Human Alkaline Ceramidase 2 Promotes the Growth, Invasion, and Migration of Hepatocellular Carcinoma Cells via Sphingomyelin Phosphodiesterase Acid-like 3B. *Cancer Sci.* 111, 2259–2274. doi:10.1111/cas.14453

Ma, X. L., Shen, M. N., Hu, B., Wang, B. L., Yang, W. J., Lv, L. H., et al. (2019). CD73 Promotes Hepatocellular Carcinoma Progression and Metastasis via Activating PI3K/AKT Signaling by Inducing Rap1-Mediated Membrane Localization of P110beta and Predicts Poor Prognosis. *J. Hematol. Oncol.* 12, 37. doi:10.1186/s13045-019-0724-7

Mallela, S. K., Mitrofanova, A., Merscher, S., and Fornoni, A. (2019). Regulation of the Amount of Ceramide-1-Phosphate Synthesized in Differentiated Human Podocytes. *Biochim. Biophys. Acta (Bba) - Mol. Cell Biol. Lipids* 1864, 158517. doi:10.1016/j.bbalip.2019.158517

Martin, G. H., Roy, N., Chakraborty, S., Desrichard, A., Chung, S. S., Woolthuis, C. M., et al. (2019). CD97 Is a Critical Regulator of Acute Myeloid Leukemia Stem Cell Function. *J. Exp. Med.* 216, 2362–2377. doi:10.1084/jem.20190598

Masuishi, Y., Nomura, A., Okayama, A., Kimura, Y., Arakawa, N., and Hirano, H. (2013). Mass Spectrometric Identification of Glycosylphosphatidylinositol-Anchored Peptides. *J. Proteome Res.* 12, 4617–4626. doi:10.1021/pr4004807

Mitrofanova, A., Mallela, S. K., Ducasa, G. M., Yoo, T. H., Rosenfeld-Gur, E., Zelnik, I. D., et al. (2019). SMPDL3b Modulates Insulin Receptor Signaling in Diabetic Kidney Disease. *Nat. Commun.* 10, 2692. doi:10.1038/s41467-019-10584-4

Waldbillig, F., Nitschke, K., Abdelhadi, A., von Hardenberg, J., Nuhn, P., Nientiedt, M., et al. (2020). Phosphodiesterase SMPDL3B Gene Expression as Independent Outcome Prediction Marker in Localized Prostate Cancer. *Int. J. Mol. Sci.* 21, 4373. doi:10.3390/ijms21124373

Xuan, L., Wang, Y., Huang, F., Fan, Z., Xu, Y., Sun, J., et al. (2020). Sorafenib Maintenance in Patients with FLT3-ITD Acute Myeloid Leukaemia Undergoing Allogeneic Haematopoietic Stem-Cell Transplantation: an Open-Label, Multicentre, Randomised Phase 3 Trial. *Lancet Oncol.* 21, 1201–1212. doi:10.1016/s1470-2045(20)30455-1

Yamauchi, T., Masuda, T., Canver, M. C., Seiler, M., Semba, Y., Shboul, M., et al. (2018). Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS. *Cancer Cell* 33, 386–400. doi:10.1016/j.ccell.2018.01.012

Yang, L., Shi, P., Zhao, G., Xu, J., Peng, W., Zhang, J., et al. (2020). Targeting Cancer Stem Cell Pathways for Cancer Therapy. *Signal. Transduction Targeted Therapy* 5, 8. doi:10.1038/s41392-020-0110-5

Yoo, T.-H., Pedigo, C. E., Guzman, J., Correa-Medina, M., Wei, C., Villarreal, R., et al. (2015). Sphingomyelinase-like Phosphodiesterase 3b Expression Levels Determine Podocyte Injury Phenotypes in Glomerular Disease. *Jasn* 26, 133–147. doi:10.1681/asn.2013111213

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 Qu 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.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fmolb.2021.741603

Mass Spectrometry-Based Targeted Serum Monomethylated Ribonucleosides Profiling for Early Detection of Breast Cancer

*Zhihao Fang[1], Yiqiu Hu[1], Jiani Chen[1], Kailun Xu[1], Kailai Wang[1], Shu Zheng[1,2] and Cheng Guo[1,2]**

[1]Cancer Institute (Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education), The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2]Cancer Center, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong, SAR China

Reviewed by:
Mengzhe Guo, Xuzhou Medical University, China
Quanqing Zhang, University of California, Riverside, United States

***Correspondence:**
Cheng Guo
cheng_guo@zju.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *15 July 2021*
Accepted: *16 August 2021*
Published: *26 August 2021*

Citation:
Fang Z, Hu Y, Chen J, Xu K, Wang K, Zheng S and Guo C (2021) Mass Spectrometry-Based Targeted Serum Monomethylated Ribonucleosides Profiling for Early Detection of Breast Cancer. Front. Mol. Biosci. 8:741603. doi: 10.3389/fmolb.2021.741603

RNA methylation plays a significant regulatory role in various of physiological activities and it has gradually become a hotspot of epigenetics in the past decade. 2′-O-methyladenosine (Am), 2′-O-methylguanosine (Gm), 2′-O-methylcytidine (Cm), 2′-O-methyluridine (Um), N^6-methyladenosine (m^6A), N^1-methylguanosine (m^1G), 5-methylcytidine (m^5C), and 5-methyluridine (m^5U) are representative 2′-O-methylation and base-methylation modified epigenetic marks of RNA. Abnormal levels of these ribonucleosides were found to be related to various diseases including cancer. Serum is an important source of biofluid for the discovery of biomarkers, and novel tumor biomarkers can be explored by measuring these ribonucleoside modifications in human serum. Herein, we developed and applied a hydrophilic interaction liquid chromatography tandem mass spectrometry (HILIC-MS/MS) method to determine the content of monomethylated ribonucleosides in human serum. The developed method enabled sensitive and accurate determination of these monomethylated ribonucleosides. By applying this robust method, we demonstrated the presence of Gm and Um in human serum for the first time, and we successfully quantified m^6A, Gm, m^1G, Cm, Um and m^5U in serum samples collected from 61 patients with breast cancer and 69 healthy controls. We discovered that the levels of Gm, m^1G, Cm, Um and m^5U in serum were all significantly decreased in breast cancer patients whereas m^6A was increased. We performed receiver operating characteristic (ROC) curve analysis, and obtained highest area under curve (AUC) value when combining these six monomethylated ribonucleosides together. These results suggest that m^6A, Gm, m^1G, Cm, Um and m^5U might have great potential to be novel biomarkers for detection of breast cancer in the early stage. In addition, this study may stimulate future investigations about the regulatory roles of monomethylated ribonucleosides on the initiation and development of breast cancer.

Keywords: HILIC-MS/MS, RNA modification, monomethylated ribonucleosides, serum, breast cancer

INTRODUCTION

Post-transcriptional modifications of RNA has great research prospects and RNA epigenetics/epitranscriptomics has been proposed (He, 2010). More than 170 different modifications of RNA have been identified in recent year and over half of them are RNA methylation modifications which are associated with the regulation of RNA functions (Jonkhout et al., 2017; Ontiveros et al., 2019). Due to the functional identification of important regulatory proteins such as methyltransferase METTL3/METTL14/WTAP complex (Liu et al., 2014) and demethylase FTO (Jia et al., 2011), RNA methylation modifications have attracted great attention and accumulating evidences have been obtained to confirm RNA methylation modifications as a novel layer of epigenetic alteration. Through their unique regulatory proteins, RNA methylation modifications play critical roles in various cellular functions, such as RNA splicing (Sun et al., 2020), stability (Zhang et al., 2019a; Chen et al., 2019; Shen et al., 2020), degradation (Ni et al., 2019; Yang et al., 2019; Chen et al., 2020b) and translation (Weng et al., 2018; Schumann et al., 2020; Song et al., 2020). Moreover, it has been revealed that RNA methylation is closely associated with the occurrence and development of human cancers.

As we known, the ribonucleoside compositions of RNA contain adenosine, guanosine, cytidine and uridine. The methylation usually occurs at nitrogen or carbon atom in the nucleobase part of the molecule to form m^6A, m^1G, m^5C, m^5U and so on. When the hydrogen on the 2′-hydroxyl (-OH) of the ribose moiety is replaced by a methyl group ($-CH_3$), it will form 2′-O-methylation ribonucleosides (Nm) including Am, Gm, Cm and Um. M^6A is the most predominant modification in mRNA, and the aberrant level of m^6A modification has great connection with the tumorigenesis and development (Ma et al., 2019; Xu et al., 2020; Chen et al., 2021; Fang et al., 2021). In the recent study of our group, it was found that the significantly elevated m^6A in human serum increased the risk of colorectal cancer and gastric cancer (Hu et al., 2021). It was reported that the dynamic level of m^1G in human serum could help early detection of breast cancer (Rashed et al., 2020) and colorectal cancer (Zhu et al., 2015). M^5C is another most abundant RNA modification which has been discovered to be a potential biomarker of various cancers (Guo et al., 2018a; Zhang et al., 2019b; Feng et al., 2020). In addition, a significant downregulation m^5U in human serum can indicate the presence of prostate cancer (Buzatto et al., 2017). For Nm, they have also been revealed to participate in the pathogenesis of various cancers and play crucial roles (Hua et al., 2018; Zhu et al., 2019; Wu et al., 2020). Recently, Li et al. reported that the level of Cm in serum is related to the reduced risk of developing esophageal squamous cell carcinoma (Li et al., 2021). Therefore, these monomethylated ribonucleosides have significant potential to be used as indicators for early detection of cancers.

Breast cancer is the most prevalent tumor with the highest mortality among women worldwide (Sung et al., 2021). Generally, the prognosis and the survival rate of breast cancer are better in the early stages, but poorer in advanced stages even after receiving the surgery and adjuvant treatment (DeSantis et al., 2016). Therefore, it is important for early detection of breast cancer. Currently, ultrasonography and mammography, together with biopsy, are used for routine screening and staging. Limitations of these inspection methods including exposure to radiation, traumatic operation and often producing inaccurate results. Besides, breast tumors must be at least a few millimeters in size to be detected. Serum biomarkers, such as carbohydrate antigen-153 (CA153) and carcinoembryonic antigen (CEA), showed low sensitivity and/or low specificity and responded late to tumor formation and recurrence (Nagrath et al., 2011). From these points of view, it is necessary to hunt for robust biomarkers for early-stage of breast cancer to elongate the survival time and reduce the suffering of patients.

Serum is easy to obtain in the clinic and contains a large number of biomolecules, so it can be used as the body fluid of choice for the discovery of biomarkers. In the past decades, a variety of analytical methods have been utilized for analyzing modified ribonucleosides (Jiang and Ma, 2009; Wei et al., 2010; Beale et al., 2018; Pero-Gascon et al., 2018). Reversed-phase liquid chromatography (RPLC) or hydrophilic interaction liquid chromatography (HILIC) coupled with tandem mass spectrometry is more favored for biomarker discovery due to its great advantages in selectivity, sensitivity, accuracy and high throughput (Zhu et al., 2015; Guo et al., 2018b; Chen et al., 2020a; Guo et al., 2020; Su et al., 2021), compared with other analytical techniques. In our study, a fast, sensitive, simple and reliable HILIC-MS/MS method for qualitative and quantitative detection of monomethylated ribonucleosides in human serum was established. We revealed the presence of Gm and Um in human serum for the first time and quantified m^6A, Gm, m^1G, Cm, Um and m^5U in serum from breast cancer patients and healthy controls. By analyzing these results, we demonstrated the differences of these modifications between breast cancer patients and healthy volunteers, and evaluate the potential of these monomethylated ribonucleosides as biomarkers for early detection of breast cancer.

MATERIALS AND METHODS

Chemicals and Reagents

Chromatographic grade acetonitrile was bought from Merck KGaA (Darmstadt, Germany). Methanol of HPLC grade was purchased from J.T.Baker (Radnor, PA, United States). Formic acid (HCOOH) was bought from Fluka (Muskegon, United States). Ammonium formate, malic acid and 5-methylcytidine (m^5C) were bought from Sigma-Aldrich (St Louis, MO, United States). 2′-O-methyladenosine (Am), N^6-methyladenosine (m^6A), 2′-O-methylguanosine (Gm), 1-methylguanosine (m^1G), 2′-O-methylcytidine (Cm), 2′-O-methyluridine (Um), 5-methyluridine (m^5U) and isotopically labeled standards $[D_3]m^6A$, $[D_3]$Um and $[^{13}C_5]m^5U$ were obtained from Toronto Research Chemicals (Toronto, Canada). $[^{13}C^{15}N_2]$Gm, $[^{13}C^{15}N_2]m^1G$ and $[^{13}C_5]$Cm were synthesized according to literature (Fu et al., 2015). Water was purified by a Milli-Q water purification device (Millipore, Milford, MA, United States).

Instrumentation

Acquity UPLC system (Waters, Milford, MA, United States) achieved by Empower Pro 6.0 software was applied for analysis. A Waters Acquity BEH Amide column (100 mm × 2.1 mm, 1.7 μm) was applied for chromatographic separation. 4000 QTRAP mass spectrometer (AB SCIEX, Foster City, CA, United States) was applied for MS detection. The mass spectrometer was equipped with electrospray ionization (ESI) positive ion mode. Multiple-reaction monitoring (MRM) was chosen to acquire data. Data acquisition and processing were controlled by Analyst 1.6.3 software.

Sample Collection

The Ethics Committee of Medical Research of the Second Affiliated Hospital, Zhejiang University School of Medicine (SAHZU) approved our study. A total of 69 healthy volunteers (mean age of 43.9 ± 11.1 years, range from 30 to 70 years) and 61 patients with breast cancer (mean age of 52.2 ± 11.9 years, range from 29 to 80 years) were recruited from SAHZU. All breast cancer patients had a diagnosis report with a pathological stage of stage I or stage II at SAHZU between June 2020 and December 2020. The exclusion criteria were as follows: 1) Co-suffering other malignant tumors. 2) Having received any type of treatment for tumors. 3) Suffering from metabolic diseases, kidney diseases or liver diseases. 4) Taking any drugs for a long time. All participating volunteers agreed the informed consent in advance. Then, the serum samples were collected in the early morning and reserved at −80°C.

Sample Preparation

At first, 10 μl of isotope-labeled internal standards (IS) mixed solution was added into 100 μl serum samples which were naturally thawed in ice. And then 330 μl pre-refrigerated acetonitrile/methanol of 2:1 (v/v) was added. After vortexed for 60 s, let it stand at −20°C for 2 h and centrifuged at 13,000 rpm, 4°C for 15 min orderly, 352 μl of the supernatant were taken out and then drain under vacuum. Then, 80 μl acetonitrile/water of 9:1 (v/v) was used to redissolve the dried samples. After vortexed for 10 s, ultrasonicated for 15 s and centrifuged at 13,000 rpm for 15 min at 4°C, 70 μl of the supernatant fraction were sucked into the sample bottle for subsequent HILIC-MS/MS detection.

HILIC-MS/MS Analysis

The mobile phase was (A) H_2O containing 10 mM ammonium formate, 0.2% formic acid and 0.06 mM malic acid, and (B) acetonitrile containing 2 mM ammonium formate, 0.2% formic acid and 0.06 mM malic acid. The eight analytes were perfectly separated at a flow rate of 0.4 ml/min by the optimized LC gradient program as follows: 0 min, 94% B; 4 min, 94% B; 6.1 min, 75% B; 6.5 min, 94% B; 8 min, 94% B. The BEH Amide column was set at 40°C and the samples temperature was maintained at 4°C. 5 μl of sample was injected each time and each sample was measured twice. To minimize the interference of the mass spectrometer, a switching valve was used and the eluents from the column were introduced into the ion source during 1.0–6.5 min.

The ion spray voltage was kept at 5.5 kV and the ion source temperature (TEM) was maintained at 550°C. Ion source gas 1 (GS1), ion source gas 2 (GS2) and curtain gas (CUR) were all set at 45 psi. The ion transitions of these eight ribonucleosides and corresponding isotope labeled internal standard (IS) were shown in **Supplementary Table S1**. The optimized MRM parameters of them including declustering potential (DP), entrance potential (EP), collision energy (CE) and collision cell exit potential (CXP) were also listed in **Supplementary Table S1**.

Method Validation

The standard working solutions of m^6A, Gm, m^1G, Cm, Um and m^5U at different concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500 nM), which were mixed with IS solution (final concentration: $[D_3]m^6A$ (5 nM), $[^{13}C^{15}N_2]Gm$ (10 nM), $[^{13}C^{15}N_2]m^1G$ (20 nM), $[^{13}C_5]Cm$ (30 nM), $[D_3]Um$ (20 nM) and $[^{13}C_5]m^5U$ (100 nM)), were made and analyzed. The calibration curves could describe as $y = ax + b$, where y represents the peak area ratio of the analyte to the corresponding IS and x denotes the concentration of the analyte. The limit of detection (LOD) and limit of quantification (LOQ) of each ribonucleoside were obtained by analyzing standard solutions with a signal-to-noise ratio of three and ten, respectively.

For the purpose of evaluating intra-day and inter-day precision, the quality control (QC) samples at three different levels of m^6A (2.5, 5, 50 nM), Gm (5, 10, 50 nM), m^1G (5, 20, 100 nM), Cm (5, 30, 150 nM), Um (5, 20, 100 nM) and m^5U (20, 100, 300 nM) were prepared in triplicate and were measured on the same day and three consecutive days, respectively. The accuracy was described as the ratio of measured value to the theoretical concentration.

For the purpose of evaluating the recovery of extraction, the serum samples were added with three different levels of m^6A (2.5, 5, 30 nM), Gm (2.5, 10, 50 nM), m^1G (5, 20, 60 nM), Cm (6, 30, 150 nM), Um (5, 20, 60 nM) and m^5U (50, 150, 300 nM). After 10 μl of IS solution (same as described above) was added, the serum samples were treated and analyzed as mentioned above. The recovery (R) of each analyte was calculated by (concentration in added serum sample–concentration in original serum sample)/added concentration × 100%.

The matrix effect was estimated by using a slope comparison method. By adding different concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500 nM) standard solution and IS to the serum extracts, the calibration curve was obtained. The ratio value of its slope to the slope of the standard solution calibration curve was matrix effect.

Statistical Analysis

Statistical analyses were achieved through SPSS 24.0 software (IBM, Armonk, NY, United States). The concentration differences of serum monomethylated modifications between healthy volunteers and breast cancer patients were accessed by Two-tailed Student's t-test, where p value less than 0.05 was considered meaningful. The area under the curve (AUC) was acquired by receiver operating characteristic (ROC) curve

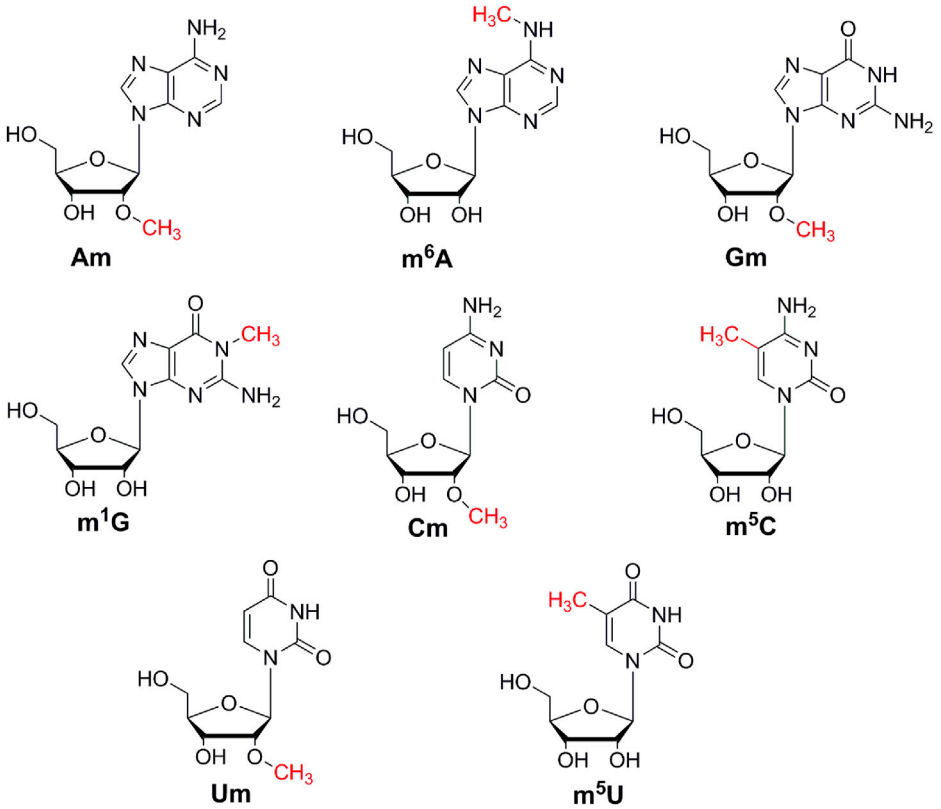


FIGURE 1 | The chemical structures of Am, m^6A, Gm, m^1G, Cm, m^5C, Um, m^5U.

analysis, and the optimal cut-off value of the methylation modification in serum for the diagnosis of breast cancer was determined by the Youden index (Youden index = sensitivity + specificity −1). Receiver operating characteristic (ROC) analysis was performed to evaluate the ability of monomethylated modifications to distinguish cancer patients from healthy controls.

RESULTS AND DISCUSSION

Optimization of Chromatographic Conditions and Mass Spectrometry Parameters

The optimization of chromatographic conditions is mainly achieved by optimizing the type of chromatographic column and the composition of the mobile phase. In order to acquire symmetrical peak shape and great separation effect in a short time, it is very important to choose a column with high separation efficiency. The chemical structures of these monomethylated ribonucleosides were illustrated in **Figure 1**. In our previous study, we found malic acid could enhance the detection of methylated nucleosides in HILIC-MS/MS (Guo et al., 2018a). Therefore, a hydrophilic interaction column of BEH Amide (100 mm × 2.1 mm, 1.7 μm, Waters) was selected for analysis, and malic acid was added into the mobile phase. As showed in

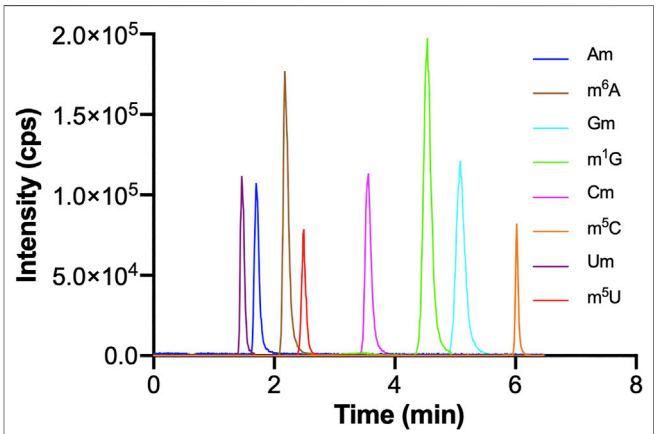


FIGURE 2 | The MRM chromatograms of Am, m^6A, Gm, m^1G, Cm, m^5C, Um, m^5U standards. The concentrations of Um and m^5U were 1,000 nM, and the concentrations of other nucleosides standards were 100 nM. The injection volume was 5.0 μl.

Figure 2, it could acquire satisfactory separation for these eight ribonucleoside modifications. Besides, the analysis could be accomplished less than 6.5 min. It meant that this analytical method was quick, high throughput and fit for large clinical practice.

TABLE 1 | Linear equations and matrix effect values of m^6A, Gm, m^1G, Cm, Um and m^5U in HILIC-MS/MS analysis.

	Linear equation	R^2 value	Linear range (nM)	Matrix effect (%)
m^6A	y = 0.3492x + 0.1053	0.9995	1–100	104.7
Gm	y = 0.1272x + 0.0308	1.0000	1–100	96.7
m^1G	y = 0.0432x + 0.0061	1.0000	1–100	94.9
Cm	y = 0.0416x + 0.0137	0.9999	1–250	94.7
Um	y = 0.1100x + 0.0312	0.9998	1–500	99.0
m^5U	y = 0.0173x - 0.0088	0.9999	1–500	99.4

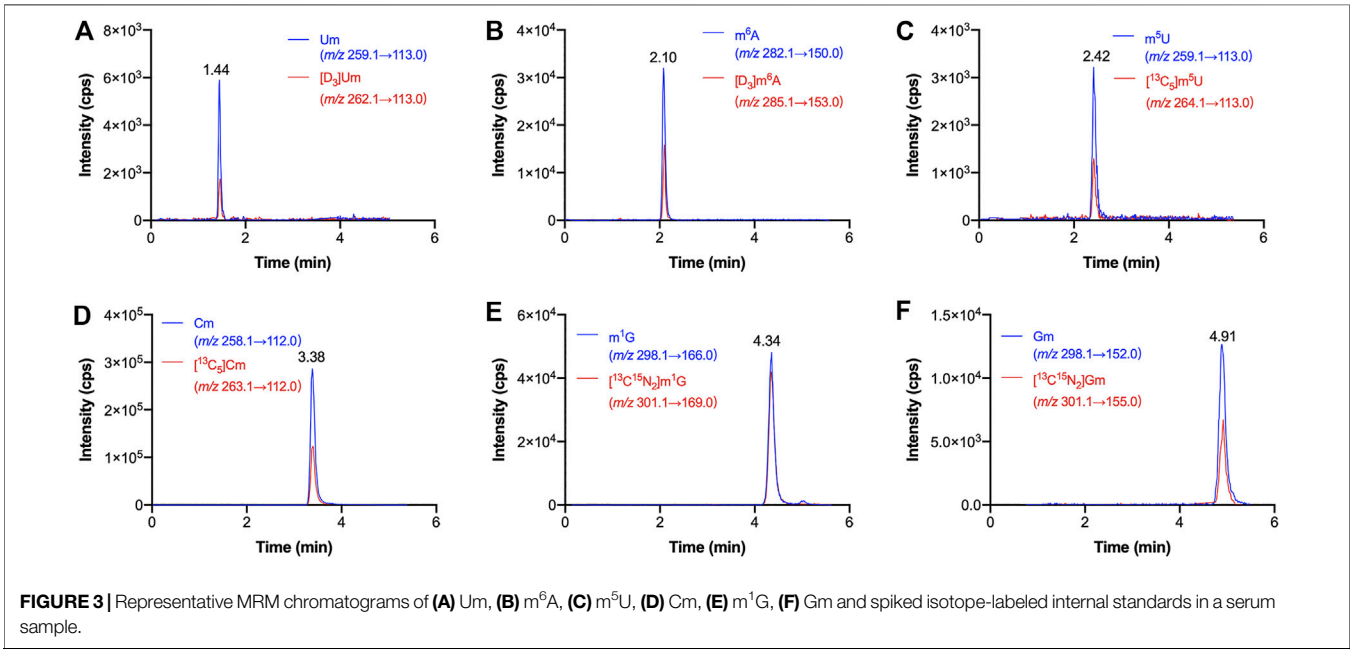


FIGURE 3 | Representative MRM chromatograms of **(A)** Um, **(B)** m^6A, **(C)** m^5U, **(D)** Cm, **(E)** m^1G, **(F)** Gm and spiked isotope-labeled internal standards in a serum sample.

For the purpose of optimizing the MRM parameters, the mass spectrometer analyzed the standard solution injected by the peristaltic pump. Abundant $[M + H]^+$ ions were observed in full scan ESI-MS. Then, the collision induced dissociation (CID) experiment was used to calculate the ion transitions of all analytes. In MS/MS, the ribose group can be easily eliminated due to the cleavage of the C-N bond. Taking Am and $[^{13}C_5]$Am as examples, abundant $[M + H]^+$ ions at *m/z* 282.1 and 287.1 were observed for Am and $[^{13}C_5]$Am, respectively. The $[M + H]^+$ ion of Am lost 146 and 151 Da was lost from $[M + H]^+$ ion of $[^{13}C_5]$Am. Therefore, ion transition *m/z* 282.1→136.0 and *m/z* 287.1→136.0 was used for the quantification of Am and $[^{13}C_5]$Am, respectively. The ion transitions of other ribonucleosides and corresponding IS were shown in **Supplementary Table S1**. In addition, the optimized parameter values of DP, EP, CE and CXP were also listed. Under these optimized conditions, The LODs and LOQs of these monomethylated ribonucleosides can reach sub femtomole level (**Supplementary Table S2**).

Validation of Analytical Method

According to the aforementioned method, the prepared calibration curve showed excellent linearities (R^2 < 0.999) in appropriate analytical ranges, and equations were showed in **Table 1**. The slope ratio values for these eight monomethylated ribonucleosides ranged from 94.7 to 104.7% (**Table 1**), which indicated the interference of matrix in this study was minimal.

As showed in **Supplementary Table S3**, the intra- and inter-day accuracy assays were in the range of 92.20–112.96% and 92.29–112.76%, respectively, and the precision of intra- and inter-day were both within 8.6%. These data indicated that sufficient reproducibility and accuracy were obtained. As showed in **Supplementary Table S4**, the recoveries ranged from 98.04 to 114.01% (RSD <10%), indicating an excellent recovery rate.

In a word, all these results mentioned above revealed that the established HILIC-MS/MS method could meet quantitative requirement of m^6A, Gm, m^1G, Cm, Um and m^5U in human serum samples, and it was quick, accurate, sensitive, reproducible and reliable.

Identification of Monomethylated Ribonucleoside Modifications in Human Serum

By using this HILIC-MS/MS method, we detected these modified ribonucleosides in serum samples from 61 patients with breast cancer and 69 healthy volunteers. The results showed that m^6A, Gm, m^1G, Cm, Um and m^5U were detected in all the serum

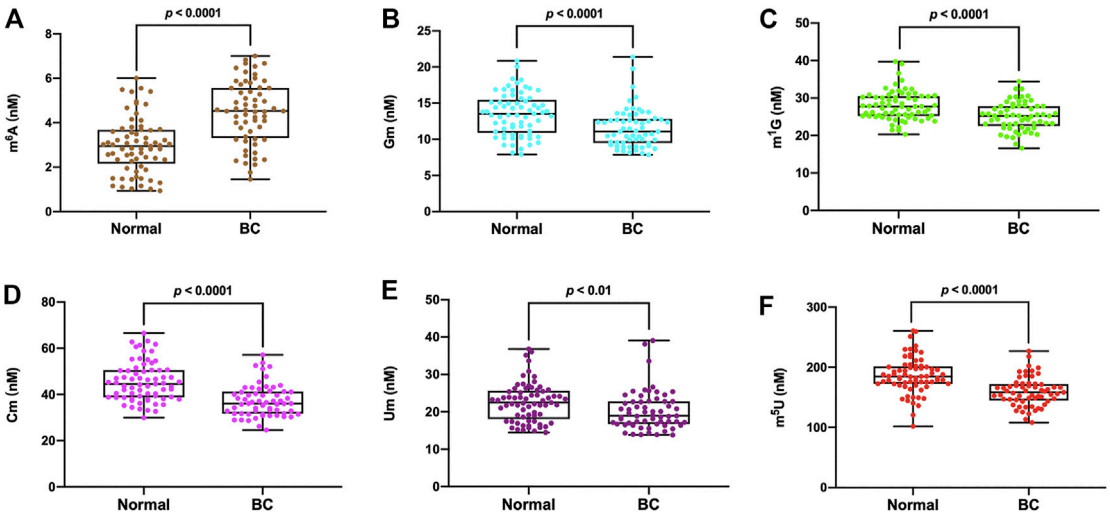


FIGURE 4 | The measured concentrations of **(A)** m^6A, **(B)** Gm, **(C)** m^1G, **(D)** Cm, **(E)** Um, **(F)** m^5U in serum samples and statistical analysis.

samples, whereas Am and m^5C could not be monitored due to their extremely low levels. As demonstrated in **Figure 3**, the retention time of Um, m^6A, m^5U, Cm, m^1G and Gm were 1.44, 2.10, 2.42, 3.38, 4.34 and 4.91 min, respectively. Of note, the retention time of these compounds were consistent with their corresponding isotope-labeled internal standard. The same tandem mass spectrometry behaviors and chromatographic retention time of each modified ribonucleoside in serum as those of the isotope-labeled IS confirmed the existence of m^6A, Gm, m^1G, Cm, Um and m^5U in human serum indubitably. It is worth nothing that the presence of Gm and Um in human serum was revealed for the first time, as far as we known.

Quantification of Monomethylated Ribonucleoside Modifications in Human Serum

The detailed concentrations of m^6A, Gm, m^1G, Cm, Um and m^5U in all the serum samples were presented in **Supplementary Table S5**. The measured concentrations (nM) in serum samples of m^6A, Gm, m^1G, Cm, Um and m^5U from healthy volunteers were in the range of 0.93–6.01, 7.91–20.85, 20.29–39.74, 29.97–66.58, 14.47–36.83 and 101.97–260.34 nM respectively, and the average concentrations were 2.97, 13.47, 27.97, 22.60 and 186.43 nM, respectively (n = 69). In the serum of breast cancer patients, the concentration of m^6A, Gm, m^1G, Cm, Um and m^5U were in the range of 1.45–7.01, 7.84–21.39, 16.59–34.41, 24.61–57.14, 13.83–39.07 and 108.20–226.84 nM respectively, and the average concentrations were 4.48, 11.55, 25.15, 37.31, 20.27 and 159.40 nM, respectively (n = 61). As illustrated in **Figure 4**, it was obvious that the levels of Gm, m^1G, Cm, Um and m^5U in serum were intensely decreased in patients with breast cancer compared to healthy controls ($p < 0.0001$ for Gm, $p < 0.0001$ for m^1G, $p < 0.0001$ for Cm, $p < 0.01$ for Um, $p < 0.0001$ for m^5U), but the concentration of m^6A in breast cancer patients was much higher than that in healthy volunteers ($p < 0.0001$).

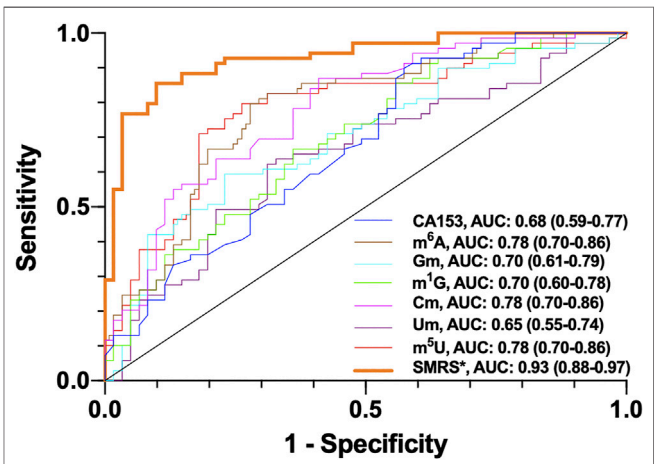


FIGURE 5 | ROC analysis for m^6A, Gm, m^1G, Cm, Um, m^5U, SMRS and CA153 in serum samples of breast cancer patients. *SMRS: serum monomethylated ribonucleosides score.

To verify the significance of m^6A, Gm, m^1G, Cm, Um and m^5U as potential breast cancer biomarkers, receiver operating characteristic (ROC) curve was plotted. Based on the Youden index, the optimal cut-off values of m^6A, Gm, m^1G, Cm, Um and m^5U were 3.76, 12.99, 25.92, 37.31,21.62 and 169.58 nM, which indicates that if the detection values break through these thresholds, it increases the risk of breast cancer. Trapezoidal rule was used to calculate area under the curve (AUC). As demonstrated in **Figure 5**, the AUC values were 0.78, 0.70, 0.70, 0.78, 0.65, 0.78 and 0.68 for m^6A, Gm, m^1G, Cm, Um, m^5U and CA153, respectively. Most of the AUC values of ribonucleosides were higher than that of CA153, implying there might be better correlation between the levels of m^6A,

Gm, m^1G, Cm and m^5U in serum and the incidence of breast cancer, compared with CA153. Moreover, when all these serum monomethylated ribonucleosides were combined, the AUC could reach 0.93, which suggested the sensitivity and specificity for breast cancer diagnosis were dramatically increased. These results indicated that lower levels of Gm, m^1G, Cm, Um and m^5U and higher levels of m^6A could be regarded as potential diagnostic indicators for the screening of breast cancer.

CONCLUSION

In our study, a fast, robust, sensitive and trustable HILIC-MS/MS method was established for analysis of monomethylated ribonucleosides. A total of 130 serum samples from two groups containing breast cancer patients and healthy control were analyzed. Six modified ribonucleosides including m^6A, Gm, m^1G, Cm, Um and m^5U were identified and quantified. Of note, to the best of our knowledge, this is the first time that Gm and Um were detected in human serum. Besides, we elucidated the differences in the contents of these monomethylated ribonucleosides between these two groups and our data, to some extent, indicated that these six modified ribonucleosides might play as potential indicators in the early detection of breast cancer.

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 Institutional Review Board of Medical Research, The Second Affiliated Hospital, Zhejiang University School of Medicine (SAHZU). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CG designed the study; ZF and YH performed the experiments; ZF, JC and KW collected the serum samples; CG and ZF analyzed and interpreted the data; CG and ZF wrote the manuscript; KX and SZ edited the manuscript. All authors commented and approved the final manuscript.

FUNDING

This study was supported by the Natural Science Foundation of Zhejiang Province (LY19B050007), Key R&D Program of Zhejiang Province (2021C03125), National Key R&D Program of China (2016YFC1302803), and National Natural Science Foundation of China (21402172).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.741603/full#supplementary-material

REFERENCES

Beale, D. J., Pinu, F. R., Kouremenos, K. A., Poojary, M. M., Narayana, V. K., Boughton, B. A., et al. (2018). Review of Recent Developments in GC-MS Approaches to Metabolomics-Based Research. *Metabolomics* 14 (11), 152. doi:10.1007/s11306-018-1449-2

Buzatto, A. Z., de Oliveira Silva, M., Poppi, R. J., and Simionato, A. V. C. (2017). Assessment of Nucleosides as Putative Tumor Biomarkers in Prostate Cancer Screening by CE-UV. *Anal. Bioanal. Chem.* 409 (13), 3289–3297. doi:10.1007/s00216-017-0297-7

Chen, H., Gao, S., Liu, W., Wong, C.-C., Wu, J., Wu, J., et al. (2021). RNA N6-Methyladenosine Methyltransferase METTL3 Facilitates Colorectal Cancer by Activating the m6A-GLUT1-mTORC1 Axis and Is a Therapeutic Target. *Gastroenterology* 160 (4), 1284–1300. e1216. doi:10.1053/j.gastro.2020.11.013

Chen, Q., Hu, Y., Fang, Z., Ye, M., Li, J., Zhang, S., et al. (2020a). Elevated Levels of Oxidative Nucleic Acid Modification Markers in Urine from Gastric Cancer Patients: Quantitative Analysis by Ultra Performance Liquid Chromatography-Tandem Mass Spectrometry. *Front. Chem.* 8, 606495. doi:10.3389/fchem.2020.606495

Chen, X., Li, A., Sun, B.-F., Yang, Y., Han, Y.-N., Yuan, X., et al. (2019). 5-methylcytosine Promotes Pathogenesis of Bladder Cancer through Stabilizing mRNAs. *Nat. Cel Biol.* 21 (8), 978–990. doi:10.1038/s41556-019-0361-y

Chen, X., Xu, M., Xu, X., Zeng, K., Liu, X., Pan, B., et al. (2020b). METTL14-mediated N6-Methyladenosine Modification of SOX4 mRNA Inhibits Tumor Metastasis in Colorectal Cancer. *Mol. Cancer* 19 (1), 106. doi:10.1186/s12943-020-01220-7

DeSantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A., and Jemal, A. (2016). Breast Cancer Statistics, 2015: Convergence of Incidence Rates between Black and white Women. *CA: A Cancer J. Clinicians* 66 (1), 31–42. doi:10.3322/caac.21320

Fang, Z., Hu, Y., Hu, J., Huang, Y., Zheng, S., and Guo, C. (2021). The Crucial Roles of N6-Methyladenosine (m6A) Modification in the Carcinogenesis and Progression of Colorectal Cancer. *Cell Biosci.* 11 (1), 72. doi:10.1186/s13578-021-00583-8

Feng, Y., Ma, C.-J., Ding, J.-H., Qi, C.-B., Xu, X.-J., Yuan, B.-F., et al. (2020). Chemical Labeling - Assisted Mass Spectrometry Analysis for Sensitive Detection of Cytidine Dual Modifications in RNA of Mammals. *Analytica Chim. Acta* 1098, 56–65. doi:10.1016/j.aca.2019.11.016

Fu, L., Amato, N. J., Wang, P., McGowan, S. J., Niedernhofer, L. J., and Wang, Y. (2015). Simultaneous Quantification of Methylated Cytidine and Adenosine in Cellular and Tissue RNA by Nano-Flow Liquid Chromatography-Tandem Mass Spectrometry Coupled with the Stable Isotope-Dilution Method. *Anal. Chem.* 87 (15), 7653–7659. doi:10.1021/acs.analchem.5b00951

Guo, C., Chen, Q., Chen, J., Yu, J., Hu, Y., Zhang, S., et al. (2020). 8-Hydroxyguanosine as a Possible RNA Oxidative Modification Marker in Urine from Colorectal Cancer Patients: Evaluation by Ultra Performance Liquid Chromatography-Tandem Mass Spectrometry. *J. Chromatogr. B* 1136, 121931. doi:10.1016/j.jchromb.2019.121931

Guo, C., Xie, C., Chen, Q., Cao, X., Guo, M., Zheng, S., et al. (2018a). A Novel Malic Acid-Enhanced Method for the Analysis of 5-Methyl-2′-Deoxycytidine, 5-Hydroxymethyl-2′-Deoxycytidine, 5-methylcytidine and 5-hydroxymethylcytidine in Human Urine Using Hydrophilic Interaction Liquid Chromatography-Tandem Mass Spectrometry. *Analytica Chim. Acta* 1034, 110–118. doi:10.1016/j.aca.2018.06.081

Guo, C., Xie, C., Ding, P., Qin, G., Mo, W., Cao, X., et al. (2018b). Quantification of Glycocholic Acid in Human Serum by Stable Isotope Dilution Ultra Performance Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry. *J. Chromatogr. B* 1072, 315–319. doi:10.1016/j.jchromb.2017.11.037

He, C. (2010). Grand challenge Commentary: RNA Epigenetics? *Nat. Chem. Biol.* 6 (12), 863–865. doi:10.1038/nchembio.482

Hu, Y., Fang, Z., Mu, J., Huang, Y., Zheng, S., Yuan, Y., et al. (2021). Quantitative Analysis of Methylated Adenosine Modifications Revealed Increased Levels of N6-Methyladenosine (m6A) and N6,2′-O-Dimethyladenosine (m6Am) in Serum from Colorectal Cancer and Gastric Cancer Patients. *Front. Cel Dev. Biol.* 9, 694673. doi:10.3389/fcell.2021.694673

Hua, W., Zhao, Y., Jin, X., Yu, D., He, J., Xie, D., et al. (2018). METTL3 Promotes Ovarian Carcinoma Growth and Invasion through the Regulation of AXL Translation and Epithelial to Mesenchymal Transition. *Gynecol. Oncol.* 151 (2), 356–365. doi:10.1016/j.ygyno.2018.09.015

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in Nuclear RNA Is a Major Substrate of the Obesity-Associated FTO. *Nat. Chem. Biol.* 7 (12), 885–887. doi:10.1038/nchembio.687

Jiang, Y., and Ma, Y. (2009). A Fast Capillary Electrophoresis Method for Separation and Quantification of Modified Nucleosides in Urinary Samples. *Anal. Chem.* 81 (15), 6474–6480. doi:10.1021/ac901216n

Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., and Novoa, E. M. (2017). The RNA Modification Landscape in Human Disease. *Rna* 23 (12), 1754–1769. doi:10.1261/rna.063503.117

Li, X., Zhao, L., Wei, M., Lv, J., Sun, Y., Shen, X., et al. (2021). Serum Metabolomics Analysis for the Progression of Esophageal Squamous Cell Carcinoma. *J. Cancer* 12 (11), 3190–3197. doi:10.7150/jca.54429

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation. *Nat. Chem. Biol.* 10 (2), 93–95. doi:10.1038/nchembio.1432

Ma, F., Liu, X., Zhou, S., Li, W., Liu, C., Chadwick, M., et al. (2019). Long Non-coding RNA FGF13-AS1 Inhibits Glycolysis and Stemness Properties of Breast Cancer Cells through FGF13-AS1/IGF2BPs/Myc Feedback Loop. *Cancer Lett.* 450, 63–75. doi:10.1016/j.canlet.2019.02.008

Nagrath, D., Caneba, C., Karedath, T., and Bellance, N. (2011). Metabolomics for Mitochondrial and Cancer Studies. *Biochim. Biophys. Acta (Bba) - Bioenerg.* 1807 (6), 650–663. doi:10.1016/j.bbabio.2011.03.006

Ni, W., Yao, S., Zhou, Y., Liu, Y., Huang, P., Zhou, A., et al. (2019). Long Noncoding RNA GAS5 Inhibits Progression of Colorectal Cancer by Interacting with and Triggering YAP Phosphorylation and Degradation and Is Negatively Regulated by the m6A Reader YTHDF3. *Mol. Cancer* 18 (1), 143. doi:10.1186/s12943-019-1079-y

Ontiveros, R. J., Stoute, J., and Liu, K. F. (2019). The Chemical Diversity of RNA Modifications. *Biochem. J.* 476 (8), 1227–1245. doi:10.1042/bcj20180445

Pero-Gascon, R., Sanz-Nebot, V., Berezovski, M. V., and Benavente, F. (2018). Analysis of Circulating microRNAs and Their Post-Transcriptional Modifications in Cancer Serum by On-Line Solid-phase Extraction-Capillary Electrophoresis-Mass Spectrometry. *Anal. Chem.* 90 (11), 6618–6625. doi:10.1021/acs.analchem.8b00405

Rashed, R., Darwish, H., Omran, M., Belal, A., and Zahran, F. (2020). A Novel Serum Metabolome Score for Breast Cancer Diagnosis. *Br. J. Biomed. Sci.* 77 (4), 196–201. doi:10.1080/09674845.2020.1784568

Schumann, U., Zhang, H.-N., Sibbritt, T., Pan, A., Horvath, A., Gross, S., et al. (2020). Multiple Links between 5-methylcytosine Content of mRNA and Translation. *BMC Biol.* 18 (1), 40. doi:10.1186/s12915-020-00769-5

Shen, C., Xuan, B., Yan, T., Ma, Y., Xu, P., Tian, X., et al. (2020). m6A-dependent Glycolysis Enhances Colorectal Cancer Progression. *Mol. Cancer* 19 (1), 72. doi:10.1186/s12943-020-01190-w

Song, P., Feng, L., Li, J., Dai, D., Zhu, L., Wang, C., et al. (2020). β-Catenin Represses miR455-3p to Stimulate m6A Modification of HSF1 mRNA and Promote its Translation in Colorectal Cancer. *Mol. Cancer* 19 (1), 129. doi:10.1186/s12943-020-01244-z

Su, X., Li, X., Wang, H., and Cai, Z. (2021). Simultaneous Determination of Methionine Cycle Metabolites, Urea Cycle Intermediates and Polyamines in Serum, Urine and Intestinal Tissue by Using UHPLC-MS/MS. *Talanta* 224, 121868. doi:10.1016/j.talanta.2020.121868

Sun, L., Wan, A., Zhou, Z., Chen, D., Liang, H., Liu, C., et al. (2020). RNA-binding Protein RALY Reprogrammes Mitochondrial Metabolism via Mediating miRNA Processing in Colorectal Cancer. *Gut* 70, 1698–1712. doi:10.1136/gutjnl-2020-320652

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

Wei, R., Li, G., and Seymour, A. B. (2010). High-throughput and Multiplexed LC/MS/MRM Method for Targeted Metabolomics. *Anal. Chem.* 82 (13), 5527–5533. doi:10.1021/ac100331b

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B. S., Dong, L., et al. (2018). METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m6A Modification. *Cell Stem Cell* 22 (2), 191–205. e199. doi:10.1016/j.stem.2017.11.016

Wu, H., Qin, W., Lu, S., Wang, X., Zhang, J., Sun, T., et al. (2020). Long Noncoding RNA ZFAS1 Promoting Small Nucleolar RNA-Mediated 2′-O-Methylation via NOP58 Recruitment in Colorectal Cancer. *Mol. Cancer* 19 (1), 95. doi:10.1186/s12943-020-01201-w

Xu, Y., Ye, S., Zhang, N., Zheng, S., Liu, H., Zhou, K., et al. (2020). The FTO/miR-181b-3p/ARL5B Signaling Pathway Regulates Cell Migration and Invasion in Breast Cancer. *Cancer Commun.* 40 (10), 484–500. doi:10.1002/cac2.12075

Yang, S., Wei, J., Cui, Y.-H., Park, G., Shah, P., Deng, Y., et al. (2019). m6A mRNA Demethylase FTO Regulates Melanoma Tumorigenicity and Response to Anti-PD-1 Blockade. *Nat. Commun.* 10 (1), 2782. doi:10.1038/s41467-019-10669-0

Zhang, Y.-F., Qi, C.-B., Yuan, B.-F., and Feng, Y.-Q. (2019b). Determination of Cytidine Modifications in Human Urine by Liquid Chromatography - Mass Spectrometry Analysis. *Analytica Chim. Acta* 1081, 103–111. doi:10.1016/j.aca.2019.07.002

Zhang, Y., Kang, M., Zhang, B., Meng, F., Song, J., Kaneko, H., et al. (2019a). m6A Modification-Mediated CBX8 Induction Regulates Stemness and Chemosensitivity of colon Cancer via Upregulation of LGR5. *Mol. Cancer* 18 (1), 185. doi:10.1186/s12943-019-1116-x

Zhu, H., Gan, X., Jiang, X., Diao, S., Wu, H., and Hu, J. (2019). ALKBH5 Inhibited Autophagy of Epithelial Ovarian Cancer through miR-7 and BCL-2. *J. Exp. Clin. Cancer Res.* 38 (1), 163. doi:10.1186/s13046-019-1159-2

Zhu, J., Djukovic, D., Deng, L., Gu, H., Himmati, F., Abu Zaid, M., et al. (2015). Targeted Serum Metabolite Profiling and Sequential Metabolite Ratio Analysis for Colorectal Cancer Progression Monitoring. *Anal. Bioanal. Chem.* 407 (26), 7857–7863. doi:10.1007/s00216-015-8984-8

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 Fang, Hu, Chen, Xu, Wang, Zheng 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fmolb.2021.707295

CDK13-Mediated Cell Cycle Disorder Promotes Tumorigenesis of High HMGA2 Expression Gastric Cancer

*Zhouying Wu[1], Min Wang[1], Feng Li[1], Feng Wang[2], Jianchao Jia[1], Zongqi Feng[1], Xue Huo[1], Jie Yang[1], Wen Jin[1], Rina Sa[1], Wenming Gao[3] and Lan Yu[1,4]**

[1]Clinical Medical Research Center/Inner Mongolia Key Laboratory of Gene Regulation of the Metabolic Diseases, Inner Mongolia People's Hospital, Hohhot, China, [2]Department of Pathology, Inner Mongolia People's Hospital, Hohhot, China, [3]Departments of Cardiology, Hohhot First Hospital, Hohhot, China, [4]Department of Endocrine and Metabolic Diseases, Inner Mongolia People's Hospital, Hohhot, China

The inhibitor of CDK4/6 has been clinically used for treating certain types of cancer which are characterized by G0/G1 acceleration induced by the CDK4/6-RB1 pathway. On the contrary, the cell cycle–related molecules are abnormal in over 50% of the patients with gastric cancer (GC), but the efficiency of inhibiting CDK4/6 does not work well as it is expected. In our study, we found HMGA2 promotes GC through accelerating the S–G2/M phase transition, instead of G0/G1. We also found CDK13 is the direct target gene of HMGA2. Importantly, we analyzed 200 pairs of GC and the adjacent tissue and proved the positive relation between HMGA2 and CDK13; moreover, high expression of both genes predicts a poorer prognosis than the expression of single gene does. We explored the effect of the novel CDK12/13 inhibiting agent, SR-4835, on high *HMGA2* expression GC and found inhibition of both genes jointly could reach a satisfied result. Therefore, we suggest that inhibition of CDK13 and HMGA2 simultaneously could be an effective strategy for high *HMGA2* expression GC. To detect the expression of both genes simultaneously and individually could be of benefit to predict prognosis for GC.

Keywords: gastric cancer, cell cycle, CDK13, SR-4835, HMGA2

OPEN ACCESS

Edited by:
William C Cho,
Queen Elizabeth Hospital, China

Reviewed by:
Sudip Banerjee,
Morehouse School of Medicine, United States
Rishi Kumar Jaiswal,
Loyola University Chicago, United States

****Correspondence:***
Lan Yu
lcyxyjzx13@163.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *07 June 2021*
Accepted: *26 July 2021*
Published: *26 August 2021*

Citation:
Wu Z, Wang M, Li F, Wang F, Jia J, Feng Z, Huo X, Yang J, Jin W, Sa R, Gao W and Yu L (2021) CDK13-Mediated Cell Cycle Disorder Promotes Tumorigenesis of High HMGA2 Expression Gastric Cancer. Front. Mol. Biosci. 8:707295. doi: 10.3389/fmolb.2021.707295

INTRODUCTION

Precisely regulated cell cycle maintains the normal cellular life, while uncontrolled cell cycle is one of the main features of all types of cancer. A series of complicated regulators, including cyclin-dependent kinases (CDKs), are involved in every detail of the cell phase transition of the cell cycle (Otto and Sicinski, 2017). Such CDKs as CDK4/6 are clearly researched (Harbour et al., 1999; Lazarov et al., 2002; Malumbres et al., 2004; Zhang et al., 2018a). Proved by the United States Food and Drug Administration, the CDK4/6 inhibitors have been commercially available and widely applied in some types of metastatic breast cancer, bringing the new landscape of treatment (Hortobagyi et al., 2016; Sledge et al., 2017; Slamon et al., 2018; Tripathy et al., 2018; Spring et al., 2020). Subsequently, the researchers have shed more light on not only the CDK4/6 inhibitor in many other types of cancer but also other members of CDKs (Min et al., 2018; Ding et al., 2020; Álvarez-Fernández and Malumbres, 2020).

CDK13, a transcription-associated CDK, was identified in 2001 and is known as the cholinesterase-related cell division controller as well as the regulator of the gene expression (Lapidot-Lifson et al., 1992; Marqués et al., 2000; Ko et al., 2001). CDK13 had been paid close

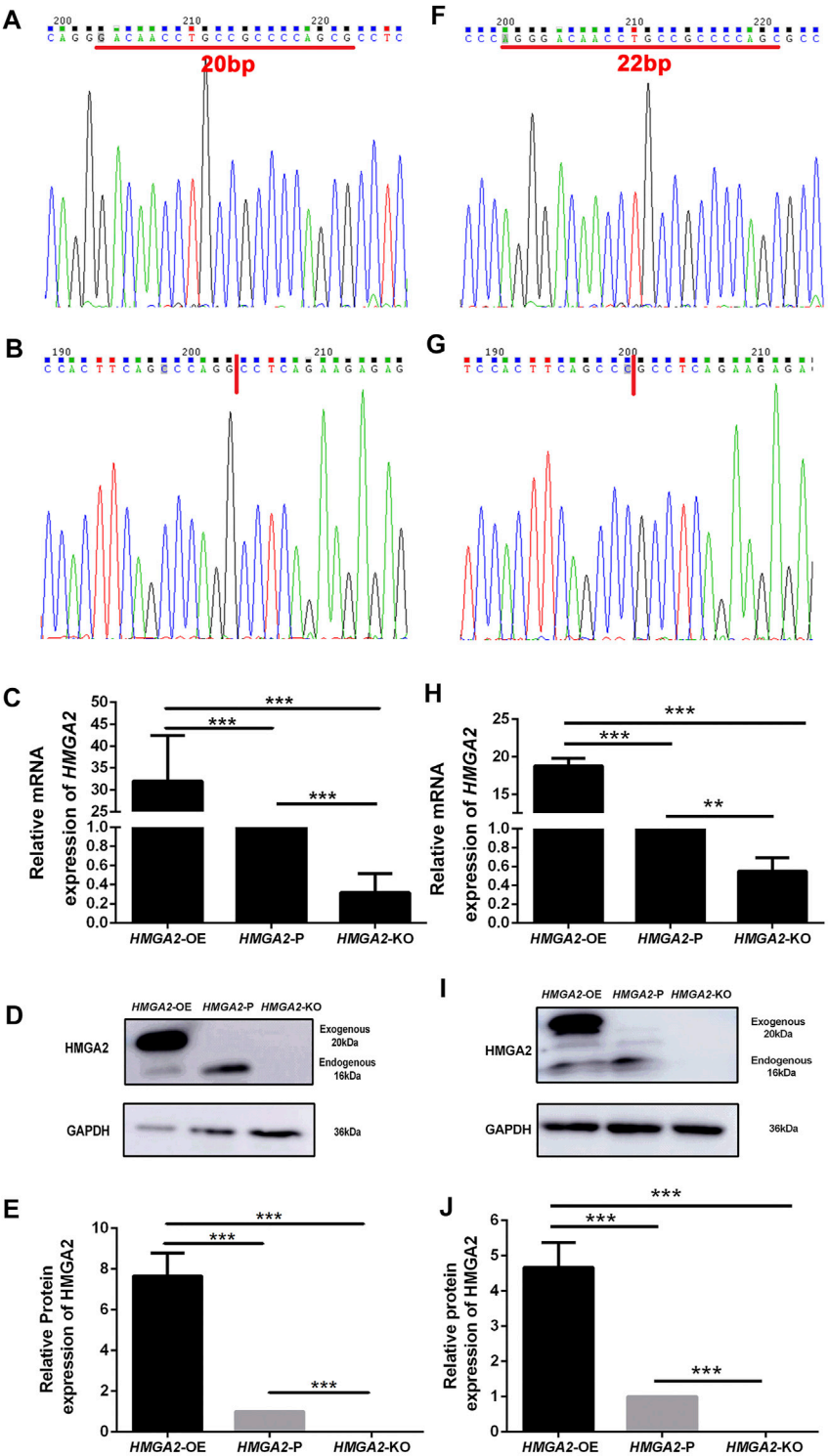


FIGURE 1 | HMGA2 was knocked out in a biallelic manner and overexpressed successfully in two human gastric cancer cell lines. Sanger sequencing results of *HMGA2* in *HMGA2*-P MKN-45 **(A)**, *HMGA2*-KO MKN-45 **(B)**, *HMGA2*-P MGC-803 **(F)**, and *HMGA2*-KO MGC-803 **(G)**. Red lines in **(A)** and **(F)** indicate the deleted bases of *HMGA2* in *HMGA2*-KO MKN-45/MGC-803, respectively. The expressions of *HMGA2* in the mRNA level in *HMGA2*-edited MKN-45 cells **(C)** and MGC-803 cells **(H)** were detected. The expressions of HMGA2 in the protein level in *HMGA2*-edited MKN-45 cells **(D,E)** and MGC-803 cells **(I,J)** are shown. ***$p < 0.001$, Student's t-test. The error bars represent SD.

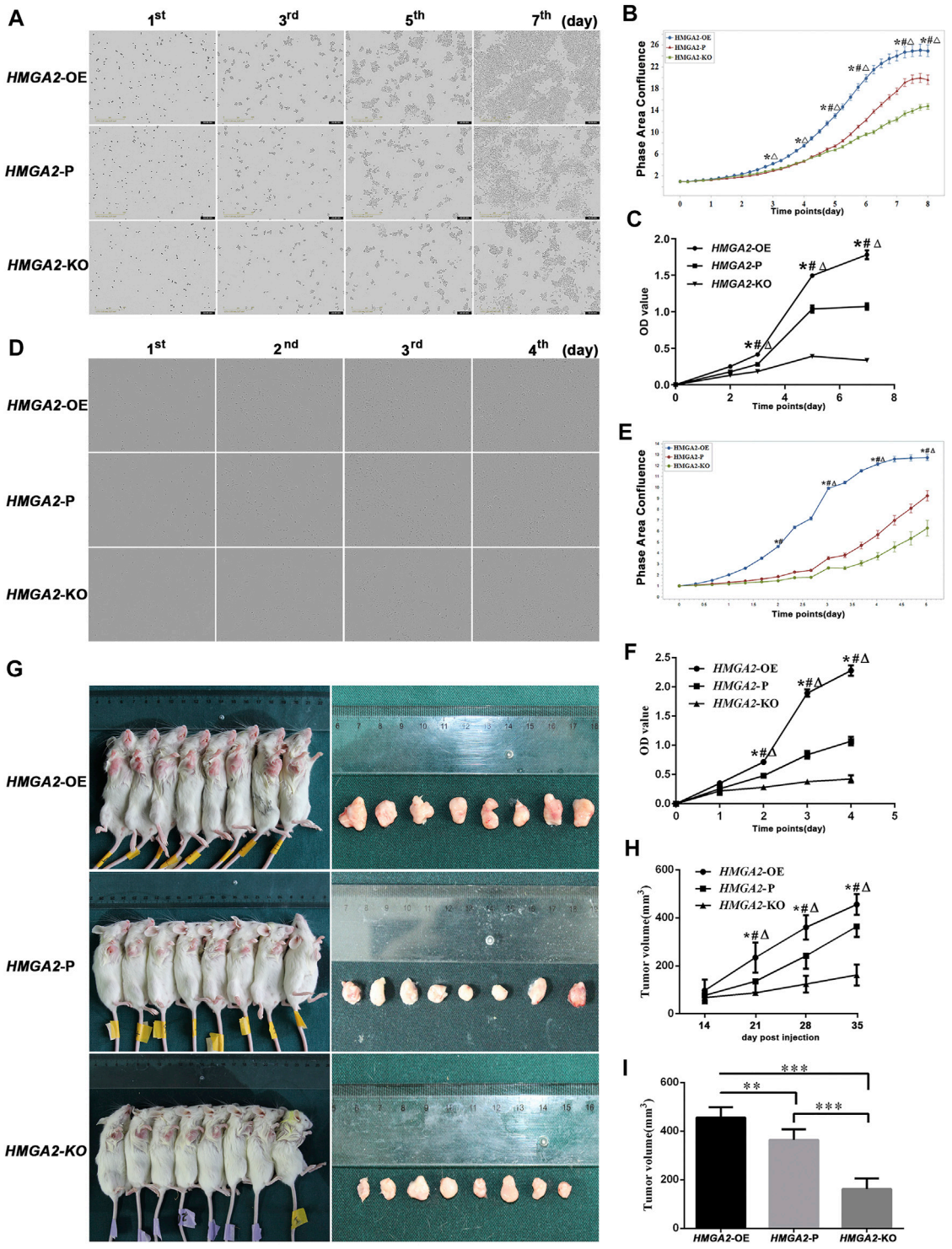


FIGURE 2 | HMGA2 promotes the proliferation of the GC cells *in vitro* and *in vivo*. **(A)** Growth comparison of *HMGA2*-edited MKN-45 cells. The images in the vertical lines are the cell proliferation at different time (first, third, fifth, and seventh days). The pictures in the horizontal lines mean different groups: **upper**, *HMGA2*-OE cells; **middle**, *HMGA2*-P cells; **bottom**, *HMGA2*-KO cells. **(B)** Continuous proliferation records of *HMGA2*-edited MKN-45 cells *via* IncuCyte S3. Abscissa: consecutive days; ordinate: phase area confluence. The phase area confluences of the initial seeded cells in different groups were set as 1, respectively. Blue curve: *HMGA2*-OE cells; red one: *HMGA2*-P cells; green one: *HMGA2*-KO cells. **(C)** MTT assay to detect the growth of *HMGA2*-edited MKN-45 cells. *X*-axis: different days of cell viability; *Y*-axis: OD value. **(D)** Growth comparison of *HMGA2*-edited MGC-803 cells. Vertical lines: different time (first, second, third, and fourth days); horizontal lines: **upper**, *HMGA2*-OE cells; **middle**, *HMGA2*-P cells; **bottom**, *HMGA2*-KO cells. **(E)** Continuous proliferation records of *HMGA2*-edited MGC-803 cells *via* IncuCyte S3.

(*Continued*)

FIGURE 2 | Abscissa: consecutive days; ordinate: phase area confluence. The phase area confluences of the initial seeded cells in different groups were set as 1, respectively. Blue curve: *HMGA2*-OE cells; red one: *HMGA2*-P cells; green one: *HMGA2*-KO cells. **(F)** Growth of *HMGA2*-edited MGC-803 cells measured by the MTT assay. *X*-axis: different days of cell viability; *Y*-axis: OD value. **(G)** Tumor-bearing mice and the isolated xenografts. The NOD/SCID mice were inoculated subcutaneously under the right axilla with *HMGA2*-edited MKN-45 cells at the number of 2×10^6 cells suspended in 100 μL PBS. The vertical lines are the sacrificed mice with xenografts and the isolated tumor correspondingly. **Upper line:** mice inoculated with *HMGA2*-OE MKN-45 cells; **middle line:** *HMGA2*-P MKN-45 cells; **bottom line:** *HMGA2*-KO MKN-45 cells. **(H)** Dynamic recording of the tumor volumes. The volumes of xenografts were measured each week on living mice. *X*-axis: different days; *Y*-axis: tumor volume (mm^3). **(I)** Isolated tumor volumes of the NOD/SCID mice inoculated with different cells. *X*-axis: mice groups of *HMGA2*-OE, *HMGA2*-P, and *HMGA2*-KO cell inoculation, respectively; *Y*-axis: tumor volume (mm^3). Symbols for **(B)**, **(C)**, **(E)**, **(F)**, and **(H)**: *, comparison of *HMGA2*-OE cells with *HMGA2*-P cells; △, comparison of *HMGA2*-OE cells with *HMGA2*-KO cells; #, comparison of *HMGA2*-P cells with *HMGA2*-KO cells; $^{*,\triangle,\#}p < 0.05$. For **(I)**, $^{**}p < 0.01$; $^{***}p < 0.001$, Student's t-test. Error bars represent SD.

attention in the children with certain types of congenital heart defects, and the heterozygous missense mutations of CDK13 would impair magnesium ion binding to ATP in these pediatric sufferers, yet no further studies were performed (Bostwick et al., 2017; Uehara et al., 2018; Hamilton and Suri, 2019; Novakova et al., 2019). The role of CDK13 in cancers, such as ovarian cancer and hepatocellular carcinoma, has been focused on since 2018; however, no underlining mechanism has been reported (Dong et al., 2018; Zeng et al., 2018; Wang et al., 2019). TCGA database showed that *CDK13* is amplified in different categories of cancer, indicating it could contribute to the tumorigenesis and development of cancer in humans. In 2019, it was reported that the triple-negative breast cancer might get benefit from the inhibitor of CDK12/CDK13, and the mechanisms were considered to be related to the enhancement of cell apoptosis by suppression of DNA damage response proteins and the cell cycle arrest induced by dysregulation of cell cycle checkpoint control proteins based on the RNA-seq data (Hopkins and Zou, 2019; Quereda et al., 2019; Tadesse et al., 2021). Although how the cell cycle checkpoint was regulated by these potential proteins has remained unclear, one point should be noticed: the inhibitor of CDK12/CDK13 might be a promising option for some types of cancer.

Gastric cancer (GC) has been in the lightening spot for years because of the high morbidity and low survival rate (Smyth et al., 2020). The uncontrolled cell proliferation of GC is mainly driven by the inordinate cell cycle progression. It was reported that the expression of the cell cycle–related molecules was abnormal in over 50% of the patients with GC (Min et al., 2018). Frustratingly, the inhibitor of CDK4/6, as a promising inhibitor to suppress the G0/G1 phase of the cell cycle, has not proved its efficiency as it is expected (Min et al., 2018). We suspect that there must be certain specific features as far as the GC cell cycle is concerned. Whether other types of CDK inhibitors could fight well against this specific cell cycle disorder of GC has not been explored, let alone the inhibitor of CDK12/CDK13.

High mobility group A2 (HMGA2) is a kind of non-histone chromosomal protein encoded by *HMGA2*. Its role is to modulate transcription by influencing the chromatin architecture through broadly binding to the chromatin and forming the multiprotein complex (Cleynen and Van de Ven, 2008). In the physiological condition, the expression of *HMGA2* is high during embryogenesis, but in most adult and differentiated tissues, the expression is almost undetectable (Zhou et al., 1996; Hammond and Sharpless, 2008; Nishino et al., 2008). Nevertheless, *HMGA2* is re-expressed in many types of cancer (Mansoori et al., 2021), and it could manipulate tumorigenesis, metastasis, and relapse *via* participating in cell cycle, apoptosis, angiogenesis, epithelial–mesenchymal transition, and chemoresistance (Zhao et al., 2018; Wu et al., 2019; Li et al., 2020a; Mansoori et al., 2020). HMGA2 could be considered a novel target gene for the precision therapy due to its vital role in cancer and the specific expression characteristics in different developmental stages (Zhu et al., 2017; Huang et al., 2018). As for its role in GC, HMGA2 is considered to be closely involved in the process of metastasis and the resistance to the medication (Wei et al., 2013; Hombach-Klonisch et al., 2014; Dong et al., 2017; Li et al., 2017; Sun et al., 2017). However, there has been rare attention to the relationship between the GC cell cycle and the expression of HMGA2.

In our study, we firstly found the GC tissues with high HMGA2 expression account for over 80% of GC. HMGA2 was then overexpressed based on the parental human GC cell lines, MKN-45 and MGC-803, and the proliferation of the cells accelerated. We found such acceleration was due to the shortened cell cycle phase transition. We further speculated that CDK13 might be the wirepuller when we analyzed our data from ChIP-seq and luciferase assay. This speculation was also reflected when we investigated the relationship between HMGA2 and CDK13 in 200 pairs of GC together with the adjacent tissue.

Furthermore, high expression of both genes predicts a poorer prognosis than the expression of single genes does. We explore the effect of the novel CDK12/13 inhibiting agent, SR-4835, on the high HMGA2 expression GC cells and consider the inhibition of both genes jointly could reach a satisfied result in high HMGA2 expression GC.

MATERIALS AND METHODS

Cell Lines and Cell Culture

Human gastric cancer cell lines MKN-45 and MGC-803 were purchased from the National Infrastructure of Cell Line Resource. Cell line identities were confirmed by STR profiling. The cells were cultured in RPMI medium 1640 (Gibco, 11875-093) and DMEM (Gibco, 11965-092) supplemented with 10% fetal bovine serum (FBS) (Gibco, 10091-148) and 1% penicillin–streptomycin (Gibco, 15140-122) and were maintained at 37°C with 5% CO_2.

Establishment of the Stable Cell Lines

The sequence of sgRNA used in the study are listed in **Table 1**. Two single-guide (sg) RNAs targeting exon 1 (within the functional AT-hook domain) of *HMGA2* were designed. The pX330 vector (Addgene, 42230) was used to produce pX330-

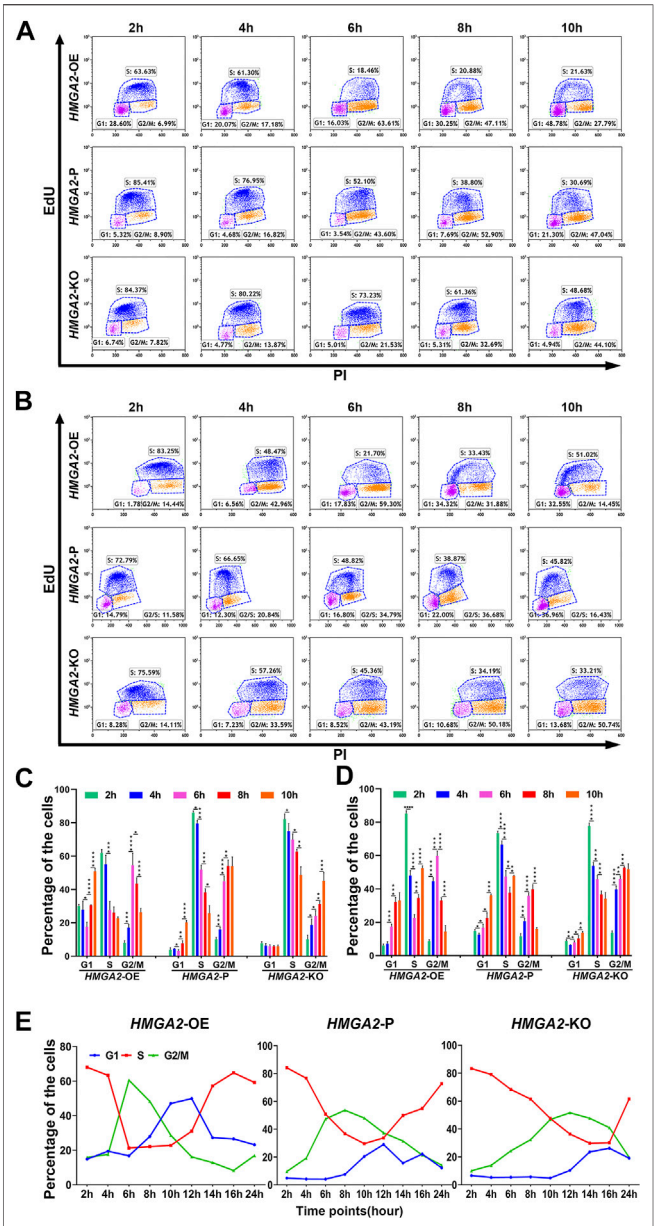


FIGURE 3 | HMGA2 shortens S–G2/M phase transition of the cell cycle in two GC cell lines. **(A, B)** Cell cycle analysis at indicated time points *via* flow cytometry. Results from the **(A)** MKN-45 cells and **(B)** MGC-803 cells. Horizontal lines: **upper**, *HMGA2*-OE cells; **middle**, *HMGA2*-P cells; **bottom**, *HMGA2*-KO cells. Vertical lines: different time points after synchronizing the cells to the G1/S boundary. The time points are 2, 4, 6, 8, and 10 h. EdU negative and PI negative indicate the cells in the G0/G1 phase; EdU positive: cells in the S phase; PI positive: cells in the G2/M phase of the cell cycle. **(C,D)** Statistical histogram of the cellular percentage in each cell cycle phase at indicated time points. **(C)** MKN-45 cells and **(D)** MGC-803 cells. *X*-axis: different cell cycle phases of the different *HMGA2*-edited cells. *Y*-axis: percentage of the cells. **Left flock:** *HMGA2*-OE cells; **middle flock:** *HMGA2*-P cells; **right flock:** *HMGA2*-KO cells. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, Student's t-test. Error bars represent SD. **(E)** Dynamic percentage of the cells in each cell cycle phase recorded at continuous time. The cell cycle of MKN-45 cells was detected for a sequence of 24 h *via* flow cytometry after synchronized to the G1/S boundary, and the detection was conducted every 2 h. **Left:** *HMGA2*-OE MKN-45 cells; **middle:** *HMGA2*-P MKN-45 cells; **right:** *HMGA2*-KO MKN-45 cells. The red, blue, and green lines represent the proportion of cells in S, G1, and G2/M phases, respectively.

HMGA2-gRNA1 and pX330-HMGA2-gRNA2 plasmids. Lipofectamine LTX and Plus Reagent (Invitrogen, 2135022) were used in transfection. The single cells were sorted *via* flow cytometry (BECKMAN COULTER, MoFlo Astrios EQs, United States). The DNA and RNA were extracted, and PCR and Sanger sequencing were performed. The expression level of *HMGA2* was verified *via* RT-PCR and western blot assay.

HMGA2-overexpressed cells were established using pCMV6-Entry-*HMGA2* (OriGene Technologies, RC210804, China).

Inhibition of CDK13

First, *CDK13* siRNA (si-h-*CDK13*, siB0804150912271, RIB BIO) or negative control siRNA (siR *NC*, siN000001-1-5, RIB BIO) was transfected into MKN-45 and MGC-803 using Lipofectamine® RNAiMAX Reagent (Invitrogen, 13778-150), respectively. Second, SR-4835 (TargetMol®, T8325/2387704-62-1) was used as the CDK13 inhibitor and added into the cell culture at 60 nM. The siRNA of CDK13 are listed in **Table 1**.

RT-PCR and Western Blot

Total RNA from the cells was extracted using Trizol™ Reagent (Invitrogen, 15596018). Reverse transcription was performed using PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, RR047A). RT-PCR was performed using TB Green® Premix Ex Taq™ II (Tli RNaseH Plus, Takara, RR820A) on the CFX96™ Real-Time System (BIO-RAD). "The primer sequence used in the study are listed in **Table 1**.

The cells were lysed in RIPA (Solarbio, #R0020) containing a protease inhibitor cocktail (Millipore, 539136). The protein concentration was determined using the Pierce™ BCA Protein Assay Kit (Thermo Scientific, 23227). Specific antibody–protein complexes were detected with the ECL-PLUS Kit (Thermo, M3121/1859022), and the images were captured *via* a gel imaging system (GE Healthcare Life Scientific, Amersham Imager 600). The primary antibodies are as follows: anti-HMGA2 (CST, #5269S), anti-CDK13 (Invitrogen, VB2774502), and anti-GAPDH (Santa Cruz Biotechnology, sc-32233).

Cell Proliferation *In Vitro* and *In Vivo*

The cell proliferation was recorded *via* an IncuCyte live-cell imaging system (Essen Bioscience, IncuCyte S3 2018B) for seven consecutive days. Meanwhile, cell viability was also determined *via* CellTiter 96 Non-Radioactive Cell Proliferation Assay Kit (Promega, G4001).

The animal experiments were approved by the animal ethics committee. Female NOD/SCID mice (7–8 weeks old, No. SCXK (Jing) 2016-0006) were purchased from Beijing Vitonlihua Experimental Animal Technology Co., Ltd. (Beijing, China), and were raised abiding by the principles of animal welfare. The processed cells were injected into the right subcutaneous axilla of the mice. The volume of the xenografts had been measured and calculated every week till the mice were sacrificed. The xenografts isolated from each group were measured and recorded.

Cell Cycle and Apoptosis Assay

For detecting the cell cycle, the cells were treated with 2 mM thymidine (Sigma T1895), synchronized to the G1/S boundary,

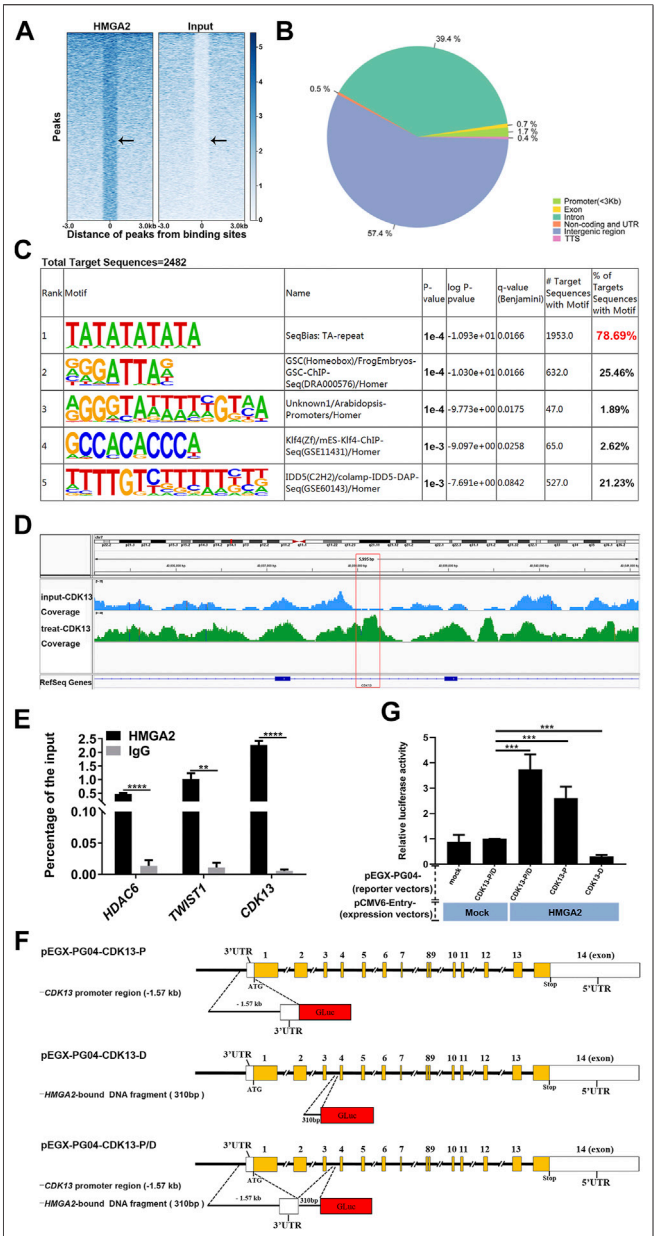


FIGURE 4 | CDK13 is the direct target of HMGA2. **(A)** Heat map of ChIP-seq peaks. **Left part:** treated sample; **right part:** input control. The black arrow points to the HMGA2 binding sites. **(B)** Pie chart of the genomic distribution of HMGA2 binding sites. Blue part: intergenic region; dark green part: intron; light green part: promoter; yellow part: exon; remaining ones: TSS, non-coding and UTR. **(C)** Known motif enrichment results of the total target sequences immune-precipitated by HMGA2. **(D)** Enrichment analysis of the HMGA2 binding peaks at intron 3 of *CDK13*. The red box represents the specific binding site of HMGA2 in intron 3 of *CDK13*. The track signal is calculated by the average read count of a 500bp window. **(E)** Confirmation of the quantity of pull-downed *CDK13 via* ChIP-qPCR. *HDAC6* and *TWIST1* act as the positive controls. Abscissa: different genes; ordinate: percentage of the input. Black columns: DNA pulled down by HMGA2; gray one: DNA pulled down by IgG. **(F)** Schematic diagrams of recombinant Gluc/SEAP dual-reporter vectors. pEGX-PG04-CDK13-P **(the upper line):** the insert fragment cloned into pEZX-PG04 is the promoter region of *CDK13*, that is, the 1.57 kb upstream region from the first ATG codon on exon 1 of the *CDK13* gene; pEGX-PG04-CDK13-D **(the middle line):**

(*Continued*)

FIGURE 4 | DNA fragment immune-precipitated by HMGA2, that is, the 310bp fragment on intron 3 of the *CDK13* gene; pEGX-PG04-CDK13-P/D **(the bottom line):** promoter region of *CDK13* plus the DNA fragment immune-precipitated by HMGA2. **(G)** Relative luciferase activity of the reporter vectors with or without HMGA2. *HMGA2*-KO MKN-45, in which endogenous HMGA2 does not express, was transfected with the HMGA2 expression vector together with luciferase vectors in the following order: pEZX-PG04-Mock *vs* pCMV6-Entry-Mock; pEZX-PG04-CDK13-P/D *vs* pCMV6-Entry-Mock; pEZX-PG04-CDK13-P/D *vs* pCMV6-Entry-HMGA2; pEZX-PG04-CDK13-P *vs* pCMV6-Entry-HMGA2; and pEZX-PG04-CDK13-D *vs* pCMV6-Entry-HMGA2. Luciferase activities were normalized to SEAP levels, and the value of the pEZX-PG04-Mock *vs* pCMV6-Entry-Mock group was set as 1. $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, Student's t-test. Error bars: SD.

and stained using Click-iT™ EdU Alexa Flour™ 488 Flow Cytometry Assay Kit (Invitrogen, C10425). For detecting apoptosis, the cells were stained by FITC Annexin VApoptosis Detection Kit I (BD Biosciences, 556547) at the indicated days after seeding. The stained cells were detected by flow cytometry (BECKMAN COULTER, Navios, United States), and the data were analyzed using Kaluza Analysis Software (version 2.1).

Chromatin Immunoprecipitation Following Sequencing Assay

According to the operating instructions, the ChIP procedure was conducted on *HMGA2*-overexpressed MKN-45 cells using HMGA2 Rabbit mAb (CST, #5269S). Histone H3 (D2B12) XP® Rabbit mAb (CST, #4620) acted as the positive control, and Normal Rabbit IgG (CST, #2729) acted as the negative control correspondingly. The results were verified by PCR. The primer we used was SimpleChIP® Human RPL30 Exon 3 Primers 1 (CST, #7014). IP efficiency was calculated with the percent input method. SimpleChIP® Plus Enzymatic Chromatin IP Kit (Magnetic Beads, CST, #9005) was used in the experiments.

The DNA was evaluated through a NanoPhotometer® spectrophotometer (IMPLEN, CA, United States) and Qubit® DNA Assay Kit in the Qubit® 2.0 Fluorometer (Life Technologies, CA, United States). The cDNA library was prepared and sequenced on Illumina's NovaSeq platform to generate 150 base pair-end reads. The raw data were processed and mapped to the human reference genome (version hg19) using Bowtie (version 2.2.5). Peaks were identified by MACS (version 2.1.1.20160309) with $p < 0.001$. We calculated the read counts of genomic 500 bp regions for the treated sample and input sample. The UCSC RefSeq Genes database was used for peak annotation.

Construction of Luciferase Vectors and Dual Luciferase Assay

GAPDH-PG04 and pEZX-PG04 vectors were purchased from GeneCopoeia. The recombined luciferase vectors were conducted and named pEZX-PG04-CDK13-P, pEZX-PG04-CDK13-D, and pEZX-PG04-CDK13-P/D, respectively. pCMV6-Entry-HMGA2 was used as the HMGA2 expression vector. The Gluc/SEAP dual-reporter vectors, together with *HMGA2* expression vectors, were transfected into *HMGA2*-knocked-out MKN-45 cells. The

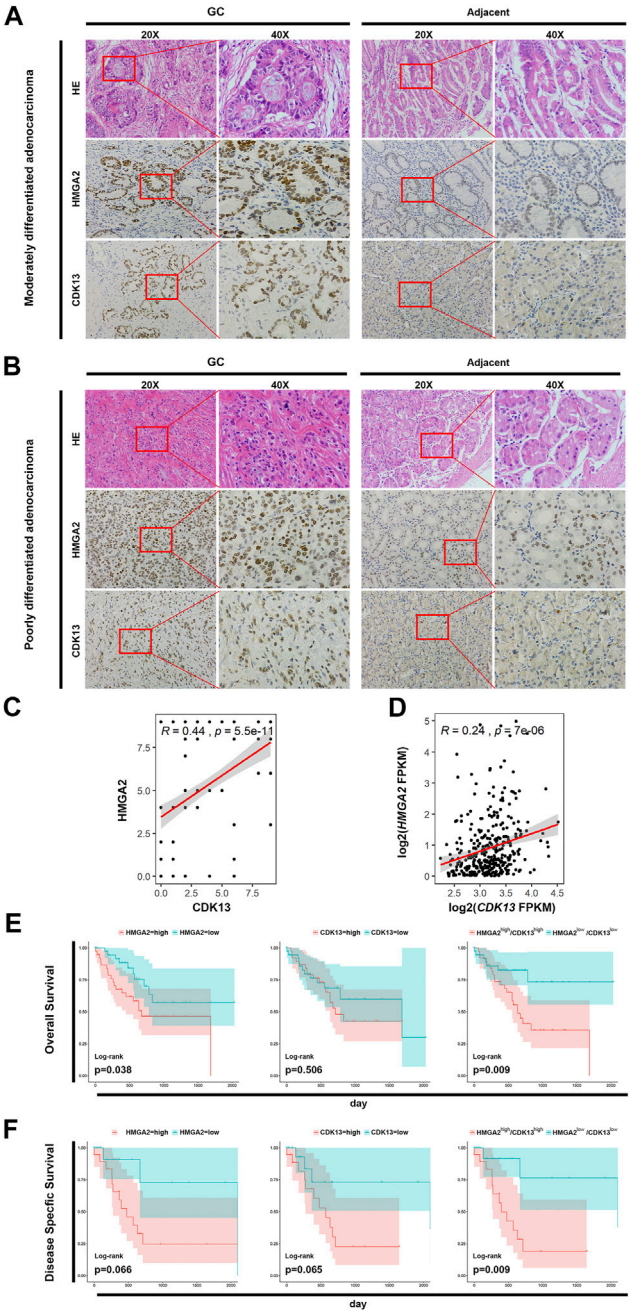


FIGURE 5 | HMGA2 and CDK13 are associated with tumorigenesis of GC. **(A)** Expression of HMGA2 and CDK13 in the tissues of the patients with moderately differentiated STAD. The **left** two columns are the HE staining and immunohistochemical results of the GC tissue. The **right** two columns are those from the adjacent tissue. For both GC and the adjacent group, respectively: the first column, ×20; the second one, images magnified from the red frames of the **left** side, ×40. Compared with the corresponding location of the adjacent tissue, the morphological disorder of cell arrangement is obvious in the GC tissue although the glands exists, which is consistent with the characteristics of this type of GC. The brown colored nuclei in the images of the second and the third horizontal line suggest HMGA2 positive and CDK13 positive, respectively. Therefore, the expressions of HMGA2 and CDK13 in GC are much higher than those in the adjacent tissue. **(B)** Expression of

(*Continued*)

FIGURE 5 | HMGA2 and CDK13 in the tissues of the patients with poorly differentiated STAD. The arrangement of images is the same as that of A. The HE staining and immunohistochemical assay were also performed on the poorly differentiated GC. The cells are more disorderly growing, losing the normal glandular morphology, and the sizes of the cells are more varied compared with those in the moderately differentiated GC, suggesting the high grade of malignancy. Positive HMGA2 or CDK13 immunohistochemical staining (in brown) localizes mainly in the nucleus of the cells. Strong positive staining in GC tissues and weak positive staining in corresponding adjacent tissues are shown. **(C)** Pearson's correlation between the expression of HMGA2 and CDK13 from the immunohistochemistry results of the 200 GC tissues was calculated. *X*-axis: expression of CDK13; *Y*-axis: expression of HMGA2; $R = 0.44$, $p = 5.5e\text{-}11$. **(D)** Pearson's correlation between *HMGA2* and *CDK13* based on the RNA-seq results of the 416 STAD patients in TCGA database was calculated. $R = 0.24$, $P = 7e\text{-}06$. **(E)** Kaplan–Meier survival analysis for overall survival (OS). The expression of *HMGA2* and *CDK13* of 76 patients with tubular STAD was stratified to OS, respectively and simultaneously. **(F)** Kaplan–Meier survival analysis for disease-specific survival (DSS). The expression of *HMGA2* and *CDK13* of 44 patients with stage IV STAD was stratified to DSS, respectively and simultaneously. For **(E)** and **(F)**: **left**, expression of *HMGA2*; **middle**, expression of *CDK13*; **right**, expression of *HMGA2* and *CDK13* simultaneously. Pink color: high expression; blue: low expression.

activities of Gluc and SEAP were measured using Secrete-Pair™ Luminescence Assay Kit (GeneCopoeia TM), and Gaussia luciferase activities were normalized to SEAP levels.

Human Tissue Specimens and Immunohistochemistry

The procedures of human tissue sampling were approved by the Medical Ethics Committee of Inner Mongolia People's Hospital. The enrolled patients signed the informed consent form beforehand. Surgical-resected tissues from the hospitalized GC patients were sectioned into slices of 5 μm thick, and 200 pairs of the tumor and the adjacent tissue were collected. All sections were stained with the primary antibodies of anti-HMGA2 (1:100, CST, 5269S) and anti-CDK13 (1:100, Invitrogen, VB2774502) using the IHC kit (for rabbit primary antibody, Bioss Antibodies, Cat: IHC001). The immunoreactivity of HMGA2 and CDK13 was evaluated based on the Quick Score System (Detre et al., 1995).

RESULTS

The Proliferation and Tumorigenicity Were Different Among *HMGA2*-OE, *HMGA2*-P, and *HMGA2*-KO Cells

HMGA2 cells in parental MKN-45 and MGC-803 cells were knocked out in a biallelic manner *via* a CRISPER-Cas9-sgRNA gene editing system individually and named *HMGA2*-KO MKN-45 and *HMGA2*-KO MGC-803. Simultaneously, *HMGA2* cells were overexpressed based on the parental MKN-45 and MGC-803 cells. They were called *HMGA2*-OE MKN-45 and *HMGA2*-OE MGC-803, correspondingly. Sanger sequencing was conducted to investigate the gene knockout results, which

TABLE 1 | Sequences of primers, sgRNA, and siRNA used in the study.

Gene	Forward (5′ to 3′)	Reverse (5′ to 3′)
HMGA2-ORF	AGAGACCCAGGGGAAGACC	AGTGGCTTCTGCTTTCTTTTGAG
HMGA2	ACGTCCGGTGTTGATGGTG	TCTTGCTGCTGCTTCCTGG
CDK13	CAAGCATAGGAGCCAAGGAGAAG	AATCAGCAAGAAGACATCGGAGTT
TWIST1	GTCACAATGCGGAGCCTAAT	AAACCCAGTCCATGGGAAAG
HDAC6	GGGCGGTGATTGGTTGG	GATTCTCTTTCCCTGGTCTTGC
β-Actin	TCCCTGGAGAAGAGCTACGAGC	TGCCACAGGACTCCATGCCCAG
sgRNA	**Forward (5′ to 3′)**	**Reverse (5′ to 3′)**
HMGA2-1	CACCGGTCCTCTCTTCTGAGGCGCT	AAACAGCGCCTCAGAAGAGAGGACC
HMGA2-2	CACCGTGGGGCGGCAGGTTGTCCCT	AAACAGGGACAACCTGCCGCCCCAC
siRNA	**Sequence (5–3′)**	
siRNA 1#	CGACGTAGTTTCATTGGAA	
siRNA 2#	GAGAAATGGTAGCCTTAAA	
siRNA 3#	GCAATATCGTCGAAAGTTA	

TABLE 2 | Expression difference of HMGA2 and CDK13 between GC and the adjacent tissue.

Percentage of positive expression	GC tissue (N = 200)	The adjacent tissue (N = 200)	*p*-Value
HMGA2	80% (160)	25% (50)	<0.001
CDK13	60% (120)	15% (30)	<0.001

showed 20-base pairs (bp) and 22-bp in *HMGA2*-KO MKN-45 and *HMGA2*-KO MGC-803 were deleted, respectively, implying the frame shift (**Figures 1A,B,F,G**). Subsequently, the expression of HMGA2 in *HMGA2*-KO MKN-45, *HMGA2*-KO MGC-803, *HMGA2*-OE MKN-45, and *HMGA2*-OE MGC-803 cells was detected using RT-PCR and western blot. The mRNA expression of *HMGA2* in *HMGA2*-OE MKN-45 cells was 30 times higher than the one in the parental cells. *HMGA2* in *HMGA2*-OE MGC-803 was 20 times higher than that in the parental MGC-803; on the contrary, the mRNA expression of *HMGA2* in *HMGA2*-KO MKN-45 cells was five times lower than that in the parental MKN-45. *HMGA2* in *HMGA2*-KO MGC-803 cells was two times lower than that in the parental MGC-803 (**Figures 1C,H**). As far as the protein expression of HMGA2 was concerned, the protein expression of HMGA2 in *HMGA2*-OE MKN-45 cells was seven times higher than that in the parental cells and the one in *HMGA2*-OE MGC-803 was five times higher than that in the parental MGC-803; on the contrary, the one was almost undetectable in both *HMGA2*-KO MKN-45 and *HMGA2*-KO MGC-803 cells (**Figures 1D,E,I,J**).

The proliferation ability of gene-edited cells and the parental cells was observed *via* IncuCyte S3 and MTT assay, respectively. As is shown in **Figures 2A,B**, no difference was found on day 0 because the number of planted cells was equal. On day 2, the descending order of proliferating speed was *HMGA2*-OE MKN-45, *HMGA2*-P MKN-45, and *HMGA2*-KO MKN-45, but without statistical significance. The growing speed of *HMGA2*-OE MKN-45 cells leapfrogged the parental cells and the *HMGA2*-KO MKN-45 cells from day 3 to day 8, and the parental cells transcended the *HMGA2*-KO MKN-45 cells from day 5 to day 8 with statistical significance, respectively (**Figures 2A,B**). The same trend was also proved *via* the MTT assay: such order began to show a statistical difference from day 3 to day 7 when the observation period was over (**Figure 2C**), suggesting that HMGA2 promotes the proliferation of the GC cells *in vitro*. Another GC cell line (MGC-803) also obtained the cell proliferating observation, presenting the same results as those in MKN-45 (**Figures 2D–F**).

To reassure the tumorigenic ability of the newly built cell lines, the *HMGA2*-edited MKN-45 cells were inoculated into the NOD/SCID mice. The xenografts were measurable at the 14th day in each group, and the tumorigenesis rate was 100%. The size of xenografts was too small to be compared among groups at the beginning until that of each group had the statistical difference at the 21st day. The descending order of xenograft size was *HMGA2*-OE, *HMGA2*-P, and *HMGA2*-KO MKN-45. When the time went by, such difference was increasingly statistically obvious (**Figures 2G–I**).

In a word, both the cellular phenotype and the xenograft animal experiment proved the success and stability of the newly gene-edited cell lines. The results confirmed that HMGA2 significantly increased cell proliferation capacity in a HMGA2-dependent manner both *in vivo* and *in vitro*.

HMGA2 Shortened the S–G2/M Phase Transition and Influenced Little on Apoptosis

The cell cycle was compared among *HMGA2*-edited MKN-45/MGC-803 cells *via* flow cytometry, respectively. Taking MKN-45, for example, most of the *HMGA2*-OE cells spent 6 h to progress

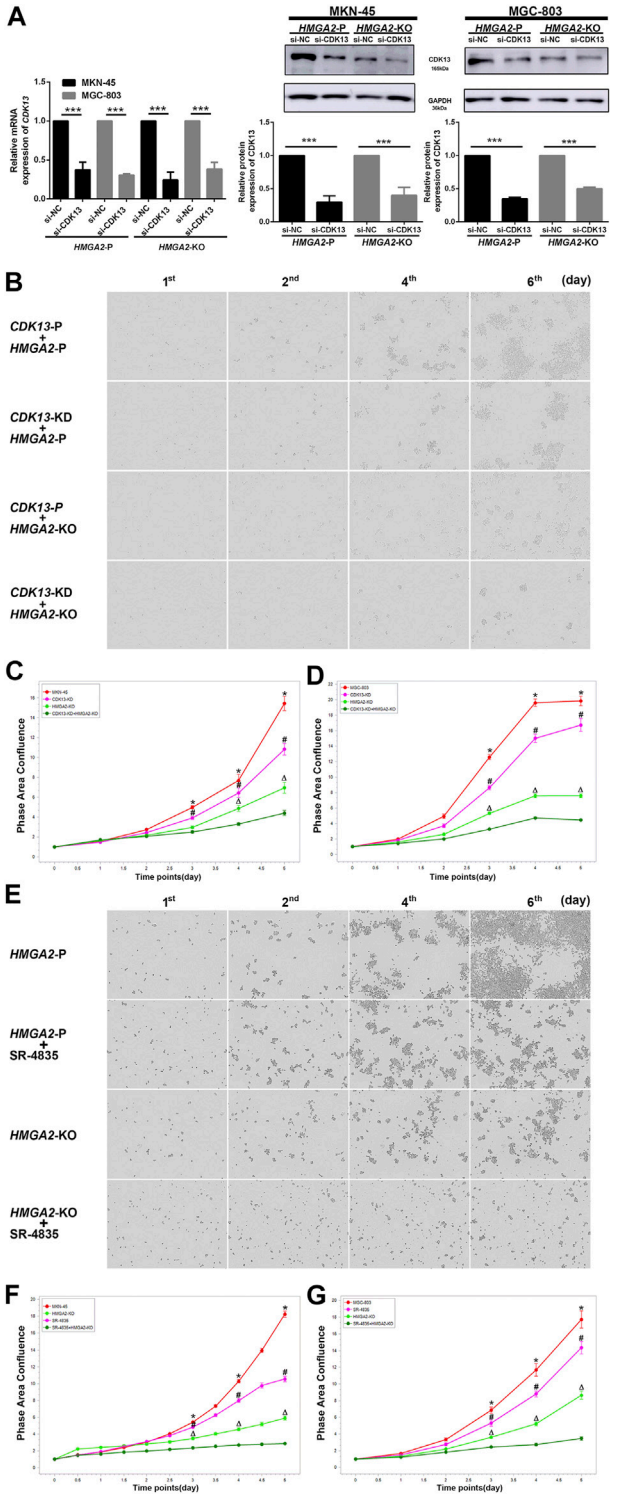


FIGURE 6 | Inhibiting CDK13 and HMGA2 simultaneously suppresses the GC cells most. **(A)** Efficiency of knocking down CDK13 in two GC cell lines. **Left:** the efficiency was verified by RT-PCR. *CDK13* was knocked down based on *HMGA2*-P and *HMGA2*-KO MKN-45/MGC-803, respectively. *X*-axis: different cell groups. *Y*-axis: expression of *CDK13*. **Middle** and **right:** western blot results of the expression of CDK13 in *HMGA2*-P and *HMGA2*-

(*Continued*)

FIGURE 6 | KO MKN-45 and MGC-803, respectively. ***$p < 0.001$. **(B)** Proliferation of the MKN-45 cells in different groups. The images in the vertical lines are the cell proliferation at different time (first, second, fourth, and sixth days). The pictures in the horizontal lines mean different groups: ***upper***, *CDK13*-P plus *HMGA2*-P cells; second, *CDK13*-KD plus *HMGA2*-P cells; third, *CDK13*-P plus *HMGA2*-KO cells; ***bottom***, *CDK13*-KD plus *HMGA2*-KO cells. KD: knockdown; KO: knockout; P: parental. **(C)** Continuous records of the proliferation of the different gene-edited MKN-45 cells. **(D)** Continuous records of the proliferation of the different gene-edited MGC-803. For **(C)** and **(D)**: abscissa, consecutive days; ordinate, phase area confluence shown *via* IncuCyte S3. Red curve: *CDK13*-P plus *HMGA2*-P cells; pink one: *CDK13*-KD plus *HMGA2*-P cells; light green one: *CDK13*-P plus *HMGA2*-KO cells; dark green one: *CDK13*-KD plus *HMGA2*-KO cells. The phase area confluences of the initial seeded cells in different groups were set as 1, respectively. *: comparison of *CDK13*-P plus *HMGA2*-P cells with *CDK13*-KD plus *HMGA2*-KO cells; #: comparison of *CDK13*-KD plus *HMGA2*-P cells with *CDK13*-KD plus *HMGA2*-KO cells; △: comparison of *CDK13*-P plus *HMGA2*-KO cells with *CDK13*-KD plus *HMGA2*-KO cells; $^{*,\triangle,\#}p < 0.05$. **(E)** Proliferation of the different treated MKN-45 cells. The images in the vertical lines are the cell proliferation at different time (first, second, fourth, and sixth days). The pictures in the horizontal lines mean different groups: **upper**, *HMGA2*-P cells; the second line, *HMGA2*-P plus SR-4835 treated cells; the third line, *HMGA2*-KO cells; **bottom**, *HMGA2*-KO plus SR-4835 treated cells. **(F)** Continuous proliferation records of the different types of MKN-45 cells. **(G)** Continuous proliferation records of the different types of MGC-803. Abscissa: consecutive days; ordinate: phase area confluence shown *via* IncuCyte S3. Red curve: *HMGA2*-P cells; pink one: *HMGA2*-P plus SR-4835 treated cells; light green one: *HMGA2*-KO cells; dark green one: *HMGA2*-KO plus SR-4835 treated cells. The phase area confluences of the initial seeded cells in the different groups were set as 1, respectively. *: comparison of *HMGA2*-P cells with *HMGA2*-KO plus SR-4835 cells; #: comparison of *HMGA2*-P plus SR-4835 cells with *HMGA2*-KO plus SR-4835 cells; △: comparison of *HMGA2*-KO cells with *HMGA2*-KO plus SR-4835 cells; $^{*,\triangle,\#}p < 0.05$.

from the S phase to G2/M; when time increased to 8 and 10 h, more cells at the G2/M phase entered the G1 phase of the next cell cycle. In detail, as time went by, the number of *HMGA2*-OE cells at the S phase gradually dropped and reached the valley bottom at 6 h, while the ones at the G2/M phase increased and arrived at the peak at the same time point (6 h); on the contrary, the G1 cell number rose with the S and G2/M phase progression (**Figure 3A** upper, **Figure 3C** left side).

Interestingly, the number of S-phase *HMGA2*-P cells did not decrease to the bottom while the observing time period was about to complete. The decreasing trend, no platform, was observed as for the S-phase proportion while the G2/M-cell proportion was increasing adversely, but the dropping trend began to appear at 10 h, showing the G2/M cells began to progress into the next G1 phase (**Figure 3A** middle, **Figure 3C** middle). Such phenomenon was also observed in the *HMGA2*-KO groups; however, the proportion of the S-phase cells was 48.68% at 10 h, suggesting it still needs even longer time for most of the *HMGA2*-KO cells to progress into the next cycle (**Figure 3A** bottom, **Figure 3C** right side). The dynamic changes of the cell cycle observed among *HMGA2*-edited MGC-803 cells are consistent with what we observed in MKN-45 (**Figures 3B,D**).

To find out the exact time point when the number of S-phase cells in the *HMGA2*-OE, *HMGA2*-P, and *HMGA2*-KO groups was the least one, cell cycles were detected every 2 h for a sequence of 24 h (**Figure 3E**). Taking the percentage of S-phase cells in each group, for example, the ration of the cells in the S phase in the *HMGA2*-OE

group began to decrease down to the valley bottom at the sixth hour, and after a flat for the next 4 h, the trend rose to the next peak, meaning the next cell cycle begins. However, such valley bottom was not the same as that in the other two groups, located at the 10th hour for *HMGA2*-P cells and the 14th hour for the *HMGA2*-KO cells, respectively. Thus, the portrait of how the expression level of HMGA2 influences the cell cycle was clearly presented. In brief, the lower the expression of HMGA2 in GC cells, the more the arrest of the cell cycle progression; furthermore, the arrest occurred in the S phase.

We investigated the apoptosis in *HMGA2*-edited MKN-45/MGC-803 cells, respectively. It turned out that there was no statistical significance among the groups (**Supplementary Figure 1**).

CDK13 Was the Direct Target of HMGA2 in Gastric Cancer

To further elucidate the underlining mechanism by which HMGA2 induced the change of phenotype in GC cells, we performed ChIP-seq analysis of *HMGA2*-OE MKN-45 cells with anti-HMGA2 antibody. The results shown in **Figures 4A–C** confirmed the trustworthy findings of ChIP-seq. The Gene Ontology analysis of HMGA2-target genes showed that CDK13 was involved in the regulation of cell population proliferation (GO: 0042127). The enrichment of *CDK13* was also proved; as is shown in **Figure 4D**, compared with those in the input group, the peaks of the genes pulled down through HMGA2 are much higher in the district of *CDK13* intron 3, which was confirmed *via* ChIP-qPCR, and the results presented that the abundance of *HDAC6* and *TWIST1* which are the well-known target genes of HMGA2 and used as the positive control was much less than the abundance of *CDK13*; this can be explained as follows: HMGA2 bound efficaciously to a certain DNA area of *CDK13* (**Figure 4E**).

The Gluc/SEAP dual-reporter assays were performed to verify whether *CDK13* expression is directly regulated by HMGA2 in the live cells. Comparing the *HMGA2*-KO cells co-transfected with pEZX-PG04-CDK13-P/D and pCMV6-Entry-Mock to the cells co-transfected with pEZX-PG04-CDK13-P/D and pCMV6-Entry-HMGA2, the stronger luciferase activity in the latter one proved that HMGA2 directly upregulated the expression of CDK13 ($p < 0.001$). Comparing the cells co-transfected with pEZX-PG04-CDK13-P and pCMV6-Entry-HMGA2 to the cells co-transfected with pEZX-PG04-CDK13-P/D and pCMV6-Entry-HMGA2, the luciferase activity was much stronger in the latter one, suggesting the DNA fragment found by ChIP is probably the enhancer of *CDK13* (**Figures 4F,G**).

HMGA2 and CDK13 Were Highly Expressed in Gastric Cancer and Related With Poorer Prognosis

The expression of both *HMGA2* and *CDK13* in the stomach adenocarcinoma (STAD) was much higher than those in the normal mucosae based on TCGA database (**Supplementary Figure 2**). To confirm such trends, cancer tissues as well as the adjacent tissues from 200 STAD patients were processed and the expression level of HMGA2 and CDK13 was explored *via* the immunohistochemical method, and the results are consistent with those from TCGA database (**Figures 5A,B**). The percentage of HMGA2 and CDK13 positive was approximately 80% in GC *vs* 25% in the adjacent tissue and 60% in GC *vs* 15% in the adjacent tissue, respectively (**Table 2**). The correlation of HMGA2 and CDK13 was statistically analyzed, and $R = 0.44$, $p = 5.5e\text{-}11041$ (**Figure 5C**). What we found in the GC patients was consistent with the results from analyzing TCGA database (**Figure 5D**).

How do *HMGA2* and *CDK13* affect the overall survival (OS)? Survival curves relating to the expression level of *HMGA2* and *CDK13* were analyzed based on TCGA database, respectively and jointly. In the patients with tubular STAD, the OS of the patients with either low *HMGA2* expression or low *CDK13* expression has more optimistic prognosis than the cases with high expression. The OS of the patients with both high *HMGA2* expression and high *CDK13* expression simultaneously was much poorer than those with low expression. Furthermore, the statistical significance of the OS difference between the simultaneous high expression of both genes and low expression was more obvious than the OS difference between the patients with high and low expressions of the single gene (**Figure 5E**). Such characteristic can also be seen in the disease-specific survival (DSS) analysis of the patients with stage IV of diffuse type STAD (**Figure 5F**). It could be inferred that the patients with high *HMGA2* and *CDK13* expression GC have poorer prognosis. Targeting HMGA2 and CDK13 associatively would be the possible promising new therapy to enhance the five-year survival rate of GC.

Synergic Inhibition of HMGA2 and CDK13 Had the Most Suppressing Efficacy on the Growth of the Gastric Cancer Cells

The *HMGA2*-KO MKN-45/MGC-803 cells together with their parental version were used to verify the anti-cancer efficacy of inhibiting HMGA2 and CDK13 associatively. They are as follows: *CDK13* knocked down only (*CDK13*-KD + *HMGA2*-P), *HMGA2* knocked out only (*CDK13*-P + *HMGA2*-KO), *CDK13*-KD together with *HMGA2*-KO (*CDK13*-KD + *HMGA2*-KO), parental (*CDK13*-P + *HMGA2*-P) MKN-45 and MGC-803. The efficiency of knocking down *CDK13 via* siRNA was confirmed by RT-PCR and western blot, and the efficiency was about 60% (**Figure 6A**). As expected, the growth of *CDK13*-KD + *HMGA2*-KO cells was suppressed the most; the ascending order of cell proliferation was *CDK13*-P + *HMGA2*-KO, *CDK13*-KD + *HMGA2*-P, and *CDK13*-P + *HMGA2*-P MKN-45/MGC-803 (**Figures 6B–D**). Subsequently, SR-4835 was used to inhibit CDK13 as an alternative method except siRNA. It turned out that no matter how CDK13 was inhibited, whether by SR-4835 or siRNA, the speed of the cell growth in different groups has maintained the same order (**Figures 6E–G**).

DISCUSSION

GC has been in the lightening spot for years because of the high morbidity and low survival rate (Smyth et al., 2020). HMGA2 has been reported as a GC-promoting gene, but how HMGA2 regulates the cell cycle in GC cells has not been illustrated clearly. Some researchers have proved that HMGA2 could promote the G1/S and G2/M phase

transitions, respectively, in the ovarian cancer and leukemia (Malek et al., 2008; Zhang et al., 2018b). In our study, the overexpression of HMGA2 accelerated the S–G2/M transition in the GC cells, instead of the G0/G1 phase, which was consistent with the findings in leukemia (Tan et al., 2016). On the contrary, some inhibitors of the cell cycle are coming into clinical practice. For instance, CDK4/6 inhibitors are providing survival benefit to the certain types of breast cancer characterized by G0/G1 acceleration induced by the CDK4/6-RB1 pathway (Goel and Tolaney, 2019; Zhang et al., 2019). We explored the anti-proliferation effect of a CDK4/6 inhibitor, palbociclib, on *HMGA2*-edited GC cells and found the cells were insensitive to palbociclib (the data are not shown). This phenomenon was also confirmed by Ahrum Min et al. (2018) who reported that out of 10 human GC cell lines, only four were sensitive to CDK4/6 inhibitors and the other six were less sensitive or even insensitive (Min et al., 2018). Thus, we deduced that CDK4/6 inhibitors could not work well in some types of GC. We then carried out transcriptome sequencing, ChIP-seq, and luciferase assay to reveal the specific cell cycle–related regulator of S–G2/M phase transition in high *HMGA2* expression GC and found HMGA2 bound to intron 3 of the *CDK13* gene directly. This is the first study that elucidated the relationship between HMGA2 and CDK13. Accordingly, we suspect CDK13 might involve in the S–G2/M transition. Quereda et al. (2019) reported CDK12/CDK13 can significantly upregulate the S–G2/M–progressing genes (Quereda et al., 2019). Interestingly, the high expression of both *HMGA2* and *CDK13* jointly predicts a poorer prognosis in our research; therefore, choosing the right cell cycle inhibitors and jointly inhibiting CDK13 and HMGA2 might be an effective strategy for the high *HMGA2* expression GC.

CDK13 has been found to have the increasingly important role in cancer biomarkers and therapeutic targets in recent years (Dong et al., 2018; Zeng et al., 2018; Quereda et al., 2019; Wang et al., 2019), while it has detailed function and the underlying mechanism has not been clearly investigated. The amino acid identity between CDK13 and CDK12 is approximately 50%; furthermore, 92% of their main functional structures are the same (Greifenberg et al., 2016). Because CDK12 and CDK13 have the identical conserved kinase domains and activating partner (cyclin K), CDK13 has been studied as the companionship of CDK12 from the beginning, and the most light has been shed on CDK12, veiling the real face of CDK13 (Greifenberg et al., 2016; Fan et al., 2020; Tadesse et al., 2021). Thus, the function and the mechanism of CDK12 have been comparatively clear; moreover, the inhibitor of CDK12 was invented and has already entered the clinical trials (Blazek et al., 2011; Dubbury et al., 2018; Liang et al., 2020). However, there is no specific inhibitor for CDK13 until THZ531, a selective CDK12/13 inhibitor, was found in 2016 (Zhang et al., 2016). Unfortunately, THZ531 was not suitable for clinical use due to the problem of bioavailability and toxic off-target (Hopkins and Zou, 2019). The recently discovered SR-4835 (a selective dual inhibitor of CDK12/CDK13) exhibits excellent anti-cancer therapeutic effects, especially when combined with PARP inhibitors (DNA-damaging chemotherapy) or anti-PD-1 (checkpoint inhibition) (Quereda et al., 2019; Li et al., 2020b). Moreover, it is reported that the IC50 of SR-4835 for CDK12 and CDK13 is different (98 nM for CDK12 *vs* 4.9 nM for CDK13). Perhaps, the IC50 difference between them could be utilized for the independent study of CDK13. In our study, we also investigated the anti-cancer effect of SR-4835 at low concentration on high *HMGA2* expression GC, and the results showed the excellent anti-cancer effect, especially when combined with HMGA2 knockout. Therefore, our findings might provide the proof for selecting the type of cell cycle inhibitor in treating GC and the basic data for molecular classification of GC.

Accumulating evidence has reported the functional difference between CDK12 and CDK13 (Quereda et al., 2019; Tadesse et al., 2021). *CDK12* was not pulled down in the ChIP experiment in our study, enlightening the thought that CDK13 indeed shares some molecular function with CDK12, but CDK13 might have its own specific role and its unique mechanism, and CDK13 could be studied unbinding to CDK12. More research studies are needed to map the independent role of CDK13, which is the main task in our next study.

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: GSE174442.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Inner Mongolia People's Hospital.

AUTHOR CONTRIBUTIONS

LY and ZW conceived and designed the study and drafted the manuscript. MW, FL, and WG performed cell and animal experiments. FW, ZF, JY, and RS performed the clinical sample collection and pathology experiments. JJ, XH, and WJ performed the statistical analysis. All authors read and critically revised the manuscript for intellectual content and approved the final manuscript.

FUNDING

This work was supported by the Science and Technology Planning Project of Inner Mongolia Science and Technology Department (Grant No. 201802153), Natural Science Foundation of Inner Mongolia (CN) (Grant Nos. 2018MS08060 and 2020MS08157), and National Natural Science Foundation of China (Grant No. 81960449). This work was also supported by the Talent Training Plan for the Key Laboratory of Inner Mongolia Science and Technology Department.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.707295/full#supplementary-material

REFERENCES

Álvarez-Fernández, M., and Malumbres, M. (2020). Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition. *Cancer cell* 37 (4), 514–529. doi:10.1016/j.ccell.2020.03.010

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., et al. (2011). The Cyclin K/Cdk12 Complex Maintains Genomic Stability *via* Regulation of Expression of DNA Damage Response Genes. *Genes Develop.* 25 (20), 2158–2172. doi:10.1101/gad.16962311

Bostwick, B. L., McLean, S., McLean, S., Posey, J. E., Streff, H. E., Gripp, K. W., et al. (2017). Phenotypic and Molecular Characterisation of CDK13-Related Congenital Heart Defects, Dysmorphic Facial Features and Intellectual Developmental Disorders. *Genome Med.* 9 (1), 73. doi:10.1186/s13073-017-0463-8

Cleynen, I., and Van de Ven, W. J. (2008). The HMGA Proteins: a Myriad of Functions (Review). *Int. J. Oncol.* 32 (2), 289–305. doi:10.3892/ijo.32.2.289

Detre, S., Saclani Jotti, G., and Dowsett, M. (1995). A "quickscore" Method for Immunohistochemical Semiquantitation: Validation for Oestrogen Receptor in Breast Carcinomas. *J. Clin. Pathol.* 48 (9), 876–878. doi:10.1136/jcp.48.9.876

Ding, L., Cao, J., Lin, W., Chen, H., Xiong, X., Ao, H., et al. (2020). The Roles of Cyclin-dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. *Ijms* 21 (6), 1960. doi:10.3390/ijms21061960

Dong, J., Wang, R., Ren, G., Li, X., Wang, J., Sun, Y., et al. (2017). HMGA2-FOXL2 Axis Regulates Metastases and Epithelial-To-Mesenchymal Transition of Chemoresistant Gastric Cancer. *Clin. Cancer Res.* 23 (13), 3461–3473. doi:10.1158/1078-0432.CCR-16-2180

Dong, X., Chen, G., Cai, Z., Li, Z., Qiu, L., Xu, H., et al. (2018). CDK13 RNA Over-editing Mediated by ADAR1 Associates with Poor Prognosis of Hepatocellular Carcinoma Patients. *Cell Physiol Biochem* 47 (6), 2602–2612. doi:10.1159/000491656

Dubbury, S. J., Boutz, P. L., and Sharp, P. A. (2018). CDK12 Regulates DNA Repair Genes by Suppressing Intronic Polyadenylation. *Nature* 564 (7734), 141–145. doi:10.1038/s41586-018-0758-y

Fan, Z., Devlin, J. R., Hogg, S. J., Doyle, M. A., Harrison, P. F., Todorovski, I., et al. (2020). CDK13 Cooperates with CDK12 to Control Global RNA Polymerase II Processivity. *Sci. Adv.* 6 (18), eaaz5041. doi:10.1126/sciadv.aaz5041

Goel, S., and Tolaney, S. M. (2019). CDK4/6 Inhibitors in Breast Cancer: a Role in Triple-Negative Disease? *Lancet Oncol.* 20 (11), 1479–1481. doi:10.1016/s1470-2045(19)30627-8

Greifenberg, A. K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C. A., et al. (2016). Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep.* 14 (2), 320–331. doi:10.1016/j.celrep.2015.12.025

Hamilton, M. J., and Suri, M. (2019). CDK13-related Disorder. *Adv. Genet.* 103, 163–182. doi:10.1016/bs.adgen.2018.11.001

Hammond, S. M., and Sharpless, N. E. (2008). HMGA2, microRNAs, and Stem Cell Aging. *Cell* 135 (6), 1013–1016. doi:10.1016/j.cell.2008.11.026

Harbour, J. W., Luo, R. X., Santi, A. D., Postigo, A. A., and Dean, D. C. (1999). Cdk Phosphorylation Triggers Sequential Intramolecular Interactions that Progressively Block Rb Functions as Cells Move through G1. *Cell* 98 (6), 859–869. doi:10.1016/s0092-8674(00)81519-6

Hombach-Klonisch, S., Natarajan, S., Thanasupawat, T., Medapati, M., Pathak, A., Ghavami, S., et al. (2014). Mechanisms of Therapeutic Resistance in Cancer (Stem) Cells with Emphasis on Thyroid Cancer Cells. *Front. Endocrinol.* 5, 37. doi:10.3389/fendo.2014.00037

Hopkins, J. L., and Zou, L. (2019). Induction of BRCAness in Triple-Negative Breast Cancer by a CDK12/13 Inhibitor Improves Chemotherapy. *Cancer cell* 36 (5), 461–463. doi:10.1016/j.ccell.2019.10.012

Hortobagyi, G. N., Stemmer, S. M., Burris, H. A., Yap, Y. S., Sonke, G. S., Paluch-Shimon, S., et al. (2016). Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N. Engl. J. Med.* 375 (18), 1738–1748. doi:10.1056/NEJMoa1609709

Huang, B., Yang, J., Cheng, Q., Xu, P., Wang, J., Zhang, Z., et al. (2018). Prognostic Value of HMGA2 in Human Cancers: A Meta-Analysis Based on Literatures and TCGA Datasets. *Front. Physiol.* 9, 776. doi:10.3389/fphys.2018.00776

Ko, T. K., Kelly, E., and Pines, J. (2001). CrkRS. *J. Cell Sci.* 114 (14), 2591–2603. doi:10.1242/jcs.114.14.2591

Lapidot-Lifson, Y., Patinkin, D., Prody, C. A., Ehrlich, G., Seidman, S., Ben-Aziz, R., et al. (1992). Cloning and Antisense Oligodeoxynucleotide Inhibition of a Human Homolog of Cdc2 Required in Hematopoiesis. *Proc. Natl. Acad. Sci.* 89 (2), 579–583. doi:10.1073/pnas.89.2.579

Lazarov, M., Kubo, Y., Cai, T., Dajee, M., Tarutani, M., Lin, Q., et al. (2002). CDK4 Coexpression with Ras Generates Malignant Human Epidermal Tumorigenesis. *Nat. Med.* 8 (10), 1105–1114. doi:10.1038/nm779

Li, W., Wang, Z., Zha, L., Kong, D., Liao, G., and Li, H. (2017). HMGA2 Regulates Epithelial-Mesenchymal Transition and the Acquisition of Tumor Stem Cell Properties through TWIST1 in Gastric Cancer. *Oncol. Rep.* 37 (1), 185–192. doi:10.3892/or.2016.5255

Li, Y., Qiang, W., Griffin, B. B., Gao, T., Chakravarti, D., Bulun, S., et al. (2020). HMGA2-mediated Tumorigenesis through Angiogenesis in Leiomyoma. *Fertil. sterility* 114 (5), 1085–1096. doi:10.1016/j.fertnstert.2020.05.036

Li, Y., Zhang, H., Li, Q., Zou, P., Huang, X., Wu, C., et al. (2020). CDK12/13 Inhibition Induces Immunogenic Cell Death and Enhances Anti-PD-1 Anticancer Activity in Breast Cancer. *Cancer Lett.* 495, 12–21. doi:10.1016/j.canlet.2020.09.011

Liang, S., Hu, L., Wu, Z., Chen, Z., Liu, S., Xu, X., et al. (2020). CDK12: A Potent Target and Biomarker for Human Cancer Therapy. *Cells* 9 (6), 9. doi:10.3390/cells9061483

Malek, A., Bakhidze, E., Noske, A., Sers, C., Aigner, A., Schäfer, R., et al. (2008). HMGA2 Gene Is a Promising Target for Ovarian Cancer Silencing Therapy. *Int. J. Cancer* 123 (2), 348–356. doi:10.1002/ijc.23491

Malumbres, M., Sotillo, R., Santamarı´a, D., Galán, J., Cerezo, A., Ortega, S., et al. (2004). Mammalian Cells Cycle without the D-type Cyclin-dependent Kinases Cdk4 and Cdk6. *Cell* 118 (4), 493–504. doi:10.1016/j.cell.2004.08.002

Mansoori, B., Mohammadi, A., Ditzel, H. J., Duijf, P. H. G., Khaze, V., Gjerstorff, M. F., et al. (2021). HMGA2 as a Critical Regulator in Cancer Development. *Genes* 12 (2), 269. doi:10.3390/genes12020269

Mansoori, B., Mohammadi, A., Naghizadeh, S., Gjerstorff, M., Shanehbandi, D., Shirjang, S., et al. (2020). miR-330 Suppresses EMT and Induces Apoptosis by Downregulating HMGA2 in Human Colorectal Cancer. *J. Cell Physiol* 235 (2), 920–931. doi:10.1002/jcp.29007

Marqués, F., Moreau, J.-L., Peaucellier, G., Lozano, J.-C., Schatt, P., Picard, A., et al. (2000). A New Subfamily of High Molecular Mass CDC2-Related Kinases with PITAI/VRE Motifs. *Biochem. Biophysical Res. Commun.* 279 (3), 832–837. doi:10.1006/bbrc.2000.4042

Min, A., Kim, J. E., Kim, Y.-J., Lim, J. M., Kim, S., Kim, J. W., et al. (2018). Cyclin E Overexpression Confers Resistance to the CDK4/6 Specific Inhibitor Palbociclib in Gastric Cancer Cells. *Cancer Lett.* 430, 123–132. doi:10.1016/j.canlet.2018.04.037

Nishino, J., Kim, I., Chada, K., and Morrison, S. J. (2008). Hmga2 Promotes Neural Stem Cell Self-Renewal in Young but Not Old Mice by Reducing p16Ink4a and p19Arf Expression. *Cell* 135 (2), 227–239. doi:10.1016/j.cell.2008.09.017

Novakova, M., Hampl, M., Vrabel, D., Prochazka, J., Petrezselyova, S., Prochazkova, M., et al. (2019). Mouse Model of Congenital Heart Defects, Dysmorphic Facial Features and Intellectual Developmental Disorders as a Result of Non-functional CDK13. *Front. Cel. Dev. Biol.* 7, 155. doi:10.3389/fcell.2019.00155

Otto, T., and Sicinski, P. (2017). Cell Cycle Proteins as Promising Targets in Cancer Therapy. *Nat. Rev. Cancer* 17 (2), 93–115. doi:10.1038/nrc.2016.138

Quereda, V., Bayle, S., Vena, F., Frydman, S. M., Monastyrskyi, A., Roush, W. R., et al. (2019). Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer cell* 36 (5), 545–558. e7. doi:10.1016/j.ccell.2019.09.004

Slamon, D. J., Neven, P., Chia, S., Fasching, P. A., De Laurentiis, M., Im, S.-A., et al. (2018). Phase III Randomized Study of Ribociclib and Fulvestrant in Hormone Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer: MONALEESA-3. *Jco* 36 (24), 2465–2472. doi:10.1200/JCO.2018.78.9909

Sledge, G. W., Jr., Toi, M., Neven, P., Sohn, J., Inoue, K., Pivot, X., et al. (2017). MONARCH 2: Abemaciclib in Combination with Fulvestrant in Women with HR+/HER2– Advanced Breast Cancer Who Had Progressed while Receiving Endocrine Therapy. *Jco* 35 (25), 2875–2884. doi:10.1200/JCO.2017.73.7585

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., and Lordick, F. (2020). Gastric Cancer. *The Lancet* 396 (10251), 635–648. doi:10.1016/S0140-6736(20)31288-5

Spring, L. M., Wander, S. A., Andre, F., Moy, B., Turner, N. C., and Bardia, A. (2020). Cyclin-dependent Kinase 4 and 6 Inhibitors for Hormone Receptor-Positive

Breast Cancer: Past, Present, and Future. *The Lancet* 395 (10226), 817–827. doi:10.1016/S0140-6736(20)30165-3

Sun, J., Sun, B., Sun, R., Zhu, D., Zhao, X., Zhang, Y., et al. (2017). HMGA2 Promotes Vasculogenic Mimicry and Tumor Aggressiveness by Upregulating Twist1 in Gastric Carcinoma. *Scientific Rep.* 7 (1), 2229. doi:10.1038/s41598-017-02494-6

Tadesse, S., Duckett, D. R., and Monastyrskyi, A. (2021). The Promise and Current Status of CDK12/13 Inhibition for the Treatment of Cancer. *Future Med. Chem.* 13 (2), 117–141. doi:10.4155/fmc-2020-0240

Tan, L., Wei, X., Zheng, L., Zeng, J., Liu, H., Yang, S., et al. (2016). Amplified HMGA2 Promotes Cell Growth by Regulating Akt Pathway in AML. *J. Cancer Res. Clin. Oncol.* 142 (2), 389–399. doi:10.1007/s00432-015-2036-9

Tripathy, D., Im, S.-A., Colleoni, M., Franke, F., Bardia, A., Harbeck, N., et al. (2018). Ribociclib Plus Endocrine Therapy for Premenopausal Women with Hormone-Receptor-Positive, Advanced Breast Cancer (MONALEESA-7): a Randomised Phase 3 Trial. *Lancet Oncol.* 19 (7), 904–915. doi:10.1016/S1470-2045(18)30292-4

Uehara, T., Takenouchi, T., Kosaki, R., Kurosawa, K., Mizuno, S., and Kosaki, K. (2018). Redefining the Phenotypic Spectrum of De Novo Heterozygous CDK13 Variants: Three Patients without Cardiac Defects. *Eur. J. Med. Genet.* 61 (5), 243–247. doi:10.1016/j.ejmg.2017.12.004

Wang, J., Zhang, Y., Lu, L., Lu, Y., Tang, Q., and Pu, J. (2019). Insight into the Molecular Mechanism of LINC00152/miR-215/CDK13 axis in Hepatocellular Carcinoma Progression. *J. Cell Biochem* 120 (11), 18816–18825. doi:10.1002/jcb.29197

Wei, C.-H., Wei, L. X., Lai, M. Y., Chen, J. Z., and Mo, X. J. (2013). Effect of Silencing of High Mobility Group A2 Gene on Gastric Cancer MKN-45 Cells. *Wjg* 19 (8), 1239–1246. doi:10.3748/wjg.v19.i8.1239

Wu, H., Zou, Q., He, H., Liang, Y., Lei, M., Zhou, Q., et al. (2019). Long Non-coding RNA PCAT6 Targets miR-204 to Modulate the Chemoresistance of Colorectal Cancer Cells to 5-fluorouracil-based Treatment through HMGA2 Signaling. *Cancer Med.* 8 (5), 2484–2495. doi:10.1002/cam4.1809

Zeng, M., Kwiatkowski, N. P., Zhang, T., Nabet, B., Xu, M., Liang, Y., et al. (2018). Targeting MYC Dependency in Ovarian Cancer through Inhibition of CDK7 and CDK12/13. *eLife* 7, 7. doi:10.7554/eLife.39030

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, K., Hong, R., Kaping, L., Xu, F., Xia, W., Qin, G., et al. (2019). CDK4/6 Inhibitor Palbociclib Enhances the Effect of Pyrotinib in HER2-Positive Breast Cancer. *Cancer Lett.* 447, 130–140. doi:10.1016/j.canlet.2019.01.005

Zhang, T., Kwiatkowski, N., Olson, C. M., Dixon-Clarke, S. E., Abraham, B. J., Greifenberg, A. K., et al. (2016). Covalent Targeting of Remote Cysteine Residues to Develop CDK12 and CDK13 Inhibitors. *Nat. Chem. Biol.* 12 (10), 876–884. doi:10.1038/nchembio.2166

Zhang, Z. C., Wang, G. P., Yin, L. M., Li, M., and Wu, L. L. (2018). Increasing miR-150 and Lowering HMGA2 Inhibit Proliferation and Cycle Progression of colon Cancer in SW480 Cells. *Eur. Rev. Med. Pharmacol. Sci.* 22 (20), 6793–6800. doi:10.26355/eurrev_201810_16147

Zhao, W., Geng, D., Li, S., Chen, Z., and Sun, M. (2018). LncRNA HOTAIR Influences Cell Growth, Migration, Invasion, and Apoptosis *via* the miR-20a-5p/HMGA2axis in Breast Cancer. *Cancer Med.* 7 (3), 842–855. doi:10.1002/cam4.1353

Zhou, X., Benson, K. F., Przybysz, K., Liu, J., Hou, Y., Cherath, L., et al. (1996). Genomic Structure and Expression of the Murine Hmgi-C Gene. *Nucleic Acids Res.* 24 (20), 4071–4077. doi:10.1093/nar/24.20.4071

Zhu, J., Wang, H., Xu, S., and Hao, Y. (2017). Clinicopathological and Prognostic Significance of HMGA2 Overexpression in Gastric Cancer: a Meta-Analysis. *Oncotarget* 8 (59), 100478–100489. doi:10.18632/oncotarget.19001

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 Wu, Wang, Li, Wang, Jia, Feng, Huo, Yang, Jin, Sa, Gao and 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.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fmolb.2021.706949

CD74 Correlated With Malignancies and Immune Microenvironment in Gliomas

Shengchao Xu[1†], Xizhe Li[2†], Lu Tang[2], Zhixiong Liu[1], Kui Yang[1] and Quan Cheng[1,3,4*]*

[1]Department of Neurosurgery, Xiangya Hospital of Central South University, Changsha, China, [2]Department of Thoracic Surgery, Xiangya Hospital of Central South University, Changsha, China, [3]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China, [4]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Hunan, China

Background: Cluster of differentiation 74 (CD74) is found to be highly involved in the development of various types of cancers and could affect the activities of infiltrated cells in the tumor microenvironment. However, these studies only focus on a few types of immune cells. Our study aims to comprehensively explore the role of CD74 in glioma prognosis and immune microenvironment.

Methods: A total of 40 glioma specimens were collected in this study. We extracted data from The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA), and Gene-Expression Omnibus (GEO) databases to explore the expression pattern of CD74 in gliomas. gene sets enrichment analysis and gene set variation analysis analyses were conducted to characterize the immune features of CD74. ESTIMATE, ssGSEA, Tumor IMmune Estimation Resource, and CIBERSORT algorithms were applied to assess the immune infiltration. Kaplan-Meier analysis was used for survival analysis. Receiver operating characteristic analysis was used to evaluate the predictive accuracy of CD74 in glioma diagnosis and prognosis.

Results: A total of 2,399 glioma patients were included in our study. CD74 was highly expressed in glioma tissue compared to normal brain tissue and its expression was significantly higher in the high-grade glioma compared to the lower grade glioma at transcriptional and translational levels. Besides, CD74 was positively associated with immune checkpoints and inflammatory cytokines as well as immune processes including cytokine secretion and leukocyte activation. The high expression of CD74 indicated a high infiltration of immune cells such as macrophages, dendritic cells, and neutrophils.

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Tathiane Malta,
University of São Paulo, Brazil
Sajib Chakraborty,
University of Dhaka, Bangladesh

***Correspondence:**
Quan Cheng
chengquan@csu.edu.cn
Kui Yang
kui.yang@csu.edu.cn

†These authors contribute equally to this manuscript.

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *08 May 2021*
Accepted: *19 August 2021*
Published: *01 September 2021*

Citation:
Xu S, Li X, Tang L, Liu Z, Yang K and Cheng Q (2021) CD74 Correlated With Malignancies and Immune Microenvironment in Gliomas. Front. Mol. Biosci. 8:706949. doi: 10.3389/fmolb.2021.706949

Abbreviations: CD74, cluster of differentiation 74; IDH, isocitrate dehydrogenase; MGMT, O6-methylguanine-DNA methyltransferase; HLA, human lymphocyte antigen; MIF, macrophage migration inhibitory factor; TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas; GEO, Gene-Expression Omnibus; GTEx, Genotype-Tissue Expression; TPM, transcripts per kilobase million; IHC, immunohistochemistry; DAB, diaminobenzidine tetrahydrochloride; GSVA, gene set variation analysis; GSEA, gene sets enrichment analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; ssGSEA, single-sample gene-set enrichment analysis; TIMER, Tumor IMmune Estimation Resource; OS, overall survival; ROC, receiver operating characteristic, LGG, lower-grade glioma; TIM-3, T cell immunoglobulin domain and mucin domain-3; AUC, area under the curve.

Moreover, patients with high expression of CD74 had poor prognoses. CD74 had moderate predictive accuracy in the diagnosis of glioblastoma and prediction of survival.

Conclusions: In conclusion, our study revealed that the high expression of CD74 was associated with poor prognosis and high immune infiltration. CD74 could be used as a potential target for glioma treatment and as a biomarker to predict the prognosis of glioma patients.

Keywords: CD74, glioma, prognosis, diagnosis, immune microenvironment

INTRODUCTION

Gliomas are primary tumors derived from the glial cells in the central nervous system, which comprise about 80% of malignant brain tumors (Goodenberger and Jenkins, 2012). According to the World Health Organization classification, gliomas are categorized into four grades: grade I, II, III, and IV (Louis et al., 2016). Although great progress has been achieved to develop novel strategies for the treatment of cancers, the prognosis of glioma patients remains unsatisfactory. The median survival time of patients with grade II, III gliomas was 11.6 and 3 years, respectively, whereas that grade IV glioma is about 15 months (Ohgaki and Kleihues, 2005; Bleeker et al., 2012; Smoll et al., 2012; Thakkar et al., 2014). Therefore, there is clear urgent to develop novel approaches for the treatment of gliomas.

Several biomarkers have been found to predict the prognoses of glioma patients with potent value. Previous studies revealed that patients with isocitrate dehydrogenase (IDH) mutant or 1p19q codeletion glioma had relatively favorable prognoses (Molenaar et al., 2014a; Molenaar et al., 2014b; Chen et al., 2019). Moreover, patients with methylated O6-methylguanine-DNA methyltransferase (MGMT) promoter gliomas benefit much more from temozolomide and radiotherapy (Hegi et al., 2005). Since the prognosis of glioma patients remains poor, more biomarkers are needed to predict the survival of glioma patients.

The cluster of differentiation 74 (CD74) is a polypeptide as the invariant chain of human lymphocyte antigen (HLA) class II. It is highly involved in the antigen presentation and the activation of $CD4^+$ T cells (Cresswell, 1994). CD74 is the cell surface membrane receptor for the cytokine macrophage migration inhibitory factor (MIF) (Farr et al., 2020). The role of the interaction between CD74 and MIF has been revealed in different types of cancers (Binsky et al., 2007; Nagata et al., 2009). The inhibition of CD74 and MIF could significantly attenuate the tumor growth of prostate cancer cells and melanoma mice model (Meyer-Siegler et al., 2006; Tanese et al., 2015). In the tumor microenvironment, the inhibition of the CD74-MIF signaling pathway could restore the antitumor activity of macrophage and dendritic cells against melanoma (Figueiredo et al., 2018).

Previous studies indicated that CD74 was highly expressed in high-grade gliomas, and it was associated with the microenvironment of glioma and could facilitate the proliferation of glioma cells (Zeiner et al., 2015; Ghoochani et al., 2016; Alban et al., 2020). However, these studies only explore the association between CD74 and different subtypes of macrophages, whereas no study has comprehensively characterized the prognostic and immune features in gliomas. Therefore, our study extracted 2,399 glioma samples from The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA), and Gene-Expression Omnibus (GEO) databases to characterize the association between CD74 and glioma prognosis as well as immune microenvironment via the application of various algorithms. Through the analysis based on a large number of samples, our study would provide a novel insight into the immune characteristics of CD74 in gliomas and reveal a potential therapeutic target as well as a biomarker for gliomas.

METHODS

Data Extraction

RNA expression of 2,399 glioma patients with corresponding clinical data was extracted from TCGA (https://portal.gdc.cancer.gov/), CGGA (http://www.cgga.org.cn/), and GEO (https://www.ncbi.nlm.nih.gov/geo/) databases. The TCGA-LGG, TCGA-GBM, CGGA325, CGGA693, CGGA301, and GSE108474 datasets were included in our study. TCGA-LGG and TCGA-GBM datasets were combined and defined as the TCGA dataset, and CGGA325 and CGGA693 datasets were combined and defined as the CGGA dataset for further analysis. There were 672, 1,013, 300, and 414 samples in the TCGA, CGGA, CGGA301, and GSE108474 datasets, respectively (**Table 1**). The batch effect was evaluated and addressed using "SVA" R package. RNA expression of normal brain tissue was obtained from Genotype-Tissue Expression (GTEx) database, which was analyzed using GEPIA online database (Tang et al., 2017). The TCGA and CGGA were RNA-seq datasets whereas CGGA301 and GSE108474 were microarray datasets. The TCGA (FPKM) dataset was downloaded format using "TCGAbiolinks" R package. CGGA (FPKM) and CGGA301 datasets were obtained from the official website. The molecular subtype of gliomas was downloaded from TCGA and CGGA databases based on definition proposed by Verhaak et al. (2010). All expression values were transformed into $\log_2$ (transcripts per kilobase million (TPM)+1). The characteristics of glioma patients included in our study were summarized in **Table 1**. The gene list of immune checkpoints and inflammatory cytokines was selected from previous studies (Wang et al., 2019; Xu et al., 2020a).

TABLE 1 | Characteristics of patients included in our study.

	TCGA (*n* = 672)	CGGA (*n* = 1,013)	CGGA301 (*n* = 300)	GSE108474 (*n* = 414)
Age				
≤41	290	469	144	—
>41	382	543	154	—
NA	0	1	2	—
Gender				
Male	387	597	179	—
Female	285	416	121	—
Subtype				
Classical	83	—	23	—
Mesenchymal	98	—	111	—
Proneural	242	—	86	—
Neural	112	—	80	—
NA	137	—	0	—
Grade				
I	0	0	0	2
II	256	290	116	86
III	265	333	57	82
IV	150	385	124	124
NA	1	5	3	120
IDH status				
Mutant	434	528	133	—
Wildtype	228	434	165	—
NA	10	51	2	—
MGMT				
Methylated	478	469	98	—
Unmethylated	157	374	187	—
NA	37	170	15	—
1p/19q status				
Codel	171	212	16	—
Non-codel	497	728	76	—
NA	4	73	208	—

IDH: isocitrate dehydrogenase; MGMT: O-6-methylguanine-DNA methyltransferase.

Glioma Specimens and Immunohistochemistry

A total of 40 glioma specimens including 21 grade II, 8 grade III, and 11 grade IV gliomas were collected to detect the protein level of CD74, which was revealed by immunohistochemistry (IHC). All the tumors were primary glioma and samples were collected when patients first received surgery without previous chemotherapy or radiotherapy. Our study was approved by the Ethics Committee of Xiangya Hospital, Central South University with written informed consent obtained. The specimens were embedded in paraffin sections. Citrate buffer (pH = 6.0) was applied for antigen retrieval. Then 0.3% H_2O_2 and 5% BSA were used for the blockade. The CD74 antibody (A9149, ABclonal, 1:100) was applied at 4 centigrade overnight. After the application of secondary antibody, diaminobenzidine tetrahydrochloride (DAB) and hematoxylin were used for staining. The score of CD74 expression was calculated by intensity score * quantity score. As for intensity scores, 0, 1, 2, and three represented negative, weak, moderate, and strong, respectively. The quantity score was determined by the proportion of stained cells, in which 0, 1, 2, 3, and four represented <10%, 10–25%, 25–50%, 50–75%, >75%, respectively. For different scores, 1–4, 5-8, and 9–12 indicated "+", "++", and "+++", respectively.

Bioinformatic Analyses

The expression of CD74 in glioma and normal tissue was compared on the GEPIA website, which contained gene expression data of TCGA and GTEx databases (Tang et al., 2017). The gene set variation analysis (GSVA) enrichment analysis was performed to show the biological processes using the "GSVA" R package (Hänzelmann et al., 2013). The co-expressed genes were screened based on the "spearman" method. Genes with $|r|>0.5$ and $p < 0.001$ were selected. The gene sets enrichment analysis (GSEA) enrichment analysis was conducted using "clusterProfiler" R package in Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. During the GSEA enrichment analysis, the value of "r" was defined as the $\log_2$ (fold change) and the *p*-value was defined as the false discovery rate (FDR). The infiltration of stromal and immune cells and tumor purity were assessed by Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm using "estimate" R package (Yoshihara et al., 2013). The estimation of immune cell infiltration was conducted by the single-sample gene-set enrichment analysis (ssGSEA), Tumor IMmune Estimation Resource (TIMER), and CIBERSORT algorithms, in which ssGSEA contained 28 immune cells, the TIMER contained six immune cells, and CIBERSORT contained 22 immune cells (Barbie et al., 2009; Newman et al., 2015; Li et al., 2020). A transcription

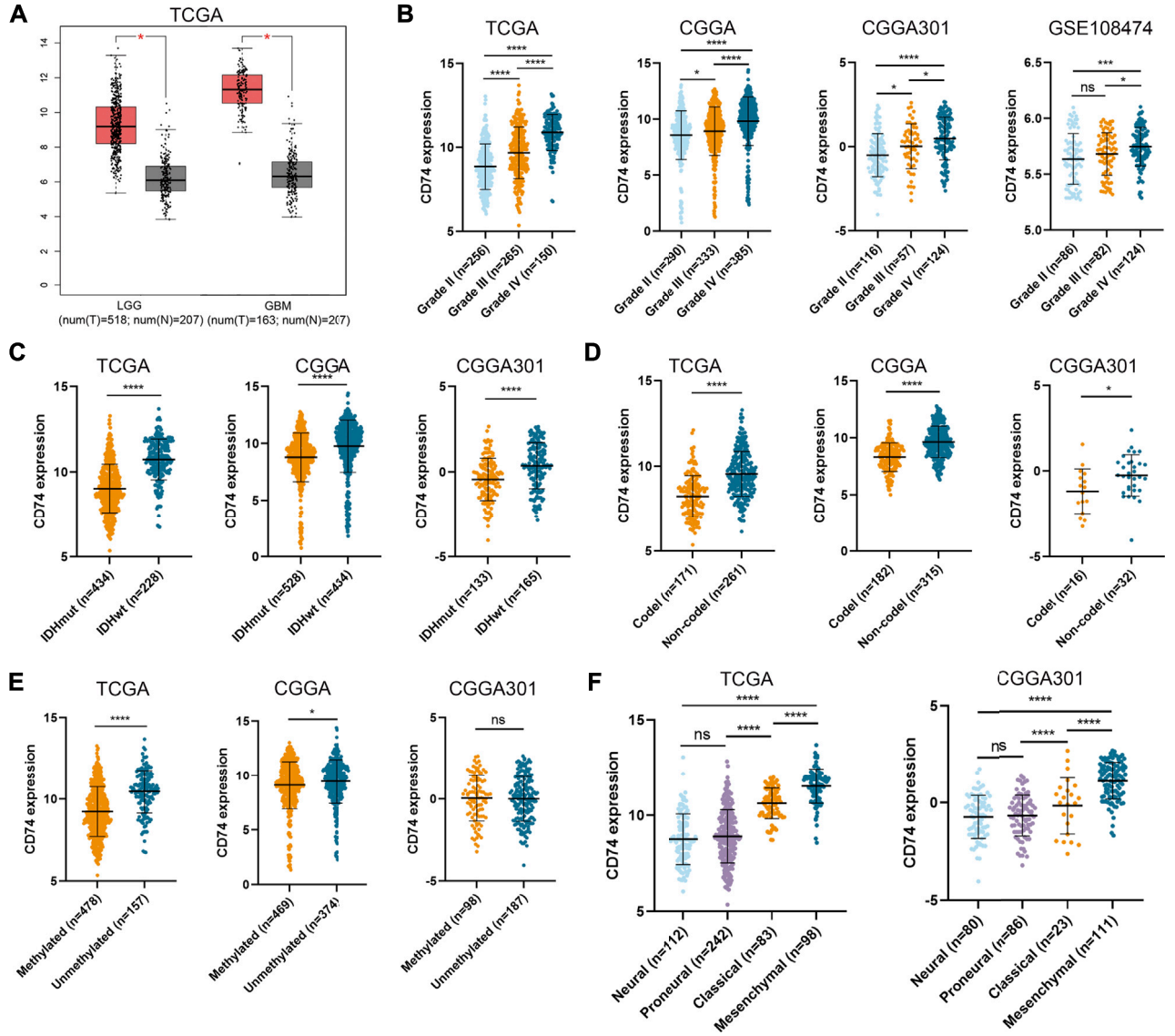


FIGURE 1 | Expression of CD74 was associated with glioma malignancies. **(A)** Expression of CD74 in glioma and normal brain tissues. The red and grey box represented tumor and normal tissues, respectively. B-E. CD74 expression in different grades **(B)**, IDH status **(C)**, 1p19q status **(D)**, and MGMT status **(E)** of gliomas. F. Expression of CD74 in different molecular subtypes of glioma. G Immunohistochemistry staining of CD74 in glioma and normal brain tissues. Data were represented as mean ± standard deviation by dot plot. IDH: isocitrate dehydrogenase; MGMT: O6-methylguanine-DNA methyltransferase. *, $p < 0.05$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, no significance.

factor (TF) - target gene interaction mapping analysis was conducted using ARACNE algorithm (Margolin et al., 2006).

Statistical Analysis

All data were analyzed and visualized using R version 3.6.0 (https://www.r-project.org/) and GraphPad Prism version 8.0.1. Shapiro-Wilk test was used to assess the normality of data. The comparison of the difference between the two groups was conducted using Student's t-test and Wilcoxon test. The one-way ANOVA test and Kruskal-Wallis test were used to compare the difference between three or more groups. Kaplan-Meier analysis was used to compare the survival time of patients in two groups. Multivariate Cox analysis was conducted to evaluate the prognostic value of CD74 in the overall survival (OS) of glioma patients in consideration of various clinical features. Receiver operating characteristic (ROC) analysis was conducted to evaluate the predictive accuracy of CD74 in the classification of glioblastoma (GBM). Time-dependent ROC analysis was adopted to assess the predictive accuracy of CD74 in the survival of glioma patients. The correlation between CD74 and other gene expressions was determined by the "spearman" method. The Two-sided p-value < 0.05 was considered statistically significant.

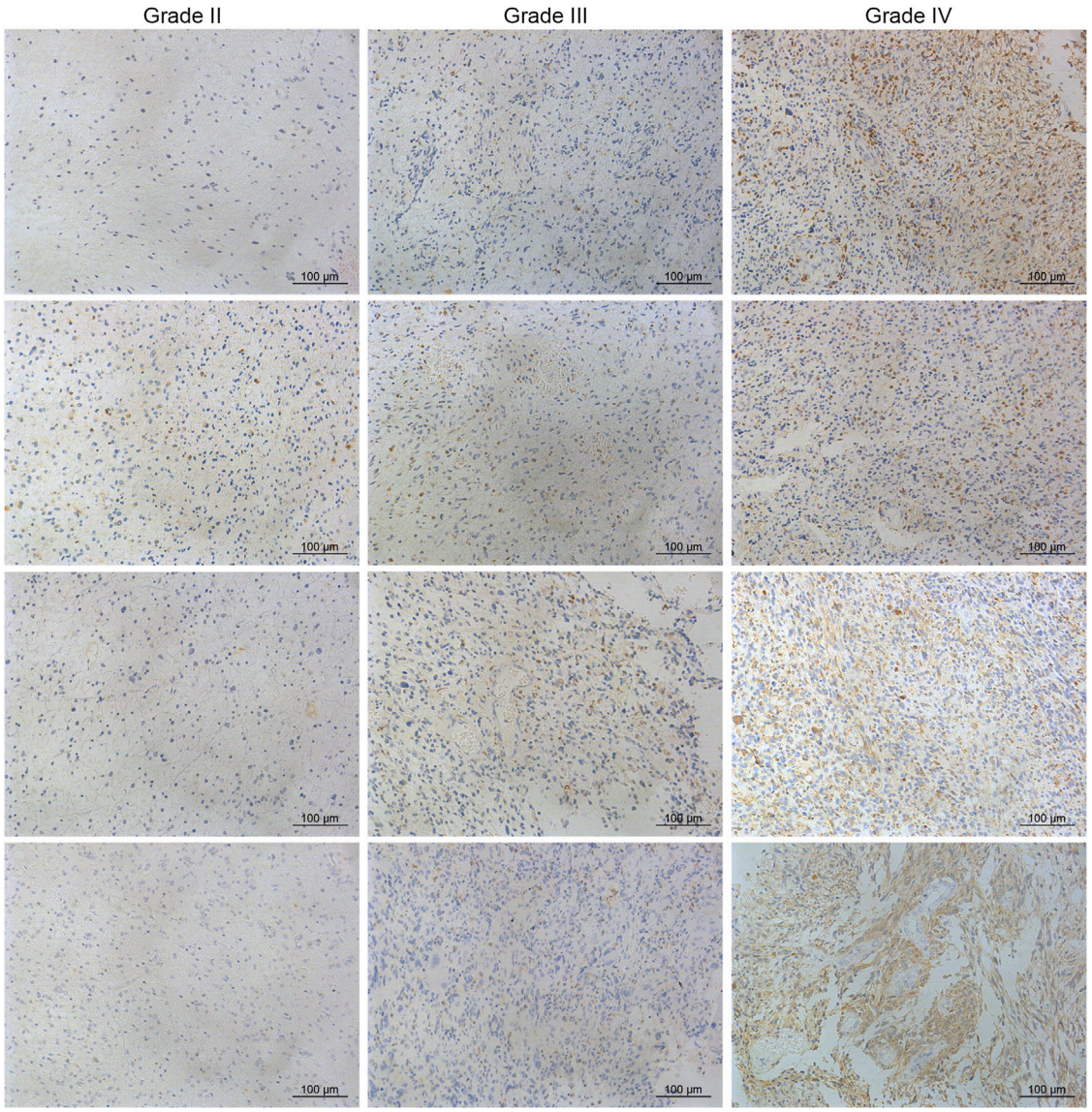


FIGURE 2 | Immunohistochemistry of CD74 in different grades of gliomas.

TABLE 2 | IHC score of CD74 in different grades of gliomas.

	+	++	+++	*p* Value
Grade II	14	5	2	<0.001
Grade III	1	5	2	
Grade IV	2	1	8	

RESULTS

High Expression of CD74 Was Associated With Glioma Malignancies

To characterize the role of CD74 in gliomas, we explored the expression pattern of CD74 in gliomas and normal tissue. In both lower-grade glioma (LGG) and GBM, the expression of CD74 was significantly higher compared to the normal tissue ($p < 0.05$) (**Figure 1A**). Shapiro-Wilk test indicated that the expression of CD74 in each dataset did not all follow normal distribution. Therefore, Kruskal-Wallis test revealed that the expression of CD74 was significantly elevated in grade IV glioma compared with grade II and III gliomas in the TCGA, CGGA, CGGA301, and GSE108474 datasets ($p < 0.05$) (**Figure 1B**). In grade II and III gliomas, CD74 had a distinct expression pattern in TCGA, CGGA, and CGGA301 datasets but not in GSE108474 dataset (**Figure 1B**). In different subtypes of glioma, the expression of CD74 was significantly higher in the IDH wildtype and 1p19q non-codeletion gliomas ($p < 0.05$) (**Figures 1C,D**). As for MGMT methylation, CD74 was highly expressed in unmethylated gliomas than methylated gliomas in TCGA and CGGA datasets ($p < 0.05$), whereas this trend was not detected in the CGGA301 dataset ($p >$

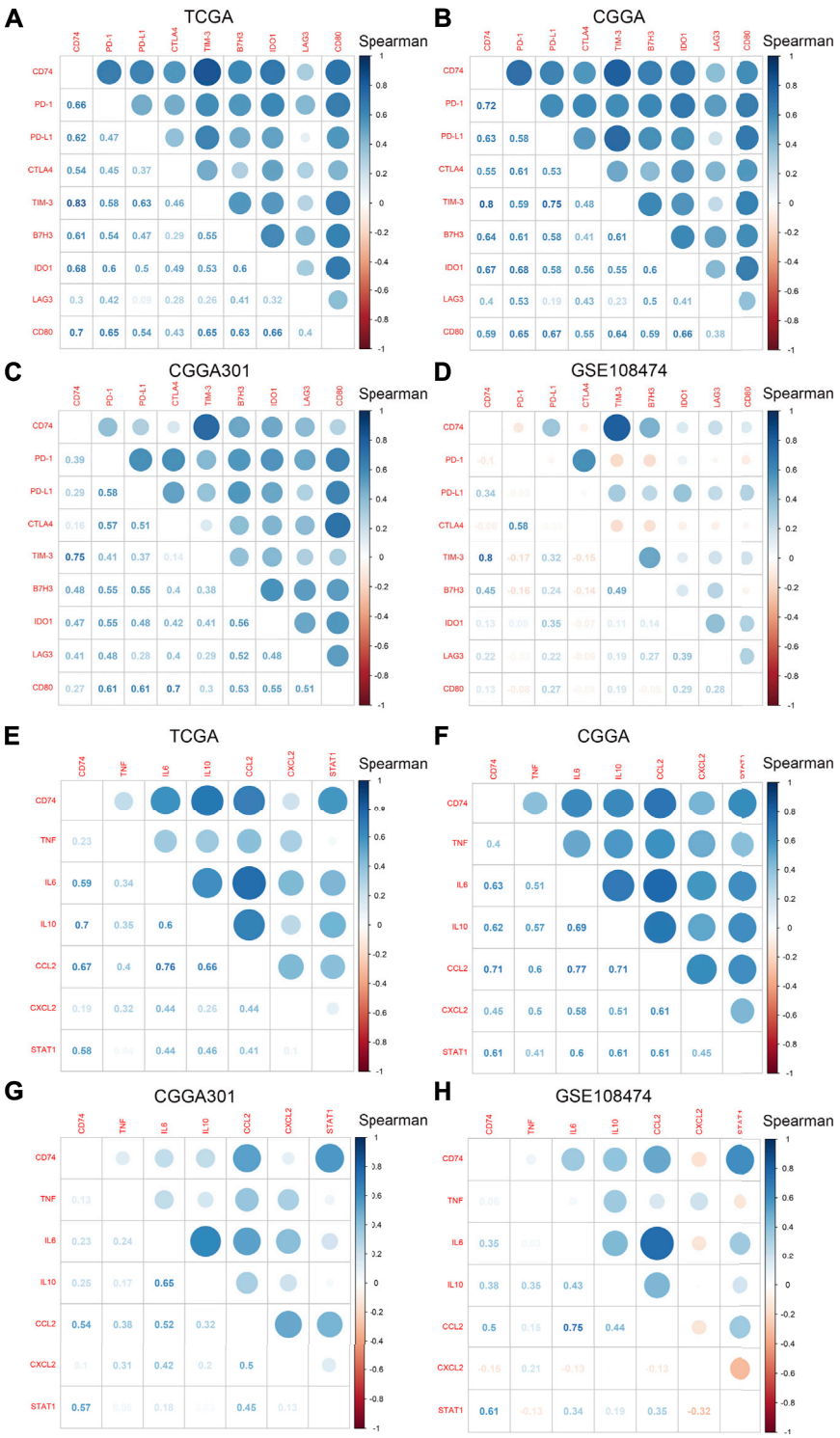


FIGURE 3 | Correlation between CD74 and immune checkpoints and inflammatory cytokines. A-D. Correlation between CD74 and immune checkpoints in TCGA **(A)**, CGGA **(B)**, CGGA301 **(C)**, and GSE108474 **(D)** datasets. E-H. Correlation between CD74 and inflammatory cytokines in TCGA **(E)**, CGGA **(F)**, CGGA301 **(G)**, and GSE108474 **(H)** datasets. Blue and red represented positive and negative correlation, respectively.

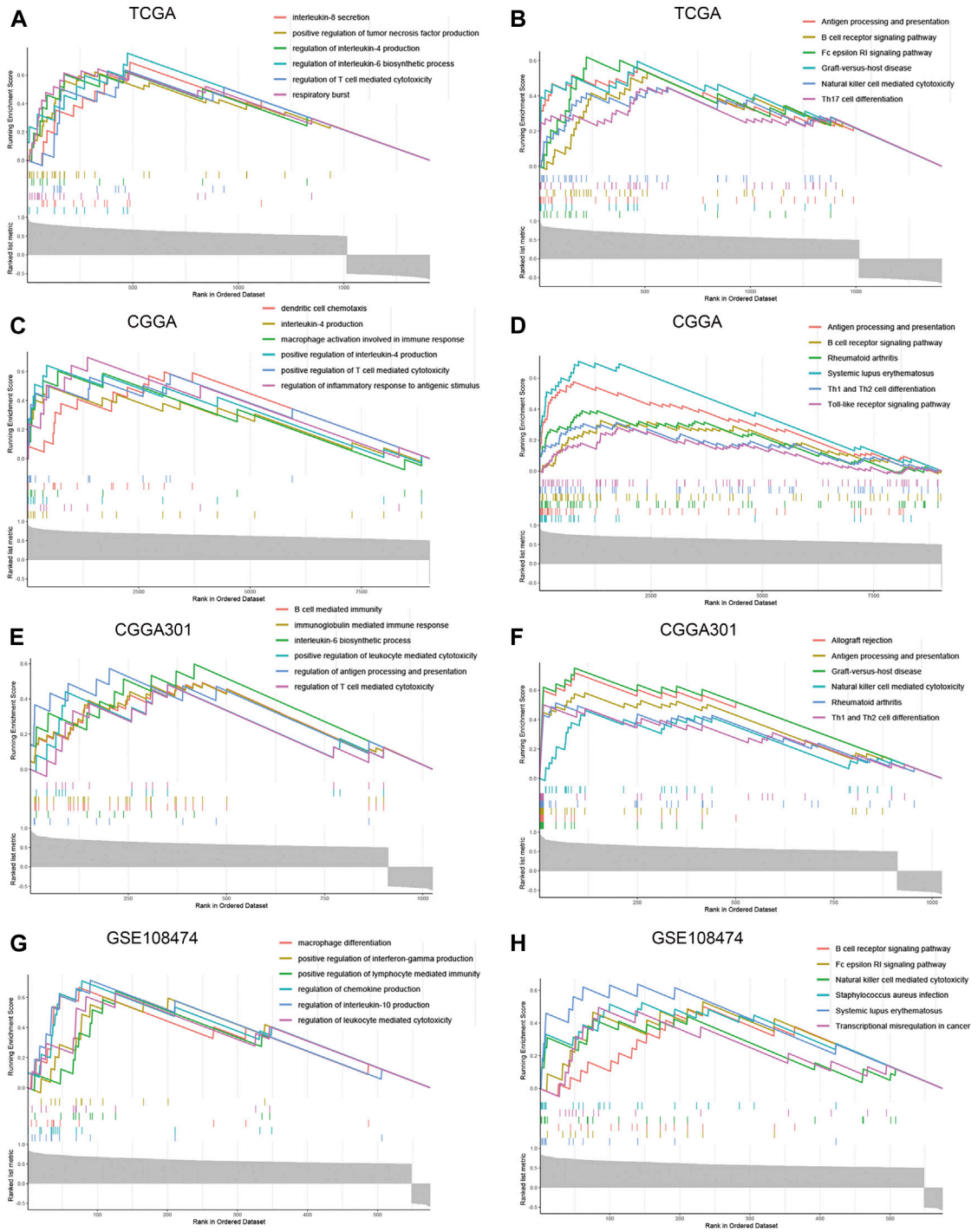


FIGURE 4 | Enrichment analysis of co-expressed genes of CD74. A-H. GSEA enrichment analysis of co-expressed genes of CD74 in GO and KEGG terms in the TCGA **(A, B)**, CGGA **(C, D)**, CGGA301 **(E, F)**, and GSE108474 **(G, H)** datasets. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

0.05) (**Figure 1E**). Besides, in different molecular subtypes of gliomas, the expression of CD74 had no significant difference between neural and proneural gliomas ($p > 0.05$) but was significantly increased in the mesenchymal subtype compared with other subtypes of glioma ($p < 0.05$) (**Figure 1F**). Since IDH mutation, MGMT methylation, and 1p19q codeletion indicated high malignancies and poor prognosis, the expression of CD74 might be associated with glioma malignancies. Apart from the RNA level, we also investigated the expression of CD74 at the protein level in different grades of gliomas. The IHC showed that

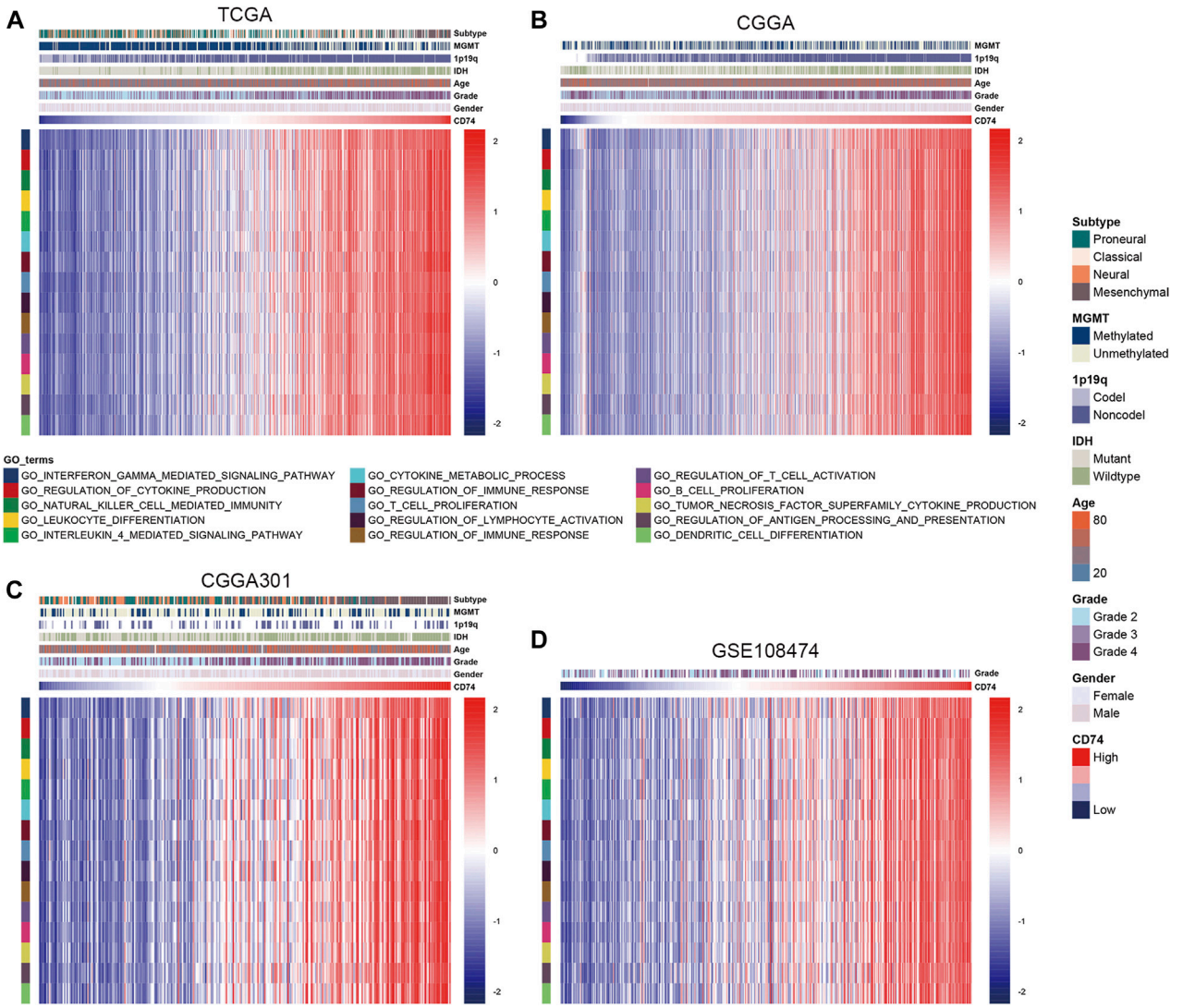


FIGURE 5 | Association between CD74 expression and immune processes. A-D. GSVA analyses revealed the association between the enrichment of immune processes and CD74 expression in the TCGA **(A)**, CGGA **(B)**, CGGA301 **(C)**, and GSE108474 **(D)** datasets. Red and blue represented relatively high and low involvement in each pathway.

the expression of CD74 was higher in the higher grade of glioma and was highest in grade IV glioma specimens (**Figure 2**; **Table 2**). These results indicated that CD74 was associated with glioma malignancies and might be used as a potential biomarker.

Expression of CD74 Correlated With Immune-Related Genes and Immune Processes

To further explore the characteristics of CD74, we assessed the correlation between CD74 and immune-related genes such as immune checkpoints and inflammatory cytokines. The gene list of immune checkpoints and inflammatory cytokines was selected from previous studies (Wang et al., 2019; Xu et al., 2020a). Results showed that CD74 was positively associated with the expression of immune checkpoints such as PD-1, PD-L1, IDO1, and B7H3 in TCGA, CGGA, and CGGA301 datasets (**Figures 3A–C**). The expression of CD74 was notably associated with the expression of T cell immunoglobulin domain and mucin domain-3 (TIM-3), which could lead to the suppression of T cells (**Figure 3D**) (Blackburn et al., 2009). Moreover, CD74 was positively correlated with the expression of inflammatory cytokines such as IL-6, IL-10, and CCL2 in the TCGA, CGGA, CGGA301, and GSE108474 datasets (**Figures 3E–H**). Then, we screened the co-expressed genes of CD74 with the cutoff point of $|r|>0.5$ and $p < 0.001$ to further characterize the role of CD74. Enrichment analysis revealed that the co-expressed genes were enriched in cytokine secretion, immune cell mediated cytotoxicity, and

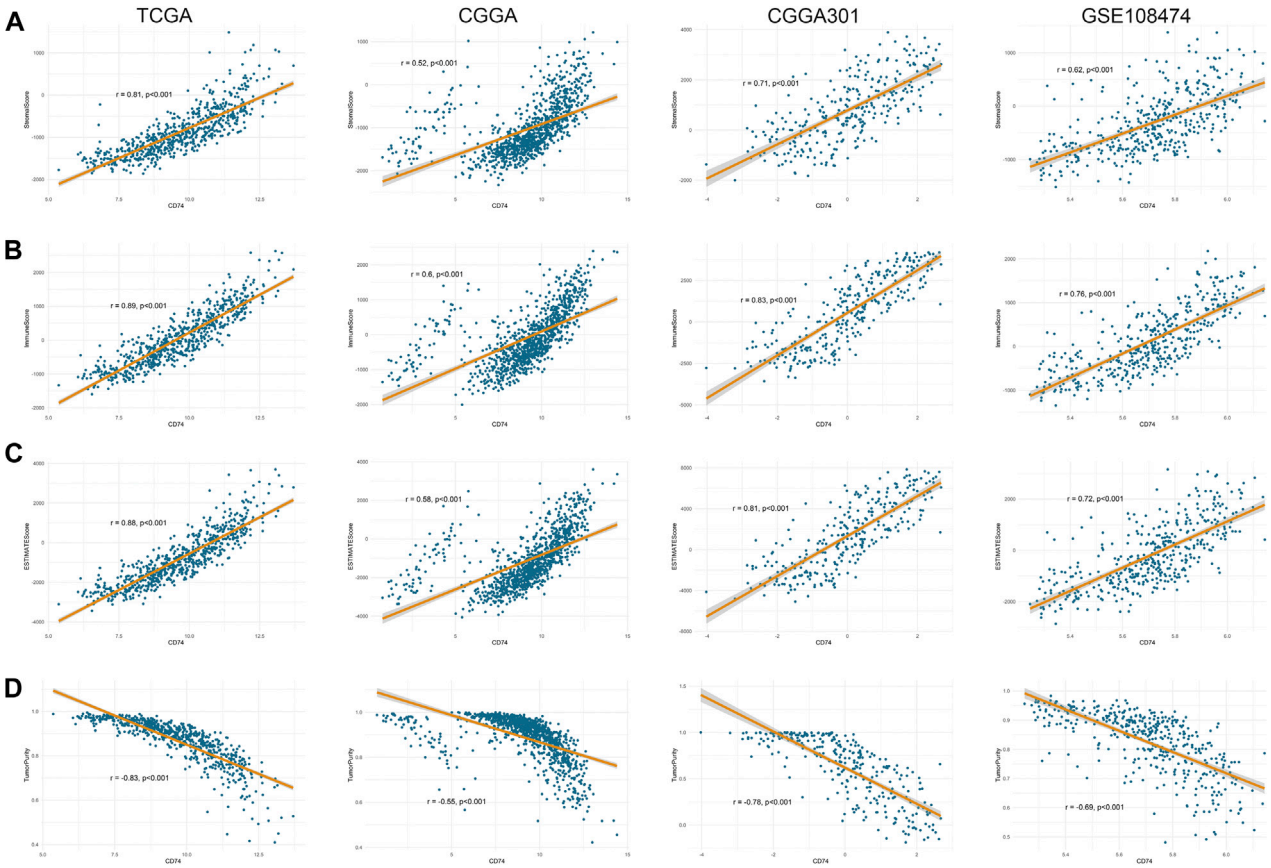


FIGURE 6 | Correlation between CD74 expression and tumor purity. A-D. Tumor purity was assessed by ESTIMATE algorithm. The correlation between CD74 expression and stromal score **(A)**, immune score **(B)**, ESTIMATE score **(C)**, and tumor purity **(D)** in the four datasets.

inflammatory response to antigenic stimulation in GO biological processes; as for in KEGG pathway, they were highly involved in antigen processing and presentation, Th1 and Th2 cell differentiation, natural killer cell mediated cytotoxicity, and autoimmune diseases such as rheumatoid arthritis and systemic lupus erythematosus (**Figures 4A–H**). Meanwhile, ARACNE algorithm was conducted to screening co-expressed genes of CD74 that were controlled by the distinct transcriptional factors, in which immune-associated genes such as IL6R, IFI16, and IFIT1B were included (**Supplementary Table S1**) (**Supplementary Figure S1**). Multivariate Cox analysis revealed that genes such as PILRA, LILRB2, and BATF were independent prognostic factors for glioma patients (**Supplementary Table S2**). Additionally, we conducted GSVA analysis to explore the association between CD74 expression and the enrichment of immune processes, which indicated that the higher expression of CD74 indicated a higher enrichment of cytokine production, lymphocyte activation, interferon related signaling pathway, and leukocyte differentiation (**Figures 5A–D**). These results demonstrated that CD74 was closely associated with immune-related genes and immune processes, indicating the immune characteristics of CD74 in gliomas.

CD74 expression indicated low tumor purity and high infiltration of immune cells in the glioma microenvironment.

Further, we employed the ESTIMATE, ssGSEA, and TIMER algorithms to further characterize the immune features of CD74. The stromal and immune scores of each sample were calculated based on their gene expression pattern. The expression of CD74 was significantly positively associated with stromal and immune scores of gliomas ($p < 0.05$) (**Figures 6A,B**). The ESTIMATE score was calculated as the sum of the stromal and immune scores, which was also positively correlated with the expression of CD74 ($p < 0.05$) (**Figure 6C**). The tumor purity was calculated based on the ESTIMATE score. CD74 was highly expressed in the low-purity tumor, which indicated a relatively high proportion of stromal and immune cells ($p < 0.05$) (**Figure 6D**). Besides, the higher expression of CD74 indicated a higher abundance of infiltrated immune cells according to the ssGSEA algorithm (**Figures 7A–D**). In CD74 high group, the abundance of macrophages, activated dendritic cells, and neutrophils was significantly increased than CD74 low group ($p < 0.05$) (**Supplementary Figure S2**). In the TIMER algorithm, the abundance of macrophage, dendritic cell, and neutrophil was markedly correlated with CD74 expression ($p < 0.05$) (**Figures**

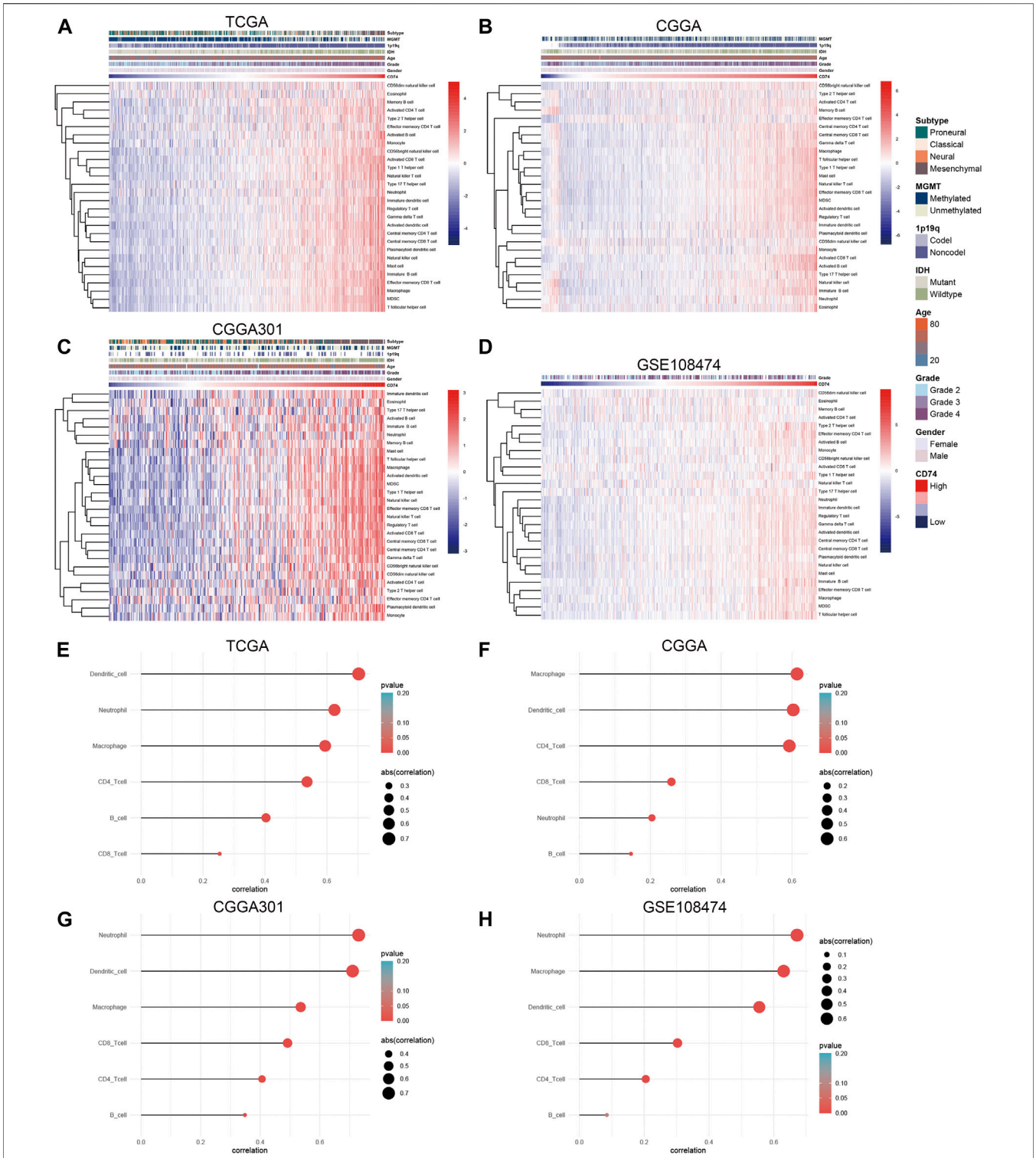


FIGURE 7 | Correlation between CD74 expression and immune infiltration. **(A–D)** The correlation between CD74 expression and immune infiltration estimated by the ssGSEA algorithm. **(E–H)** The correlation between CD74 expression and immune infiltration estimated by the TIMER algorithm.

7E–H). These findings were further verified by the CIBERSORT algorithm, which indicated a higher abundance of activated dendritic cells, macrophages and neutrophils ($p < 0.05$) (**Supplementary Figure S3**). These results suggested that CD74 correlated with the infiltration of immune cells such as macrophages, dendritic cells, and neutrophils in the glioma microenvironment.

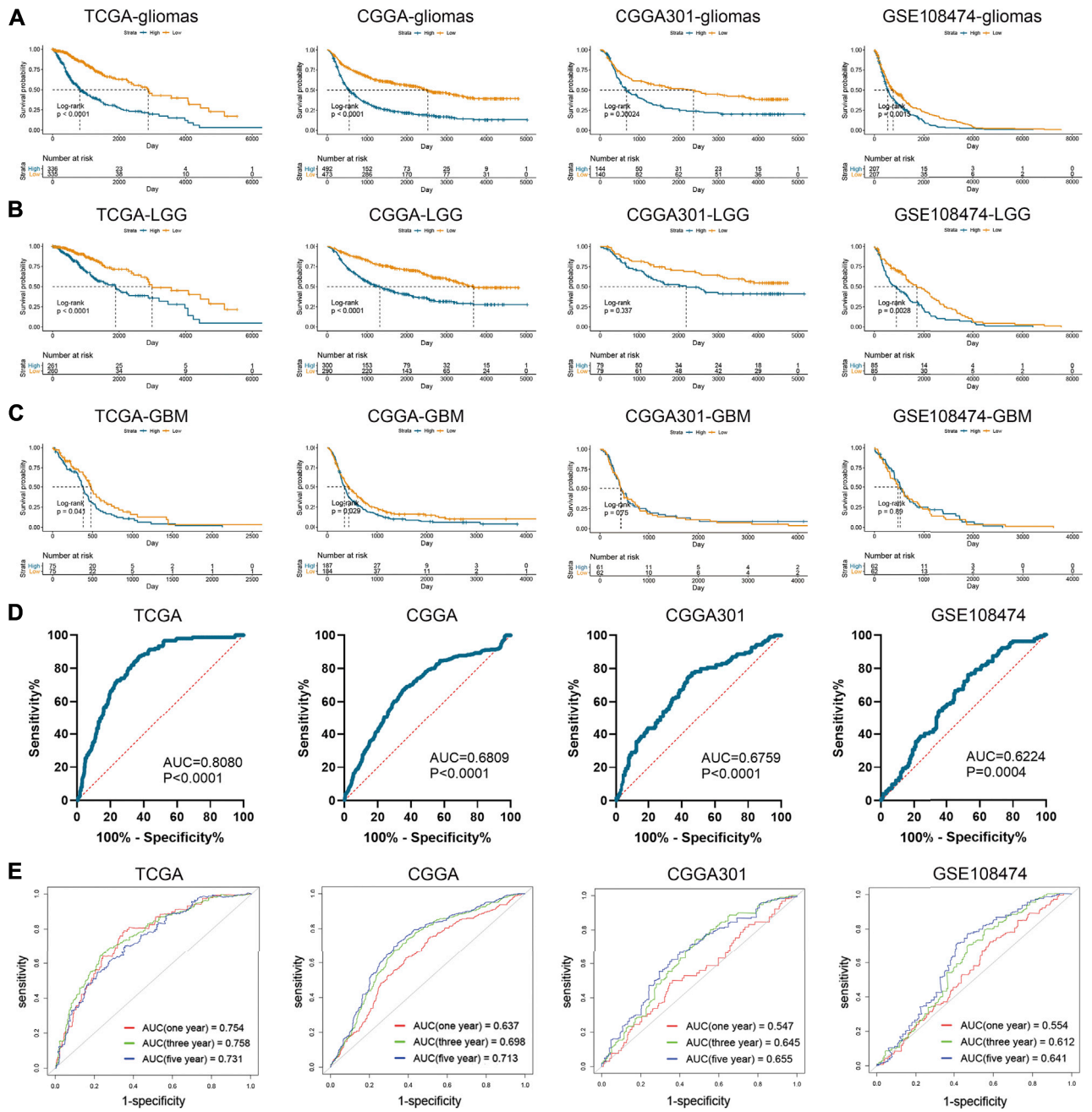


FIGURE 8 | Prognostic and predictive values of CD74 in gliomas. A-C. Kaplan-Meier analysis of CD74 expression in the overall survival of patients with gliomas **(A)**, lower-grade gliomas **(B)**, and glioblastoma **(C)**. D. Patients were divided into glioblastoma and lower-grade glioma groups. ROC analysis was conducted to evaluate the predictive accuracy of CD74 expression for the diagnosis of glioblastoma. E. ROC analysis of CD74 expression in predicting one-, three, and 5-year overall survival of glioma patients. ROC: receiver operating characteristic.

Prognostic and Predictive Values of CD74 in Gliomas

Apart from the immune features, CD74 also exhibited prognostic and predictive values in the OS of glioma patients. Kaplan-Meier analysis revealed that the high expression of CD74 indicated poor prognosis in patients with gliomas and LGG ($p < 0.05$) (**Figures 8A,B**). While in patients with GBM, this finding was consistent in TCGA and CGGA datasets ($p < 0.05$), and no significant difference was detected in CGGA301 and GSE108474 datasets ($p > 0.05$) (**Figure 8C**). Multivariate Cox analysis showed that CD74 and tumor grade were independent risk factors for glioma patients in TCGA, CGGA, and GSE108474 datasets ($p < 0.05$) whereas no significant difference was detected in the CGGA301 cohort ($p > 0.05$) (**Table 3**). Besides, tumor grade was the

TABLE 3 | Multivariate analysis of CD74 and clinical features in four datasets.

Variables	TCGA (n = 672)		CGGA (n = 1,013)		CGGA301 (n = 300)		GSE108474 (n = 414)	
	HR (95%CI)	p Value	HR (95%CI)	p Value	HR (95%CI)	p Value	HR (95%CI)	p Value
CD74	1.177 (1.045–1.326)	0.007	1.070 (1.019–1.121)	0.018	1.147 (0.896–1.468)	0.277	2.253 (1.219–4.164)	0.010
Age	1.040 (1.028–1.053)	<0.001	1.007 (1–1.015)	0.046	1.025 (0.987–1.064)	0.201	—	—
Grade	1.885 (1.447–2.456)	<0.001	2.149 (1.854–2.491)	<0.001	2.275 (1.473–3.514)	<0.001	1.602 (1.359–1.889)	<0.001
IDH	2.043 (1.277–3.27)	0.003	1.164 (0.927–1.46)	0.191	0.601 (0.245–1.475)	0.266	—	—
1p19q	1.614 (0.944–2.762)	0.080	2.574 (1.845–3.592)	<0.001	4.13 (1.352–12.611)	0.013	—	—
MGMT	1.321 (0.941–1.854)	0.108	1.122 (0.932–1.35)	0.223	2.365 (1.174–4.767)	0.016	—	—

HR: hazard ratio; IDH: isocitrate dehydrogenases; MGMT: O-6-methylguanine-DNA methyltransferase.

independent risk factor for glioma patients in the four cohorts ($p < 0.05$). Then we applied CD74 as the potential biomarker to facilitate the diagnosis of GBM, results showed that CD74 had moderate predictive accuracy with the area under the curve (AUC) of 0.808, 0.681, 0.676, and 0.622 in TCGA, CGGA, CGGA301, and GSE108474 datasets, respectively ($p < 0.05$) (**Figure 8D**). Moreover, CD74 exhibited moderate accuracy in predicting the OS of glioma patients. The AUC of one-, three-, and 5 year survival was 0.754, 0.758, and 0.731, respectively in the TCGA dataset; that in the CGGA cohort was 0.637, 0.698, and 0.713, respectively; that in CGGA301 dataset was 0.547, 0.645, and 0.655, respectively; that in GSE108474 dataset was 0.554, 0.612, and 0.641, respectively (**Figure 8E**). These results indicated that CD74 could serve as a biomarker to predict the prognosis and facilitate the diagnosis of gliomas.

DISCUSSION

Given the fact that glioma patients suffer from poor prognosis and limited therapies, there is a clear urgent need to find novel biomarkers and develop therapeutic strategies. Our study revealed that CD74 was highly expressed in the high-grade glioma and correlated with poor prognosis of glioma patients. Besides, CD74 was positively associated with immune infiltration in the glioma immune microenvironment. These results indicated that CD74 could be used as a potential target for the treatment of glioma and as a novel biomarker to predict prognoses of glioma patients.

Previous studies revealed that CD74 was associated with glioma malignancy and activity of glioma-associated macrophages (Zeiner et al., 2015; Ghoochani et al., 2016; Alban et al., 2020). Our study collected 40 glioma specimens and extracted 2,399 glioma samples from the online databases to explore the expression pattern of CD74 in gliomas. The high expression of CD74 was significantly higher in glioma tissue compared to the normal tissue. CD74 was highly expressed in grade IV, IDH wildtype, 1p19q non-codeletion, and MGMT unmethylated gliomas. Moreover, CD74 was highly expressed in the mesenchymal subtype, which was the most malignant molecular subtype of glioma (Wang et al., 2021). Besides, the protein level of CD74 was markedly elevated in the higher-grade glioma. These results indicated that the high expression of CD74 was associated with a relatively malignant type of glioma. Therefore, CD74 might serve as a biomarker to predict the malignancy of glioma.

As a pivotal component in antigen presentation, CD74 was assumed to have an impact on immune responses and immune infiltrations. A previous study suggested that CD74 was associated with poor prognosis and high tumor-infiltrating leucocyte in breast cancer, which was consistent with our finding (Wang et al., 2017). Our study found that the expression of CD74 was positively associated with the expression of immune checkpoints and inflammatory cytokines. Then, we screened the co-expressed genes of CD74 to further characterize the immune feature of CD74 in gliomas. The GSEA and GSVA analyses showed that the positively co-expressed genes of CD74 were enriched in immune responses and autoimmune diseases. These results preliminarily demonstrated the immune characteristics of CD74 in gliomas. However, it should be noted that these algorithms were conducted based on the expression of pathway-related genes, which could not determine the actual status of these immune processes in different samples. Additional experiments were needed to explore the regulatory effect of CD74 on specific immune processes.

Immune microenvironment was implicated to play an important role in cancer progression (Quail and Joyce, 2017; Zhang et al., 2020a; Zhang et al., 2020b; Wang et al., 2020). CD74 was the receptor of MIF and the CD74-MIF signaling pathway regulated the activity of macrophages and other immune cells (Borghese and Clanchy, 2011; Figueiredo et al., 2018). Therefore, we also explored the correlation between CD74 and the abundance of infiltrated immune cells. The tumor purity of each glioma sample was calculated based on the ESTIMATE algorithm. Since the high immune infiltration indicated low tumor purity, the negative correlation between CD74 and tumor purity suggested that high expression of CD74 was associated with high immune infiltration. With the application of two algorithms, we found that CD74 was markedly positively associated with the infiltration of immune cells such as macrophages, dendritic cells, and neutrophils. Previous studies indicated that macrophages in tumor microenvironment commonly referred to M2 subtype, which exhibited pro-tumoral effects and induced immunosuppressive context (Najafi et al., 2019; Xu et al., 2020b). Moreover, high densities of infiltrating neutrophils were associated with advanced stages of cancers including glioma (Shaul and Fridlender, 2019). Additionally, Zhang et al. (2017) revealed that tumor purity

was lowest in grade IV glioma and highest in grade II glioma, and the low tumor purity indicated a high enrichment of macrophages, activated dendritic cells, and neutrophils (Zhang et al., 2017). Therefore, the infiltration of macrophages, dendritic cells, and neutrophils indicated high malignancy of glioma, which was consistent with our findings that CD74 was positively associated with the malignancy of glioma and the infiltration of macrophages, dendritic cells, and neutrophils. However, it should be noted that the infiltration of immune cells was assessed using different algorithms. These algorithms were conducted based on the expression of genes that were associated with the activity of related immune cells. Therefore, it was hard to determine the exact population of immune cells in different samples. However, through this large-scale analysis, the immune characteristics of CD74 were preliminarily demonstrated, which could provide a reference for further investigation of the regulatory association between CD74 and immune cells.

The low tumor purity has been implicated to be associated with poor prognosis in glioma patients (Zhang et al., 2017; Xu et al., 2021; Zhang et al., 2021; Kang et al., 2021). Since high expression of CD74 indicated low tumor purity, we further explored the clinical significance of CD74 in glioma patients. In glioma and LGG patients, CD74 was significantly associated with survival rates. However, the finding was diverse in GBM patients. This might be due to the intense malignancy of GBM and the single biomarker had a limited ability to stratify the survival of GBM patients. Moreover, multivariate analyses indicated that CD74 was an independent risk factor for glioma patients in the TCGA, CGGA, and GSE108474 datasets, but not in the CGGA301 dataset. Since we included the status of IDH, 1p19q, and MGMT in multivariate analyses, the number of samples that had complete information in the CGGA301 dataset was relatively limited as shown in **Table 1**, which might lead to the discrepancy of the results in the CGGA301 dataset compared with other datasets. Additionally, CD74 exhibited moderate predictive accuracy in the diagnosis of GBM and the prognosis of glioma patients. Compared with integrated signature, the predictive accuracy of CD74 was relatively low. Cheng et al. constructed a risk signature based on the expression of eight immune-related genes, which could predict the prognosis and the efficacy of chemoradiotherapy in GBM patients (Cheng et al., 2016). Similarly, Zuo et al. (2019) reported a six-gene signature that could predict the OS and therapeutic responses of GBM patients (Zuo et al., 2019). The AUC value of their constructed signatures in predicting long-term survival was higher than that of our study. Therefore, the predictive value of a single biomarker was limited. However, CD74 still had some merits in predicting the diagnosis and prognosis of gliomas and could be used as a potential target for glioma treatment.

However, additional experiments would help provide a better understanding of the underlying mechanism of CD74 in the development and progression of gliomas. Although CD74 was proven to be associated with several immune processes, the effect of CD74 on specific immune cells required further investigation.

CONCLUSION

To sum up, we conducted a large-scale analysis to characterize the immune and clinical features of CD74 in gliomas. The high expression of CD74 was associated with poor prognosis and high immune infiltration. CD74 could be used as a potential target for glioma treatment and as a biomarker to predict the prognosis of glioma patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The data are accessible in TCGA and CGGA databases.

ETHICS

The studies involving human participants were approved by the Ethics Committee of Xiangya Hospital, Central South University with written informed consent obtained.

AUTHOR CONTRIBUTIONS

QC and KY conceived, designed, and supervised the study. SX and XL drafted the manuscript. SX, LT and ZL collected the data. SX performed data analysis and visualization. XL facilitated the revision of the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (81703622, 81873635, 82073893); Natural Science Foundation of Hunan Province (2018JJ3838, 2019JJ50948); China Postdoctoral Science Foundation (2018M633002); Key Research and Development Program of Hunan Province (2018SK2101); Xiangya Hospital Central South University postdoctoral foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.706949/full#supplementary-material

REFERENCES

Alban, T. J., Bayik, D., Otvos, B., Rabljenovic, A., Leng, L., Jia-Shiun, L., et al. (2020). Glioblastoma Myeloid-Derived Suppressor Cell Subsets Express Differential Macrophage Migration Inhibitory Factor Receptor Profiles that Can Be Targeted to Reduce Immune Suppression. *Front. Immunol.* 11, 1191. doi:10.3389/fimmu.2020.01191

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

Binsky, I., Haran, M., Starlets, D., Gore, Y., Lantner, F., Harpaz, N., et al. (2007). IL-8 Secreted in a Macrophage Migration-Inhibitory Factor- and CD74-dependent Manner Regulates B Cell Chronic Lymphocytic Leukemia Survival. *Proc. Natl. Acad. Sci.* 104 (33), 13408–13413. doi:10.1073/pnas.0701553104

Blackburn, S. D., Shin, H., Haining, W. N., Zou, T., Workman, C. J., Polley, A., et al. (2009). Coregulation of CD8+ T Cell Exhaustion by Multiple Inhibitory Receptors during Chronic Viral Infection. *Nat. Immunol.* 10 (1), 29–37. doi:10.1038/ni.1679

Bleeker, F. E., Molenaar, R. J., and Leenstra, S. (2012). Recent Advances in the Molecular Understanding of Glioblastoma. *J. Neurooncol.* 108 (1), 11–27. doi:10.1007/s11060-011-0793-0

Borghese, F., and Clanchy, F. I. (2011). CD74: an Emerging Opportunity as a Therapeutic Target in Cancer and Autoimmune Disease. *Expert Opin. Ther. Targets* 15 (3), 237–251. doi:10.1517/14728222.2011.550879

Chen, H., Thomas, C., Munoz, F. A., Alexandrescu, S., Horbinski, C. M., Olar, A., et al. (2019). Polysomy Is Associated with Poor Outcome in 1p/19q Codeleted Oligodendroglial Tumors. *Neuro Oncol.* 21 (9), 1164–1174. doi:10.1093/neuonc/noz098

Cheng, W., Ren, X., Zhang, C., Cai, J., Liu, Y., Han, S., et al. (2016). Bioinformatic Profiling Identifies an Immune-Related Risk Signature for Glioblastoma. *Neurology* 86 (24), 2226–2234. doi:10.1212/wnl.0000000000002770

Cresswell, P. (1994). Assembly, Transport, and Function of MHC Class II Molecules. *Annu. Rev. Immunol.* 12, 259–291. doi:10.1146/annurev.iy.12.040194.001355

Farr, L., Ghosh, S., and Moonah, S. (2020). Role of MIF Cytokine/CD74 Receptor Pathway in Protecting against Injury and Promoting Repair. *Front. Immunol.* 11, 1273. doi:10.3389/fimmu.2020.01273

Figueiredo, C. R., Azevedo, R. A., Mousdell, S., Resende-Lara, P. T., Ireland, L., Santos, A., et al. (2018). Blockade of MIF-CD74 Signalling on Macrophages and Dendritic Cells Restores the Antitumour Immune Response against Metastatic Melanoma. *Front. Immunol.* 9, 1132. doi:10.3389/fimmu.2018.01132

Ghoochani, A., Schwarz, M. A., Yakubov, E., Engelhorn, T., Doerfler, A., Buchfelder, M., et al. (2016). MIF-CD74 Signaling Impedes Microglial M1 Polarization and Facilitates Brain Tumorigenesis. *Oncogene* 35 (48), 6246–6261. doi:10.1038/onc.2016.160

Goodenberger, M. L., and Jenkins, R. B. (2012). Genetics of Adult Glioma. *Cancer Genet.* 205 (12), 613–621. doi:10.1016/j.cancergen.2012.10.009

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

Hegi, M. E., Diserens, A.-C., Gorlia, T., Hamou, M.-F., de Tribolet, N., Weller, M., et al. (2005). MGMTGene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* 352 (10), 997–1003. doi:10.1056/nejmoa043331

Kang, Y., Huang, J., Liu, Y., Zhang, N., Cheng, Q., and Zhang, Y. (2021). Integrated Analysis of Immune Infiltration Features for Cervical Carcinoma and Their Associated Immunotherapeutic Responses. *Front. Cel Dev. Biol.* 9, 573497. doi:10.3389/fcell.2021.573497

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 (W1), W509–W514. doi:10.1093/nar/gkaa407

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). ARACNE: an Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 7 Suppl 1 (Suppl. 1), S7. doi:10.1186/1471-2105-7-S1-S7

Meyer-Siegler, K. L., Iczkowski, K. A., Leng, L., Bucala, R., and Vera, P. L. (2006). Inhibition of Macrophage Migration Inhibitory Factor or its Receptor (CD74) Attenuates Growth and Invasion of DU-145 Prostate Cancer Cells. *J. Immunol.* 177 (12), 8730–8739. doi:10.4049/jimmunol.177.12.8730

Molenaar, R. J., Radivoyevitch, T., Maciejewski, J. P., van Noorden, C. J. F., and Bleeker, F. E. (2014). The Driver and Passenger Effects of Isocitrate Dehydrogenase 1 and 2 Mutations in Oncogenesis and Survival Prolongation. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1846 (2), 326–341. doi:10.1016/j.bbcan.2014.05.004

Molenaar, R. J., Verbaan, D., Lamba, S., Zanon, C., Jeuken, J. W. M., Boots-Sprenger, S. H. E., et al. (2014). The Combination of IDH1 Mutations and MGMT Methylation Status Predicts Survival in Glioblastoma Better Than Either IDH1 or MGMT Alone. *Neuro Oncol.* 16 (9), 1263–1273. doi:10.1093/neuonc/nou005

Nagata, S., Jin, Y.-F., Yoshizato, K., Tomoeda, M., Song, M., Iizuka, N., et al. (2009). CD74 Is a Novel Prognostic Factor for Patients with Pancreatic Cancer Receiving Multimodal Therapy. *Ann. Surg. Oncol.* 16 (9), 2531–2538. doi:10.1245/s10434-009-0532-3

Najafi, M., Hashemi Goradel, N., Farhood, B., Salehi, E., Nashtaei, M. S., Khanlarkhani, N., et al. (2019). Macrophage Polarity in Cancer: A Review. *J. Cel Biochem* 120 (3), 2756–2765. doi:10.1002/jcb.27646

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

Ohgaki, H., and Kleihues, P. (2005). Population-based Studies on Incidence, Survival Rates, and Genetic Alterations in Astrocytic and Oligodendroglial Gliomas. *J. Neuropathol. Exp. Neurol.* 64 (6), 479–489. doi:10.1093/jnen/64.6.479

Quail, D. F., and Joyce, J. A. (2017). The Microenvironmental Landscape of Brain Tumors. *Cancer Cell* 31 (3), 326–341. doi:10.1016/j.ccell.2017.02.009

Shaul, M. E., and Fridlender, Z. G. (2019). Tumour-associated Neutrophils in Patients with Cancer. *Nat. Rev. Clin. Oncol.* 16 (10), 601–620. doi:10.1038/s41571-019-0222-4

Smoll, N. R., Gautschi, O. P., Schatlo, B., Schaller, K., and Weber, D. C. (2012). Relative Survival of Patients with Supratentorial Low-Grade Gliomas. *Neuro-Oncology* 14 (8), 1062–1069. doi:10.1093/neuonc/nos144

Tanese, K., Hashimoto, Y., Berkova, Z., Wang, Y., Samaniego, F., Lee, J. E., et al. (2015). Cell Surface CD74-MIF Interactions Drive Melanoma Survival in Response to Interferon-γ. *J. Invest. Dermatol.* 135 (11), 2775–2784. doi:10.1038/jid.2015.204

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

Thakkar, J. P., Dolecek, T. A., Horbinski, C., Ostrom, Q. T., Lightner, D. D., Barnholtz-Sloan, J. S., et al. (2014). Epidemiologic and Molecular Prognostic Review of Glioblastoma. *Cancer Epidemiol. Biomarkers Prev.* 23 (10), 1985–1996. doi:10.1158/1055-9965.epi-14-0275

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

Wang, K. Y., Huang, R. Y., Tong, X. Z., Zhang, K. N., Liu, Y. W., Zeng, F., et al. (2019). Molecular and Clinical Characterization of TMEM71 Expression at the Transcriptional Level in Glioma. *CNS Neurosci. Ther.* 25 (9), 965–975. doi:10.1111/cns.13137

Wang, Y., Wang, Y., Xu, C., Liu, Y., and Huang, Z. (2020). Identification of Novel Tumor-Microenvironment-Regulating Factor that Facilitates Tumor Immune Infiltration in Colon Cancer. *Mol. Ther. - Nucleic Acids* 22, 236–250. doi:10.1016/j.omtn.2020.08.029

Wang, Z.-Q., Milne, K., Webb, J. R., and Watson, P. H. (2017). CD74 and Intratumoral Immune Response in Breast Cancer. *Oncotarget* 8 (8), 12664–12674. doi:10.18632/oncotarget.8610

Wang, Z., Zhang, H., Xu, S., Liu, Z., and Cheng, Q. (2021). The Adaptive Transition of Glioblastoma Stem Cells and its Implications on Treatments. *Sig Transduct Target. Ther.* 6 (1), 124. doi:10.1038/s41392-021-00491-w

Xu, S., Tang, L., Dai, G., Luo, C., and Liu, Z. (2020). Expression of m6A Regulators Correlated with Immune Microenvironment Predicts Therapeutic Efficacy and Prognosis in Gliomas. *Front. Cel Dev. Biol.* 8, 594112. doi:10.3389/fcell.2020.594112

Xu, S., Tang, L., Dai, G., Luo, C., and Liu, Z. (2021). Immune-related Genes with APA in Microenvironment Indicate Risk Stratification and Clinical Prognosis in Grade II/III Gliomas. *Mol. Ther. - Nucleic Acids* 23, 1229–1242. doi:10.1016/j.omtn.2021.01.033

Xu, S., Tang, L., Li, X., Fan, F., and Liu, Z. (2020). Immunotherapy for Glioma: Current Management and Future Application. *Cancer Lett.* 476, 1–12. doi:10.1016/j.canlet.2020.02.002

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

Zeiner, P. S., Preusse, C., Blank, A.-E., Zachskorn, C., Baumgarten, P., Caspary, L., et al. (2015). MIF Receptor CD74 Is Restricted to Microglia/Macrophages, Associated with a M1-Polarized Immune Milieu and Prolonged Patient Survival in Gliomas. *Brain Pathol.* 25 (4), 491–504. doi:10.1111/bpa.12194

Zhang, C., Cheng, W., Ren, X., Wang, Z., Liu, X., Li, G., et al. (2017). Tumor Purity as an Underlying Key Factor in Glioma. *Clin. Cancer Res.* 23 (20), 6279–6291. doi:10.1158/1078-0432.ccr-16-2598

Zhang, N., Dai, Z., Wu, W., Wang, Z., Cao, H., Zhang, Y., et al. (2021). The Predictive Value of Monocytes in Immune Microenvironment and Prognosis of Glioma Patients Based on Machine Learning. *Front. Immunol.* 12, 656541. doi:10.3389/fimmu.2021.656541

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

Zhang, Y., Yang, M., Ng, D. M., Haleem, M., Yi, T., Hu, S., et al. (2020). Multi-omics Data Analyses Construct TME and Identify the Immune-Related Prognosis Signatures in Human LUAD. *Mol. Ther. - Nucleic Acids* 21, 860–873. doi:10.1016/j.omtn.2020.07.024

Zuo, S., Zhang, X., and Wang, L. (2019). A RNA Sequencing-Based Six-Gene Signature for Survival Prediction in Patients with Glioblastoma. *Sci. Rep.* 9 (1), 2615. doi:10.1038/s41598-019-39273-4

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 Xu, Li, Tang, Liu, Yang and Cheng. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fmolb.2021.720020

Comprehensive Analysis of HDAC Family Identifies *HDAC1* as a Prognostic and Immune Infiltration Indicator and *HDAC1*-Related Signature for Prognosis in Glioma

Yuxiang Fan[1], *Xinyu Peng*[2], *Yubo Wang*[1], *Baoqin Li*[3] *and Gang Zhao*[1]*

[1]*Department of Neurosurgery, The First Hospital of Jilin University, Changchun, China,* [2]*Department of Hepatobiliary Pancreatic Surgery, The First Hospital of Jilin University, Changchun, China,* [3]*Department of Spine Surgery, The First Hospital of Jilin University, Changchun, China*

Background: The histone deacetylase (HDAC) family limited accessibility to chromatin containing tumor suppressor genes by removing acetyl groups, which was deemed a path for tumorigenesis. Considering glioma remained one of the most common brain cancers with a dichotomy prognosis and limited therapy responses, HDAC inhibitors were an area of intensive research. However, the expression profiles and prognostic value of the HDACs required more elucidation.

Methods: Multiple biomedical databases were incorporated, including ONCOMINE, GEPIA, TCGA, CGGA, GEO, TIMER, cBioPortal, and Metascape, to study expression profiles, prognostic value, immune infiltration, mutation status, and enrichment of HDACs in glioma. STRING and GeneMANIA databases were used to identify *HDAC1*-related molecules. LASSO regression, Cox regression, Kaplan-Meier plot, and receiver operating characteristic (ROC) analyses were performed for *HDAC1*-related signature construction and validation.

Results: *HDAC1* was significantly overexpressed in glioma, while *HDAC11* was downregulated in glioblastoma. Except for *HDAC 6/9/10*, the HDAC family expression was significantly associated with glioma grade. Most of the HDAC family also correlated with glioma genetic mutations. Higher *HDAC1* expression level predicted more dismal overall survival (OS) ($p < 0.0001$) and disease-free survival (DFS) ($p < 0.0001$), but a higher level of *HDAC11* held more favorable OS ($p = 2.1e{-}14$) and DFS ($p = 4.8e{-}08$). *HDAC4*

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong, SAR China

Reviewed by:
Ajay Chatrath,
University of Virginia, United States
Monal Sharma,
Duke University, United States

***Correspondence:**
Gang Zhao
zhaogjdyy@126.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *03 June 2021*
Accepted: *19 July 2021*
Published: *01 September 2021*

Citation:
Fan Y, Peng X, Wang Y, Li B and Zhao G (2021) Comprehensive Analysis of HDAC Family Identifies HDAC1 *as a Prognostic and Immune Infiltration Indicator and* HDAC1*-Related Signature for Prognosis in Glioma. Front. Mol. Biosci. 8:720020. doi: 10.3389/fmolb.2021.720020*

Abbreviations: AUC, area under the curve ; CNS, central nervous system; CI, confidence interval; DFS, disease-free survival; FC, fold change; FKBP3, FK506 binding protein 3; GEPIA, gene expression profiling interactive analysis; GTEx, genotype-tissue expression project; GBM, glioblastoma multiforme; HR, hazard ratio; HAT, histone acetyltransferase; HDAC, histone deacetylase; IDH1, isocitrate dehydrogenase 1; KM, Kaplan-Meier; LGG, lower-grade glioma; MDSC, myeloid-derived suppressive cell; NFAT, nuclear factor of activated T cell; OS, overall survival; PPARδ, peroxisome proliferator-activated receptor δ; PGC1α, peroxisome proliferator-activated receptor γ coactivator1 α; PHF21A, PHD finger protein 21A; PH, proportional hazards; ROC, receiver operating characteristic; Tregs, regulatory T cells; RBL1, retinoblastoma transcriptional corepressor like 1; RUNX1, runt-related transcription factor 1; RUNX1T1, RUNX1 partner transcriptional co-repressor 1; STRING, search tool for the retrieval of interacting genes; SATB1, special AT-rich sequence-binding protein 1; TCGA, the cancer genome atlas; LASSO, the least absolute shrinkage and selection operator; TIMER, tumor immune estimation resource.

displayed the highest mutation ratio, at 2.6% of the family. The prognostic value of *HDAC1* was validated with ROC achieving 0.70, 0.77, 0.75, and 0.80 as separability for 1-, 3-, 5-, and 10-years OS predictions in glioma, respectively. Moreover, *HDAC1* expression positively correlated with neutrophil (r = 0.60, p = 2.88e-47) and $CD4^+$ T cell infiltration (r = 0.52, p = 3.96e-35) in lower-grade glioma. The final *HDAC1*-related signature comprised of *FKBP3*, *HDAC1* (Hazard Ratio:1.49, 95%Confidence Interval:1.20–1.86), *PHF21A*, *RUNX1T1*, and *RBL1*, and was verified by survival analysis ($p < 0.0001$) and ROC with 0.80, 0.84, 0.83, and 0.88 as separability for 1-, 3-, 5-, and 10-years OS predictions, respectively. The signature was enriched in chromatin binding.

Conclusion: HDAC family was of clinical significance for glioma. Most of the HDAC family significantly correlated with the glioma grade, *IDH1* mutation, and *1p/19q* codeletion. *HDAC1* was both a prognostic and immune infiltration indicator and a central component of the *HDAC1*-related signature for precise prognosis prediction in glioma.

Keywords: glioma, HDAC1, HDAC family, immune infiltration, prognosis, signature

INTRODUCTION

Glioma, characterized by its dichotomy prognosis, is one of the most common primary brain tumors in adults (Schiff et al., 2019; Wen et al., 2020). Patients diagnosed with glioblastoma multiforme (GBM, WHO IV grade) typically hold a median survival time of merely 14 months, whereas most cases of low-grade glioma, like pilocytic astrocytoma (WHO I grade), attained clinical cure after surgical resection (Louis et al., 2016; Wen et al., 2020). The malignancy might come from a highly unstable genome and extensive epigenetic deregulation. Since histone modifications were one of the key mechanisms in epigenetics, investigations related to the imbalance between histone acetyltransferases (HATs) and histone deacetylases (HDACs) in glioma are emerging (Kunadis et al., 2021).

Histone acetylation by HATs relaxed chromatin structure and facilitated the transcriptional complex accessing the core histone (Ruijter et al., 2003). By manipulating acetyl groups, HATs and HDACs reached a balance in the regulation of histone structure under physical conditions (Eyüpoglu and Savaskan, 2016). However, a carcinogenic transformation would initiate once the equilibrium was disrupted (Minucci and Pelicci, 2006). One possible mechanism behind the transformation was mediated by the activation of oncogenic genes like *c-Myc* and the repression of the tumor suppressor gene when the HDACs took dominance (Bolden et al., 2006; Nguyen et al., 2020). Therefore, it was of necessity to explore the HDACs in the context of glioma based on their oncogenic properties. The HDAC family so far encompassed four classes, which were further categorized by their dependency on zinc, namely zinc-dependent class I (HDAC1, 2, 3, and 8), class IIa (HDAC4, 5, 7, and 9), class IIb (HDAC 6 and 10), and class IV (HDAC11), and zinc-independent class III (SIRT proteins) (Chen et al., 2020). This study would mainly discuss the HDAC family.

In this research, the HDAC family was systemically analyzed across varieties of databases to evaluate its clinical and prognostic value for glioma. It helped discern the differentially expressed HDAC family genes with significance in glioma compared to normal, as well as the general expression level of each HDAC family member in glioma. The potential ties of the HDAC family to glioma were assessed by detecting the associations between the HDAC family's expression and glioma grade, and genetic mutations, and by conducting survival analysis on the HDAC family. Genetic alterations, interactive network, and functional enrichment annotations were additionally acquired for the HDAC family. The prognostic gene *HDAC1* and the *HDAC1*-related genes were prepared for the final *HDAC1*-related signature construction. In conclusion, the HDAC family was of prognostic significance and clinical interest for glioma. It exposed the pivotal role of *HDAC1* in glioma, as both an independent prognostic and immune infiltration biomarker and a central component of the *HDAC1*-related signature for precise prognosis prediction.

MATERIALS AND METHODS

Oncomine

The HDAC family member expression profiles in cancers were analyzed in the ONCOMINE database (www.ONCOMINE.org), which enabled public access to resourceful genome-wide cancer microarray data that originated from various studies (Rhodes et al., 2004). Significance thresholds were set as a p-value less than 0.01, a fold-change over 2, and gene rank within the top 10%. The detailed information for each study included was listed in **Table 1**.

Gene Expression Profiling Interactive Analysis

The mRNA sequencing data of the HDAC family together with its corresponding clinical information of glioma patients were

TABLE 1 | Studies on HDAC family in ONCOMINE.

	Study (Glioma vs. Normal)	Fold change	*p*-valueadjusted	*t*-test	Cases (Tumor/Normal)	References
HDAC1	Atypical teratoid tumor	4.597	8.32E-04	5.136	5/4	Pomeroy Nature 2002/01/24
	Anaplastic oligoastrocytoma	2.560	4.41E-04	5.148	4/6	French Cancer Res 2005/12/15
	Glioblastoma	3.131	3.08E-08	14.61	542/10	TCGA 2013/06/03
HDAC2	Desmoplastic medulloblastoma	3.133	1.09E-04	6.188	14/4	Pomeroy Nature 2002/01/24
	Oligodendroglioma	3.235	6.99E-04	8.816	3/7	Shai Oncogene 2003/07/31
	Glioblastoma	−3.683	2.70E-10	−11.1	22/3	Lee Cancer Cell 2006/05/01
	Anaplastic oligoastrocytoma	−4.399	5.95E-04	−5.14	6/4	Bredel Cancer Res 2005/10/01
HDAC4	Glioblastoma	−2.147	1.70E-13	−15.6	80/4	Murat J Clin Oncol 2008/06/20
HDAC5	Anaplastic oligoastrocytoma	−2.092	4.64E-05	−8.11	6/4	Bredel Cancer Res 2005/10/01
	Glioblastoma	−2.378	1.46E-07	−12.2	27/4	Bredel Cancer Res 2005/10/01
	Anaplastic oligodendroglioma	−2.436	0.003	−7.56	3/4	Bredel Cancer Res 2005/10/01
	Glioblastoma	−2.221	6.80E-20	−11.7	81/23	Sun Cancer Cell 2006/04/01
	Glioblastoma	−2.742	3.08E-08	−14.9	542/10	TCGA 2013/06/03
HDAC6	Glioblastoma	3.221	8.88E-13	8.831	81/23	Sun Cancer Cell 2006/04/01
	Anaplastic astrocytoma	2.605	2.34E-06	5.48	19/23	Sun Cancer Cell 2006/04/01
HDAC11	Anaplastic oligoastrocytoma	−4.196	1.09E-05	−9.32	4/6	French Cancer Res 2005/12/15
	Anaplastic oligodendroglioma	−3.927	7.32E-06	−9.90	23/6	French Cancer Res 2005/12/15
	Glioblastoma	−3.539	3.14E-14	−11.0	81/23	Sun Cancer Cell 2006/04/01

retrieved from GEPIA (http://gepia2.cancer-pku.cn/), a web-based interactive tool providing comprehensive and customizable analyses with The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression Project (GTEx) RNA sequencing data as resources (Tang et al., 2017). In this study, differential expression analysis comparing 681 gliomas (518 cases of lower-grade glioma, 163 cases of glioblastoma) with 207 normal samples was performed, and survival analysis with Kaplan-Meier (KM) plot and survival heatmap was also included.

TCGA, Chinese Glioma Genome Atlas, and Gene Expression Omnibus

The glioma cohort in the TCGA research program (https://www.cancer.gov/tcga), including 505 cases of lower-grade glioma (LGG) and 155 cases of GBM RNA sequencing counts data with clinical information, was acquired with the R package "TCGAbiolinks" (Colaprico et al., 2016). The dataset was mainly used as a developing cohort for the construction of the *HDAC1*-related signature. The TCGA developing cohort was then randomly split at a ratio of 3:7 using the R package "caret" for internal validation (Kuhn, 2008).

The glioma RNA-seq dataset "mRNAseq_693" recruited in the CGGA (http://www.cgga.org.cn/) contains 693 glioma samples. It was selected to conduct clinical correlation analysis by linking RNA-seq data with pathological grade and typical mutation statuses, such as *IDH1* mutation and *1p/19q* codeletion. Moreover, the "mRNAseq_693" dataset was employed as verification for previous results from the differential expression analysis and the survival analysis in GEPIA.

The microarray dataset GSE16011 from the Gene Expression Omnibus (GEO) database was gathered as an external validation cohort for the *HDAC1*-related signature (Gravendeel et al., 2009). It consisted of 284 samples, including 117 LGG cases, 156 GBM cases, and normal controls. Patients with clear mutation records were kept.

cBioPortal

The mutation profiles of each HDAC family member were further analyzed in glioma with help from the cBioPortal for Cancer Genomics (http://cbioportal.org), which was an intuitive web tool for exploring analysis and visualization on cancer genomic data collected from several platforms (Gao et al., 2013). The detailed mutation statuses of the HDAC family in glioma pathological subtypes were also shown.

Search Tool for the Retrieval of Interacting Genes

STRING (https://string-db.org) pictured both physical and functional protein-protein interaction (PPI) networks based on current knowledge and prediction via systemic co-expression analysis and text-mining of literature (Szklarczyk et al., 2017). The PPI network analysis was used to identify genes associated with the HDAC family members and *HDAC1*. The HDAC family PPI network was visualized by the Cytoscape app (version 3.7.2).

Metascape

Metascape (http://metascape.org) served as a web-based portal mainly for gene annotation, interactome analysis, and functional enrichment analysis (Zhou et al., 2019). The HDAC family and the *HDAC1*-related gene signature were uploaded to query for the functional interpretation of those genes in fixed combination, thereby enlightening future investigation.

Tumor Immune Estimation Resource

TIMER database (http://timer.cistrome.org) was introduced to measure immune infiltration in the tumor microenvironment and attain a better comprehension of tumor-immune interactions (Li et al., 2020). Considering feasibility, "immune privilege" in the central nervous system (CNS), and loading capacity for each immune infiltration estimation analysis, only a fraction of typical infiltrated immune cells in glioma, including neutrophils, macrophages, myeloid-derived suppressive cells (MDSCs),

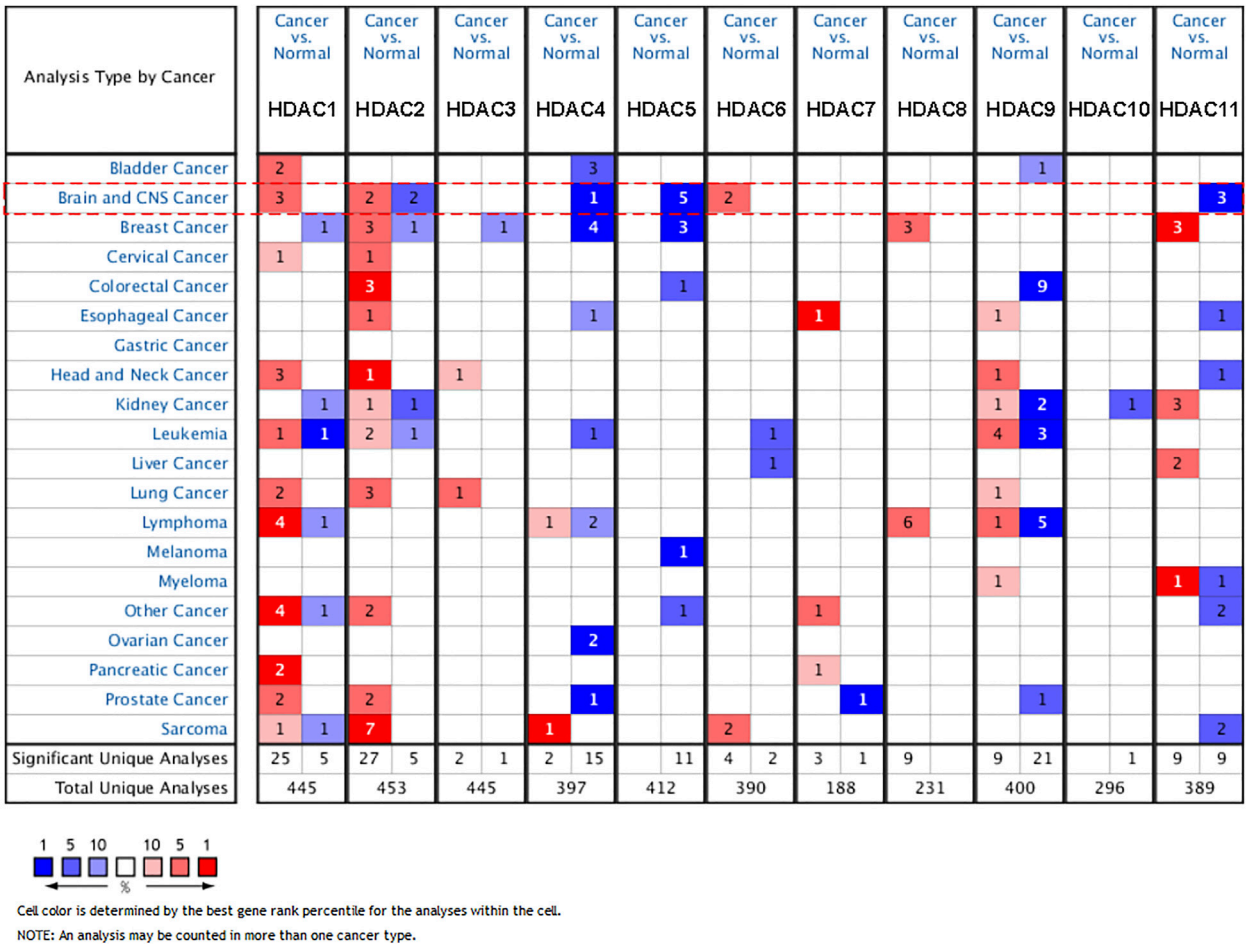


FIGURE 1 | Expression microarray of HDAC family across varieties of cancers and corresponding control groups in ONCOMINE. The number in cells indicated the significant datasets showing the differentially expressed genes of the HDAC family. The red color in the cell was for upregulation, and the blue color was for downregulation. The student's t-test was used to compare the different transcriptional values. The threshold was set as *p*-value < 0.01, fold-change = 2, gene rank < 10%.

$CD4^+$ and $CD8^+$ T cells, and regulatory T cells (Tregs), was filtered out to check for potential links with *HDAC1* expression.

GeneMANIA

Except for STRING, GeneMANIA (http://www.genemania.org) was also used for identifying possible *HDAC1*-related genes as candidate genes for further *HDAC1*-related gene signature construction (Franz et al., 2018). The network depicting genes functionally close to *HDAC1* was illustrated by recognizing patterns of gene co-annotation in the Gene Ontology or by using enrichment analysis.

Prognostic *HDAC1*-Related Gene Signature Development and Validation

The time-dependent receiver operating characteristic (ROC) analysis was conducted to test the predictive value for *HDAC1*, *HDAC2*, and *HDAC11*. Further, the *HDAC1*-related genes derived from GeneMANIA and STRING were regarded as predictive candidates for the *HDAC1*-related signature development. LASSO regression, univariate, and multivariate Cox regression analyses were consecutively applied to calculate the prognostic risk. Only genes with prognostic value and significance when fitting into the signature would be selected. The coefficients in the calculated regression model for each included gene were utilized to calculate the risk score, $Riskscore = \sum_{i=1}^{n} \beta_i \times gene_i$, where β represented the coefficient. The predictive nature of the signature components was represented by Hazard Ratio (HR) and 95% Confidence Interval (95%CI). The ROC analysis and survival analysis were performed for the signature discriminating capacity evaluation in the TCGA developing cohort and the GEO validation cohort.

Statistical Analysis

The statistical analyses and graphs associated with *HDAC1*-related gene signature development and validation were achieved in R 3.6.2 (R Core Team, 2019) and RStudio (version 1.1.463). The Least Absolute Shrinkage and

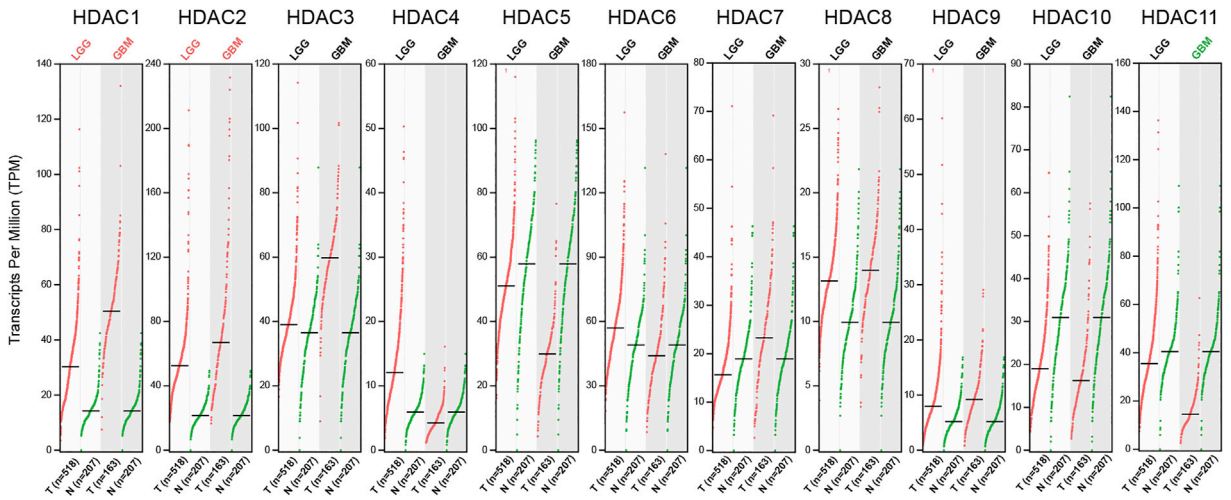


FIGURE 2 | RNA-seq data of HDAC family in LGG and GBM accessed from GEPIA. The Scatter plots showed expression comparisons of the HDAC family between the tumor and the normal groups. The statistically significant comparison with $|Log_2FC|$ over 1 was marked either in red (upregulation) or green (downregulation) ($q < 0.01$).

Selection Operator (LASSO) regression analysis, Cox regression analysis, and Kaplan-Meier survival analysis along with proportional hazards (PH) test were performed with the R package "survminer" and the R package "survival." Wilcoxon test and Kruskal Wallis test were used for statistical comparisons. In this study, a *p*-value less than 0.05 was regarded as statistically significant, and the *p*-values were adjusted with the "BH" method if involved with multiple hypothesis testing.

RESULTS

Differential Expression Analysis of HDAC Family Genes in Glioma

The transcriptional profiles of the HDAC family in brain and CNS cancers were investigated in the ONCOMINE database (**Figure 1**). It revealed that *HDAC1* and *HDAC6* were significantly over-expressed and within the top 5% gene rank in glioma compared to the normal group, which was evidenced by three and two studies, respectively. In contrast, under-expressed *HDAC4*, *HDAC5*, and *HDAC11* ranking within the top 1% gene rank were observed in glioma. It showed that two studies supported significantly overexpressed *HDAC2* in CNS cancer, but an equal number of studies concluded otherwise. The evidence with differential significance and top gene ranking for the rest of the HDAC family members remained scarce. The detailed study information regarding the HDAC family in ONCOMINE was summarized in **Table 1**.

HDAC1 mRNA over-expression in anaplastic oligoastrocytoma ($n = 4$) reached a fold change (FC) of 2.56 compared with the normal tissues ($n = 6$) ($p = 4.41E-04$) (French et al., 2005). In GBM samples ($n = 542$) from the TCGA cohort, *HDAC1* showed 3.13-fold increase of expression ($p = 3.08E-08$). Over-expressed *HDAC6* showed FCs of 3.22 and 2.61 for GBM ($n = 81$) ($p = 8.88E-13$) and anaplastic astrocytoma ($n = 19$) ($p = 2.34E-06$) groups, respectively (Sun et al., 2006).

Moreover, it was detected that the *HDAC4* expression in GBM ($n = 80$) was 2.15-fold lower than the control group ($n = 4$) ($p = 1.70E-13$) (Murat et al., 2008). The comparisons of *HDAC5* expression between anaplastic oligoastrocytoma ($n = 6$), anaplastic oligodendroglioma ($n = 3$), GBM ($n = 27$), and normal ($n = 4$) exposed a 2.10-fold ($p = 4.64E-05$), 2.44-fold ($p = 0.003$), 2.38-fold ($p = 1.46E-07$) decrease, respectively (Bredel et al., 2005). *HDAC5* mRNA level in GBM ($n = 81$) was downregulated with a FC of 2.22 in Sun's 2006 study ($p = 6.80E-20$), and a fold change of 2.74 in the GBM cohort ($n = 542$) from TCGA ($p = 3.08E-08$) (Sun et al., 2006). *HDAC11* displayed 4.20-fold downregulation in anaplastic oligoastrocytoma ($n = 4$) ($p = 1.09E-05$), and 3.93-fold downregulation in anaplastic oligodendroglioma ($n = 23$) in the French 2005 study ($p = 7.32E-06$) (French et al., 2005). Similarly, *HDAC11* was found to transcriptionally decrease with a FC of 3.54 in GBM ($n = 81$) ($p = 3.14E-14$) (Sun et al., 2006).

The studies focusing on *HDAC2* were of interest in ONCOMINE. One medulloblastoma study was filtered out. However, the rest of the three studies led to conflicting conclusions. As oligodendroglioma ($n = 3$) was compared with normal ($n = 7$) in Shai's 2003 study, *HDAC2* was upregulated and showed a fold change of 3.24 ($p = 6.99E-04$). The other two studies found that *HDAC2* significantly under-expressed in anaplastic oligoastrocytoma ($n = 6$) with a fold change of 4.40 ($p = 5.95E-04$), and in GBM ($n = 22$) with a fold change of 3.68 ($p = 2.70E-10$).

To fully investigate the HDAC family expression in glioma, the GEPIA database, which incorporated the GTEx and the TCGA data, was explored for verification. In the differentially expressed gene analysis, each HDAC family member was studied in both LGG ($n = 518$) and GBM ($n = 163$) subsets with thresholds of $|Log_2FC|$ over 1 and q-value less than 0.01. Of the entire HDAC family, only *HDAC1*, *HDAC2*, and

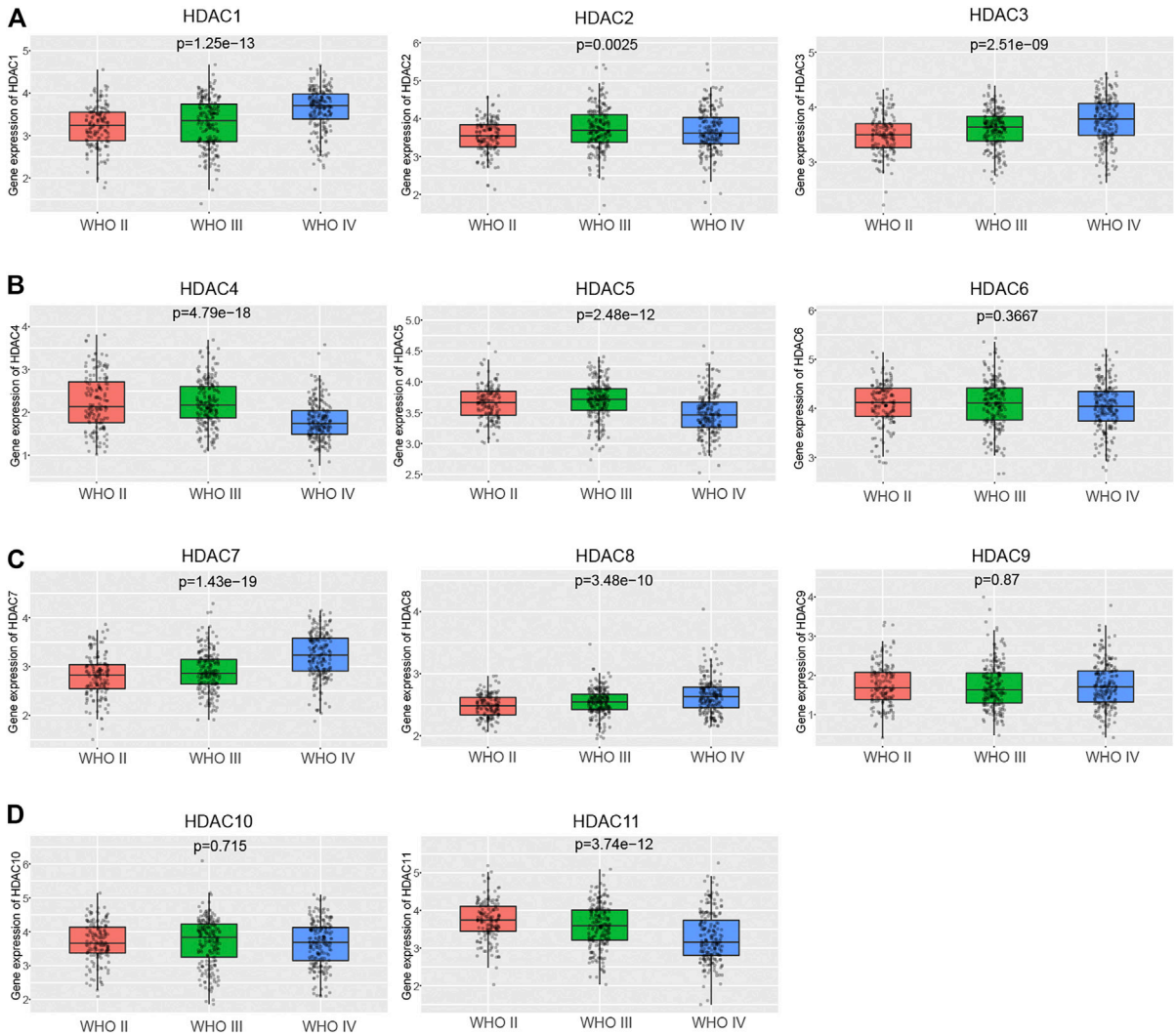


FIGURE 3 | Links between HDAC family and glioma grade in CGGA. **(A–D)** The box plots depicted that the expression of the HDAC family changed with glioma grade.

HDAC11 were tested to have more than a 2-fold alteration in expression with significance compared to normal tissue (**Figure 2**). *HDAC1* and *HDAC2* mRNA level was upregulated in both LGG and GBM, which was over 2-fold higher than the normal, whereas *HDAC11* expression level was more than 2-fold lower in the GBM group only.

Links Between HDAC Family and Glioma Grade, and Genetic Mutations

The relationship between the HDAC family and clinical features was studied. It was of priority to check whether the HDAC family's expression altered with glioma pathological grade using the glioma cohort in CGGA. Except for *HDAC6* (p = 0.367), *HDAC9* (p = 0.870), and *HDAC10* (p = 0.715), the majority of HDAC family expression levels correlated significantly with glioma grade (**Figure 3**). Specifically, the mRNA level of *HDAC1* (p = 1.25e−13), *HDAC3* (p = 2.51e−09), *HDAC7* (p = 1.43e−19), and *HDAC8* (p = 3.48e−10) showed an increasing trend with glioma grade progressing (**Figures 3A,C**). On the contrary, *HDAC11* (p = 3.74e−12) expression was negatively associated with glioma grade (**Figure 3D**). *HDAC2* (p = 0.0025) exhibited the highest expression level in the WHO III group but the lowest one in the WHO II group (**Figure 3A**). It also revealed that *HDAC4* (p = 4.79e−18) and *HDAC5* (p = 2.48e−12) expressed the most in the WHO III group but the least in the WHO IV group (**Figure 3B**).

Since the prognosis of glioma mainly depends on the pathological grade and genetic mutations, among which isocitrate dehydrogenase 1 (*IDH1*) and *1p/19q* status represented primary concern, further investigation was conducted in an attempt to research relations between HDAC family and *IDH1* and *1p/19q* based on the CGGA glioma cohort. The HDAC family, other than *HDAC6* (p = 0.187), *HDAC9* (p = 0.0526), and *HDAC10* (p = 0.91), are differentially expressed in the *IDH1* mutant and *IDH1* wildtype group with

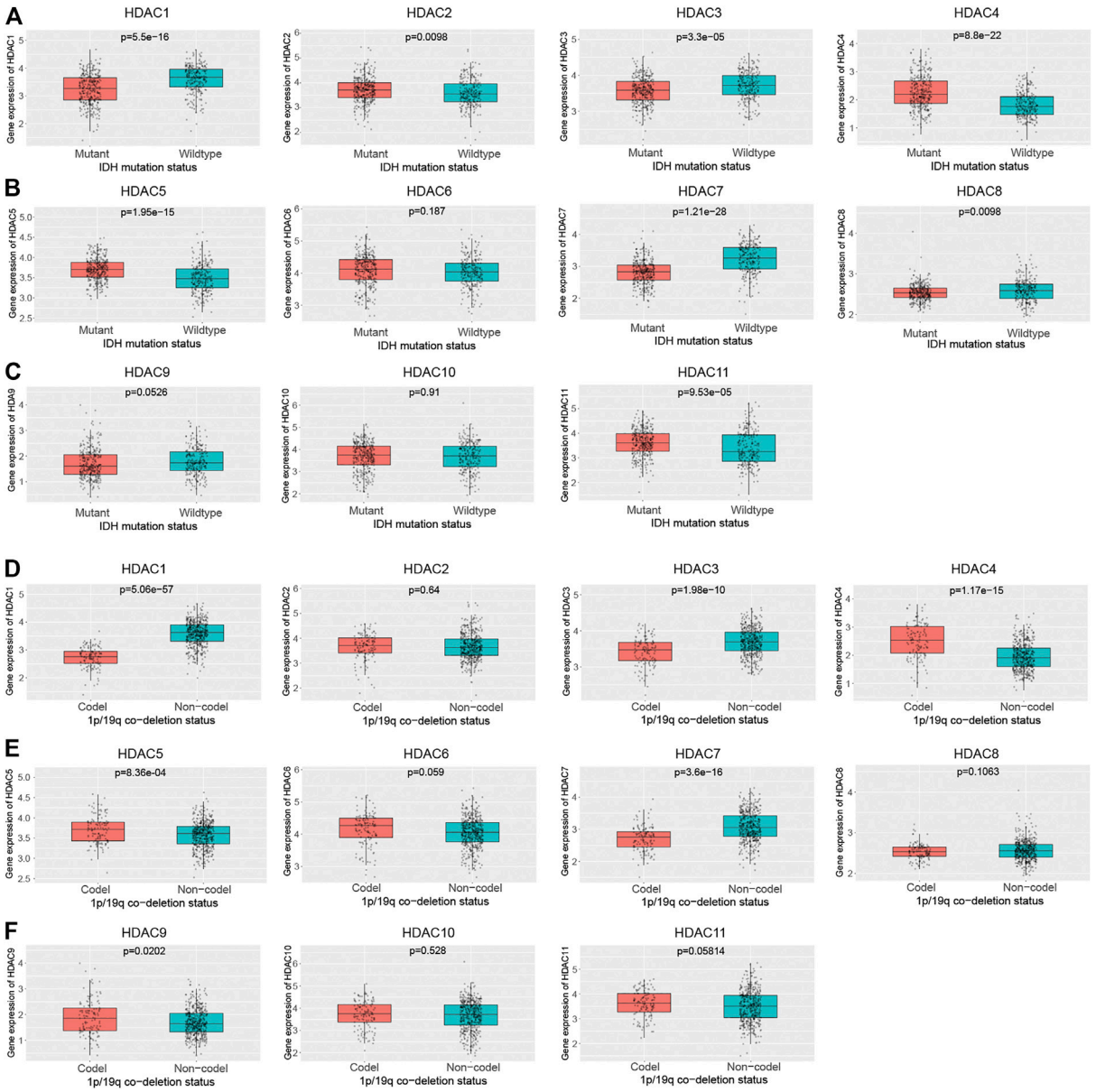


FIGURE 4 | Correlation between HDAC family and glioma genetic mutation in CCGA. **(A–F)** The box plots showed that the expression of the HDAC family varied according to the glioma mutations.

significance (**Figure 4**). As shown in the figure, *HDAC1* (p = 5.5e−16), *HDAC3* (p = 3.3e−05), *HDAC7* (p = 1.21e−28), and *HDAC8* (p = 0.0098) expression levels were significantly higher in the *IDH1* wildtype than the mutant (**Figures 4A–C**). However, *HDAC2* (p = 0.0098), *HDAC4* (p = 8.80e−22), *HDAC5* (p = 1.95e−15), and *HDAC11* (p = 9.53e−05) significantly expressed more transcripts in the *IDH1* mutant group compared to the wildtype (**Figures 4A–C**).

The HDAC family members were also characterized by *1p/19q* mutation (**Figure 4**). It turned out to have the same expression levels regarding the *1p/19q* co-deletion statuses in *HDAC2* (p = 0.64), *HDAC6* (p = 0.059), *HDAC8* (p = 0.106), and *HDAC10* (p = 0.528), and *HDAC11* (p = 0.0581) (**Figures 4D–F**). *HDAC1* (p = 5.06e−57), *HDAC3* (p = 1.98e−10), and *HDAC7* (p = 3.6e−16) showed higher expression levels in the *1p/19q* non-codeletion group in contrast to the codeletion group (**Figures 4D,E**). While *HDAC4* (p = 1.17e−15), *HDAC5* (p = 8.36e-04), and *HDAC9* (p = 0.0202) showed higher expression levels in the *1p/19q* codeletion group (**Figures 4D–F**).

Prognostic Characteristics of HDAC Family Genes in Glioma

Apart from the links with glioma grade and genetic mutations, survival analysis based on GEPIA (n = 676) and CGGA (n = 404)

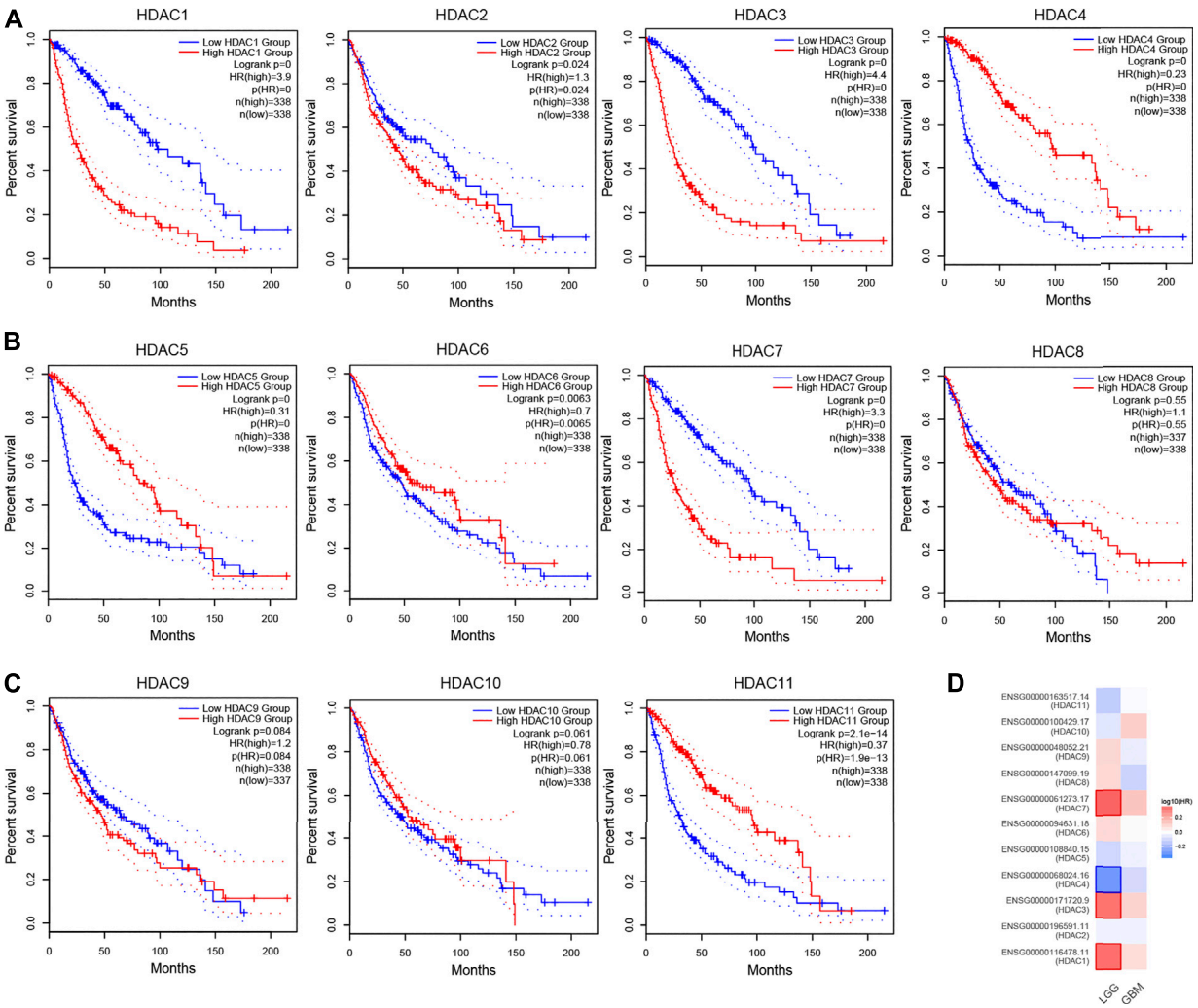


FIGURE 5 | Prognostic evaluation (OS) of HDAC family in glioma based on GEPIA. **(A–C)** The KM survival curves showed distinct survival possibility predicted by varying expression of HDAC family. **(D)** The heatmap was used for the link between the HDAC family members' expression and the OS of LGG and GBM patients.

data enabled the prognostic evaluation for the clinical value of the HDAC family. The survival analysis mainly focused on overall survival (OS) and disease-free survival (DFS).

The OS analysis based on GEPIA revealed that the glioma patients with high expression levels of *HDAC1* (HR:3.9, $p <$ 0.0001), *HDAC2* (HR:1.3, $p = 0.024$), *HDAC3* (HR:4.4, $p <$ 0.0001), and *HDAC7* (HR:3.3, $p < 0.0001$) would face with more risks compared to the ones with low expression of these genes (**Figures 5A,B**). However, the patients would reap OS benefits if they expressed high levels of *HDAC4* (HR:0.23, $p <$ 0.0001), *HDAC5* (HR:0.31, $p < 0.0001$), *HDAC6* (HR:0.7, $p =$ 0.0063), and *HDAC11* (HR:0.37, $p =$ 2.1e−14) (**Figures 5A–C**). The transcriptional levels of *HDAC8* ($p = 0.55$), *HDAC9* ($p =$ 0.084), and *HDAC10* ($p = 0.061$) imposed less influence on OS than the other HDAC family members (**Figures 5B,C**). Further research was performed to detect the impact that the HDAC family exerted on the OS of patients with LGG and GBM, in which *HDAC1*, *HDAC3*, and *HDAC7* remained the top three risk genes, whereas *HDAC4* was a remarkably protective gene (**Figure 5D**).

The survival analysis was also applied in the CGGA cohort to validate the prognostic value of the HDAC family. Only *HDAC1* ($p < 0.0001$), *HDAC3* ($p = 0.012$), *HDAC7* ($p < 0.0001$), and *HDAC8* ($p = 0.0056$) were tested with significance for a favorable prognosis if modestly expressed (**Figures 6A,B**). In addition, overexpressed *HDAC4* ($p < 0.0001$), *HDAC5* ($p < 0.0001$), *HDAC6* ($p = 0.0012$), and *HDAC11* ($p < 0.0001$) enabled the patients to survive longer (**Figures 6A–C**). The other HDACs showed neither survival benefits nor risks.

The DFS-oriented study using GEPIA data showed that more transcripts of *HDAC1* (HR:3.0, $p < 0.0001$), *HDAC3* (HR:2.7, $p =$ 2.1e−15), *HDAC7* (HR:2.1, $p =$ 3e−09), and *HDAC9* (HR:1.5, $p =$ 0.0026) accompanied with less DFS probabilities in glioma patients (**Figures 7A–C**). However, extended DFS would be observed in patients with high expression of *HDAC4* (HR:0.35, $p =$ 1.1e−16), *HDAC5* (HR:0.43, $p =$ 9.4e−12), *HDAC6* (HR:0.73,

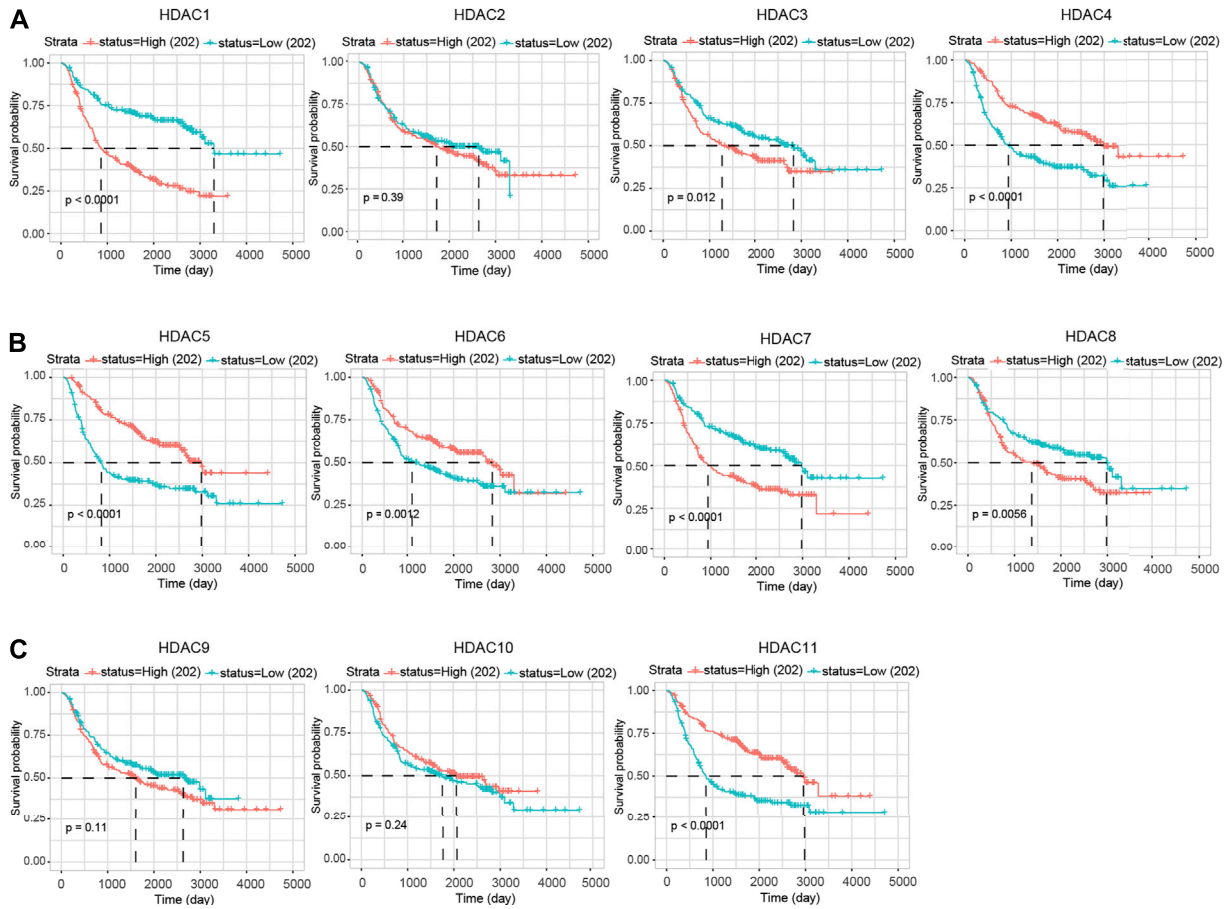


FIGURE 6 | Prognostic feature (OS) of HDAC family in glioma based on CGGA. **(A–C)** The KM survival analysis was performed for verification regarding the prediction efficiency of the HDAC family.

p = 0.014), *HDAC10* (HR:0.72, p = 0.0092), and *HDAC11* (HR: 0.5, p = 4.8e−08). The rest of the HDAC family displayed no risks or benefits for DFS in glioma patients (**Figures 7A–C**). Similarly, *HDAC1*, *HDAC3*, and *HDAC7* accounted for the major risk factors for LGG, but *HDAC4* and *HDAC5* enhanced DFS probabilities (**Figure 7D**).

Genetic Mutations, Interactive Network, and Functional Enrichment Analysis of HDAC Family

The genetic mutations of the HDAC family were analyzed with the TCGA data available at the cBioPortal database. Each of the HDAC family members harbored genetic mutations, in which *HDAC4* was the most prominent with a mutation ratio of 2.6% (**Figure 8A**). It was followed by the mutation ratio of *HDAC9*, *HDAC10*, and *HDAC6* being 1.5, 1.5, and 1.3%, respectively (**Figure 8A**). Interestingly, the mutation ratio of *HDAC1* and *HDAC11* were the same at 0.5% (**Figure 8A**). In terms of mutation statuses in the detailed glioma subtypes, astrocytoma and GBM shared a similar distribution of mutation patterns. However, oligoastrocytoma only showed three kinds of genetic alterations, "mutation," "deep deletion," and "multiple alterations," and oligodendroglioma harbored one more type compared to oligoastrocytoma, "amplification" (**Figure 8B**).

Further, interaction analysis was performed in the HDAC family using the STRING web tool. Beyond the HDAC family were *MTA1*, *MTA2*, *RBBP7*, *NCOR2*, and *ANKRA2* that were incorporated in the PPI network (**Figure 8C**). The functional enrichment annotated the interactive network mainly as "histone deacetylation" (**Figure 8D**). Nevertheless, enrichment results, such as "oligodendrocyte differentiation," "histone methyltransferase," and "heat shock protein binding," were also noticeable.

Associate Between Prognostic *HDAC1* and Immune Infiltration in Glioma

Based on the results summarized in **Table 2**, *HDAC1*, *HDAC2*, and *HDAC11* were selected as the predictive genes with significant clinical value considering their significant and consistent performance in the previous analyses. ROC analysis was applied to assess the

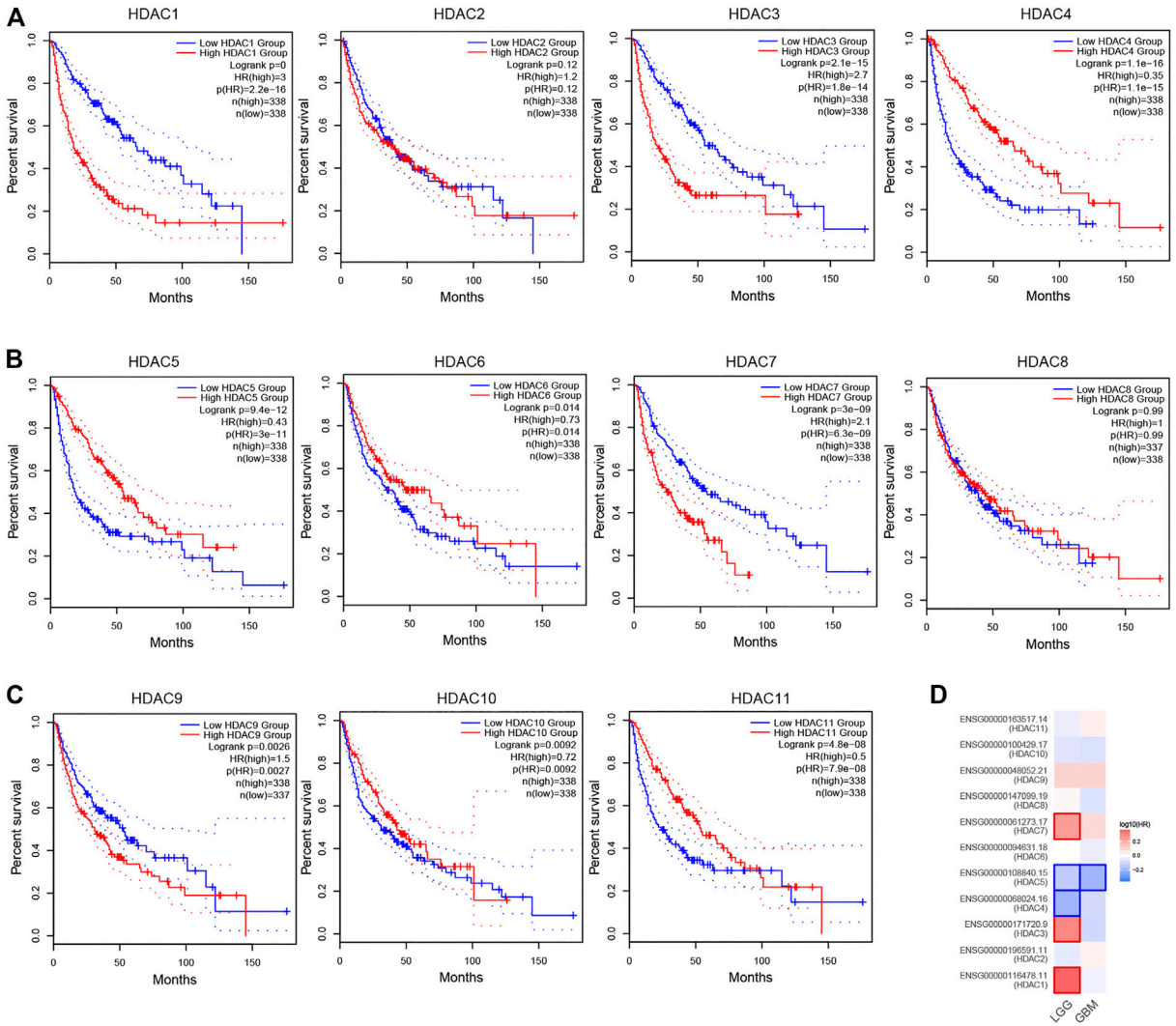


FIGURE 7 | Prognostic value (DFS) of HDAC family in glioma based on GEPIA. **(A–C)** The KM survival curves revealed contrasting survival possibility predicted by varying expression of HDAC family. **(D)** The heatmap was used for detecting the relations between the HDAC family members' expression and the DFS of LGG and GBM patients.

discrimination efficiency of the three genes in the TCGA glioma cohort. The separability of the *HDAC1* turned out to be valid, reaching 0.70, 0.77, 0.75, and 0.80 for 1-, 3,- 5-, and 10-years OS predictions in glioma, respectively (**Figure 9A**). However, the area under the ROC curve (AUC) of the *HDAC2* and the *HDAC11* was calculated with no value exceeding 0.5 throughout the time points in the study (**Figures 9B,C**).

Immune infiltration presumably involved with glioma progression and prognosis. Since *HDAC1* was filtered out to be the prognostic gene for glioma, the association between the immune infiltration and the *HDAC1* expression was analyzed by the TIMER database. Considering the limited immune cells in the CNS system due to the BBB, neutrophils, macrophages, MDSCs, $CD4^+$ T cells, $CD8^+$ T cells, and Tregs were regarded as the major infiltrated immune cell types in glioma for the evaluation. The correlation test was categorized by LGG and GBM.

It revealed that each result of the immune correlation analysis was of significance ($p < 0.0\ 1$) but a relatively strong relation ($r > 0.5$) was limited either in LGG or GBM (**Figures 9D,E**). Intriguingly, *HDAC1* mRNA level was relatively closely and positively tied to neutrophil infiltration in the LGG group (r = 0.60, p = 2.88e-47) (**Figure 9D**). The *HDAC1* expression also positively associated with $CD4^+$ T cell infiltration (r = 0.52, p = 3.96e-35) (**Figure 9D**) in the LGG group. The negative relation between $CD8^+$ T cell infiltration and *HDAC1* expression level was noticeable (r = -0.27, p = 1.17e-09) (**Figure 9D**). In the GBM group, the strongest relationship was between macrophage and *HDAC1* expression (r = 0.19, p = 2.81e-01) (**Figure 9E**). Additionally, *HDAC1* was used to explore the potential *HDAC1*-related genes in the GeneMANIA (**Figure 9F**). The

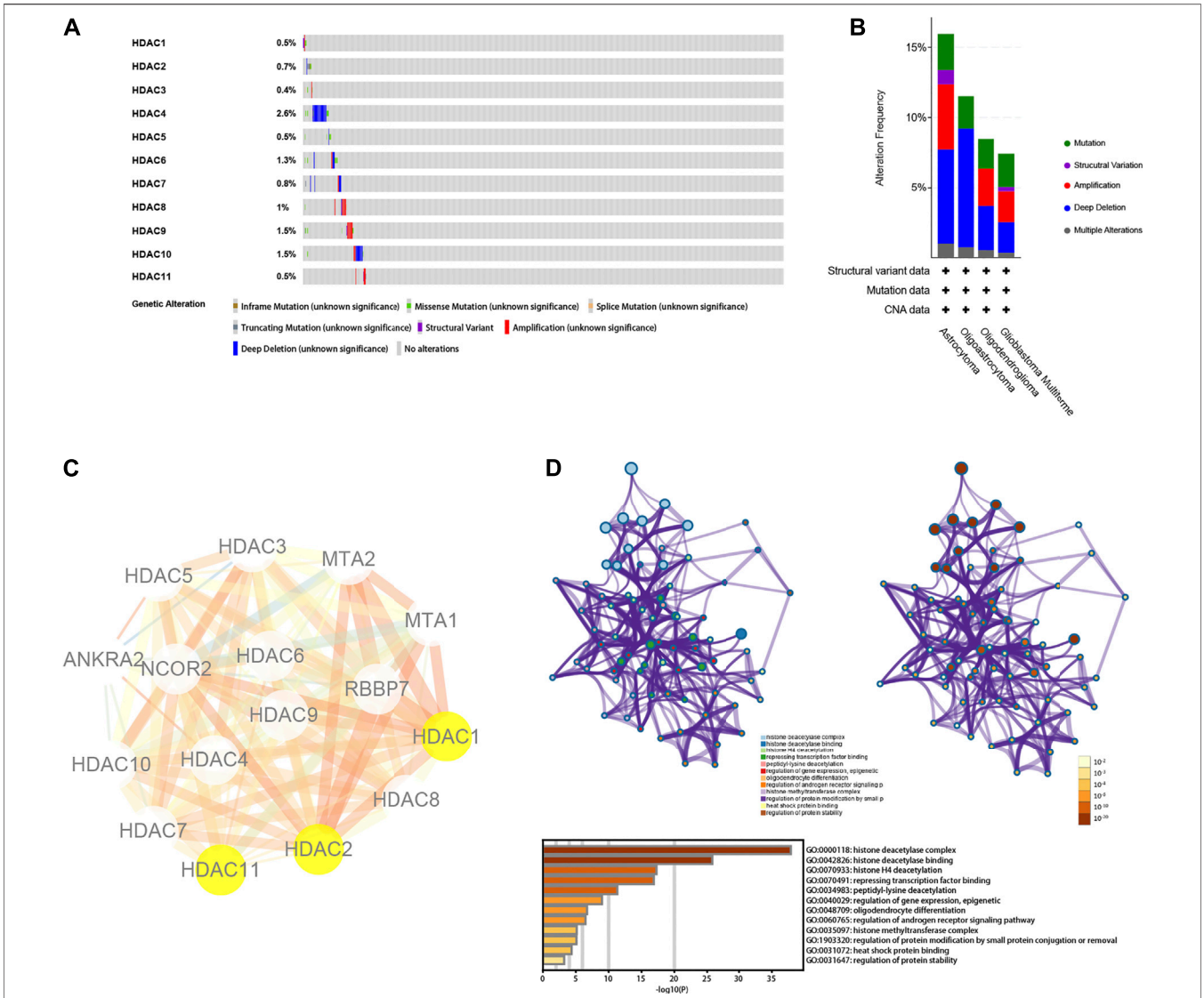


FIGURE 8 | Genetic mutation, interactive network, and functional enrichment of HDAC family. **(A,B)** Summary for genetic alterations of HDAC family in glioma. **(C)** Protein-protein interactive network of HDAC family by STRING. The size of the edges in the network was set to change according to the co-expression value among the nodes. The color of the edges in the network altered with the combined scores from the low to the high, which were all the evidence scores (including transferred scores). **(D)** Functional enrichment plot and annotation by Metascape.

entire *HDAC1*-centered interactive network (**Figure 9F**) together with supplementary genes acquired from the STRING in the same way as the GeneMANIA, the candidate risk genes for *HDAC1*-related signature construction, were thus recruited. The *HDAC1*-related genes were listed in **Supplementary Table S1**.

Development and Validation of *HDAC1*-Related Gene Signature

Fifteen genes were screened out of the initial *HDAC1*-related genes through LASSO regression analysis (**Supplementary Figure S1**). Univariate and multivariate Cox regression analyses were then performed for the signature development (**Supplementary Table S2**). The *HDAC1*-related gene signature was ultimately constructed in the TCGA cohort and comprised of five promising prognostic genes: *HDAC1* (HR: 1.4938, 95%CI:1.1997–1.8600), *RUNX1T1* (HR:0.7349, 95%CI: 0.6483–0.8330), *FKBP3* (HR:0.6382, 95%CI:0.4782–0.8517), *RBL1* (HR:1.7148, 95%CI:1.3826–2.1268), and *PHF21A* (HR: 0.4956, 95%CI:0.3874–0.6340) (**Figure 10A**). The signature model satisfied the proportional hazards (PH) assumption (**Supplementary Figure S2**). And it was also adjusted and tested with other prognostic factors, including gender, grade, *IDH1* mutation, and *1p/19q* codeletion, to be an independent prognostic indicator (Riskscore, HR:1.0935, 95%CI: 1.0045–1.1904, p = 0.039) (**Supplementary Figure S3**). The signature risk score was calculated as follows: Riskscore =

TABLE 2 | Summary of the HDAC family analysis overall result significance *p*-value.

	HDAC1	HDAC2	HDAC3	HDAC4	HDAC5	HDAC6	HDAC7	HDAC8	HDAC9	HDAC10	HDAC11
Gene expression (glioma vs. control)											
ONCOMINE	3.08e-08	2.70e-10	—	1.70e-13	6.80e-20	8.88e-13	—	—	—	—	3.14e-14
GEPIA (|Log2FC|>1)	<0.0100	<0.0100	>0.0100	>0.0100	>0.0100	>0.0100	>0.0100	>0.0100	>0.0100	>0.0100	(GBM) < 0.01
Clinical trait correlation											
CGGA (WHO grade II - III - IV)	1.25e–13	0.0025	2.51e–09	4.79e–18	2.48e–12	0.3670	1.43e–19	3.48e–10	0.8700	0.7150	3.74e–12
CGGA (*IDH1* mutation - *IDH1* wildtype)	5.50e–16	0.0098	3.30e–05	8.80e–22	1.95e–15	0.1870	1.21e–28	0.0098	0.0526	0.9100	9.53e–05
CGGA (*1p/19q* codeletion - non-codeletion)	5.06e–57	0.6400	1.98e–10	1.17e–15	8.36e-04	0.0590	3.60e–16	0.1063	0.0202	0.5280	0.0580
Survival analysis (high-expression vs. low-expression)											
GEPIA (Overall survival)	<0.0001	0.0240	<0.0001	<0.0001	<0.0001	0.0063	<0.0001	0.5500	0.0840	0.0610	2.10e–14
CGGA (Overall survival)	<0.0001	0.3900	0.01200	<0.0001	<0.0001	0.0012	<0.0001	0.0056	0.1100	0.2400	<0.0001
GEPIA (Disease Free survival)	<0.0001	0.1200	2.10e–15	1.10e–16	9.40e–12	0.01400	3.00e–09	0.9900	0.0026	0.0092	4.80e–08

0.4013×*HDAC1*-0.3081×*RUNX1T1*-0.4491×*FKBP3*+0.5393×*RBL1*-0.7020×*PHF21A*.

The expression profiles of the *HDAC1*-related signature components were investigated in the TCGA glioma cohort (**Figure 10B**). *HDAC1* and *RBL1* were upregulated with the increasing risk score, while the expression levels of *FKBP3*, *RUNX1T1*, and *PHF21A* reversed (**Figure 10B**). To validate the prognostic efficiency, the signature was tested by survival analysis and ROC curve in the TCGA developing cohort. It showed that the high-risk group which scored high according to the equation held poorer survival probabilities than the low-risk ($p < 0.0001$) (**Figure 10C**). Additionally, the discrimination of the signature was measured by the AUC being 0.800, 0.841, 0.829, and 0.882 for 1-, 3-, 5- and 10-years OS, respectively (**Figure 10D**).

The signature was also validated internally and externally. As for internal validation, the TCGA developing cohort was randomly divided into two sets, one for training containing 30% of the cohort cases, and the other as the testing set. It classified the data with the 1-year OS AUC being 0.860, the 3-years OS AUC being 0.844, and the 5-years OS AUC being 0.816 in the training set (**Figure 10E**). Moreover, the 1-year OS AUC being 0.778, the 3-years OS AUC being 0.836, and the 5-years OS AUC being 0.832 were measured in the testing set (**Figure 10F**). In the GEO validation cohort used for external validation, the high-risk group consistently had a higher survival risk compared to the low one ($p < 0.0001$) (**Figure 11A**). And the discrimination of the signature was confirmed in the GEO cohort regarding being 0.695, 0.834, 0.831, and 0.826 for 1-, 3-, 5-, and 10-years OS, respectively (**Figure 11B**). The signature was annotated as "chromatin binding" by the functional enrichment analysis in Metascape (**Figure 11C**). The clinical baseline information of the TCGA developing cohort and the GEO validation cohort were listed in **Supplementary Table S3**.

DISCUSSION

Differentially expressed HDAC family members were identified with significance in ONCOMINE, namely the upregulated *HDAC1* and *HDAC6*, as well as the downregulated *HDAC4*, *HDAC5,* and *HDAC11* in glioma (**Figure 1**). After ruling out the unrelated medulloblastoma study, the contradictory expression profiles in the three studies on *HDAC2* resulted in a perplexing conclusion, of which one indicated overexpression while the other two opposed. However, the study size of the three studies was too limited to conclude firmly regarding *HDAC2* expression level in glioma. The Lee 2006 study recruiting 22 cases of GBM which was larger than the size of the other two *HDAC2* studies might suggest that *HDAC2* was under-expressed in glioma (**Table 1**).

Further validation of the HDAC family differential expression was conducted in GEPIA. Only the comparison with a change over 2-fold and q-value less than 0.01 was marked in the exhibition (**Figure 2**). The *HDAC1* and *HDAC2* comparisons in glioma, and the *HDAC11* comparison in GBM, were remarkable for their significantly contrasting expression. The remaining groups were neither of significance nor holding differential changes exceeding 2-fold. A previous study composed of 20 low-grade and 23 high-grade glioma patients concluded that class II and IV HDACs expressed less in GBM compared to low-grade glioma and normal tissue (Lucio-Eterovic et al., 2008; Williams et al., 2017). However, it seemed too bold and assertive to firmly conclude with such a limited sample size.

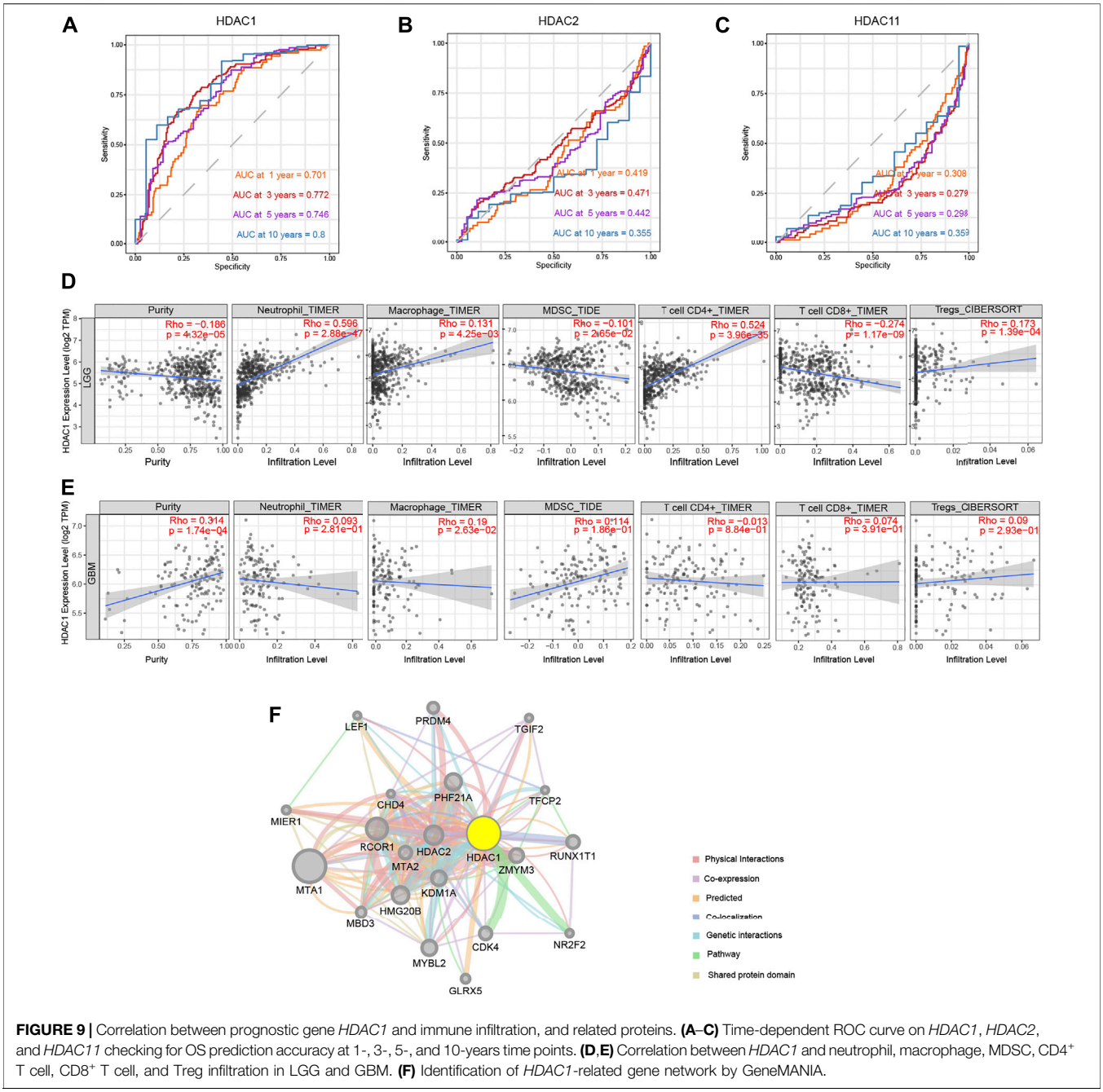


FIGURE 9 | Correlation between prognostic gene *HDAC1* and immune infiltration, and related proteins. **(A–C)** Time-dependent ROC curve on *HDAC1*, *HDAC2*, and *HDAC11* checking for OS prediction accuracy at 1-, 3-, 5-, and 10-years time points. **(D,E)** Correlation between *HDAC1* and neutrophil, macrophage, MDSC, $CD4^+$ T cell, $CD8^+$ T cell, and Treg infiltration in LGG and GBM. **(F)** Identification of *HDAC1*-related gene network by GeneMANIA.

Besides, the single-centered data might be insufficient and unpersuasive since the differential expression results from multiple sources in this study only agree with a few genes to be further analyzed. Considering the findings from ONCOMINE, GEPIA, and CGGA, *HDAC1*, *HDAC2*, and *HDAC11* were thus verified for their differential expression in glioma (**Figures 1**, **2**, **3**; **Table 2**).

It was assumed that the differentially expressed HDAC family might contribute to the clinical and genetic features of glioma. The investigation on the potential links between the HDAC family and the pathological grade, and the genetic alterations of glioma, revealed that the expression levels of *HDAC1*, *HDAC2*, and *HDAC11* significantly altered with the glioma grade (*HDAC1*, p = 1.25e−13; *HDAC2*, p = 0.0025; *HDAC11*, p = 3.74e−12) and the *IDH1* mutation status (*HDAC1*, p = 5.5e−16; *HDAC2*, p = 0.0098; *HDAC11*, p = 9.53e−05) (**Figures 3A,D**; **Figures 4A,C**; **Table 2**). The transcriptional level of *HDAC1* (p = 5.06e−57), but not *HDAC2* (p = 0.64) and *HDAC11* (p = 0.058), significantly varied with the *1p/19q* mutation (**Figures 4D,E**).

Given that low pathological grade, *IDH1* mutant, and *1p/19q* codeletion were favorable prognostic factors, the fact that *HDAC1*

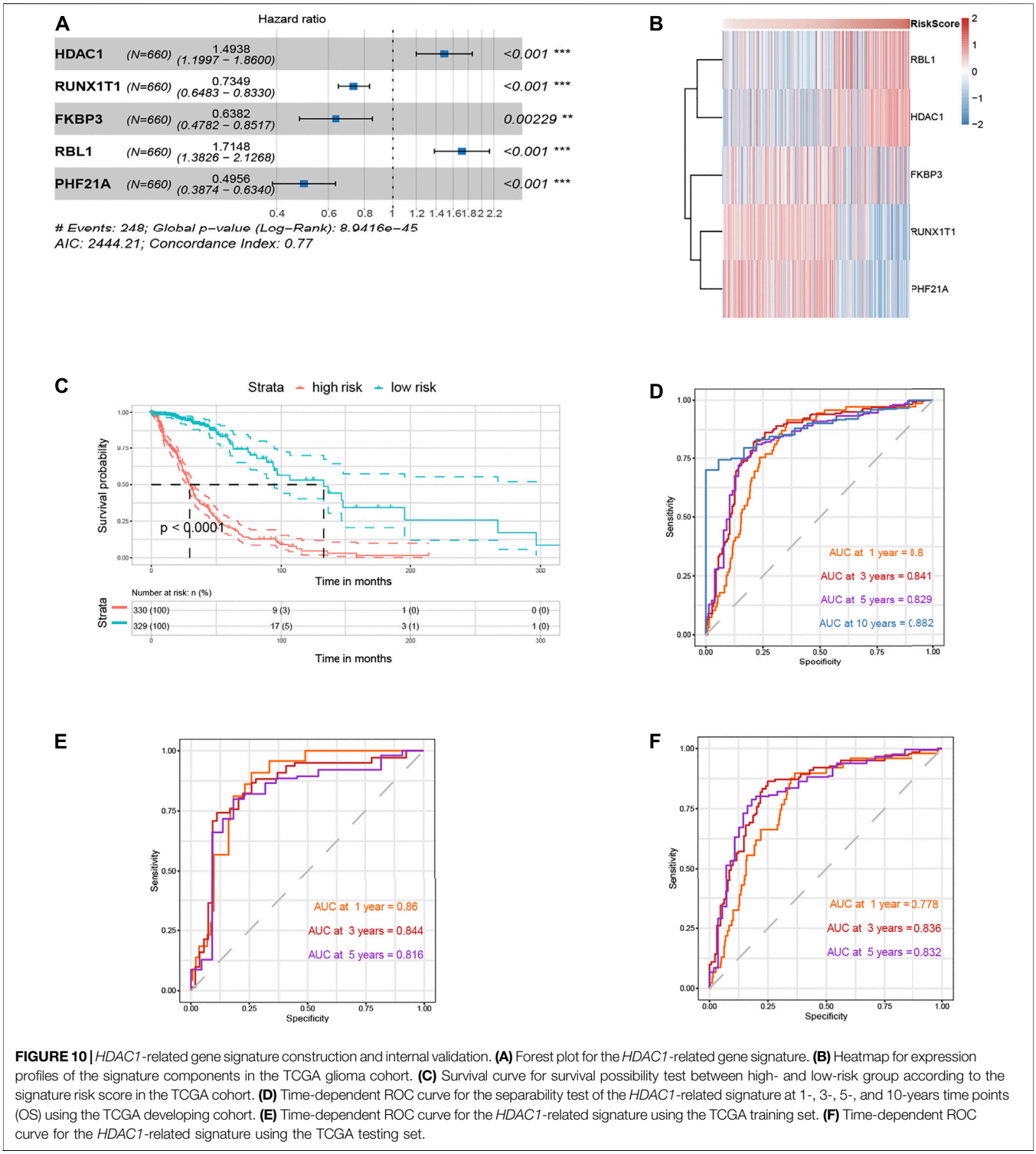


FIGURE 10 | *HDAC1*-related gene signature construction and internal validation. **(A)** Forest plot for the *HDAC1*-related gene signature. **(B)** Heatmap for expression profiles of the signature components in the TCGA glioma cohort. **(C)** Survival curve for survival possibility test between high- and low-risk group according to the signature risk score in the TCGA cohort. **(D)** Time-dependent ROC curve for the separability test of the *HDAC1*-related signature at 1-, 3-, 5-, and 10-years time points (OS) using the TCGA developing cohort. **(E)** Time-dependent ROC curve for the *HDAC1*-related signature using the TCGA training set. **(F)** Time-dependent ROC curve for the *HDAC1*-related signature using the TCGA testing set.

overexpressed in the WHO III and WHO IV group (**Figure 3A**), in the *IDH1* wildtype group (**Figure 4A**), and the *1p/19q* non-codeletion (**Figure 4D**) implied that *HDAC1* tended to be a detrimental prognostic biomarker in glioma (Lapointe et al., 2018). It also suggested the role of *HDAC11* as a protective factor concerning its opposite expression pattern relative to *HDAC1* (**Figures 3D**, **4C**,F). However, *HDAC2* expressed the most in the WHO III group but the least in the WHO IV group, indicating a baffling part *HDAC2* played in the glioma progression (**Figure 3A**). It was still reasonable that *HDAC2* was involved with the *IDH1* mutation (**Figure 4A**).

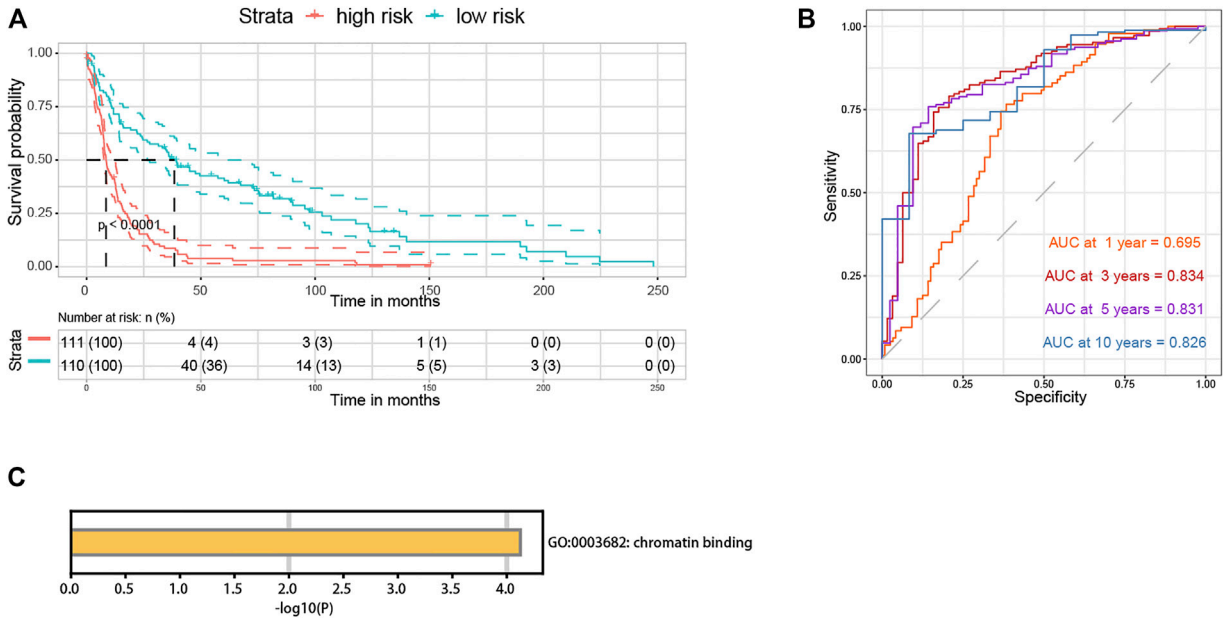


FIGURE 11 | *HDAC1*-related gene signature external validation and enrichment analysis. **(A)** Survival curve for survival possibility test between high- and low-risk group according to the signature risk score in the GEO cohort. **(B)** Time-dependent ROC curve for the separability test of the *HDAC1*-related signature at 1-, 3-, 5-, and 10-years time points (OS) using the GEO validation cohort. **(C)** Functional enrichment analysis of the *HDAC1*-related signature.

Consistent with the earlier results and assumption, the OS-oriented survival analysis in both the GEPIA (n = 676) and the CGGA (n = 404) showed that *HDAC1* overexpression (GEPIA, HR:3.9, $p < 0.0001$; CGGA, $p < 0.0001$) brought risks to glioma patients (**Figures 5A**, **6A**), while *HDAC11* served as a favorable prognostic indicator (GEPIA, HR:0.37, p = 2.1e−14; CGGA, $p <$ 0.0001) (**Figures 5C**, **6C**). Although *HDAC2* (HR:1.3, p = 0.024) significantly distinguished the group with poorer prognosis in GEPIA (**Figure 5A**), it failed to impose any disadvantages to the *HDAC2*-overexpressed group (p = 0.39) when validated (**Figure 6A**). Moreover, *HDAC1* (HR:3.0, $p < 0.0001$) acted as a hazard predictive biomarker, and *HDAC11* (HR:0.5, p = 4.8e−08) was still a protective factor in the DFS analysis (**Figures 7A,C**). The relatively low mutation ratio of *HDAC1* and *HDAC11* probably suggested the stability of the genes (**Figure 8A**).

In contrast to less efficient discrimination of *HDAC2* and *HDAC11*, the AUC value of *HDAC1* over 0.7 throughout the time points manifested that *HDAC1* was identified to be a promising prognostic biomarker for glioma (**Figures 9A–C**; **Table 2**). It was found recently that knockdown of *HDAC1* with siRNA reduced LN18 GBM cell proliferation, leaving cell viability unaffected (Was et al., 2019). The result that simultaneous inhibition of *HDAC1* and *HDAC2* led to a significant drop in GBM cell proliferation synergistically suggested an efficient combination anticancer strategy (Was et al., 2019). More in-depth research found the involvement of *HDAC1* together with *HDAC2* in the regulation of a transcription factor *c-Myc* (Nguyen et al., 2020). The selective and broad inhibition of *HDAC1* and *HDAC2* disrupted *c-Myc* regulation on aerobic glycolysis enhancing oxidative metabolism, followed by peroxisome proliferator-activated receptor γ coactivator1 α (PGC1α) and peroxisome proliferator-activated receptor δ (PPARδ), thereby extending overall survival of patient-derived xenograft (Nguyen et al., 2020). The additional mechanism of *HDAC1*-involved invasive and proliferative phenotype in GBM cells could be attributed to the interaction between *HDAC1* and phosphorylated special AT-rich sequence-binding protein 1 (SATB1) (Han et al., 2013).

It also exposed the moderate relationship between *HDAC1* expression and neutrophil infiltration (r = 0.60, p = 2.88e-47) in LGG, as well as $CD4^+$ T cell infiltration (r = 0.52, p = 3.96e-35) (**Figure 9D**). Although the studies on how the *HDAC1* expression predicted immune infiltration in LGG were urgently needed, it was demonstrated that IFN-β silenced interleukin-8 (IL-8) transcription by increasing *HDAC1* expression level in GBM cells (Nozell et al., 2006). Given that IL-8 was one of the chemokines that recruited migrating neutrophils, impaired IL-8 release likely resulted in reduced neutrophil infiltration, which might explain the absence of relation between *HDAC1* and neutrophil infiltration in the GBM.

The components of the final *HDAC1*-related signature could be the reason why the separability of the signature exhibited robust efficiency (1-year OS AUC = 0.80, 3-years OS AUC = 0.84, 5-years OS AUC = 0.829, 10-years OS AUC = 0.882) (**Figures 10D,E**). In the signature, *HDAC1* (HR:1.4938) and *RBL1* (HR: 1.7148) were deemed a hazard, while *RUNX1T1* (HR:0.7349), *FKBP3* (HR:0.6382), and *PHF21A* (HR:0.4956) promoted survival possibilities for glioma patients (**Figure 10A**). The

HDAC1-related signature was adjusted as an independent risk indicator and eventually validated both internally and externally showing reliable discrimination for glioma prognosis prediction (**Supplementary Figure S3**; **Figures 10D–F**; **Figures 11A,B**).

Of note, *HDAC1* was transcriptionally regulated by the nuclear factor of activated T cell (NFAT), which played a role in glioma stem cell growth and mesenchymal transition (Song et al., 2020). Retinoblastoma transcriptional corepressor like 1 (RBL1), known for its modulation in the G1/S cell cycle, behaved oppositely and functioned as a tumor suppressor in a GBM model, conflicts of which could come from either the species' differences or some other regulations unidentified (Naert et al., 2020). RUNX1 partner transcriptional co-repressor 1 (RUNX1T1) earned prestige for its fusion with Runt-related transcription factor 1 (RUNX1) in acute myeloid leukemia (Beghini, 2019). Related to HDAC class I signaling, FK506 Binding Protein 3 (FKBP3) regulated *HDAC2* expression contributing to the drug resistance in tumor cells (Tong et al., 2019). PHD finger protein 21A (PHF21A) was mainly studied in neurodevelopmental disorders (Vallianatos and Iwase, 2015). Though the research on *RUNX1T1*, *FKBP3*, and *PH21A* in glioma was definite, the signature was annotated as "chromatin binding" by the functional enrichment analysis (**Figure 10E**).

However, the prognostic *HDAC1* per se and the *HDAC1*-related gene signature were yet to be employed and verified universally. Considering some potential barriers when promoting sequencing techniques in clinical oncology testing, it would require more time for the real arrival of the molecular prognostic biomarker era. In addition, it remains a question to be solved regarding how *HDAC1* was involved with the glioma progression and the modulation of immune infiltration in glioma. Acetylation and deacetylation were assumed to stand downstream in the HDAC-involved glioma modulation and thereby possibly influence glioma prognosis, which also warranted large-scale data to testify in the future. How epigenetics like methylation affected glioma through regulation of the HDACs remains another interesting topic.

CONCLUSION

Conclusively, the comprehensive analysis of the HDAC family exposed that the HDAC family components were of prognostic significance for glioma. Overexpressed *HDAC1* in glioma positively altered its expression with glioma grade, *IDH1* wildtype, and *1p/19q* non-codeletion, while *HDAC11* was downregulated and acted oppositely compared to *HDAC1*. The majority of the HDAC family significantly correlated with the grade, *IDH1* mutation, and *1p/19q* codeletion in glioma. *HDAC1* was manifested to serve as a prognostic biomarker for glioma, and an indicator for neutrophil and $CD4^+$ T cell infiltration in LGG. Ultimately, the *HDAC1*-centered signature composed of the *HDAC1*-related genes was developed and validated for precise prognosis prediction in glioma. It was tempting to speculate an optimistic future for glioma not only by implementing *HDAC1* and other HDACs as prognostic biomarkers but also by targeting *HDAC1* and its closely related genes.

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

The study and conception were designed by YF, YW, and GZ; the coding and data analysis was done by YF, XP, and YW; figures and tables were prepared by YF, XP, and BL; the manuscript was drifted by YF; reviewing and editing was completed by YW and GZ. All authors contributed to the manuscript revision. All authors read and approved the final version of the manuscript.

FUNDING

We acknowledge the funding support from the National Natural Science Foundation of China (Grant Number: 81772684, Beijing, China).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.720020/full#supplementary-material

Supplementary Figure S1 | LASSO regression of the HDAC1-related genes. The trajectory plot on the left panel was for independent variables sourced from the HDAC1-related genes. The plot on the right panel showed that each lambda and confidence interval. LASSO, least absolute shrinkage, and selection operator.

Supplementary Figure S2 | Candidate HDAC1-related genes for the signature under proportional hazard assumption.

Supplementary Figure S3 | The HDAC1-related signature adjusted to be an independent risk indicator.

Supplementary Table S1 | HDAC1-related genes collected from STRING and GeneMANIA.

Supplementary Table S2 | HDAC1-related genes with significance in univariate and multivariate Cox regression analyses.

REFERENCES

Beghini, A. (2019). Core Binding Factor Leukemia: Chromatin Remodeling Moves towards Oncogenic Transcription. *Cancers* 11 (12), 1973. doi:10.3390/cancers11121973

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

Bredel, M., Bredel, C., Juric, D., Harsh, G. R., Vogel, H., Recht, L. D., et al. (2005). Functional Network Analysis Reveals Extended Gliomagenesis Pathway Maps and Three Novel MYC-Interacting Genes in Human Gliomas. *Cancer Res.* 65 (19), 8679–8689. doi:10.1158/0008-5472.CAN-05-1204

Chen, R., Zhang, M., Zhou, Y., Guo, W., Yi, M., Zhang, Z., et al. (2020). The Application of Histone Deacetylases Inhibitors in Glioblastoma. *J. Exp. Clin. Cancer Res.* 39 (1), 138. doi:10.1186/s13046-020-01643-6

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

Eyüpoglu, I. Y., and Savaskan, N. E. (2016). Epigenetics in Brain Tumors: HDACs Take Center Stage. *Cn* 14 (1), 48–54. doi:10.2174/1570159x13666151030162457

Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., et al. (2018). GeneMANIA Update 2018. *Nucleic Acids Res.* 46 (W1), W60–W64. doi:10.1093/nar/gky311

French, P. J., Swagemakers, S. M. A., Nagel, J. H. A., Kouwenhoven, M. C. M., Brouwer, E., van der Spek, P., et al. (2005). Gene Expression Profiles Associated with Treatment Response in Oligodendrogliomas. *Cancer Res.* 65 (24), 11335–11344. doi:10.1158/0008-5472.CAN-05-1886

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. Signaling* 6 (269), pl1. doi:10.1126/scisignal.2004088

Gravendeel, L. A. M., Kouwenhoven, M. C. M., Gevaert, O., de Rooi, J. J., Stubbs, A. P., Duijm, J. E., et al. (2009). Intrinsic Gene Expression Profiles of Gliomas Are a Better Predictor of Survival Than Histology. *Cancer Res.* 69 (23), 9065–9072. doi:10.1158/0008-5472.CAN-09-2307

Han, S., Xia, J., Qin, X., Han, S., and Wu, A. (2013). Phosphorylated SATB1 Is Associated with the Progression and Prognosis of Glioma. *Cell Death Dis* 4, e901. doi:10.1038/cddis.2013.433

Kuhn, M. (2008). Building Predictive Models inRUsing thecaretPackage. *J. Stat. Soft.* 28 (5), 1–26. doi:10.18637/jss.v028.i05

Kunadis, E., Lakiotaki, E., Korkolopoulou, P., and Piperi, C. (2021). Targeting post-translational Histone Modifying Enzymes in Glioblastoma. *Pharmacol. Ther.* 220, 107721. doi:10.1016/j.pharmthera.2020.107721

Lapointe, S., Perry, A., and Butowski, N. A. (2018). Primary Brain Tumours in Adults. *The Lancet* 392 (10145), 432–446. doi:10.1016/s0140-6736(18)30990-5

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 (W1), W509–W514. doi:10.1093/nar/gkaa407

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Lucio-Eterovic, A. K., Cortez, M. A., Valera, E. T., Motta, F. J., Queiroz, R. G., Machado, H. R., et al. (2008). Differential Expression of 12 Histone Deacetylase (HDAC) Genes in Astrocytomas and normal Brain Tissue: Class II and IV Are Hypoexpressed in Glioblastomas. *BMC Cancer* 8, 243. doi:10.1186/1471-2407-8-243

Minucci, S., and Pelicci, P. G. (2006). Histone Deacetylase Inhibitors and the Promise of Epigenetic (And More) Treatments for Cancer. *Nat. Rev. Cancer* 6 (1), 38–51. doi:10.1038/nrc1779

Murat, A., Migliavacca, E., Gorlia, T., Lambiv, W. L., Shay, T., Hamou, M.-F., et al. (2008). Stem Cell-Related "Self-Renewal" Signature and High Epidermal Growth Factor Receptor Expression Associated with Resistance to Concomitant Chemoradiotherapy in Glioblastoma. *Jco* 26 (18), 3015–3024. doi:10.1200/JCO.2007.15.7164

Naert, T., Dimitrakopoulou, D., Tulkens, D., Demuynck, S., Carron, M., Noelanders, R., et al. (2020). RBL1 (P107) Functions as Tumor Suppressor in Glioblastoma and Small-Cell Pancreatic Neuroendocrine Carcinoma in Xenopus Tropicalis. *Oncogene* 39 (13), 2692–2706. doi:10.1038/s41388-020-1173-z

Nguyen, T. T. T., Zhang, Y., Shang, E., Shu, C., Torrini, C., Zhao, J., et al. (2020). HDAC Inhibitors Elicit Metabolic Reprogramming by Targeting Super-enhancers in Glioblastoma Models. *J. Clin. Invest.* 130 (7), 3699–3716. doi:10.1172/JCI129049

Nozell, S., Laver, T., Patel, K., and Benveniste, E. N. (2006). Mechanism of IFN-β-Mediated Inhibition ofIL-8Gene Expression in Astroglioma Cells. *J. Immunol.* 177 (2), 822–830. doi:10.4049/jimmunol.177.2.822

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

Ruijter, A. J. M. d., Gennip, A. H. v., Caron, H. N., Kemp, S., and Kuilenburg, A. B. P. v. (2003). Histone Deacetylases (HDACs): Characterization of the Classical HDAC Family. *Biochem. J.* 370 (Pt 3), 737–749. doi:10.1042/BJ20021321

Schiff, D., Van den Bent, M., Vogelbaum, M. A., Wick, W., Miller, C. R., Taphoorn, M., et al. (2019). Recent Developments and Future Directions in Adult Lower-Grade Gliomas: Society for Neuro-Oncology (SNO) and European Association of Neuro-Oncology (EANO) Consensus. *Neuro Oncol.* 21 (7), 837–853. doi:10.1093/neuonc/noz033

Song, Y., Jiang, Y., Tao, D., Wang, Z., Wang, R., Wang, M., et al. (2020). NFAT2-HDAC1 Signaling Contributes to the Malignant Phenotype of Glioblastoma. *Neuro Oncol.* 22 (1), 46–57. doi:10.1093/neuonc/noz136

Sun, L., Hui, A.-M., Su, Q., Vortmeyer, A., Kotliarov, Y., Pastorino, S., et al. (2006). Neuronal and Glioma-Derived Stem Cell Factor Induces Angiogenesis within the Brain. *Cancer Cell* 9 (4), 287–300. doi:10.1016/j.ccr.2006.03.003

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi:10.1093/nar/gkw937

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

Tong, J., Shen, Y., Chen, X., Wang, R., Hu, Y., Zhang, X., et al. (2019). FKBP3 Mediates Oxaliplatin Resistance in Colorectal Cancer Cells by Regulating HDAC2 Expression. *Oncol. Rep.* 42 (4), 1404–1412. doi:10.3892/or.2019.7259

Vallianatos, C. N., and Iwase, S. (2015). Disrupted Intricacy of Histone H3K4 Methylation in Neurodevelopmental Disorders. *Epigenomics* 7 (3), 503–519. doi:10.2217/epi.15.1

Was, H., Krol, S. K., Rotili, D., Mai, A., Wojtas, B., Kaminska, B., et al. (2019). Histone Deacetylase Inhibitors Exert Anti-tumor Effects on Human Adherent and Stem-like Glioma Cells. *Clin. Epigenet* 11 (1), 11. doi:10.1186/s13148-018-0598-5

Wen, P. Y., Weller, M., Lee, E. Q., Alexander, B. M., Barnholtz-Sloan, J. S., Barthel, F. P., et al. (2020). Glioblastoma in Adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) Consensus Review on Current Management and Future Directions. *Neuro Oncol.* 22 (8), 1073–1113. doi:10.1093/neuonc/noaa106

Williams, M. J., Singleton, W. G. B., Lowis, S. P., Malik, K., and Kurian, K. M. (2017). Therapeutic Targeting of Histone Modifications in Adult and Pediatric High-Grade Glioma. *Front. Oncol.* 7, 45. doi:10.3389/fonc.2017.00045

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 © 2021 Fan, Peng, Wang, Li and Zhao. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fmolb.2021.686718

Upregulation of Translationally Controlled Tumor Protein Is Associated With Cervical Cancer Progression

Xiaoyu Zhu[1†], Ji Ren[1†], Dianqin Xu[2], Di Cheng[3], Wei Wang[3], Jie Ren[2], Ziwen Xiao[2], Hongmei Jiang[1], Yan Ding[4] and Yujie Tan[1,2]**

[1]School of Laboratory Medicine, Guizhou Medical University, Guiyang, China, [2]Affiliated Hospital of Guizhou Medical University, Guiyang, China, [3]Affiliated Oncology Hospital of Guizhou Medical University, Guiyang, China, [4]Taihe Hospital, Affiliated to Hubei Medical University, Shiyan, China

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Rodney Hull,
University of Pretoria, South Africa
Hongying Hao,
University of Louisville, United States

***Correspondence:**
Yan Ding
dyywzx@163.com
Yujie Tan
tanyujie@gmc.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 27 March 2021
Accepted: 11 August 2021
Published: 13 September 2021

Citation:
Zhu X, Ren J, Xu D, Cheng D, Wang W, Ren J, Xiao Z, Jiang H, Ding Y and Tan Y (2021) Upregulation of Translationally Controlled Tumor Protein Is Associated With Cervical Cancer Progression. Front. Mol. Biosci. 8:686718. doi: 10.3389/fmolb.2021.686718

Outside a few affluent countries with adequate vaccination and screening coverage, cervical cancer remains the leading cause of cancer-related deaths in women in many countries. Currently, a major problem is that a substantial proportion of patients are already at an advanced cancer stage when diagnosed. There is increasing evidence that indicates the involvement of translationally controlled tumor protein 1 (TPT1) overexpression in cancer development, but little is known about its implication in cervical cancer. We assessed the levels of TPT1 in surgical tissue and sera of patients with cervicitis, cervical intraepithelial neoplasia III, and cervical cancer, as well as in normal and cancerous cervical cell lines. Gene sets, pathways, and functional protein interactions associated with TPT1 were identified using the TCGA data cohort of cervical cancer. We found that the TPT1 expression was significantly increased in cervical cancer tissue compared to all nonmalignant cervical tissues, including samples of cervicitis, cervical intraepithelial neoplasia III, and normal controls. Serum level of TPT1 was also increased in cervical cancer patients compared to healthy subjects. Furthermore, elevated TPT1 expression was significantly correlated with lymph node metastasis and a low differentiation degree of the cancer. In the cancerous tissues and cell lines, selective markers of PI3K/AKT/mTOR pathway over-activation, apoptosis repression, and EMT were detected, and their interaction with TPT1 was supported by biometrics analyses. Our results, for the first time, demonstrate a strong correlation of upregulated TPT1 expression with cervical cancer progression, suggesting that TPT1 might provide a potential biomarker for cervical cancer progression.

Keywords: cervical cancer, TPT1, PI3K/Akt/mTOR pathway, diagnosis, biometrics

INTRODUCTION

Translationally controlled tumor protein 1 (TPT1/TCTP, aka histamine-releasing factor HRF and fortilin) is a highly conserved protein abundantly present in all eukaryotic organisms (Li et al., 2001) and is involved in almost all fundamental biological processes underpinning growth, stress response, and survival (Bommer and Telerman, 2020). In addition to its intracellular functions, TPT1 can be

secreted to stimulate histamine production of basophils to elicit allergic reactions and inflammation (Kawakami et al., 2019).

TPT1 is frequently elevated in cancers, leading to suppression of apoptosis, promotion of metastasis, and resistance to anticancer therapy (Koziol and Gurdon, 2012; Lee et al., 2020). TPT1 overexpression in circulation or tissues of patients has been documented in leukemia (Yağcı et al., 2013) and cancers in most main human organs such as the lung (Chen et al., 2013; Sun et al., 2019), liver (Chan et al., 2012; Lin et al., 2020), colon (Bommer et al., 2017), prostate (Kaarbø et al., 2013; Rocca et al., 2015), breast (Neuhäuser et al., 2019), etc. (Koziol and Gurdon, 2012; Ambrosio et al., 2015; Chen et al., 2015; Phanthaphol et al., 2017; Bommer and Telerman, 2020). The pro-survival roles of TPT1 manifest themselves in multiple and correlated facets of tumorigenicity and malignant transformation, including genome stability (Li et al., 2017), cell cycle, apoptosis machinery (Nagano-Ito and Ichikawa, 2012), mitotic/meiotic progression (Bommer and Telerman, 2020), PI3K/AKT/TOR signaling (Bommer et al., 2015), epithelial-to-mesenchymal transition (EMT) (Mishra et al., 2018), induction of pluripotent stemness (Amson et al., 2013), autophagy (Lee et al., 2020), and the reciprocal constraint between TPT1 and the p53-dependent tumor suppressor pathway (Amson et al., 2011).

Recently, inhibition of TPT1 to attain tumor reversion has provided a new approach for cancer therapy (Tuynder et al., 2002). The TPT1 gene was identified as the most differentially downregulated in revertant human leukemia cell U937 and breast cancer cells such as MCF7, compared with their malignant counterparts (Tuynder et al., 2002). Inhibition of TPT1 by anti-sense cDNA or siRNA *in vitro* or injection of antagonistic drugs, such as anti-histaminic compounds into tumor-bearing mice, suppressed the malignant phenotype of cancers of the breast, lung, and colon and melanoma (Tuynder et al., 2004).

However, hampered by the inherent difficulties of research resources, our knowledge of the association between TPT1 and human cervical carcinoma is largely absent. As per the global cancer statistics, cervical cancer is ranked as the fourth most common cancer in women (https://www.who.int/health-topics/cervical-cancer#tab = tab_1) (Das, 2021). Diagnosis and treatment of the preinvasive lesions at cervical intraepithelial neoplasia (CIN) stages would render the disease curable (Castle et al., 2017) and possess tremendous importance in the cervical cancer control agenda (Preti et al., 2020). Although the cervical cancer screening and prophylactic vaccination against HPV have achieved some success in prevention and early treatment in the most affluent countries, cervical cancer remains the leading cause of mobility and mortality for women in countries with inadequate health services and vaccination coverage (Bray et al., 2018; Beddoe, 2019; Canfell et al., 2020).

While no clinical evidence is available, a few works have used human papillomavirus (HPV)-infected cervical cancer cell lines to examine the role of TPT1 in cell survival. Forced TPT1 overexpression protects HeLa cells from cytotoxic drug-induced cell death, *via* interfering with the mitochondria-mediated apoptosis pathway (Jung et al., 2014). TPT1 expression is inducible by incubation of HeLa cells under stimulation of serum, which can be reversed by rapamycin treatment, implying that TPT1 is tightly regulated by the growth factor/PI3K/AKT/mTOR network (Bommer et al., 2015). In another study using a mouse cervical cancer model inoculated with HeLa cells, TPT1 protein expression increased in the advanced tumor but not in the stage of tumor initiation. The authors speculated that the TPT1 level is inversely related to the apoptotic activity (Hou and Zhai, 2015). However, the function of TPT1 in inflammation restriction *via* negatively regulating autophagy (Bae et al., 2017) may overrun its pro-proliferative action. TPT1 is identified as one of the most upregulated genes in SiHa cells treated with anti-inflammatory peptide Annexin A1 (*ANXA1*) and might have contributed to the inhibition of cell proliferation (Prates et al., 2015).

In the present study, by analyzing both tissue and serum TPT1 levels from clinical subjects and online databases, we attempt to clarify the relevance of TPT1 to cervical cancer development and the potential value of serum TPT1 as a facile indicator of cancer diagnosis and prognosis.

MATERIALS AND METHODS

Patient Samples

Ethics approval for the study was obtained from the Human Ethics Committee of the Guizhou Medical University. A waiver of documentation of the written informed consent form of the subjects was issued by the Human Ethics Committee of the Guizhou Medical University, and oral informed consent was obtained from the patients.

Surgically resected cervical tissue samples were collected from patients treated in the Gynaecology Department at the Affiliated Hospital of Guizhou Medical University between January and September 2018. Information of samples used in different experiments, such as diagnosis, pathologic staging, and case number, is summarized in **Tables 1**, **2**, **3**. Samples were divided into four groups according to the pathological diagnosis, including cervical cancer (n = 36, age 31–70, mean age 48.50 ± 9.35), cervical intraepithelial neoplasia grade III (CINIII. n = 35, age 27–70, mean age 44.74 ± 11.23), cervicitis (n = 30, age 22–68, mean age 42.87 ± 10.56), and the control group (n = 30, age 30–54, mean age 45.13 ± 4.91). The control group included patients who underwent hysterectomy because of fibroid uteri but for whom the cervix tissues were confirmed to be normal by histopathological diagnosis postoperatively. In parallel, serum samples were collected from patients with cervical cancer (n = 40, age 26–69, mean age 47.55 ± 10.92), patients with CINIII (n = 40, age 23–60, mean age 42.68 ± 9.08), and healthy women (n = 40, age 31–56, mean age 45.18 ± 7.45). There was no statistically significant difference in age between these groups ($p > 0.05$). Another set of tissue specimens was obtained from cervical cancer patients (n = 20, age 31–68, mean age 48.85 ± 8.74) at the surgery and paired into cancerous and paracancerous groups. Cancer tissue and paracancerous tissue adjacent to and at the outside of the edge of cancer were dissected by the responsible surgeon and evaluated by tissue appearance, including color, texture, and hardness. All samples were confirmed through pathological examinations performed independently by two chief pathologists at the hospital.

The study had the following inclusion criteria: 1) diagnosis of cervical cancer and CINIII were based on the Consensus

TABLE 1 | TPT1 staining positivity rate (%) in cervical tissues of different groups.

Group	Number of cases	Positive cases	Positive rate (%)	*p*-value
Control	30	2	6.7	N/A
Cervicitis	30	8	27	a
CINIII	35	16	46	a
Cervical cancer	36	25	69	a, b, c

a, b, c, p < 0.05 comparing to the control, cervicitis, and CINIII samples, respectively.

TABLE 2 | Comparison of TPT1 content in serum of different groups of patients (means ± SD).

Group	Number of cases	TPT1 (pg/ml)	*p*-value
Normal	40	52.48 ± 15.01	N/A
CINIII	40	54.71 ± 15.89	ns
Cervical cancer	40	66.40 ± 30.98	0.028[a]

[a]*p < 0.05; ns, not significant.*

TABLE 3 | Correlation between TPT1 expression level and patient clinicopathological features.

Variables	Number of cases	TPT1 expression		*p*-value
		High	Low	
Ages				
≤45	7	2	5	0.081
>45	13	9	4	
FIGO stage				
I	12	4	8	0.199
II	8	5	3	
III	0	0	0	
IV	0	0	0	
Lymph node metastasis				
Positive	7	6	1	0.019[a]
Negative	13	4	9	
Differentiated degree				
Low	5	4	1	0.035[a]
Medium	15	4	11	
High	0	0	0	

[a]*p < 0.05.*

Guidelines for the Management of Abnormal Cervical Cancer Screening Tests in China, combined with medical history, cytology, and pathology biopsy evidence, and 2) the patients had not received chemo-/radio-/hormone therapy or taken other medications prior to the hospital admission. Subjects were excluded if the patient had concurrent conditions of infectious disease, autoimmune disease, metastatic malignancy in the pelvic area, or cancers in other organs such as the breast, liver, kidney, and spleen or had received immunosuppressive therapy within a half year prior to the hospital admission.

Cell Culture

The normal human cervical epithelial cell line HCerEpic (Beijing Beina Chuanglian Biotechnology Institute, China) and the human cervical carcinoma SiHa and HeLa cells (Nanjing Kebai Biotech Co., Ltd., China) were used. The cells were cultured in DMEM (Gibco, United States) supplemented with 10% FBS (Biyuntian Biotech Co., Ltd., China) and grown in an incubator at 37°C in 5% CO_2 with 97% humidity. Cell medium was replenished every 2–3 days and passaged with 0.25% trypsin when the cell confluency reached 80–90%.

Immunohistochemistry of TPT1 Protein Expression

The formalin-fixed, paraffin-embedded tissue samples were freshly sliced and subjected to immunohistochemistry. Tissue sections were sequentially deparaffinized, rehydrated, and treated in citric buffer for 3 min for antigen retrieval and 3% hydrogen peroxide for 30 min to block endogenous peroxide activity. After blocking with goat serum for 30 min and washing with PBS 3 times, sections were incubated with anti-human TPT1 primary antibody (1:1,000, ab133568, Abcam, United States) at 4°C overnight. Next, sections were similarly washed and incubated with HRP-conjugated goat-anti-rabbit immunoglobulin (PV-6001, Beijing Zhongshan Jinqiao Biotech Co., Ltd.) at room temperature for 2 h. After washing with PBS, the sections were developed in chromogen solution of 3′-3′-diaminobenzidine (DAB, Beijing Suolaibao Biotec Co., Ltd., China) for 2–3 min, followed by hematoxylin counterstaining, hydrochloric acid differentiation, bluing, and dehydration, and finally mounted for visual inspection under a microscope.

Slides were examined in five randomly selected high-resolution fields (400 x), and the staining score method that combines quantitatively counting the number of positively stained cells and qualitative evaluation of color intensity was adopted, using validated liver cancer slides as the positive control (Chan et al., 2012) and PBS to replace the primary antibody as the negative control. For the color intensity score, slides were assigned a score of 0 (unstained), 1 (weak), 2 (moderate), and 3 (strong). The staining area score was given for the percentage of positive cell count in the investigated areas: 0 (0%), 1 (1–25%), 2 (26–50%), 3 (50–75%), and 4 (>75%). The final immunoreactivity score (IRS) was presented as the staining intensity score × staining area score. An IRS of "0–4" was defined as negative, and an IRS of "5–12" was defined as positive (Chen et al., 2017).

Measurement of Serum TPT1 Using Enzyme-Linked Immunosorbent Assay

A commercially available ELISA kit (CSB-EL024134HU, Wuhan Huamei Biotech Co., Ltd., China) was used for serum TPT1

detection, following the manufacturer's instruction. Briefly, 100 µl diluents of test samples and protein standards with concentrations ranging from 0 to 800 pg/ml were added into the 96-well plate provided in the kit. After incubation at 37°C for 2 h, 100 µl biotinylated antibody was added to each well and incubated at 37°C for 1 h. Then the plate was subjected to a thorough wash and spin-drying before 100 µl of horseradish peroxidase-labeled avidin was added and the plate was incubated at 37°C for 1 h. After another round of washes and spin-drying, 90 µl of the substrate solution was added, and the plate was kept from light at 37°C for 15–30 min. Upon termination of the reaction with 50 µl stop solution, the optical density (OD) at 450 nm was measured using a microplate reader. The readings of OD were normalized with a background value obtained from blank sample wells, and the sample concentration was calculated according to the standard curve.

Western Blot

50–80 mg cervical tissues were cut into small pieces and ground in liquid nitrogen to extract cell lysates for protein assessment. HCerEpic, SiHa, and HeLa cells in the logarithmic growth phase were seeded into 6-well plates at a density of $1x10^5$ cells per well and cultured for 24 h. Proteins were extracted from the cells lysed in 100 µl of RIPA buffer on ice for 30 min. After centrifugation at 13,400 g at 4°C for 10 min, the supernatant was taken to determine the protein concentration using a BCA kit (Beijing Suolaibao Biotech Co., Ltd., China). 30 µg of total protein was resolved by SDS-PAGE with certain percentages of polyacrylamide separation gel corresponding to the molecular weight of the target protein, that is, 12% for TPT1, BAX, and BCL2, 10% for vimentin and p53, and 8% for all the others. The rabbit-anti-human monoclonal antibodies, anti-TPT1 (ab133568), anti-BAX (ab32503), anti-BCL2 (ab32124), anti-p53 (ab7899), and anti–N-cadherin (ab76011), were purchased from Abcam, United States. The rabbit anti-human polyclonal anti-mTOR (#2927), anti–p-mTOR (#2971), and anti-AKT (#9272) were from Cell Signalling Technology, United States. The rabbit-anti-human polyclonal anti-PI3KCA (bs-2067R) was from Beijing Biosynthesis Biotech Co., Ltd. Other anti-human monoclonal antibodies used were rabbit anti-vimentin (BF0071, Affinity Biosciences Ltd.), rabbit anti–E-cadherin (BF0219, Affinity Biosciences Ltd.), and mouse anti-GAPDH (AP0063, Nanjing Bioworld Biotech Co., Ltd.). GAPDH served as an internal reference. Membranes were then probed with HRP-conjugated goat-anti-rabbit (PMK-013–090, Pumei Biotech Co., Ltd., China) or goat-anti-mouse (EarthOx, United States) secondary antibody. Quantitative analysis was performed using ImageJ software, and the densitometry calculation was normalized against GAPDH to indicate the relative abundance of the target protein.

RNA Extraction and Quantitative RT-PCR

An Axy Prep kit (AP-MN-MS-50, Axygen Biotech Co., Ltd., China) was used to extract total RNA from tissues or cultured cells prepared as described above for WB. The purity and concentration of RNA were measured using NanoDrop One. The reverse transcription was performed using a PrimeScrip TM RT reagent kit with a gDNA Eraser (Perfect Real Time, Takara, Japan) according to the manufacturer's protocols.

Real-time qRT-PCR was performed using TB green Premix Rex Taq TM II (Tli RNase H Plus, Takara, Japan). The relative mRNA transcription of *TPT1* was calculated using the $2^{-\Delta\Delta CT}$ method, taking the gene of human housekeeping enzyme *HPRT* as the endogenous control. Primer sequences were as follows: forward 5′-CAGTAATCACTGGTGTCGAT-3 and reverse 5′-GGATGT GCTTGATTTGTTCT-3′ for TPT1, and forward 5′-ATGGCG ACCCGCAGCCCT-3′ and reverse 5′-CCATGAGGAATAAACACCCT-3' for HPRT. The PCR condition was 40 cycles of 30 s 95°C, 5 s at 95°C and 30 s at 60°C.

Biometrics

The RNA-Seq gene data of *TPT1* transcription in a cohort of 305 cervical cancer tissues were acquired from the TCGA database. The samples were divided into two groups with phenotype labels of high (n = 153) or low (n = 152) *TPT1* transcription, defined by the threshold at the median *TPT1* value. Gene Set Enrichment Analysis (GSEA3.0, UC San Diego, CA, United States) (http://software.broadinstitute.org/gsea/index.jsp) was applied to calculate and predict gene sets statistically correlating with the two phenotypes. The hallmark gene sets "h.all.v7.2.symbols.gmt" in the Molecular Signatures Database (MSigDB) on the GSEA website (http://software.broadinstitute.org/gsea/msigdb/index.jsp) were applied as reference genes. Each gene in a certain gene set was ranked by its expression difference from the cumulative running sum values, weighted using the default statistical setting of 1,000 genome permutations per analysis. An enrichment score (ES) for each gene set was recorded as the maximum deviation from zero. The p-value and the normalized enrichment score (NES) were used to rank the enrichment pathways of each phenotype, where a p-value < 0.05 calculated by Student's t-test and an FDR (false discovery rates) q-value ≤ 0.25 were the statistical cut-offs for significant enrichment.

The online Search Tool of the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org/) was used to construct the protein–protein interaction (PPI) map to calculate the strength of the association between TPT1 and key proteins in apoptosis, PI3K, and EMT pathways. Information sources of primary (experiments), predicted (databases), and textmining results in *Homo sapiens* were searched to gain an integrated overview of the pathway network. The connectivity degree was identified as network edges (lines connecting pairs of proteins) with the quantified confidence score. The minimum confidence score was set at 0.7 (high confidence).

Statistical Analysis

SPSS 24.0 software was used to analyze the data, and correlation analysis was performed using GraphPad Prism 8.0. Two-tailed Student's t-test and ANOVA (analysis of variance) were used for differential comparison between two and multiple groups, respectively. Logarithmic transformation

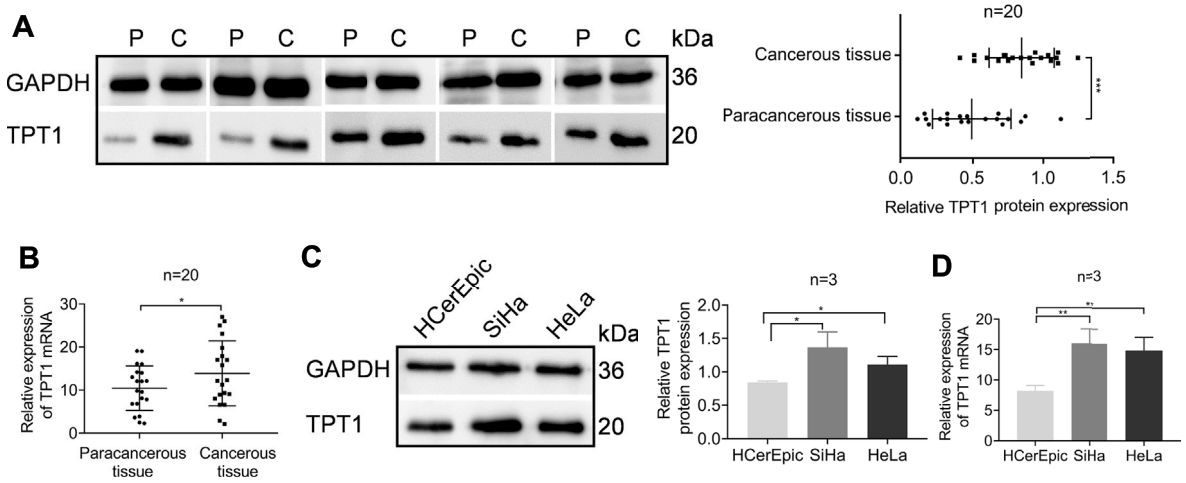


FIGURE 1 | Elevated TPT1 expression in cervical cancer tissues and cells. **(A)** Protein expression of TPT1 in tissues of cervical cancer patients, assessed by WB and pairwise compared to paracancerous tissue **(left)** and quantified **(right)**. P: paracancerous tissue; C: cancerous tissue. **(B)** Relative TPT1 gene transcription in the paired cervical cancer tissues, measured by qRT-PCR. **(C)** TPT1 protein expression in normal human cervical epithelial cell line HCerEpic and cervical carcinoma cell lines SiHa and HeLa, analyzed by WB **(left)** and quantified **(right)**. **(D)** Relative TPT1 gene transcription in the cell lines, measured by qRT-PCR. WB images are representative of all samples tested. All experiments were repeated at least three times. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

was applied to variables that did not follow normal distribution and the nonhomogeneous variances. Numerical data conforming to the normal distribution were presented as "mean ± SD," while the categorical data were expressed as the number of cases or the percentage rate. The chi-squared test (χ^2 test) and the Kruskal–Wallis rank sum test were used for comparison between two- and multi-component data, respectively. A p-value < 0.05 was considered statistically different.

RESULTS

TPT1 Expression Is Elevated in Cervical Cancer Tissues and Cell Lines at Both Protein and mRNA Levels

Despite the ample evidence and discussion of the association between TPT1 overexpression and cancer, it is rarely studied in cervical cancer. We collected cervical cancer tissues from the local hospital to characterize the status of TPT1 expression. We found that the TPT1 protein abundance was almost doubled in the cancerous tissues compared to the paired paracancerous tissues (**Figure 1A**). To a lesser extent, the TPT1 mRNA transcription, revealed by qRT-PCR, was also higher (increased by 32.95%) in the cancerous than in the adjacent paracancerous tissues, despite a relatively wide overlap between the two sample groups (**Figure 1B**). TPT1 protein expression was also significantly higher in both human cervical carcinoma SiHa and HeLa cells than in the human normal cervical epithelial cell line HCerEpic (**Figure 1C**). In alignment with it, the TPT1 gene transciption was remarkably upregulated in cervical cancer cell lines (increased by 125% in SiHa cells and 111% in HeLa cells, compared to HCerEpic cells) (**Figure 1D**).

The Protein Abundance of TPT1 Correlates Positively With Malignant Transformation, Lymph Node Metastasis, and the Low Differentiation Degree of Cervical Cancer

Like many other common cancers, cervical cancer usually develops over a long progressive course. To get more insights into the trajectory of TPT1 expression in cervical cancer evolvement, we compared TPT1 protein levels in cervix tissues from age-matched patients with cervicitis, CINIII, and cervical cancer, by immunohistochemical staining (IHC). The HPV infection–induced cervical squamous intraepithelial neoplasia (CIN) is graded into CINI, CINII, and CINIII according to the depth of the lesion within the afflicted cervix surface. CINIII includes carcinoma *in situ* and has a high risk of developing into cancer. Whereas CIN is tightly associated with HPV infection and can involve inflammation, cervicitis is referred to as a mostly benign inflammation of the cervix that can result from a wide range of stimulations such as infection. For the control, due to the lack of cervix tissue from a healthy population, we collected cervix tissues upon hysterectomy for treatment of fibroid uteri which had been confirmed to be pathologically normal. Fibroid uteri are noncancerous myomas in the uterus and rarely include cervix lesions. A continuous increase in the TPT1 positive rate was observed in the order of control (6.7%), cervicitis (27%), CINIII (46%), and cervical cancer (69%) (**Figure 2A**; **Table 1**). TPT1 expression in circulation was also determined in CINIII and cervical cancer patients, with sera from age-matched healthy women as the normal control. Although the stepwise enhancement of TPT1 secretion from the normal baseline to CINIII and from CINIII to cancer did not reach the threshold for statistical significance, the increase was significant in cancer compared to the normal control (26.5%, $p = 0.028$) (**Figure 2B**; **Table 2**).

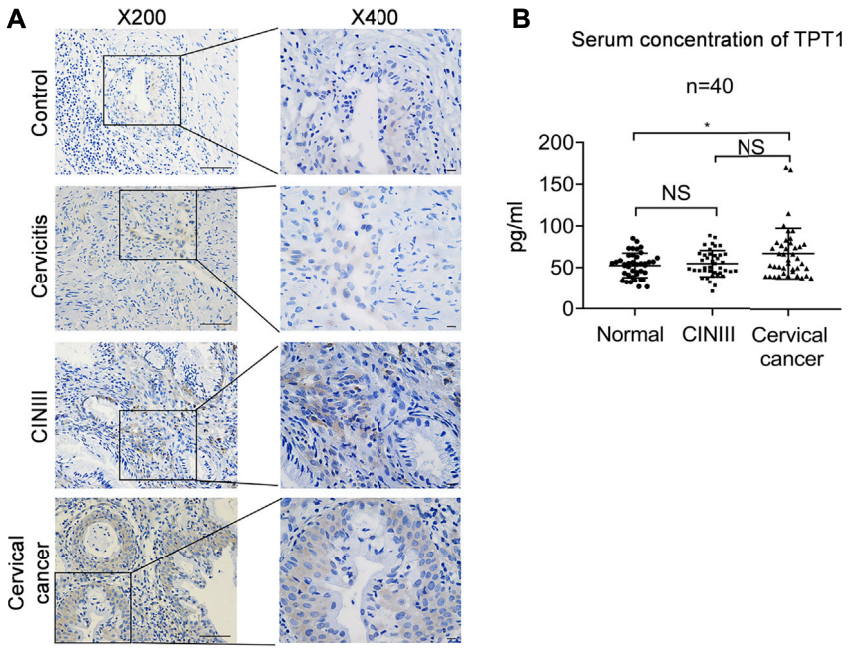


FIGURE 2 | Increased abundance of TPT1 protein in the tissue and sera of cervical cancer compared to cervicitis, CINIII, and controls. **(A)** Tissue IHC staining of TPT1 protein in the control and different cervical lesions including cervicitis, CINIII, and cervical cancer. Scale bars: x200, 50 μm; x400: 10 μm. **(B)** Serum TPT1 concentration in CINIII and cervical cancer patients, compared to normal healthy controls, measured by ELISA. IHC images are representative of all samples tested. *$p <$ 0.05; NS: statistically not significant.

Among the cancerous tissues in which we quantified TPT1 abundance by WB, we further stratified the high (≥median TPT1 value) or low (<median TPT1 value) TPT1 protein expression with clinical parameters including age, FIGO (International Federation of Gynaecology and Obstetrics) stage, lymph node metastasis, and degree of cancer differentiation (**Table 3**). The result demonstrated that a high TPT1 protein level significantly correlated with lymph node metastasis (p = 0.019) and a low degree of tumor differentiation (p = 0.035), implying its link with EMT and cancer aggressiveness. No correlation was observed between the TPT1 status and age or the FIGO stage.

Biometrics Analysis Reveals Differentially Expressed Gene Sets and Pathways Correlated With TPT1 Expression in Cervical Cancer

The data of *TPT1* gene transcription in cervical cancer tissues were downloaded from the TCGA database and analyzed by GSEA to identify gene sets and pathways associated with differential TPT1 expression. GSEA provided us with a gene transcript profiling in certain categorical phenotype contexts, that is, high (≥median value) or low (<median value) *TPT1* transcription. Gene sets positively enriched in the low *TPT1* phenotype were genes encoding proteins involved in cell cycle arrest including G2/M checkpoint, mitotic spindle, and E2F transcription factors (**Figure 3A**). Genes upregulated by Wnt signaling activation through β-catenin accumulation were also associated with low *TPT1* levels (**Figure 3A**). Meanwhile, genes encoding proteins involved in oxidative phosphorylation, subgroups of genes regulated by MYC versions 1 and 2 (MYC targets V1 and V2), EMT-related genes, and upregulated genes during angiogenesis were significantly enriched in the high *TPT1* group (**Figure 3B**). These genes are implicated in the metabolic reprogramming, cell proliferation, transforming activity, and invasion of cervical cancer, potentially in concert with TPT1 activity.

We next sought an understanding of how the TPT1 protein is functionally involved in pathways of apoptosis, proliferation, and EMT, the most frequently dysregulated pathways during cervical cancer development. A map of protein–protein interactions (PPIs) between TPT1 and key factors regulating apoptosis, PI3K cascade, and intercellular adhesion was created by mining the Search Tool for the Retrieval of Interacting Genes (STRING) to predict their functional cooperation in biological events (**Figure 4**). The map shows that TPT1 directly interacts with TP53 (a combined association score of 0.770) and BCL2 (a combined association score of 0.736). The proapoptotic protein BAX and the antiapoptotic BCL2 are both transcriptional targets of p53. The interactions among them suggest that abnormal TPT1 activity could impair the cascade of apoptosis, *via* dysregulation of p53 and BCL2. With a very high combined association score above 0.9, TP53 and BCL2 lead the interaction network to a cluster of proteins related to PI3K/AKT/mTOR signal transduction (PTEN, AKT1, mTOR. PI3CA, PI3CB, and PI3CG), which is critical in bioprocesses of cell cycle and proliferation. Members of the catenin family (CTNNA1) and the cadherin superfamily (CDH15) are also linked with the

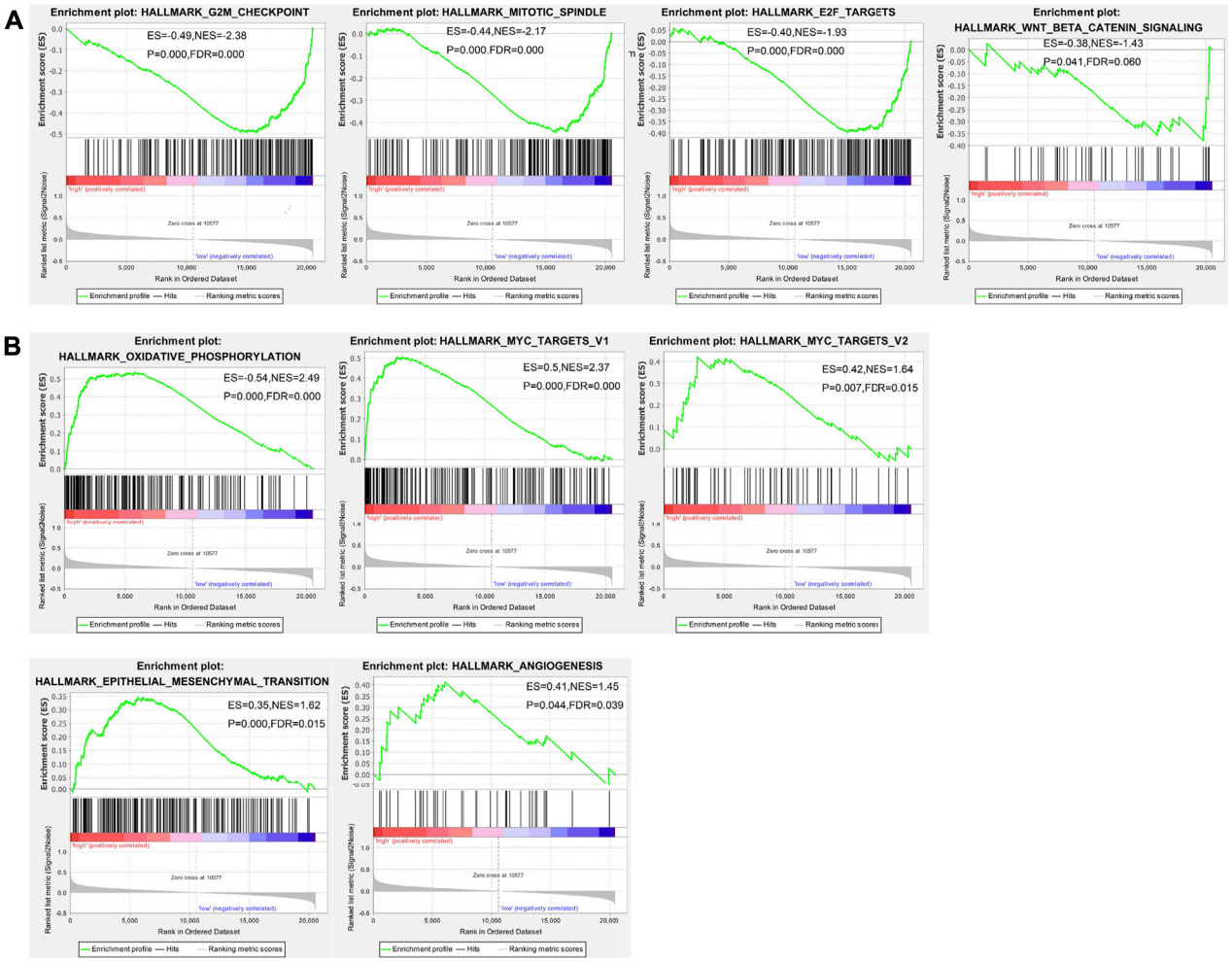


FIGURE 3 | GSEA enrichment analysis of TCGA data on gene sets and pathways associated with low **(A)** or high **(B)** *TPT1* expression in cervical cancer. The horizontal bar, gradient-filled from red to blue, represents where genes in each gene set appear in the ranked list of genes. Genes on the left side (red) correlate most strongly with the phenotype. The vertical black lines represent the running enrichment scores as the projection of individual genes onto the horizontal ranked gene list. The bottom ranking matric in gray, moving from above zero (positively correlated) to below zero (negatively correlated), measures a gene's correlation with the phenotype profile. Statistical variables of GSEA analysis are displayed in each image. ES, enrichment score for the gene set, reflecting the degree to which this gene set is over-represented at the peak (furthest from 0.0) of the entire ranked list of genes; *p*-value, statistical assessment of the significance against null distribution; NES, normalized enrichment score calculated by adjusting ES for gene set size or multiple hypothesis testing across analyzed gene sets; FDR (false discovery rate *q*-value), the estimated probability that a given NES represents a false positive finding. $p < 0.05$ and FDR < 0.25 were considered statistically significant.

cluster at very high association scores—all above 0.9. These factors play roles in cell adhesion, thereby being involved in cancer metastasis.

Key Factors in Pathways Related to TPT1 Regulation and Function Are Altered in Cervical Cancer Tissues and Cells

TPT1 participates in the precise and accurate regulation of various bioprocesses. Although in this work we could not directly examine its mechanistic role in cervical cancer development, we compared the activities of the PI3K/AKT/mTOR pathway, apoptosis, and EMT markers between cancerous and paracancerous tissues of cervical cancer patients (**Figure 5**). According to reports in the literature and our GSEA/STRING data, these factors are closely linked with TPT1 regulation and functions. As anticipated, in comparison to the paracancerous control, the cancerous tissues had significantly higher protein expression of PI3K, AKT, mTOR, and p-mTOR, accompanied by a lower BAX:BCL2 ratio (decreased by 37.80%) that indicated a status with favored growth signaling and restrained apoptosis. Moreover, the expression switch from E-cadherin to N-cadherin in cancer tissues suggested the activation of EMT, along with the overexpression of vimentin, an important EMT marker. Meanwhile, p53 expression was significantly reduced (**Figures 5A,B**). It is known that p53 and TPT1 form an antagonizing regulatory circuit where p53 directly represses TPT1 transcription and TPT1 promotes p53 degradation (Amson et al., 2011). Similar signatures of PI3K/AKT/mTOR

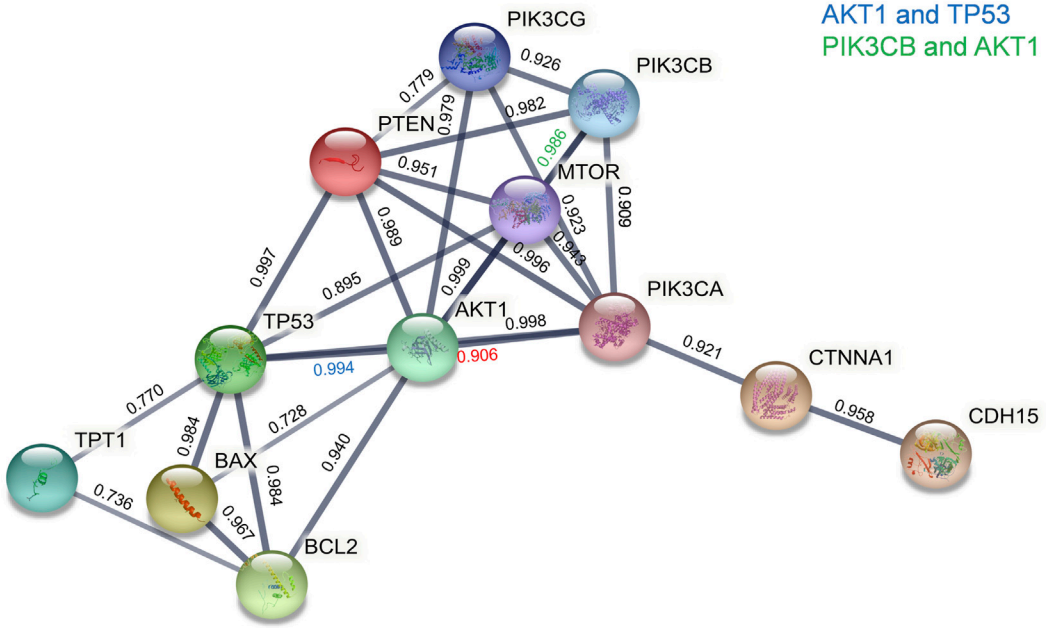


FIGURE 4 | STRING PPI network analyses of functional connections of TPT1 with key proteins of PI3K/AKT/mTOR signaling, apoptosis, and EMT pathways. Line thickness indicates the strength of data support, and circles in different colors represent individual proteins with the abbreviated name. The number above each line is the combined association score that indicates a probability of the existence of a real functional association between assigned proteins, calculated from online sources of textmining, experiments, and databases. The threshold of the confidence score was set at 0.7 (high).

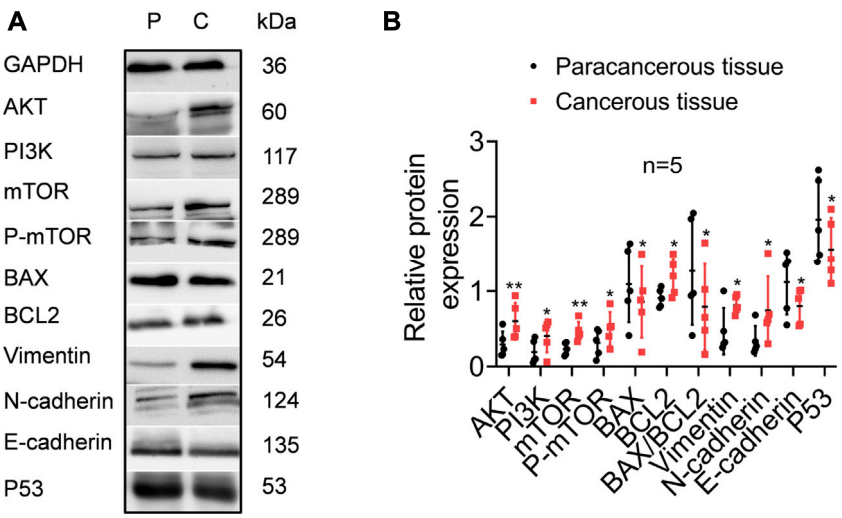


FIGURE 5 | Identification of key proteins involved in PI3K/AKT/mTOR pathway over-activation, apoptosis suppression, EMT, and p53 degradation in cancerous tissues of cervical cancer patients, in comparison with the paracancerous tissues. **(A)** WB images that are representative of all samples tested. **(B)** Densitometry quantification of the WB results from the experiment performed in A. The ratio of BAX:BCL2 is calculated to highlight the apoptosis activity. *$p < 0.05$. **$p < 0.01$. *** $p < 0.001$.

over-activation, blocking of apoptosis, and EMT activity were detected in both SiHa and HeLa cells, in comparison to HCerEpic cells, except that the change in vimentin was not evident (**Figures 6A,B**). Besides, the level of p53 in the cancer cell lines did not display an inversed relationship with TPT1 expression, as we might have expected. The causative relationships between these pathways with TPT1 overexpression in cervical cancer warrant verification.

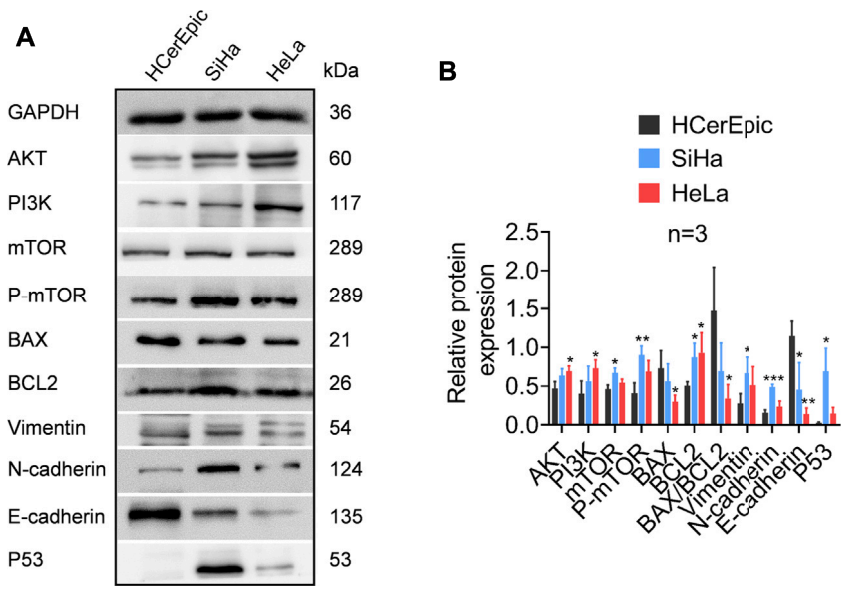


FIGURE 6 | Identification of proteins involved in PI3K/AKT/mTOR pathway over-activation, apoptosis suppression, and EMT in cervical carcinoma cell lines SiHa and HeLa, in comparison with the non-malignant cervical epithelial cell line HCerEpic. **(A)** WB images that are representative of all replicates of the experiment. **(B)** Densitometry quantification of the WB results from the experiment performed in A. The ratio of BAX:BCL2 is calculated to highlight the apoptosis activity. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

DISCUSSION

TPT1 mRNA and viral particles share a similar structure and trigger an innate immune reaction in a similar way (Amson et al., 2013). The dual function of TPT1 in defense against infection and promoting cell survival, two events deterministic for the cellular fate in cervical cancer, underlines its importance in cervical cancer development. However, to the best of our knowledge, the present study is the first to report its relevance to cervical cancer progression in patients. We also sought to address the differential expression of TPT1 in noncancerous cervical lesions including cervicitis and pre-cancerous CINIII neoplasia. The data suggested a continuum of increased TPT1 expression leading up to malignant transformation. Bearing in mind that TPT1 might execute different functions in the context of benign cervical inflammation, its steady increase through cancer development suggests that TPT1 is an important signature manifesting itself long before cancer diagnosis. The finding that high expression of TPT1 significantly correlated with both lymph node metastasis and cancer cell differentiation implies that it may promote cancer progression. However, this was not echoed in FIGO staging, which could be explained by the small number of sample cases and the fact that no samples at FIGO stages III and IV were available.

The invasion of the HPV oncogenes is the major etiological cause of cervical cancer and the precursor epithelial lesions. Malignant transformation of cervical cancer ensues from failure in the defense against the virus infection, manifesting as interrupted cellular machinery, including DNA replication, inefficient immune response, chronic inflammation (Boccardo et al., 2010), disruption of p53 expression and function (Ruttkay-Nedecky et al., 2013), and abnormal PI3K/AKT/mTOR signaling (Zhang et al., 2015; Bossler et al., 2019). The close relationship between TPT1 and HPV infection is also shown in a proteomics profiling of lung A549 cells stably infected by HPV16E6/E7, where the cyto-protective TPT1 was remarkably upregulated (Ciotti et al., 2009). For the small number of HPV-negative cervical cancers, over-activation of the PI3K/AKT/mTOR pathway or mutations of its regulatory factors, such as PTEN, EGFR, and HER2, accounted for the virus-independent etiology of carcinogenesis (Tjalma, 2018).

In our study, biometrics data mining substantially compensated for the limitation of the sample size and the scarcity of experimental reference. Both GSEA enrichment analysis and STRING mapping provided information that has resonance with the near-ubiquitous engagement of TPT1 in tumorigenesis found in other cancers (Amson et al., 2013; Lee et al., 2020), including growth promotion *via* PI3K signaling, antiapoptotic activities, EMT, and the reciprocal interaction with p53. Although we could not clarify whether the increased TPT1 was a phenotype secondary to other alterations in cervical cancer, it is plausible to speculate that TPT1 was proactively involved in the examined pathways. Unlike the cervical cancer cell lines, where a dramatic upregulation of TPT1 at both mRNA and protein levels was detected, the relative TPT1 gene transcription in about half of the cancerous tissue samples did not distinguish themselves from their paracancerous controls. Nevertheless, the mean value still met the statistical requirement of significance. Regardless of the small sample size, it might imply a predominance of post-transcriptional tuning of TPT1 in the cervical cancer niche. However, before further evidence becomes available, we ought not to rush into any speculation.

We also found that the serum TPT1 concentrations had a tendency to increase in CINIII and cancer patients and was significantly higher in cancer patients than in the healthy population. Monitoring the alteration of the TPT1 protein at the systemic level in cervical cancer is

especially meaningful as it could considerably alleviate the diagnostic burden and may serve as a convenient biomarker to assist early detection of cancer or treatment evaluation.

The multiplicity of TPT1 function in the pathway network, evident in most common cancers and tentatively inferred by our data, offers a great impetus to investigate its therapeutic potential in cervical cancer management. A combined mechanism that evokes an intracellular reprogramming to restore the antitumor defense system was proposed (Tuynder et al., 2002). Moreover, depletion of TPT1 in HeLa cells or mice not only promoted apoptosis by BCL2 inhibition but enhanced the overall autophagy flux (Bae et al., 2017), two attractive strategies against cancer. Interestingly, some existing drugs are emerging as efficient anticancer agents *via* TPT1 inhibition or knockdown. For example, the antidepressant Sertraline and the antipsychotic drug Thioridazine were tested successfully *ex vivo* on primary AMT cells, reducing cell viability by inactivating TPT1 (RaP et al., 2013). Similar treatment on the colon cancer cell line HCT116 retrieved wild-type p53 function and provoked apoptosis in the cells (Amson et al., 2011). Inhibition of TPT1 by binding an anti-malaria drug, Dihydroartemisinin (DHA), to its phosphorylated form in breast cancer cell lines reduced cell growth and induced apoptosis (Lucibello et al., 2015).

In conclusion, we demonstrate a positive correlation between TPT1 expression and cervical cancer progression. Studies with larger cohorts and cellular and animal models are intensely ongoing, conducted by our team, to further consolidate the results and unravel the underlying mechanisms. In particular, evaluation of the significance of TPT1 delivered *via* circulation would hugely benefit cervical cancer diagnosis. Three merits of bringing TPT1 onboard in the battle against cervical cancer are in sight—opening an earlier window for cervical cancer intervention, circumvention of the obstacle of tissue sampling, and a prompt establishment of a tractable therapy option using current anti-TPT1 drugs.

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 Human Ethics Committee of the Guizhou 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

YT conceived and designed the experiments; XZ and JR performed the experiments; XZ analyzed the data; ZX, JR, WW, HJ, DC, and DX contributed to the sample collection; YT and YD provided the experimental device/tools; XZ and JR were the major contributors in writing the article draft; YT and YD supervised the data interpretation and edited the article. All authors have reviewed, edited, and approved the final article.

FUNDING

The present research project and its follow-up studies have been funded by the National Natural Science Foundation of China (NSFC, grant number 81960473), the Supplemental Fund Supporting the NSFC Project from the Guiyang Science and Technology Bureau (grant number Zhuke (2017) 30–36), and the Guizhou Provincial Education Research Project (grant number KY (2021)).

ACKNOWLEDGMENTS

We heartily appreciate the language proofreading provided by Dou-Dou Liu, a native English speaker.

REFERENCES

Ambrosio, M. R., Rocca, B. J., Barone, A., Onorati, M., Mundo, L., Crivelli, F., et al. (2015). Expression of Translationally Controlled Tumor Protein in Human Kidney and in Renal Cell Carcinoma. *Biomed. Res. Int.* 2015, 730390. doi:10.1155/2015/730390

Amson, R., Pece, S., Lespagnol, A., Vyas, R., Mazzarol, G., Tosoni, D., et al. (2011). Reciprocal repression between P53 and TCTP. *Nat. Med.* 18 (1), 91–99. doi:10.1038/nm.2546

Amson, R., Pece, S., Marine, J.-C., Fiore, P. P. D., and Telerman, A. (2013). TPT1/ TCTP-regulated pathways in phenotypic reprogramming. *Trends Cel Biol.* 23 (1), 37–46. doi:10.1016/j.tcb.2012.10.002

Bae, S.-Y., Byun, S., Bae, S. H., Min, D. S., Woo, H. A., and Lee, K. (2017). TPT1 (tumor protein, translationally-controlled 1) negatively regulates autophagy through the BECN1 interactome and an MTORC1-mediated pathway. *Autophagy* 13 (5), 820–833. doi:10.1080/15548627.2017.1287650

Beddoe, A. M. (2019). Elimination of cervical cancer: challenges for developing countries. *ecancer* 13, 975. doi:10.3332/ecancer.2019.975

Boccardo, E., Lepique, A. P., and Villa, L. L. (2010). The role of inflammation in HPV carcinogenesis. *Carcinogenesis* 31 (11), 1905–1912. doi:10.1093/carcin/ bgq176

Bommer, U.-A., Iadevaia, V., Chen, J., Knoch, B., Engel, M., and Proud, C. G. (2015). Growth-factor dependent expression of the translationally controlled tumor protein TCTP is regulated through the PI3-K/Akt/mTORC1 signalling pathway. *Cell Signal.* 27 (8), 1557–1568. doi:10.1016/j.cellsig.2015.04.011

Bommer, U.-A., Vine, K. L., Puri, P., Engel, M., Belfiore, L., Fildes, K., et al. (2017). Translationally controlled tumor protein TCTP is induced early in human colorectal tumors and contributes to the resistance of HCT116 colon cancer cells to 5-FU and oxaliplatin. *Cell Commun Signal* 15 (1), 9. doi:10.1186/s12964-017-0164-3

Bommer, U. A., and Telerman, A. (2020). Dysregulation of TCTP in Biological Processes and Diseases. *Cells* 9 (7), 1632. doi:10.3390/cells9071632

Bossler, F., Hoppe-Seyler, K., and Hoppe-Seyler, F. (2019). PI3K/AKT/mTOR Signaling Regulates the Virus/Host Cell Crosstalk in HPV-Positive Cervical Cancer Cells. *Int. J. Mol. Sci.* 20 (9), 2188. doi:10.3390/ijms20092188

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

Canfell, K., Kim, J. J., Brisson, M., Keane, A., Simms, K. T., Caruana, M., et al. (2020). Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *The Lancet* 395 (10224), 591–603. doi:10.1016/s0140-6736(20)30157-4

Castle, P. E., Murokora, D., Perez, C., Alvarez, M., Quek, S. C., and Campbell, C. (2017). Treatment of cervical intraepithelial lesions. *Int. J. Gynecol. Obstet.* 138 (Suppl. 1), 20–25. doi:10.1002/ijgo.12191

Chan, T. H. M., Chen, L., Liu, M., Hu, L., Zheng, B.-j., Poon, V. K.-M., et al. (2012). Translationally controlled tumor protein induces mitotic defects and chromosome missegregation in hepatocellular carcinoma development. *Hepatology* 55 (2), 491–505. doi:10.1002/hep.24709

Chen, C., Deng, Y., Hua, M., Xi, Q., Liu, R., Yang, S., et al. (2015). Expression and clinical role of TCTP in epithelial ovarian cancer. *J. Mol. Hist.* 46 (2), 145–156. doi:10.1007/s10735-014-9607-y

Chen, M., Zhang, Y., and Zheng, P.-S. (2017). Tafazzin (TAZ) promotes the tumorigenicity of cervical cancer cells and inhibits apoptosis. *PLoS One* 12 (5), e0177171. doi:10.1371/journal.pone.0177171

Chen, W., Wang, H., Tao, S., Zheng, Y., Wu, W., Lian, F., et al. (2013). Tumor protein translationally controlled 1 is a p53 target gene that promotes cell survival. *Cell Cycle* 12 (14), 2321–2328. doi:10.4161/cc.25404

Ciotti, M., Marzano, V., Giuliani, L., Nuccetelli, M., D'Aguanno, S., Azzimonti, B., et al. (2009). Proteomic investigation in A549 lung cell line stably infected by HPV16E6/E7 oncogenes. *Respiration* 77 (4), 427–439. doi:10.1159/000176209

Das, M. (2021). WHO launches strategy to accelerate elimination of cervical cancer. *Lancet Oncol.* 22 (1), 20–21. doi:10.1016/s1470-2045(20)30729-4

Hou, L. J., and Zhai, J. J. (2015). Aberrant expression profile of translationally controlled tumor protein and tumor-suppressive microRNAs in cervical cancer. *J. BUON* 20 (6), 1504–1509.

Jung, J., Kim, H. Y., Maeng, J., Kim, M., Shin, D. H., and Lee, K. (2014). Interaction of translationally controlled tumor protein with Apaf-1 is involved in the development of chemoresistance in HeLa cells. *BMC Cancer* 14, 165. doi:10.1186/1471-2407-14-165

Kaarbø, M., Storm, M. L., Qu, S., Wæhre, H., Risberg, B., Danielsen, H. E., et al. (2013). TCTP is an androgen-regulated gene implicated in prostate cancer. *PLoS One* 8 (7), e69398.

Kawakami, Y., Kasakura, K., and Kawakami, T. (2019). Histamine-Releasing Factor, a New Therapeutic Target in Allergic Diseases. *Cells* 8 (12), 1515. doi:10.3390/cells8121515

Koziol, M. J., and Gurdon, J. B. (2012). TCTP in development and cancer. *Biochem. Res. Int.* 2012, 105203. doi:10.1155/2012/105203

Lee, J. S., Jang, E. H., Woo, H. A., and Lee, K. (2020). Regulation of Autophagy Is a Novel Tumorigenesis-Related Activity of Multifunctional Translationally Controlled Tumor Protein. *Cells* 9 (1), 257. doi:10.3390/cells9010257

Li, F., Zhang, D., and Fujise, K. (2001). Characterization of fortilin, a novel antiapoptotic protein. *J. Biol. Chem.* 276 (50), 47542–47549. doi:10.1074/jbc.m108954200

Li, Y., Sun, H., Zhang, C., Liu, J., Zhang, H., Fan, F., et al. (2017). Identification of translationally controlled tumor protein in promotion of DNA homologous recombination repair in cancer cells by affinity proteomics. *Oncogene* 36 (50), 6839–6849. doi:10.1038/onc.2017.289

Lin, Z., Zhang, X., Wang, J., Liu, W., Liu, Q., Ye, Y., et al. (2020). Translationally controlled tumor protein promotes liver regeneration by activating mTORC2/AKT signaling. *Cell Death Dis* 11 (1), 58. doi:10.1038/s41419-020-2231-8

Lucibello, M., Adanti, S., Antelmi, E., Dezi, D., Ciafrè, S., Carcangiu, M. L., et al. (2015). Phospho-TCTP as a therapeutic target of Dihydroartemisinin for aggressive breast cancer cells. *Oncotarget* 6 (7), 5275–5291. doi:10.18632/oncotarget.2971

Mishra, D. K., Srivastava, P., Sharma, A., Prasad, R., Bhuyan, S. K., Malage, R., et al. (2018). Translationally controlled tumor protein (TCTP) is required for TGF-β1 induced epithelial to mesenchymal transition and influences cytoskeletal reorganization. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1865 (1), 67–75. doi:10.1016/j.bbamcr.2017.09.014

Nagano-Ito, M., and Ichikawa, S. (2012). Biological effects of Mammalian translationally controlled tumor protein (TCTP) on cell death, proliferation, and tumorigenesis. *Biochem. Res. Int.* 2012, 204960. doi:10.1155/2012/204960

Neuhäuser, K., Küper, L., Christiansen, H., and Bogdanova, N. (2019). Assessment of the role of translationally controlled tumor protein 1 (TPT1/TCTP) in breast cancer susceptibility and ATM signaling. *Clin. Translational Radiat. Oncol.* 15, 99–107. doi:10.1016/j.ctro.2019.01.006

Phanthaphol, N., Techasen, A., Loilome, W., Thongchot, S., Thanan, R., Sungkhamanon, S., et al. (2017). Upregulation of TCTP is associated with cholangiocarcinoma progression and metastasis. *Oncol. Lett.* 14 (5), 5973–5979. doi:10.3892/ol.2017.6985

Prates, J., Franco-Salla, G. B., Dinarte Dos Santos, A. R., da Silva, W. A., da Cunha, B. R., Tajara, E. H., et al. (2015). ANXA1Ac2-26 peptide reduces ID1 expression in cervical carcinoma cultures. *Gene* 570 (2), 248–254. doi:10.1016/j.gene.2015.06.021

Preti, M., Rosso, S., Micheletti, L., Libero, C., Sobrato, I., Giordano, L., et al. (2020). Risk of HPV-related extra-cervical cancers in women treated for cervical intraepithelial neoplasia. *BMC Cancer* 20 (1), 972. doi:10.1186/s12885-020-07452-6

Robert, A., Salvatore, P., Di Fiore, P. P., Karp, J., Marine, J.-C., and Telerman, A. (2013). "Tumor reversion: From bench to potential clinical applications using sertraline and thioridazine. *Cancer Res.* 73 (8), 4338. doi:10.1158/1538-7445.AM2013-4338

Rocca, B. J., Ginori, A., Barone, A., Calandra, C., Crivelli, F., De Falco, G., et al. (2015). Translationally controlled tumor protein in prostatic adenocarcinoma: correlation with tumor grading and treatment-related changes. *Biomed. Res. Int.* 2015, 985950. doi:10.1155/2015/985950

Ruttkay-Nedecky, B., Jimenez, A. M. J., Nejdl, L., Chudobova, D., Gumulec, J., Masarik, M., et al. (2013). Relevance of infection with human papillomavirus: The role of the p53 tumor suppressor protein and E6/E7 zinc finger proteins. *Int. J. Oncol.* 43 (6), 1754–1762. doi:10.3892/ijo.2013.2105

Sun, R., Lu, X., Gong, L., and Jin, F. (2019). TCTP promotes epithelial–mesenchymal transition in lung adenocarcinoma. *Ott* Vol. 12, 1641–1653. doi:10.2147/ott.s184555

Tjalma, W. (2018). HPV negative cervical cancers and primary HPV screening. *Facts Views Vis. Obgyn* 10 (2), 107–113.

Tuynder, M., Fiucci, G., Prieur, S., Lespagnol, A., Géant, A., Beaucourt, S., et al. (2004). Translationally controlled tumor protein is a target of tumor reversion. *Proc. Natl. Acad. Sci.* 101 (43), 15364–15369. doi:10.1073/pnas.0406776101

Tuynder, M., Susini, L., Prieur, S., Besse, S., Fiucci, G., Amson, R., et al. (2002). Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1. *Proc. Natl. Acad. Sci.* 99 (23), 14976–14981. doi:10.1073/pnas.222470799

Yağcı, M., Yegin, Z. A., Akyürek, N., Kayhan, H., Ozkurt, Z. N., Sucak, G. T., et al. (2013). TCTP/HRF pathway and angiogenesis: a feasible intercourse in chronic lymphocytic leukemia. *Leuk. Res.* 37 (6), 665–670. doi:10.1016/j.leukres.2013.01.005

Zhang, L., Wu, J., Ling, M. T., Zhao, L., and Zhao, K.-N. (2015). The role of the PI3K/Akt/mTOR signalling pathway in human cancers induced by infection with human papillomaviruses. *Mol. Cancer* 14, 87. doi:10.1186/s12943-015-0361-x

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 Zhu, Ren, Xu, Cheng, Wang, Ren, Xiao, Jiang, Ding and Tan. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fmolb.2021.744363

SNX20AR/MiRNA-301a-3p/SNX20 Axis Associated With Cell Proliferation and Immune Infiltration in Lung Adenocarcinoma

Yixiao Yuan[1,2†], Xiulin Jiang[3†], Lin Tang[1,2], Juan Wang[1,2], Qianqian Liu[1,2], Xiaolan Zou[2] and Lincan Duan[1,2]*

[1]Department of Graduate School of Kunming Medical University, Kunming, China, [2]Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University, Kunming, China, [3]Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China

Lung cancer is the most common tumor with severe morbidity and high mortality. Increasing evidence has demonstrated that SNX20 plays crucial roles in the progression of human cancer. However, the functions and mechanism of SNX20 in LUAD are still barely known. Here, we employ the TCGA, GEO and CCLE databases to examine the expression of SNX20 in human varies cancer, the results shown that SNX20 is down-regulated in lung Adenocarcinoma, SNX20 level was significantly positive correlated with poor prognosis and lung cancer immune cell infiltration. We found that over-expression of SNX20 significantly restrain NSCLC cell proliferation and migration. Subsequently, we discover a network regulating SNX20 in LUAD, further study found that the decreased of the SNX20 likely caused by DNA hypermethylation. Furthermore, we identified that SNX20AR/miRNA-301a-3p mediated decreased of SNX20 correlated with lung cancer progression and cancer immune infiltration in LUAD. Our findings suggested that ncRNAs play a crucial role in the regulatory network of SNX20. Collectively, our findings demonstrate the suppressor roles of the SNX20AR/miRNA-301a-3p/SNX20 axis in Lung Adenocarcinoma, represent that SNX20 have the potential of as an effective therapeutic target in future.

Keywords: lncRNA, miRNA-301a-3p, SNX20, LUAD, immune infiltration, cell proliferation

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Abhishek Jauhari,
University of Pittsburgh, United States
Sayan Chakraborty,
*Institute of Molecular and Cell Biology (A*STAR), Singapore*

***Correspondence:**
Lincan Duan
duanmumuhuosan@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *20 July 2021*
Accepted: *08 September 2021*
Published: *17 September 2021*

Citation:
Yuan Y, Jiang X, Tang L, Wang J, Liu Q, Zou X and Duan L (2021) SNX20AR/MiRNA-301a-3p/SNX20 Axis Associated With Cell Proliferation and Immune Infiltration in Lung Adenocarcinoma. Front. Mol. Biosci. 8:744363. doi: 10.3389/fmolb.2021.744363

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in the world (Sung et al., 2021). Lung adenocarcinoma is the most common molecular subtype of NSCLC cancer, and LUAD accounts for almost 50 percent of lung cancers. Although there are many treatments for lung cancer, the incidence rate and mortality rate of LUAD patients remains very high (Martin and Leighl, 2017). Therefore, it's extremely urgent to discerned accurate and sensitive immune-related biomarkers and elucidate the molecular mechanisms participate in LUAD progression.

SNX20, as a member of sorting nexins proteins family, play crucial role in the functional organization (Kurten et al., 1996; Carlton et al., 2004). Previous studies demonstrated that sorting nexins was divided into different groups according to the special functional domains, mainly ranging from BAR, PDZ, FERM-like, RGS to SH3 domains (Zheng, 2001; Wassmer et al., 2009; van Weering et al., 2012). In humans, the SNX subgroup mainly including the sorting nexin-20, sorting nexin-21 and the SNX-PXB proteins (Zeng et al., 2002). These proteins has a C-terminal PX-associated B (PXB) domain which is essential for protein interactions

(Clairfeuille et al., 2015). It has been shown that SNX20 able to regulation the endosomal trafficking and endothelial cell adhesion (Schaff et al., 2008). Finally, the SNX20 gene locus has recently been correlated to inflammatory bowel disease (Brant et al., 2017). Until now, there was no study focused on the function of SNX20 in LUAD progression. In this study, we aimed to investigate the role of SNX20 in Lung adenocarcinoma progression and tumor-infiltrating lymphocytes.

In this work, we find that SNX20 was significantly decreased in LUAD and it's low expression was correlated with poor prognosis, pathological stage and lymph node metastasis. GSEA analysis suggested that high SNX20 expression was mainly enriched with the immune-related signaling pathways, such as JAK STAT signaling pathway, T cell receptor signaling pathway and Toll like receptor signaling pathway. Additionally, SNX20 was positive correlated with the different immune infiltrations and immune check point related gene expression. We show that elevated the SNX20 expression was significantly inhibits the cell proliferation and migration of NSCLC cells. Subsequently, in this study, we identified a lncRNA, termed as SNX20AR (SNX20 associated lncRNA: ENSG00000258168) was high expressed in NSCLC cancerous tissues and predicts poor prognosis. We identified that SNX20AR/miRNA-301a-3p mediated decreased of SNX20 correlated with lung cancer progression and cancer immune infiltration in LUAD. Our findings suggested that ncRNAs play a crucial role in the regulatory network of SNX20. Collectively, our findings demonstrate the suppressor roles of the SNX20AR/miRNA-301a-3p/SNX20 axis in LUAD, represent that SNX20 have the potential of as an effective therapeutic target in future.

MATERIALS AND METHODS

The Expression, Prognosis, Clinical Information and Immune Infiltration Analysis

We mainly using the following databases to analysis the expression, prognosis, clinical information and Immune infiltration of SNX20 in cancers. The detail information of databases used in this study are as follows: Oncomine database (https://www.oncomine.org/) (Rhodes et al., 2004), TIMER (http://timer.cistrome.org/) (Li et al., 2017), UALCAN (http://ualcan.path.uab.edu/) (Chandrashekar et al., 2017), Kaplan-Meier plotter (http://kmplot.com/) (Nagy et al., 2021), TISIDB database (http://cis.hku.hk/TISIDB/) (Ru et al., 2019), the Human Disease Methylation Database (http://bio-bigdata.hrbmu.edu.cn/) (Lv et al., 2012), the PrognoScan database (http://dna00.bio.kyutech.ac) (Mizuno et al., 2009), The Linked Omics database (http://www.linkedomics.org/) (Vasaikar et al., 2018). GEPIA database (http://gepia.cancer-pku.cn/) (Tang et al., 2017) and cbioportal (http://www.cbioportal.org/) (Cerami et al., 2012).

Prediction and Construction the ceRNA Network

We employ the starbase (www.starbase.sysu.edu.cn) (Rhodes et al., 2004), Targetscan (http://www.targetscan.org/) (Li et al., 2017), miRDB (http://mirdb.org) (Chandrashekar et al., 2017) and miRbase (http://microrna.sanger.ac.uk/) (Nagy et al., 2021) to predict the potential miRNAs that able to binds with miRNAs. We also using the starbase (Rhodes et al., 2004) to analysis the expression and prognosis of miRNAs and lncRNAs. The starbase (www.starbase.sysu.edu.cn) (Nagy et al., 2021) and lncRNASNP (http://bioinfo.life.hust.edu.cn/lncRNASNP/) (Lv et al., 2012) was apply to analysis the upstream lncRNAs that bindings with the miRNAs. Furthermore, we using the starbase to analysis the correlation between the LncRNA/miRNA/mRNA.

Analysis the Protein and Gene Interation Networks

We employed the STRING (www.string-db.org) (Mering, 2003) and GeneMANIA (http://www.genemania.org) (Warde-Farley et al., 2010) perform protein-protein and gene-gene interaction network analysis of SNX20.

Analysis the Cell Localization and Coding Potential of lncRNA

We employed the lncLocator (http://www.csbio.sjtu.edu.cn/bioinf/lncLocator2/) (Cao et al., 2018) and CPC2 (http://cpc2.cbi.pku.edu.cn) analysis the Cell localization and coding potential of lncRNA.

Plasmids Construction and Cell Culture

The cDNA for SNX20 was constructed employ pCDH-CMV-MCS-EF1-Puro vector. The NSCLC related cells mainly purchased from The Kunming Institute of Zoology (KIZ) of Chinese Academy of Sciences (CAS). BEAS-2B and NSCLC related cells lines were cultured using the RPMI1640 medium, this medium contain the 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The cDNA for SNX20 primer sequences are list follows: SNX20-PCR-F: ATGGCAAGTCCAGAGCAC CCTG, SNX20-PCR-R: TCAGGGTGTGGCGTCAGGAGC CGGAGCCA.

Quantitative Real-Time PCR

The qRT-PCR assay was performed as documented (Jiang et al., 2018). For Real-time RT-PCR assay, indicated cells were lysed by RNAiso Plus (Takara Bio, Beijing, China, Cat# 108-95-2). Total RNA was extracted according to the manufacturer's protocol, and then reverse transcribed using RT reagent Kit (Takara Bio, Beijing, China, Cat# RR047A; TIANGEN Biotech, Beijing, China, Cat# KR211-02). Real-time PCR was performed by FastStart Universal SYBR Green Master Mix (Roche, Cat# 04194194001; TIANGEN Biotech, Beijing, China, Cat# FP411-02) using an Applied Biosystems 7,500 machine. The primers and antibodies used in this study are shown in xx. The primer sequences are list follows SNX20-F: ACCTGACGGGCACTT AGACA, SNX20-R: AGAGCAGTTTGACGTGCTTCC; β-actin-F: CTTCGCGGGCGACGAT, β-actin-R: CCATAG GAATCCTTCTGACC. The expression quantification was obtained with the $2-\Delta\Delta Ct$ method.

Cell Proliferation, Colony Formation and Cell Migration Assays

Cell proliferation, colony formation, tumor sphere formation assay was performed as documented (62). Briefly, for cell proliferation assay, indicated cells were plated into 12-well plates at a density of 1.5 × 104, the cell numbers were subsequently counted each day using an automatic cell analyzer countstar (Shanghai Ruiyu Biotech Co., China, IC 1000). For colony formation assay, indicated cells were seeded in 6-well plate with 600 cells per well supplemented with 2 ml cell culture medium, and the cell culture medium was changed every 3 days for 2~3 weeks. Indicated cells were fixed with 4% PFA and stained with 0.5% crystal violet.

Cell migration assays was performed as documented (62). Briefly, to produce a wound, the monolayer cells in 6-well plate were scraped in a straight line with pipette tips. Plate was then washed with warm PBS to remove detached cells. Photographs of the scratch were taken at indicated time points using Nikon inverted microscope (Ti-S). Gap width was calculated with GraphPad Prism software. For trans-well assay, 2.5–3×104 cells in 100 μL serum free medium were plated in an 8.0 μm, 24-well plate chamber insert (Corning Life Sciences, catalog no. 3422), with medium containing 10% FBS at the bottom of the insert. Cells were incubated for 24 h, and then fixed with 4% paraformaldehyde for 20 min. After washing, cells were stained with 0.5% crystal violet blue. The positively stained cells were examined under the microscope.

Dual-Luciferase Assay

Briefly, putative binding sites for miR-301a-3p on the 3′-UTR of SNX20 was predicted by starbase dataBase (http://starbase.sysu.edu.cn/). Wild-type and mutant SNX20 (mut- SNX20 or mut-SNX20) fragments were constructed and inserted downstream of the luciferase reporter gene in the reporter plasmid pGL3 plasmid (Promega). HEK-293T cells (2 × 104 cells/well) were seeded in a 24-well plate and co-transfected with 3′-UTR SNX20 construct and miR-301a-3p mimics or miR Ctrl using Lipofectamine 3,000. Both firefly and Renilla luciferase expressions were measured post-transfection using the Dual Luciferase Kit (Promega) according to the manufacturer's instructions.

Western Blotting

The Western Blotting and Immunohistochemistry staining assay was performed as documented (Griffiths-Jones, 2006). Briefly, Cell lysates were collected, perform the Western blot, primary antibody overnight incubation and second antibody incubation. Finally, develop using instrument.The detail information of antibodies employ in our study are as follows: SNX20 antibody (SNX20, FNab08087, Rabbit 1:1,000,) and β-actin (Catalog number: 66009-1-Ig, 1:20,000, Proteinch, Shanghai, China).

Statistical Analysis

The significance of the data between two assays groups was decided by Student's t-test, $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), was considered significantly.

RESULTS

SNX20 Was Decreased in Human Cancers

In order to examine the mRNA of SNX20 expression pattern in multifarious cancer, we employed the TIMER tools to analysis the expression of SNX20, the result shown that SNX20 was low expression in LUAD, LUSC, and PAAD, higher expression was observed in BRCA, CHOL, ESCA, GBM, HNSC, KIRC, and KIRP (**Figure 1A**). To further verify the results, we using the combine the TCGA and GTE databases to figure out the SNX20 expression. As is show in **Figure 1B**, the SNX20 was significantly up-regulation in COAD, GBM, KIRC, KIRP, LAML, PAAD, SKCM, SARC, and TGCT cancer than match healthy tissue. Besides, we found that the SNX20 was down regulated in NSCLC cells lines observe in CCLE network tools (**Figure 1C**). Above all, our findings indicated that the SNX20 may plays different roles in the progression of different cancers.

Correlations of SNX20 Expression With Pathological Stages in Cancers

We employ the GEPIA tools to examine the relationship between the expression of SNX20 and the human cancers pathological stage. Interestingly, we find that the expression of SNX20 was markedly positive with the pathological stage of KIRC, SARC and negative with the pathological stage of OV and THYM (**Figures 2A–D**). These results suggested that SNX20 plays different roles in different human cancers.

The Prognostic Values of SNX20 in Human Cancer

The prognostic value of SNX20 expression in human cancers was analyzed by several databases. In GEPIA, we found that lower SNX20 expression was associated with poorer overall survival (OS) in CESC, LUAD and SARC, the patients with higher SNX20 expression had poor OS in LGG and UVM (**Figures 3A–E**). Additionally, high expression of SNX20 has a better DFS observed in SKCM, LIHC and UCEC (**Figures 3F–H**). Lower expression of SNX20 was related to poor DSS in CESC, HNSC, LGG, LUAD, SKCM, UCEC and UVM (**Supplementary Figure S1A**), and linkage to poor PFS in CESC, HNSC, KIRP and UCEC (**Supplementary Figure S1B**). The above results proved that SNX20 expression closely related to the prognosis of various cancer types.

Associations Between SNX20 and Clinical Characteristics of LUAD Patients

Considering the significance of SNX20R in cancers, next, we want to exploration the correlation between SNX20 expression and clinical features in LUAD. First, we found that the lower expression was observed in the three GEO datasets (**Figures 4A–C**). Next, we find that the RNA of SNX20 was significantly lower in LUAD by perform the UALCAN tools analysis ((**Figure 4D**). In addition, we also find the expression of SNX20 was decreased with the elevation of stage nodal

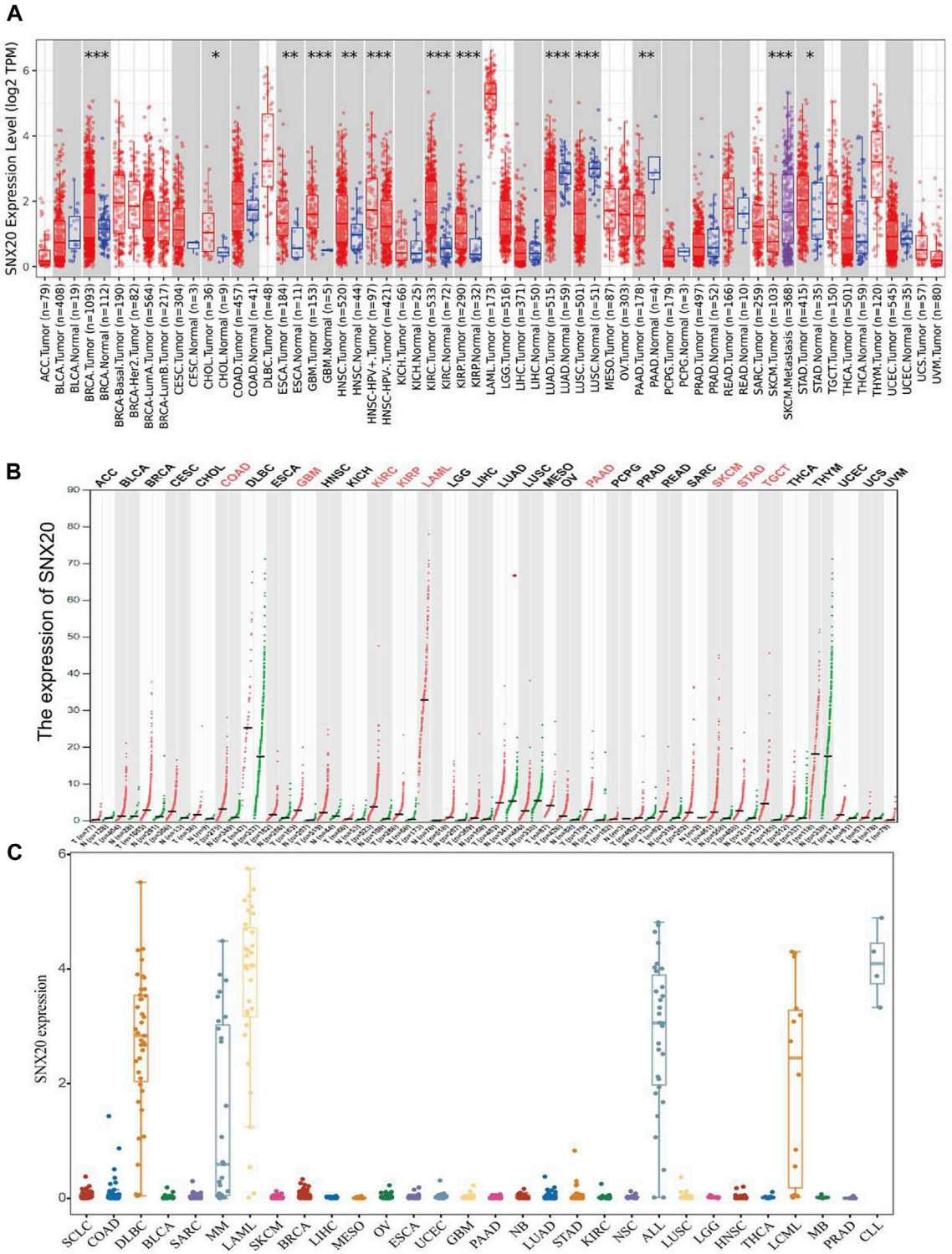


FIGURE 1 | Expression analysis for SNX20 in human cancers. **(A)** SNX20 expression of different tumor types in the TIMER database. **(B)** SNX20 expression in TCGA and GTEx datas. **(C)** SNX20 expression of different tumor cells lines in the CCLE database.

metastasis and tumor stage (**Figures 4E,F**). Surprisingly, we find higher expression of SNX20 has the better OS, PFS and PPS (**Figures 4G–I**). Finally, we adopt the GEO dataset to verify above results, the studies shown that elevated the SNX20 expression display the better OS and RFS in NSCLC (**Figures 4J,K**). ROC curve analysis of SNX20 showed an AUC value of 0.884 in TCGA

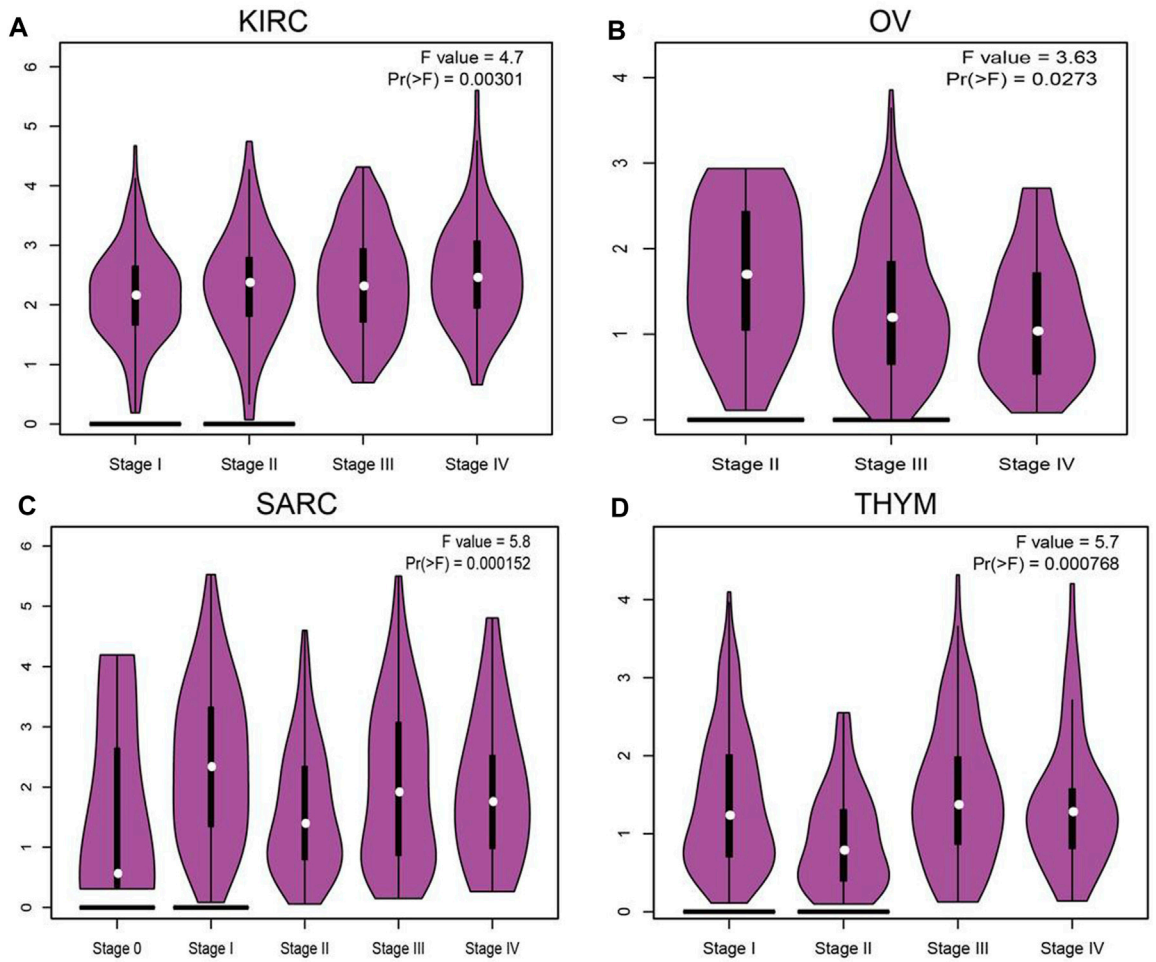


FIGURE 2 | The pathological stage analysis for SNX20 in human cancers. **(A–D)** The pathological stage of SNX20 in KIRC **(A)**, OV **(B)**, SARC **(C)**, and THYM **(D)**.

LUAD patients (**Figure 4L**). These results indicate that SNX20 has the potential to act as a detection index for the diagnosis of lung cancer with high sensitivity and specificity.

Analysis the Gene Mutation of SNX20 in NSCLC

For explore the gene mutation information about the SNX20, we employ the cBioportal tools preform comprehensive analysis regard to the SNX20. The result shown that the mutation rate of SNX20 reached 1.8% in NSCLC (**Figures 5A,B**), the results also display the mutation of SNX20 in different NSCLC molecular Subtypes (**Figure 5C**), Next, we also examine the mutation type and base mutation in NSCLC, we found that Missense substitution and base G > A reached the highest mutation rate in NSCLC (**Figures 5D,E**). Overall, these results emphasize the gene mutation of SNX20 may be contribute to the SNX20 low expression in NSCLC.

Correlations of SNX20 Expression With DNA Methylation

As a one of the crucial epigenetic modification, DNA Methylation plays an significant roles in regulation gene expression. In order to explore the molecular mechanism of the SNX20 aberrantly up-regulated in LUAD, we analyzed the promoter methylation levels of SNX20 in LUAD. We found that there are many methylation sites in the promoter region of SNX20, and the differential methylation regions were indicated in the heatmaps (**Figure 6A**). Importantly, by using the shiny methylation analysis resource tool (SMART) analysis (Tay et al., 2014) we uncovered that the methylation of SNX20 was significantly higher in LUAD cancerous tissues compared to that in normal tissues (**Figure 6B**). Consistently, we found that the methylation levels on the specific methylation site (cg06207201) within SNX20 promote region negatively correlated with its expression in LUAD (**Figure 6C**). Furthermore, we showed that the elevated

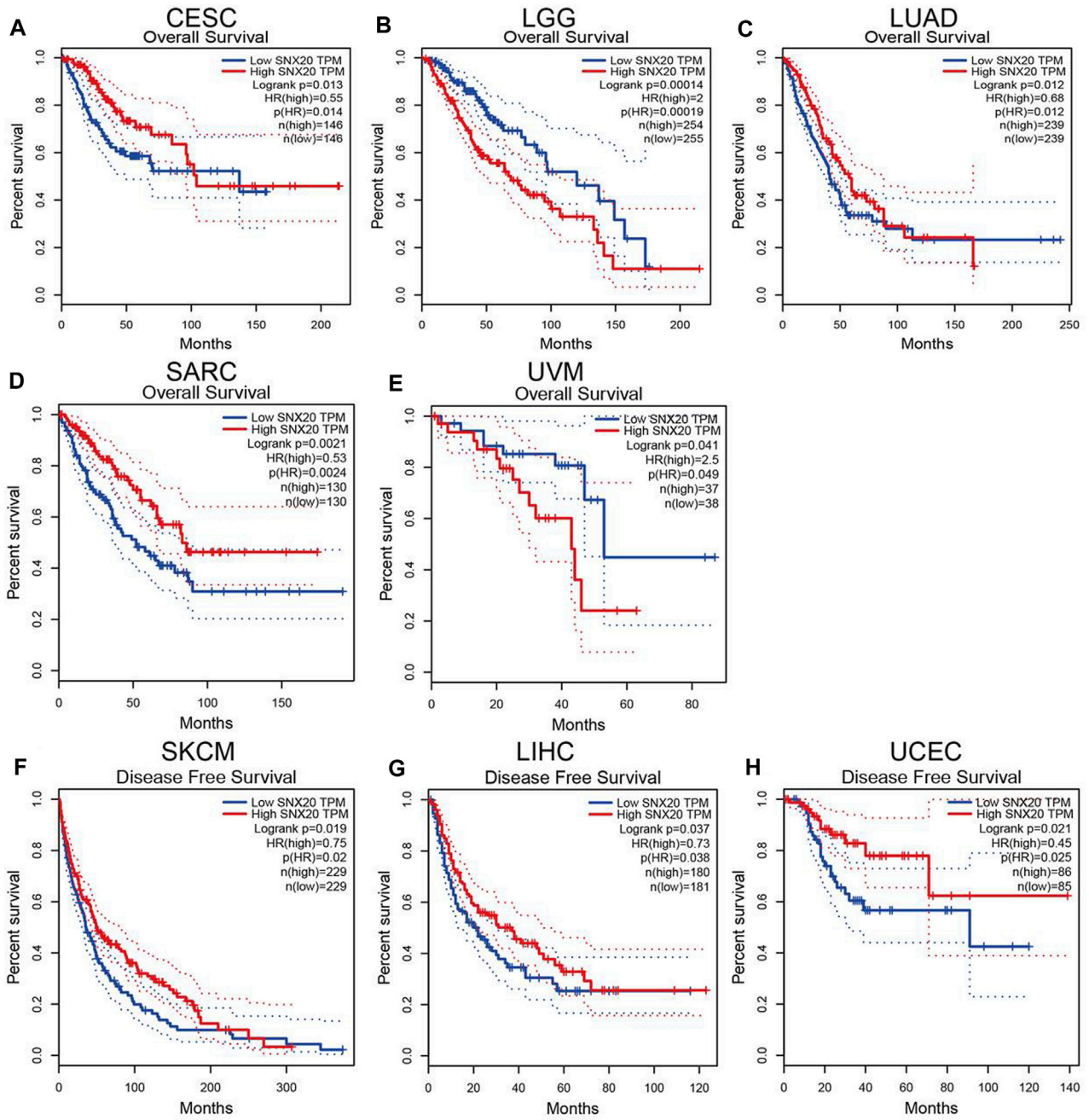


FIGURE 3 | The prognostic values in various cancer subgroups of SNX20. **(A–H)** Prognostic HR of SNX20 in different cancers for OS **(A–E)**, DFS **(F–H)**.

methylation levels on cg06207201 site correlates with worse OS in the TCGA-LUAD cohorts, using the methSurv dataset (**Figure 6D**). Additionally, we also found that another two methylation sites (cg01144086 and cg08330349) also negatively with the expression of SNX20 in LUAD (**Figure 6E**). Above all, there results suggested the DNA methylation plays crucial roles in the modulates the expression of SNX20 in LUAD.

Analysis of the Function of SNX20 in LUAD

To further study the function of SNX20, we employed the Linkedomics to perform a correlation analysis of SNX20. Two heatmaps were constructed to illustrate the genes whose expression was most positively and negatively correlated with that of SNX20 (**Figures 7A,B**). GO annotation revealed that these genes participate in various immune response, including T cell activation, interferon γ production, adaptive immune response,

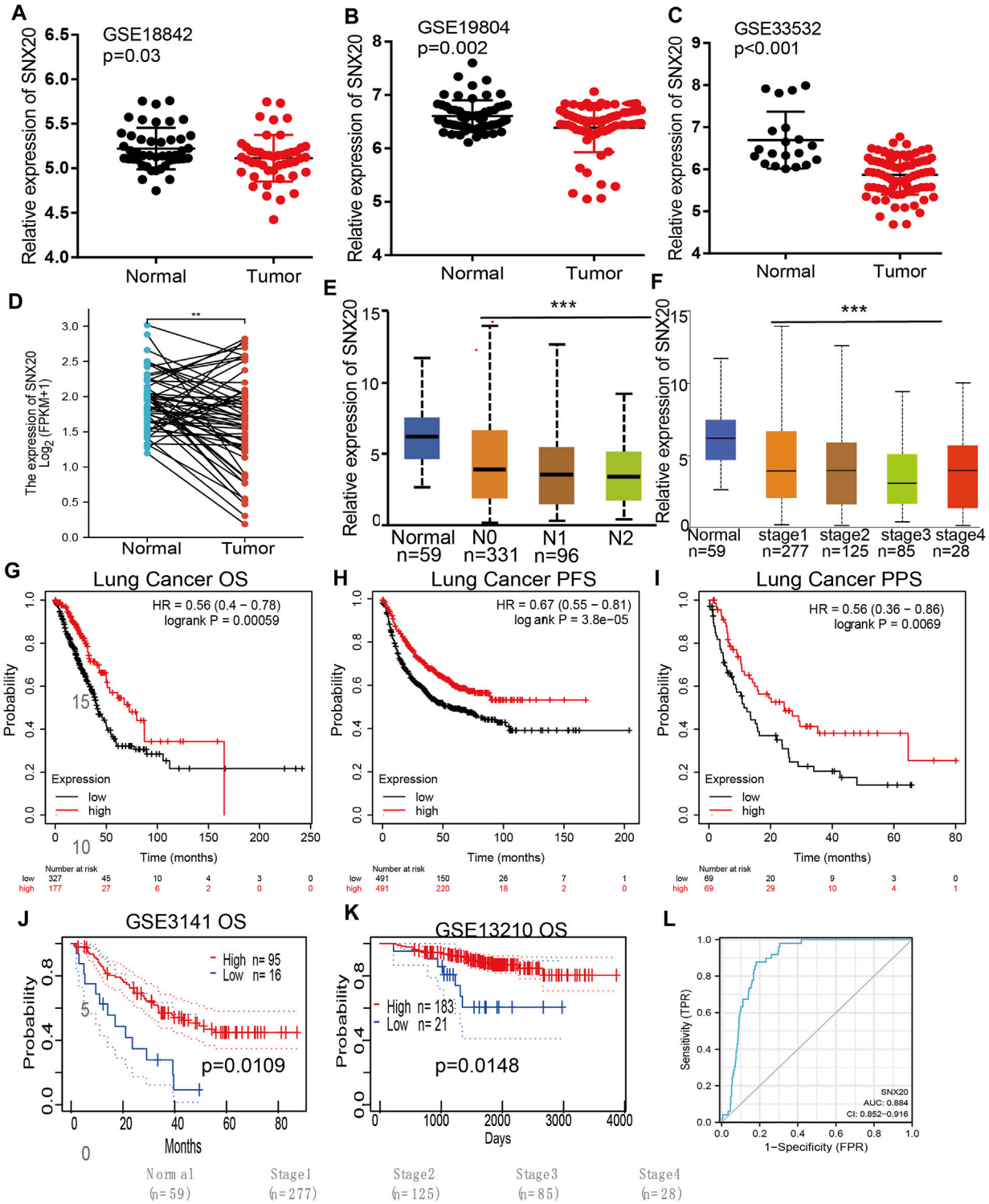


FIGURE 4 | SNX20 was highly expression in NSCLC. **(A–C)** The expression of SNX20 in GEO database. **(D)** The pathological stage analysis for SNX20 in LUAD determined by GEPIA database. **(E)** The expression of SNX20 in 58 pairs of LUAD tissues and adjacent normal tissues. **(F)** The mRNA of SNX20 in patient's tumor stage by UALCAN database. **(G–I)** The different survival state in LUAD, includings OS, PFS, PPS. **(J–K)** The different survival state in GEO databases, includings OS, RFS. **(L)** ROC curve analyses and AUC values for SNX20 in TCGA LUAD patients.

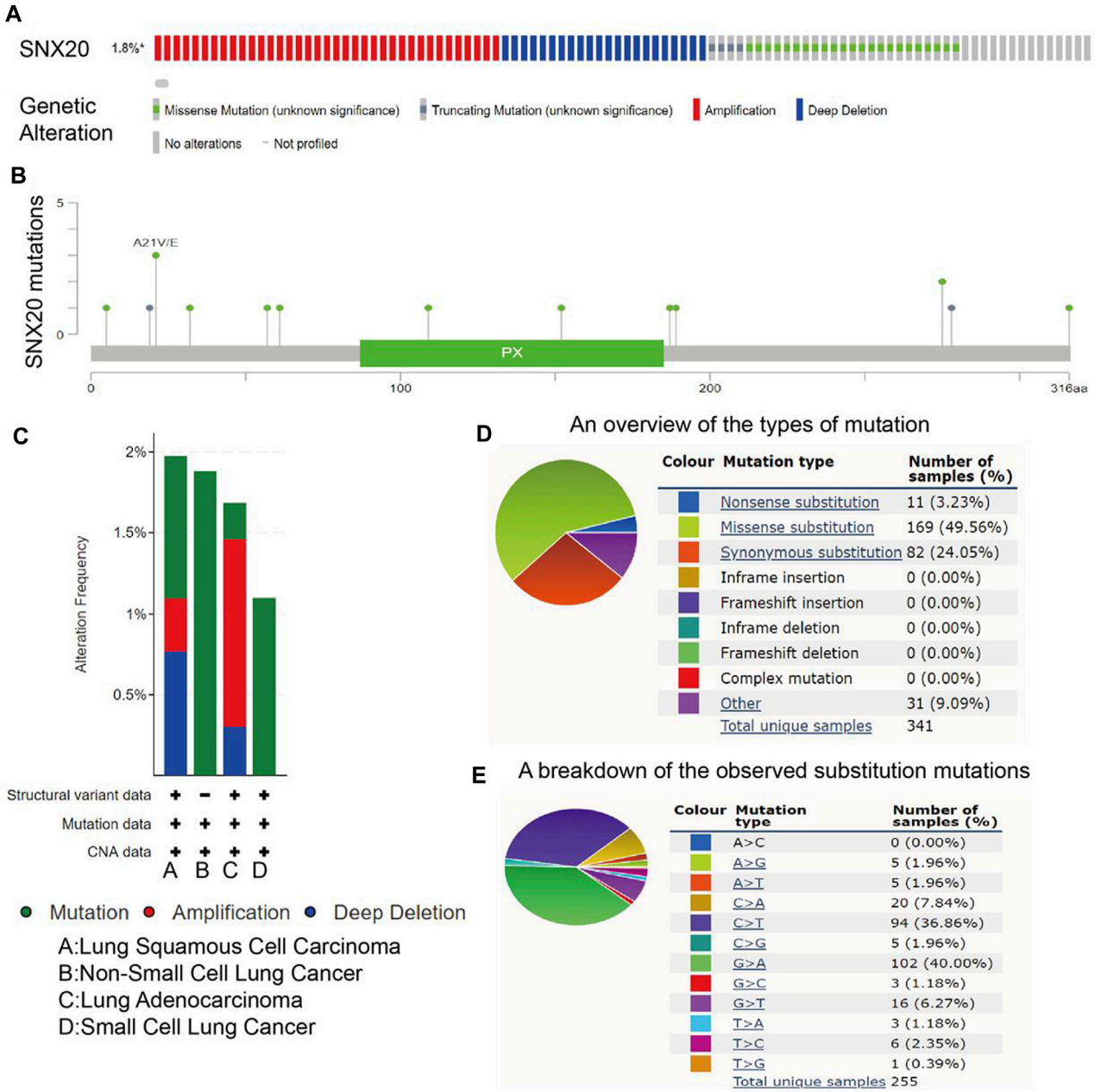


FIGURE 5 | Analysis of the gene mutation of SNX20. **(A–B)** The picture indicated that SNX20 mutations (TCGA) using the cBioportal. **(C)** The picture indicated that gene mutation of SNX20 in different NSCLC histological subtype. **(D–E)** The picture indicated that different mutation types of SNX20 in NSCLC.

leukocyte proliferation, regulation of defense response to virus by virus, myeloid dendritic cell activation, interleukin-4 production, leukocyte activation involved in inflammatory response, lymphocyte activation involved in immune response, response to chemokine, immune response regulating signaling pathway and regulation of leukocyte activation (**Figure 7C**). KEGG pathway analysis showed the enrichment in intestinal immune network for IgA production, primary immunodeficiency, hematopoietic cell lineage, T cell receptor signaling pathway, cell adhesion molecules, Th1 and Th2 cell differentiation, Th17 cell differentiation, Natural killer cell medicated cytotoxicity and Fc epsilon RI signaling pathway (**Figure 7D**). Furthermore, we also employed the GeneMania and STRING databases construction the gene and protein interaction networks of SNX20 in LUAD. The results indicated that the twenty frequently altered genes was correlated with SNX20, including SELPLG, ENOSF1 and NCF (Figure xx), the significantly correlated with the SNX20 protein mainly including ZNRF2, METTL6, FKBP14, HECA and SELPLG (**Figures 7E,F**).

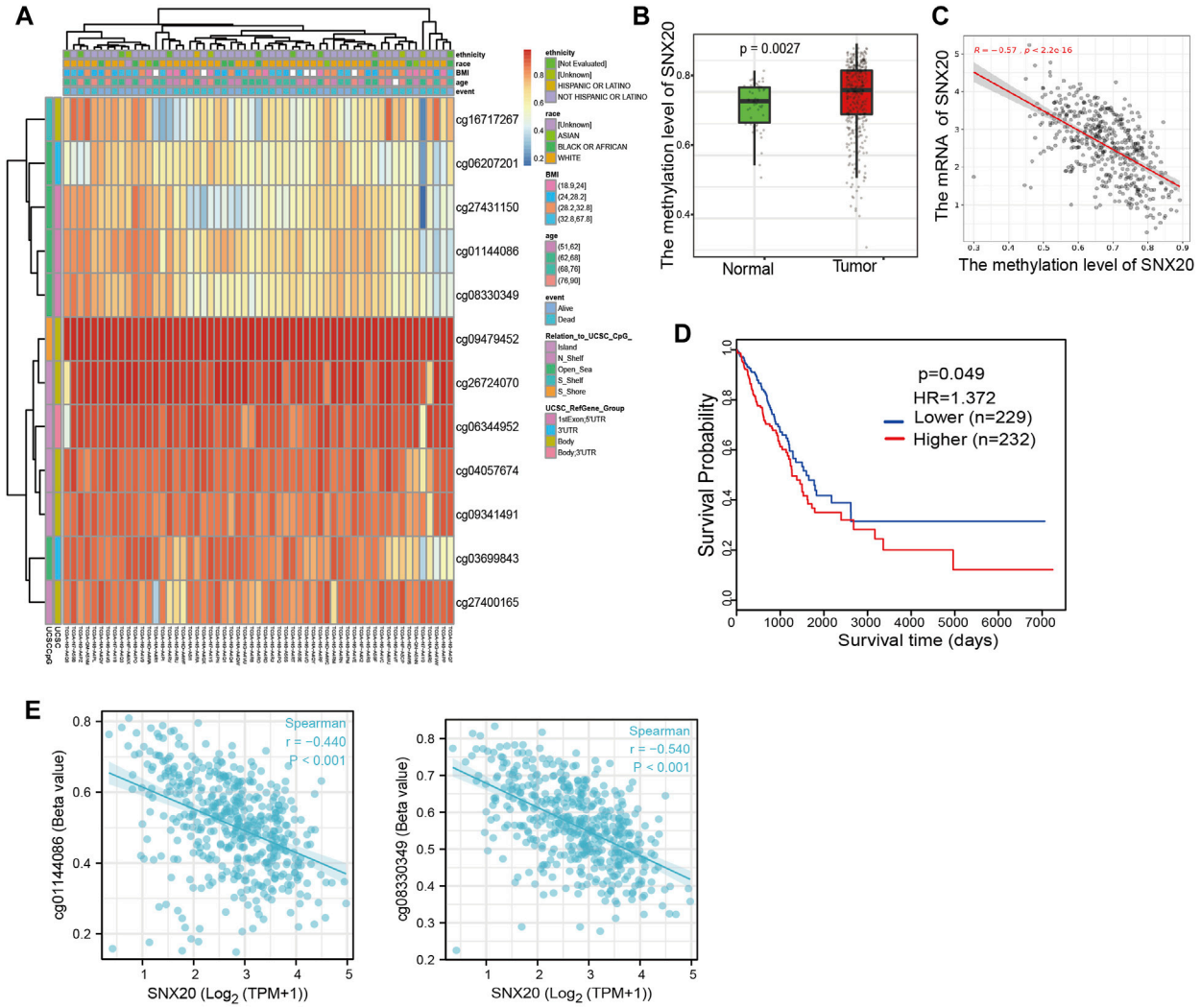


FIGURE 6 | Analysis of the DNA methylation level of SNX20. **(A)** The differential methylation sites linked to SNX20 were show as heatmaps. **(B)** The expression and methylation of SNX20 in LUAD tissues and normal tissues. **(C)** The correlation between the methylation level and mRNA expression of SNX20. **(D)** Survival state of the methylation level of SNX20 in the TCGA LUAD dataset. **(E)** The correlation between the diverse methylation sites and SNX20 expression in LUAD.

Identification of SNX20-Associated Signaling Pathways Using GSEA

To further explore the molecular mechanisms affected by SNX20 in Lung adenocarcinoma, we perform the GSEA enriched analysis and find SNX20 mainly participate in the immune related biology processes. GO enriched results indicated that SNX20 were enriched mainly in regulation of adaptive immune response, positive regulation of immune effector process, CD4 positive β T cell activation and B cell differentiation (**Figure 8A**). Similarly, The KEGG enriched results shown that SNX20 were enriched mainly involve in chemokine signaling pathway, JAK STAT signaling pathway, T cell receptor signaling pathway and Toll like receptor signaling pathway (**Figure 8B**). These results strongly indicated that SNX20 was mainly involved in regulation the immune response of lung adenocarcinoma.

Analysis the Upstream miRNA of SNX20

To investigate whether some miRNAs regulates the SNX20 expression by miRNA sponge manner, we employed the starbase (Li et al., 2014), Targetscan (Agarwal, 2015), miRDB (Wong and Wang, 2015) and miRbase (Griffiths-Jones, 2006) to predict the potential miRNAs that able to binds with miR-301a-3p. The results find that has-miR-34c-5p, has-miR-301a-3p, has-miR-338-5p, and has-miR-3614-5p may be bind with SNX20. Next, we employed the Starbase and Kmplot databases analysis the expression and prognosis of SNX20 in Lung cancer (**Figures 9A,B**). According to

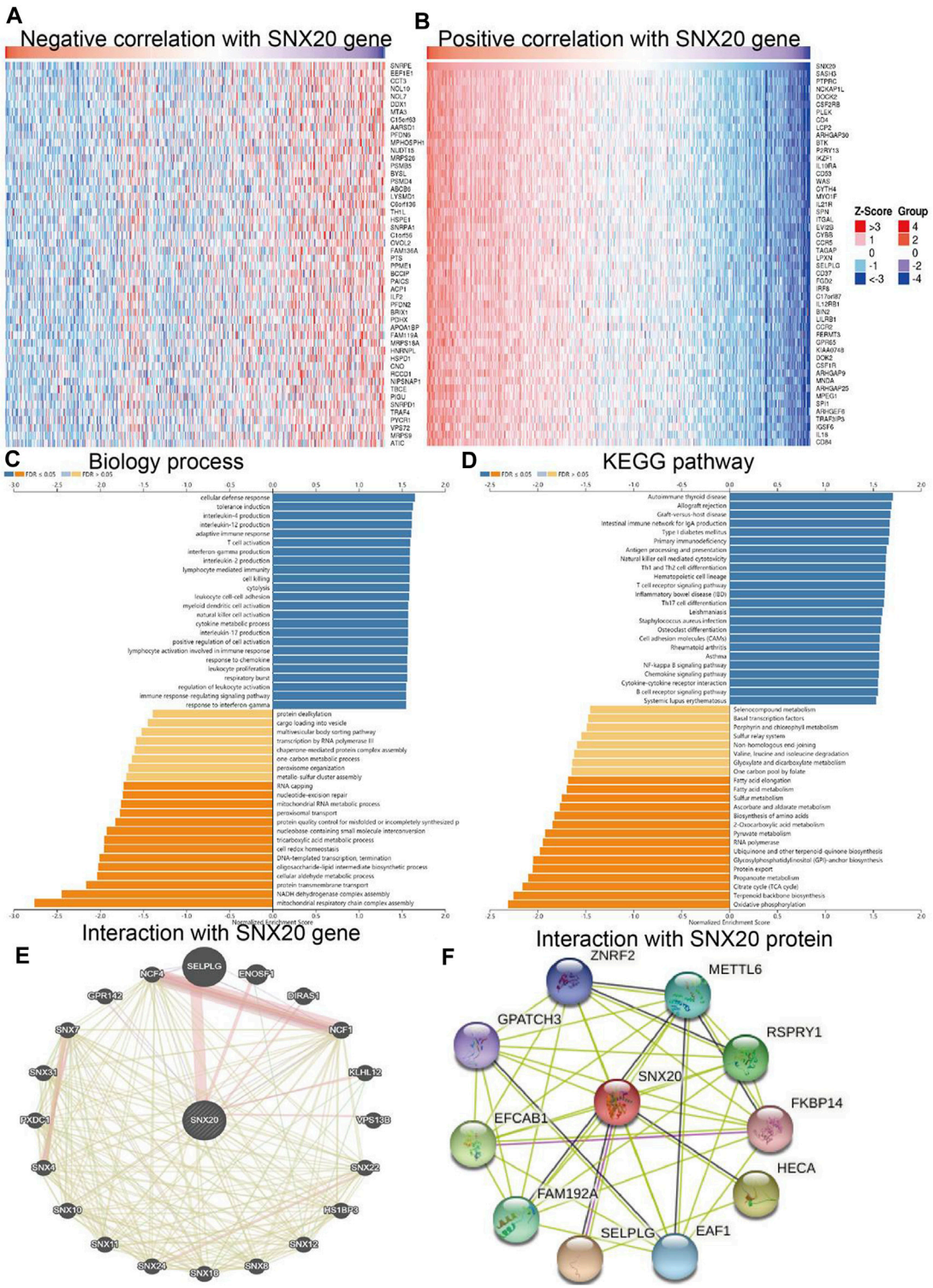


FIGURE 7 | GO and KEGG enrichment analysis for SNX20. **(A–B)** Heat maps of genes positively and negatively correlated with SNX20 (top 50). **(C–D)** Biology process and KEGG pathway analysis of SNX20 by GSEA. **(E)** The gene interaction meshwork of SNX20 was constructed using GeneMania. **(F)** Employ STRING to construction the protein interaction meshwork of SNX20.

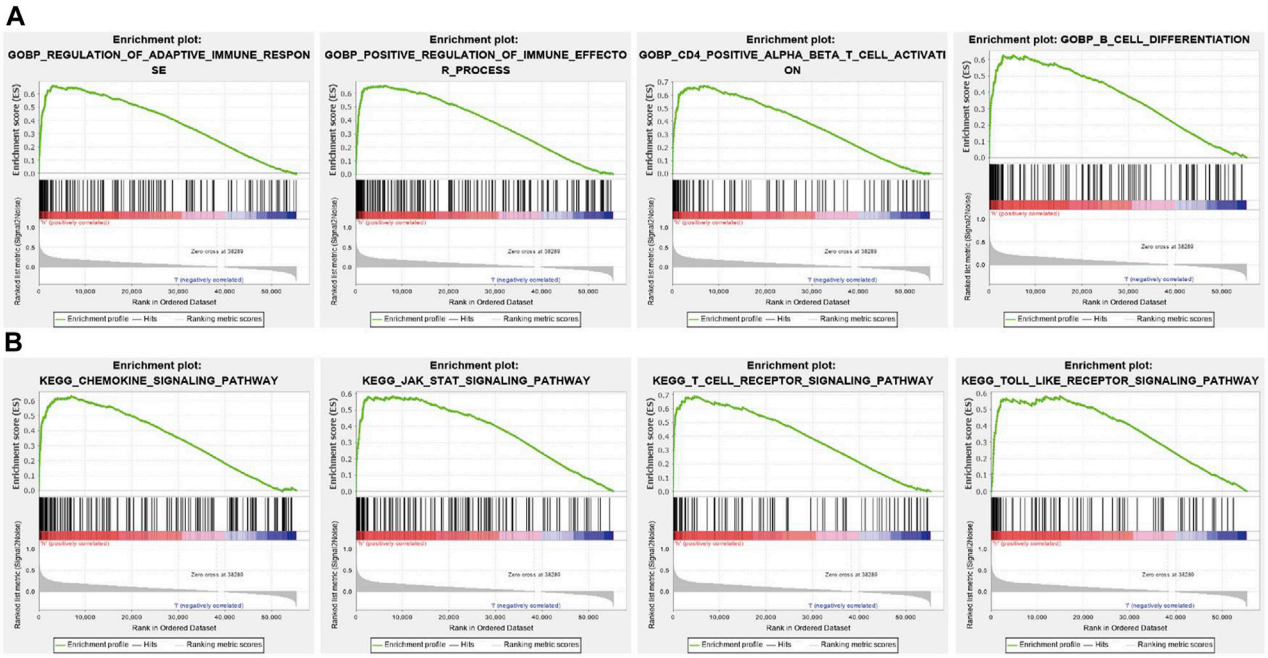


FIGURE 8 | GSEA for SNX20. **(A)** The GO enriched result for SNX20 expression. **(B)** The KEGG pathway enriched result for SNX20 expression.

competing endogenous RNAs hypothesis, the miRNA expression would be negative with the mRNA expression (Tay et al., 2014). Among there miRNAs, only has-miR-301a-3p expression was negatively correlated with SNX20 in LUAD patients (**Figure 9C**), we also analysis the base pair relationship between the miR-301a-3p and SNX20 (**Figure 9D**). By perform comprehensive analysis the expression, prognosis and expression correlation, we confirmation the miR-301a-3p may be as aN miRNA sponge for SNX20 (**Figure 9E**). Furthermore, we adopt the NSCLC GEO dataset to verify the expression of miR-301a-3p in lung cancer, the analysis results shown that miR-301a-3p expression lower in the whole blood and tissue in human lung cancer (**Figures 9F,G**). To investigate the effects of miR-301a-3p on the expression of target gene, we conducted overexpression analysis. Overexpressing miR-301a-3p significantly reduced both mRNA and protein expressions of SNX20 in A549 cell (**Figures 9H–J**). The luciferase assay showed that transfection of miR-30a-5p mimics significantly reduced the relative luciferase activity of SNX20-3UTR-WT-treated lung cancer cells, but did not affect that of SNX20-3UTR-MUT-treated lung cancer cells (**Figure 9K**). All these results confirmed the tumor suppressor roles of miRNA-30a-5p in the lung cancer progression.

SNX20AR Functions as a Sponge for miR-301a-3p

To explore the upstream lncRNAs that binding with miR-301a-3p, we employed the starbase (Nagy et al., 2021) and lncRNASNP (Gong et al., 2015) analysis the upstream lncRNAs. We find 4 possible lncRNA, including the AC01086.4, XIST, SNX20AR and SLC26A4-AS1. We first examine the expression and prognosis of lncRNAs by employed the starbase and Kmplot (**Figures 10A,B**). According to competing endogenous RNAs hypothesis, the miRNA expression would be negative with the lncRNA expression, and lncRNA should positive correlation with the mRNA expression (Vasaikar et al., 2018). Owing to the mRNA was decreased expression in LUAD and miRNA was highly expression in LUAD, the lncRNA should be low expression in LUAD. By perform comprehensive analysis the expression, prognosis and expression correlation, we confirmation the SNX20AR may be as an lncRNA sponge for miR-301a-3p. Next, the correlation analysis indicated that the SNX20AR expression not only negatively correlated with miR-301a-3p, but also positively correlated with SNX20 in LUAD (**Figures 10C,D**).We also analysised the base pair relationship between the miR-301a-3p and SNX20AR (**Figure 10F**). Finally, we find that the SNX20AR expression was decreased with the cancer stage elevated, and adopt the NSCLC GEO dataset to verify the expression of SNX20AR in lung cancer, the analysis results shown that SNX20AR expression lower in the in human lung cancer (**Figures 10G,H**). In order to explored the molecular characteristics of SNX20AR, we perform the related analysis by employed the lncLocator (Cao et al., 2018) and Coding Potential Calculator (CPC) (Kong et al., 2007) find that the SNX20AR not only mainly in the cytoplasm, but also no possess protein coding potential (**Figures 10I,J**). Thus, the data show that SNX20AR may be as a sponge for miR-33a-5p in LUAD.

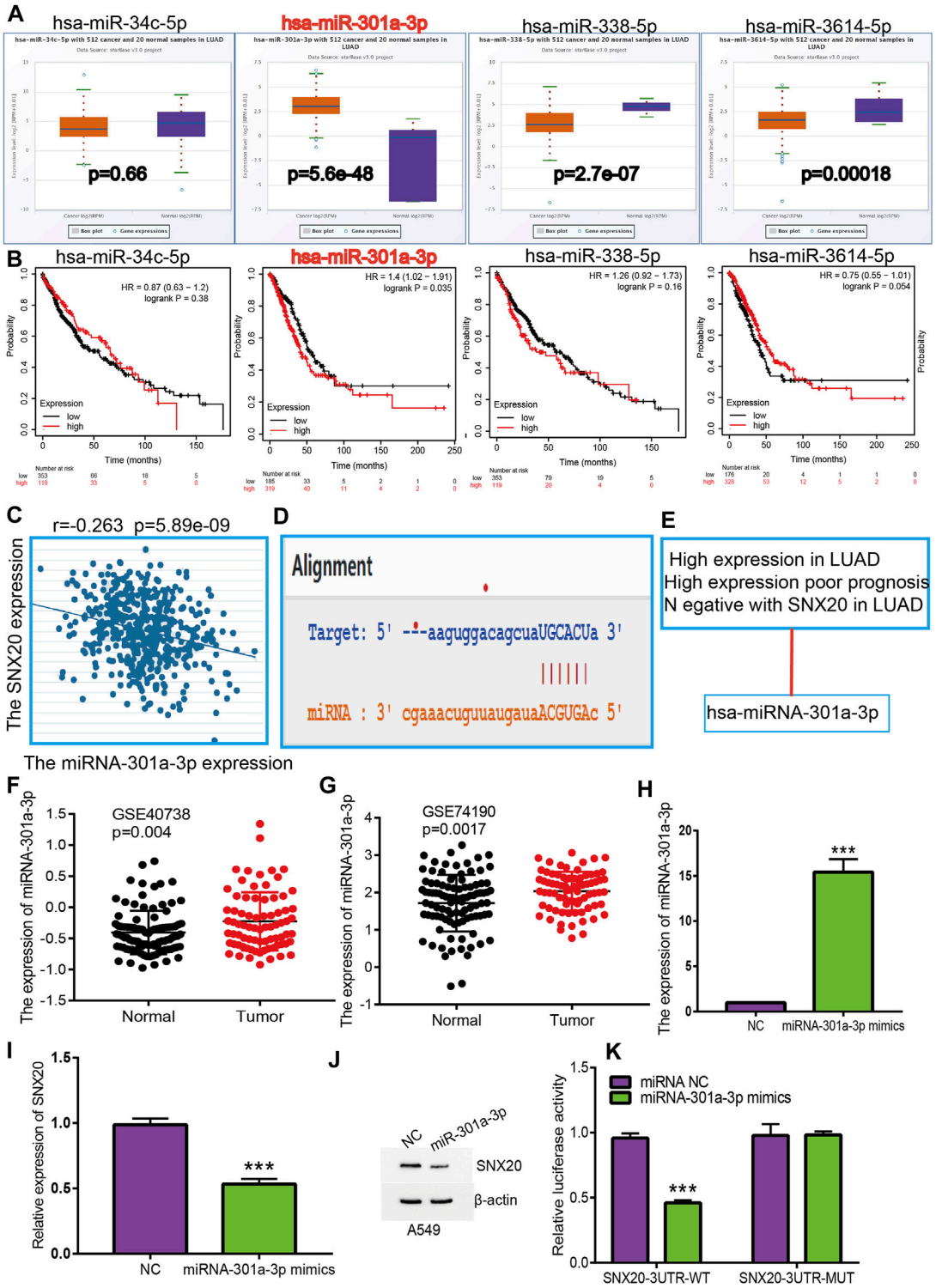


FIGURE 9 | Identification of miRNA-301a-3p could binding the SNX20 in LUAD. **(A)** The expression of miRNA-34c-5p, miRNA-301a-3p, miRNA-338-5p, and miRNA-3614-5p in TCGA-LUAD. **(B)** The prognostic values of miRNA-34c-5p, miRNA-301a-3p, miRNA-338-5p and miRNA-3614-5p in TCGA-LUAD. **(C)** The correlation between expression of miRNA-301a-3p and SNX20 in LUAD. **(D)** Base pairing between miRNA-301a-3p and the SNX20L 3′ UTR predicted by starBase. **(E)** The reasons for obtaining miRNA-301a-3p. **(F)** The expression of miRNA-301a-3p in GSE40738 and control group examine *via* GEO database. **(G)** The expression of miRNA-301a-3p in GSE74190 and control normal samples determined by GEO database. **(H)** The expression of miRNA-301a-3p in A549 cell after transfection miRNA-301a-3p mimics examined by using qRT-PCR assay. **(I)** The expression of SNX20 in A549 cell after transfection miRNA-301a-3p mimics examined by using qRT-PCR assay. **(J)** The protein of SNX20 in A549 cell after transfection miRNA-301a-3p mimics examined by using Western blot assay. **(K)** The relative luciferase activities were analyzed in A549 cell cotransfected with the miR-301a-3p mimics or mimics NC and the SNX20 3′UTR wild-type (WT) or mutant (MUT) luciferase reporter vectors.

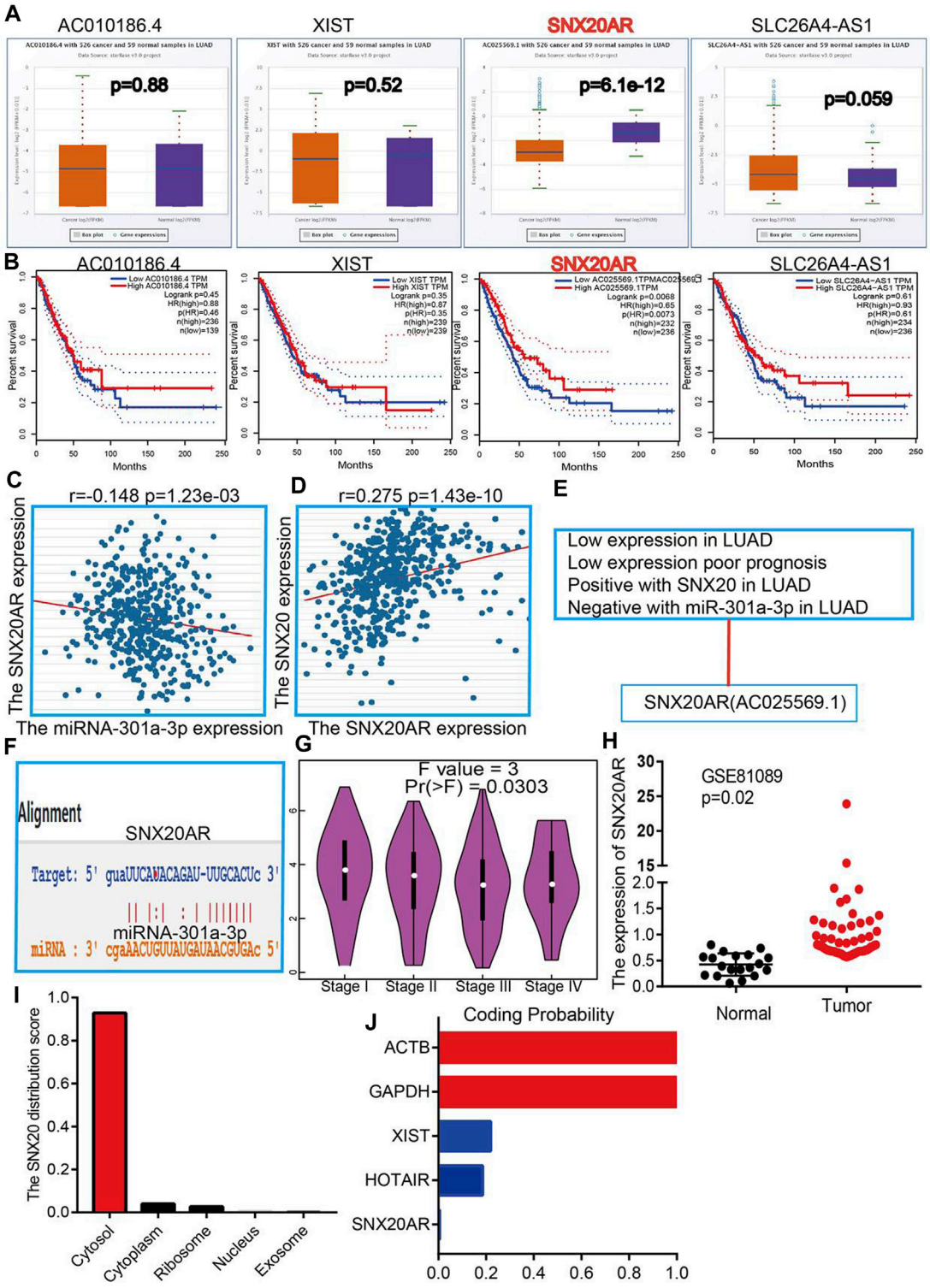


FIGURE 10 | Forecast the Upstream lncRNAs of miRNA-301a-3p. **(A)** The expression of AC010186.4, XIST, SNX20AR and SLC26A4-AS1 in TCGA-LUAD. **(B)** The prognostic values of AC010186.4, XIST, SNX20AR and SLC26A4-AS1 in TCGA-LUAD. **(C)** The correlation between expression of miRNA-301a-3p and SNX20AR in LUAD. **(D)** The correlation between expression of SNX20 and SNX20AR in LUAD **(E)** The reasons for obtaining SNX20A. **(F)** Base pairing between miRNA-301a-3p and the SNX20AR predicted by starBase. **(G)** The pathological stage analysis for SNX20AR in LUAD determined by GEPIA database. **(H)** The expression of SNX20AR in GSE81089 and control normal samples determined by GEO database. **(I–J)** Analysis the Cell localization and coding potential of lncRNA by employed the lncLocator and Coding Potential Calculator.

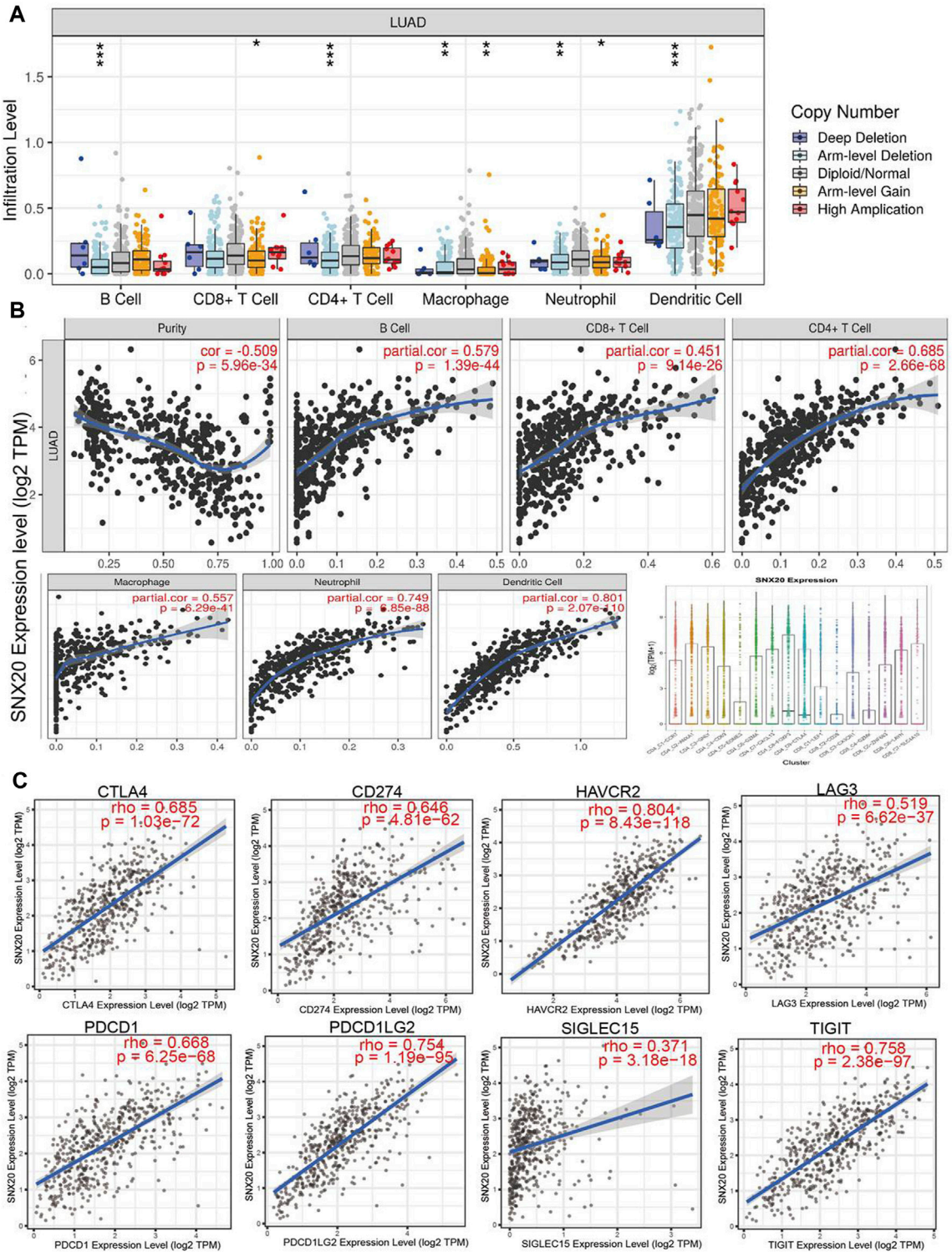


FIGURE 11 | The relationship between immune infiltration and the expression of SNX20 in LUAD. **(A)** The SNX20 gene copy numbers affect the infiltration level of different immune cells in LUAD. **(B)** SNX20 is positively linked to the infiltration of different immune cells employed the TIMER databases. **(C)** SNX20 is positively linked to the immune checkpoints related gene employed the TIMER databases.

TABLE 1 | The cox proportional hazard model of SNX20 and different tumor-infiltrating immune cells in LUAD by perform the TIMER databases analysis.

	LUAD				555
	Coef	HR	95%CI_l	95%CI_u	*p*.value
B_cell	−4.888	0.008	0.001	0.098	0
CD8_Tcell	0.466	1.594	0.268	9.488	0.608
CD4_Tcell	3.414	30.401	1.974	468.135	0.014
Macrophage	−0.272	0.762	0.061	9.502	0.833
Neutrophil	0.672	1.958	9.037	104.214	0.74
Dendritic	0.595	1.813	0.414	7.935	0.43
SNX20	−0.382	0.682	0.452	1.029	0.068

TABLE 2 | The correlation between SNX20 expression and different tumor lymphocyte infiltration in human cancer by perform the TISIDB databases analysis.

	LUAD	
	R	*p*
Activated CD8 T cell (Act _CD8)	0.542	***
Central memory CD8 T cell (Tcm _CD8)	0.498	***
Activated CD4 T cell (Act _CD4)	0.842	***
Central memory CD4T cell (Tcm _CD4)	0.404	***
Effector memory CD4 T cell (Tem _CD4)	0.415	***
T follicular helper cell (Tfh)	0.517	***
Gamma delta T cell (Tgd)	0.846	***
Type 1 T helper cell (Th1)	0.423	***
Type 17 T helper cell (Th17)	0.779	***
Type 2 T helper cell (Th2)	0.485	***
Regulatory T cell (Treg)	0.3	***
Activated B cell (Act _B)	0.782	***
Immature B cell (Imm _B)	0.683	***
Memory B cell (Mem _B)	0.797	***
natural killer cell (NK)	0.464	***
CD56bright natural killer cell (CD56bright)	0.726	***
CD56dim natural killer cell (CD56dim)	0.406	***
Myeloid derived suppressor cell (MDSC)	0.182	***
Natural killer T cell (NKT)	0.853	***
Activated dendtritic cell (Act _DC)	0.716	***
Plasmacytoid dendtritic cell (pDC)	0.62	***
Immature dendtritic cell (iDC)	0.516	***
Macrophage (Macrophage)	0.229	***
Eosinophi (Eosinophil)	0.736	***
Mast (Mast)	0.58	***
Monocyte (Monocyte)	0.46	***
Neutrophil (Neutrophil)	0.425	***

*, **, *** *indicate* p<*0.05*, p<*0.01*, p<*0.001*, *respectively.*

SNX20 Expression is Correlated With Immune Infiltration and Immune Checkpoints Related Gene Expression in LUAD

To explore the significances of SNX20 in the tumor microenvironment of LUAD. We employed the TIMER to analysis the correlation of SNX20 level with immune infiltration in Lung adenocarcinoma. We find that SNX20 CNV was significantly affect the immune infiltration levels of the immune cells, including the B cell, CD8$^+$T cells, CD4$^+$T cells, Macrophage, Neutrophil and Dendritic cell (**Figure 11A**). In addition, we also show that the expression of SNX20

TABLE 3 | The correlation analysis between SNX20 and different immune cells related gene markers.

Description	Gene markers	LUAD			
		None		Purity	
		Cor	*p*	Cor	p
B cell	CD19	0.523	***	0.394	***
	CD79A	0.482	***	0.354	***
CD8$^+$ T cell	CD8A	0.607	***	0.511	***
	CD8B	0.495	***	0.41	***
Dendritic cell	ITGAX	0.756	***	0.702	***
	NRP1	0.314	***	0.298	***
	CD1C	0.553	***	0.491	***
	HLA-DPA1	0.752	***	0.704	***
	HLA-DRA	0.737	***	0.675	***
	HLA-DQB1	0.757	***	0.488	***
	HLA-DPB1	0.571	***	0.702	***
M1 Macrophage	PTGS2	−0.124	*	−0.146	*
	IRF5	0.596	***	0.54	***
	NOS2	0.176	**	0.091	0.56
M2 Macrophage	MS4A4A	0.689	***	0.623	***
	VSIG4	0.658	***	0.611	***
	CD163	0.7	***	0.643	***
Monocyte	CSF1R	0.848	***	0.816	***
	CD86	0.81	***	0.764	***
	KIR2DS4	0.258	***	0.189	***
	KIR3DL3	0.106	***	0.08	***
	KIR3DL2	0.347	***	0.265	***
	KIR3DL1	0.22	***	0.154	***
Natural killer cell	KIR2DL4	0.247	***	0.167	**
	KIR2DL3	0.284	***	0.208	***
	KIR2DL1	0.209	***	0.154	***
Neutrophils	CCR7	0.717	***	0.626	***
	ITGAM	0.785	***	0.753	***
	CEACAM8	0.313	***	0.321	***
T cell (general)	CD3D	0.686	***	0.585	***
	CD3E	0.759	***	0.683	***
	CD2	0.8	***	0.736	***
T cell exhaustion	CTLA4	0.685	***	0.591	***
	LAG3	0.519	***	0.434	***
	HAVCR2	0.804	***	0.755	***
	GZMB	0.428	***	0.314	***
	PDCD1	0.668	***	0.592	***
TAM	CCL2	0.498	***	0.404	***
	IL10	0.623	***	0.535	***
	CD68	0.689	***	0.641	***
Tfh	BCL6	0.144	***	0.148	***
	IL21	0.427	***	0.383	***
Th1	TBX21	0.661	***	0.578	***
	STAT4	0.609	***	0.51	***
	STAT1	0.517	***	0.45	***
	IFNG	0.489	***	0.401	***
Th2	GATA3	0.544	***	0.439	***
	STAT6	0.294	***	0.345	***
	STAT5A	0.785	***	0.739	***
Th17	STAT3	0.216	***	0.252	***
	IL17A	0.256	***	0.189	***
Treg	FOXP3	0.764	***	0.707	***
	CCR8	0.776	***	0.728	***
	STAT5B	0.476	***	0.496	***
	TGFB1	0.577	***	0.52	***

*, **, *** *indicate* p<*0.05*, p<*0.01*, p<*0.001*, *respectively.*

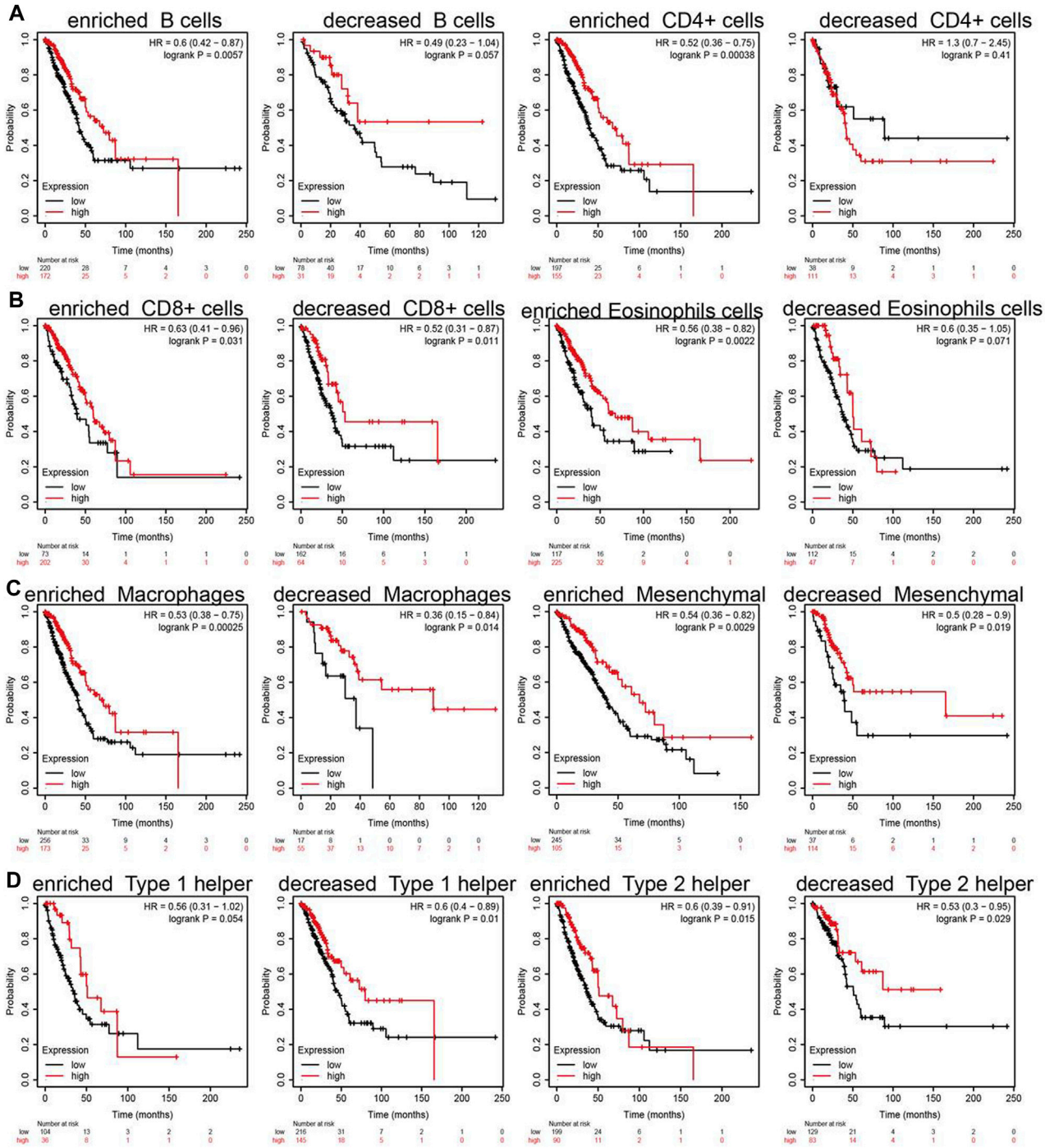


FIGURE 12 | Overall survival curves based on the expression of SNX20 in immune cell subgroups in LUAD. **(A–D)** Correlations between SNX20 expression and OS in different immune cell infiltration group in LUAD patients.

were negatively correlated with tumor purity (r = −0.509, p = 5.96e-34) and positively associated with the immune infiltration of B cells (r = 0.579, p = 1.39e-44), $CD8^+$ T cell (r = 0.451, p = 9.14e-26), $CD4^+$ T cell (r = 0.685, p = 2.66e-28), Macrophage (r = 0.557, p = 6.29e-41), Neutrophil (r = 0.801, p = 2.07e-110) and Dendritic cell (r = 0.328, p = 5.46E-10) in LUAD (**Figure 11B**).

Additionally, we also employed the cox proportional hazard model examine the prognostic value of SNX20 expression and tumor infiltration immune cells in LUAD. The result suggested that B cells (p = 0) and $CD4^+$ T cells (p = 0.014) were significantly associated with clinical prognosis in LUAD (**Table 1**).

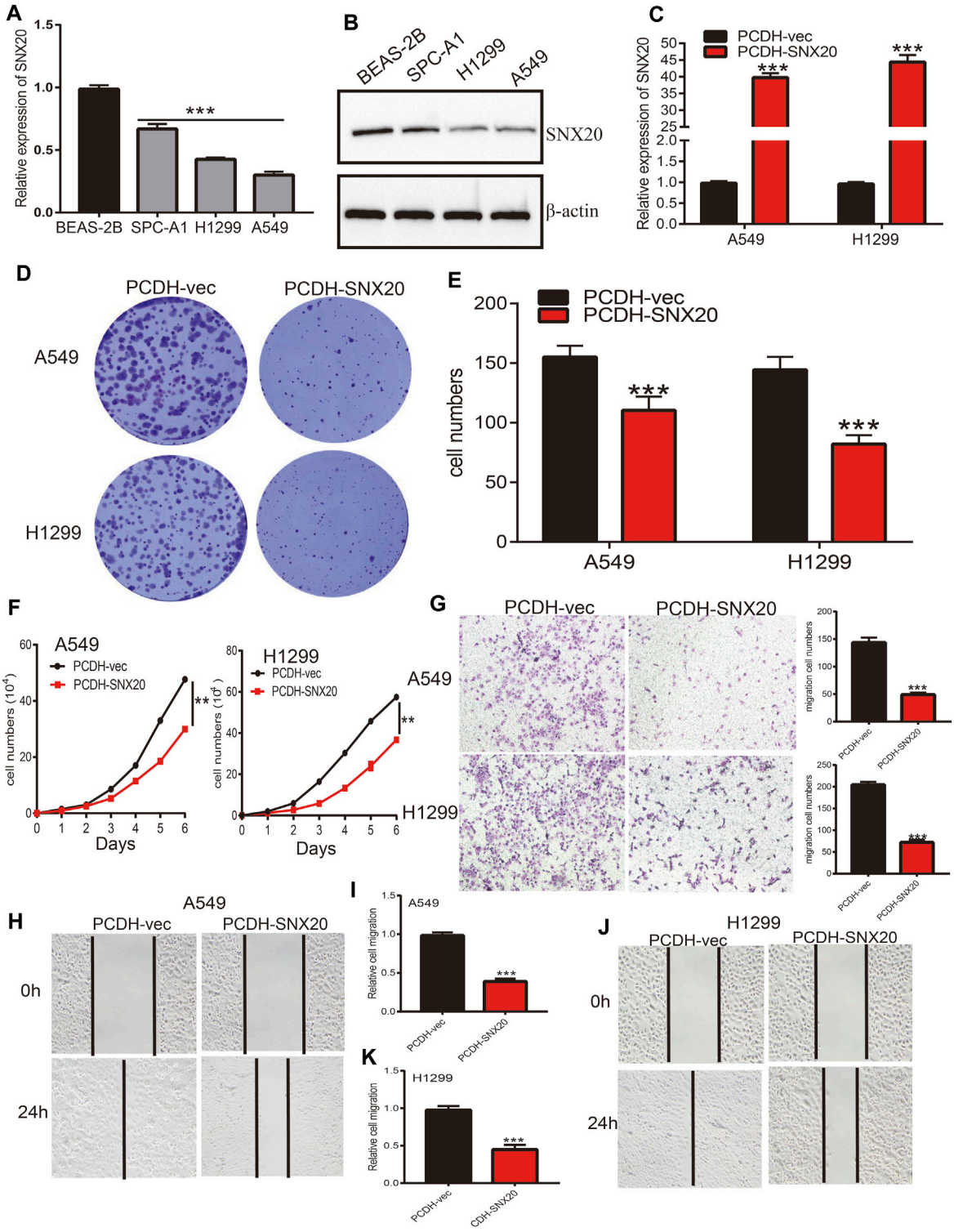


FIGURE 13 | Over-expression of SNX20 on cell proliferation and migration ability of LUAD Cell. **(A)** The mRNA of SNX20 in NSCLC cell lines examined by qPCR assay. **(B)** The protein of SNX20 in NSCLC cell lines examined by Western blot assay. **(C)** Establishment of SNX20 over-expression in NSCLC cells lines and verified by PCR assay. **(D–E)** The colony formation experiment was employed detect over-expression of SNX20 on the growth of NSCLC cells. **(F)** The growth curve experiment was employed detect over-expression of SNX20 on the growth of NSCLC cells. **(G)** The transwell experiment was employed detect over-expression of SNX20 on the migration of NSCLC cells. **(H–K)** The wound healing experiment was employed detect over-expression of SNX20 on the migration of NSCLC cells. **(I)** is the quantification data for **(H)**, **(K)** is the quantification data for **(J)**.

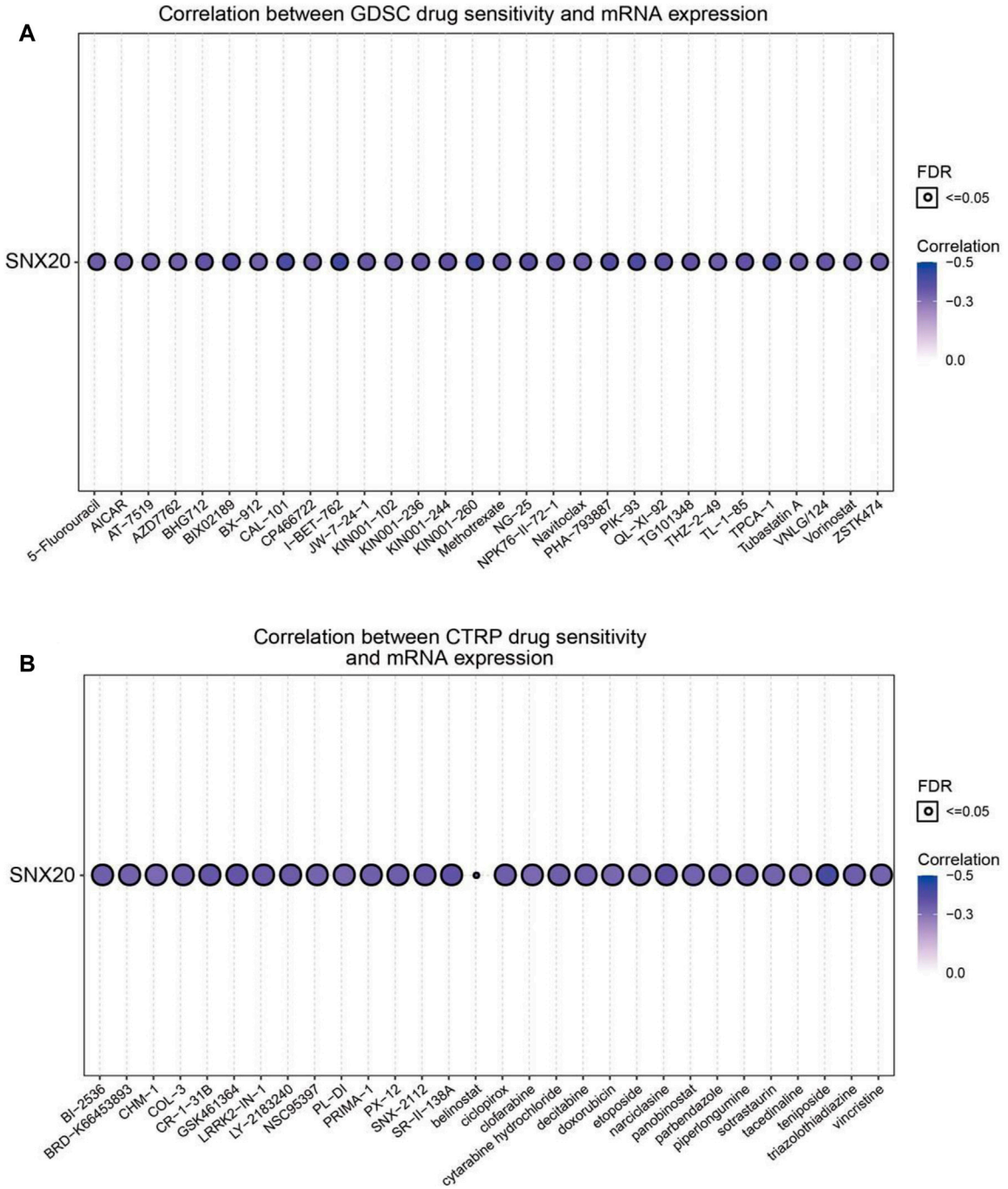


FIGURE 14 | Analysis the correlation between the SNX20 expression and Drug sensitivity in diverse human cancer. **(A)** The correlation between the SNX20 expression and Drug sensitivity in diverse human cancer analysis by employed GDSC database. **(B)** The correlation between the SNX20 expression and Drug sensitivity in diverse human cancer analysis by employed CTRP database.

Undoubtedly, studies has been demonstrated that immune checkpoints play an crucial roles in the development of cancer immunotherapy (Haanen and Robert, 2015). We employed the TIMER database to exploration the correlation between the SNX20 expression and immune check point related gene. The results suggested that SNX20 was markedly positive with the expression of immune check point related gene, including the CTLA4 (r = 0.685, p = 1.03e-72), CD274 (r = 0.646, p = 4.81e-62), HAVCR2 (r = 0.804, p = 8.43e-118), LAG3 (r = 0.519, p = 6.62e-37), PDCD1 (r = 0.668, p = 6.25e-68), PDCD1LG2 (r = 0.754, p = 1.19e-95), SIGLEC15 (r = 0.371, p = 3.18e-18) and TIGIT (r = 0.758, p = 2.38e-97) (**Figure 11C**). Finally, we also

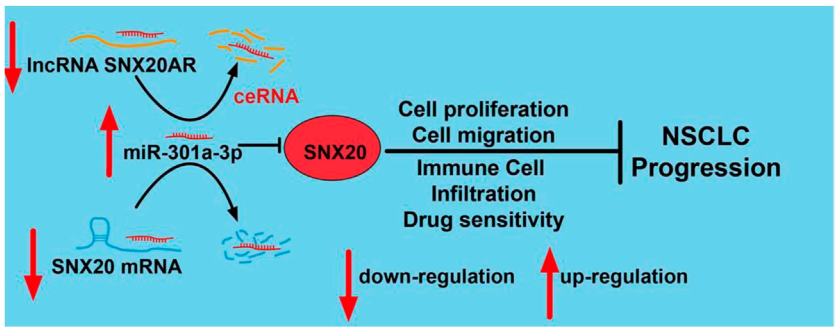


FIGURE 15 | Schematic diagram of the molecular mechanism of SNX20AR/miRNA-301a-3p/SNX20 Axis Associated With Cell Proliferation and Immune Infiltration In Lung Adenocarcinoma.

employed TISIDB database analysis the relationship between SNX20 level and 28 tumors immune infiltrating cell subtypes and the immune regulator (**Table 2**). The results suggested that SNX20 was positively with the 28 tumor immune infiltrating cell, immune regulator and MHC molecular (**Supplementary Figure S2**). These findings indicate that SNX20 plays significant role in cancer immune regulation of LUAD.

Analysis the Correlation Between the SNX20 Expressions and Immune Cell Type Markers

We estimated the correlation between SNX20 expression and different immune cell gene marker in LUAD by employed the TIMER database. Our results showed that the SNX20 expression was strongly correlation with the different immune markers. For instances, SNX20 expression was strongly correlation with $CD8^+$ T markers, CD8A (r = 0.607), CD8B (r = 0.495). The correlation with between the SNX20 expression and immune cells markers as is show in the **Table 3**. These findings indicated that SNX20 was participate in regulate the tumor immune infiltration in lung adenocarcinoma.

Analysis the Prognosis of SNX20 Based on the Different Immune Cells

Due to the SNX20 expression can affect the immune infiltration of immune cells. We further explored the expression of SNX20 and the different immune cells infiltration whether influence the prognosis of lung cancer patients. The analysis results suggested that patients with the high expression of SNX20 and enriched the B cells, $CD4^+$ cells, $CD8^+$ cells, Eosinophils cells, Macrophages cells, Mesenchymal stem cells, Type 1 helper cells and Type 2 helper cells will display a better prognosis. While, The highly expression SNX20 and decreased the B cells, $CD4^+$ cells, $CD8^+$ cells, Eosinophils cells, Macrophages cells, Mesenchymal stem cells, Type 1 helper cells and Type 2 helper cells will display a poor prognosis (**Figures 12A–D**). Together, these results suggested that SNX20 expression and different immune cells infiltration would be affect the prognosis of LUAD.

Over-Expression of SNX20 Suppress Malignant Phenotype of LUAD

In order to explored the function of SNX20 in LUAD progression. We first examine the SNX20 mRNA expression in NSCLC cells lines. The result show that the mRNA and protein levels of SNX20 were decreased in NSCLC cells, especially in A549 and H1299 cells (**Figures 13A,B**). Considering the SNX20 was low expression in NSCLC cells, we construction stable overexpression SNX20 cells and using the qRT-PCR assay detection the over-expression efficiency (**Figure 13C**). Furthermore, we perform the gain of function to examine elevated the SNX20 expression whether affect the proliferation and migration ability of NSCLC cells. The growth curve and clone information experimental result shown that elevated SNX20 was significantly suppressed the cell growth ability of NSCLC cells (**Figures 13D–F**). Similarly, decreased the migration ability of NSCLC cells was observed in over-expression of SNX20 NSCLC cells (**Figures 13G–K**). The above findings indicated that SNX20 plays tumor suppressor role in the LUAD progression.

Analysis the Correlation Between the SNX20 Expression and Drug Sensitivity

Above results suggested that SNX20 may plays oncogene roles in the cancer progression, so we next explored the correlation between SNX20 expression and different drug sensitivity in different cancer cell lines from the GDSC and CTRP database. The result indicated that SNX20 expression was negatively correlated with drug sensitivity of I-BET-762, KIN001-260, CAL-101, PIK-93, PHA-793887, BIX02189, TPCA-1, NG-25, QL-XI-92, KIN001-244, TG101348, TL-1-85, BHG712 and NPK76-II-72-1in GDSC database ($r < -0.36$, **Figure 14A**). In CTRP database, we observed SNX20 expression was negatively correlated with the drug sensitivity of teniposide, BRD-K30748066, SR-II-138A, Narciclasine, GSK461364, CR-1-31B,

LY-2183240, LRRK2-IN-1, SNX-2112, piperlongumine, ciclopirox, PX-12 and triazolothiadiazine ($r < -0.35$, **Figure 14B**). In summary, these results demonstrate that SNX20 was significantly related to diverse drug sensitivity in the different cancer cell lines.

DISCUSSION

Studies have shown that SNX20 plays an crucial roles in endosome-associated scaffolds (Clairfeuille et al., 2015). However, the functions and molecular of SNX20 in the progression of LUAD was remain puzzled. In this study, we analysis the SNX20 expression in human cancers, the results shown that SNX20 was elevated in the nine cancers. However, SNX20 was low expression in LUAD and LUSC, The GEO cohort also find that SNX20 was low expressed in LUAD, using the TCGA LUAD and GEO dataset analysis found that elevated the SNX20 expression display a better prognosis, SNX20 expression was associated with the tumor stage and lymph node metastasis (**Figure 4**). Fan et al. (2020) found that high SNX20 expression has a better overall survival than patients with low expression SNX20 group. However, no study focus on the functions analysis of the SNX20 co-expression genes. In our study, we employed the LinkedOmics tools to examine the SNX20 co-expression in LUAD. Next, we perform the GO and KEGG enriched analysis the SNX20 co-expression genes, the GO term mainly involve in the T cell activation, adaptive immune response, leukocyte proliferation, regulation of defense response to virus by virus, myeloid dendritic cell activation, interleukin-4 production, leukocyte activation involved in inflammatory response, lymphocyte activation involved in immune response, response to chemokine, immune response regulating signaling pathway and regulation of leukocyte activation (**Figure 7C**). The KEGG pathway mainly involve in the intestinal immune network for IgA production, primary immunodeficiency, hematopoietic cell lineage, T cell receptor signaling pathway, cell adhesion molecules, Th1 and Th2 cell differentiation, Th17 cell differentiation, Natural killer cell medicated cytotoxicity and Fc epsilon RI signaling pathway (**Figure 7D**). Finally, our GSEA results shown that SNX20 expression was participate in the adaptive immune response, positive regulation of immune effector process, CD4 positive β T cell activation and B cell differentiation (**Figure 8A**). Similarly, The KEGG enriched results shown that SNX20 were enriched mainly involve in chemokine signaling pathway, JAK STAT signaling pathway, T cell receptor signaling pathway and Toll like receptor signaling pathway (**Figure 8B**). These results strongly indicated that SNX20 was mainly involved in regulation the immune response of lung adenocarcinoma. With regard to the function of SNX20 in LUAD, analysis the protein-protein interaction network shown that SNX20 was significantly correlated with the NCF1, KLHL12, ENOSF1 and HS1BP3. These proteins all play crucial roles in the regulation of immune microenvironment. It has been shown that depletion of NCF1 able to inhibits the cell growth of melanoma (Kelkka et al., 2013). As a one of E3 ligase, KLHL12 play crucial role in the regulation of wnt and Planar Cell Polarity Signaling (Shami Shah et al., 2019). Søreng (2017) found that HS1BP3 was negative regulate the autophagy process *via* inhibits of PLD1 activity and localization. Yang et al. (2015) suggested that ENOSF1 could be as potential serum biomarkers for gastric cancer. The above results indicated that SNX20 may not only involve in cell proliferation in LUAD, but also have important functions in the regulation of immune microenvironment.

To explore the correlation between SNX20 and immune infiltration in LUAD, we employed the TIMER database to examine the immune infiltration related analysis, the results shown that SNX20 CNV was significantly affect the immune infiltration levels of the immune cells, including the B cell, $CD8^+$T cells, $CD4^+$T cells, Macrophage, Neutrophil and Dendritic cell (**Figure 11A**). In addition, we also find that the expression of SNX20 were negatively correlated with tumor purity and positively associated with the immune infiltration of B cells ($r = 0.579$, $p = 1.39e\text{-}44$), $CD8^+$ T cell ($r = 0.451$, $p = 9.14e\text{-}26$), $CD4^+$ T cell ($r = 0.685$, $p = 2.66e\text{-}28$), Macrophage ($r = 0.557$, $p = 6.29e\text{-}41$), Neutrophil ($r = 0.801$, $p = 2.07e\text{-}110$) and Dendritic cell ($r = 0.328$, $p = 5.46E\text{-}10$) in LUAD (**Figure 11B**). Additionally, our study suggested that SNX20 was markedly positive with the expression of immune check point related gene, including the CTLA4 ($r = 0.685$, $p = 1.03e\text{-}72$), CD274 ($r = 0.646$, $p = 4.81e\text{-}62$), HAVCR2 ($r = 0.804$, $p = 8.43e\text{-}118$), LAG3 ($r = 0.519$, $p = 6.62e\text{-}37$), PDCD1 ($r = 0.668$, $p = 6.25e\text{-}68$), PDCD1LG2 ($r = 0.754$, $p = 1.19e\text{-}95$), SIGLEC15 ($r = 0.371$, $p = 3.18e\text{-}18$) and TIGIT ($r = 0.758$, $p = 2.38e\text{-}97$) (**Figure 11C**). These findings indicate that SNX20 plays significant role in cancer immune regulation of LUAD.

Perform the TIMER tools analysis shown that SNX20 expression was markedly correlation with the varies immune cells gene markers, including the B cells (CD19), $CD8^+$ T Cell (CD8A, CD8B), dendritic cell (ITGAX, NRP1, CD1C, HLA-DAP1, HLA-DRA, HLA-DQB1, HLA-DPB1), M1 Macrophage cells (PTGS2, IRF5, NOS2) M2 Macrophage cells (MS4A4A, CD163), Monocyte cells (CSF1R, CD86, KIR2DS4, KIR3DL3, KIR3DL2, KIR3DL1), Natural killer cells (KIR2DL4,KIR2DL3, KIR2DL1), Neutrophils cells (CCR7, ITGAM, CEACAM8), T cells (CD3D, CD3E, CD2), T cell exhaustion cells (CTLA4, LAG3, HAVCR2, GZMB, PDCD1), TAM cells (CCL2, IL10, CD68), Tfh cells (BCL6, IL21), Th1 cells (TBX21, STAT4, STAT1, IFNG), Th2 cells (GATA3, STAT6, STAT5A), Th17 cells (STAT3, IL17A), Treg cells (FOXP3, CCR8, STAT5B, TGFB1). These studies indicated that SNX20 plays crucial role in the regulation immune infiltrating cells of LUAD (**Table 2**). We also that SNX20 expression and different immune cells infiltration would be affect the prognosis of LUAD.

We also determined the upstream regulation mechanism of SNX20 in the progression of LUAD. As the important content of epigenetic modification, non-coding RNAs was shown plays significant roles in the regulation gene expression. LncRNA was reported that as an miRNA sponge to competitively binding miRNA and result in change the expression of downstream target genes. For example, the PDL1 related lncRNA was shown that through elevated the c-Myc expression and participate in the progression of lung

adenocarcinoma (Qu et al., 2021). Chen et al. (2020) found that LINC00173.v1 *via* inhibits the miR-511-5p to upregulation of VEGFA expression, result in boost the vascular endothelial cells growth and migration in lung squamous cell carcinoma. Yan et.al observed that lncRNA JPX *via* restrain the expression of miR-33a-5p and subsequently up-regulated the Twist1 expression, result in activating Wnt/β-catenin signaling to involve in the progression of lung cancer (Pan et al., 2020). In this study, we first predicted and comprehensive analysis the upstream miRNAs of SNX20 in LUAD. Combine with The 4 databases predicted 4 potential upstream miRNAs, while only miR-301a-3p was highly expression in LUAD and it's high expression was negative associated with SNX20 in LUAD. Next, we further examine the upstream lncRNAs of miR-301a-3p, by perform the expression and correlation analysis, only SNX20 meet the established conditions. The SNX20AR was significantly decreased in LUAD and it's low expression was associated with the poor prognosis. Moreover, we find SNX20AR expression was not only markedly positive with the SNX20, but also negative with the miRNA-301a-3p expression in LUAD. Above all, although the SNX20 related ceRNA network was analysed *via* bioinformatics analysis, the accurate and credible assays are needed to verify the proposed hypothesis. We also found that over-expression of SNX20 was significantly inhibits the cell proliferation and migration of NSCLC cells.

This work not only deep the understanding of the potential mechanism of lung cancer progression, but also raised our awareness about the tumor-immune microenvironment of lung cancer. Moreover, we revealed that SNX20 was down-regulation in LUAD. The upstream complex molecular regulation mechanism of SNX20 was revealed by us, that is SNX20AR/miRNA-301a-3p (**Figure 15**). Finally, our work demonstrates that SNX20 was further implicated in alteration of the tumor microenvironment. Over-expression of SNX20 inhibits cell proliferation and cell migration in NSCLC cells. These findings suggested that SNX20 may be plays tumor suppressor roles in the progression of lung cancer and represent an effective target of immunotherapy in the 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 author.

AUTHOR CONTRIBUTIONS

LD, XJ, and YY designed this work, performed related assay and analyzed data. LT, JW, QL, and XZ contributed to study materials, LD supervised and wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by National Nature Science Foundation of China (82160508), Yunnan Applied Basic Research Projects (YNWRMY-2019-067), Yunnan Province Specialized Training Grant for High-Level Healthcare Professionals (D-201614) and Yunnan Province Applied Basic Research foundation (2019FE001) to LD.

ACKNOWLEDGMENTS

The authors would like to thank support from department of Department of Thoracic Surgery II, The Third Affiliated Hospital of Kunming Medical University (Yunnan Tumor Hospital), Kunming, China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.744363/full#supplementary-material

REFERENCES

Agarwal, V.,(2015). Predicting Effective microRNA Target Sites in Mammalian mRNAs, *Elife* 4, e05005. doi:10.7554/elife.05005

Brant, S. R., Okou, D. T., Simpson, C. L., Cutler, D. J., Haritunians, T., Bradfield, J. P., et al. (2017). Genome-Wide Association Study Identifies African-specific Susceptibility Loci in African Americans with Inflammatory Bowel Disease. *Gastroenterology* 152 (1), 206–e2. doi:10.1053/j.gastro.2016.09.032

Cao, Z., Pan, X., Yang, Y., Huang, Y., and Shen, H.-B. (2018). The lncLocator: a Subcellular Localization Predictor for Long Non-coding RNAs Based on a Stacked Ensemble Classifier. *Bioinformatics* 34 (13), 2185–2194. doi:10.1093/bioinformatics/bty085

Carlton, J., Bujny, M., Peter, B. J., Oorschot, V. M. J., Rutherford, A., Mellor, H., et al. (2004). Sorting Nexin-1 Mediates Tubular Endosome-To-TGN Transport through Coincidence Sensing of High- Curvature Membranes and 3-phosphoinositides. *Curr. Biol.* 14 (20), 1791–1800. doi:10.1016/j.cub.2004.09.077

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

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, J., Liu, A., Wang, Z., Wang, B., Chai, X., Lu, W., et al. (2020). LINC00173.v1 Promotes Angiogenesis and Progression of Lung Squamous Cell Carcinoma by Sponging miR-511-5p to Regulate VEGFA Expression. *Mol. Cancer* 19 (1), 98. doi:10.1186/s12943-020-01217-2

Clairfeuille, T., Norwood, S. J., Qi, X., Teasdale, R. D., and Collins, B. M. (2015). Structure and Membrane Binding Properties of the Endosomal Tetratricopeptide Repeat (TPR) Domain-Containing Sorting Nexins SNX20 and SNX21. *J. Biol. Chem.* 290 (23), 14504–14517. doi:10.1074/jbc.m115.650598

Fan, L., Li, L., Huang, C., Huang, S., Deng, J., and Xiong, J. (2020). Increased SNX20 and PD-L1 Levels Can Predict the Clinical Response to PD-1 Inhibitors in Lung Adenocarcinoma. *Ott* 13, 10075–10085. doi:10.2147/ott.s262909

Gong, J., Liu, W., Zhang, J., Miao, X., and Guo, A. Y. (2015). lncRNASNP: a Database of SNPs in lncRNAs and Their Potential Functions in Human and Mouse. *Nucleic Acids Res.* 43 (issue), D181–D186. doi:10.1093/nar/gku1000

Griffiths-Jones, S., (2006). miRBase: microRNA Sequences, Targets and Gene Nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi:10.1093/nar/gkj112

Haanen, J. B. A. G., and Robert, C. (2015). Immune Checkpoint Inhibitors. *Prog. Tumor Res.* 42, 55–66. doi:10.1159/000437178

Jiang, L.-P., Fan, S.-Q., Xiong, Q.-X., Zhou, Y.-C., Yang, Z.-Z., Li, G.-F., et al. (2018). GRK5 Functions as an Oncogenic Factor in Non-small-cell Lung Cancer. *Cell Death Dis* 9 (3), 295. doi:10.1038/s41419-018-0299-1

Kelkka, T., Pizzolla, A., Laurila, J. P., Friman, T., Gustafsson, R., Källberg, E., et al. (2013). Mice Lacking NCF1 Exhibit Reduced Growth of Implanted Melanoma and Carcinoma Tumors. *PLoS One* 8 (12), e84148. doi:10.1371/journal.pone.0084148

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). CPC: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine. *Nucleic Acids Res.* 35, W345–W349. doi:10.1093/nar/gkm391

Kurten, R. C., Cadena, D. L., and Gill, G. N. (1996). Enhanced Degradation of EGF Receptors by a Sorting Nexin, SNX1. *Science* 272 (5264), 1008–1010. doi:10.1126/science.272.5264.1008

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 (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

Lv, J., Liu, H., Su, J., Wu, X., Liu, H., Li, B., et al. (2012). DiseaseMeth: a Human Disease Methylation Database. *Nucleic Acids Res.* 40 (issue), D1030–D1035. doi:10.1093/nar/gkr1169

Martin, P., and Leighl, N. B. (2017). Review of the Use of Pretest Probability for Molecular Testing in Non-small Cell Lung Cancer and Overview of New Mutations that May Affect Clinical Practice. *Ther. Adv. Med. Oncol.* 9 (6), 405–414. doi:10.1177/1758834017704329

Mering, C. v., (2003). STRING: a Database of Predicted Functional Associations between Proteins. *Nucleic Acids Res.* 31 (1), 258–261. doi:10.1093/nar/gkg034

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

Nagy, Á., Munkácsy, G., and Győrffy, B. (2021). Pancancer Survival Analysis of Cancer Hallmark Genes. *Sci. Rep.* 11 (1), 6047. doi:10.1038/s41598-021-84787-5

Pan, J., Fang, S., Tian, H., Zhou, C., Zhao, X., Tian, H., et al. (2020). lncRNA JPX/miR-33a-5p/Twist1 axis Regulates Tumorigenesis and Metastasis of Lung Cancer by Activating Wnt/β-Catenin Signaling. *Mol. Cancer* 19 (1), 9. doi:10.1186/s12943-020-1133-9

Qu, S., Jiao, Z., Lu, G., Yao, B., Wang, T., Rong, W., et al. (2021). PD-L1 lncRNA Splice Isoform Promotes Lung Adenocarcinoma Progression *via* Enhancing C-Myc Activity. *Genome Biol.* 22 (1), 104. doi:10.1186/s13059-021-02331-0

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

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

Schaff, U. Y., Shih, H. H., Lorenz, M., Sako, D., Kriz, R., Milarski, K., et al. (2008). SLIC-1/sorting Nexin 20: a Novel Sorting Nexin that Directs Subcellular Distribution of PSGL-1. *Eur. J. Immunol.* 38 (2), 550–564. doi:10.1002/eji.200737777

Shami Shah, A., Batrouni, A. G., Kim, D., Punyala, A., Cao, W., Han, C., et al. (2019). PLEKHA4/kramer Attenuates Dishevelled Ubiquitination to Modulate Wnt and Planar Cell Polarity Signaling. *Cel Rep.* 27 (7), 2157–2170. doi:10.1016/j.celrep.2019.04.060

Søreng, K., (2017). HS1BP3 Inhibits Autophagy by Regulation of PLD1. *Autophagy* 13 (5), 985–986. doi:10.1080/15548627.2017.1291483

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, 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

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The Multilayered Complexity of ceRNA Crosstalk and Competition. *Nature* 505 (7483), 344–352. doi:10.1038/nature12986

van Weering, J. R. T., Sessions, R. B., Traer, C. J., Kloer, D. P., Bhatia, V. K., Stamou, D., et al. (2012). Molecular Basis for SNX-BAR-Mediated Assembly of Distinct Endosomal Sorting Tubules. *Embo J* 31 (23), 4466–4480. doi:10.1038/emboj.2012.283

Vasaikar, S. V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: Analyzing Multi-Omics Data within and across 32 Cancer Types. *Nucleic Acids Res.* 46 (D1), D956–d963. doi:10.1093/nar/gkx1090

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

Wassmer, T., Attar, N., Harterink, M., van Weering, J. R. T., Traer, C. J., Oakley, J., et al. (2009). The Retromer Coat Complex Coordinates Endosomal Sorting and Dynein-Mediated Transport, with Carrier Recognition by the Trans-golgi Network. *Dev. Cel* 17 (1), 110–122. doi:10.1016/j.devcel.2009.04.016

Wong, N., and Wang, X. (2015). miRDB: an Online Resource for microRNA Target Prediction and Functional Annotations. *Nucleic Acids Res.* 43 (issue), D146–D152. doi:10.1093/nar/gku1104

Yang, J., Xiong, X., Wang, X., Guo, B., He, K., and Huang, C. (2015). Identification of Peptide Regions of SERPINA1 and ENOSF1 and Their Protein Expression as Potential Serum Biomarkers for Gastric Cancer. *Tumor Biol.* 36 (7), 5109–5118. doi:10.1007/s13277-015-3163-2

Zeng, W., Yuan, W., Wang, Y., Jiao, W., Zhu, Y., Huang, C., et al. (2002). Expression of a Novel Member of Sorting Nexin Gene Family, SNX-L, in Human Liver Development. *Biochem. Biophysical Res. Commun.* 299 (4), 542–548. doi:10.1016/s0006-291x(02)02695-5

Zheng, B., (2001). RGS-PX1, a GAP for Galpha S and Sorting Nexin in Vesicular Trafficking. *Science* 294 (5548), 1939–1942. doi:10.1126/science.1064757

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 Yuan, Jiang, Tang, Wang, Liu, Zou and Duan. 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.

GLOSSARY

ACC Adrenocortical carcinoma
BLCA Bladder Urothelial Carcinoma
BRCA Breast invasive carcinoma
CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL Cholangiocarcinoma
COAD Colon adenocarcinoma
DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA Esophageal carcinoma
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
LGG Brain Lower Grade Glioma
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
STAD Stomach adenocarcinoma
TGCT Testicular Germ Cell Tumors
THCA Thyroid carcinoma
THYM Thymoma
UCEC Uterine Corpus Endometrial Carcinoma
UCS Uterine Carcinosarcoma
UVM Uveal Melanoma

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fmolb.2021.727481

Characterization of METTL7B to Evaluate TME and Predict Prognosis by Integrative Analysis of Multi-Omics Data in Glioma

Xiaochuan Chen[1†], Chao Li[2†], Ying Li[1], Shihong Wu[1], Wei Liu[1], Ting Lin[1], Miaomiao Li[1], Youliang Weng[1], Wanzun Lin[3] and Sufang Qiu[1,4*]*

[1]Department of Radiation Oncology, Fujian Medical University Cancer Hospital, Fujian Cancer Hospital, Fuzhou, China, [2]Department of Oncology, Sanming Second Hospital, Sanming, China, [3]Department of Radiation Oncology, Shanghai Proton and Heavy Ion Center, Fudan University Cancer Hospital, Shanghai, China, [4]Fujian Provincial Key Laboratory of Translational Cancer Medicine, Fuzhou, China

OPEN ACCESS

Edited by:
Florence Le Calvez-Kelm, International Agency For Research On Cancer (IARC), France

Reviewed by:
Shilpita Karmakar, Jackson Laboratory, United States
Anupam Nath Jha, Tezpur University, India

***Correspondence:**
Sufang Qiu
sufangqiu@fjmu.edu.cn
Wanzun Lin
linwanzun@qq.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 18 June 2021
Accepted: 24 August 2021
Published: 17 September 2021

Citation:
Chen X, Li C, Li Y, Wu S, Liu W, Lin T, Li M, Weng Y, Lin W and Qiu S (2021) Characterization of METTL7B to Evaluate TME and Predict Prognosis by Integrative Analysis of Multi-Omics Data in Glioma. Front. Mol. Biosci. 8:727481. doi: 10.3389/fmolb.2021.727481

Glioma is the most common and aggressive type of primary brain malignant tumor with limited treatment approaches. Methyltransferase-like 7B (METTL7B) is associated with the pathogenesis of several diseases but is rarely studied in glioma. In this study, 1,493 glioma samples (data from our cohort, TCGA, and CGGA) expressing METTL7B were used to explore its prognostic value and mechanism in the immune microenvironment. Results showed that high expression of METTL7B is associated with poor prognosis and abundant immunosuppressive cells. Further, functional enrichment showed that METTL7B is involved in the negative regulation of immunity and carcinogenic signaling pathways. Moreover, a METTL7B-related prognostic signature constructed based on multi-omics showed a good prediction of the overall survival (OS) time of glioma patients. In conclusion, METTL7B is a potential prognostic biomarker. In addition, the prognostic prediction model constructed in this study can be used in clinical setups for the development of novel effective therapeutic strategies for glioma patients and improving overall survival.

Keywords: glioma, prognosis, immunosuppressive microenvironment, multi-omics, METTL7B

INTRODUCTION

Glioma is the most common and aggressive type of primary brain malignant tumor (Jiang et al., 2016; Lapointe et al., 2018). However, current therapy approaches, for glioma including surgery, chemotherapy, and radiotherapy, are not fully effective. Therefore, treatment of glioma is a challenge resulting in high mortality rates. WHO classification system defines diffuse low-grade glioma (LGG) as the WHO grade II/III based on histological type (Wesseling and Capper 2018).

Abbreviations: ALD1, associated with lipid droplets; CGGA, Chinese glioma genome atlas; EMT, epithelial-mesenchymal transition; FDR, false discovery rate; GBM, glioblastoma; GEPIA, gene expression profiling interactive analysis; GSEA, gene set enrichment analysis; GO, gene ontology; H-score, histochemistry score; IDH, isocitrate dehydrogenase; IHC, immunohistochemistry; LGG, low-grade glioma; METTL7B, methyltransferase-like 7B; NES, normalized enrichment score; OS, overall survival; ROC, receiver operating characteristic; TAM, tumor-associated macrophage; TMB, tumor mutation burden; TME, tumor immune microenvironment; WES, whole exome sequencing; TCGA, The Cancer Genome Atlas.

Most LGGs show recurrence and gradually transform into higher-grade gliomas (GBM) leading to death (Louis et al., 2007; Liu et al., 2019). The median survival time of GBM patients is approximately 14.6 months, and the 5-year survival rate is less than 10% (Stupp et al., 2005; Stupp et al., 2009). Therefore, novel effective biomarkers for the prediction of the prognosis of patients with glioma should be explored to improve clinical outcomes.

Human methyltransferase-like (MettL) proteins are involved in methylation reactions. Although members of this protein family play key biological functions, for instance, METTL2B, METTL3, METTL8, and METTL16 are RNA methyltransferases and are implicated in tumorigenesis, the roles of METTL proteins are not clear (Xu K et al., 2017; Xu L et al., 2017; Deng et al., 2018; Ignatova et al., 2019). Methyltransferase-like 7B (METTL7B), also known as associated with LD protein 1 (ALD1), is localized on chromosome 12. Studies report that METTL7B is implicated in diseases, such as infection Abdel-Hameed et al. (2014), non-alcoholic steatohepatitis lipid metabolism Thomas et al. (2013), and several tumors, which include primary thyroid cancer (PTC), lung adenocarcinoma (LUAD), and non-small cell lung cancer (NSCLC). In PTC, METTL7B is upregulated and promotes tumor invasion and malignancy by activating the TGF-β1-induced EMT (Cai et al., 2018). In LUAD, a recent study demonstrates that METTL7B is overexpressed in NSCLC tumor tissues and promotes tumorigenesis by regulating cell cycle progression (Ali et al., 2020). Moreover, METTL7B promotes tumorigenesis by regulating cell cycle progression in non-small cell lung cancer (Liu et al., 2020). However, the oncogenic role and prognostic value of METTL7B in glioma has not been reported previously.

This study aimed to elucidate the efficacy of METTL7B, as a potential diagnostic and prognostic biomarker for glioma. Expression levels of METTL7B mRNA in glioma tissues and normal brain tissues were determined. Further, survival analysis, independent prognostic analysis, ROC curve analysis, and clinical correlation analysis were used to determine the clinical and prognostic value of METTL7B. In addition, gene set enrichment analysis (GSEA) was used for functional and pathway analysis. Immune infiltration correlation analysis was performed to determine the role of METTL7B in the tumor immune microenvironment. Multi-omics data (WES and DNA methylation array) were then analyzed. Finally, a prognosis prediction model was constructed based on METTL7B-related significant alterations. The findings from this study form a basis for the development of an effective predictor for the prognosis of gliomas.

MATERIALS AND METHOD

Data Retrieval

Data were retrieved from the GEPIA database (http://gepia.cancer-pku.cn/index.html) (Tang et al., 2017). Expression data were used to explore differences in expression levels of METTL7B mRNA in glioma tissues and normal brain tissues. RNA expression data and corresponding clinical data of 666 and 693 glioma (LGG + GBM) patients were retrieved from the TCGA database using the UCSC website (https://xenabrowser.net/) (Kent et al.,2002) and from the CGGA website (http://www.cgga.org.cn/), respectively. In addition, somatic mutation data (VarScan2 Variant Aggregation) and DNA methylation data (the Illumina 450K methylation array) of glioma were retrieved from the TCGA database (https://portal.gdc.cancer.gov/).

A total of 134 glioma specimen tissues (G6042) were purchased from Servicebio Company (Wuhan, China) and used to further validate METTL7B expression and its prognostic value in glioma.

Immunohistochemistry (IHC) Analysis

Glioma tissues sections were analyzed through IHC using anti-human METTL7B (Proteintech, Cat #17001-1-AP). HRP-linked secondary antibodies (Abcam, Cat #ab205718, UK) were then used followed by DAB treatment. Images were obtained under a microscope (3DHISTECH, Hungary) at ×20 magnification. Histochemistry score (H-score) was used to detect and quantify the expression level of METTL7B. H-score was calculated as follows: H-score = (percentage of cells of weak intensity × 1) + (percentage of cells of moderate intensity × 2) + (percentage of cells of strong intensity × 3).

Single-Cell Level Analysis

We obtained GBM single-cell sequencing data (GSE131928, GSE139448, GSE84465) from the online database of the Tumor Immune Single-Cell Hub (TISCH) (http://tisch.comp-genomics.org/) (Neftel et al., 2019; Sun et al., 2021), which was used to classify malignant cells, immune cells, and stromal cells by hierarchical clustering. Then, the expression of METTL7B in these cells was evaluated, and the results were illustrated by heatmaps.

Gene Set Enrichment Analysis

GSEA was used to compare expression levels between a priori defined set of genes and high and low METTL7B expression groups in the enrichment of MSigDB Collection (c2.cp.kegg and c5.go.bp. v7.2. symbols.gmt). High and low METTL7B were then used as a phenotype label and gene set permutations were carried out 1,000 times for each analysis. False discovery rate (FDR) and normalized enrichment score (NES) were used to classify gene ontology (GO) and KEGG pathways enriched in differential phenotype.

DNA Methylation Analysis

Limma package in R was used to normalize the gene methylation matrix. The Pearson correlation coefficient was used to determine the association between the gene expression and DNA methylation level of METTL7B. The Kaplan–Meier curves of differential METTL7B DNA promoter CpG sites in glioma patients were constructed using the R survival package.

Somatic Mutation Analysis

WES somatic mutations data of both high METTL7B (n = 327) and low METTL7B groups (n = 329) were used to detect the SNVs, SNPs, and INDELs using VarScan2.39 software. Differentially mutated genes, which were defined with a

p value lower than 0.05, were analyzed using Fisher's exact test. R maftools package was used for visualization of somatic mutations and calculation of TMB score.

Analysis of Immune Infiltration

The abundance of immune cells was determined using CIBERSORT (22 immune cell types) and xCell (64 immune and stromal cell types) algorithms (Newman et al., 2015). Mann–Whitney U test was performed to compare differential immune cell distribution between high and low METTL7B expression groups. Further, the expression of genes which negatively regulate the Cancer-Immunity Cycle Chen and Mellman (2013) was determined in low and high METTL7B expression groups. Immunosuppressive gene signatures were retrieved from the Tracking Tumor Immunophenotype website (http://biocc.hrbmu.edu.cn/TIP/index.jsp) (Xu et al., 2018).

Construction and Evaluation of Prognosis Prediction Model

Survival time and status were determined to assess the prognosis of glioma patients. Cases that included their overall survival were considered. Prognosis prediction models were constructed based on gene expression values of METTL7B, differential immune cells and immunosuppressive gene sets, gene mutation, and methylation probe signals. Cases with all characteristics being investigated were considered for the model with comprehensive characteristics from multi-omics data, and 545 same cases were obtained.

Genes were expressed as 0 (wild) and 1 (mutation) based on somatic mutation. Immune cell fraction level was expressed as 0 or 1 (not = 0). Univariate Cox regression analysis was performed to determine prognostic factors. Significant predictive features were further identified by LASSO-COX analysis. A prognosis signature was constructed based on coefficients from LASSO-COX analysis. SurvivalROC package in R was used to measure predicting ability of the prognosis model using the Kaplan–Meier survival curves and time-dependent receiver operating characteristic (ROC) curves. The R rms package was used to construct a prognostic nomogram based on the model and clinical information of glioma patients. Further, calibration plots for 3 and 5 years were constructed to validate predicted and actual probabilities.

Statistical Analysis

Expression of METTL7B in tumor and normal tissues was estimated using Wilcoxon signed-rank test. OS of participants were compared between high and low METTL7B expression groups through Kaplan–Meier analysis using the R Survival and Survminer package. Univariate Cox analysis was used to determine potential prognostic factors, whereas multivariate Cox analysis was performed to identify METTL7B expression as an independent risk factor for OS in glioma patients. The receiver operating characteristic (ROC) curve was used to evaluate the diagnostic value of METTL7B expression using the R survivalROC package. The area under curve (AUC) represented the diagnostic value. Correction between clinical pathologic features and METTL7B expression was analyzed using Wilcoxon signed-rank test or Kruskal–Wallis test. $p < 0.05$ was considered statistically significant. All data analyses were carried out using R software (version 3.6.0) and AdobeIllustratorCS6.

RESULTS

Clinical Prognostic Value of METTL7B Expression in TCGA and CGGA Database

Gene expression data of 518 LGG and 163 GBM samples from TCGA and 207 normal brain tissues from GTEx portal were dissected using GEPIA. Expression levels of METTL7B mRNA were higher in LGG and GBM tissues, compared with the expression levels in normal tissues (**Figure 1A**). IHC staining data of LGG and GBM were used to evaluate the expression level of METTL7B in glioma tissues. Overall, positive staining for METTL7B was detected in GBM tissue, whereas LGG samples showed negative staining for METTL7B (**Figures 1B, C**). The Kaplan–Meier analysis of the TCGA and CGGA dataset (including GBM and LGG) showed that a high expression level of METTL7B was associated with poor prognosis (**Figure 1D**). This differential expression was validated using our dataset (**Figure 1E**). In addition, the prognostic value of METTL7B for different WHO grades was evaluated. LGG patients with high METTL7B expression levels showed significantly shorter OS time compared with patients with low METTL7B expression level in CGGA cohorts (**Figure 1F**). The High METTL7B expression level of GBM patients was associated with poor OS in TCGA and CGGA cohorts (**Figure 1G**).

Univariate Cox analysis showed that METTL7B expression level (HR = 1.644; 95% CI = 1.536–1.759; $p < 0.001$), grade, age, IDH/codel subtype, and MGMT promoter status were significantly correlated with poor OS (**Figure 2A**). Multivariate Cox analysis showed that high METTL7B expression level (HR = 1.157; 95% CI = 1.048–1.277; $p = 0.004$) in the TCGA database was independently associated with poorer OS. This finding implied that METTL7B was a potential independent prognostic indicator for glioma (**Figure 2B**). These findings were determined using data from the CGGA database (**Figures 2C,D**). Moreover, subgroup analysis in different WHO grade indicated that METTL7B was an independent predictor of poor prognosis (**Supplementary Table S1–S2**). In addition, METTL7B was a significant predictor of 1-year (AUC = 0.812), 3-year (AUC = 0.875), and 5-year survival (AUC = 0.812) using receiver operating characteristic curve analysis of TCGA data (**Figure 2E**). This finding was further validated using data retrieved from the CGGA database (**Figure 2F**). More interestingly, we found that the expression level of METTL7B was higher in malignant cells compared with immune cells and stromal cells in the GBM patients by using the TISCH database (**Figure 2G**).

Relationship Between METTL7B Expression and Clinical-Pathological Features

To determine the role of METTL7B in tumorigenesis and tumor development, we explored relationships between

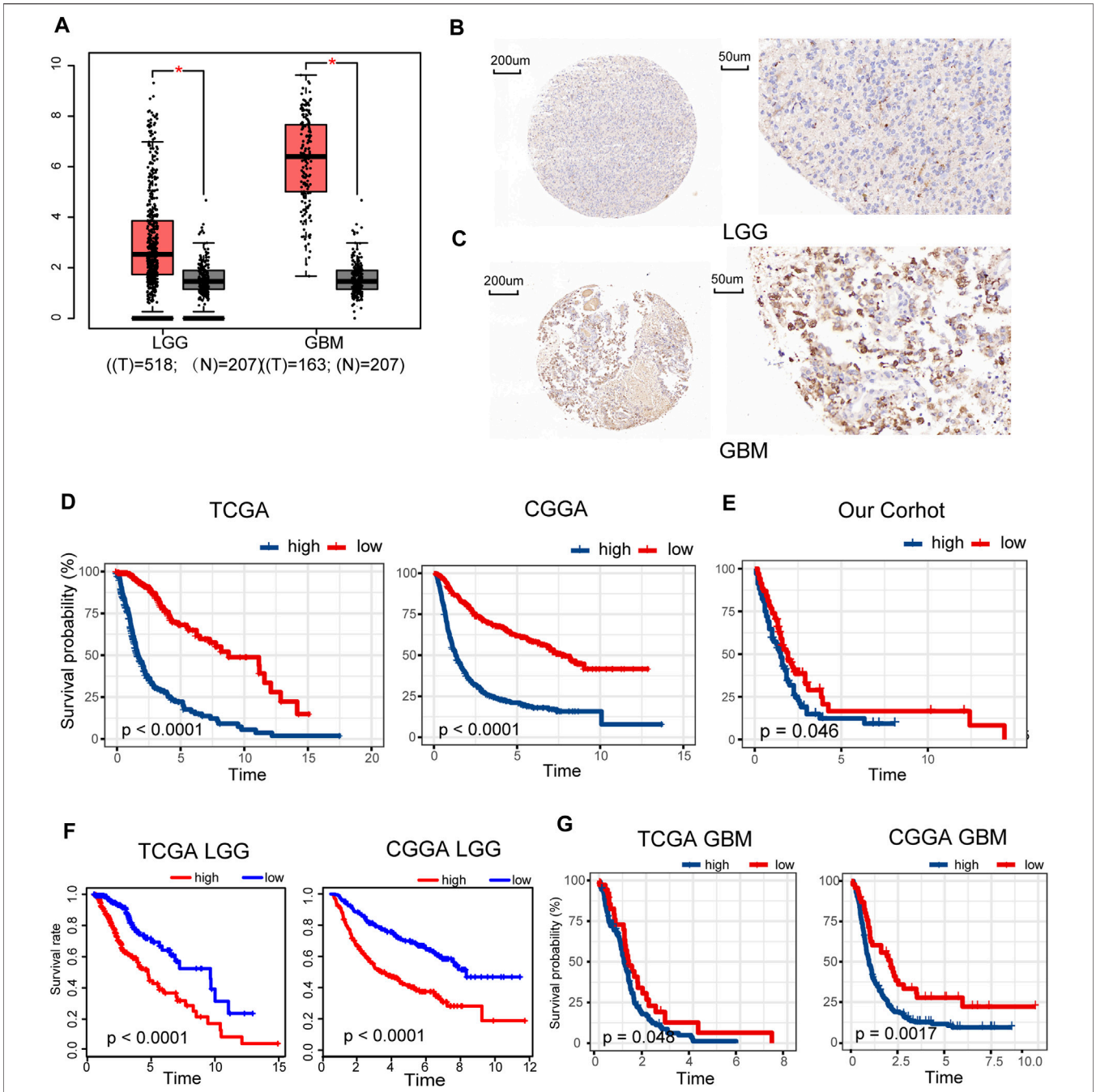


FIGURE 1 | Clinical prognostic value of METTL7B expression in LGG and GBM patients. **(A)** METTL7B expression of normal brain tissues from GTEx data (n = 207) compared with expression in LGG (n = 518) and GBM (n = 163) samples retrieved from TCGA database. **(B–C)** Representative specimens exhibiting METTL7B IHC labeling pattern in LGG and GBM. **(D–E)** Survival analysis of patients with glioma in high METTL7B and low METTL7B groups in the TCGA, CGGA, and our cohort. **(F)** KM survival curve of patients with LGG in high METTL7B and low METTL7B groups in TCGA and CGGA datasets. **(G)** KM survival curve of patients with GBM in high METTL7B and low METTL7B groups in TCGA and CGGA datasets. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

METTL7B and clinic-pathological features of gliomas, including WHO grade, IDH status, IDH/codel subtype. Glioma patients with high WHO grades, IDH-wild-type, showed significantly high expression of METTL7B. These correlations were confirmed using CGGA datasets and our datasets (**Figures 3A,B**). Moreover, subgroup analysis in different WHO grade indicated that the levels of METTL7B expression were higher in the IDHwt subtype compared with the IDHmut-non-codel and IDHmut-codel subtype (**Figure 3C**).

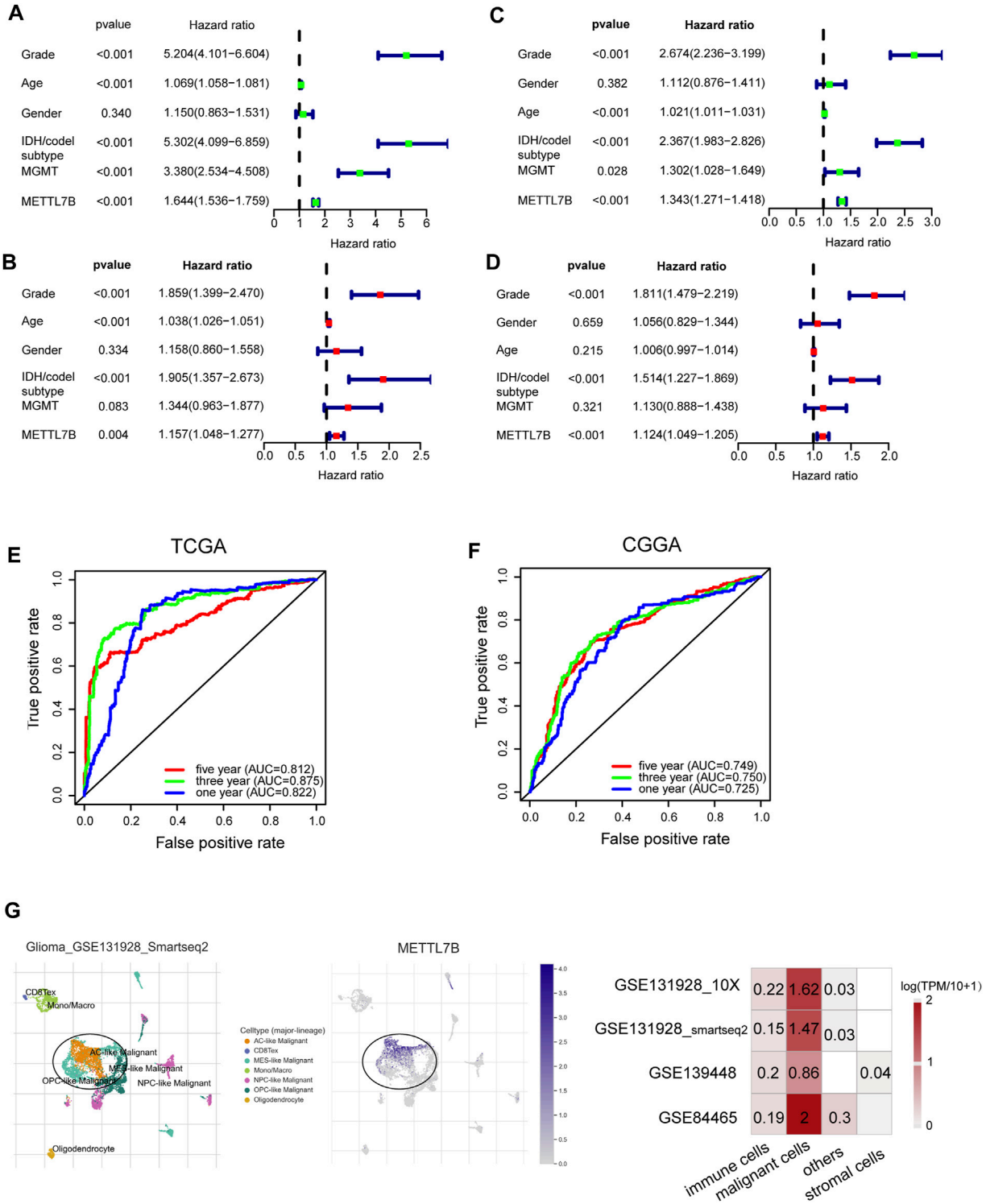


FIGURE 2 | Relationship between METTL7B expression and prognosis of glioma patients. **(A–D)** Univariate and multivariate Cox analyses evaluating the independent prognostic value of METTL7B in terms of OS in glioma patients using TCGA and CGGA datasets. **(E–F)** Receiver operator characteristic curve analysis of METTL7B in TCGA and CGGA datasets. AUC, area under the curve. **(G)** Single-cell level analysis evaluating the expression of METTL7B.

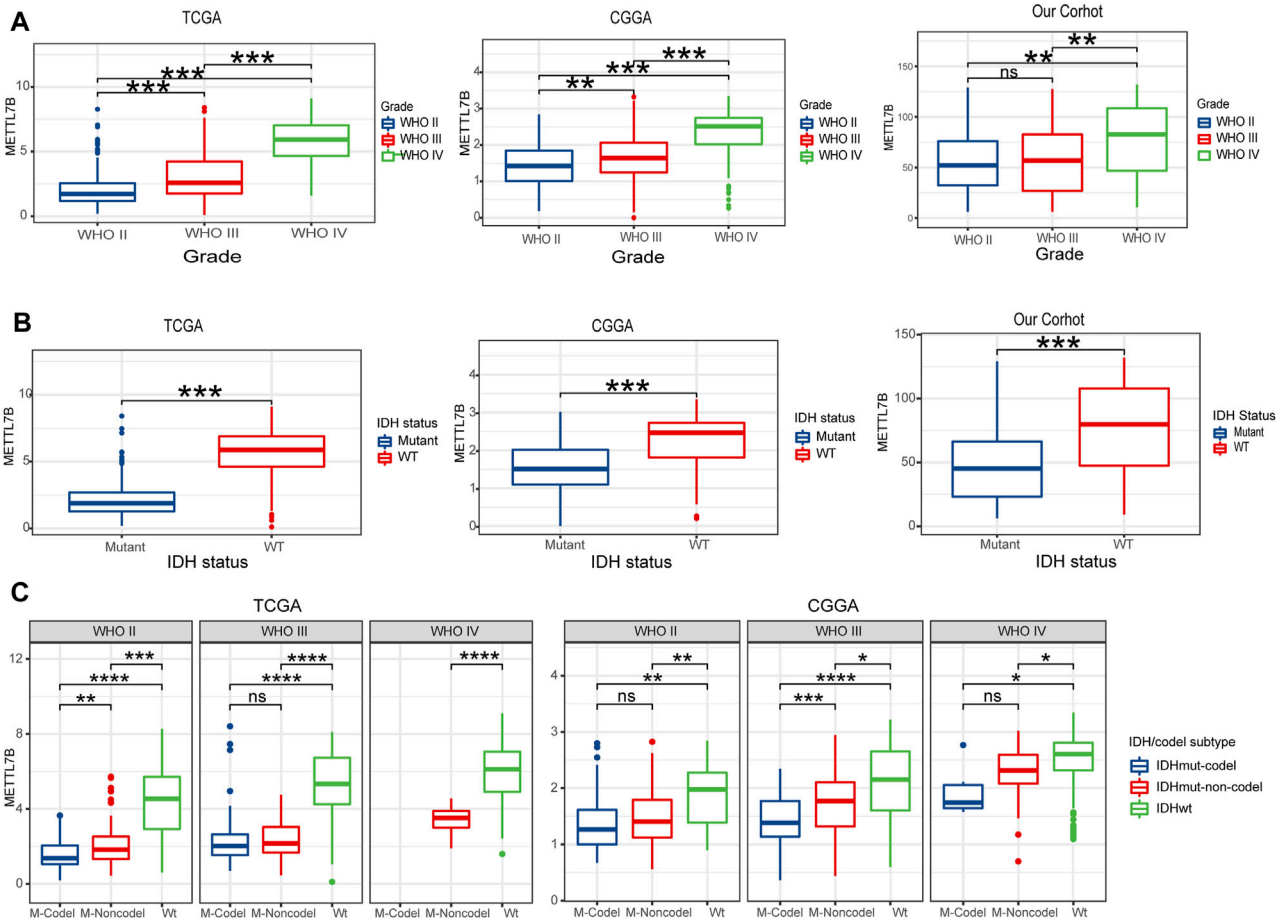


FIGURE 3 | Relationship between METTL7B expression and clinical-pathological features. **(A)** Expression levels of METTL7B in gliomas with different WHO grades. **(B)** Expression levels of METTL7B in gliomas with different IDH status. **(C)** Expression levels of METTL7B with IDH/codel subtype in different WHO grades; *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.

Gene Set Enrichment Analysis of METTL7B Expression Data

GO and signaling pathways in low and high METTL7B expression groups were analyzed using gene set enrichment analysis using MSigDB Collection (c2.cp.kegg and c5.go.bp. v7.2. symbols) (FDR <0.05). Normalized enrichment score (NES) was used to determine the most significantly enriched GO and signaling pathways. In our study, negative regulation of immune effector process, negative regulation of T cell proliferation, negative regulation of B cell activation, negative regulation of CD4 positive alpha beta T cell activation, positive regulation of I kappab kinase NF kappab signaling, positive regulation of vasculature development, apoptosis, p53 signaling pathway, JAK STAT signaling pathway, cell adhesion molecules cams, and ECM receptor interaction pathways were enriched in METTL7B high expression phenotype (**Supplementary Figure S1**).

Differences in DNA Methylation Related to METTL7B Expression

Hypermethylation in CpG islands and hypomethylation in CpG poor regions were defined as abnormal DNA methylation was positively correlated with enhanced tumorigenesis and tumor progression (Koch et al., 2018; Soozangar et al., 2018). Illumina Infinium 450k DNA methylation data from TCGA portal were used to identify and compare the effects of DNA methylation patterns in high and low METTL7B groups. METTL7B expression showed significant negative correlation ($r = -0.65$, $p < 0.0001$) with METTL7B DNA methylation (**Figure 4A**). The Pearson correlation analysis showed that methylation of METTL7B CpG sites was highly correlated with METTL7B expression. In addition, methylation of pall CpG sites, except for CpG sites cg05567435, negatively correlated with the expression of METTL7B (**Table 1**). Significant CpG sites ($|r| > 0.5$, $p < 0.0001$ (**Figures 4B–D**) were identified. Kaplan–Meier analysis was used to assess prognostic values of these significant METTL7B DNA CpG sites in patients with glioma. Analysis showed that high levels of these CpG sites were associated with good OS (**Figures 4E–G**).

Differences in Somatic Mutations Related to METTL7B Expression

Somatic mutations for high and low METTL7B expression cohorts were explored to find relevant genetic alterations. The

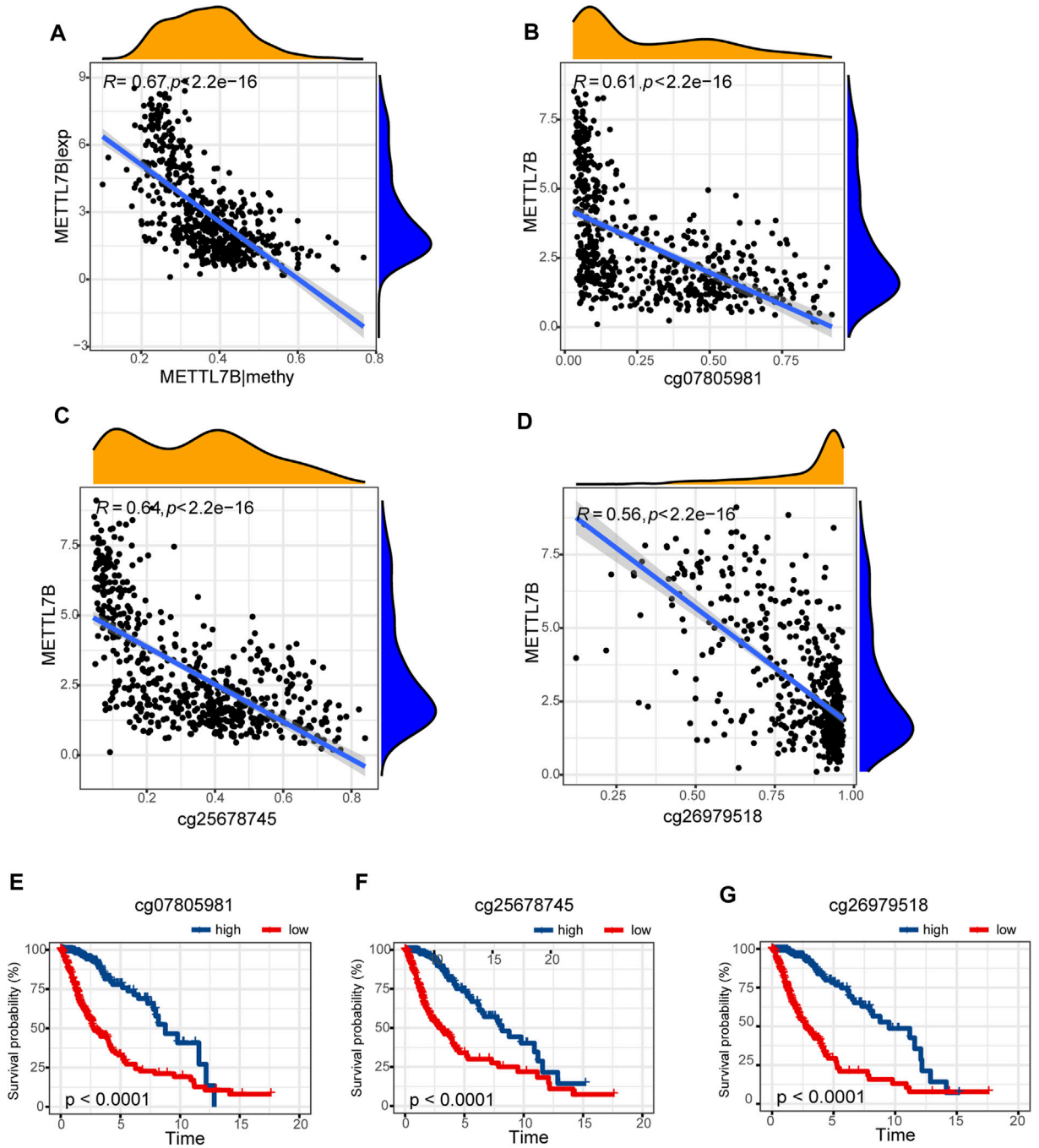


FIGURE 4 | Differences in DNA methylation related to METTL7B expression. **(A)** Expression of METTL7B was negatively regulated by METTL7B DNA methylation. **(B–D)** Expression of METTL7B was significantly negatively regulated by METTL7B DNA promoter CpG sites, including cg07805981, cg25678745, and cg26979518. **(E–G)** Kaplan-Meier analysis showed that high levels of these CpG sites were associated with good OS.

top 30 most frequently mutated genes in the corresponding cohorts are shown in **Figures 5A,B**. Tumor driver genes (such as TTN, EGFR, and PTEN) showed high mutation rates in the high METTL7B cohort compared with the rates in the low METTL7B cohort. On the contrary, IDH1 mutation showed the highest mutation frequency in low METTL7B cohort compared with high METTL7B cohort. These differential mutation rates were validated using our dataset (**Table 2**). Our findings are consistent with reports from previous studies that IDH1 mutation are correlated with more favorable OS implying

TABLE 1 | The methylation of CpGs sites negatively correlated with METTL7B.

Tag	*p* value	Cor
cg25678745	7.42E-73	−0.654
cg26979518	1.89E-69	−0.642
cg07805981	3.72E-48	−0.552
cg07530577	2.03E-26	−0.419
cg15702701	1.92E-10	−0.259
cg26537248	2.57E-10	−0.257
cg18610738	2.22E-05	−0.174
cg05567435	1.57E-01	−0.059

that IDH1 mutations play a critical role in glioma patients (**Figures 5C, D**). Moreover, top 10 genes with differential mutation frequencies between the two cohorts were identified using Fisher's exact test (**Table 2**). Association between METTL7B expression and survival of gliomas patients with different TMB scores was further analyzed. We found that TMB exhibits a positive correlation with the level of METTL7B (**Figure 5E**). Samples in TCGA were divided into four groups based on TMB score and METTL7B expression level. The resulting groups were high TMB score with high or low METTL7B expression and low TMB score with high or low METTL7B expression. Analysis of these groups showed that low METTL7B with low TMB score group had a higher OS compared with high METTL7B group with high TMB score (**Figure 5F**).

Estimation of Immune Cell-Type Fractions in Glioma

CIBERSORT method in combination with LM22 signature matrix and xCell (http://xcell.ucsf.edu/) was used to estimate differences in the distribution of immune cell types in the tumor microenvironment of low METTL7B and high METTL7B groups. Results obtained from 697 samples from TCGA and 693 samples from CGGA using CIBERSORT analysis are shown in **Figures 6A,B**. Patients with high METTL7B expression levels showed significantly higher expression levels of immunosuppressive cells (such as Tregs, TAMs, and neutrophils), NK cells, and rested T cells. On the contrary, patients with high METTL7B expression levels showed significantly lower levels of activated NK cells (**Figure 6A**). Similar to TCGA results, levels of TAMs and Tregs in the high METTL7B group were higher compared with levels in the low METTL7B group in the CGGA cohort (**Figure 6B**).

To further explore the potential function of immune cells in tumor infiltration, the expression model of TIICs in glioma with different METTL7B cohorts was studied using xCell (**Supplementary Figure S1**). Factors with a significant difference in infiltration ratio were selected. Analysis showed that TAMs were significantly higher in the low METTL7B expression group compared with the high METTL7B expression group (**Figure 6C**). TAMs originating from monocytes are an important type of immune cells in the tumor microenvironment, accounting for 50% of total immune cell counts. In addition, TAMs play an important role in neoplasia, metastasis, immune escape, and tumor angiogenesis (Liu and Cao 2015; de Groot and Pienta 2018). These findings imply that METTL7B plays an important role in the tumor immune microenvironment.

High METTL7B Expression Indicates an Immunosuppressive Microenvironment

Cancer-Immunity Cycle provides a theoretical basis for cancer immunotherapy research (Chen and Mellman 2013). In this study, we explored the expression of genes negatively regulating The Cancer-Immunity Cycle in low and high METTL7B groups. The findings showed that these genes were mostly upregulated in the high METTL7B group (**Figures 7A,B**). This implies that patients in this group have lower activities of the immune microenvironment.

Moreover, we explored the relationship between several immune-associated molecules and the expression of METTL7B. PD1 and PD-L1 levels were high in the high METTL7B group compared with the levels in the low METTL7B group (**Figures 7C,D**). In addition, the expression of crucial immune checkpoints (i.e., CTLA-4 and LAG-3) in the high group was significantly higher compared with the levels in the low METTL7B group (**Figures 7E,F**). In addition, the expression levels of immunosuppressive cytokines were significantly higher in the high METTL7B group compared with levels in the low METTL7B group (**Figures 8G,H**). These findings imply that the high METTL7B expression level promotes the immunosuppressive microenvironment through upregulation of immune checkpoints and immunosuppressive cytokines.

High METTL7B Predicts Response of Immune Checkpoint Blockade (ICB) Therapy

To further determine the effect of high METTL7B expression on immunotherapy of glioma patients, ImmuCellAI (http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/), was used to predict immune checkpoint blockade (ICB) therapy response. Immune cell abundance estimated using TCGA and CGGA databases was used for the prediction. The results showed that glioma (LGG and GBM) patients with high expression levels of METTL7B have a better response to immunotherapy and GBM have worse response, compared to LGG (**Figure 8**).

Construction and Validation of Prognosis Prediction Model

The above findings show significant METTL7B-related alterations in multi-omics characteristics including expression profile change, DNA methylation, somatic mutation, and immune cells. A total of 31 upregulated immunosuppressive genes and immune checkpoints were analyzed in the high METTL7B cohort in TCGA and CGGA datasets to explore expression changes. The top 10 frequently mutated genes were

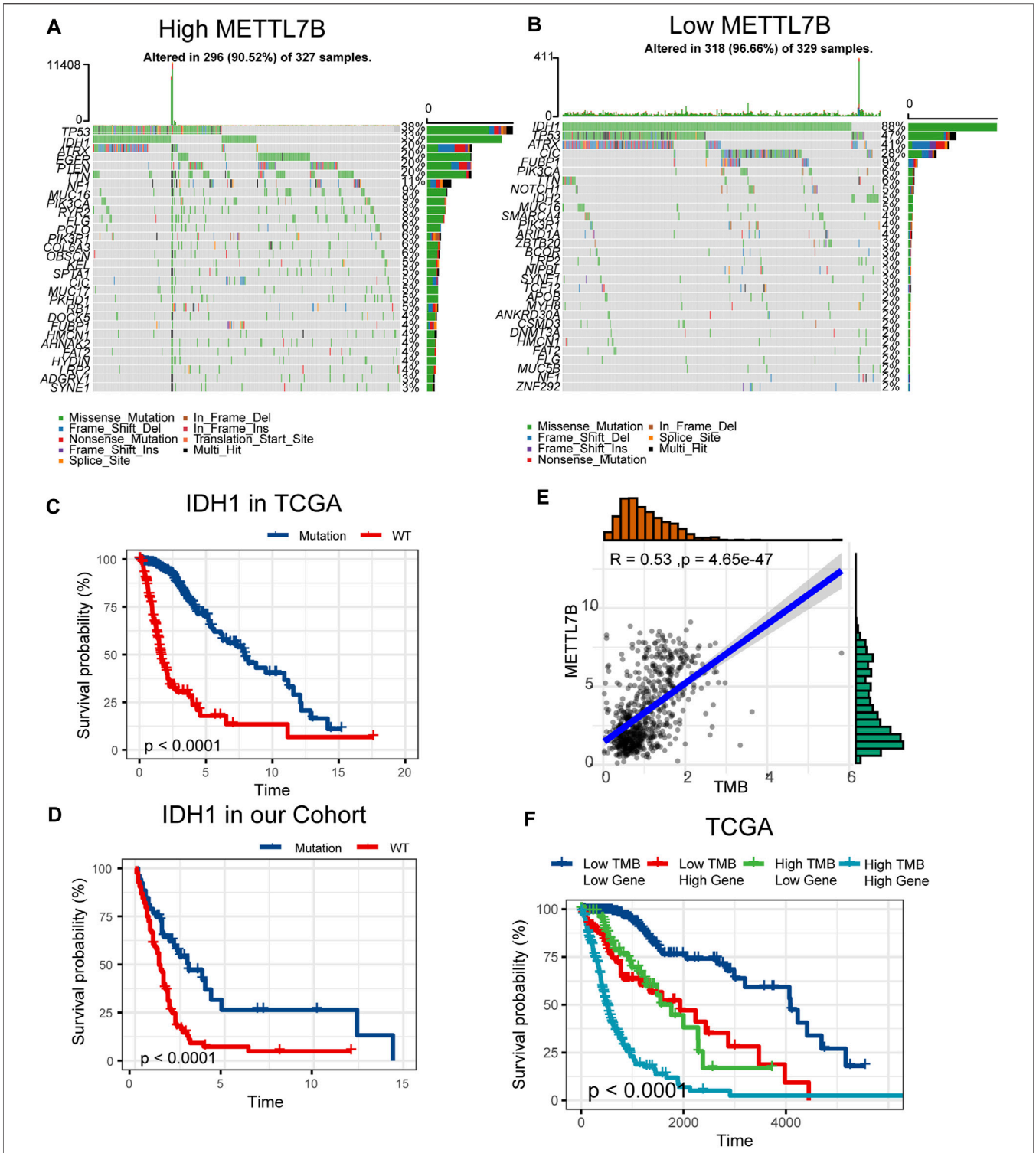


FIGURE 5 | Differences in somatic mutations related to METTL7B expression. **(A–B)** Top 30 most frequently mutated genes between high and low METTL7B groups. **(C–D)** Survival analysis of gliomas with different IDH1 status in the TCGA dataset and our cohort. **(E)** Correlation curve between METTL7B and the level of TMB. **(F)** Survival analysis of gliomas between different TMB scores and high or low METTL7B expression levels.

TABLE 2 | The most differential frequently mutated genes.

Gene	Low num	High num	*p* value
FUBP1	26	10	0.009
TP53	152	121	0.018
MUC16	15	29	0.029
IDH1	289	108	<0.0001
ATRX	125	62	<0.0001
CIC	89	16	<0.0001
TTN	20	64	<0.0001
PTEN	3	64	<0.0001
EGFR	1	64	<0.0001
NF1	6	32	<0.0001

identified in the low METTL7B and high METTL7B cohorts to explore somatic mutations. Analysis of DNA methylation showed a total of 8 differential methylation probes at the regions of METTL7B. Moreover, a total of 21 differential immune cells were detected in differential METTL7B cohorts for the TCGA and CGGA datasets.

Univariate Cox regression analysis and LASSO-COX analysis were used to determine key prognostic biomarkers, and a risk signature was constructed. TCGA samples were randomly divided into training and independent test sets. Multi-omics data were missing from the other databases; therefore, they were not included in the analysis. The formula for the risk signature was determined using corresponding coefficients: risk signature = 0.1596 × (METTL7B) + 0.2345 × CD274 + 0.0149 × PDCD1 + −0.2305 × VTCN1 + 0.0026 × IL10 + 0.0823 × EZH2 + −0.7471 × cg07805981 + −0.3115 × cg25678745 + -0.6065 × cg26979518 + -0.3416 × Basophils (Xcell) + 0.0664 × $CD4^+$ memory T cells (Xcell) + -0.1906 × eosinophils (Xcell) + -0.5024 × IDH1|snv + −0. 7,348 × CIC|snv.

Kaplan–Meier analysis showed significantly poorer prognosis for glioma patients with high risk signature ($p < 0.001$, **Figures 9A,B**). ROC curve analysis showed that the risk signature had better-predictive power for 1-year, 3-year, and 5-year survival (AUC = 0.904.0.981.0.859 **Figure 9C**) compared with the use of METTL7B expression level (AUC = 0.781.0.841.0.812 **Figure 2F**) alone. This finding was further validated using the test set (**Figure 9D**). A prognostic nomogram was used as a quantitative tool to predict the survival of patients. The nomogram combined risk signature with clinical information of glioma patients (**Figure 9E**). Moreover, the calibration curve of the nomogram showed consistency between prediction and observation (C-index = 0.88) (**Figure 9F**). The training nomogram was not significantly different in 3-year (AUC = 0.982) and 5-year (AUC = 0.825) survival predictions (**Figure 9G**) compared with the risk signature constructed based on multi-omics alterations. However, the test nomogram displayed better-predictive power of 3- and 5-year survival (AUC = 0.907.0.871) compared with the risk signature constructed using the test set (**Figure 9G**).

DISCUSSION

This study presents the clinical prognostic value of METTL7B expression in glioma. Further, a significant association between METTL7B expression and clinical-pathological features is reported. Functional enrichment analysis showed that METTL7B negatively regulates immunity and carcinogenic signaling pathways. Moreover, a high METTL7B expression level is associated with high levels of immunosuppressive cells implying that it promotes the immunosuppressive microenvironment. A METTL7B-related prognostic signature was constructed and validated based on the multi-omics. The prognostic signature showed a high value in predicting the overall survival (OS) time of glioma patients. In conclusion, these findings provide insights on the pathologic role of METTL7B in promoting tumor progression and its potential value as a new diagnostic and prognostic biomarker for glioma.

METTL7B is an alkyl thiol methyltransferase that metabolizes hydrogen sulfide (H_2S) (Maldonato et al., 2021). H_2S is detected to induce NSCLC migration and invasion, as well as the epithelial-mesenchymal transition (EMT) process (Wang et al., 2020). In the advanced stage of thyroid cancer, METTL7B plays an important role in regulating EMT (Cai, Chen, Chen, Li, Du, and Zhou 2018). Moreover, METTL7B knockdown promoted cell cycle arrest at G0/G1phase and induced cellular apoptosis in clear cell renal cell carcinoma (Li et al., 2021). In our study, KEGG analysis demonstrated that METTL7B is involved in cell adhesion molecules cams and apoptosis. These findings imply that METTL7B may promote the growth and metastasis of glioma cells by being involved in the metabolism of H_2S.

Integrative risk signature comprising immune-related alterations and epigenetic regulation was used to estimate the role of METTL7B due to the complex pathogenesis of glioma. Treg cells, TAMs, and neutrophils are the most abundant immunosuppressive cells in the tumor microenvironment. Treg cells are recruited into the tumor microenvironment (TME) and inhibit anti-tumor immune responses, thus affecting the effectiveness of cancer immunotherapy (Yano et al., 2019). TAMs are major components in TME. Monocytes undergo reprogramming into TAMs after recruitment at the tumor site leading to the gain of protumoral functions such as supporting tumor growth; promoting angiogenesis, tumor invasion, and metastases; and suppressing T cells responsible for antitumoral responses (Netea-Maier et al., 2018). High levels of immunosuppressive cells (Treg cells, TAMs, and neutrophils) were observed in the high METTL7B group. In addition, immunosuppressive cells were significantly enriched in negative regulation of immunity pathway indicating that METTL7B can be used to predict the immune microenvironment.

Tumor immunosuppressive cytokines and immune checkpoints play an important role in tumorigenesis and cancer development by enhancing tumor immune escape.

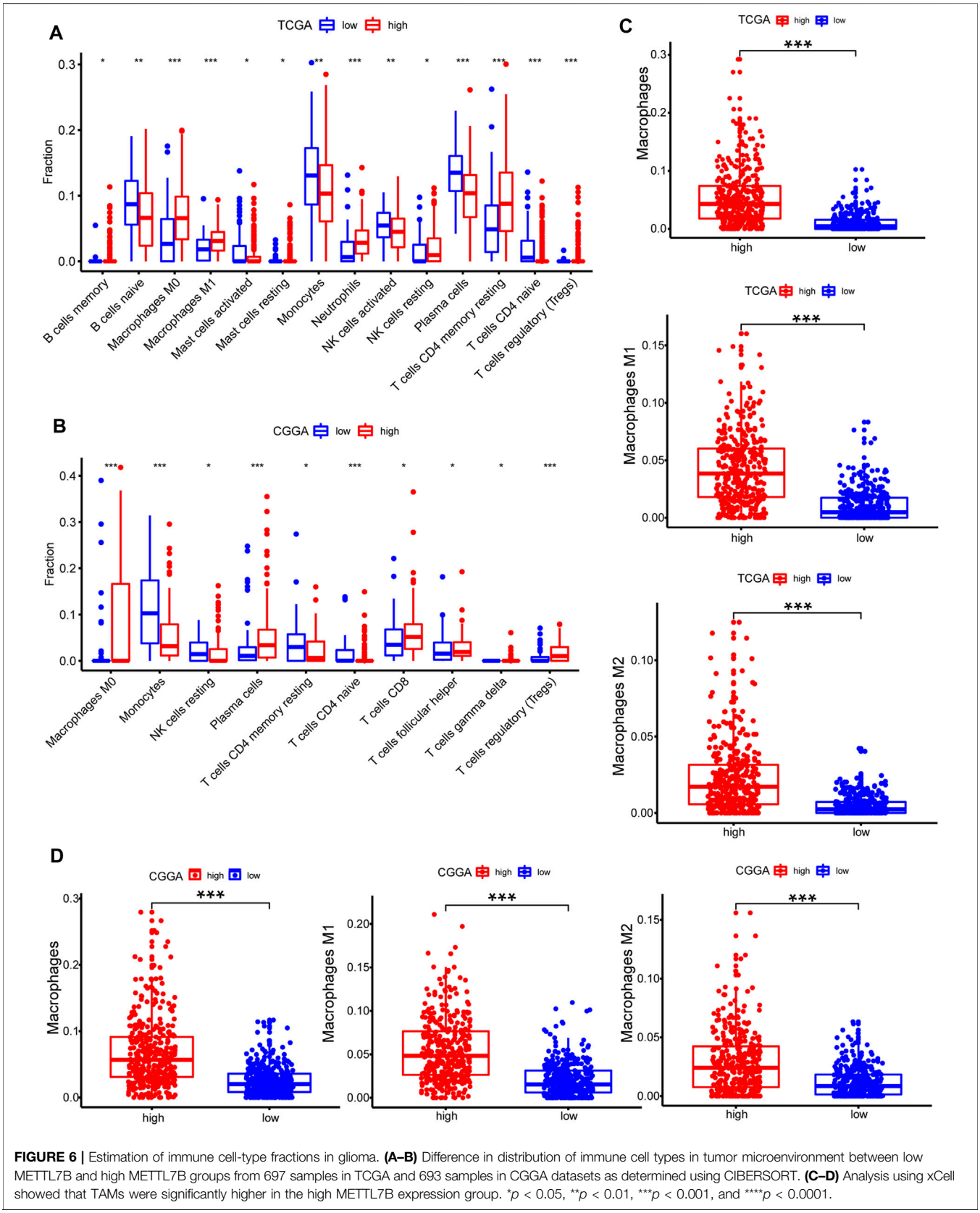


FIGURE 6 | Estimation of immune cell-type fractions in glioma. **(A–B)** Difference in distribution of immune cell types in tumor microenvironment between low METTL7B and high METTL7B groups from 697 samples in TCGA and 693 samples in CGGA datasets as determined using CIBERSORT. **(C–D)** Analysis using xCell showed that TAMs were significantly higher in the high METTL7B expression group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

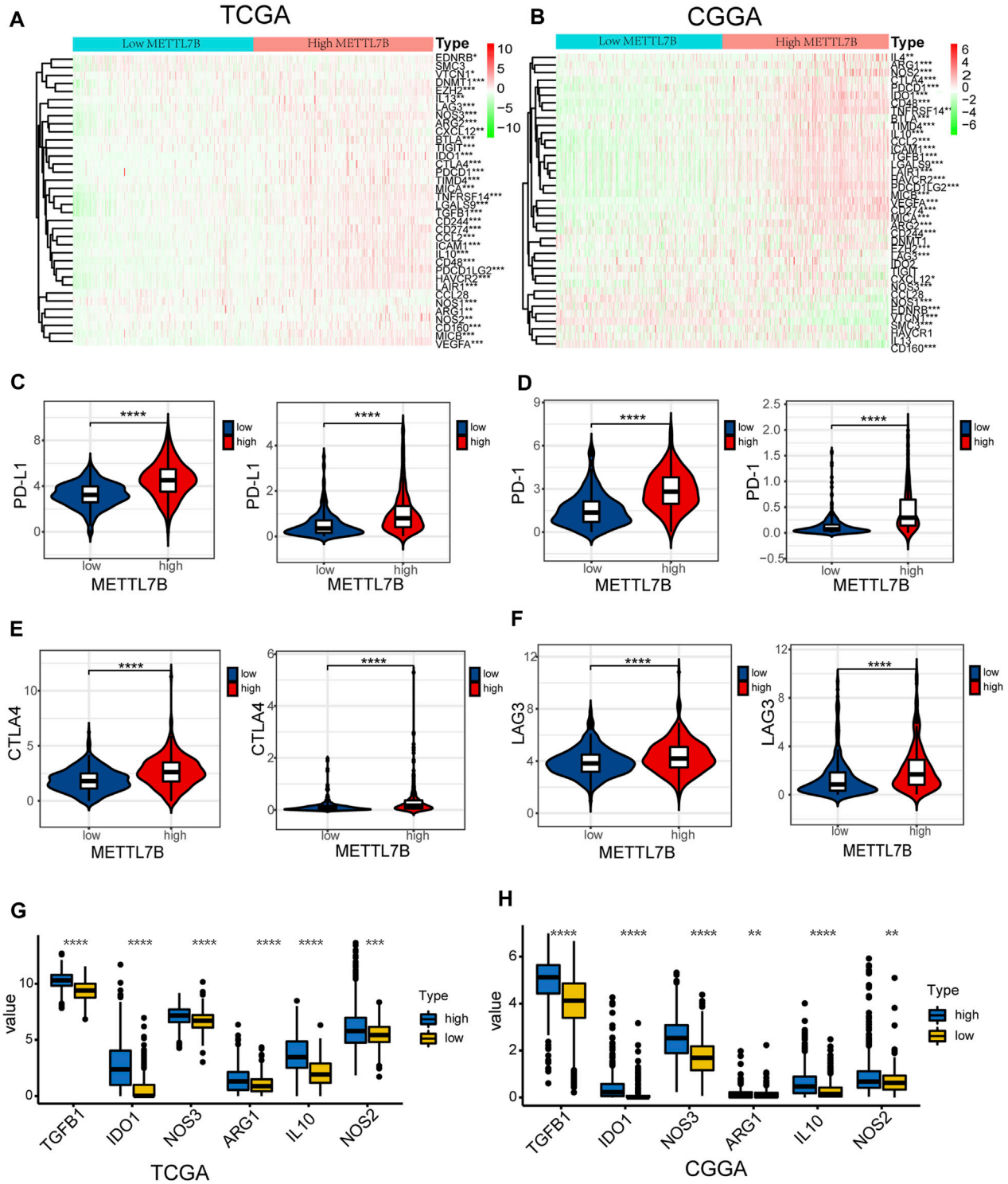


FIGURE 7 | High METTL7B expression indicates an immunosuppressive microenvironment. **(A–B)** Heatmap of profiles of genes involved in negative regulation of the Cancer-Immunity Cycle in high and low METTL7B groups in the TCGA and CGGA datasets. **(C–F)** PD-L1, PD-1, CTLA4, and LAG3 expression levels in high and low METTL7B groups. **(G–H)** Tumor immunosuppressive cytokine expression in high and low METTL7B groups. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

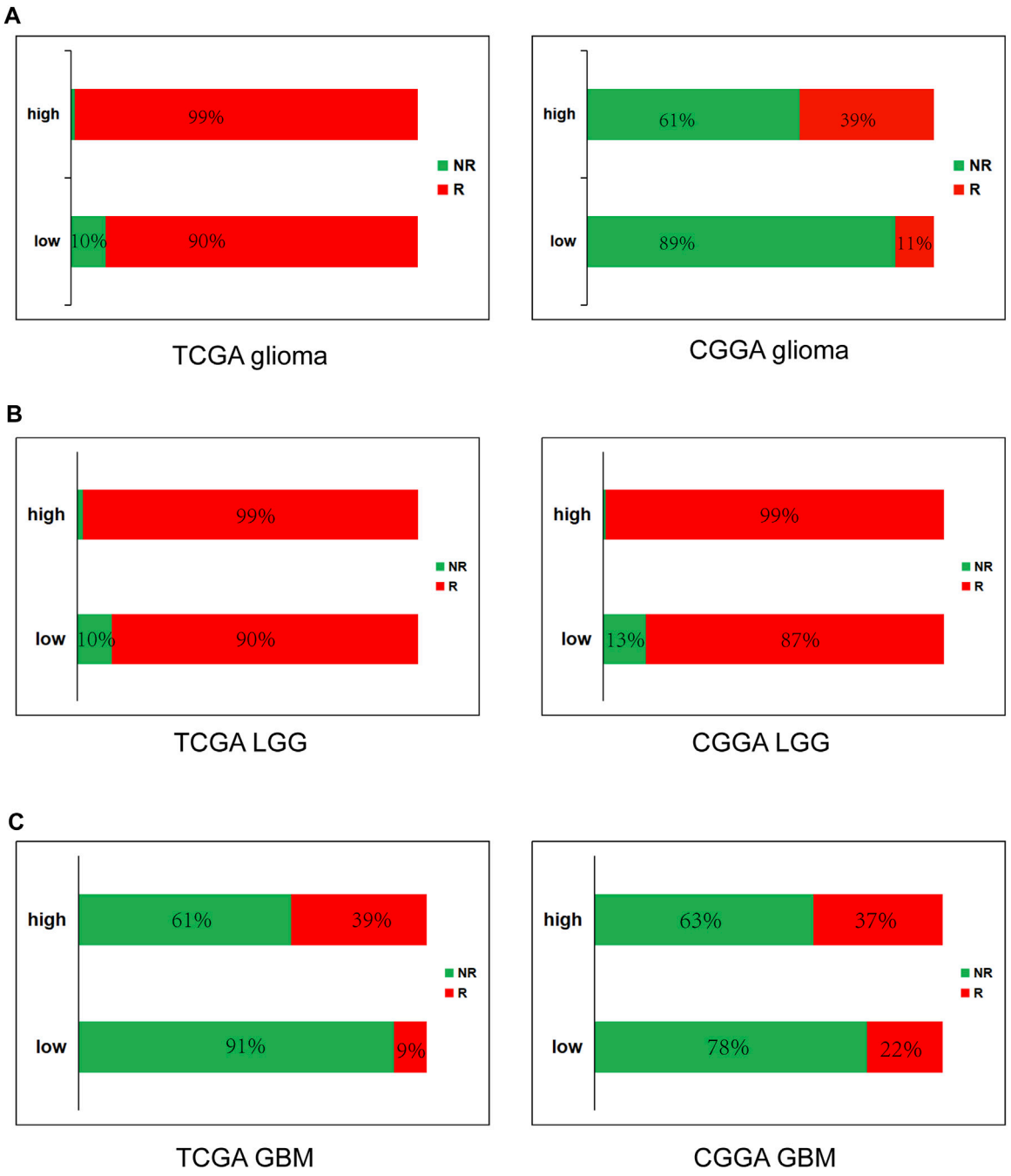


FIGURE 8 | High METTL7B expression is correlated with response of ICB therapy as shown using ImmuCellAI in glioma **(A)**, LGG **(B)** and GBM **(C)**.

Increased expression of tumor immunosuppressive cytokines is a key feature of immune cell exhaustion. TGF-β has been shown to suppress the immune response by inhibiting NK-cell activity, decreasing cytokine production, inhibiting dendritic cell maturation, and altering T-cell cytotoxic properties (Haque and Morris 2017). Previous studies report that M2-macrophages, Tregs, and Th2-cells produce IL-10, which is implicated in the impairment of proliferation, cytokine production, and migratory capacities of effector T cells (Barbi et al., 2014). Stimulation of immune checkpoint targets is the main mechanism through which tumors escape immune cells attack. In our study, immunosuppressive cytokines and immune checkpoints like PD1, PD-L1, LAG3, and CTLA-4

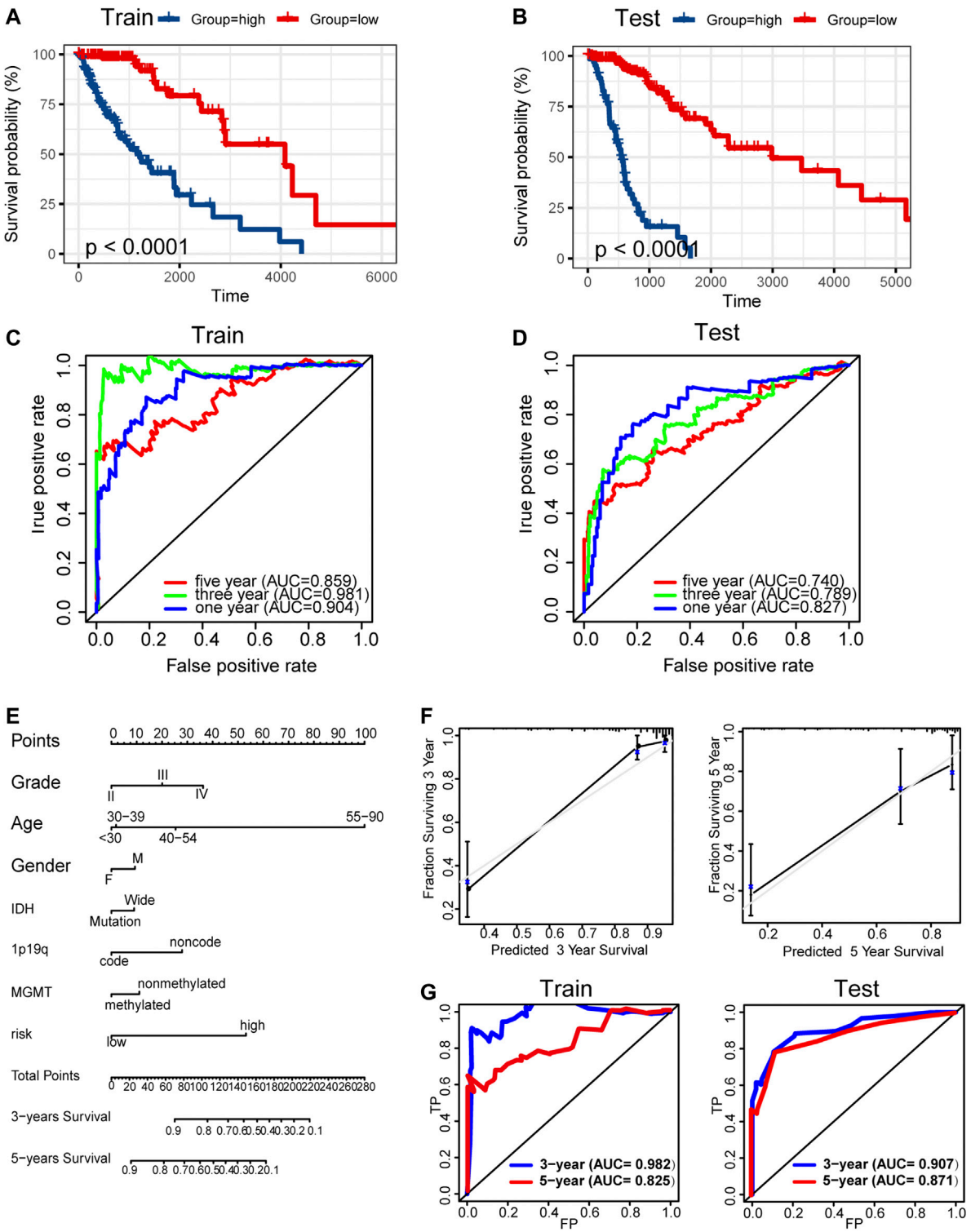


FIGURE 9 | Construction and validation of prognosis prediction model. **(A–B)** Kaplan–Meier overall survival curves for gliomas assigned to high and low risk groups based on risk score in the training and test sets. **(C–D)** ROC curves showing predictive efficiency of the risk signature on 1-year, 3-year, and 5-year survival rate. **(E)** Nomogram for predicting 3- and 5-year overall survival with risk signature and clinical information in the training set. **(F)** Calibration curve of the nomogram using the training set. **(G)** ROC curves of the prediction index value in predicting 3- and 5-year overall survival of the training and test nomogram.

were upregulated in the high METTL7B group. These findings imply that METTL7B helps cancer cells in evading natural anti-tumor immune responses by decreasing the activity of NK or CTL cells, enhancing suppressive cells (Tregs, TAMs, and neutrophils), and increasing immunosuppressive molecular factors. Moreover, analysis using ImmuCellAI showed that patients with high METTL7B had a better response to immunotherapy. Therefore, targeting METTL7B may have significant clinical implications in improving immunotherapy.

A previous study reports that abnormal DNA methylation is implicated in the induction and progression of glioma (Mathur et al., 2020). In our study, METTL7B methylation showed a negative correlation with METTL7B mRNA expression in gliomas. Low methylation of METTL7B in glioma tissues is attributed to the negative correlation between METTL7B methylation and expression levels. We further explored specific CpG sites in METTL7B DNA promoter at which methylation is significantly correlated with METTL7B mRNA expression. Notably, cg07805981, cg25678745, and cg26979518 showed significant associations with METTL7B expression. Therefore, we evaluated the prognostic value of these three CpG sites. METTL7B hypermethylation showed a significantly high OS in patients with glioma. In summary, METTL7B expression was negatively regulated by METTL7B methylation, and METTL7B methylation status is a potential prognostic factor of OS.

Pathogenesis of tumor may be related to the variation of the exon group due to its complexity and involvement of multiple genes. Several genes, including IDH, TP53, PTEN, and EGFR, undergo mutations in gliomas (Wong et al., 1992; Cancer Genome Atlas Research Network, 2008). These mutations occur in a defined order during progression to a high-grade tumor. IDH mutations occur early in the development of glioma from a stem cell that gives rise to both astrocytes and oligodendrocytes (Yan et al., 2009). TP53 mutation is a relatively early event during the development of an astrocytoma, whereas loss or mutation of PTEN and amplification of EGFR are characteristics of higher-grade tumors (Weber et al., 1996; Furnari et al., 2007; Ohgaki and Kleihues 2007). In our study, tumors with IDH1 mutations showed a better outcome compared with those with wild-type IDH genes. Previous studies report that METTL7B can be induced by mutant P53 and wild-type P53 protein through interaction with the upstream promoter region of METTL7B (Neilsen et al., 2011). KEGG analysis in our study revealed that METTL7B is involved in p53 pathway. However, fewer p53 mutations were identified in the high METTL7B group which can be attributed to the complexity of high-grade gliomas. TMB, a novel biomarker for the prediction of immune responses, is effective in various cancers, such as breast cancer (Park et al., 2018; Thomas et al., 2018). In our study, the high METTL7B group with high TMB showed high OS compared with the low METTL7B group with low TMB. Similar results have been reported in most malignancies that higher TMB induces local immune recognition and improves prognosis.

Bioinformatics is a flourishing study approach. Through data analysis, many potential tumor markers can be found for the study of antitumor treatments. Previous studies have shown that m6A related genes may predict the prognosis or be used for the diagnosis of glioma (Dong and Cui 2020). Ye et al. found a novel ferroptosis-related gene signature for prognostic prediction in glioma patients and revealed the relationship between ferroptosis-related genes and immune checkpoint molecules (Chen et al., 2021). A hypoxia risk model based on five hypoxia-associated genes, which served as an independent prognostic indicator and reflected overall immune response intensity in the glioma microenvironment (Lin et al., 2020). The discovery of these biomarkers enables us to understand more about the mechanism of glioma development, thus aiding the clinical diagnosis and treatment of glioma. However, the prognosis of tumors is complex and cannot be predicted accurately by a single index. The combined analysis of multiple indicators or multi-omics will improve the accuracy of prognosis prediction. In the present study, a risk signature constructed based on multi-omics characteristics showed a superior prediction capability with a higher AUC value compared with the risk signature constructed from METTL7B levels alone. In addition, a comprehensive nomogram, in which combined risk signature was built using clinical information of glioma patients, has a favorable advantage in prognostic prediction.

Although this study provides a more integrative view of the association between METTLL7B expression and glioma and a prognostic model with high predictive capabilities, it has limitations that should be explored further. Notably, matching multi-omics data was missing in other data sources. This prevented us from examining the robustness of the model when used for other data.

In summary, our findings revealed that METTL7B was overexpressed in glioma and identified it as a promising prognostic biomarker. Moreover, the METTL7B expression level was positively correlated with immunosuppressive cells implying that it may play an important role in regulating the microenvironment. Our prognostic prediction model can be used in clinical applications to improve OS and for the development of new effective therapeutic strategies for glioma 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 authors.

AUTHOR CONTRIBUTIONS

SQ and WL conceived the project and designed this study. XC analyzed the data and wrote the manuscript. YL and SW helped to analyze the data. WL helped to perform immunohistochemical analysis. ML, TL, and YW helped in the acquisition of data. CL

helped to revise the manuscript. All authors approved the final version for submission.

FUNDING

The project was funded by Joint Funds for the Innovation of Science and Technology, Fujian province (2018Y9105); United Fujian Provincial Health and Education Project for Tackling the Key Research, China (2019-WJ-03); Science and Technology Program of Fujian Province, China (2018Y2003); National Natural Science Foundation of China (11974077, U1605253, 82072986).

ACKNOWLEDGMENTS

We thank Zhang and Pan for useful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.727481/full#supplementary-material

REFERENCES

Abdel-Hameed, E. A., Ji, H., Sherman, K. E., and Shata, M. T. M. (2014). Epigenetic Modification of FOXP3 in Patients with Chronic HIV Infection. *J. Acquir Immune Defic Syndr.* 65, 19–26. doi:10.1097/qai.0b013e3182a1bca4

Ali, J., Liu, W., Duan, W., Liu, C., Song, J., Ali, S., et al. (2020). METTL7B (Methyltransferase-like 7B) Identification as a Novel Biomarker for Lung Adenocarcinoma. *Ann. Transl Med.* 8, 1130. doi:10.21037/atm-20-4574

Barbi, J., Pardoll, D., and Pan, F. (2014). Treg Functional Stability and its Responsiveness to the Microenvironment. *Immunol. Rev.* 259, 115–139. doi:10.1111/imr.12172

Cai, W. Y., Chen, X., Chen, L. P., Li, Q., Du, X. J., and Zhou, Y. Y. (2018). Role of Differentially Expressed Genes and Long Non-coding RNAs in Papillary Thyroid Carcinoma Diagnosis, Progression, and Prognosis. *J. Cel. Biochem.* 119, 8249–8259. doi:10.1002/jcb.26836

Cancer Genome Atlas Research Network (2008). Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways. *Nature* 455, 1061–1068. doi:10.1038/nature07385

Chen, D. S., and Mellman, I. (2013). Oncology Meets Immunology: the Cancer-Immunity Cycle. *Immunity* 39, 1–10. doi:10.1016/j.immuni.2013.07.012

Chen, Z., Wu, T., Yan, Z., and Zhang, M. (2021). Identification and Validation of an 11-Ferroptosis Related Gene Signature and its Correlation with Immune Checkpoint Molecules in Glioma. *Front. Cel Dev. Biol.* 9, 652599. doi:10.3389/fcell.2021.652599

de Groot, A. E., and Pienta, K. J. (2018). Epigenetic Control of Macrophage Polarization: Implications for Targeting Tumor-Associated Macrophages. *Oncotarget* 9 (9), 20908–20927. doi:10.18632/oncotarget.24556

Deng, X., Su, R., Weng, H., Huang, H., Li, Z., and Chen, J. (2018). RNA N6-Methyladenosine Modification in Cancers: Current Status and Perspectives. *Cell Res* 28, 507–517. doi:10.1038/s41422-018-0034-6

Dong, Z., and Cui, H. (2020). The Emerging Roles of RNA Modifications in Glioblastoma. *Cancers (Basel)* 12, 12. doi:10.3390/cancers12030736

Furnari, F. B., Fenton, T., Bachoo, R. M., Mukasa, A., Stommel, J. M., Stegh, A., et al. (2007). Malignant Astrocytic Glioma: Genetics, Biology, and Paths to Treatment. *Genes Develop.* 21 (21), 2683–2710. doi:10.1101/gad.1596707

Haque, S., and Morris, J. C. (2017). Transforming Growth Factor-β: A Therapeutic Target for Cancer. *Hum. Vaccin. Immunother.* 13, 1741–1750. doi:10.1080/21645515.2017.1327107

Ignatova, V. V., Jansen, P. W. T. C., Baltissen, M. P., Vermeulen, M., and Schneider, R. (2019). The Interactome of a Family of Potential Methyltransferases in HeLa Cells. *Sci. Rep.* 9 (9), 6584. doi:10.1038/s41598-019-43010-2

Jiang, T., Mao, Y., Ma, W., Mao, Q., You, Y., Yang, X., et al. (2016). CGCG Clinical Practice Guidelines for the Management of Adult Diffuse Gliomas. *Cancer Lett.* 375, 263–273. doi:10.1016/j.canlet.2016.01.024

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The Human Genome Browser at UCSC. *Genome Res.* 12, 996–1006. doi:10.1101/gr.229102

Koch, A., Joosten, S. C., Feng, Z., de Ruijter, T. C., Draht, M. X., Melotte, V., et al. (2018). Analysis of DNA Methylation in Cancer: Location Revisited. *Nat. Rev. Clin. Oncol.* 15, 459–466. doi:10.1038/s41571-018-0004-4

Lapointe, S., Perry, A., and Butowski, N. A. (2018). Primary Brain Tumours in Adults. *Lancet* 392, 432–446. doi:10.1016/S0140-6736(18)30990-5

Li, W., Xu, S., Peng, N., Zhang, Z., He, H., Chen, R., et al. (2021). Downregulation of METTL7B Inhibits Proliferation of Human Clear Cell Renal Cancer Cells *In Vivo* and *In Vitro*. *Front. Oncol.* 11, 634542. doi:10.3389/fonc.2021.634542

Lin, W., Wu, S., Chen, X., Ye, Y., Weng, Y., Pan, Y., et al. (2020). Characterization of Hypoxia Signature to Evaluate the Tumor Immune Microenvironment and Predict Prognosis in Glioma Groups. *Front. Oncol.* 10, 796. doi:10.3389/fonc.2020.00796

Liu, D., Li, W., Zhong, F., Yin, J., Zhou, W., Li, S., et al. (2020). METTL7B Is Required for Cancer Cell Proliferation and Tumorigenesis in Non-small Cell Lung Cancer. *Front. Pharmacol.* 11, 178. doi:10.3389/fphar.2020.00178

Liu, X., Li, Y., Li, S., Fan, X., Sun, Z., Yang, Z., et al. (2019). IDH Mutation-specific Radiomic Signature in Lower-Grade Gliomas. *Aging* 11, 673–696. doi:10.18632/aging.101769

Liu, Y., and Cao, X. (2015). The Origin and Function of Tumor-Associated Macrophages. *Cell Mol Immunol* 12, 1–4. doi:10.1038/cmi.2014.83

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., et al. (2007). The 2007 WHO Classification of Tumours of the central Nervous System. *Acta Neuropathol.* 114, 97–109. doi:10.1007/s00401-007-0243-4

Maldonato, B. J., Russell, D. A., and Totah, R. A. (2021). Human METTL7B Is an Alkyl Thiol Methyltransferase that Metabolizes Hydrogen Sulfide and Captopril. *Sci. Rep.* 11 (11), 4857. doi:10.1038/s41598-021-84218-5

Mathur, R., Zhang, Y., Grimmer, M. R., Hong, C., Zhang, M., Bollam, S., et al. (2020). MGMT Promoter Methylation Level in Newly Diagnosed Low-Grade Glioma Is a Predictor of Hypermutation at Recurrence. *Neuro Oncol.* 22, 1580. doi:10.1093/neuonc/noaa059

Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* 178, 835–e21. doi:10.1016/j.cell.2019.06.024

Neilsen, P. M., Noll, J. E., Suetani, R. J., Schulz, R. B., Al-Ejeh, F., Evdokiou, A., et al. (2011). Mutant P53 Uses P63 as a Molecular Chaperone to Alter Gene Expression and Induce a Pro-invasive Secretome. *Oncotarget* 2, 1203–1217. doi:10.18632/oncotarget.382

Netea-Maier, R. T., Smit, J. W. A., and Netea, M. G. (2018). Metabolic Changes in Tumor Cells and Tumor-Associated Macrophages: A Mutual Relationship. *Cancer Lett.* 413, 102–109. doi:10.1016/j.canlet.2017.10.037

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, 453–457. doi:10.1038/nmeth.3337

Ohgaki, H., and Kleihues, P. (2007). Genetic Pathways to Primary and Secondary Glioblastoma. *Am. J. Pathol.* 170, 1445–1453. doi:10.2353/ajpath.2007.070011

Park, S. E., Park, K., Lee, E., Kim, J.-Y., Ahn, J. S., Im, Y.-H., et al. (2018). Clinical Implication of Tumor Mutational burden in Patients with HER2-Positive Refractory Metastatic Breast Cancer. *Oncoimmunology* 7, e1466768. doi:10.1080/2162402x.2018.1466768

Soozangar, N., Sadeghi, M. R., Jeddi, F., Somi, M. H., Shirmohamadi, M., and Samadi, N. (2018). Comparison of Genome-wide Analysis Techniques to DNA Methylation Analysis in Human Cancer. *J. Cel Physiol* 233, 3968–3981. doi:10.1002/jcp.26176

Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., et al. (2005). Radiotherapy Plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N. Engl. J. Med.* 352, 987–996. doi:10.1056/NEJMoa043330

Stupp, R., Hegi, M. E., Mason, W. P., van den Bent, M. J., Taphoorn, M. J., Janzer, R. C., et al. (2009). Effects of Radiotherapy with Concomitant and Adjuvant

Temozolomide versus Radiotherapy Alone on Survival in Glioblastoma in a Randomised Phase III Study: 5-year Analysis of the EORTC-NCIC Trial. *Lancet Oncol.* 10, 459–466. doi:10.1016/s1470-2045(09)70025-7

Sun, D., Wang, J., Han, Y., Dong, X., Ge, J., Zheng, R., et al. (2021). TISCH: a Comprehensive Web Resource Enabling Interactive Single-Cell Transcriptome Visualization of Tumor Microenvironment. *Nucleic Acids Res.* 49, D1420–d1430. doi:10.1093/nar/gkaa1020

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

Thomas, A., Klein, M. S., Stevens, A. P., Reinders, Y., Hellerbrand, C., Dettmer, K., et al. (2013). Changes in the Hepatic Mitochondrial and Membrane Proteome in Mice Fed a Non-alcoholic Steatohepatitis Inducing Diet. *J. Proteomics* 80, 107–122. doi:10.1016/j.jprot.2012.12.027

Thomas, A., Routh, E. D., Pullikuth, A., Jin, G., Su, J., Chou, J. W., et al. (2018). Tumor Mutational burden Is a Determinant of Immune-Mediated Survival in Breast Cancer. *Oncoimmunology* 7, e1490854. doi:10.1080/2162402x.2018.1490854

Wang, M., Yan, J., Cao, X., Hua, P., and Li, Z. (2020). Hydrogen Sulfide Modulates Epithelial-Mesenchymal Transition and Angiogenesis in Non-small Cell Lung Cancer via HIF-1α Activation. *Biochem. Pharmacol.* 172, 113775. doi:10.1016/j.bcp.2019.113775

Weber, R. G., Sabel, M., Reifenberger, J., Sommer, C., Oberstrass, J., Reifenberger, G., et al. (1996). Characterization of Genomic Alterations Associated with Glioma Progression by Comparative Genomic Hybridization. *Oncogene* 13 (13), 983–994.

Wesseling, P., and Capper, D. (2018). WHO 2016 Classification of Gliomas. *Neuropathol. Appl. Neurobiol.* 44, 139–150. doi:10.1111/nan.12432

Wong, A. J., Ruppert, J. M., Bigner, S. H., Grzeschik, C. H., Humphrey, P. A., Bigner, D. S., et al. (1992). Structural Alterations of the Epidermal Growth Factor Receptor Gene in Human Gliomas. *Proc. Natl. Acad. Sci.* 89, 2965–2969. doi:10.1073/pnas.89.7.2965

Xu, K., Yang, Y., Feng, G.-H., Sun, B.-F., Chen, J.-Q., Li, Y.-F., et al. (2017). Mettl3-mediated m6A Regulates Spermatogonial Differentiation and Meiosis Initiation. *Cel Res* 27, 1100–1114. doi:10.1038/cr.2017.100

Xu, L., Liu, X., Sheng, N., Oo, K. S., Liang, J., Chionh, Y. H., et al. (2017). Three Distinct 3-methylcytidine (m3C) Methyltransferases Modify tRNA and mRNA in Mice and Humans. *J. Biol. Chem.* 292, 14695–14703. doi:10.1074/jbc.M117.798298

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 78, 6575–6580. doi:10.1158/0008-5472.can-18-0689

Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., et al. (2009). IDH1 and IDH2 Mutations in Gliomas. *N. Engl. J. Med.* 360, 765–773. doi:10.1056/NEJMoa0808710

Yano, H., Andrews, L. P., Workman, C. J., and Vignali, D. A. A. (2019). Intratumoral Regulatory T Cells: Markers, Subsets and Their Impact on Anti-tumor Immunity. *Immunology* 157, 232–247. doi:10.1111/imm.13067

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 Chen, Li, Li, Wu, Liu, Lin, Li, Weng, Lin and Qiu. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fmolb.2021.735263

Significance of Parkinson Family Genes in the Prognosis and Treatment Outcome Prediction for Lung Adenocarcinoma

Yanqi Li[1], Xiao Lu[1], Jiao Zhang[1], Quanxing Liu[1], Dong Zhou[1], Xufeng Deng[1], Yuan Qiu[2], Qian Chen[3], Manyuan Li[1], Guixue Yang[1], Hong Zheng[1] and Jigang Dai[1]**

[1]Department of Thoracic Surgery, Xinqiao Hospital, Army Medical University (Third Military Medical University), Chongqing, China,
[2]Department of General Surgery, Xinqiao Hospital, Army Medical University (Third Military Medical University), Chongqing, China,
[3]Cancer Center of Daping Hospital, Third Military Medical University (Army Medical University), Chongqing, China

Epidemiological investigations have shown that patients with Parkinson's disease (PD) have a lower probability of developing lung cancer. Subsequent research revealed that PD and lung cancer share specific genetic alterations. Therefore, the utilisation of PD biomarkers and therapeutic targets may improve lung adenocarcinoma (LUAD) diagnosis and treatment. We aimed to identify a gene-based signature from 25 Parkinson family genes for LUAD prognosis and treatment choice. We analysed Parkinson family gene expression and protein levels in LUAD, utilising multiple databases. Least absolute shrinkage and selection operator (LASSO) regression was used to construct a prognostic model based on the TCGA-LUAD cohort. We validated the model in external GEO cohorts. Immune cell infiltration was compared between risk groups, and GEO data was used to explore the model's predictive ability for LUAD treatment response. Nearly all Parkinson family genes exhibited significant differential expression between LUAD and normal tissues. LASSO regression confirmed that our seven Parkinson family gene-based signature had excellent prognostic performance for LUAD, as validated in three GEO cohorts. The high-risk group was clearly associated with low tumour immune cell infiltration, suggesting that immunotherapy may not be an optimal treatment choice. This is the first Parkinson family gene-based model for the prediction of LUAD prognosis and treatment outcome. The association of these genes with poor prognosis and low immune infiltration requires further investigation.

Keywords: Parkinson gene family, LUAD, prognosis, tumour mutation burden, neoantigen, immunotherapy

OPEN ACCESS

Edited by:
Florence Le Calvez-Kelm, International Agency For Research On Cancer (IARC), France

Reviewed by:
Jasim Khan, University of Alabama at Birmingham, United States
Abhishek Jauhari, University of Pittsburgh, United States

****Correspondence:***
Hong Zheng
ziecoe@tmmu.edu.cn
Jigang Dai
daijigang@tmmu.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *02 July 2021*
Accepted: *07 September 2021*
Published: *20 September 2021*

Citation:
Li Y, Lu X, Zhang J, Liu Q, Zhou D, Deng X, Qiu Y, Chen Q, Li M, Yang G, Zheng H and Dai J (2021) Significance of Parkinson Family Genes in the Prognosis and Treatment Outcome Prediction for Lung Adenocarcinoma. Front. Mol. Biosci. 8:735263. doi: 10.3389/fmolb.2021.735263

INTRODUCTION

Parkinson's disease (PD) is the most common neurodegenerative motor disorder. It develops as a result of the premature death of dopamine-containing neurons in a part of the midbrain called the substantia nigra. This leads to a loss of dopaminergic neurons within the substantia nigra pars compacta, depletion of dopamine in the striatum, and the presence of Lewy bodies (Jankovic, 2008). In contrast to the excessive neuronal cell death observed in PD, cancer develops from unrestricted cell proliferation and resistance to cell death (Filippou and Outeiro, 2021). Interestingly, with the development of a more comprehensive understanding of both diseases, an intimate link between PD

and lung cancer has been gradually revealed. Most epidemiological studies and meta-analyses have reported a lower incidence of lung cancer in PD patients compared to that in the general population (Catalá-López et al., 2014; Ong et al., 2014; Peretz et al., 2016), highlighting a significant overlap between genes upregulated in PD and downregulated in lung cancer or vice versa (Ibáñez et al., 2014). The intriguing overlap of genes implicated in these two completely different diseases suggests that studying these genes may help improve lung cancer diagnosis, prognosis, and treatment.

Although there is a limited number of studies on Parkinson family genes in cancer, insightful findings have been reported in recent years. Mitophagy, a selective form of autophagy, is the major pathway for the degradation of dysfunctional or superfluous mitochondria in eukaryotic cells (Georgakopoulos et al., 2017), playing a central role in mitochondrial quality control and protection against damaged mitochondria (Tatsuta and Langer, 2008; Youle and Narendra, 2011). The Parkinson family genes PARK2 (PRKN) and PARK6 (PINK1) are considered the main regulators of mitophagy (Yan et al., 2020; Xie et al., 2021). The dysregulation of PRKN- and PINK1-mediated mitophagy has thus been suggested as one of the possible mechanisms underlying the pathogenesis of PD (Lin and Beal, 2006). This notion has been preliminarily validated in mouse models (Lu et al., 2014). Interestingly, mitophagy also has a significant impact on the occurrence and development of tumours. A recent study of hepatocellular carcinoma (HCC) suggested that mitophagy triggered by the accumulation of PINK1 and PRKN translocation can promote the apoptosis of HCC cells, suppressing the growth of patient-derived tumour xenografts (Chen et al., 2019). Further, mitophagy can inhibit the growth of pancreatic tumours by attenuating mitochondrial iron accumulation, inflammasome activation, and other processes, with PINK1 and PRKN depletion confirmed to promote KRAS-driven pancreatic tumourigenesis in mouse models (Kang et al., 2019). PARK18 (EIF4G1), an important part of the EIF4F complex, is required for cap-dependent mRNA translation (Jaiswal et al., 2018). Inevitably, EIF4G1 is involved in various cancer-related processes, such as the activation of the mTOR signalling pathway and hypoxia-inducible factor-1α (HIF-1α)-related processes (Glück et al., 2018; Lu et al., 2021). Other family members such as PARK22 (CHCHD2), which plays an important role in the switch between catabolism and anabolism (Zacksenhaus et al., 2017), PARK7 (DJ-1), which is involved in ferroptosis (Cao et al., 2020), ubiquitination-related regulatory genes PARK15 (FBXO7) and PARK5 (UCHL1) (Goto et al., 2015; Teixeira et al., 2016; Liu et al., 2020), as well as PARK10 (USP24), which is related to cancer-associated acetylation (Wang et al., 2017). In general, the significance of Parkinson family genes in cancer remains to be further explored and could be of major relevance for our understanding of cancer progression.

We have focused on exploring the role of Parkinson family genes in the occurrence and development of cancer. Our previous studies showed that PARK6 (PINK1) can promote migration and proliferation of lung cancer cells by regulating autophagy (Lu et al., 2020), and PARK7 (DJ-1) is necessary for the transcription of HIF-1α and survival of colorectal cancer cells (Lin et al., 2018; Zheng et al., 2018). However, there has been no relevant research on the overall prognostic significance of Parkinson family genes in cancer. Herein, we provide a preliminary analysis of Parkinson family genes in the prognosis and treatment of lung adenocarcinoma (LUAD). The current findings will help us further understand the role of Parkinson family genes in cancer, aid in LUAD prognosis and treatment, and indicate possible directions for future research on this gene family.

MATERIALS AND METHODS

Differential Gene Expression Analysis

We downloaded the normalised gene expression data of cancer and normal tissues from the Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression Project (GTEx) databases in UCSC Xena (http://xena.ucsc.edu/). Data was graphically displayed using the ggplot2 R package. For differences in protein expression between cancer and normal tissues, we used the clinical proteomic tumor analysis consortium (CPTAC) analysis tool of the UALCAN database (http://ualcan.path.uab.edu/index.html), and no LUAD protein expression data were found for PARK6 (PINK1), PARK10 (ELAVL4), and PARK16 (SLC41A1). The expression of Parkinson family genes at different stages of LUAD was explored using the GEPIA2 database (http://gepia2.cancer-pku.cn/#index).

Survival Analysis

We obtained and downloaded the gene expression and detailed pathological data of 436 LUAD patients and 425 LUSC patients (primary solid tumour samples with detailed prognostic information and a survival time of up to 15 years were included) from the SangerBox database (http://www.sangerbox.com/) to analyse the impact of Parkinson family genes on cancer prognosis. Data was graphically displayed with the help of the survival R package. The forest plot was constructed using GraphPad Prism 8.

Prognostic Model Construction and Verification

Based on the expression and prognostic value of 25 Parkinson family genes in 436 LUAD and 425 LUSC patient samples, we used the survival R package to construct a Least absolute shrinkage and selection operator (LASSO) regression model. Risk factor analysis was performed using the Hiplot online analysis platform (https://hiplot.com.cn/). A prognostic nomogram model for the 436 LUAD patients was constructed using the rms R package, and the bootstrap method was applied to assess consistency. To validate the model in external data sets, we selected three Gene Expression Omnibus (GEO) chips (GSE37745, GSE31210, and GSE30219) of LUAD patients with detailed prognostic information. The receiver operating characteristic (ROC) curve and time-dependent ROC curve of the working characteristics of subjects were established using R in order to evaluate the survival prediction accuracy of the seven-gene signature for the three chips.

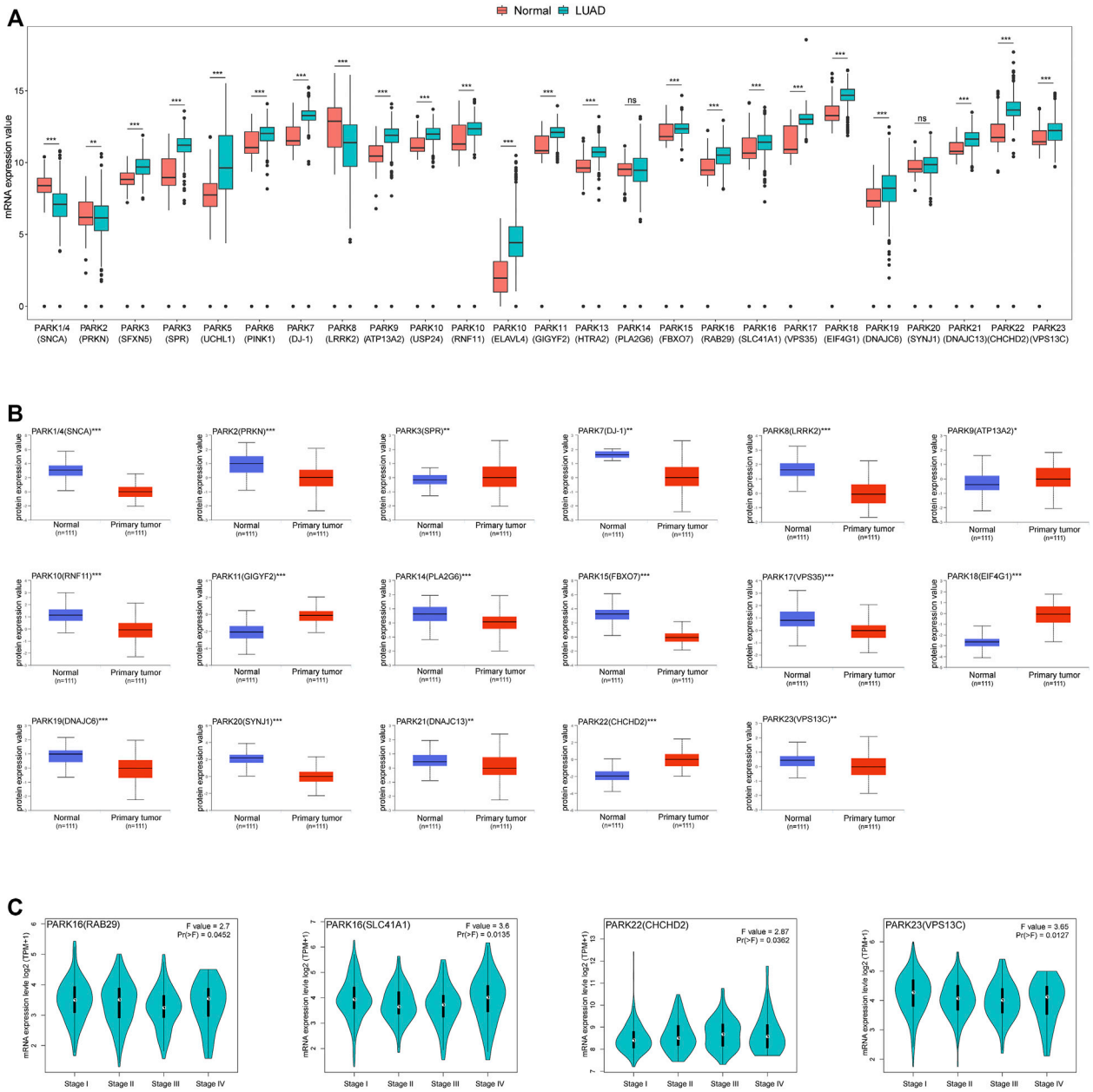


FIGURE 1 | Expression of Parkinson family genes in LUAD. **(A)** Differences in mRNA expression of 25 Parkinson family genes in LUAD and normal tissues from TCGA and GTEx databases (normal = 397, LUAD = 513). **(B)** Differences in the protein expression levels of 17 Parkinson family genes between LUAD and normal tissues in the UALCAN database. **(C)** Expression differences for four Parkinson family genes at different disease stages in the GEPIA2 database. *, **, and *** represent $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Bioinformatics Analysis

Univariate and multivariate analyses were conducted using the IBM SPSS Statistics 19. The tumour mutation burden (TMB) and neoantigen (NEO) data of 436 LUAD patients were obtained from the Cancer Imaging Archive (TCIA) database (https://tcia.at/home) (Van Allen et al., 2015; Hugo et al., 2016), and both data were available only for 363 patients. We explored mutations of the seven genes screened via LASSO regression (TCGA, PanCancer Atlas) using the cBioportal database (https://www.cbioportal.org/). Employing the HitPredict database (http://www.hitpredict.org/) (Patil and Nakamura, 2005; Patil et al., 2011; López et al., 2015), we searched for interaction partners of the seven genes. The protein network interaction map was constructed using Cytoscape_v3.7.1. GO and KEGG pathway enrichment analysis of the seven genes and Gene Set Enrichment Analysis (GSEA) analysis of high- and low-risk LUAD patients were carried out using the OmicShare 6.2.1 online tool.

Immune Cell Infiltration Analysis

In the TIMER database (http://cistrome.dfci.harvard.edu/TIMER/), we explored the impact of mutations in the seven genes on the degree of infiltration for six immune cell types within the tumour microenvironment (TME). Based on the ESTIMATE database (https://bioinformatics.mdanderson.org/estimate/), we further analysed the immune scores of 436 LUAD patients in different risk groups. Next, we used seven analysis methods (TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC) to determine the immune cell infiltration status within the TME of patients in the TIMER2 database (http://timer.cistrome.org/). In addition, we explored differences between the high- and low-risk groups in the steps of the cancer immunity cycle for these patients using the TIP database (http://biocc.hrbmu.edu.cn/TIP/) (Xu et al., 2018).

Analysis of the Efficacy and Response to Immunotherapy and Targeted Therapy

Gene expression and detailed pathological data of anti-PD-1-treated LUAD patients were obtained from the GSE135222 dataset. The correlation between risk scores and various immunosuppressive molecules was graphically displayed utilising the online analysis tool Hiplot. R packages were used to analyse the correlation between risk score and multiple treatment targets.

RESULTS

Expression of Parkinson Family Genes in LUAD

In order to explore the expression of Parkinson family genes in LUAD, we obtained the normalised cancer and normal tissue gene expression data of LUAD patient samples from the TCGA and GTEx databases in UCSC Xena. All Parkinson family genes exhibited significant differential expression, with the exception of PARK14 (PLA2G6) and PARK20 (SYNJ1) (**Figure 1A**). Similar results were obtained for LUSC (**Supplementary Figure S1**). Furthermore, using the UALCAN database, we analysed the protein expression of Parkinson family genes between LUAD and normal tissue. Seventeen genes exhibited significant differences in expression at the protein level (**Figure 1B**). Finally, we explored the expression of Parkinson family genes at different TNM stages in the GEPIA2 database and found that four genes had significant expression differences during the progression of LUAD (**Figure 1C**). In summary, all Parkinson family genes exhibited significant differential expression in LUAD, suggestive of their involvement in the occurrence and development of lung cancer.

Prognostic Value of Parkinson Family Genes in LUAD

Next, we selected 436 LUAD patients from TCGA to investigate the impact of Parkinson family genes on the prognosis of LUAD. We found that 12 genes had significant differences based on LUAD prognosis (**Figures 2A,B**). Similarly, 11 genes had a significant impact on the prognosis of LUSC patients (**Supplementary Figures S2A,B**). Moreover, we performed GSEA analysis to explore the influence of deregulation of these genes on 50 hallmark gene sets. Results showed increased expression for these genes associated with poor prognosis can obviously activate a variety of cancer-related pathways (WNT, P53 and NOTCH pathway for LUAD; MYC, G2M checkpoint and Oxidative phosphorylation for LUSC). Interesting, whether in LUAD or LUSC, the impact of the increase of protective prognostic genes and risk genes on the 50 hallmark gene sets were obviously different and even opposed to each other, this finding may provide some help for follow-up in-depth study (**Supplementary Figure S3, S4**).

However, not all differentially expressed Parkinson genes (in **Figure 1A**) accurately predicted overall survival in LUAD patients. Therefore, we sought to create a gene signature from 25 Parkinson family genes that had the most pronounced impact on LUAD prognosis. To this end, we constructed a LASSO regression model based on the expression and prognosis data of 436 LUAD patients. We obtained a seven-gene prognostic signature (**Figures 2C,D**). The complete names of the seven genes and their main functions are listed in **Table 1**. Meanwhile, we further analyzed the correlation between the 7 genes and genes in the non-small cell lung cancer pathway (map05223) from KEGG database. The results showed that the 7 genes were related to multiple cancer progression marker genes respectively, this result suggest these genes were involved in the development of lung adenocarcinoma via different mechanisms (**Supplementary Figure S5**). We also constructed a LASSO regression model based on the data of LUSC patients. However, limited by data quality, the results indicated that a combination of two genes had the best performance (**Supplementary Figures S2C,D**). Therefore, we focused on the seven-gene signature for the prognosis and treatment outcome prediction in LUAD.

Our Seven-Gene Signature can be Used as an Independent Prognostic Indicator for LUAD

To better assess the prognostic value of our seven-gene signature in LUAD, we first conducted risk factor analysis based on the genes. The occurrence of death was significantly correlated with a higher risk factor score (**Figure 3A**). Furthermore, we divided 436 patients into high- and low-risk groups based on risk factor score and analysed prognosis in the two groups. The high-risk group exhibited a significantly poorer prognosis among patients with different TNM stages (**Figure 3B**). Univariate and multivariate analyses also indicated that the signature-based risk factor score was a superior prognostic indicator compared to TNM staging (**Figures 3C,D**). Taken together, the seven-gene signature of Parkinson family genes can be used as an independent prognostic marker for LUAD.

Validation of the Prognostic Value of Our Seven-Gene Signature for LUAD

In order to validate and further assess the prognostic value of our seven-gene signature for LUAD, we constructed a nomogram based on data from the TCGA database in order to predict patient

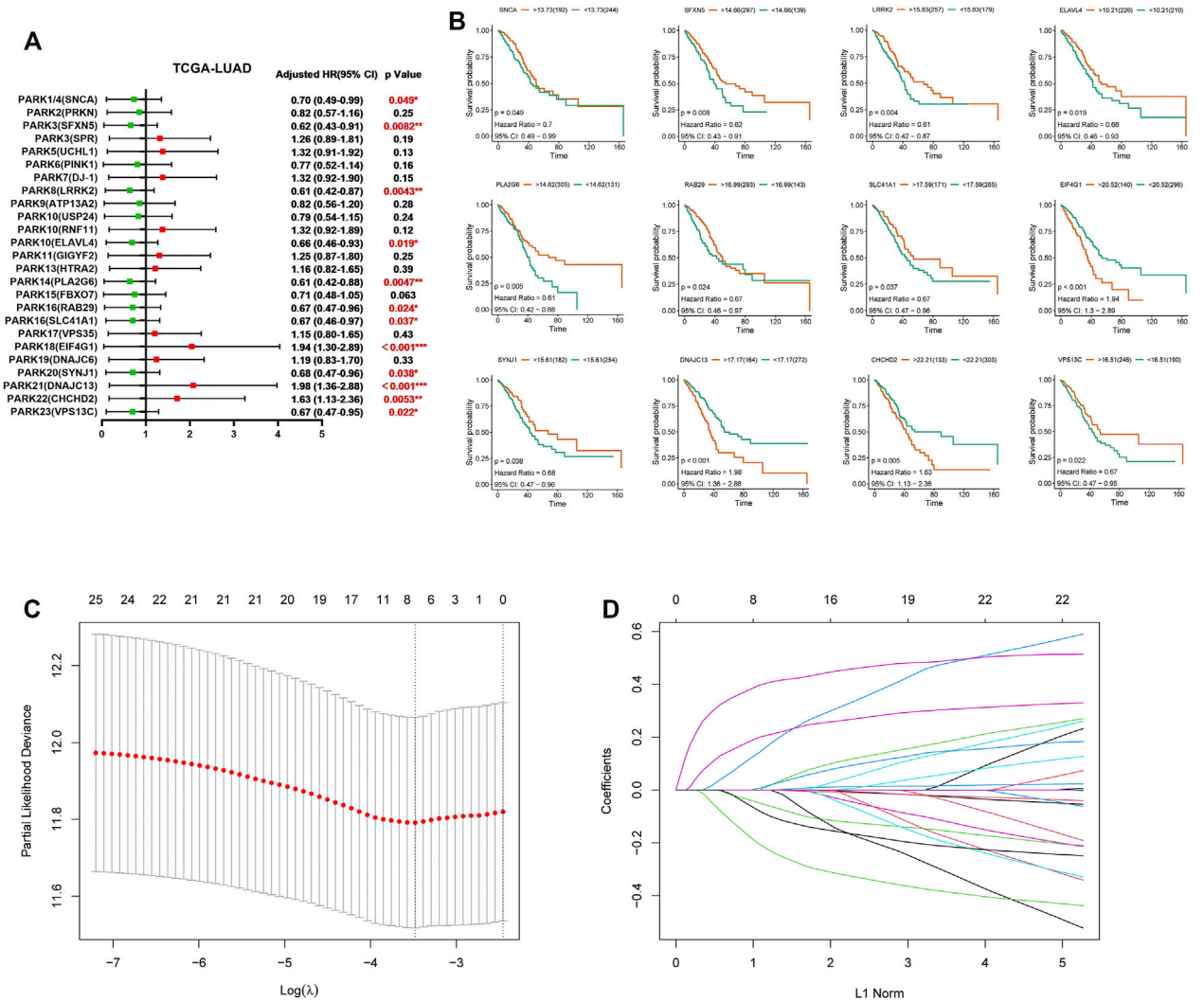


FIGURE 2 | The clinical significance of Parkinson family genes in the prognosis of LUAD and the LASSO regression model. **(A)** LUAD prognostic forest plot of 25 Parkinson family genes from the TCGA database. **(B)** 12 Parkinson family genes with significant differences based on the prognosis of 436 LUAD patients. **(C,D)** LASSO regression prognostic model for the 436 LUAD patients.

TABLE 1 | LASSO regression result of Parkinson's disease gene family in TCGA-LUAD dataset.

Gene symbol	Full name	Function
PARK3(SFXN5)	sideroflexin-5	mitochondrial amino-acid transporter
PARK14(PLA2G6)	calcium-independent phospholipase A2	involved in cellular membrane homeostasis, mitochondrial integrity and signal transduction
PARK16(RAB29)	ras-related protein Rab-7L1	Involved in vesicle trafficking
PARK16(SLC41A1)	solute carrier family 41 member 1	a predominant Mg2+ efflux system at the plasma membrane
PARK18(EIF4G1)	eukaryotic translation initiation factor 4 gamma 1	component of the protein complex eIF4F
PARK21(DNAJC13)	dnaJ homolog subfamily C member 13	involved in membrane trafficking through early endosomes
PARK22(CHCHD2)	coiled-coil-helix-coiled-coil-helix domain-containing protein 2	transcription factor

survival probability by weighing age, gender, stage, T, N, M, and the signature-based risk score (**Figure 4A**). Further, we applied the bootstrap method to evaluate the nomogram's predictive performance. The calibration curves indicated that the nomogram-predicted probability matched the actual 3- and 5-years survival (**Figures 4B,C**). Subsequently, we selected three

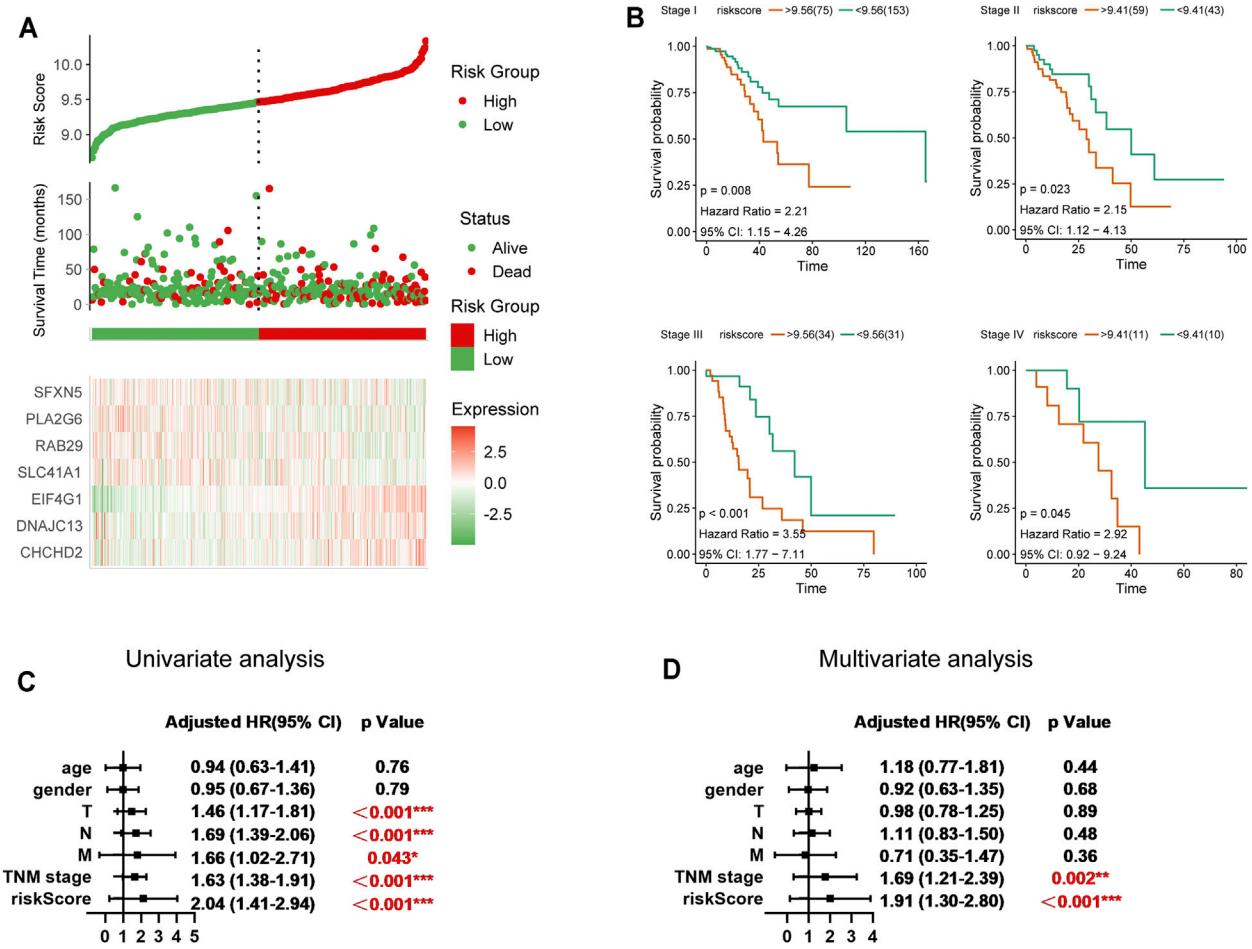


FIGURE 3 | Significance of the seven-gene prognostic model in LUAD. **(A)** Risk factor analysis for the 436 LUAD patients (cut-off value = 9.48). **(B)** Prognostic differences between high- and low-risk groups in different TNM stages in LUAD. **(C,D)** Univariate and multivariate analysis of the seven-gene prognostic risk model.

GEO datasets (GSE37745, GSE31210, and GSE30219) with detailed prognostic information for external testing (**Table 2**). The seven-gene signature risk scores suggested a significantly better diagnostic performance than TNM staging based on the ROC curve, also showing excellent prognostic performance at different survival time points in the time-dependent ROC curve (**Figures 4D–F**). Taken together, risk evaluation based on the seven-gene signature has important clinical significance in the diagnosis and treatment of LUAD.

Functional Enrichment Analysis of the Seven Signature Genes

We explored the possible mechanisms underlying the functional association of signature genes with poor prognosis. Utilising the HitPredict database, we searched for interaction partners of their protein products, which are displayed in the protein interaction network diagram (**Figure 5A**). We then performed GO and KEGG enrichment analyses for these genes. GO enrichment analysis indicated the genes' involvement in diverse tumourigenesis- and development-associated signalling, such as metabolism-related pathways, ubiquitination pathways, and RNA transcription-related processes (**Figures 5B–D**). Furthermore, KEGG pathway enrichment analysis also revealed that immune and cell proliferation-related pathways (RIG-I-like receptor, MAPK, mTOR, and ErBb) were enriched by these genes (**Figure 5E**).

Signature Gene Expression Affects the TME in LUAD

In view of the fact that various immune-related processes were also observed among functional enrichment results (**Figure 5E**), we sought to explore the influence of the seven signature genes on the TME. Employing the TIMER database, we obtained the effect of various single gene mutations on the infiltration of six immune cell types (B cells, CD8+T cells, CD4+T cells, macrophages, neutrophils, and dendritic cells) into the TME. The results indicated that mutations of signature genes decreased the degree of infiltration for almost all six immune cell types (**Figures 6A–G**). We also explored the link between risk factor score and TMB (or NEO) based on TCIA data. A high-risk score

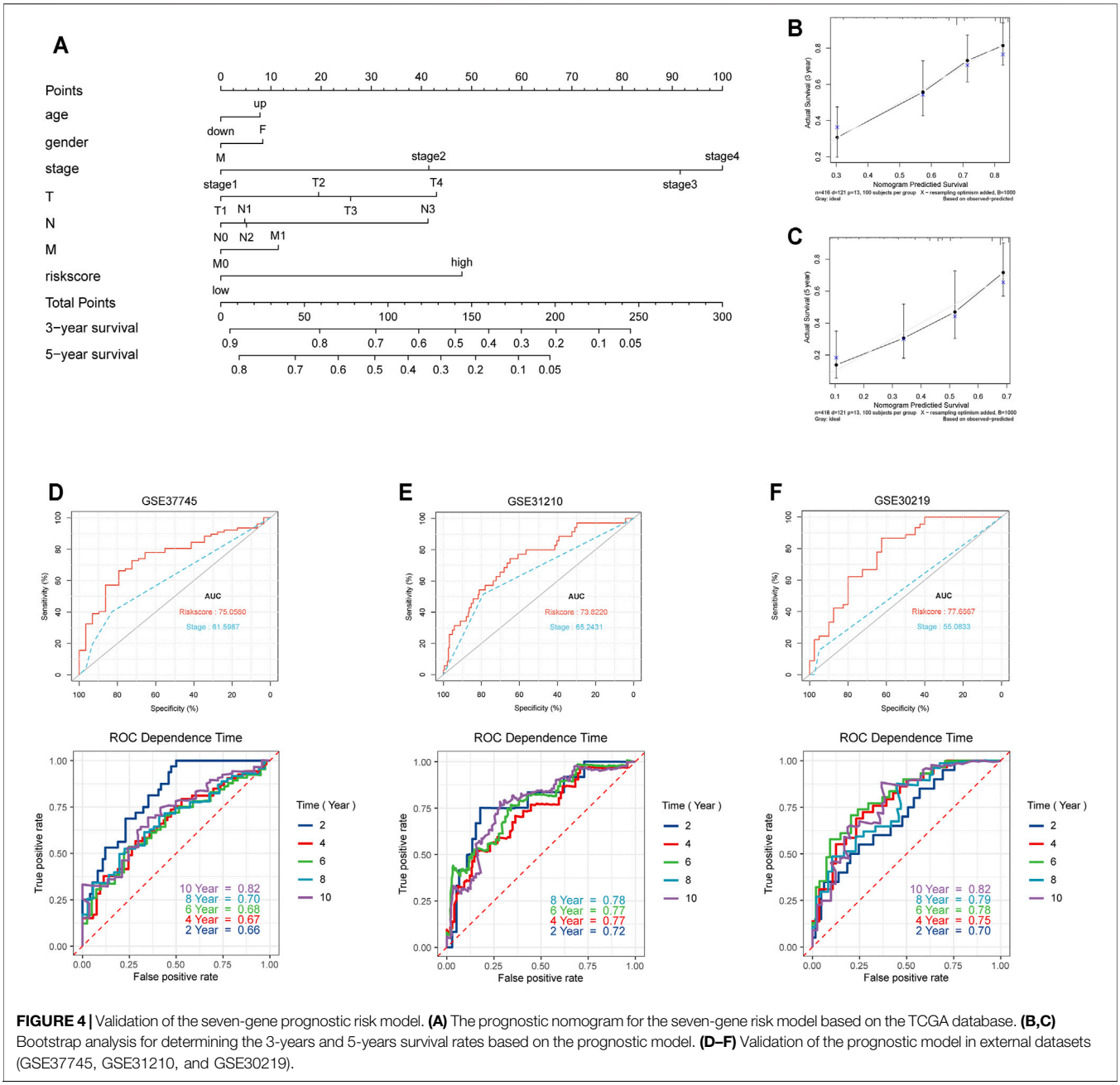


FIGURE 4 | Validation of the seven-gene prognostic risk model. **(A)** The prognostic nomogram for the seven-gene risk model based on the TCGA database. **(B,C)** Bootstrap analysis for determining the 3-years and 5-years survival rates based on the prognostic model. **(D–F)** Validation of the prognostic model in external datasets (GSE37745, GSE31210, and GSE30219).

was clearly related to a high TMB and high NEO (**Figure 6H**). Mutation data from the cBioportal database indicated that mutations in these seven genes were rare among TCGA-LUAD patients (**Figure 6I**).

We further explored the effect of the seven-gene signature on the TME. Estimation analysis indicated that the immune score of the high-risk group was significantly lower than that of the low-risk group for the 436 LUAD patients (**Figure 7A**). We further observed the different infiltration levels of 22 immune cell types between the low- and high-risk groups using CIBERSORT (**Figure 7B**). Meanwhile, in order to eliminate possible errors caused by different analysis methods, we employed another six methods (**Supplementary Figure S6**) and summarised the results, revealing the same trend and significant differences (**Table 3**). However, these results revealed an interesting phenomenon. In the high-risk group, the infiltration of DC cells, B cells, CD4+ T cells, and CD8+ T cells remained lower, while M1 pro-inflammatory macrophages were upregulated, and M2 anti-inflammatory macrophages and Treg cells were downregulated.

The cancer immunity cycle reflects a complex interaction network involving the chemokine system, other immunomodulators, and the various immune cell types. Based on TIP data, we found that high risk was associated with the

TABLE 2 | GEO dataset for the prognostic value of the 7 genes in LUAD.

Features	GSE37745	GSE31210	GSE30219
Platforms	GPL570	GPL570	GPL570
Samples			
adeno	106	226	85
else	90	20	222
Age			
<60	64	105	118
≥60	132	136	174
Gender			
female	89	130	43
male	107	116	250
Stage			
I	130	168	170
II	35	58	59
III	27	0	55
IV	4	0	8

release of cancer cell antigens (Step 1), reduced priming and activation (Step 3), recruitment of multiple immune cells (Step 4), and the infiltration of immune cells into tumours (Step 5) (**Figure 7C**). It is worth noting that Th1 cells were significantly downregulated in the high-risk group, while Th2 cells were upregulated, although the difference was not statistically significant. This partly explains why the various effector cells of cellular immunity exhibited low infiltration with an opposite trend of change compared to regulatory cells. The detailed underlying mechanism remains to be further explored. Similarly, GSEA analysis showed that signature-based high risk was associated with a significant inhibition of tumour immune-related processes (**Figure 7D**). Exploring the mechanism underlying the increase in M2 cells and Treg cells in the high-risk group may help explain the low infiltration of effector cells.

Utility of the Seven-Gene Signature for the Prediction of Immunotherapy and Targeted Therapy Response in LUAD

Finally, we sought to explore the utility of our seven-gene signature as a predictor of immunotherapy and targeted therapy response in LUAD. A high-risk score was significantly negatively correlated with a variety of immunosuppressive molecules (**Figure 8A**). Furthermore, we explored the correlation between risk score and the efficacy of PD-L1 immunotherapy in LUAD. Data from GSE135222 indicated that after PD-1 immune checkpoint inhibitor treatment, the risk score of dying patients was higher than that of surviving patients, and there was a negative correlation between the survival time for surviving patients and the risk score (**Figure 8B**). Limited by the sample size, the results were not statistically significant, and we did not find other available immunotherapy cohorts with detailed prognostic and gene expression data. Nevertheless, we believe that these results would be useful for treatment selection. We analysed the correlation between risk score and the expression of multiple driver genes for targeted therapy in TCGA patients, observing a clear correlation (**Figure 8C**). Furthermore, based on the Genomics of Drug Sensitivity in Cancer (GDSC) database, we predicted the impact of 7 genes on the half-inhibitory concentration (IC50) of some common chemotherapeutics (platinum, paclitaxel, etc) and targeted agents for non-small cell lung cancer (NSCLC). Interestingly, the result showed high expression of SLC41A1, RAB29 and PLA2G6 may be positively correlated with resistance of various common chemotherapeutics, whereas high level of CHCHD2, EIF4G1, DNAJC13 and SFXN5 may be involved in resistance of target therapy for drive gene (**Supplementary Figure S7**). However, we did not find available targeted therapy datasets for validation. In general, our findings indicated that targeted therapy may be more appropriate than immunotherapy for high-risk patients as a precision medicine approach for LUAD.

DISCUSSION

Lung cancer remains the major contributor to the cancer disease burden worldwide, ranking first in cancer-associated mortality. The biggest reason for this is that lung cancer patients have already entered the middle or late stages of disease upon diagnosis. Advanced disease has a strong tendency for metastasis and relapse. Although progress has been made in diagnosis and treatment strategies over the past decades, patient prognosis remains very poor, with a 5-years survival rate of only about 20% (Siegel et al., 2021). As LUAD is the major histologic subtype of lung cancer, it is necessary to explore more effective and sensitive biomarkers for its prognosis and treatment.

Instead of basing LUAD prognosis on a single gene, we screened 25 Parkinson family genes and established a seven-gene prognostic signature (PARK3 [SFXN5], PARK14 [PLA2G6], PARK16 [RAB29], PARK16 [SLC41A1], PARK18 [EIF4G1], PARK21 [DNAJC13], PARK22 [CHCHD2]) for LUAD. Whether in the TCGA-LUAD cohort or the three GEO cohorts for external verification, the seven-gene signature exhibited excellent performance for LUAD prognosis. Although previous studies have shown that the signature genes were related to a variety of cancers (liver cancer, breast cancer, oral cancer, and others), there have been few studies on the mechanism through which they participate in cancer progression (Li et al., 2017; Khowal et al., 2018; Uddin et al., 2018; Yao et al., 2019). Migration, proliferation, energy metabolism, and autophagy are all considered regulatory targets, but the underlying mechanism remains to be further elucidated (Hosgood et al., 2008; Wei et al., 2015; Jaiswal et al., 2018; Ma et al., 2020; Wang et al., 2021). In addition, HIF-1α, a key regulator of the tumour hypoxic response, may be implicated in the signature genes' association with poor prognosis. Studies have shown that in non-small cell lung cancer (NSCLC), CHCHD2 and HIF-1α co-localise in both the cytoplasm and the nucleus, with CHCHD2 overexpression significantly promoting that

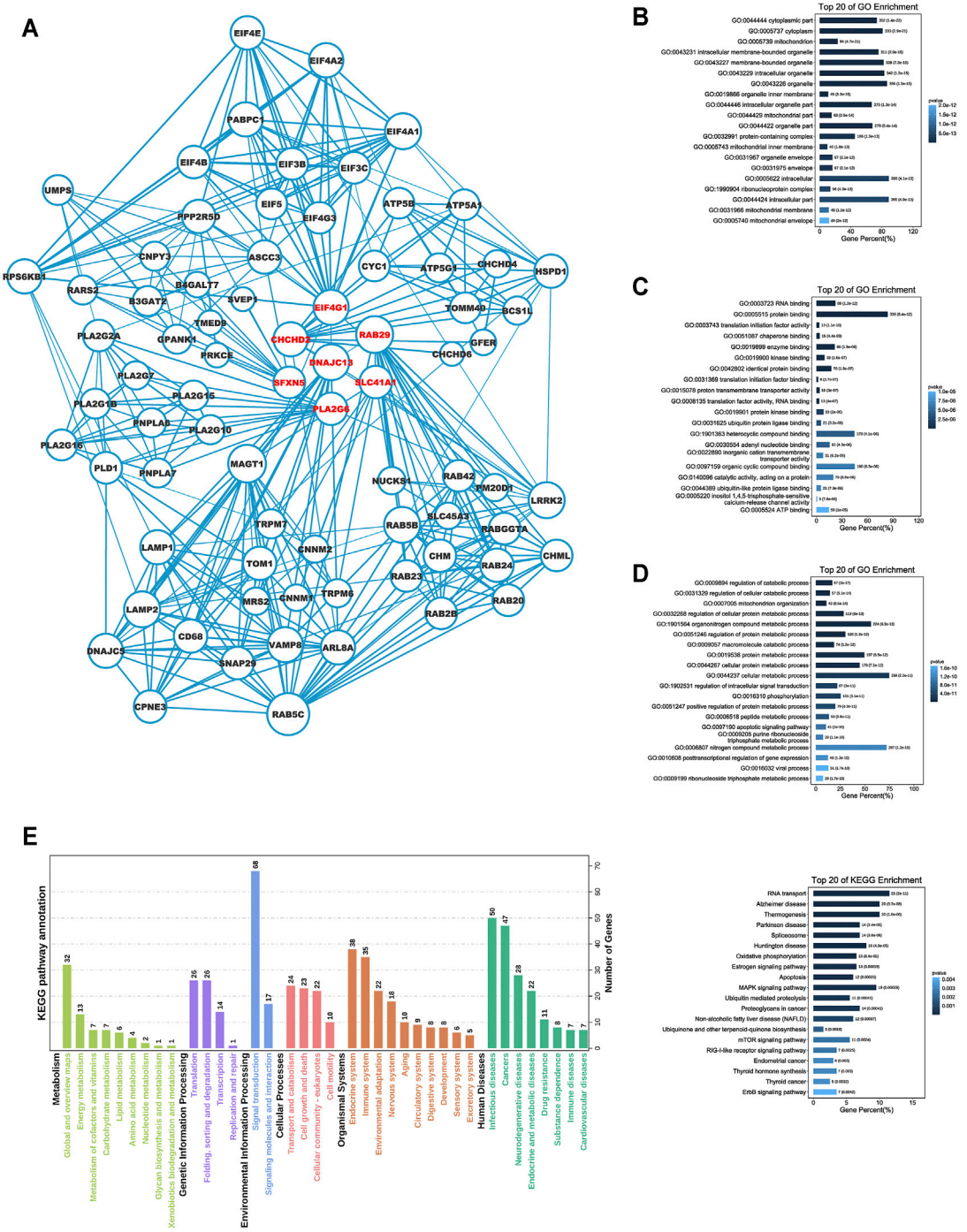


FIGURE 5 | Functional enrichment analysis for the seven Parkinson family genes. **(A)** The protein network interaction map for seven Parkinson family genes based on the HitPredict database. **(B–D)** GO enrichment analysis of the seven Parkinson family genes. **(E)** KEGG pathway enrichment analysis of the seven Parkinson family genes.

of HIF-1α (Yin et al., 2020). EIF4G1 was also confirmed as involved in HIF-1α overexpression in NSCLC (Glück et al., 2018). Furthermore, studies have suggested that EIF4G1 can also promote NSCLC progression by regulating the expression and phosphorylation of mTOR (Ser 2448) (Lu et al., 2021). In general, the current understanding of the signature genes' functions in cancer is relatively limited, necessitating further investigation.

As the seven genes belong to the Parkinson family, understanding how they contribute to PD may be conducive

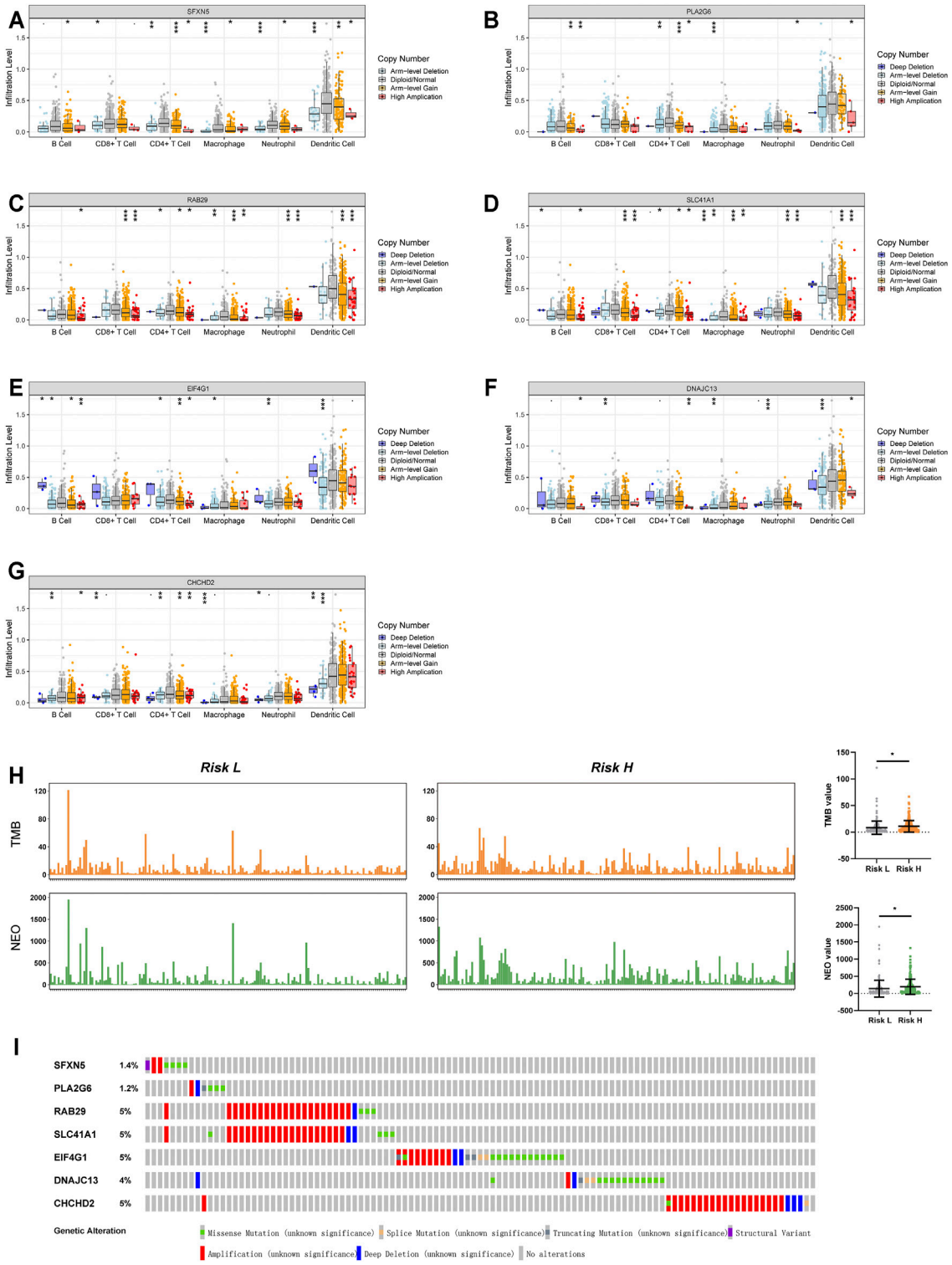


FIGURE 6 | The effect of mutations in the seven Parkinson family genes on immune cell infiltration within the TME. **(A–G)** The effect of mutations in SFXN5, PLA2G6, RAB29, SLC41A1, EIF4G1, DNAJC13, and CHCHD2 on the infiltration of six immune cell types (B cell, CD8+T cell, CD4+T cell, macrophage, neutrophil, dendritic cell) in the TME based on the TIMER database. **(H)** Mutations of TMB and NEO in high- and low-risk groups in the TCGA database. **(I)** Mutation status of the seven Parkinson family genes in LUAD data from TCGA.

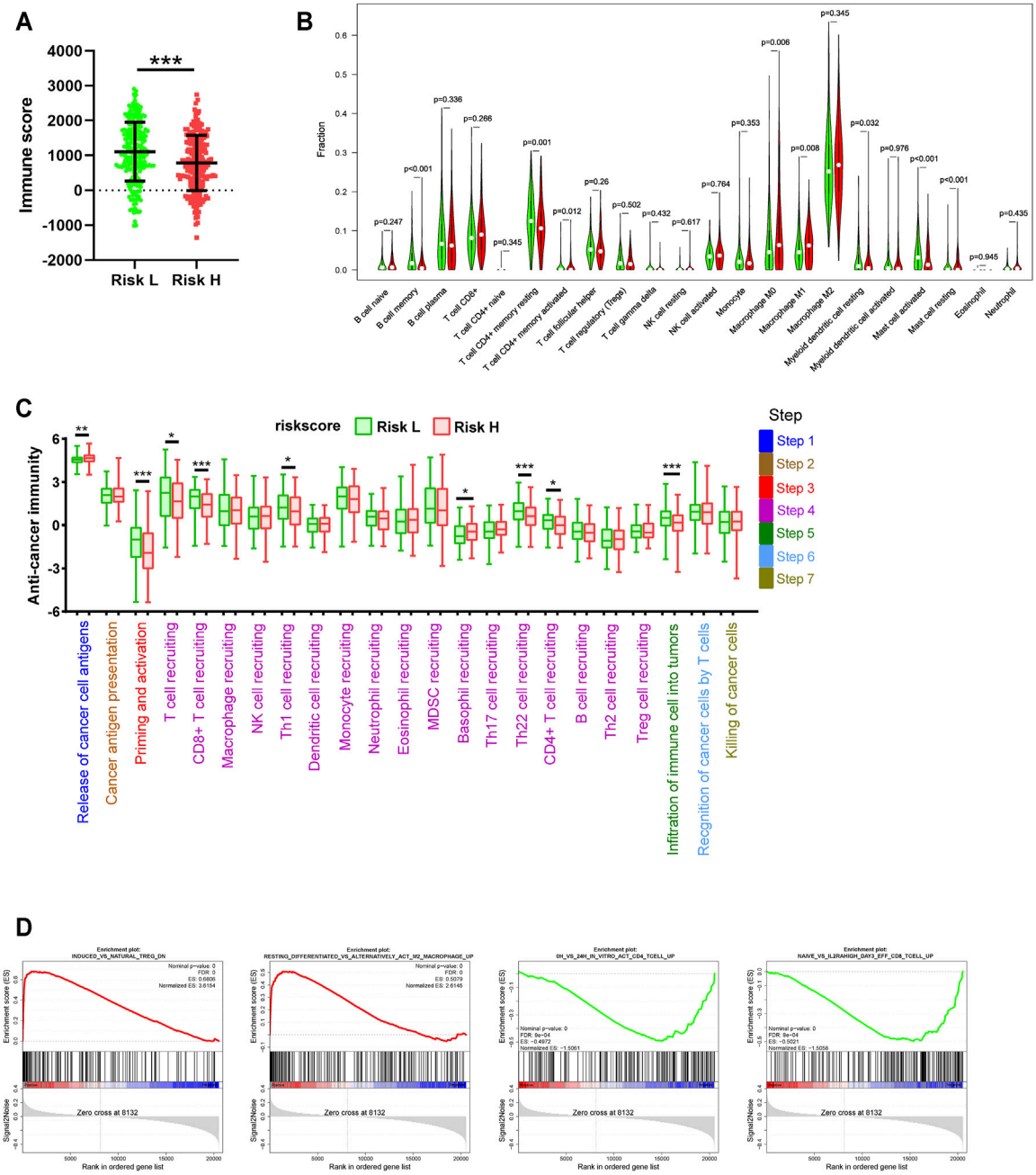


FIGURE 7 | Differences in tumor immune-related processes between high- and low-risk groups. **(A)** Estimation analysis of the immune scores for high- and low-risk groups. **(B)** Differences in the infiltration status of 22 immune cell types between high- and low-risk groups via the CIBERSORT method. **(C)** Differences in the cancer immunity cycle between high- and low-risk groups. **(D)** GSEA analysis for high- and low-risk groups based on TCGA data.

for future studies in cancer. Unfortunately, the research on SFXN5, SCL41A1, EIF4G1, and DNAJC13 in PD is limited (Deng et al., 2015; Wang et al., 2015; Puschmann, 2017), and their specific mechanisms in PD pathogenesis remain unclear. Nevertheless, some studies have provided valuable insights. PLA2G6 is believed to induce PD mainly by disturbing lipid metabolism in neurons. When lipid metabolism is dysregulated, the transient interaction between α-synuclein and the synaptic vesicle membrane composed of phospholipids and other lipids was affected, resulting in α-synuclein aggregation and the resultant pathological α-synuclein conversion (Mori et al., 2020). At the same time, oxidative stress and inflammation caused by an imbalance of lipid metabolism also contribute to the occurrence and development of PD (Alecu and Bennett, 2019). In addition, RAB29 is believed to cause PD via

TABLE 3 | The correlation between the risk score and immune cell infiltration level.

	TIMER	CIBERSORT.ABS	QUANTISEQ	MCPCOUNTER	XCELL	EPIC
B_cell	−0.29***	Naïve: −0.06, Memory: −0.23***, Plasma: −0.14**	−0.25***	−0.24***	Naïve: −0.12*, Memory: −0.22***, Plasma: −0.14**	−0.23***
CD4+T_cell	−0.27***	Naïve: −0.0018, Resting_memory: −0.27***, Activatied_memory: 0.11*	0.052	Null	Naïve: −0.25***, Non_regulatory: −0.14**, Central_memory: −0.27***, Effector_memory: −0.21***	−0.075
CD8+T_cell	−0.041	-0.070	−0.057	−0.064	Naive: 0.0055, Centra memory: −0.16***, Effector memory: −0.052	−0.15**
Macrophage	−0.14**	M0: 0.077, M1: 0.027*, M2: −0.17***	M1: 0.080, M2: −0.31***	0.064	M1: 0.020*, M2: −0.26***	−0.15**
Dendritic_cell	−0.12*	Resting: −0.18***, Activated: −0.036	0.12**	−0.26***	Activated: -0.18***, Plasmacytoid: −0.051	Null
Tregs_cell	Null	−0.15**	−0.22***	Null	−0.037	Null
Mast_cell	Null	Resting: 0.14** Activated: −0.31***	Null	Null	-0.17***	Null
Eosinophil	Null	−0.069	Null	Null	−0.29***	Null
Endothelial_cell	Null	Null	Null	−0.20***	−0.16**	−0.18***

*, **, and *** represent $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively.*

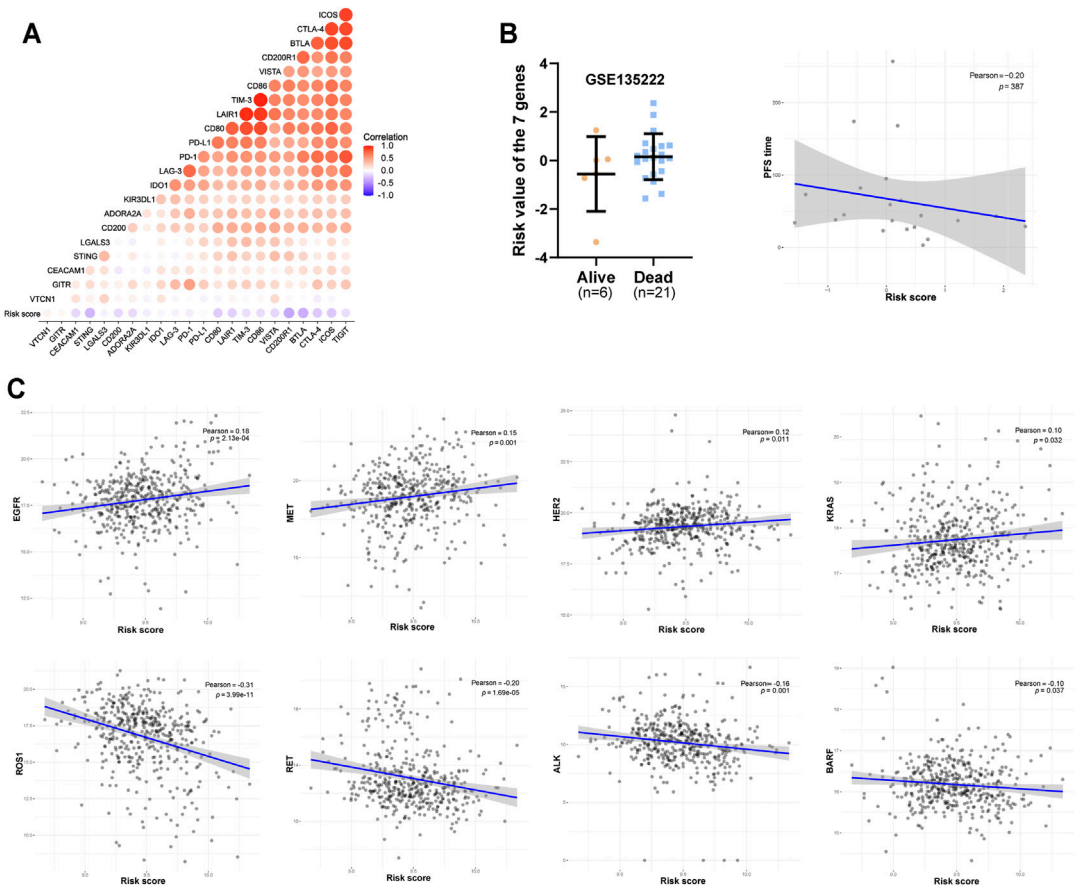


FIGURE 8 | Predictive value of the seven-gene signature for LUAD immunotherapy and targeted therapy outcomes. **(A)** Heat map of the correlation between risk score and the expression of multiple immunosuppressive regulatory molecules. **(B)** The correlation between the risk score and the efficacy of PD-1 immunotherapy based on the treatment data of GSE135222. **(C)** Correlation between risk score and treatment outcome of LUAD patients.

lysosomal dysfunction, while CHCHD2 may cause PD by impairing mitochondrial function (Zhou et al., 2019; Mazza et al., 2021). These processes are also of major relevance in tumours, providing an avenue for their further investigation in cancer.

Finally, we analysed the predictive merit of our seven-gene signature for LUAD treatment outcome. We observed that a high risk based on the signature may be related to low immune cell infiltration in LUAD, which was not underpinned by the polarisation of M2 macrophages and

Treg cell infiltration. We further found that PD-1 immunotherapy may not be optimal for high-risk patients, as a high risk was related to low immune scores, suggesting targeted therapy as a more suitable option. Recent research has shown that tumour-associated macrophages (TAMs) are critical mediators of the PD-1/PD-L1 axis, and the high infiltration of M2 macrophages is significantly correlated with high PD-L1 expression within the TME (Zhu et al., 2020). Furthermore, SPP1 can promote the polarisation of M2 macrophages, and knockdown of SPP1 significantly suppressed PD-L1 expression on LUAD cells (Zhang et al., 2017). Other specific immunoregulatory mechanisms of signature genes required further investigation. Nevertheless, we consider the current findings to be of considerable value for the treatment of high-risk LUAD patients.

In summary, our study is the first to explore the role of Parkinson family genes in the prognosis and treatment of LUAD. Our research showed that the combination of seven Parkinson family genes may help predict LUAD prognosis as well as the treatment outcome for high-risk patients. Our study provides novel insight into the significance of Parkinson family genes in LUAD. We hope that the current findings will be of value for the clinical diagnosis and treatment of LUAD, and, more importantly, will further the exploration of Parkinson family genes in cancer. However, our results need more support and validation from clinical and animal models, continue to explore the unknown function and specific mechanisms of Parkinson family genes in tumors is a potential direction.

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

JD and HZ contributed substantially to the conception and design, and gave final approval of the version to be published. YL, QL, XL, JZ, ML, and GY contributed to the analysis and interpretation of the data. YL, HZ, and JD drafted the manuscript. XD, QC, DZ, and YQ critically revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the grants from the National Natural Science Foundation of China (81972190), Natural Science Foundation of Chongqing (cstc2018jcyjAX0205), and the Miaopu Talent Grant from Army Medical University (2019R057).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.735263/full#supplementary-material

REFERENCES

Alecu, I., and Bennett, S. A. L. (2019). Dysregulated Lipid Metabolism and its Role in α-Synucleinopathy in Parkinson's Disease. *Front. Neurosci.* 13, 328. doi:10.3389/fnins.2019.00328

Cao, J., Chen, X., Jiang, L., Lu, B., Yuan, M., Zhu, D., et al. (2020). DJ-1 Suppresses Ferroptosis through Preserving the Activity of S-Adenosyl Homocysteine Hydrolase. *Nat. Commun.* 11, 1251. doi:10.1038/s41467-020-15109-y

Catalá-López, F., Suárez-Pinilla, M., Suárez-Pinilla, P., Valderas, J. M., Gómez-Beneyto, M., Martinez, S., et al. (2014). Inverse and Direct Cancer Comorbidity in People with Central Nervous System Disorders: A Meta-Analysis of Cancer Incidence in 577,013 Participants of 50 Observational Studies. *Psychother Psychosom* 83, 89–105. doi:10.1159/000356498

Chen, Y., Chen, H.-N., Wang, K., Zhang, L., Huang, Z., Liu, J., et al. (2019). Ketoconazole Exacerbates Mitophagy to Induce Apoptosis by Downregulating Cyclooxygenase-2 in Hepatocellular Carcinoma. *J. Hepatol.* 70, 66–77. doi:10.1016/j.jhep.2018.09.022

Deng, H., Wu, Y., and Jankovic, J. (2015). TheEIF4G1gene and Parkinson's Disease. *Acta Neurol. Scand.* 132, 73–78. doi:10.1111/ane.12397

Filippou, P. S., and Outeiro, T. F. (2021). Cancer and Parkinson's Disease: Common Targets, Emerging Hopes. *Mov Disord.* 36, 340–346. doi:10.1002/mds.28425

Georgakopoulos, N. D., Wells, G., and Campanella, M. (2017). The Pharmacological Regulation of Cellular Mitophagy. *Nat. Chem. Biol.* 13, 136–146. doi:10.1038/nchembio.2287

Glück, A. A., Orlando, E., Leiser, D., Poliaková, M., Nisa, L., Quintin, A., et al. (2018). Identification of a MET-eIF4G1 Translational Regulation axis that Controls HIF-1α Levels under Hypoxia. *Oncogene* 37, 4181–4196. doi:10.1038/s41388-018-0256-6

Goto, Y., Zeng, L., Yeom, C. J., Zhu, Y., Morinibu, A., Shinomiya, K., et al. (2015). UCHL1 Provides Diagnostic and Antimetastatic Strategies Due to its Deubiquitinating Effect on HIF-1α. *Nat. Commun.* 6, 6153. doi:10.1038/ncomms7153

Hosgood, H. D., Menashe, I., Shen, M., Yeager, M., Yuenger, J., Rajaraman, P., et al. (2008). Pathway-based Evaluation of 380 Candidate Genes and Lung Cancer Susceptibility Suggests the Importance of the Cell Cycle Pathway. *Carcinogenesis* 29, 1938–1943. doi:10.1093/carcin/bgn178

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, 35–44. doi:10.1016/j.cell.2016.02.065

Ibáñez, K., Boullosa, C., Tabarés-Seisdedos, R., Baudot, A., and Valencia, A. (2014). Molecular Evidence for the Inverse Comorbidity between central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-Analyses. *Plos Genet.* 10, e1004173. doi:10.1371/journal.pgen.1004173

Jaiswal, P. K., Koul, S., Shanmugam, P. S. T., and Koul, H. K. (2018). Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1) Is Upregulated during Prostate Cancer Progression and Modulates Cell Growth and Metastasis. *Sci. Rep.* 8, 7459. doi:10.1038/s41598-018-25798-7

Jankovic, J. (2008). Parkinson's Disease: Clinical Features and Diagnosis. *J. Neurol. Neurosurg. Psychiatry* 79, 368–376. doi:10.1136/jnnp.2007.131045

Kang, R., Xie, Y., Zeh, H. J., Klionsky, D. J., and Tang, D. (2019). Mitochondrial Quality Control Mediated by PINK1 and PRKN: Links to Iron Metabolism and Tumor Immunity. *Autophagy* 15, 172–173. doi:10.1080/15548627.2018.1526611

Khowal, S., Naqvi, S. H., Monga, S., Jain, S. K., and Wajid, S. (2018). Assessment of Cellular and Serum Proteome from Tongue Squamous Cell Carcinoma Patient Lacking Addictive Proclivities for Tobacco, Betel Nut, and Alcohol: Case Study. *J. Cel. Biochem.* 119, 5186–5221. doi:10.1002/jcb.26554

Li, M., Li, C., Liu, W.-X., Liu, C., Cui, J., Li, Q., et al. (2017). Dysfunction of PLA2G6 and CYP2C44-Associated Network Signals Imminent Carcinogenesis from

Chronic Inflammation to Hepatocellular Carcinoma. *J. Mol. Cel Biol* 9, 489–503. doi:10.1093/jmcb/mjx021

Lin, M. T., and Beal, M. F. (2006). Mitochondrial Dysfunction and Oxidative Stress in Neurodegenerative Diseases. *Nature* 443, 787–795. doi:10.1038/nature05292

Lin, Y., Chen, Q., Liu, Q.-x., Zhou, D., Lu, X., Deng, X.-f., et al. (2018). High Expression of DJ-1 Promotes Growth and Invasion via the PTEN-AKT Pathway and Predicts a Poor Prognosis in Colorectal Cancer. *Cancer Med.* 7, 809–819. doi:10.1002/cam4.1325

Liu, S., González-Prieto, R., Zhang, M., Geurink, P. P., Kooij, R., Iyengar, P. V., et al. (2020). Deubiquitinase Activity Profiling Identifies UCHL1 as a Candidate Oncoprotein that Promotes TGFβ-Induced Breast Cancer Metastasis. *Clin. Cancer Res.* 26, 1460–1473. doi:10.1158/1078-0432.CCR-19-1373

López, Y., Nakai, K., and Patil, A. (2015). HitPredict Version 4: Comprehensive Reliability Scoring of Physical Protein-Protein Interactions from More Than 100 Species. *Database* 2015, bav117. doi:10.1093/database/bav117

Lu, W., Karuppagounder, S. S., Springer, D. A., Allen, M. D., Zheng, L., Chao, B., et al. (2014). Genetic Deficiency of the Mitochondrial Protein PGAM5 Causes a Parkinson's-like Movement Disorder. *Nat. Commun.* 5, 4930. doi:10.1038/ncomms5930

Lu, X., Liu, Q.-X., Zhang, J., Zhou, D., Yang, G.-X., Li, M.-Y., et al. (2020). PINK1 Overexpression Promotes Cell Migration and Proliferation via Regulation of Autophagy and Predicts a Poor Prognosis in Lung Cancer Cases. *Cancer. Manag. Res.* Vol. 12, 7703–7714. doi:10.2147/CMAR.S262466

Lu, Y., Yu, S., Wang, G., Ma, Z., Fu, X., Cao, Y., et al. (2021). Elevation of EIF4G1 Promotes Non-small Cell Lung Cancer Progression by Activating mTOR Signalling. *J. Cel Mol Med* 25, 2994–3005. doi:10.1111/jcmm.16340

Ma, L., Zheng, L. H., Zhang, D. G., and Fan, Z. M. (2020). CHCHD2 Decreases Docetaxel Sensitivity in Breast Cancer via Activating MMP2. *Eur. Rev. Med. Pharmacol. Sci.* 24, 6426–6433. doi:10.26355/eurrev_202006_21541

Mazza, M. C., Nguyen, V., Beilina, A., Karakoleva, E., Coyle, M., Ding, J., et al. (2021). Combined Knockout of Lrrk2 and Rab29 Does Not Result in Behavioral Abnormalities *In Vivo*. *J. Parkinson. Dis.* 11, 569–584. doi:10.3233/JPD-202172

Mori, A., Imai, Y., and Hattori, N. (2020). Lipids: Key Players that Modulate α-Synuclein Toxicity and Neurodegeneration in Parkinson's Disease. *Int. J. Mol. Sci.* 21, 3301. doi:10.3390/ijms21093301

Ong, E. L., Goldacre, R., and Goldacre, M. (2014). Differential Risks of Cancer Types in People with Parkinson's Disease: A National Record-Linkage Study. *Eur. J. Cancer* 50, 2456–2462. doi:10.1016/j.ejca.2014.06.018

Patil, A., Nakai, K., and Nakamura, H. (2011). HitPredict: a Database of Quality Assessed Protein-Protein Interactions in Nine Species. *Nucleic Acids Res.* 39, D744–D749. doi:10.1093/nar/gkq897

Patil, A., and Nakamura, H. (2005). Filtering High-Throughput Protein-Protein Interaction Data Using a Combination of Genomic Features. *Bmc Bioinformatics* 6, 100. doi:10.1186/1471-2105-6-100

Peretz, C., Gurel, R., Rozani, V., Gurevich, T., El-Ad, B., Tsamir, J., et al. (2016). Cancer Incidence Among Parkinson's Disease Patients in a 10-yrs Time-Window Around Disease Onset: A Large-Scale Cohort Study. *Parkinsonism Relat. Disord.* 28, 68–72. doi:10.1016/j.parkreldis.2016.04.028

Puschmann, A. (2017). New Genes Causing Hereditary Parkinson's Disease or Parkinsonism. *Curr. Neurol. Neurosci. Rep.* 17, 66. doi:10.1007/s11910-017-0780-8

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

Tatsuta, T., and Langer, T. (2008). Quality Control of Mitochondria: Protection against Neurodegeneration and Ageing. *Embo J.* 27, 306–314. doi:10.1038/sj.emboj.7601972

Teixeira, F. R., Randle, S. J., Patel, S. P., Mevissen, T. E. T., Zenkeviciute, G., Koide, T., et al. (2016). Gsk3β and Tomm20 Are Substrates of the SCFFbxo7/PARK15 Ubiquitin Ligase Associated with Parkinson's Disease. *Biochem. J.* 473, 3563–3580. doi:10.1042/BCJ20160387

Uddin, M., Balaravi Pillai, B., Tha, K., Ashaie, M., Karim, M., and Chowdhury, E. (2018). Carbonate Apatite Nanoparticles-Facilitated Intracellular Delivery of siRNA(s) Targeting Calcium Ion Channels Efficiently Kills Breast Cancer Cells. *Toxics* 6, 34. doi:10.3390/toxics6030034

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., et al. (2015). Genomic Correlates of Response to CTLA-4 Blockade in Metastatic Melanoma. *Science* 350, 207–211. doi:10.1126/science.aad0095

Wang, F., Liao, Y., Zhang, M., Zhu, Y., Wang, W., Cai, H., et al. (2021). N6-methyladenosine Demethyltransferase FTO-Mediated Autophagy in Malignant Development of Oral Squamous Cell Carcinoma. *Oncogene* 40, 3885–3898. doi:10.1038/s41388-021-01820-7

Wang, L., Cheng, L., Li, N.-N., Yu, W.-J., Sun, X.-Y., and Peng, R. (2015). Genetic Analysis ofSLC41A1in Chinese Parkinson's Disease Patients. *Am. J. Med. Genet.* 168, 706–711. doi:10.1002/ajmg.b.32365

Wang, S.-A., Wang, Y.-C., Chuang, Y.-P., Huang, Y.-H., Su, W.-C., Chang, W.-C., et al. (2017). EGF-mediated Inhibition of Ubiquitin-specific Peptidase 24 Expression Has a Crucial Role in Tumorigenesis. *Oncogene* 36, 2930–2945. doi:10.1038/onc.2016.445

Wei, Y., Vellanki, R. N., Coyaud, É., Ignatchenko, V., Li, L., Krieger, J. R., et al. (2015). CHCHD2 Is Coamplified with EGFR in NSCLC and Regulates Mitochondrial Function and Cell Migration. *Mol. Cancer Res.* 13, 1119–1129. doi:10.1158/1541-7786.MCR-14-0165-T

Xie, Y., Liu, J., Kang, R., and Tang, D. (2021). Mitophagy in Pancreatic Cancer. *Front. Oncol.* 11, 616079. doi:10.3389/fonc.2021.616079

Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., et al. (2018). TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 78, 6575–6580. doi:10.1158/0008-5472.CAN-18-0689

Yan, C., Gong, L., Chen, L., Xu, M., Abou-Hamdan, H., Tang, M., et al. (2020). PHB2 (Prohibitin 2) Promotes PINK1-PRKN/Parkin-dependent Mitophagy by the PARL-PGAM5-PINK1 axis. *Autophagy* 16, 419–434. doi:10.1080/15548627.2019.1628520

Yao, Y., Su, J., Zhao, L., Li, R., Liu, K., and Wang, S. (2019). CHCHD2 Promotes Hepatocellular Carcinoma and Indicates Poor Prognosis of Hepatocellular Carcinoma Patients. *J. Cancer* 10, 6822–6828. doi:10.7150/jca.31158

Yin, X., Xia, J., Sun, Y., and Zhang, Z. (2020). CHCHD2 Is a Potential Prognostic Factor for NSCLC and Is Associated with HIF-1a Expression. *Bmc Pulm. Med.* 20, 40. doi:10.1186/s12890-020-1079-0

Youle, R. J., and Narendra, D. P. (2011). Mechanisms of Mitophagy. *Nat. Rev. Mol. Cel Biol* 12, 9–14. doi:10.1038/nrm3028

Zacksenhaus, E., Shrestha, M., Liu, J. C., Vorobieva, I., Chung, P. E. D., Ju, Y., et al. (2017). Mitochondrial OXPHOS Induced by RB1 Deficiency in Breast Cancer: Implications for Anabolic Metabolism, Stemness, and Metastasis. *Trends Cancer* 3, 768–779. doi:10.1016/j.trecan.2017.09.002

Zhang, Y., Du, W., Chen, Z., and Xiang, C. (2017). Upregulation of PD-L1 by SPP1 Mediates Macrophage Polarization and Facilitates Immune Escape in Lung Adenocarcinoma. *Exp. Cel Res.* 359, 449–457. doi:10.1016/j.yexcr.2017.08.028

Zheng, H., Zhou, C., Lu, X., Liu, Q., Liu, M., Chen, G., et al. (2018). DJ-1 Promotes Survival of Human colon Cancer Cells under Hypoxia by Modulating HIF-1α Expression through the PI3K-AKT Pathway. *Cancer Manag. Res.* 10, 4615–4629. doi:10.2147/CMAR.S172008

Zhou, W., Ma, D., Sun, A. X., Tran, H.-D., Ma, D.-l., Singh, B. K., et al. (2019). PD-linked CHCHD2 Mutations Impair CHCHD10 and MICOS Complex Leading to Mitochondria Dysfunction. *Hum. Mol. Genet.* 28, 1100–1116. doi:10.1093/hmg/ddy413

Zhu, Z., Zhang, H., Chen, B., Liu, X., Zhang, S., Zong, Z., et al. (2020). PD-L1-Mediated Immunosuppression in Glioblastoma Is Associated with the Infiltration and M2-Polarization of Tumor-Associated Macrophages. *Front. Immunol.* 11, 588552. doi:10.3389/fimmu.2020.588552

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 Li, Lu, Zhang, Liu, Zhou, Deng, Qiu, Chen, Li, Yang, Zheng 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fmolb.2021.751770

Immunohistochemical Analysis of CYP11B2, CYP11B1 and β-catenin Helps Subtyping and Relates With Clinical Characteristics of Unilateral Primary Aldosteronism

*Luyan Sun[1†], Yiran Jiang[1†], Jing Xie[2†], Hongyuan Zhu[1], Luming Wu[1], Xu Zhong[1], Weiwei Zhou[1], Tingwei Su[1] and Weiqing Wang[1,3]**

[1]Shanghai Key Laboratory for Endocrine Tumors, Shanghai Clinical Centre for Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, [2]Department of Pathology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, [3]Laboratory for Endocrine and Metabolic Diseases of Institute of Health Science, Shanghai Jiaotong University School of Medicine and Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

OPEN ACCESS

Edited by:
Tianhui Chen,
University of Chinese Academy of Sciences, China

Reviewed by:
Yu Liu,
Ningbo University, China
Fang Li,
Dalian Medical University, China

***Correspondence:**
Weiqing Wang
wqingw61@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 02 August 2021
Accepted: 07 September 2021
Published: 24 September 2021

Citation:
Sun L, Jiang Y, Xie J, Zhu H, Wu L, Zhong X, Zhou W, Su T and Wang W (2021) Immunohistochemical Analysis of CYP11B2, CYP11B1 and β-catenin Helps Subtyping and Relates With Clinical Characteristics of Unilateral Primary Aldosteronism. Front. Mol. Biosci. 8:751770. doi: 10.3389/fmolb.2021.751770

Background: Primary aldosteronism is caused by aldosterone overproduction. While conventional hematoxylin-eosin staining can demonstrate morphological abnormality, it cannot provide any functional histopathological information. We aimed to identify the diagnostic, functional and prognostic value of CYP11B2, CYP11B1, and β-catenin immunostaining in unilateral hyperaldosteronism.

Method: A total of 134 patients with unilateral hyperaldosteronism were recruited in our study. The expression of CYP11B2, CYP11B1, and β-catenin was evaluated semiquantitatively on 134 patients' sections using immunohistochemistry technology and the relationship with clinical data was assessed.

Results: Patients were classified into four subtypes based on CYP11B2 staining as below: (1)118 patients with unilateral single aldosterone-producing adenoma (APA), (2)11 with unilateral multiple APA, (3)four with aldosterone-producing cell cluster (APCC), and (4)one with an undefined source. Adjusted CYP11B2 H-score was correlated with serum aldosterone, aldosterone to renin ratio (ARR), and serum potassium. In the abnormal β-catenin staining group, hypertension duration, aldosterone, ARR, cortisol, tumor diameter, tumor area, and CYP11B2 H-score were significantly higher than those of the wild-type group. Serum potassium level was significantly lower in the abnormal β-catenin staining group. Age, gender, BMI, family history of hypertension, adjusted CYP11B2 and CYP11B1 H-scores differed significantly between complete clinical success and incomplete clinical success groups. Age, gender and family history of hypertension were independently associated with complete clinical success based on multivariate logistic regression analysis.

Conclusion: CYP11B2 immunostaining could improve the differential diagnosis of unilateral hyperaldosteronism. Adjusted CYP11B2 H-score could be used as a

histopathological marker to reflect the severity of unilateral APA. Dysregulation of Wnt/β-catenin signaling and impaired β-catenin degradation may provoke the proliferation and enhance the steroidogenic ability of APA tumor cells, indicating that the Wnt pathway might be a potential, actionable, therapeutic target in the treatment of hyperaldosteronism. Age, sex and family history of hypertension were independent predictors of clinical outcome after adrenalectomy for unilateral hyperaldosteronism.

Keywords: hyperaldosteronism, CYP11B2, CYP11B1, beta catenin, immunohistochemistry, prognosis

INTRODUCTION

Primary aldosteronism (PA) is a group of disorders caused by an autonomous overproduction of aldosterone from the adrenal glands, independent of the renin-angiotensin system (Stowasser and Gordon, 2016). PA is the most common form of secondary hypertension, with an estimated prevalence reaching more than 10% in the hypertensive population (Rossi et al., 2006) and even approximately 20% in resistant hypertension patients (Calhoun et al., 2002).

Conventional hematoxylin-eosin (HE) staining performed in most pathological laboratories provides only morphological information; however, it is unsuitable for functional histopathological analyses (Nakamura et al., 2014). For instance, the differential distinction between an adenoma and a nodular hyperplasia is unclear on HE staining because many aldosterone-producing adenomas (APAs) also exhibit peritumoral zona glomerulosa (ZG) hyperplasia (Ganguly, 1998; Enberg et al., 2004; Boulkroun et al., 2010; Yamazaki et al., 2017). CYP11B2 and CYP11B1, which catalyze the final biosynthesis steps of aldosterone and cortisol, respectively, share a 93% homogeneity at the amino acid sequence level (Gomez-Sanchez et al., 2014). With the successful development of specific antibodies against CYP11B2 and CYP11B1 (Gomez-Sanchez et al., 2014), immunohistochemistry is now a promising technique for precisely locating the culprit responsible for aldosterone excess and for broadening our knowledge regarding the spectrum of PA. Nishimoto et al. (Nishimoto et al., 2010) have revealed an entity of subcapsular CYP11B2-expressing cells, termed aldosterone-producing cell cluster (APCC), which are 200–1,300 μm wide and 100–500 μm deep and can sustain aldosterone secretion in spite of a suppressed renin-angiotensin-aldosterone system. Later, Omata et al. (Omata et al., 2018) demonstrated that bilateral idiopathic hyperaldosteronism may result from the accumulation and enlargement of APCCs which harbor somatic aldosterone-driver gene mutations rather than diffuse aldosterone-producing ZG hyperplasia. Hence, CYP11B2 immunostaining can markedly improve the accuracy of PA histopathological diagnosis.

The Wnt/β-catenin signaling pathway is critical for regulating the proliferation, differentiation, and tumorigenesis of the adrenal cortex (El Wakil and Lalli, 2011; Clevers and Nusse, 2012). And several studies have reported mutaion burden of CTNNB1 gene which encodes β-catenin in APA (Scholl et al., 2015; Akerstrom et al., 2016). The frequency of somatic CTNNB1 mutation in APA ranges from 2.1 to 5.1% (Scholl et al., 2015; Akerstrom et al., 2016). Most CTNNB1 mutations affect the serine/threonine residues in exon 3, which thereby hamper the proteasomal degradation of β-catenin under resting conditions, leading to abnormal intracellular accumulation of β-catenin. Actually, constitutive activation of Wnt/β-catenin signaling is associated with aldosterone overproduction and treatment of H295R cell line, which already has an activating mutation of CTNNB1 endogenously, with siRNA against β-catenin or β-catenin inhibitor can significantly decrease aldosterone secretion (Berthon et al., 2014). However, the mechanism underlying how β-catenin provokes mineralocorticoid overproduction is undetermined; possible explanations include the indirect activation of CYP11B2 promoters (Berthon et al., 2014) or increased expression of LHCGR and GNRHR by β-catenin in pregnant and postmenopausal patients (Teo et al., 2015).

Laparoscopic adrenalectomy is the first-line choice for treating unilateral PA and excess aldosterone secretion can be resolved in most cases (Stowasser and Gordon, 2016). But not all PA patients can achieve hypertension cure or remission after the surgery, the factors that may affect clinical outcome include age, gender, family history of essential hypertension, hypertension duration, BMI, antihypertensive drug dosage and so on (Steichen et al., 2012; Williams et al., 2017). Nevertheless, individual prediction of the clinical outcome based on the above factors is not completely accurate, and more predictors based on new emerging technology need to be identified (Stowasser and Gordon, 2016). To date, the association between immunostaining features and prognosis of unilateral PA after surgery still remains an uncharted area and warrants further investigation.

Therefore, the aim of our current study is to utilize CYP11B2 immunostaining to subtype clinically diagnosed unilateral PA and explore the functional and prognostic significance of CYP11B2, CYP11B1 and β-catenin immunostaining.

PATIENTS AND METHODS

Patients

We recruited 134 patients referred to the Shanghai Clinical Centre for Endocrine and Metabolic Diseases in Ruijin Hospital Affiliated to Shanghai Jiaotong University School of Medicine from February 2010 to January 2013. The study protocol was approved by the local ethics committee and informed consent was obtained from all patients.

All patients underwent a diagnostic workup that included measurement of serum ions (Na^+, K^+, Cl^-, P, and Ca^{2+}), serum cortisol, plasma renin activity (PRA), and serum aldosterone. Diagnostic workup of PA was confirmed according to the 2008 Endocrine Society Clinical Practice Guideline as previously described (Funder et al., 2008; Jiang et al., 2015). First, patients were required to withdraw mineralocorticoid receptor antagonists and non-potassium sparing diuretics for at least 4 weeks. In addition, β-blockers, angiotensin-converting enzyme inhibitors, and angiotensin II receptor blockers were withdrawn for at least 2 weeks. Only non-dihydropyridine calcium blockers and/or α-receptor blockers were permitted to control blood pressure. Oral potassium supplementation was added if serum potassium level was less than 3.5 mmol/L. Second, aldosterone to renin ratio (ARR) was calculated for detecting suspicious cases of PA. Patients were kept in an upright posture for 2 h and then seated for 15 min before blood samples were collected in the morning. Third, a saline infusion test was performed in patients with ARR >30 (ng/dl)/(ng/mL h) as the confirmatory test for PA. Serum aldosterone was measured in a supine position before and after the infusion of 2L 0.9% NaCl solution from 8:00 am to 12:00 pm. Failure in suppressing post infusion aldosterone level to 10 ng/dl confirmed the diagnosis of hyperaldosteronism.

Adrenal computed tomography (CT) and adrenal venous sampling (AVS) were performed in patients with PA to differentiate between bilateral and unilateral PA. More details of AVS are described in the Supplementary Material. Patients with unilateral aldosterone overproduction underwent laparoscopic adrenalectomy.

Patients aging from 18 to 70 years old who underwent confirmation workup of unilateral PA (saline infusion test, CT scan and AVS) were included in our study. Patients who are concomitant with other endocrine diseases such as Cushing syndrome and pheochromocytoma, who have chronic renal failure, chronic heart failure, liver cirrhosis and have cardiovascular event attack within 6 months were excluded in our study. Also, pregnant patients were excluded. 156 patients were screened in our study. 22 patients were excluded according to exclusion criteria.

The diagnosis of adenoma and UAH were established according to the four corners criteria: 1) a biochemical confirmation of PA; 2) aldosterone secretion is lateralized, which is validated by AVS; 3)adenoma or nodular hyperplasia at histopathology; and crucially important, 4) with cure or improvement in blood pressure, aldosterone, plasma renin activity, ARR and serum potassium after adrenalectomy (Stowasser and Gordon, 2016). Pathological diagnosis was performed by two experienced pathologists. Primary Aldosteronism Surgical Outcome (PASO) criteria were used for outcomes and follow-up assessment of adrenalectomy for unilateral PA (Williams et al., 2017). Complete clinical success was defined as patients with normalized blood pressure without any use of antihypertensive medicine. Incomplete clinical success was defined as patients requiring antihypertensive medicine to control BP, with partial and absent clinical success defined according to PASO criteria. Blood pressure, serum ions (Na^+, K^+, Cl^-, P, and Ca^{2+}), aldosterone and plasma renin activity were measured as routine.

Laboratory Assays

Laboratory assays were performed as previously described (Jiang et al., 2015). All tests were performed in a CAP (College of American Pathologists, No. 7217913)-accredited laboratory. Serum aldosterone and PRA were measured by radioimmunoassay following the manufacturer's instructions (Beckman Coulter Corp., Brea, CA, United States). Plasma cortisol was measured on Access Immunoassay Systems (Beckman Coulter Corp., Brea, CA, United States).

Immunohistochemistry

Immunohistochemistry was performed according to the standard histopathological protocol described in our previous study (Wu et al., 2018). Antibodies against CYP11B2 (mouse monoclonal, 1: 200; Merck Millipore, MABS1251, RRID: AB_2650562), CYP11B1 (rat monoclonal, 1:200; Merck Millipore, MABS502, RRID: AB_2650563), and β-catenin (rabbit polyclonal, 1:400; Cell Signaling Technology, Cat #9562, RRID: AB_331,149) were used. A standard avidin-biotin-peroxidase complex technique was used to demonstrate the primary antibody binding (Vector pk-7200; Vector Laboratories, Inc., Burlingame, CA, United States). Here, an APA was defined as a CYP11B2-positive, well-circumscribed, round or ovoid-shaped nodular lesion composed of a mixture of ZG-like and zona fasciculata (ZF)-like cells, often with a fibrous capsule around it. In contrast, an APCC was defined as a CYP11B2-positive cell cluster, cuneiform or trapezoid in shape, and morphologically identical to adjacent ZG cells without a fibrous capsule. McCarty H-score was used to evaluate the immunoreactivity semi-quantitatively in our study, considering both the percentage of positively stained cells and intensity of their immunopositivity (McCarty et al., 1985; Nakamura et al., 2014; Ono et al., 2014; Monticone et al., 2015). More details of assessing McCarty H-score are described in the **Supplementary Material**.

As for β-catenin immunostaining, both the percentage of immunopositive cells and immunointensity of the stained cells were also recorded. Patients were categorized into the abnormal β-catenin staining group for focal cytoplasmic staining (<30% but strong staining), diffuse cytoplasmic staining (30–70% and at least distinct staining, >70% regardless of staining intensity), focal nuclear staining (<5% of nuclei but strong staining), or diffuse nuclear staining (>5% and at least distinct staining); otherwise, they were categorized into the wild-type staining group (Wu et al., 2018).

Statistical Analysis

Continuous variables were expressed as median (interquartile intervals: 25–75%), while categorical variables were expressed as frequency and percentage. The Mann-Whitney *U* test was used for the comparison of continuous parameters between two different groups, and the chi-squared test was introduced for the comparison of categorical variables. Spearman's rank correlation was employed to explore the association between clinical and immunostaining parameters. We performed

TABLE 1 | Clinical characteristics of the study population.

	All patients (n = 134)
Age (years)	48 (40, 55)
Male (%)	42.5
Duration of hypertension (months)	78 (36, 120)
Family history of hypertension (%)	53.7
SBP (mmHg)	165 (150, 180)
DBP (mmHg)	100 (95, 110)
BMI (kg/m^2)	23.19 (21.22, 25.51)
Aldosterone (ng/dl)	44.75 (28.32, 68.27)
Renin activity (ng/ml·h)	0.10 (0.03, 0.22)
ARR	468.11 (170.90, 1929.33)
Serum potassium (mmol/L)	2.47 (2.10, 2.86)
Prevalence of Hypokalemia (%)	97.0
Serum cortisol (μg/dl)	11.49 (9.08, 14.92)
Tumor diameter (cm)	1.50 (1.10, 2.00)
Tumor area (cm^2)	1.77 (0.95, 3.14)

Data are expressed as median with interquartile range or proportion of patients (%). Hypokalemia is defined as spontaneous serum potassium concentration <3.5 mmol/L. Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; ARR, aldosterone to renin ratio.

univariate and multivariate logistic regression analyses to determine the predictors associated with complete clinical success. All statistical analyses were two-sided and performed using IBM SPSS Statistics (version 22.0) software. $p < 0.05$ was considered significant.

RESULTS

Clinical and Hormonal Data of Patients With Unilateral PA

Overall 134 patients were enrolled in our study (**Table 1**). The median age at presentation was 48 years, with an approximately 78 months median duration of hypertension. The median systolic blood pressure (SBP) and diastolic blood pressure (DBP) were 165 and 100 mmHg, respectively. The median body mass index (BMI) was 23.19 kg/m^2. Male patients accounted for 42.5% of all the subjects and 53.7% of the patients had a family history (first-degree relatives) of hypertension. The median tumor size on CT scan was 1.5 cm, and the median tumor area was 1.77 cm^2. Additionally, the serum aldosterone concentration (median 44.75 ng/dl, range 28.32–68.27) and ARR (median 468.11, range 170.90–1929.33) were high and PRA was suppressed (median 0.10 ng/ml h, range 0.03–0.22). The median value of serum cortisol among the patients with PA was 11.49 μg/dl (range 9.08–14.92).

Immunohistopathological Subtyping of Unilateral PA Using CYP11B2 Immunohistochemistry

HE staining revealed that 77.6% (104/134) of the patients were diagnosed with adenoma and 22.4% (30/104) with unilateral adrenal hyperplasia (UAH). Among 104 patients with adenoma, 91 showed positive CYP11B2 immunostaining in their tumors; whereas 11 patients showing multiple CYP11B2-positive immunostaining tumors were considered as multiple APAs (**Figure 1**). One patient showed APCC in the ZG with CYP11B2-negative unilateral adrenocortical adenomas. The remaining one patient showed neither positive CYP11B2 immunostaining adenoma nor APCC on his section (**Supplementary Figure S1**). Among 30 patients with UAH, 27 patients showed one positive CYP11B2 staining nodule; consequently, their diagnosis was changed to APA. Three patients showed APCC in their adrenal cortex without CYP11B2-positive adrenocortical nodules or diffuse CYP11B2-positive ZG hyperplasia.

Therefore, the final diagnosis of these 134 patients with unilateral PA was classified into four subtypes according to CYP11B2 immunostaining: 1).118 unilateral single APA, 2).11 patients with unilateral multiple APAs, 3). four patients with APCC, and 4). one patient with unilateral PA of undefined source (**Table 2**).

CYP11B2/1 Immunostaining Score and Correlation Between CYP11B2/1 Staining Score and Clinical Characteristics

In the semiquantitative analysis of CYP11B2 immunostaining, illustration of various intensities of CYP11B2 staining in unilateral APA specimens is demonstrated in **Supplementary Figure S2**. Tumor area was calculated according to the diameter of the adrenocortical tumor on CT with the assumption that the tumor was approximately spherical in shape. As both tumor area and steroidogenic enzyme immunoreactivity are determinants of the overall tumor steroid hormone production, we investigated the functional significance of CYP11B2 by multiplying the H-score of the steroidogenic enzyme by tumor area and correlated the product with clinical parameters of the 118 unilateral single APA patients. The H-score of CYP11B2 adjusted for tumor area was positively correlated with serum aldosterone (r = 0.399, $p = 0.000$) and ARR (r = 0.301, $p = 0.001$), and inversely correlated with serum potassium (r = −0.236, $p = 0.010$). The CYP11B2 immunoreactivity adjusted for tumor area was not correlated with serum cortisol (r = 0.158, $p = 0.097$) or PRA (r = −0.168, $p = 0.073$; **Table 3**).

On the other hand, the correlation between CYP11B1 staining and clinical parameters revealed that the adjusted CYP11B1 H-score was positively correlated with serum cortisol (r = 0.434, $p = 0.000$), serum aldosterone (r = 0.308, $p = 0.001$), and ARR (r = 0.302, $p = 0.001$), and inversely correlated with PRA (r = −0.201, $p = 0.031$). The CYP11B1 immunoscore adjusted for tumor area was not correlated with serum potassium (r = −0.079, $p = 0.398$; **Table 3**).

β-catenin Immunostaining and the Difference in Clinical Characteristics Between Wild-type and Abnormal β-catenin Staining Groups

Wild-type staining of β-catenin was observed in 64.4% (76/118) of patients with APA and the remaining 35.6% (42/118) showed abnormal β-catenin staining (**Supplementary Table S1**). We

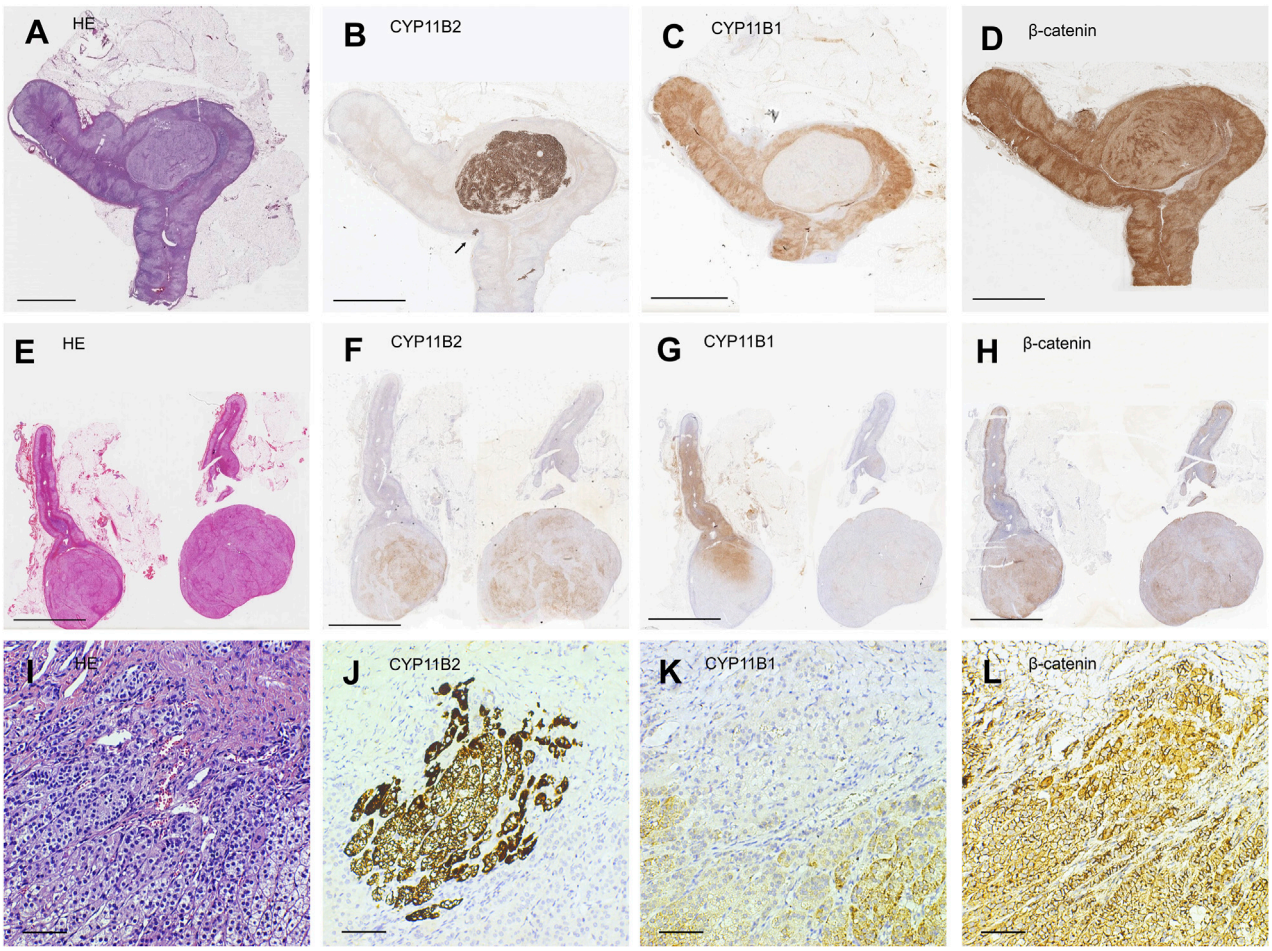


FIGURE 1 | Representative illustration of different subtypes of unilateral primary aldosteronism. **(A,E,I)** Hematoxylin-eosin staining. **(B,F,J)** Immunostaining for CYP11B2. **(C,G,K)** Immunostaining for CYP11B1. **(D,H,L)** Immunostaining for β-catenin. Scale bars, 5 mm **(A–H)** and 50 μm **(I–L)**. **(A–D)** Representative illustration of aldosterone-producing adenoma (APA). **(E–H)** Representative illustration of multiple APA. **(I–L)** Representative illustration of an aldosterone-producing cell cluster (APCC) whose site is indicated by a black arrow in **(B)**.

TABLE 2 | Subtype classification based on CYP11B2 immunostaining result of unilateral primary aldosteronism.

Hematoxylin-eosin staining	n	CYP11B2 immunostaining result	Subtype classification
adenoma	91	Single CYP11B2-positive adenoma	Single APA
adenoma	11	Multiple CYP11B2-positive adenomas	Multiple APAs
adenoma	1	CYP11B2 negative in adenoma, APCC in cortex	APCC
adenoma	1	CYP11B2 negative in whole section	Untargetable culprit
UAH	27	Single CYP11B2-positive nodule	Single APA
UAH	3	CYP11B2 negative in nodular hyperplasia, APCC in cortex	APCC

UAH, unilateral adrenal hyperplasia; n, number of patients; APCC, aldosterone-producing cell cluster; APA, aldosterone-producing adenoma.

further compared the preoperative clinical and endocrinological parameters of the patients between the wild-type staining and abnormal staining groups (**Supplementary Table S1**). Preoperative serum aldosterone (p = 0.004), ARR (p = 0.029), serum cortisol (p = 0.000), duration of hypertension (p = 0.019), tumor diameter (p = 0.000), and tumor area (p = 0.000) were significantly higher in the abnormal staining group than that in the wild-type group. Serum potassium level (p = 0.032) was significantly lower in the abnormal staining group than that in the wild-type staining group. However, there were no significant differences in SBP (p = 0.505), DBP (p = 0.504), age (p = 0.146), sex (p = 0.471), family history of hypertension (p = 0.763), BMI (p = 0.800), and PRA (p = 0.208) between these two groups. As for the immunoreactivity score, the CYP11B2 score (p = 0.000) was

TABLE 3 | Results of Spearman's correlation test between immunohistochemical staining scores and clinical parameters.

		Aldosterone	PRA	ARR	Potassium	Cortisol
Adjusted CYP11B2 score	r	0.399	–0.168	0.301	–0.236	0.158
	P	0.000	0.073	0.001	0.010	0.097
Adjusted CYP11B1 score	r	0.308	–0.201	0.302	–0.079	0.434
	P	0.001	0.031	0.001	0.398	0.000

Abbreviations: r, Spearman correlation coefficient; PRA, plasma renin activity; ARR, aldosterone to renin ratio.

significantly higher in the abnormal staining group than that in the wild-type staining group, whereas the CYP11B1 score was similar ($p = 0.262$) between these two groups (**Figure 2**).

Comparison Between and Predictive Factors Associated With Different Clinical Outcome Groups

Ninety-six of 118 APA patients were successfully followed up to assess the outcome after adrenalectomy for unilateral primary aldosteronism and they were categorized into complete and incomplete clinical success groups as described in the Methods section. Clinical characteristics and immunostaining features of the two groups are compared in **Table 4**. Gender and family history of hypertension differed significantly between the two groups. In the complete clinical success group, the age ($p < 0.05$), BMI ($p < 0.05$), adjusted CYP11B2 H-score ($p < 0.05$) and adjusted CYP11B1 H-score ($p < 0.05$) were significantly lower than those in the incomplete clinical success group. There were no significant differences in the duration of hypertension ($p = 0.30$), SBP ($p = 0.07$), DBP ($p = 0.25$), serum aldosterone ($p = 0.83$), PRA ($p = 0.49$), ARR ($p = 0.58$), serum potassium ($p = 0.41$), serum cortisol ($p = 0.07$), tumor diameter ($p = 0.21$), tumor area ($p = 0.21$) and β-catenin staining status ($p = 0.64$) between the two groups. Unadjusted and adjusted logistic regression analyses were used to identify factors associated with different clinical outcomes (**Supplementary Table S2**). The unadjusted logistic regression analysis showed that age, sex, BMI and family history of hypertension were associated with complete clinical success, but adjusted CYP11B2 H-score and adjusted CYP11B1 H-score were not. Finally, we identified three independent predictive factors associated with complete clinical success using an adjusted analysis: age, sex and family history of hypertension, while BMI, adjusted CYP11B2 H-score and adjusted CYP11B1 H-score were not independently associated with complete clinical success (**Supplementary Table S2**).

DISCUSSION

Herein, our study firstly evaluated the association between steroidogenic enzyme immunoreactivity, β-catenin immunostaining status, clinical parameters and outcome in patients with unilateral PA. Differential diagnosis between unilateral APA and UAH can sometimes be puzzling and difficult when using the traditional HE staining technique, because the ZG of the peritumoral tissue adjacent to APA can sometimes display different levels of hyperplasia, and HE staining cannot identify the precise lesion that is responsible for aldosterone overproduction (Ganguly, 1998; Enberg et al., 2004; Boulkroun et al., 2010; Nakamura et al., 2014; Yamazaki et al., 2017). *In situ* hybridization of CYP11B2 has been used to identify aldosterone-producing cells; however, this technique cannot be performed at a large-scale as it is time-consuming and difficult to implement. Using a specific monoclonal antibody for CYP11B2, immunohistochemistry staining of steroidogenic enzyme can be an advantageous tool to precisely localize the cells producing aldosterone and markedly improve the diagnostic accuracy as it is faster, more convenient and less expensive than *in situ* hybridization (Nanba et al., 2013; Gomez-Sanchez et al., 2014). Here, 30 patients were initially diagnosed with UAH based on HE staining, and then 27 of them were re-diagnosed with APA after CYP11B2 immunostaining, as positive nodular staining was observed. Moreover, we found more than one APAs on the tissue sections of 11 patients via CYP11B2 immunostaining, many of which were undetectable under CT examination. Yamazaki et al. (Yamazaki et al., 2017) classified CT-negative PA into two subtypes based on CYP11B2 immunostaining: multiple adrenocortical micronodules and diffuse hyperplasia. They also reported that the ZG of adjacent adrenals near micronodules often demonstrated various degrees of paradoxical hyperplasia, which is histopathologically considered hyperplasia but is negative for CYP11B2 staining. Although all patients in our study showed positive CT findings, our result is consistent with that of Yamazaki et al. Paradoxical hyperplasia of ZG cells was also observed in our study. Interestingly, neither diffuse CYP11B2-positive hyperplasia of ZG cells nor positive nodular staining was observed in the remaining three UAH patients; but several APCCs within the adrenal cortex were observed on their sections. Nishimoto et al. (Nishimoto et al., 2010) firstly identified a novel cell cluster positive for CYP11B2 immunostaining, which they termed as APCC. Several studies (Nishimoto et al., 2010; Nanba et al., 2017; Omata et al., 2017; Omata et al., 2018) have demonstrated that APCCs frequently harbor aldosterone-driver somatic mutations mostly in the *CACNA1D* gene, and can autonomously secrete mineralocorticoid independent of the renin-angiotensin system under normal and pathological conditions. Omata et al. (Omata et al., 2018) elucidated that idiopathic hyperaldosteronism (IHA) may result from the accumulation and enlargement of APCCs instead of hyperplasia of ZG cells. Moreover, Nanba et al. (Nanba et al., 2013) observed APCCs in unilateral PA patients with ipsilateral adrenocortical tumor negative for CYP11B2 staining. According to these findings, we hypothesized that APCCs may be the source of aldosterone excess in certain

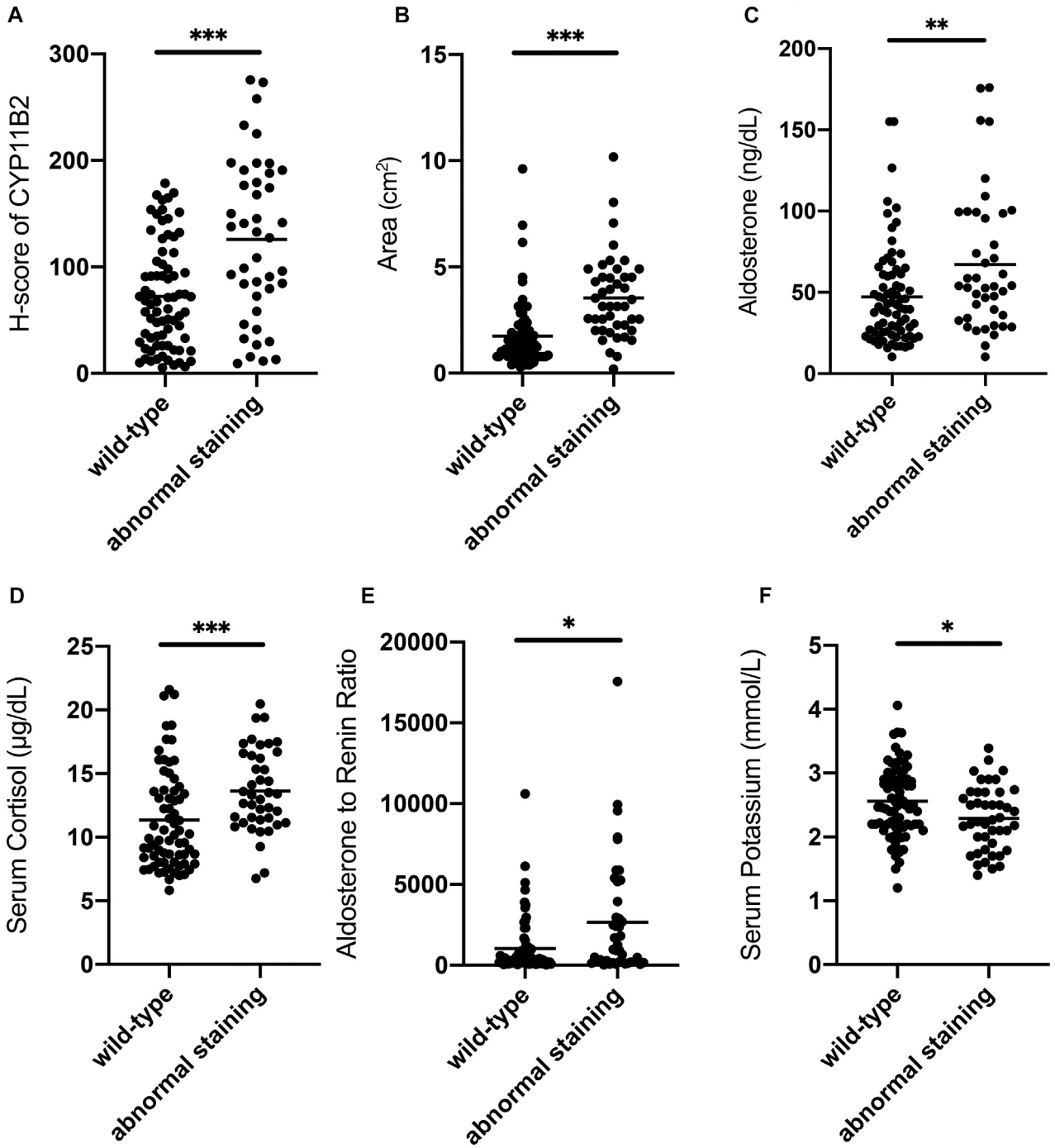


FIGURE 2 | Comparison of clinical characteristics between wild-type staining and abnormal staining groups of β-catenin in patients. **(A)** CYP11B2 H-score of the abnormal staining group is significantly higher than that of the wild-type staining group. **(B)** Tumor area of the abnormal staining group is significantly higher than that of the wild-type staining group. **(C)** Serum aldosterone of the abnormal staining group is significantly higher than that of the wild-type staining group. **(D)** Serum cortisol of the abnormal staining group is significantly higher than that of the wild-type staining group. **(E)** ARR of the abnormal staining group is significantly higher than that of the wild-type staining group. **(F)** Serum potassium of the abnormal staining group is significantly lower than that of the wild-type staining group. Bars in the figure represent the median of each group data. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

patients with unilateral PA rather than diffuse hyperplasia of aldosterone-producing ZG cells or adenomas. Additionally, one patient in our study had a CYP11B2-negative adrenocortical tumor and one APCC on his section, indicating that APCC rather than the adrenocortical tumor found by CT may be responsible for the aldosterone overproduction. Thus, our

TABLE 4 | Comparison of clinical characteristics and immunohistochemical parameters between different clinical outcome groups in APA patients.

	Complete clinical success (n = 51)	Incomplete clinical success (n = 45)	*P*
Age (years)	46 (35, 54)	51 (42, 56)	<0.05
Male (%)	33.3	60.0	<0.01
Duration of hypertension (months)	84 (24, 120)	96 (48, 132)	0.30
Family history of hypertension (%)	39.2	77.8	<0.001
SBP (mmHg)	160 (150, 180)	173 (150, 195)	0.07
DBP (mmHg)	100 (90, 110)	100 (100, 115)	0.25
BMI (kg/m^2)	22.86 (20.20, 24.89)	23.44 (22.23, 26.64)	<0.05
Aldosterone (ng/dl)	43.02 (28.63, 73.91)	50.37 (26.16, 71.13)	0.83
Renin activity (ng/ml·h)	0.08 (0.03, 0.21)	0.11 (0.03, 0.28)	0.49
ARR	659.27 (169.90, 2,358.00)	328.24 (153.74, 2,283.51)	0.58
Serum potassium (mmol/L)	2.40 (2.10, 2.70)	2.50 (2.10, 2.96)	0.41
Prevalence of Hypokalemia (%)	98.0	95.6	0.16
Serum cortisol (μg/dl)	10.97 (8.06, 13.69)	12.24 (10.42, 16.12)	0.07
Tumor diameter (cm)	1.50 (1.00, 2.00)	1.60 (1.25, 2.22)	0.21
Tumor area (cm^2)	1.77 (0.79, 3.14)	2.01 (1.23, 3.87)	0.21
Adjusted CYP11B2 score	95.18 (39.16, 251.11)	186.40 (80.58, 415.28)	<0.05
Adjusted CYP11B1 score	128.59 (58.56, 232.48)	173.22 (113.70, 329.29)	<0.05
β-catenin wild-type staining (%)	64.7	60.0	0.64

Data are expressed as median with interquartile range or proportion of patients (%).
Hypokalemia is defined as spontaneous serum potassium concentration <3.5 mmol/L. Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; ARR, aldosterone to renin ratio.

findings, in accordance with other studies, indicated that CYP11B2 staining can more precisely pinpoint the culprit responsible for aldosterone overproduction compared to traditional HE staining and effectively differentiate between different subtypes of unilateral PA. Further, APCC may play a crucial role not only in normal blood pressure maintenance (Nanba et al., 2017; Omata et al., 2017) but also in both unilateral and bilateral, CT-negative and CT-positive hyperaldosteronism (Nanba et al., 2013; Yamazaki et al., 2017; Omata et al., 2018).

Immunoreactivity was assessed using the McCarty H-score, which is now considered the best available approach for semi-quantitatively evaluating the immunoreactivity of steroidogenic enzymes in an objective manner in a specific tumor sample (McCarty et al., 1985). In our study, the CYP11B2 H-score correlated with serum aldosterone and serum potassium, but not with PRA, ARR, or serum cortisol. Meanwhile, the CYP11B1 H-score correlated with serum cortisol, but not with serum aldosterone, PRA, or ARR (data not shown). Several studies (Nanba et al., 2013; Nakamura et al., 2014; Ono et al., 2014) have shown that steroidogenic enzyme H-score level alone does not necessarily represent the overall steroid hormone production capacity of an adrenocortical tumor because it only represents the hormone production ability of the tumor per unit area and per cell, and there may be several other factors such as tumor area and upstream steroidogenic enzymes of CYP11B2/1 that may also play a role in hormone production. Therefore, the functional significance of CYP11B2/1 immunostaining was investigated by correlating the H-score adjusted for tumor area with the endocrine data. Through this adjustion, the adjusted CYP11B2 score was correlated with serum aldosterone, ARR, and serum potassium, and the adjusted CYP11B1 score was correlated with serum 8:00 am cortisol, serum aldosterone, ARR, and PRA. These correlationships provide further evidence that adjusted H-score can be used as a new marker to reflect clinical manifestations and severity of unilateral PA.

APA is a heterogenous lesion composed of cells originating from different adrenocortical zones. Both ZF-like cells (cells with large, vacuolated, lipid-laden cytoplasm and central, round nuclei similar to that of ZF) and ZG-like cells (cells with relatively small, lipid-poor cytoplasm and a higher nuclear to cytoplasmic ratio similar to that of ZG) are observed in the majority of APAs (Enberg et al., 2004; Nakamura et al., 2014; Gioco et al., 2015; Fallo et al., 2017). Our results were consistent with the histopathological findings and provide further immunohistochemical evidence on this perspective. Functionally, CYP11B2 is restricted in ZG, catalysing the final steps of mineralocortical biosynthesis, and CYP11B1 is restricted in ZF and zona reticularis (ZR), catalysing the final step of glucocorticoid formation (Stowasser and Gordon, 2016). Our results suggest that APA can simultaneously produce cortisol in patients with PA, which may lead to clinical or subclinical manifestation of Cushing syndrome (Goupil et al., 2015; Stowasser and Gordon, 2016; Fallo et al., 2017). A positive correlationship between the adjusted CYP11B1 score and serum cortisol was revealed in our study, which emphasized the importance of careful monitoring of the hypothalamic-pituitary-adrenal axis before and after surgery in patients with unilateral PA.

Interestingly, the adjusted CYP11B1 score not only correlated with serum cortisol, but also with serum aldosterone, ARR, and PRA. This correlation raised a question on the role of the CYP11B1 enzyme in the biosynthesis of aldosterone. We suggest that both CYP11B1 and CYP11B2 can convert deoxycorticosterone to corticosterone via 11β-hydroxylase, and corticosterone is the precursor of aldosterone. Although CYP11B2 is the only enzyme that can convert corticosterone to aldosterone, CYP11B1 may still contribute to aldosterone

overproduction via synthesizing enough precursor, i.e. corticosterone, as a substrate for the biological activity of CYP11B2 (Nakamura et al., 2014; Ono et al., 2014).

The Wnt/β-catenin signaling pathway plays a pivotal role in the regulation of proliferation, differentiation, and function of the normal adrenal cortex and adrenocortical tumorigenesis (El Wakil and Lalli, 2011; Clevers and Nusse, 2012). Aberrant cytoplasmic and nuclear accumulation of β-catenin reflects the constitutive activation of this pathway (Tissier et al., 2005). The prevalence of abnormal β-catenin staining in our study was 35.6%, similar to that reported by Tissier et al. (Tissier et al., 2005). Further, we observed that the abnormal staining group secreted more aldosterone and cortisol than the wild-type staining group. Our results showed that the tumor area and CYP11B2 H-score were significantly higher in the abnormal staining group than those in the wild-type group. Several factors such as McCarty H-score and tumor area are involved in the overall steroid hormone generation of adrenocortical adenoma as described above. General steroid hormone production of APA simultaneously increases as tumor area becomes larger. Besides, higher CYP11B2 H-score implies that constitutive β-catenin activation increases the aldosterone synthesis capacity of APA per unit area and per cell. Studies have revealed that β-catenin can modulate the membrane potential and intracellular ion homoeostasis in different cell types through interacting with ion channels and transporters (Lesage et al., 2004; Wilmes et al., 2012; Zhao et al., 2014). Hence, we hypothesized that an aberrant cytoplasmic accumulation of β-catenin may depolarize the membrane potential and increase the cytosolic Ca^{2+} concentration of APA cells, thereby enhancing the steroidogenic ability. However, this hypothesis needs deeper exploration. Therefore, both the higher CYP11B2 H-score and tumor area effect synergistically give rise to higher serum aldosterone concentration in the abnormal staining group, whereas the higher serum cortisol in the abnormal staining group could mainly result from the tumor area effect rather than the similar CYP11B1 H-score compared with the wild-type group. Nevertheless, the underlying mechanism of why aberrant activation of the Wnt/β-catenin signaling pathway selectively enhances CYP11B2 expression rather than CYP11B1 in patients with APA warrants further investigation.

Laparoscopic adrenalectomy is now considered the standard approach in the treatment of unilateral PA as it is minimally invasive and can normalize aldosterone secretion, decrease blood pressure, and reduce antihypertensive medication dosage (Funder et al., 2008; Stowasser and Gordon, 2016). However, normotension without antihypertensive drug usage cannot always be achieved after adrenalectomy for unilateral PA. Complete clinical success was achieved in 37% of patients in the PASO study (ranging from 17 to 62% between different centers) (Williams et al., 2017). In the current study, complete clinical success was attained in 53.1% patients, similar with the PASO study. Laparoscopic adrenalectomy can confer a decrease in SBP about 25–40 mmHg and a reduction in the number of antihypertensive medications prescribed about 1–2 drug classes averagely (Steichen et al., 2012; Williams et al., 2017). Clinical characteristics predicting hypertension cure include sex, age, family history of hypertension, hypertension duration, BMI, estimated glomerular filtration rate, evidence of arteriolosclerosis and so on (Steichen et al., 2012; Stowasser and Gordon, 2016; Williams et al., 2017). Using the multivariate logistic regression model, our results proposed that sex, age and family history of hypertension were independently associated with clinical outcome. Patients who were younger, female and without a family history of hypertension had a higher likelihood of being cured of hypertension after adrenalectomy. Although the adjusted CYP11B2 and CYP11B1 H-scores were not independent predictors associated with clinical success based on multivariate analysis in our study, our results demonstrated that the adjusted H-scores in the complete success group were significantly lower than those in the incomplete success group, indicating that patients with lower immunostaining scores might benefit more from surgery for unilateral PA. It is reported that hyperaldosteronism patients have an increased rate of cardiovascular events than matched essential hypertension patients, showing that aldosterone per se can cause target organ damage independent from its hypertensive effect (Mulatero et al., 2013). The pathogenetic role of aldosterone excess include chronic inflammation, fibrosis, endothelial dysfunction, glucose and lipid metabolism disorder (Kozakova et al., 2003; Fallo et al., 2007; Staermose et al., 2009; Tsuchiya et al., 2009). Through these machanisms, aldosterone oversecretion can result in irreversible vascular and renal damage. Several studies have demonstrated media-to-lumen ratio of small arteries and estimated glomerular filtration rate are independently associated with the outcome following adrenalectomy for unilateral PA (Rossi et al., 2008; Group et al., 2009). As mentioned above, the adjusted H-scores are positively correlated with serum aldosterone. We postulate that patients with higher scores may develop more deleterious vascular remodeling and more impaired renal function due to the persistent secondary hypertension and thereby decrease the likelihood of complete clinical success (Rossi et al., 2008; Group et al., 2009). The main limitaion of our study is that the sample size is not large enough, which maybe undermines the statistical validity of multivariate analysis. Larger studies are needed to identify the association between immunohistochemical features and clinical outcome after adrenalectomy for unilateral PA.

In conclusion, CYP11B2 immunohistochemical staining is a useful histopathological tool for precisely locating the cells responsible for aldosterone production, and assists in the differential diagnosis and subtyping of unilateral PA. In a small portion of patients with unilateral PA, increased autonomous aldosterone production is not caused by the CT-detectable adenomas but might be driven by subcapsular APCCs. Adjusted CYP11B2 and CYP11B1 H-scores are correlated with serum aldosterone and other endocrine parameters in patients with APA. The aberrant constitutive activation of β-catenin can provoke the proliferation activity and enhance the steroidogenic ability of tumor cells, which suggests that a Wnt pathway inhibitor may be a potential, actionable therapeutic option for the treatment of hyperaldosteronism. Being younger, female and without a family history of hypertension predicts a better outcome after

adrenalectomy in primary aldosteronism patients. However, further studies are required to elucidate the prognostic significance of immunostaining on surgical outcome of unilateral PA.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The ethics committee of Ruijin Hospital, Shanghai Jiaotong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by YJ, LS, and JX. The first draft of the manuscript was written by YJ and LS, and all authors commented on previous versions of the manuscript. All authors have read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (81800683, 81902717), Shanghai Key Project of Precise Diagnosis and Treatment of Difficult Diseases (SHDC2020CR 2002A), Shanghai Key Support Project of Clinical Research (SHDC2020CR6015), Shanghai Health and Medical Development Foundation (805344), Training Program of the Excellent Young Talents of the Shanghai Municipal Health System (2017YQ005), Shanghai Municipal Health Commission (201740040, 201840049) and Shanghai Sailing Program (19YF1429300, 19YF1429400).

ACKNOWLEDGMENTS

We thank Yu Zhu, MD, Department of Urinary Surgery; Fukang Sun, MD, Department of Urinary Surgery; and Zhiyuan Wu, MD, Department of Radiology for their generous help in the diagnosis and treatment of patients with primary aldosteronism during our research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.751770/full#supplementary-material

REFERENCES

Åkerström, T., Maharjan, R., Sven Willenberg, H., Cupisti, K., Ip, J., Moser, A., et al. (2016). Activating Mutations in CTNNB1 in Aldosterone Producing Adenomas. *Sci. Rep.* 6, 19546. doi:10.1038/srep19546

Berthon, A., Drelon, C., Ragazzon, B., Boulkroun, S., Tissier, F., Amar, L., et al. (2014). WNT/β-catenin Signalling Is Activated in Aldosterone-Producing Adenomas and Controls Aldosterone Production. *Hum. Mol. Genet.* 23 (4), 889–905. doi:10.1093/hmg/ddt484

Boulkroun, S., Samson-Couterie, B., Dzib, J.-F. G., Lefebvre, H., Louiset, E., Amar, L., et al. (2010). Adrenal Cortex Remodeling and Functional Zona Glomerulosa Hyperplasia in Primary Aldosteronism. *Hypertension* 56 (5), 885–892. doi:10.1161/hypertensionaha.110.158543

Calhoun, D. A., Nishizaka, M. K., Zaman, M. A., Thakkar, R. B., and Weissmann, P. (2002). Hyperaldosteronism Among Black and white Subjects with Resistant Hypertension. *Hypertension* 40 (6), 892–896. doi:10.1161/01.hyp.0000040261.30455.b6

Clevers, H., and Nusse, R. (2012). Wnt/β-Catenin Signaling and Disease. *Cell* 149 (6), 1192–1205. doi:10.1016/j.cell.2012.05.012

El Wakil, A., and Lalli, E. (2011). The Wnt/beta-Catenin Pathway in Adrenocortical Development and Cancer. *Mol. Cell Endocrinol.* 332 (1-2), 32–37. doi:10.1016/j.mce.2010.11.014

Enberg, U., Volpe, C., Hoog, A., Wedell, A., Farnebo, L., Thoren, M., et al. (2004). Postoperative Differentiation between Unilateral Adrenal Adenoma and Bilateral Adrenal Hyperplasia in Primary Aldosteronism by mRNA Expression of the Gene CYP11B2. *Eur. J. Endocrinol.* 151 (1), 73–85. doi:10.1530/eje.0.1510073

Fallo, F., Castellano, I., Gomez-Sanchez, C. E., Rhayem, Y., Pilon, C., Vicennati, V., et al. (2017). Histopathological and Genetic Characterization of Aldosterone-Producing Adenomas with Concurrent Subclinical Cortisol Hypersecretion: a Case Series. *Endocrine* 58 (3), 503–512. doi:10.1007/s12020-017-1295-4

Fallo, F., Dellamea, P., Sonino, N., Bertello, C., Ermani, M., Vettor, R., et al. (2007). Adiponectin and Insulin Sensitivity in Primary Aldosteronism. *Am. J. Hypertens.* 20 (8), 855–861. doi:10.1016/j.amjhyper.2007.03.012

Funder, J. W., Carey, R. M., Fardella, C., Gomez-Sanchez, C. E., Mantero, F., Stowasser, M., et al. (2008). Case Detection, Diagnosis, and Treatment of Patients with Primary Aldosteronism: An Endocrine Society Clinical Practice Guideline. *J. Clin. Endocrinol. Metab.* 93 (9), 3266–3281. doi:10.1210/jc.2008-0104

Ganguly, A. (1998). Primary Aldosteronism. *N. Engl. J. Med.* 339 (25), 1828–1834. doi:10.1056/NEJM199812173392507

Gioco, F., Seccia, T. M., Gomez-Sanchez, E. P., Rossi, G. P., and Gomez-Sanchez, C. E. (2015). Adrenal Histopathology in Primary Aldosteronism. *Hypertension* 66 (4), 724–730. doi:10.1161/HYPERTENSIONAHA.115.05873

Gomez-Sanchez, C. E., Qi, X., Velarde-Miranda, C., Plonczynski, M. W., Parker, C. R., Rainey, W., et al. (2014). Development of Monoclonal Antibodies against Human CYP11B1 and CYP11B2. *Mol. Cell Endocrinol.* 383 (1-2), 111–117. doi:10.1016/j.mce.2013.11.022

Goupil, R., Wolley, M., Ahmed, A. H., Gordon, R. D., and Stowasser, M. (2015). Does Concomitant Autonomous Adrenal Cortisol Overproduction Have the Potential to Confound the Interpretation of Adrenal Venous Sampling in Primary Aldosteronism? *Clin. Endocrinol.* 83 (4), 456–461. doi:10.1111/cen.12750

Jiang, Y., Zhang, C., Wang, W., Su, T., Zhou, W., Jiang, L., et al. (2015). Diagnostic Value of ACTH Stimulation Test in Determining the Subtypes of Primary Aldosteronism. *J. Clin. Endocrinol. Metab.* 100 (5), 1837–1844. doi:10.1210/jc.2014-3551

Kozàkovà, M., Buralli, S., Palombo, C., Bernini, G., Moretti, A., Favilla, S., et al. (2003). Myocardial Ultrasonic Backscatter in Hypertension. *Hypertension* 41 (2), 230–236. doi:10.1161/01.hyp.0000052542.68896.2b

Lesage, F., Hibino, H., and Hudspeth, A. J. (2004). Association of -catenin with the -subunit of Neuronal Large-Conductance Ca2+-Activated K+ Channels. *Proc. Natl. Acad. Sci.* 101 (2), 671–675. doi:10.1073/pnas.0307681100

McCarty, K. S., Jr., Miller, L. S., Cox, E. B., Konrath, J., and McCarty-Sr, K. S. (1985). Estrogen Receptor Analyses. Correlation of Biochemical and Immunohistochemical Methods Using Monoclonal Antireceptor Antibodies. *Arch. Pathol. Lab. Med.* 109 (8), 716–721.

Monticone, S., Castellano, I., Versace, K., Lucatello, B., Veglio, F., Gomez-Sanchez, C. E., et al. (2015). Immunohistochemical, Genetic and Clinical Characterization of Sporadic Aldosterone-Producing Adenomas. *Mol. Cell Endocrinol.* 411 (1-2), 146–154. doi:10.1016/j.mce.2015.04.022

Mulatero, P., Monticone, S., Bertello, C., Viola, A., Tizzani, D., Iannaccone, A., et al. (2013). Long-term Cardio- and Cerebrovascular Events in Patients with Primary Aldosteronism. *J. Clin. Endocrinol. Metab.* 98 (12), 4826–4833. doi:10.1210/jc.2013-2805

Nakamura, Y., Maekawa, T., Felizola, S. J. A., Satoh, F., Qi, X., Velarde-Miranda, C., et al. (2014). Adrenal CYP11B1/2 Expression in Primary Aldosteronism: Immunohistochemical Analysis Using Novel Monoclonal Antibodies. *Mol. Cell Endocrinol.* 392 (1-2), 73–79. doi:10.1016/j.mce.2014.05.002

Nanba, K., Tsuiki, M., Sawai, K., Mukai, K., Nishimoto, K., Usui, T., et al. (2013). Histopathological Diagnosis of Primary Aldosteronism Using CYP11B2 Immunohistochemistry. *J. Clin. Endocrinol. Metab.* 98 (4), 1567–1574. doi:10.1210/jc.2012-3726

Nanba, K., Vaidya, A., Williams, G. H., Zheng, I., Else, T., and Rainey, W. E. (2017). Age-Related Autonomous Aldosteronism. *Circulation* 136 (4), 347–355. doi:10.1161/circulationaha.117.028201

Nishimoto, K., Nakagawa, K., Li, D., Kosaka, T., Oya, M., Mikami, S., et al. (2010). Adrenocortical Zonation in Humans under normal and Pathological Conditions. *J. Clin. Endocrinol. Metab.* 95 (5), 2296–2305. doi:10.1210/jc.2009-2010

Omata, K., Anand, S. K., Hovelson, D. H., Liu, C.-J., Yamazaki, Y., Nakamura, Y., et al. (2017). Aldosterone-Producing Cell Clusters Frequently Harbor Somatic Mutations and Accumulate with Age in Normal Adrenals. *J. Endocr. Soc.* 1 (7), 787–799. doi:10.1210/js.2017-00134

Omata, K., Satoh, F., Morimoto, R., Ito, S., Yamazaki, Y., Nakamura, Y., et al. (2018). Cellular and Genetic Causes of Idiopathic Hyperaldosteronism. *Hypertension* 72 (4), 874–880. doi:10.1161/HYPERTENSIONAHA.118.11086

Ono, Y., Nakamura, Y., Maekawa, T., Felizola, S. J. A., Morimoto, R., Iwakura, Y., et al. (2014). Different Expression of 11β-Hydroxylase and Aldosterone Synthase between Aldosterone-Producing Microadenomas and Macroadenomas. *Hypertension* 64 (2), 438–444. doi:10.1161/HYPERTENSIONAHA.113.02944

Rossi, G. P., Bernini, G., Caliumi, C., Desideri, G., Fabris, B., Ferri, C., et al. (2006). A Prospective Study of the Prevalence of Primary Aldosteronism in 1,125 Hypertensive Patients. *J. Am. Coll. Cardiol.* 48 (11), 2293–2300. doi:10.1016/j.jacc.2006.07.059

Rossi, G. P., Bolognesi, M., Rizzoni, D., Seccia, T. M., Piva, A., Porteri, E., et al. (2008). Vascular Remodeling and Duration of Hypertension Predict Outcome of Adrenalectomy in Primary Aldosteronism Patients. *Hypertension* 51 (5), 1366–1371. doi:10.1161/HYPERTENSIONAHA.108.111369

Scholl, U. I., Healy, J. M., Thiel, A., Fonseca, A. L., Brown, T. C., Kunstman, J. W., et al. (2015). Novel Somatic Mutations in Primary Hyperaldosteronism Are Related to the Clinical, Radiological and Pathological Phenotype. *Clin. Endocrinol.* 83 (6), 779–789. doi:10.1111/cen.12873

Staermose, S., Marwick, T. H., Gordon, R. D., Cowley, D., Dowling, A., and Stowasser, M. (2009). Elevated Serum Interleukin 6 Levels in Normotensive Individuals with Familial Hyperaldosteronism Type 1. *Hypertension* 53 (4), e31–32. doi:10.1161/HYPERTENSIONAHA.108.128512

Steichen, O., Zinzindohoué, F., Plouin, P.-F., and Amar, L. (2012). Outcomes of Adrenalectomy in Patients with Unilateral Primary Aldosteronism: a Review. *Horm. Metab. Res.* 44 (3), 221–227. doi:10.1055/s-0031-1299681

Stowasser, M., and Gordon, R. D. (2016). Primary Aldosteronism: Changing Definitions and New Concepts of Physiology and Pathophysiology Both inside and outside the Kidney. *Physiol. Rev.* 96 (4), 1327–1384. doi:10.1152/physrev.00026.2015

Teo, A. E. D., Garg, S., Haris Shaikh, L., Zhou, J., Karet Frankl, F. E., Gurnell, M., et al. (2015). Pregnancy, Primary Aldosteronism, and AdrenalCTNNB1Mutations. *N. Engl. J. Med.* 373 (15), 1429–1436. doi:10.1056/NEJMoa1504869

Tissier, F., Cavard, C., Groussin, L., Perlemoine, K., Fumey, G., Hagneré, A.-M., et al. (2005). Mutations of β-Catenin in Adrenocortical Tumors: Activation of the Wnt Signaling Pathway Is a Frequent Event in Both Benign and Malignant Adrenocortical Tumors. *Cancer Res.* 65 (17), 7622–7627. doi:10.1158/0008-5472.CAN-05-0593

Tsuchiya, K., Yoshimoto, T., and Hirata, Y. (2009). Endothelial Dysfunction Is Related to Aldosterone Excess and Raised Blood Pressure. *Endocr. J.* 56 (4), 553–559. doi:10.1507/endocrj.k09e-014

Williams, T. A., Lenders, J. W. M., Mulatero, P., Burrello, J., Rottenkolber, M., Adolf, C., et al. (2017). Outcomes after Adrenalectomy for Unilateral Primary Aldosteronism: an International Consensus on Outcome Measures and Analysis of Remission Rates in an International Cohort. *Lancet Diabetes Endocrinol.* 5 (9), 689–699. doi:10.1016/S2213-8587(17)30135-3

Wilmes, J., Haddad-Tóvolli, R., Alesutan, I., Munoz, C., Sopjani, M., Pelzl, L., et al. (2012). Regulation of KCNQ1/KCNE1 by β-catenin. *Mol. Membr. Biol.* 29 (3-4), 87–94. doi:10.3109/09687688.2012.678017

Wu, L., Xie, J., Jiang, L., Su, T., Ye, L., Zhou, W., et al. (2018). Feminizing Adrenocortical Carcinoma: The Source of Estrogen Production and the Role of Adrenal-Gonadal Dedifferentiation. *J. Clin. Endocrinol. Metab.* 103 (10), 3706–3713. doi:10.1210/jc.2018-00689

Wu, V.-C., Chueh, S.-C., Chang, H.-W., Lin, L.-Y., Liu, K.-L., Lin, Y.-H., et al. (2009). Association of Kidney Function with Residual Hypertension after Treatment of Aldosterone-Producing Adenoma. *Am. J. Kidney Dis.* 54 (4), 665–673. doi:10.1053/j.ajkd.2009.06.014

Yamazaki, Y., Nakamura, Y., Omata, K., Ise, K., Tezuka, Y., Ono, Y., et al. (2017). Histopathological Classification of Cross-Sectional Image Negative Hyperaldosteronism. *J. Clin. Endocrinol. Metab.* 102 (4), 2016–2986. doi:10.1210/jc.2016-2986

Zhao, C., Zhang, J., Li, K., Yang, J., Yu, H., Duan, S., et al. (2014). β-Catenin Regulates Membrane Potential in Muscle Cells by Regulating the α2 Subunit of Na,K-ATPase. *Eur. J. Neurosci.* 40 (1), 2216–2224. doi:10.1111/ejn.12564

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 Sun, Jiang, Xie, Zhu, Wu, Zhong, Zhou, Su 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fmolb.2021.719982

Identification of Mast Cell-Based Molecular Subtypes and a Predictive Signature in Clear Cell Renal Cell Carcinoma

*Hanxiang Liu and Yi Yang**

Pediatric Urology, Shengjing Hospital of China Medical University, Shenyang, China

Background: Kidney renal clear cell carcinoma (KIRC) is a common malignant tumor of the urinary system. Surgery is the preferred treatment option; however, the rate of distant metastasis is high. Mast cells in the tumor microenvironment promote or inhibit tumorigenesis depending on the cancer type; however, their role in KIRC is not well-established. Here, we used a bioinformatics approach to evaluate the roles of mast cells in KIRC.

Methods: To quantify mast cell abundance based on gene sets, a single-sample gene set enrichment analysis (ssGSEA) was utilized to analyze three datasets. Weighted correlation network analysis (WGCNA) was used to identify the genes most closely related to mast cells. To identify new molecular subtypes, the nonnegative matrix factorization algorithm was used. GSEA and least absolute shrinkage and selection operator (LASSO) Cox regression were used to identify genes with high prognostic value. A multivariate Cox regression analysis was performed to establish a prognostic model based on mast cell-related genes. Promoter methylation levels of mast cell-related genes and relationships between gene expression and survival were evaluated using the UALCAN and GEPIA databases.

Results: A prolonged survival in KIRC was associated with a high mast cell abundance. KIRC was divided into two molecular subtypes (cluster 1 and cluster 2) based on mast cell-related genes. Genes in Cluster 1 were enriched for various functions related to cancer development, such as the TGFβ signaling pathway, renal cell carcinoma, and mTOR signaling pathway. Based on drug sensitivity predictions, sensitivity to doxorubicin was higher for cluster 2 than for cluster 1. By a multivariate Cox analysis, we established a clinical prognostic model based on eight mast cell-related genes.

Conclusion: We identified eight mast cell-related genes and constructed a clinical prognostic model. These results improve our understanding of the roles of mast cells in KIRC and may contribute to personalized medicine.

Keywords: renal clear cell carcinoma, mast cell, WGCNA, TCGA, arrayexpress, ICGC, clinical prognostic model

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong, SAR China

Reviewed by:
Yanqing Li,
Shandong University, China
Ibrahim Y. Hachim,
University of Sharjah, United Arab Emirates

***Correspondence:**
Yi Yang
yangyilab@163.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *03 June 2021*
Accepted: *13 September 2021*
Published: *27 September 2021*

Citation:
Liu H and Yang Y (2021) Identification of Mast Cell-Based Molecular Subtypes and a Predictive Signature in Clear Cell Renal Cell Carcinoma. Front. Mol. Biosci. 8:719982. doi: 10.3389/fmolb.2021.719982

INTRODUCTION

Clear cell renal cell carcinoma (KIRC) accounts for approximately 65–70% of all renal cell carcinomas (Warren and Harrison, 2018; Siegel et al., 2020). Metastasis is the main cause of death in patients with KIRC (Li et al., 2018). The early clinical features are not obvious and are difficult to identify. Therefore, some patients with KIRC have metastases when they are first diagnosed (Motzer et al., 1996). Although surgical treatment achieves good results, the 5-years survival rate for patients with metastatic KIRC is still low (Heidenreich et al., 2012; Sara et al., 2016). Studies on immune checkpoint inhibitors have made significant advances for KIRC treatment; however, the response to immunotherapy in patients with KIRC varies greatly across individuals (Fang et al., 2020). Therefore, it is necessary to identify therapeutic targets and effective predictors for early diagnosis and treatment.

Mast cells are one of the main components of the tumor immune microenvironment. The mast cell density is elevated in various types of tumors (Marone et al., 2016). Mast cells can be attracted by chemotactic molecules produced by tumor cells, thus producing a variety of angiogenic and lymphangiogenic factors, thereby contributing to tumor growth and metastasis (Boesiger et al., 1998; Abdel-Majid and Marshall, 2004; Taskinen et al., 2008; Detoraki et al., 2010; Melillo et al., 2010; Theoharides et al., 2010; Sismanopoulos et al., 2012). While many studies have demonstrated that mast cells can promote tumor development, others have shown that mast cells have tumor-inhibitory effects (Dabiri et al., 2004; Amini et al., 2007). A poor prognosis in KIRC has been linked to the existence of mast cells (Hiroshi et al., 1999; Melillo et al., 2010; Strouch et al., 2010; David et al., 2011; Rao et al., 2016). Previous studies have shown that mast cells can be used as targets for immunotherapy of solid tumors (Oldford and Marshall, 2015). Beuselinck et al. used unsupervised transcriptome analysis to identify four robust KIRC subtypes that were associated with different responses to sunitinib treatment (Beuselinck et al., 2015). Zhao et al. classified KIRC in the Chinese population into three classes based on gene expression, which provides practical guidelines on clinical treatment of patients with KIRC (Zhao et al., 2020). However, few studies have examined the role of mast cells in KIRC and the molecular mechanisms underlying their effects.

In this study, we used a bioinformatics approach to evaluate the prognostic value of mast cell-related genes in KIRC. In particular, we used a single-sample gene set enrichment analysis (ssGSEA) to quantify mast cell abundance in three KIRC datasets. Then, a series of statistical analyses, including a univariate Cox regression analysis, weighted correlation network analysis (WGCNA), GSEA, Least Absolute Shrinkage and Selection Operator (LASSO) Cox analysis, and Kaplan–Meier survival analysis, were performed to identify mast cell-related genes that may regulate the development of KIRC and to develop clinical prognostic models. These results will improve our understanding of the role of mast cells in KIRC and provide a basis for personalized treatment.

TABLE 1 | Univariate Cox regression analysis of fibroblast abundance in The Cancer Genome Atlas (TCGA), E-MTAB-1980, and International Cancer Genome Consortium (ICGC) cohorts.

Datasets	HR	HR.95 L	HR.95H	*p*-value
TCGA	0.004201883	0.000320462	0.055094827	3.08E-05
E-MTAB-1980	4.95E-05	2.81E-08	0.087285488	0.009339429
ICGC	0.164918533	0.000142324	191.0997492	0.616586053

METHODS

Processing of KIRC Patient Data Set

Clinical information and KIRC transcriptome sequencing data were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/repository), including data for 539 KIRC and 72 normal cases. The E-MTAB-1980 dataset ($n = 101$) was downloaded from the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/). Similarly, sample information for KIRC ($n = 91$) was downloaded from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/). Immune-related genes were derived from the Immunology database and Analysis Portal (ImmPort) database (https://www.immport.org/home). A mast cell gene set (**Supplementary Table S1**) was obtained from a previous study (Bindea et al., 2013).

Quantification of Mast Cell Abundance

Mast cell abundance was quantified in three bladder datasets using ssGSEA based on the mast cell gene set using the GSVA R package (Hänzelmann et al., 2013).

Identification of Mast Cell-Related Genes and Molecular Subtypes

WGCNA was performed using the R package "WGCNA" (Langfelder and Horvath, 2008) to identify highly correlated gene modules among samples and these modules were used for subsequent analyses. WGCNA was based on 1670 immune-related genes from TCGA-KIRC, and the relationships between single genes and mast cell density were quantified by gene significance. Module membership was evaluated as the correlation between the gene expression profiles and module characteristic genes. The total number of non-gray modules was eight. The brown module was most highly correlated with the mast cell density ($r = 0.58$, p = 9e-46). This module contained 258 mast cell-related genes. Among the 258 mast cell-related genes, 250 were consistently found in all ArrayExpress and the International Cancer Genome Consortium (ICGC) datasets and were used for an non-negative matrix factorization (NMF) clustering analysis. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the 250 genes were performed using the clusterProfiler package in R (Yu et al., 2012). We identified functional pathways related to cluster 1 in TCGA dataset and used h.all.v7.1.symbols.gmt as the reference gene set for GSEA. The analysis was performed using 1000 permutations with a <0.05 false discovery rate (FDR) as the screening threshold, and GSEA version 4.0.1. ESTIMATE (Yoshihara et al., 2013) and

CIBERSORT (Newman et al., 2019) algorithms were used to explore the relationship between molecular subtypes and tumor immune microenvironment.

Chemotherapeutic Response and Immunotherapeutic Response Prediction

The responses to doxorubicin and sunitinib, two commonly used chemotherapeutic drugs, were predicted for each sample according to the Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/) using the R package "pRRophetic." Based on 10-fold cross-validation of the GDSC training set, the prediction accuracy was evaluated, and ridge regression was used to estimate the IC_{50} values for the samples. Repeated gene expression estimates were summarized as an average value, and default values were used for all parameters setting the tissue type to "allSolidTumours" and using "combat" for batch effect removal (Geeleher et al., 2014). All parameters were set to the default values. We then compared the TCGA KIRC expression profile of cluster 1 and 2 with another published dataset that contained the data of 47 patients with melanoma who responded to immunotherapies using subclass mapping method (https://cloud.genepattern.org/gp) (Lu et al., 2019).

Construction and Verification of Clinical Prognostic Model Based on Mast Cell-Related Genes

To identify the mast cell-related genes most closely related to prognosis, a Cox regression analysis with the LASSO penalty was performed using the R package "glmnet." To construct the optimal clinical prognostic model of mast cell-related genes, a multiple regression analysis was used. We developed the formula for the risk score as follows: Risk Score = coef (gene 1) × expr (gene 1) + coef (gene 2) × expr (gene 2) + coef (gene 3) × expr (gene 3) + ... + coef (gene N) × expr (gene N). Patients were classified into a high-risk and low-risk group according to the median value of the risk scores of all samples in each dataset.

Survival Analysis and Methylation Analysis

To evaluate the relationship between survival and the expression of eight genes in the model, the GEPIA database (http://gepia.cancer-pku.cn/) was used (Tang et al., 2017). The UALCAN database (http://ualcan.path.uab.edu/) was used to analyze the promoter methylation levels of genes used to construct a clinical prognostic model (Chandrashekar et al., 2017).

Statistical Analyses

Differences in overall survival (OS) among groups were compared using the Kaplan-Meier analysis and log-rank test. A multi-time receiver operating characteristic (ROC) analysis and area under the curve (AUC) were used to evaluate signal specificity and sensitivity. R (version 4.0.2) was used for all statistical analyses. Statistical significance was set at $p < 0.05$.

RESULTS

Mast Cell Abundance is Beneficial for Survival of Patients With KIRC

To quantify mast cell abundance based on a mast cell gene set in three KIRC datasets (TCGA, ArrayExpress, and ICGC cohorts), ssGSEA was used. A univariate Cox analysis showed that the mast cell gene set was a protective factor for KIRC (**Table 1**). In addition, we divided the samples in the three data sets into low abundance and high abundance groups based on ssGSEA scores. A high abundance of mast cells was beneficial for the survival of patients with KIRC (**Figures 1A–C**).

Identification of Mast Cell-Related Genes and Molecular Subtypes

To identify genes related to mast cells, WGCNA was used. The genes were clustered into eight modules (**Figure 1D**). As determined by Pearson's correlation coefficients (**Figure 1E**), the brown module was most highly correlated with mast cell abundance (r: 0.58, p = 9e-46). A functional enrichment analysis showed that the genes in the brown module are enriched for the following GO terms: regulation of epithelial cell proliferation, epithelial cell proliferation, regulation of chemotaxis, receptor ligand activity, and signaling receptor activator activity (**Figure 1F**). A KEGG pathway analysis showed that the mast cell-related genes were involved in the MAPK signaling pathway, Rap 1 signaling pathway, cytokine–cytokine receptor interaction, and PI3K-Akt signaling pathway (**Figure 1G**). Furthermore, to obtain survival-related mast cell-related genes, we used a univariate Cox regression analysis. Among 250 mast cell-related genes, 103 were related to survival with a threshold of $p < 0.05$ (**Supplementary Table S2**). An NMF clustering analysis divided these 103 genes in TCGA-KIRC into two molecular subtypes (Cluster 1 and 2) with different molecular and clinical characteristics. **Figure 2A** shows a heatmap of expression differences between clusters 1 and 2 in the TCGA cohort. The immune score for cluster 2 was significantly higher than that of cluster 1 ($p < 0.05$), with no significant differences in the stromal score and tumor purity between the two subgroups (**Figures 2B–D**). We also divided ArrayExpress-KIRC and ICGC-KIRC data into cluster 1 and cluster 2. **Figure 3A** shows a heatmap for the two subgroups in the ArrayExpress cohort. The stromal score for the cluster 1 subgroup was higher than that for the cluster 2 subgroup (**Figure 3B**). The tumor purity was lower for the cluster 1 subgroup than the cluster 2 subgroup (**Figure 3C**). There was no significant difference in immune scores between the two subgroups (**Figure 3D**). **Figure 4A** shows a heatmap for the two subgroups in the ICGC cohort; the immune score, stromal score, and tumor purity did not differ significantly between cluster 1 and cluster 2 (**Figures 4B–D**). Differences in the tumor immune microenvironment were observed between the two molecular subtypes in the three KIRC cohorts (**Figures 5A–C**). As determined by a Kaplan-Meier analysis, survival time was longer in cluster 1 than in cluster 2 (TCGA, $p < 0.001$;

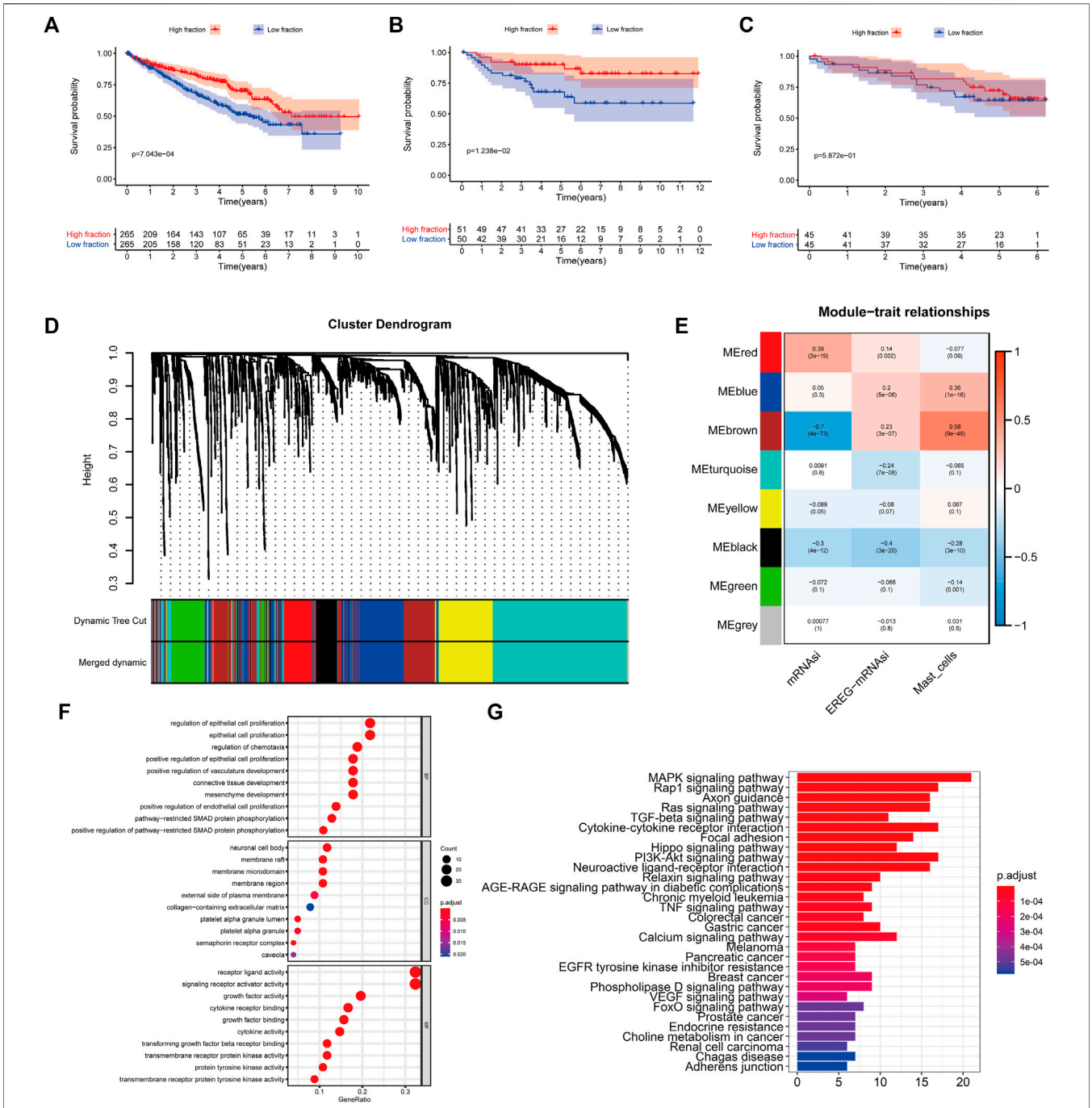


FIGURE 1 | (A–C) Kaplan–Meier curves for patients with bladder cancer (BLCA) showed that in the six cohorts, patients with a low fibroblast abundance have a better prognosis than that of patients with a high fibroblast abundance [**(A)**: The Cancer Genome Atlas (TCGA); **(B)**: E-MTAB-1980; **(C)**: International Cancer Genome Consortium (ICGC)] **(C)** Using weighted correlation network analysis (WGCNA), eight modules were identified. **(D)** The brown module was most highly correlated with mast cells (cor: 0.58, p = 9e-46). **(E, F)** Functional enrichment analysis of 258 mast cell-related genes.

ArrayExpress cohort, $p < 0.001$) (**Figures 5D–F**). A GSEA showed that cluster 1 is enriched for the upregulation of inositol phosphate metabolism, adipocytokine signaling pathway, endocytosis, phosphatidylinositol signaling system, TGFβ signaling pathway, renal cell carcinoma, mTOR signaling pathway, vasopressin regulated water reabsorption, fatty acid metabolism, leukocyte transendothelial migration, and focal adhesion (**Figure 6**).

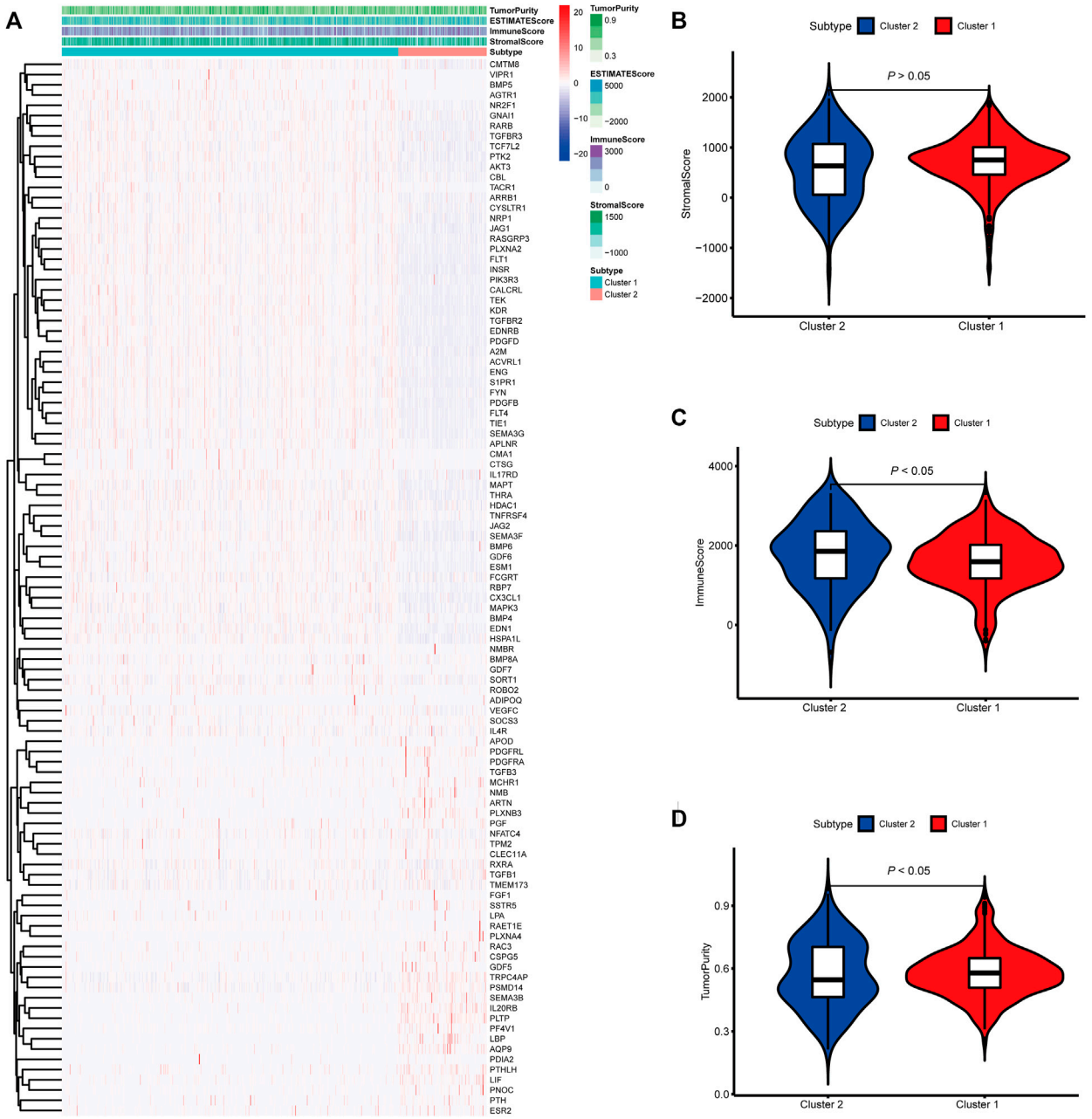


FIGURE 2 | Molecular subtypes identified based on mast cell-related genes in The Cancer Genome Atlas (TCGA) cohort. **(A)** Heatmap of differences between cluster 1 and cluster 2. **(B–D)** Differential analyses of the immune score, stromal score, and tumor purity between cluster 1 and cluster 2.

Cluster 2 is More Sensitive to Immuno- and Chemotherapies

To predict the response to immunotherapy, subclass mapping was applied to compare the expression profiles of the two KIRC subtypes with a published dataset for patients with melanoma treated by immunotherapy (Ro et al., 2017). In the TCGA cohort, cluster 2 was more likely to respond to anti-CTLA4 treatment (p = 0.010). However, based on corrected p-values, cluster 2 was not sensitive to CTLA4-R (**Figure 7A**). We also used GDSC data to predict the IC_{50} of doxorubicin for cluster 1 and cluster 2 in the three cohorts. Sensitivity to doxorubicin was significantly higher for cluster 2 than for cluster 1 (TCGA cohort, p = 0.016; ArrayExpress cohort, p = 0.003; ICGC cohort, p = 0.0002) (**Figures 7B–D**).

Construction of a Clinical Prognostic Model Based on Mast Cell-Related Genes

Based on 103 mast cell-related genes, we used the LASSO Cox regression algorithm to identify 46 genes with high

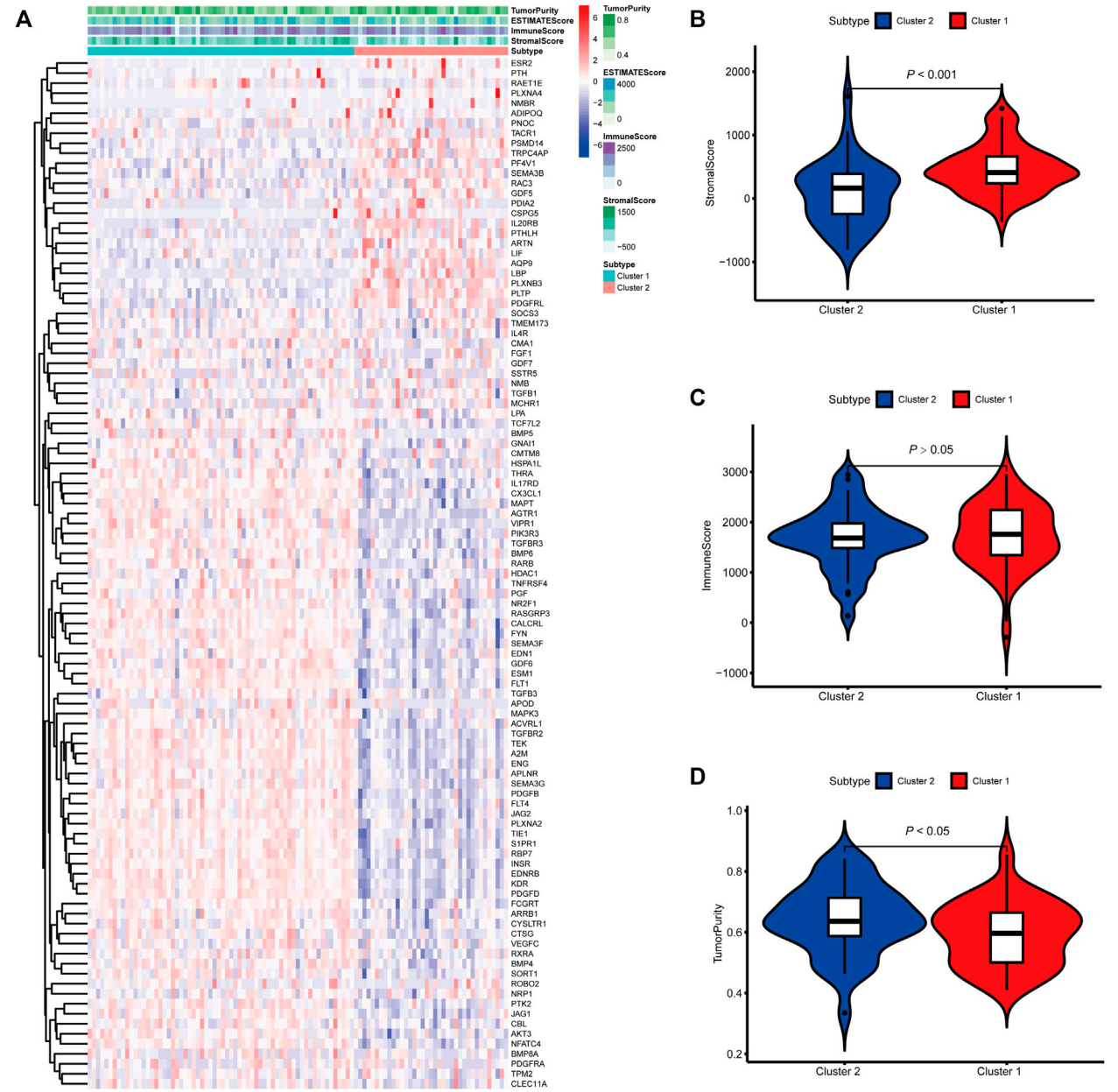


FIGURE 3 | Molecular subtypes identified based on mast cell-related genes in the E-MTAB-1980 cohort. **(A)** Heatmap of differences between cluster 1 and cluster 2. **(B–D)** Differential analyses of the immune score, stromal score, and tumor purity between cluster 1 and cluster 2.

prognostic value using $p < 0.01$ as a threshold (**Supplementary Table S3**). Finally, by a multivariate Cox regression analysis, we identified eight genes for the construction of a clinical prognostic model of mast cell-related genes (**Table 2**). The risk score was calculated as follows: [TRPC4AP expression level × (0.0286)] + [TEK expression level × (−0.0636)] + [IL17RD expression level × (−0.1686)] + [PTH expression level × (1.4582)] + [PDIA2 expression level × (0.1083)] + [SOCS3 expression level × (0.0053)] + [FCGRT expression level × (−0.0117)] + [GDF5 expression level × (0.8070)].

Application of the Prognostic Model to Patients With KIRC

At a ratio of 1:1, patients with KIRC in the TCGA cohort were divided into training and test sets. Based on the risk score calculated from the clinical prognostic model based on mast

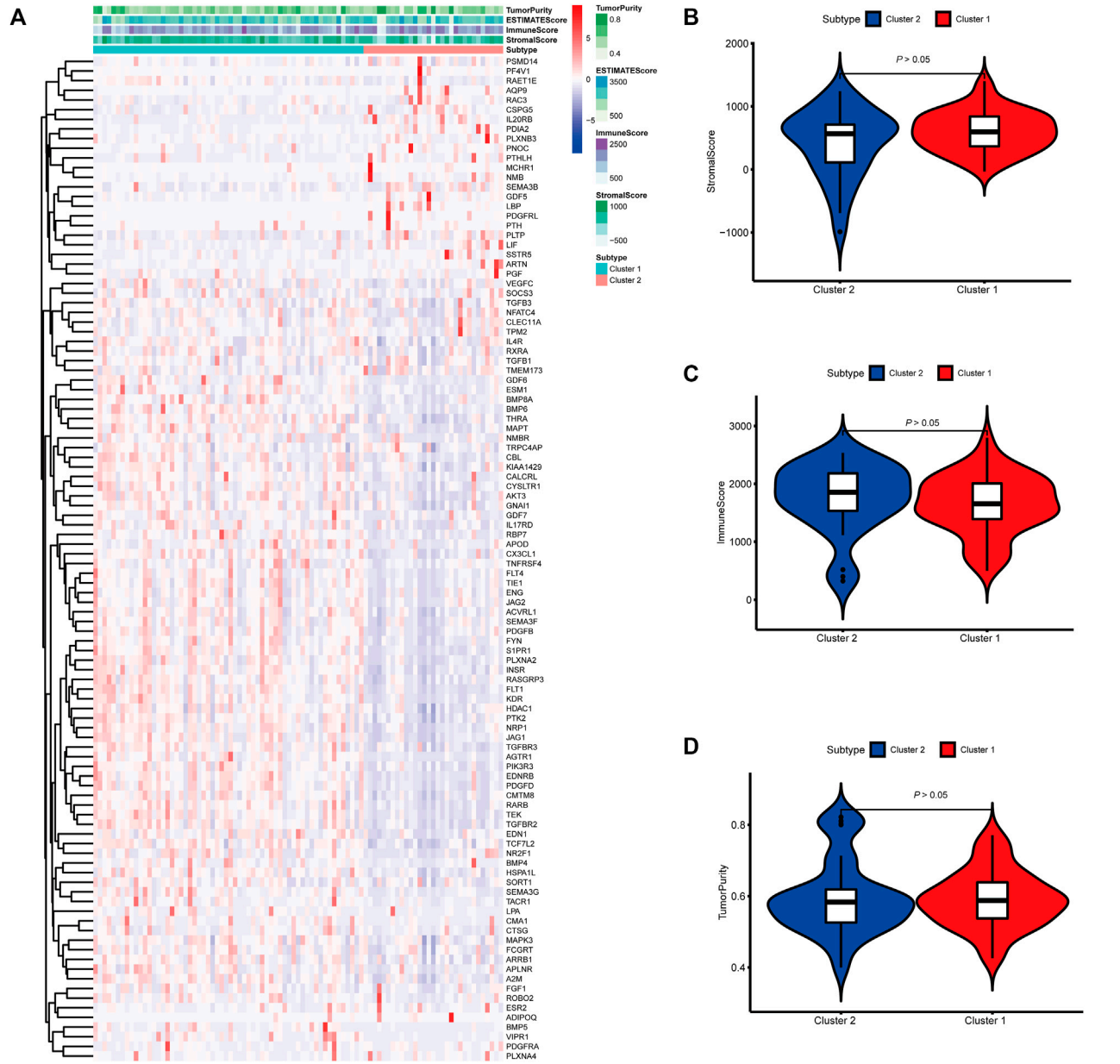


FIGURE 4 | Molecular subtypes identified based on mast cell-related genes in the International Cancer Genome Consortium (ICGC) cohort. **(A)** Heatmap of differences between cluster 1 and cluster 2. **(B–D)** Differential analyses of the immune score, stromal score, and tumor purity between cluster 1 and cluster 2.

cell-related genes, patients with KIRC were divided into high-risk and low-risk groups (**Figure 8B**). As the risk score increased, the survival time of patients with KIRC decreased (**Figure 8G**). A time-dependent ROC curve analysis supported the predictive value of the model. These results show that our prognostic indicators have a good performance (**Figure 8L**).

Validation of the Clinical Prognostic Model

To determine the reliability of the clinical prognostic model across populations, we applied the formula to the TCGA cohort, TCGA testing cohort, ArrayExpress cohort, and ICGC cohort, yielding similar results to those obtained for the training set. Patients with KIRC were divided into high-risk or low-risk groups based on the risk score calculated from the model (**Figure 8A, C–E**). A lower risk was associated with a longer survival time (**Figure 8B, H–J**). Additionally, we verified the predictive accuracy of the clinical prognostic model in a joint analysis of the TCGA cohort, TCGA testing group, ArrayExpress cohort, and ICGC cohort (**Figure 8C, M–O**). Therefore, the newly

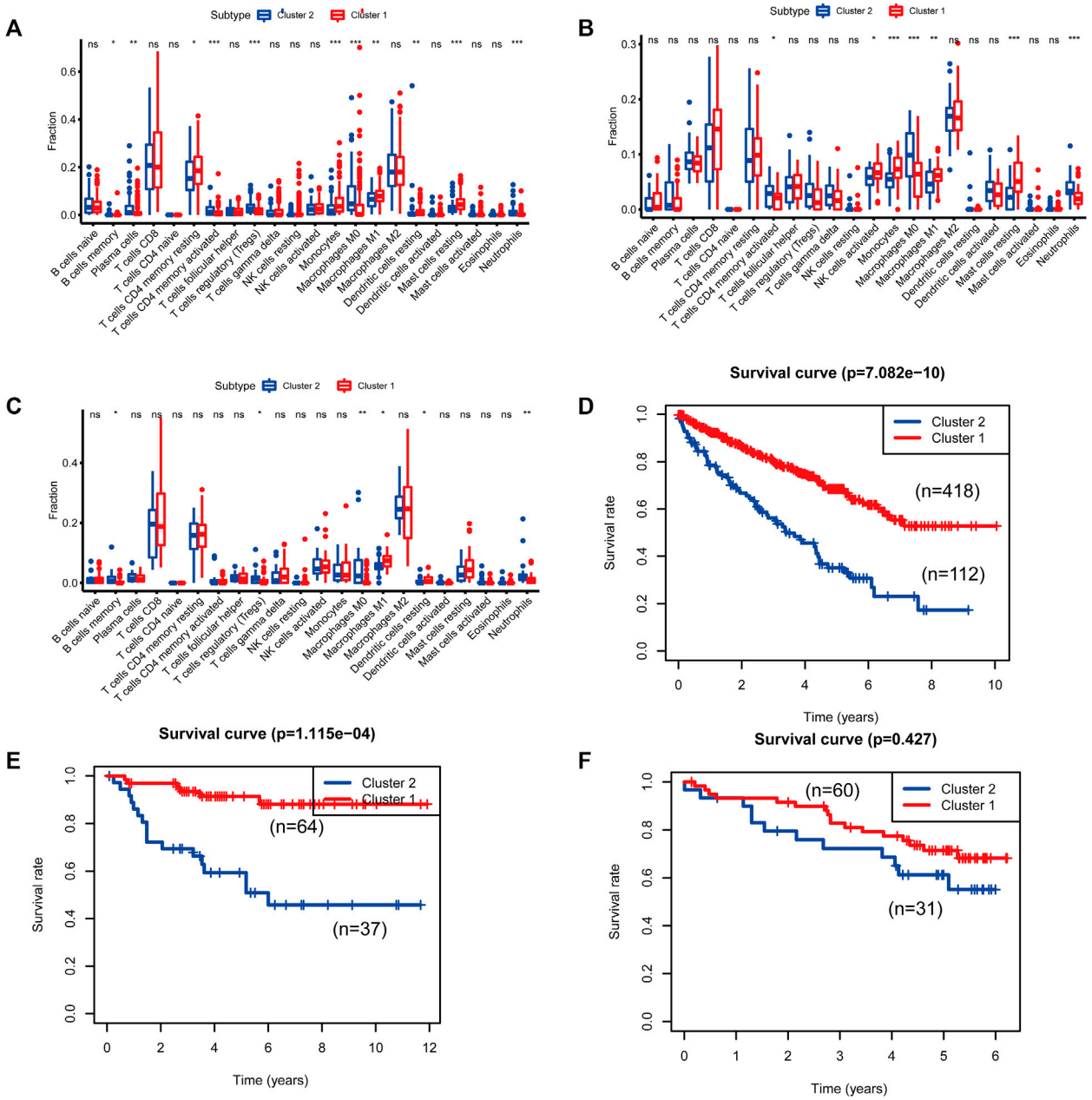


FIGURE 5 | Differences in immune cell populations and survival between the two molecular subtypes of cluster 1 and cluster 2. [**(A, D)**: The Cancer Genome Atlas (TCGA); **(B, E)**: E-MTAB-1980; **(C, F)**: International Cancer Genome Consortium (ICGC)].

developed clinical prognostic model is generalizable to different populations.

Survival Analysis and Methylation Analysis of Eight Genes Included in the Prognostic Model

To analyze the correlation between OS in patients with KIRC and the expression of the eight mast cell genes included in the model, the GEPIA database was utilized. OS was better for patients with low TEK expression than with high TEK expression ($p < 0.001$). Patients with high TEK expression had better disease-free survival (DFS) than that of patients with low TEK expression ($p = 0.00043$). Patients with high IL17RD expression had better OS than that of patients with low IL17RD expression ($p < 0.001$). Patients with high IL17RD expression had better OS than that of patients with low IL17RD expression ($p = 0.00045$). Patients with high FCGRT expression had better OS than that of patients with

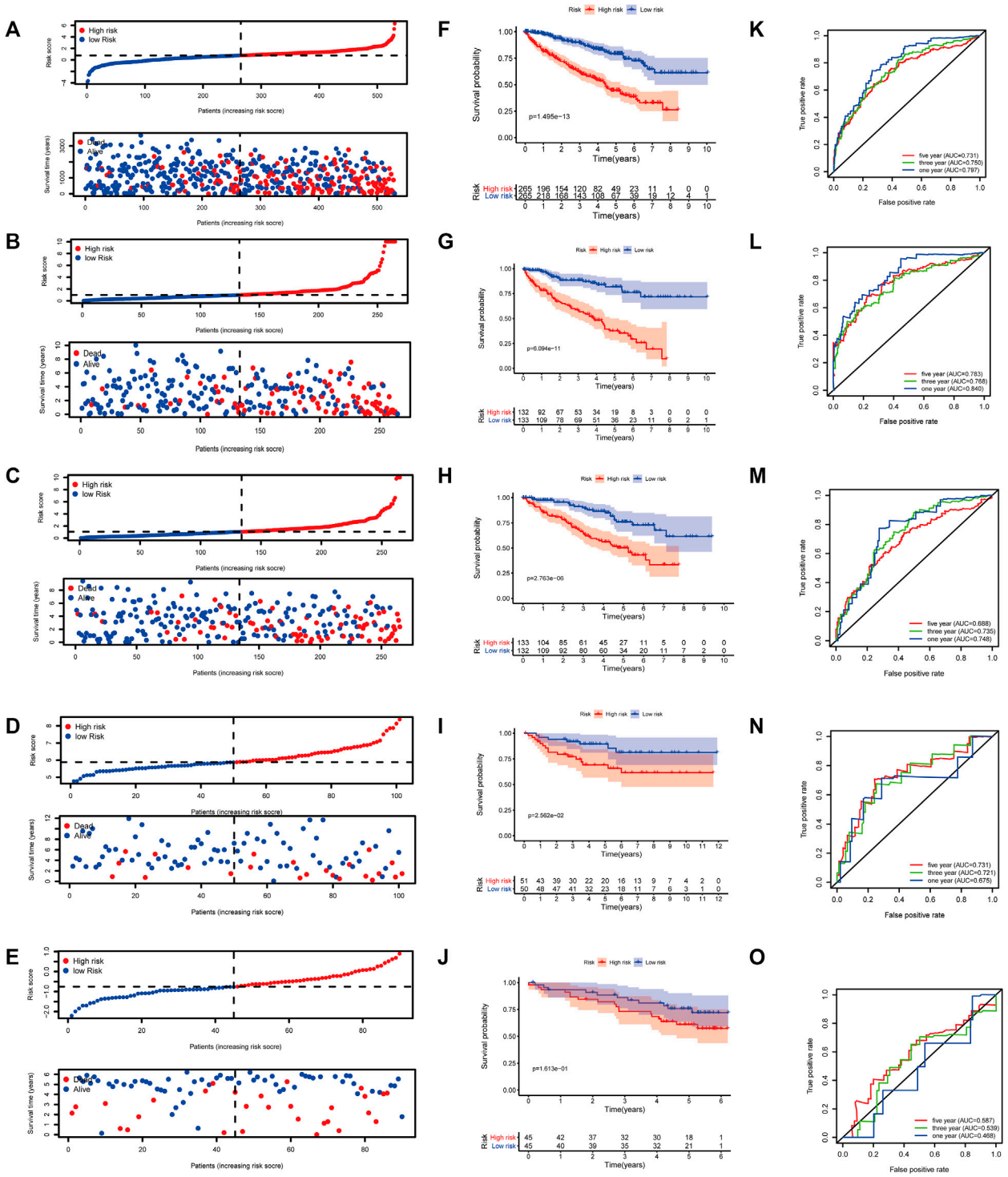


FIGURE 6 | Gene set enrichment analysis (GSEA) of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway differences between cluster 1 and cluster 2. [**(A)**: inositol phosphate metabolism, **(B)**: adipocytokine signaling pathway, **(C)**: endocytosis, **(D)**: phosphatidylinositol signaling system, **(E)**: TGFβ signaling pathway, **(F)**: renal cell carcinoma, **(G)**: mTOR signaling pathway, **(H)**: vasopressin regulated water reabsorption, **(I)**: fatty acid metabolism, **(J)**: leukocyte transendothelial migration, **(K)**: focal adhesion]. NES: normalized enrichment score.

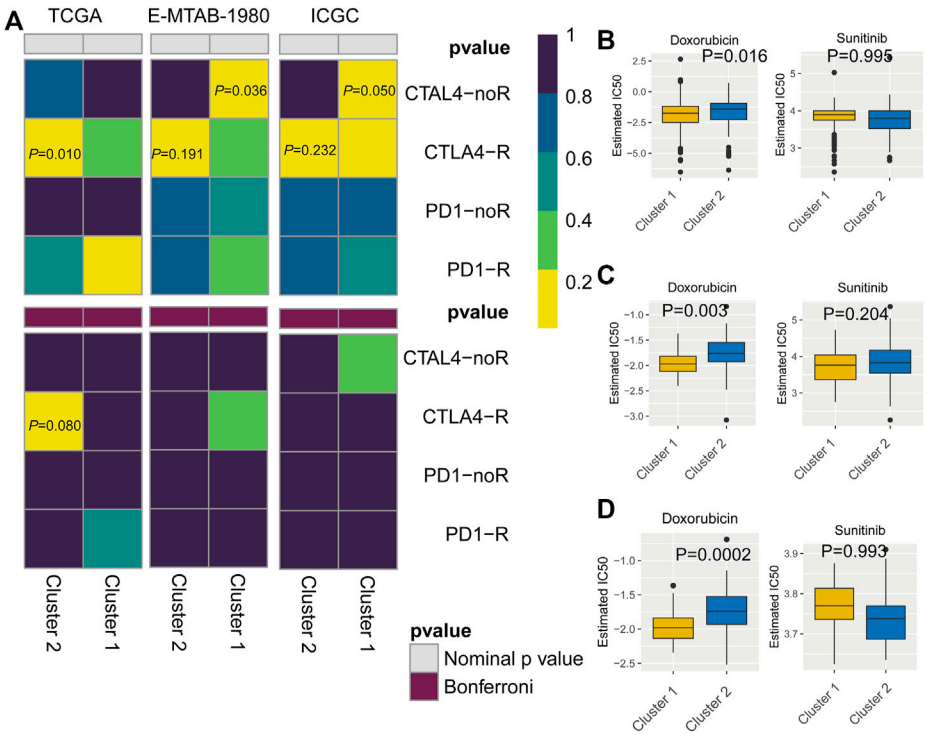


FIGURE 7 | (A) Subclass mapping analysis showed that cluster 2 is sensitive to CTLA4-R. [The Cancer Genome Atlas (TCGA): PCTLA4-R = 0.010; E-MTAB-1980: PCTLA4-R = 0.191; International Cancer Genome Consortium (ICGC): PCTLA4-R = 0.232] Based on corrected *p*-values, cluster 1 is not sensitive to CTLA4-R. **(B)** Box plot of estimated IC_{50} values for sunitinib and doxorubicin in cluster 1 and cluster 2. [**(A)**: TCGA; **(B)**: E-MTAB-1980; **(C)**: International Cancer Genome Consortium (ICGC)].

TABLE 2 | Multivariate Cox regression analysis of genes related to mast cells used to construct the model.

Gene	Coefficient	HR	HR.95 L	HR.95H	*p*-value
TRPC4AP	0.02857631	1.02898853	1.00929749	1.04906374	0.00374682
TEK	−0.0635723	0.93840629	0.90164062	0.97667113	0.00182356
IL17RD	−0.1686320	0.84481971	0.75159926	0.94960225	0.00470115
PTH	1.45824552	4.29841145	0.82061110	22.5153437	0.08435432
PDIA2	0.10826512	1.11434314	0.99652289	1.24609344	0.05758127
SOCS3	0.00534328	1.00535758	1.00240904	1.00831479	0.00036303
FCGRT	−0.0117308	0.98833770	0.97729089	0.99950937	0.04080188
GDF5	0.80699657	2.24116669	1.40590172	3.57267371	0.00069420

low FCGRT expression ($p < 0.001$). In contrast, patients with low PDIA2 expression had better OS than that of patients with high PDIA2 expression ($p < 0.001$). Patients with low PDIA2 expression had a better DFS than that of patients with high PDIA2 expression ($p < 0.001$). Patients with low SOCS3 expression had better OS than that of patients with high SOCS3 expression ($p = 0.00013$). Patients with low GDF5 expression had better OS than that of patients with high GDF5 expression ($p = 0.00013$) (**Figure 9**). In addition, we analyzed the promoter methylation levels of eight genes using the UALCAN database. The promoter methylation levels of FCGRT ($p < 0.001$), PDIA2 ($p < 0.001$), PTH ($p < 0.001$), and TRPC4AP ($p < 0.001$) were decreased in KIRC and those of GDF5 ($p < 0.001$) and TEK ($p = 0.024$) were increased in KIRC (**Figure 10**).

DISCUSSION

KIRC is the third most common malignant tumor of the urinary system. In 2020, it accounted for an estimated 14,830 deaths, with approximately 73,750 new cases in the United States (Siegel et al., 2020). Surgery remains the best treatment option. However, most patients eventually develop distant metastases (Rao et al., 2018). At present, radiotherapy, chemotherapy, targeted therapy, and immunotherapy are not effective in KIRC (Lalani et al., 2018).

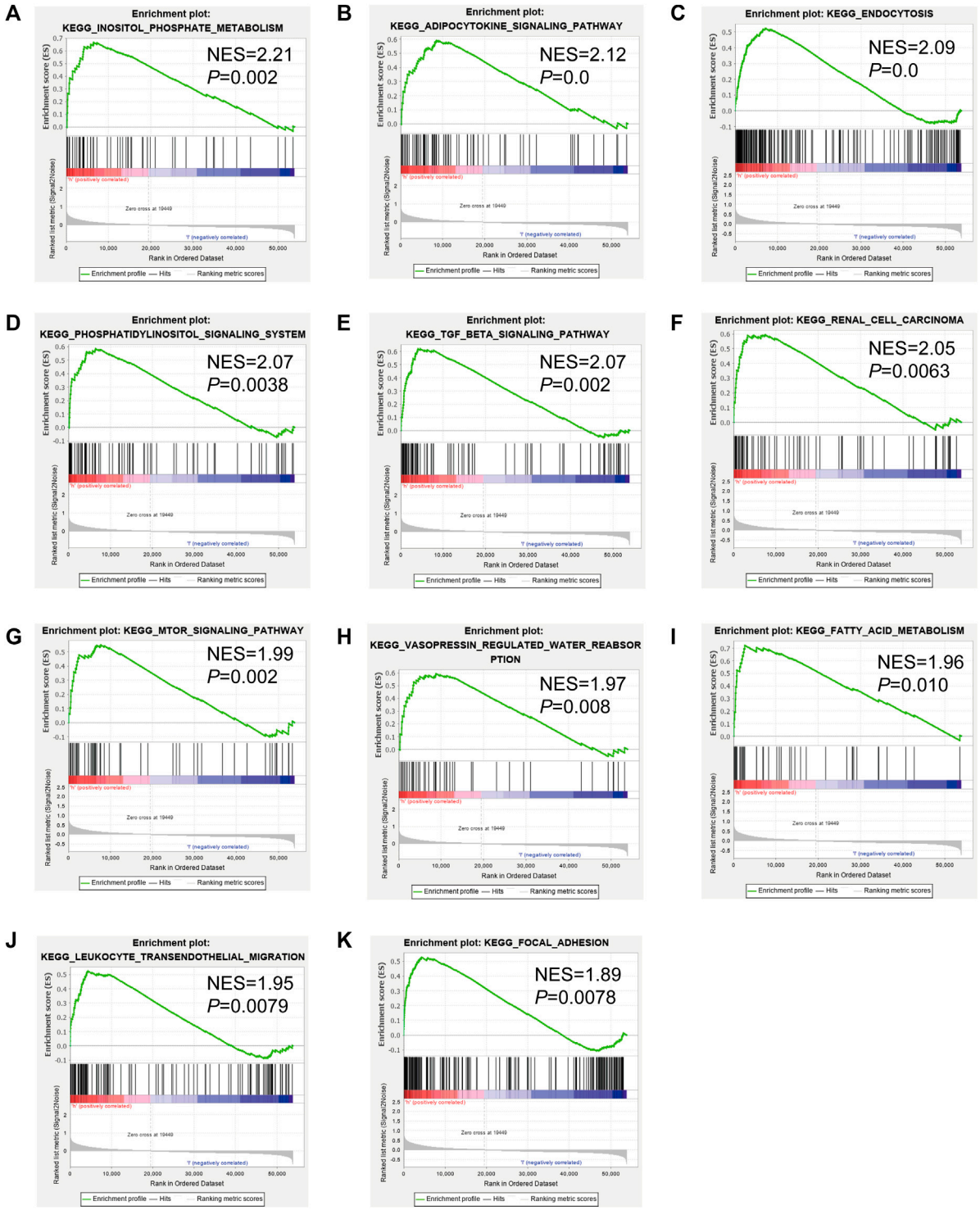


FIGURE 8 | (A–E) Distribution of patients according to the risk index. **(F–J)** Risk score calculated from the clinical prognostic model can predict survival. **(K–O)** Receiver operating characteristic (ROC) curve to verify the prognostic value of the model. [**(A, F, K)**: The Cancer Genome Atlas (TCGA); **(B, G, L)**: TCGA training group; **(C, H, M)**: TCGA testing group; **(D, I, N)**: E-MTAB-1980; **(E, J, O)**: International Cancer Genome Consortium (ICGC)].

Therefore, a deeper understanding of the molecular mechanisms underlying KIRC is necessary for development of effective early diagnostic methods and prognostic markers.

Mast cells are important components of the immune microenvironment of tumor tissues and can promote or inhibit tumorigenesis by releasing various factors (Varricchi et al., 2017).

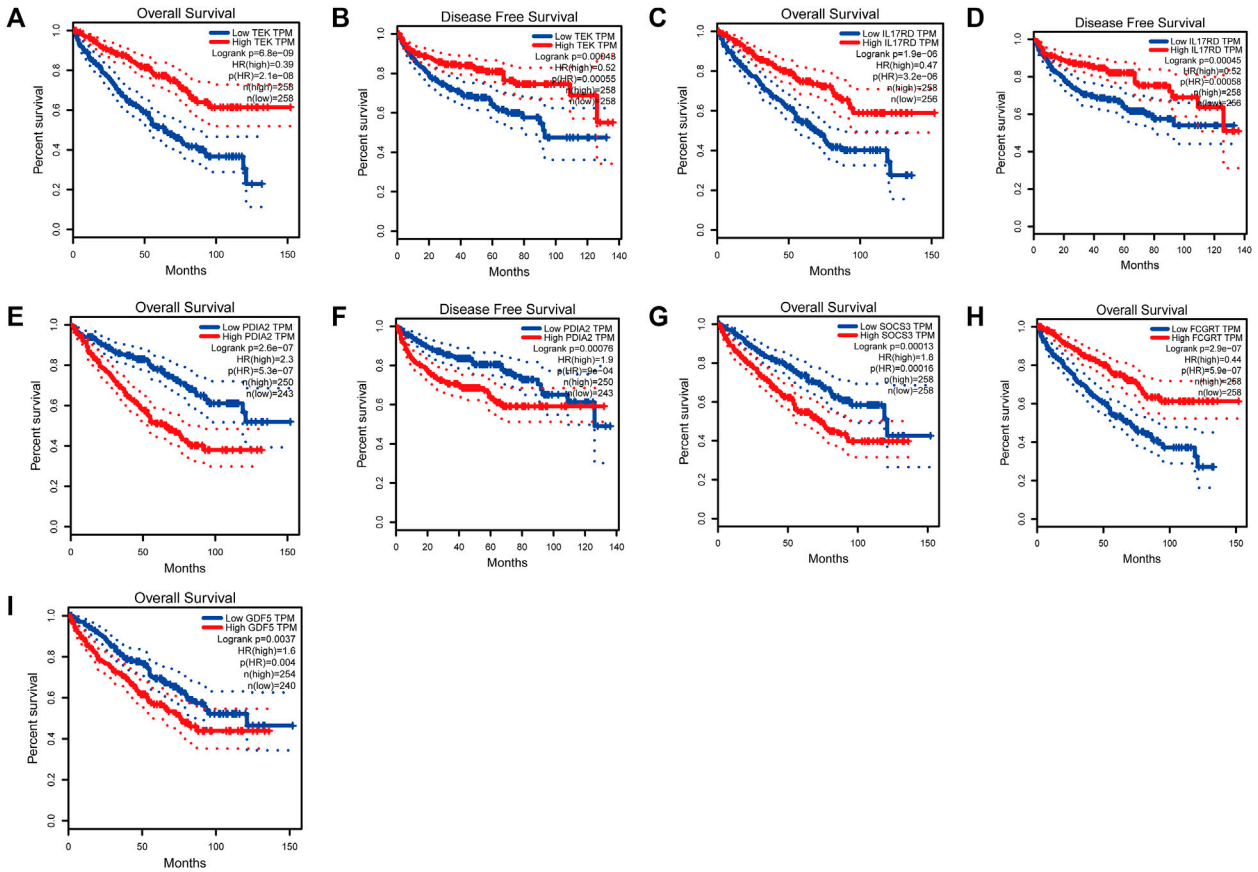


FIGURE 9 | Analysis of overall survival (OS) and disease-free survival (DFS) by the application of the constructed model based on mast cell-related genes of patients with kidney renal clear cell carcinoma (KIRC). (TEK: $p_{OS} < 0.001$, $p_{DFS} = 0.00043$; IL17RD: $p_{OS} < 0.001$, $p_{DFS} = 0.00045$; FCGRT: $p_{OS} < 0.001$; PDIA2: $p_{OS} < 0.001$, $p_{DFS} = 0.00076$; SOCS3: $p_{OS} = 0.00013$; GDF5: $p_{OS} = 0.00013$).

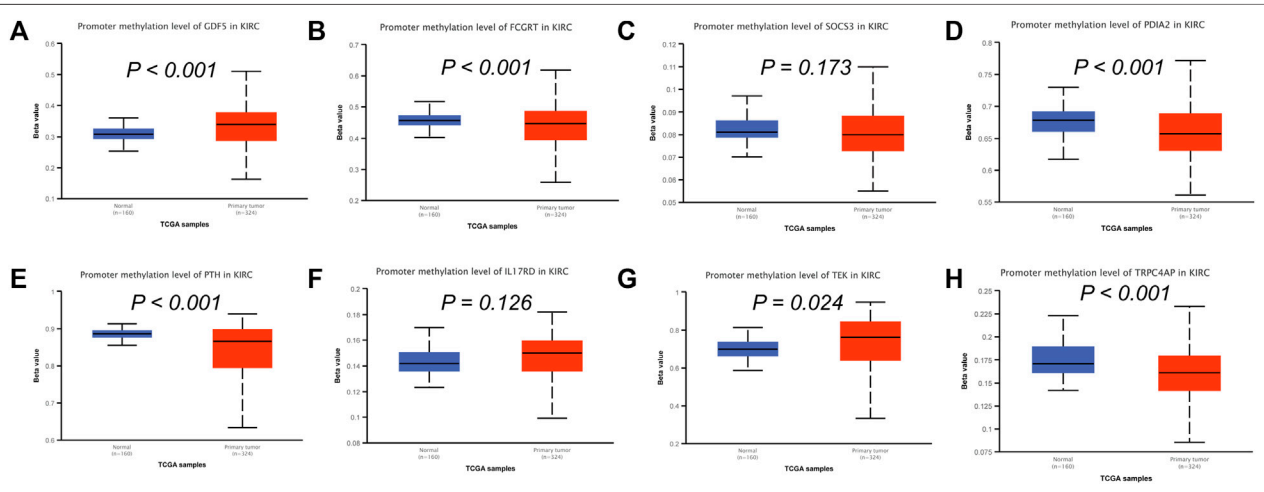


FIGURE 10 | Analysis of promoter methylation levels of mast cell-related genes in the constructed model. **(A–H)** GDF5: $p < 0.001$; SOCS3: $p = 0.173$; FCGRT: $p < 0.001$; PDIA2: $p < 0.001$; PTH: $p < 0.001$; IL17RD: $p = 0.126$; TEK: $p = 0.024$; TRPC4AP: $p < 0.001$.

We quantified the abundance of mast cells in three cohorts and found that the mast cell gene set used in the study is a protective factor in KIRC. We identified mast cell genes that were most closely related to the ssGSEA score by WGCNA. A functional enrichment analysis showed that mast cell-related genes are related to epithelial cell proliferation regulation, epithelial cell proliferation, chemotaxis

regulation, receptor ligand activity, signal receptor activator activity, MAPK signaling pathway, Rap1 signaling pathway, cytokine receptor interaction, and PI3K/Akt signaling pathway. Furthermore, after screening for survival-related mast cell-related genes, we divided KIRC into two molecular subtypes, cluster 1 and cluster 2, based on these genes. Predictive analyses of the responses to immuno-chemotherapies indicated that cluster 2 is sensitive to anti-CTLA4 treatment based on the *p*-value but not based on the corrected *p*-value. The mechanism underlying the observed difference in sensitivity requires further research. In addition, sensitivity to doxorubicin was higher for cluster 2 than cluster 1.

Finally, we constructed a clinical prognostic model based on mast cell-related genes using the LASSO Cox regression model and multivariate Cox regression model to predict the prognosis and survival time of patients with KIRC. Time-dependent ROC supported the predictive ability of the model.

The clinical prognostic model was based on eight mast cell-related genes. The tyrosine kinase receptor TEK, mainly expressed on endothelial cells, is activated by Angiopoietin-1. Endothelial cell survival and vascular maturation are promoted by the activation and phosphorylation of TEK, leading to downstream signal transduction (Eklund and Saharinen, 2013). Moreover, TEK promotes immune responses, the activation of mast cells, and the adhesion of mast cells to VCAM-1 (Kanemaru et al., 2015). Low TEK expression promotes AKT phosphorylation, the epithelial–mesenchymal transition, and the proliferation and migration of KIRC cells and inhibits the apoptosis of KIRC cells (Chen et al., 2021). In addition, the mitogen-activated protein kinase (MAPK) pathway is related to senescence, apoptosis, cell proliferation, differentiation, and migration (Sun et al., 2015). Cytokine signal transduction 3 (SOCS3) is an inhibitor of IL-6 and a negative regulator of cytokine signal transduction. SOCS3 not only inhibits cytokine-mediated JAK/STAT signal transduction, but also maintains the MAPK pathway, thereby promoting the growth of KIRC and angiogenesis (Oguro et al., 2013). Very few studies of KIRC have focused on TRPC4AP, IL17RD, PTH, PDIA2, FCGRT, and GDF5, and the role of these mast cell-related genes in KIRC requires further research.

Epigenetic changes often occur in KIRC and may be important events in its development (Joosten et al., 2018). Abnormal DNA methylation is a common type of epigenetic change, including genome-wide changes and regional variation (Jones and Baylin, 2002; Feinberg and Tycko, 2004). Abnormal DNA methylation can induce the abnormal expression of cancer-related genes and is the most common epigenetic change in tumorigenesis. Changes in DNA methylation during tumor progression affect target tumor cells; additionally, the immune system may undergo methylation changes during immune responses (Li et al., 2017). In our study, the promoter methylation levels of FCGRT, PDIA2, PTH, and TRPC4AP were reduced in KIRC. In contrast, promoter methylation levels of GDF5 and TEK were elevated in KIRC. We believe that a decrease in TEK resulting from an increase in promoter methylation levels may promote the proliferation and migration of KIRC cells, ultimately leading to the occurrence and progression of KIRC. Of course, the mechanism underlying the changes in the methylation levels of these genes in KIRC needs further verification.

The prognostic value of the newly established model was supported by an analysis of the OS of patients with KIRC in a training group, in which patients classified as high-risk had a shorter survival time. In addition, we used the risk scores calculated from the prognosis model to generate a risk curve to monitor disease progression. The ROC curve showed that our clinical prognostic model had a high predictive value. All results were verified by ArrayExpress and ICGC cohorts. Therefore, this model may be valuable for evaluating the prognosis of patients with KIRC.

This study had some limitations. First, all data were collected from TCGA, ArrayExpress, and ICGC, but lack of a support from hospital centre. Second, experimental studies of the functions of mast cell-related genes were not conducted. Therefore, further verification is needed to clarify the molecular mechanisms underlying KIRC and the roles of mast cells.

CONCLUSION

We found a correlation between prognosis and mast cell abundance in KIRC. By WGCNA, genes related to mast cells were identified, and two molecular subtypes (cluster 1 and cluster 2) were identified. Patients in cluster 2 were more likely to benefit from immunotherapy. The newly developed clinical prognostic model based on eight mast cell-related genes may contribute to the monitoring and the prediction of survival. More broadly, our research provides a basis for personalized medicine in KIRC.

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

HL was responsible for writing the manuscript, YY analyzed the data and concept proposal, and supervised the manuscript.

FUNDING

This work was supported in part by the Key Research and Development Project of Liaoning Province (2020JH 2/10300145) and 345 Talent Project of Shengjing Hospital, LiaoNing Revitalization Talents Program (XLYC1908008).

ACKNOWLEDGMENTS

We thank Editage for editing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.719982/full#supplementary-material

REFERENCES

Abdel-Majid, R. M., and Marshall, J. S. (2004). Prostaglandin E2Induces Degranulation-independent Production of Vascular Endothelial Growth Factor by Human Mast Cells. *J. Immunol.* 172, 1227–1236. doi:10.4049/jimmunol.172.2.1227

Amini, R.-M., Aaltonen, K., Nevanlinna, H., Carvalho, R., Salonen, L., Heikkilä, P., et al. (2007). Mast Cells and Eosinophils in Invasive Breast Carcinoma. *BMC Cancer* 7, 165. doi:10.1186/1471-2407-7-165

Beuselinck, B., Job, S., Becht, E., Karadimou, A., Verkarre, V., Couchy, G., et al. (2015). Molecular Subtypes of clear Cell Renal Cell Carcinoma Are Associated with Sunitinib Response in the Metastatic Setting. *Clin. Cancer Res.* 21, 1329–1339. doi:10.1158/1078-0432.CCR-14-1128

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

Boesiger, J., Tsai, M., Maurer, M., Yamaguchi, M., Brown, L. F., Claffey, K. P., et al. (1998). Mast Cells Can Secrete Vascular Permeability Factor/Vascular Endothelial Cell Growth Factor and Exhibit Enhanced Release after Immunoglobulin E-dependent Upregulation of Fcε Receptor I Expression. *J. Exp. Med.* 188, 1135–1145. doi:10.1084/jem.188.6.1135

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

Chen, S., Yu, M., Ju, L., Wang, G., Qian, K., Xiao, Y., et al. (2021). The Immune-related Biomarker TEK Inhibits the Development of clear Cell Renal Cell Carcinoma (ccRCC) by Regulating AKT Phosphorylation. *Cancer Cel Int* 21, 119. doi:10.1186/s12935-021-01830-1

Dabiri, S., Huntsman, D., Makretsov, N., Cheang, M., Gilks, B., Badjik, C., et al. (2004). The Presence of Stromal Mast Cells Identifies a Subset of Invasive Breast Cancers with a Favorable Prognosis. *Mod. Pathol.* 17, 690–695. doi:10.1038/modpathol.3800094

David, Chang., Ying, Ma., Baoan, Ji., Wang, H., Deng, D., Liu, Y., et al. (2011). Mast Cells in Tumor Microenvironment Promotes the *In Vivo* Growth of Pancreatic Ductal Adenocarcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 17 (22), 7015–7023. doi:10.1158/1078-0432.CCR-11-0607

Detoraki, A., Granata, F., Staibano, S., Rossi, F. W., Marone, G., and Genovese, A. (2010). Angiogenesis and Lymphangiogenesis in Bronchial Asthma. *Allergy* 65, 946–958. doi:10.1111/j.1398-9995.2010.02372.x

Eklund, L., and Saharinen, P. (2013). Angiopoietin Signaling in the Vasculature. *Exp. Cel Res.* 319, 1271–1280. doi:10.1016/j.yexcr.2013.03.011

Fang, J., Hu, M., Sun, Y., Zhou, S., and Li, H. (2020). Expression Profile Analysis of m6A RNA Methylation Regulators Indicates They Are Immune Signature Associated and Can Predict Survival in Kidney Renal Cell Carcinoma. *DNA Cel Biol.* 39, 2194–2211. doi:10.1089/dna.2020.5767

Feinberg, A. P., and Tycko, B. (2004). The History of Cancer Epigenetics. *Nat. Rev. Cancer* 4, 143–153. doi:10.1038/nrc1279

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, R47, 2014 . -03-. doi:10.1186/gb-2014-15-3-r47

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

Heidenreich, A., Wilop, S., Pinkawa, M., Porres, D., and Pfister, D. (2012). Surgical Resection of Urological Tumor Metastases Following Medical Treatment. *Dtsch. Arztebl. Int.* 109, 631–637. doi:10.3238/arztebl.2012.0631

Hiroshi, Y., Masakatsu, K., Hideo, T., Nakano, Y., Matsui, S., Monden, T., et al. (1999). Mast Cell Infiltration Around Gastric Cancer Cells Correlates with Tumor Angiogenesis and Metastasis. *Gastric Cancer* 2, 26–32. doi:10.1007/s101200050017

Jones, P. A., and Baylin, S. B. (2002). The Fundamental Role of Epigenetic Events in Cancer. *Nat. Rev. Genet.* 3, 415–428. doi:10.1038/nrg816

Joosten, S. C., Smits, K. M., Aarts, M. J., Melotte, V., Koch, A., Tjan-Heijnen, V. C., et al. (2018). Epigenetics in Renal Cell Cancer: Mechanisms and Clinical Applications. *Nat. Rev. Urol.* 15, 430–451. doi:10.1038/s41585-018-0023-z

Kanemaru, K., Noguchi, E., Tokunaga, T., Nagai, K., Hiroyama, T., Nakamura, Y., et al. (2015). Tie2 Signaling Enhances Mast Cell Progenitor Adhesion to Vascular Cell Adhesion Molecule-1 (VCAM-1) through α4β1 Integrin. *PLOS ONE* 10, e0144436. doi:10.1371/journal.pone.0144436

Lalani, A. A., Mcgregor, B. A., Albiges, L., Choueiri, T. K., Motzer, R., Powles, T., et al. (2018). Systemic Treatment of Metastatic clear Cell Renal Cell Carcinoma in 2018: Current Paradigms, Use of Immunotherapy, and Future Directions. *Eur. Urol.* 75, 100–110. doi:10.1016/j.eururo.2018.10.010

Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Li, L., Zheng, H., Huang, Y., Huang, C., Zhang, S., Tian, J., et al. (2017). DNA Methylation Signatures and Coagulation Factors in the Peripheral Blood Leucocytes of Epithelial Ovarian Cancer. *Carcinogenesis* 38, 797–805. doi:10.1093/carcin/bgx057

Li, Y., Gong, Y., Ning, X., Peng, D., Liu, L., He, S., et al. (2018). Downregulation of Cldn7 Due to Promoter Hypermethylation Is Associated with Human clear Cell Renal Cell Carcinoma Progression and Poor Prognosis. *J. Exp. Clin. Cancer Res.* 37, 276. doi:10.1186/s13046-018-0924-y

Lu, X., Jiang, L., Zhang, L., Zhu, Y., Hu, W., Wang, J., et al. (2019). Immune Signature-Based Subtypes of Cervical Squamous Cell Carcinoma Tightly Associated with Human Papillomavirus Type 16 Expression, Molecular Features, and Clinical Outcome. *Neoplasia* 21, 591–601. doi:10.1016/j.neo.2019.04.003

Marone, G., Varricchi, G., Loffredo, S., and Granata, F. (2016). Mast Cells and Basophils in Inflammatory and Tumor Angiogenesis and Lymphangiogenesis. *Eur. J. Pharmacol.* 778, 146–151. doi:10.1016/j.ejphar.2015.03.088

Melillo, R. M., Guarino, V., Avilla, E., Galdiero, M. R., Liotti, F., Prevete, N., et al. (2010). Mast Cells Have a Protumorigenic Role in Human Thyroid Cancer. *Oncogene* 29, 6203–6215. doi:10.1038/onc.2010.348

Motzer, R. J., Bander, N. H., and Nanus, D. M. (1996). Renal-cell Carcinoma. *N. Engl. J. Med.* 335, 865–875. doi:10.1056/NEJM199609193351207

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, 773–782. doi:10.1038/s41587-019-0114-2

Oguro, T., Ishibashi, K., Sugino, T., Hashimoto, K., Tomita, S., Takahashi, N., et al. (2013). Humanised Antihuman IL-6R Antibody with Interferon Inhibits Renal Cell Carcinoma Cell Growth *In Vitro* and *In Vivo* through Suppressed Socs3 Expression. *Eur. J. Cancer* 49, 1715–1724. doi:10.1016/j.ejca.2012.11.038

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

Rao, A., Wiggins, C., and Lauer, R. C. (2018). Survival Outcomes for Advanced Kidney Cancer Patients in the Era of Targeted Therapies. *Ann. Transl. Med.* 6, 165. doi:10.21037/atm.2018.04.44

Rao, Q., Chen, Y., Yeh, C.-R., Ding, J., Li, L., Chang, C., et al. (2016). Recruited Mast Cells in the Tumor Microenvironment Enhance Bladder Cancer Metastasis via Modulation of ERβ/CCL2/CCR2 EMT/MMP9 Signals. *Oncotarget* 7, 7842–7855. doi:10.18632/oncotarget.5467

Ro, W., Che, P. L., Reube, A., Spence, C. N., Prieto, P. A., Mille, J. P., et al. (2017). Integrated Molecular Analysis of Tumor Biopsies on Sequential Ctla-4 and Pd-1 Blockade Reveals Markers of Response and Resistance. *Sci. Transl. Med.* 9, eaah3560. doi:10.1126/scitranslmed.aah3560

Sara, V., Tornberg, H., Nisen, H., Visapää, H., Järvinen, R., Mirtti, T., et al. (2016). Outcome of Surgery for Patients with Renal Cell Carcinoma and Tumour Thrombus in the Era of Modern Targeted Therapy. *Scand. J. Urol.* 50 (5), 380–386. doi:10.1080/21681805.2016.1217558

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

Sismanopoulos, N., Delivanis, D. A., Alysandratos, K. D., Angelidou, A., Vasiadi, M., Therianou, A., et al. (2012). IL-9 Induces Vegf Secretion from Human Mast Cells and IL-9/IL-9 Receptor Genes Are Overexpressed in Atopic Dermatitis. *PLOS ONE* 7, e33271. doi:10.1371/journal.pone.0033271

Strouch, M. J., Cheon, E. C., Salabat, M. R., Krantz, S. B., Gounaris, E., Melstrom, L. G., et al. (2010). Crosstalk between Mast Cells and Pancreatic Cancer Cells Contributes to Pancreatic Tumor Progression. *Clin. Cancer Res.* 16, 2257–2265. doi:10.1158/1078-0432.CCR-09-1230

Sun, Y., Liu, W.-Z., Liu, T., Feng, X., Yang, N., and Zhou, H.-F. (2015). Signaling Pathway of Mapk/erk in Cell Proliferation, Differentiation, Migration, Senescence and Apoptosis. *J. Receptors Signal Transduction* 35, 600–604. doi:10.3109/10799893.2015.1030412

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

Taskinen, M., Karjalainen-Lindsberg, M.-L., and Leppä, S. (2008). Prognostic Influence of Tumor-Infiltrating Mast Cells in Patients with Follicular Lymphoma Treated with Rituximab and Chop. *Blood* 111, 4664–4667. doi:10.1182/blood-2007-11-125823

Theoharides, T. C., Zhang, B., Kempuraj, D., Tagen, M., Vasiadi, M., Angelidou, A., et al. (2010). IL-33 Augments Substance P-Induced VEGF Secretion from Human Mast Cells and Is Increased in Psoriatic Skin. *Proc. Natl. Acad. Sci.* 107, 4448–4453. doi:10.1073/pnas.1000803107

Varricchi, G., Galdiero, M. R., Loffredo, S., Marone, G., Iannone, R., Marone, G., et al. (2017). Are Mast Cells Masters in Cancer. *Front. Immunol.* 8, 424. doi:10.3389/fimmu.2017.00424

Warren, A. Y., and Harrison, D. (2018). Who/isup Classification, Grading and Pathological Staging of Renal Cell Carcinoma: Standards and Controversies. *World J. Urol.* 36, 1913–1926. doi:10.1007/s00345-018-2447-8

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, 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

Zhao, Q., Xue, J., Hong, B., Qian, W., Liu, T., Fan, B., et al. (2020). Transcriptomic Characterization and Innovative Molecular Classification of clear Cell Renal Cell Carcinoma in the Chinese Population. *Cancer Cel Int* 20, 461. doi:10.1186/s12935-020-01552-w

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 Liu 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 Molecular Biosciences

REVIEW
published: 29 September 2021
doi: 10.3389/fmolb.2021.736330

Clinical Application Value of Circulating Cell-free DNA in Hepatocellular Carcinoma

Yuyuan Zhang[1,2,3†], Zaoqu Liu[1,2,3†], Kun Ji[1,2,3], Xin Li[1,2,3], Caihong Wang[4], Zhigang Ren[5], Yang Liu[6], Xinju Chen[7], Xinwei Han[1,2,3], Lingfang Meng[8], Lifeng Li[9,10] and Zhen Li[1,2,3]**

[1]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Interventional Institute of Zhengzhou University, Zhengzhou, China, [3]Interventional Treatment and Clinical Research Center of Henan Province, Zhengzhou, China, [4]Department of Magnetic Resonance, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, China, [5]Department of Infections Disease, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, China, [6]Department of Radiation Oncology, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China, [7]First Ward of Spleen, Stomach, Liver and Gall, The First Affiliated Hospital of Henan University of TCM, Zhengzhou, China, [8]Department of Ultrasound, Zhengzhou Sixth People's Hospital, Henan Infectious Disease Hospital, Zhengzhou, China, [9]Internet Medical and System Applications of National Engineering Laboratory, Zhengzhou, China, [10]Cancer Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:
Florence Le Calvez-Kelm,
International Agency For Research On Cancer (IARC), France

Reviewed by:
Emmanuelle Gormally,
Catholic University of Lyon, France
Kenneth. S Hettie,
Stanford University, United States

***Correspondence:**
Lifeng Li
lilifeng0317@163.com
Zhen Li
lzlyct@620.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 05 July 2021
Accepted: 14 September 2021
Published: 29 September 2021

Citation:
Zhang Y, Liu Z, Ji K, Li X, Wang C, Ren Z, Liu Y, Chen X, Han X, Meng L, Li L and Li Z (2021) Clinical Application Value of Circulating Cell-free DNA in Hepatocellular Carcinoma. Front. Mol. Biosci. 8:736330. doi: 10.3389/fmolb.2021.736330

Hepatocellular carcinoma (HCC) is one of the most common cancers worldwide and a leading cause of cancer-related deaths. Due to late diagnosis, early intrahepatic metastasis and nonresponse to systemic treatments, surgical resection and/or biopsy specimens remain the gold standard for disease staging, grading and clinical decision-making. Since only a small amount of tissue was obtained in a needle biopsy, the conventional tissue biopsy is unable to represent tumor heterogeneity in HCC. For this reason, it is imperative to find a new non-invasive and easily available diagnostic tool to detect HCC at an early stage and to monitor HCC recurrence. The past decade has witnessed considerable evolution in the development of liquid biopsy technologies with the emergence of next-generation sequencing. As a liquid biopsy approach, molecular analysis of cell-free DNA (cfDNA), characterized by noninvasiveness and real-time analysis, may accurately represent the tumor burden and comprehensively reflect genetic profile of HCC. Therefore, cfDNA may be used clinically as a predictive biomarker in early diagnosis, outcome assessment, and even molecular typing. In this review, we provide an update on the recent advances made in clinical applications of cfDNA in HCC.

Keywords: cfDNA, CtDNA, Hepatocellular carcinoma, clinical application, biomarkers

Abbreviations: AFP, α-fetoprotein; ARMS, amplification-refractory mutation system; AUC, area under the curve; cfDNA, cell-free DNA; CNV, copy number variation; CTCs, circulating tumor cells; ctDNA, circulating tumor DNA; EPIC-Seq, the Epigenetic Profile Inference from the CfDNA Deep Sequencing; HCC, Hepatocellular carcinoma; mmPCR-NGS: massively multiplexed PCR and next-generation sequencing; NGS, next-generation sequencing; qPCR, quantitative real-time PCR; RFS, progression-free survival; ROC, receiver operator characteristic curve; sWGS, shallow WGS; TACE, transplantation or transarterial chemoembolization; TCGA, The Cancer Genome Atlas; UMIs, unique molecule identifiers; WGS, whole-genome sequencing.

INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most common cancer worldwide and the fourth most frequently reported cause of cancer death by 2018 (Bray et al., 2018). HCC ranks third of cancer-related mortality in China, and the major risk factors for HCC are hepatitis B virus (HBV) or hepatitis C virus (HCV) infection, exposure to aflatoxin B1, alcohol consumption, and metabolic disorders (Chen et al., 2016). Currently, benefiting from liver resection, ablation, liver transplantation, the 5-years survival rate of early HCC (BCLC stage A) can reach 50–75%. Unfortunately, owing to the paucity of specific symptoms and early intra/extrahepatic metastases, most patients have already reached an advanced cancer stage at the time of first HCC diagnosis, giving rise to fewer than 40% of HCC patients eligible for surgical intervention (Forner et al., 2018). Therefore, it is critical to find a robust method to detect patients with HCC at earlier stages, monitor tumor recurrence and even better predict response to treatment in a dynamic and real-time manner.

The diagnosis and surveillance of HCC primarily depends on findings from imaging analysis and alpha-fetoprotein (AFP) levels. Unfortunately, this dynamic imaging has limitations in accuracy and sensitivity when referring to small or hypovascularized lesions. Biopsy is recognized as the standard diagnosis method, while the problem of invasiveness and the false positivity remains to be solved. Lately, circulating free DNA (cfDNA) has emerged as a promising alternative in tumor diagnosis, recurrence surveillance and druggable targets identification (Benesova et al., 2013; Marzese et al., 2013; Diaz and Bardelli, 2014). CfDNA is the fragmented DNA in the blood circulation which can be detected in healthy individuals and patients with cancer, and ctDNA is the fraction of cfDNA specifically derived from primary or metastatic tumors, with the concentration ranging from 0.01 to 90% (Jen et al., 2000). Nowadays, both qualitative and quantitative analysis of cfDNA is utilized in cancer based on underlying genetic predisposition, including detection of genomic changes, mutational analysis, oncogenic pathway determination, prediction/monitoring of treatment response, drug resistance alterations, and identification of mechanisms of malignant/metastatic transformation (Kaseb et al., 2019).

HCC is a highly heterogeneous disease attributing to the accumulation of somatic genomic aberrations in passenger and driver genes as well as epigenetic modifications (Forner et al., 2018). For many years, the recognition of genomic aberrations in HCC mainly relied on liver resection/biopsies. Since invasiveness, traditional fine-needle biopsy does not entirely avoid risks and potential complications, such as pain (84%) (Eisenberg et al., 2003), bleeding (Russo et al., 2018) and needle tract seeding (2.7%) (Silva et al., 2008). In addition, insufficient material for clinical sequencing occurs in 20–25% of needle biopsies (Zill et al., 2015), resulting in the mental and financial pressure of patients to some extent. Intrahepatic metastasis occurs early in the progression of HCC and thus information acquired from a needle biopsy of a single tumor lesion might fail to reflect the tumor burden (Swanton, 2012). In turn, analysis of cfDNA may overcome these limitations and subsequently provide the genomic profiles of all lesions (both primary lesion and metastasis), and this method can be utilized to track genomic evolution systematically and dynamically (Kaseb et al., 2019). With increasing availability and reliability in high-throughput technology, plasma cfDNA may meet the demands of disease surveillance, management of different stages and personal precision medicine for patients with HCC. This review will provide an update on the advances made in the clinical application of cfDNA in HCC in recent years.

cfDNA

Biological Basis

cfDNA is double-stranded DNA measuring approximately 150–200 base pairs which exists in the plasma or serum (Fan et al., 2010). Typically, the concentration of cfDNA is low (10–15 ng/ ml on average) for healthy people, with a short half-life of between 16 min and 2.5 h (Lo et al., 1999; Diehl et al., 2008). While, increasing concentration of cfDNA can be observed under some physical and pathological circumstances, such as exercise, inflammation, surgery, autoimmune disease, and transplantation (Diehl et al., 2008). These characteristics indicate that cfDNA can provide more real-time information regarding the cancer status than routine serum biomarkers, such as AFP, CEA and CA-199.

Although the clinical benefits of cfDNA have been increasingly recognized recently, many aspects of the biological characteristics of tumor-derived cfDNA still have not been elucidated clearly. Firstly, the origin of cfDNA has not been definitively clarified with several possible mechanisms proposed (**Figure 1**). Two possibilities exist for the main origins of cfDNA: cellular breakdown mechanisms (such as apoptosis and necrosis) and active DNA release mechanisms (such as exosomes, virtosomes and argonaute) (Aucamp et al., 2018). Tumor-associated DNA is primarily derived from apoptosis of primary, metastatic and circulating tumor cells (CTCs), as well as active release from proliferative tumor cells, which is known as circulating tumor DNA (ctDNA) (Crowley et al., 2013). In general, ctDNA only accounts for a small fraction of the total cfDNA and there is no way to isolate ctDNA, especially from other cfDNA. Only the emergence of tumor-related molecular alterations in cfDNA indicates the presence of ctDNA (Ye et al., 2019).

In addition, there are still some different views concerning circulating DNA fragment patterns derived from tumors such as the concentration of total ctDNA and the proportion of tumor-derived ctDNA fragments (Umetani et al., 2006a; Gao et al., 2010; Mouliere et al., 2011). A number of studies reported the finding of increased integrity of tumor-derived plasma DNA (Wang et al., 2003; Jiang et al., 2006). For example, a study reported that the increased DNA integrity in plasma DNA is associated with cancer, and excellent performance was achieved in detecting the cancer group from the nonneoplastic group. On the other hand, there is also seemingly contradictory evidence that plasma DNA molecules released by tumors might be shorter (Diehl et al., 2005; Mouliere et al., 2011). For example, a study performed by massively parallel sequencing suggested that the size of plasma DNA molecules harboring tumor-associated genomic alterations

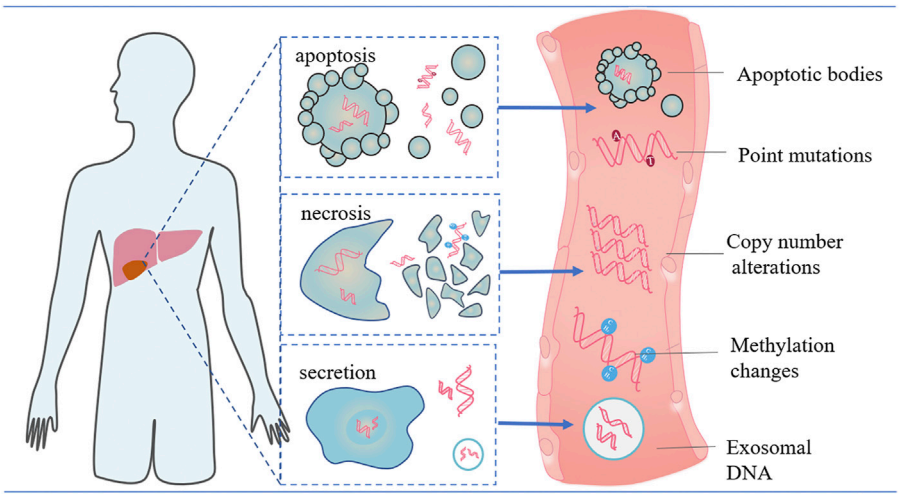


FIGURE 1 | Origin and range alterations of cfDNA. Cell-free DNA (cfDNA) is passively released by apoptotic or necrotic tumor cells and actively secreted by exosomes, which are membrane -bound vesicles released by tumor cells. Of these cells, apoptosis serves as the main source of cfDNA in both normal and diseased tissues. Analysis of these molecules can be employed for early tumor detection and may provide a prognostic treatment strategy for HCC patients.

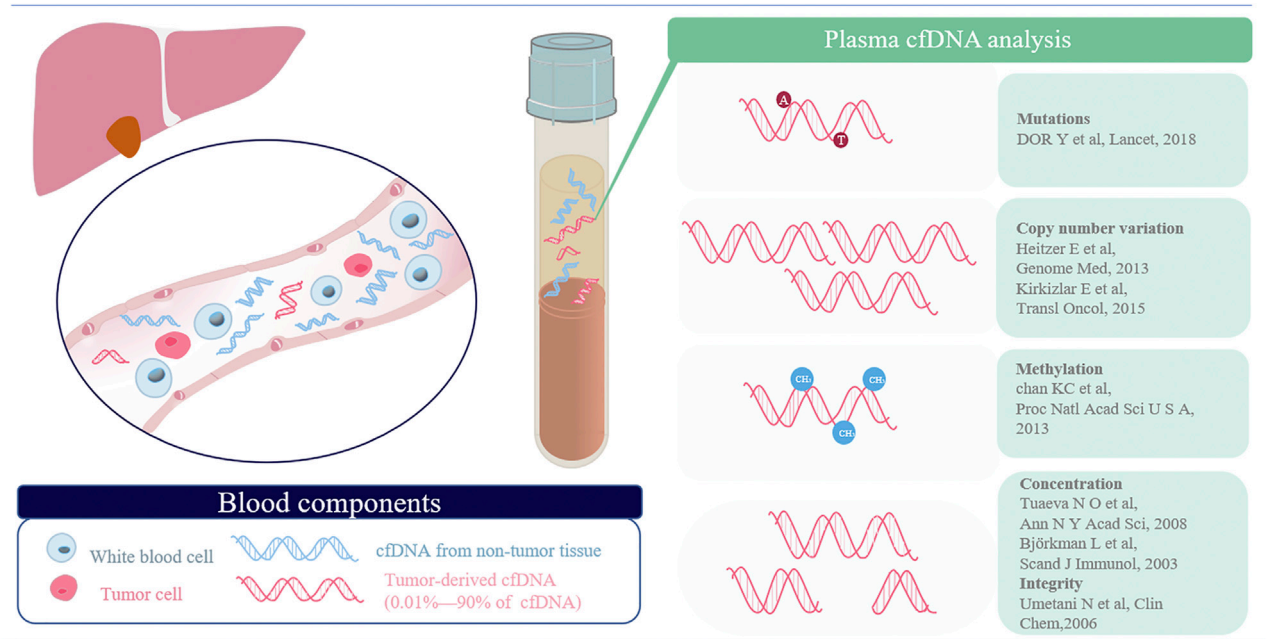


FIGURE 2 | Overview of the methods employed in the detection of different cancer-associated alterations. cfDNA in the plasma captures different cancer-associated changes, including mutations, copy number aberrations, alterations in DNA methylation and altered DNA fragmentation patterns (concentration and integrity).

was shorter (Jiang et al., 2015). Overall, cfDNA in the plasma is vulnerable to various assay platforms and physiological state, and thus it is necessary to comprehensively analyze tumor-derived plasma DNA rather than treating without distinction.

Technology Platform for ctDNA

CtDNA accounts for only a small percentage of the total cfDNA in the peripheral blood and is highly variable from patient to patient (0.01–90%), which therefore necessitates the utilization of hypersensitive and highly specific approaches (Corcoran and Chabner, 2018). Typically, alterations of cfDNA in patients with cancers involve quantity and quality (Gao et al., 2017). The former refers to the total concentration and integrity, and the latter means genetic or epigenetic alterations containing single nucleotide mutations, copy number variations and methylation changes and so on. In summary, the techniques for the analysis of

cfDNA can be summarized as targeted methods, such as digital PCR, BEAMing digital PCR, and amplification-refractory mutation system (ARMS)-PCR, as well as untargeted methods, such as whole-genome sequencing (WGS) and next-generation sequencing (NGS). Although the former method can screen cfDNA aberrations dynamically and has a very high sensitivity, it detects mutations only in a set of predefined genes based on prior knowledge (Oh et al., 2010). A more comprehensive view of the entire genomic landscape may be exhibited using the latter method, which may provide the potential of identifying drug resistance genes and recognizing novel actionable targets (Slatko et al., 2018). In general, the rational combination of the two techniques would accomplish the purpose, providing a noninvasive approach to guide the clinical management of patients. In the following part, we will detailedly introduce the techniques for detecting aberrant cfDNA in the plasma (**Figure 2**).

Concentrations

In 1977, for the first time, researchers revealed that cfDNA increased in patients with tumors (Leon et al., 1977). After this seminal work, a significant volume of researches have attempted to detect cancers by quantifying the absolute concentration of cfDNA in the serum/plasma (Mehra et al., 2018; Valpione et al., 2018; Wu et al., 2019). At present, the level of cfDNA in the plasma was estimated by several classical methods, including fluorescence-based quantitative real-time PCR (qPCR), histochemical techniques (such as Hoechst staining and PicoGreen staining) and optical imaging techniques (such as ultraviolet measurement) (Björkman et al., 2003; Tuaeva et al., 2008). It appears that the level of cfDNA in the plasma may reflect the tumor burden (Yang et al., 2011) and may have diagnostic value for HCC (Huang et al., 2012; Yan et al., 2018; Qu et al., 2019). However, an increase in the concentration alone cannot be regarded as an effective and cancer-specific metric for diagnosis and classification because the cfDNA in the plasma is vulnerable to various assay platforms and physiological state like active ongoing hepatitis and hepatic dysfunction (Yan et al., 2018).

Integrity

There is another quantitative characterization of tumors associated with cfDNA: assessment of the integrity index. Currently, the most popular method was published by Umetani. By developed the qPCR method for ALU repeats, this method accomplished the assessment of the integrity without DNA purification and then established the integrity index (Umetani et al., 2006b). In the ALU-qPCR, DNA integrity was determined as the ratio of ALU247 primers to ALU115 primers, and the ALU115 primers represent the absolute amount of DNA. Plasma cfDNA integrity has been applied to many types of cancers, such as breast cancer (Umetani et al., 2006a), colon cancer (Agostini et al., 2011) and liver cancer (Huang et al., 2016a). Compared with concentrations of cfDNA in the plasma, cfDNA integrity is less affected by much influential factors. In previous studies (El-Shazly et al., 2010; Chen et al., 2012; Huang et al., 2016a), changes of integrity in cfDNA is associated with large tumor size and vascular invasion, and it exerted a crucial role in distinguishing cancer from healthy people and surveillance of residual disease after surgery.

Mutations

Studies identifying and characterizing somatic genetic mutations in HCC have been updated annually. There are two types of methods for detecting mutation: targeted methods monitoring interesting gene loci and nontargeted massively parallel sequencing. When the proportion of cfDNA derived from tumor cells accounts for only a small part of the cfDNA in the plasma, abundant errors are expected to be present in the PCR amplification and sequencing procedures. To overcome this problem, unique molecule identifiers (UMIs) have emerged for detecting single nucleotide mutations at a fraction as low as 1 in 1 million molecules (Kinde et al., 2011). In the clinical practice, compared with tumor tissue level mutational heterogeneity, single specific ctDNA alterations do not have sufficiently high sensitivity or specificity as a diagnostic biomarker for HCC. In past literatures, a positive rate of 20–56% for plasma ctDNA was reported (Huang et al., 2016b; Jiao et al., 2018a). Improvement in NGS technology allowed comprehensive analysis of the mutational landscape of ctDNA, expanded the list of mutation biomarker candidates, and promisingly improved the detection rate of tumor-related mutation with the positive rate of 60–80% (Ng et al., 2018; Yang et al., 2019).

Methylation

Epigenetic events play a crucial role in hepatocarcinogenesis. Methylation is a covalent modification pattern that can exist stably in the plasma after the release of defective cells (Dor and Cedar, 2018). Currently, the detection of methylation primarily consists of two types based on the size of the targeted region: target-region methylation detection (methylation-specific PCR) and genome-wide methylation detection (bisulfite sequencing) (Chan et al., 2013). Although concentrated targeted regions are advantageous and inexpensive, the former method cannot precisely show the tumor state because methylation regions (especially CpG islands) are easily affected by gene expression levels. The latter method has wide coverage of the sequenced regions, as well as strong prospects for clinical application. At present, there is ample data to support that DNA methylation changes of specific genes in cfDNA, including hypermethylation of TGR5 (Han et al., 2014), MT1M (Ji et al., 2014) and the RASSF1A promoters (Wong et al., 2000), as well as hypomethylation of LINE-1 elements (Tangkijvanich et al., 2007). Not surprisingly, the combinations of cfDNA methylation in panels can improve the diagnostic performance with high sensitivity and specificity of over 90% (Yang et al., 2019). In summary, the methylation panels appear to be a potential detection marker in the clinical utility accompany with promising application prospects.

Clinical Applications in HCC

CfDNA was originally discovered in 1948 (Mandel and Metais, 1948). With the recent development of genomics and molecular

TABLE 1 | | Clinical Applications of cfDNA in HCC.

Applications	Methods	DNA aberrations	
Detection or screening early HCC			
Integrity	qPCR amplification of b-globin gene	Increased in HBV associated HCC patients.	Chen et al. (2012)
	qPCR	Decreased in HCC patients, increased after hepatectomy in cancer patients and the AUCs for detecting HCC by cfDNA integrity and AFP were 0.705 and 0.605, respectively.	Agostini et al. (2011)
	real-time PCR amplification	Significantly higher in HCC patients than in HCV carriers without known HCC.	Tokuhisa et al. (2007)
Concentration	MiSeq sequencing	There is no difference between HCC patients and healthy people.	Liao et al. (2016)
Mutation	MiSeq sequencing	Genetic mutations were detected in plasma samples of two patients (4.9%) for the TERT genetic mutation, four (9.8%) patients for CTNNB1 mutation, and two (4.9%) patients for the TP53 mutation.	Liao et al. (2016)
	PCR-based sequencing	The median sensitivity of CancerSEEK among the eight cancer types evaluated was 70%, the specificity was 99%.	Cohen et al. (2018)
Methylation	MSRE-qPCR	The combination analysis of these four genes resulted in an increased AUC of 0.933 with 92.7% sensitivity, 81.9% specificity in discriminating HCC from normal control.	Huang et al. (2011)
	the 5 hmC-Seal technique	A 32-gene diagnostic model accurately distinguished early HCC (stage 0/A) from non-HCC (validation set: AUC = 88.4%), showing superior performance over AFP.	Cai et al. (2019a)
guiding treatment			
Mutation	next-generation sequencing	CtDNA derived from noninvasive blood tests can provide exploitable genomic profiles in patients with HCC.	Ikeda et al. (2018)
Prognosis assessment			
	exome sequencing	Multivariate analysis identified ctDNA (OR 6.10; 95% CI, 1.11–33.33, $p = 0.038$) as an independent predictor of microscopic vascular invasion of the portal vein (VP).	Pezzuto et al. (2018)
Concentration			
Mutation	high-resolution melting PCR (HRM-PCR) and COLD-PCR	Mutated p53 genes could be used as a biomarker of tumor recurrence during the clinical evolution of the transplanted patients.	García-Fernández et al. (2016)
	whole exome sequencing	Real-time track the therapeutic responses in the longitudinal monitoring.	Cai et al. (2017)
Methylation	Targeted bisulfite sequencing	Methylation pattern was highly correlated with tumor burden, treatment response, and stage and can effectively predict prognosis and survival ($p < 0.001$).	Xu et al. (2017)
	real-time quantitative methylation specific PCR (RTQ-MSP)	After surgical resection, the median p16INK4a methylation indices in plasma and buffy coat concordantly decreased 12- and 15-fold.	Wong et al. (2003)
	Methylation Specific PCR (MSP)	Examination of LINE-1 hypomethylation and RASSF1A promoter hypermethylation was effective in predicting early recurrence of HCC after curative resection.	Liu et al. (2017)

CNV, copy number variation; HCC, hepatocellular carcinoma; qPCR, quantitative polymerase chain reactions.

genetics, cfDNA research has significantly increased for a variety of clinical and research purposes (Bettegowda et al., 2014), and cfDNA analysis is increasingly recognized as an effective and noninvasive tool in the diagnosis, personalized treatment, and follow-up of HCC patients (**Table 1**).

Early Detection

The effective management of patients with HCC relies on the early diagnosis of the disease (Lim et al., 2017), while biomarkers for early detection are still deficient. Most recently, improvement of the next-generation sequencing technology and better understanding of genetic or epigenetic alteration of HCC have allowed comprehensive analysis of the landscape of cfDNA. The non-invasive method was also attached great importance in early detection of HCC.

Increasing concentration of cfDNA was observed in the status of illness, thus quantitative measurement of untargeted cfDNA may have diagnostic value for HCC(Tokuhisa et al., 2007). However, due to factors of various assay platforms and physiological state of different patients, there were wide variations in the concentration of cfDNA even among healthy controls. A recent study enrolled 24 HCC patients and 62 hepatitis B virus-related liver fibrosis patients, and constructed a model including age, cfDNA, and AFP, had an area of 0.98 (95% confidence interval 0.92–1.00) under the ROC for the diagnosis of HCC, with 87.0% sensitivity and 100% specificity. Overall, it would not be recommended for using quantitative measurement independently, whereas the combination of cfDNA with other proteins or genetic biomarkers is expected to be a clinical tool for the early diagnosis of HCC (Liao et al., 2016). Compared with the concentration, detecting the cfDNA integrity index is another approach employed in quantitative analysis, which is considered more stable. A study showed that serum DNA integrity in HCC patients was notably higher than that in HBV carriers or healthy controls (Chen et al., 2012). In another study, found that cfDNA integrity was significantly decreased in HCC patients compared with patients with benign diseases and healthy individuals. Additionally, cfDNA integrity (AUC = 0.705) had a higher diagnostic performance than AFP (AUC = 0.605) (Huang et al., 2016a). Considering the heterogeneity of integrity in different studies, it is necessary to comprehensively analyze aberrantly short and long DNA molecules in the plasma of patients.

Cancer-specific genetic aberrations in cfDNA are detectable by liquid biopsy as biomarkers for the diagnosis and monitoring of HCC. As the high heterogeneity in HCC, single gene variation in cfDNA do not have sufficiently high sensitivity and specificity as a biomarker for HCC(Yang et al., 2017; Jiao et al., 2018b; Marchio et al., 2018). For example, a study showed that there were tumor-associated mutations for HCC in only 8 of 41 patient (19.5%) plasma samples, including mutation to such genes as TP53, hTERT, and CTNNB1 (Liao et al., 2016). To achieve an improved ability to diagnose early HCC, different panels of multigene utilizing NGS technology have broader application in the world. A study of 30 biopsy proven HCC patients were prospectively recruited. Using a panel of 46 genes frequently altered in HCCs, deep sequencing of the DNA from the biopsies, cfDNA, and matched germline was performed and ctDNA was detected in 63% of the patients (Labgaa et al., 2018). Joshua D. Cohen et al. developed a method, termed CancerSeek, which combined the 8 serum protein biomarkers with 16 tumor-associated genes in the cfDNA to detect eight types of early (stage I and II) cancers involving 1,005 patients. Among these cancers, CancerSeek has demonstrated accuracy in the diagnosis of early HCC with a sensitivity of 98% and specificity greater than 99% (Cohen et al., 2018). In summary, large panels of targeted NGS or allele-specific assays targeting hotspot mutations are required for early detection, and ctDNA combined with other biomarkers (AFP or other novel circulating molecular biomarkers) may play a crucial supplementary role as a diagnostic biomarker in HCC.

DNA methylation is a core mechanism of epigenetic regulation of gene expression and cell type-, tissue- and organ-specific methylation signature have been exploited as an HCC marker (Villanueva et al., 2015; Lehmann-Werman et al., 2016; Hlady and Robertson, 2018). A single methylation marker candidate study found that the methylation of RASSF1A was detected in 90% of cfDNA in HCC patients. It showed an accuracy rate of 77.5 and 72.5% as a diagnostic marker of HCC in healthy people and HCV carriers, respectively (Zhong et al., 2003). Similar to mutation in cfDNA, combinations of cfDNA methylation in panels can also improve the diagnostic performance in HCC. A model constructed with four methylation genes (APC, GSTP1, RASSF1A, and SFRP1) had a sensitivity of 92.7% and a specificity of 81.9% for detecting HCC(Huang et al., 2011). Furthermore, a recent study developed a 32-gene diagnostic model that accurately distinguished early HCC (BCL 0/A stage) from non-HCC and 15 AFP-negative patients, exhibiting superior performance over AFP (AUC = 88.4%). The model also exhibited superior ability in diagnosing HCC from high-risk populations, such as those with hepatitis and hepatic cirrhosis (AUC = 84.6%) (Cai et al., 2019a). Overall, ctDNA methylation panels appear to have the strongest potential for clinical utility in the early detection of HCC, and all the biomarkers requires prospective validation for the robust performance.

Medicine Guidance

Another important application of cfDNA is the identification of tumor gene profiling for individualized treatment, which has become a fundamental practice in cancer medicine (Harding et al., 2019). In the recent decade, various advances in treatment bring great promises and new opportunities for HCC therapeutics. Sorafenib and Lenvatinib in first-line showed efficacy, and regorafenib, cabozantinib as well as ramucirumab in second-line provided more chances for sequential systemic therapy in advanced HCC. Recently, immunotherapy has emerged as one of the most promising approaches to extend current options for needed HCC treatment (Liu et al., 2021a). Frustratingly, less than 40% of HCC patients are eligible for potentially curative therapies. Therefore, it is imperative to detect predictive biomarkers including FGFR4, Cdk5, PD-1, PD-L1, and tumor mutation burden (TMB) which may optimize the therapy strategy and forecast the benefit to patients receiving target and immune therapy. The limited applications of molecular classification using biopsy tissue samples may be attributable intra-tumoral or intertumoral heterogeneity in HCC patients; therefore, one piece of tumor tissue failed to present the complete molecular profile of one HCC patient (Liu et al., 2021b; Liu et al., 2021c). In contrast, numerous studies have demonstrated high concordance rates between plasma and tissue samples, particularly for alterations in key genes (Adalsteinsson et al., 2017; Strickler et al., 2018). Therefore, combined with the features of safety, low cost, operability and dynamic monitoring, cfDNA analysis may possess the potential to provide individualized treatment guidance for HCC patients.

Sadakatsu (Ikeda et al., 2018) implemented NGS to identify actionable mutations in 12 HCC patients, and ultimately, 556 exons of 68 genes were sequenced. A total of 79% of patients harbored at least one actionable mutation alteration, and the median of drug target mutation alterations was 2 (0–5), which suggested that performing liquid biopsies and detecting cfDNA may play a key role in guiding therapeutic decision-making. Furthermore, a patient with a *CDKN2A*-inactivating and a *CTNNB1*-activating mutation received palbociclib (CDK4/6 inhibitor) and celecoxib (*COX-2*/*Wnt* inhibitor); another patient with a PTEN-inactivating and a *MET-a*ctivating mutation received matched treatment: sirolimus (mechanistic target of rapamycin inhibitor) and cabozantinib (MET inhibitor). Eventually, the treatment outcomes of two patients were described as good. Another study prospectively enrolled 121 patients to identify predictors of primary resistance to systematic therapy using cfDNA. Patients with mutations in the PI3K/MTOR pathway had significantly shorter progression-free survival (PFS) than those without these mutations after tyrosine kinase inhibitors (2.1 *vs.* 3.7 months, $p < 0.001$), but not after immune checkpoint inhibition (CPI) (von Felden et al., 2021). Because of the low long-term response to sorafenib in HCC (Méndez-Blanco et al., 2018) and large individual differences in response to immunotherapy (Llovet et al., 2018), few investigations into the use of cfDNA to guide therapy in HCC have been reported.

Prognosis Evaluation

Earlier studies showed that the levels of cfDNA in the plasma were significantly correlated with tumor burden (Cai et al., 2017; Pezzuto et al., 2018) with a high specificity and short half-life

periods, meaning that cfDNA has a unique strength of prognostic value. There is evidence that mutations in cfDNA are closely related to vascular invasions, and patients with cfDNA mutations are more frequently reported to have shorter recurrence-free survival (RFS) (Cai et al., 2019b). García-Fernández et al. (2016) found that the presentation of TP53 mutations in cfDNA might be used as a biomarker of tumor recurrence in patients with transplanted HCC. Furthermore, Rui-Hua Xu et al. (2017) constructed a prognostic prediction model based on methylation alterations in cfDNA to effectively predict prognosis and survival ($p < 0.001$), which was highly correlated with tumor burden, treatment response, and stage. The above findings demonstrated that changes in cfDNA, including concentration, mutation and methylation, may provide references for the prediction of HCC prognosis.

Benefiting from a short half-life, cfDNA exhibits a fast response to changes in tumor burden after treatment, and thus it is possible to improve prognostication with the detection of tumor-specific molecular alterations. One study reported that the level of cfDNA decreased significantly after surgery, indicating that the dynamic monitoring of cfDNA alterations could provide additional information regarding the presence of residual lesions and early relapse after surgery (Corcoran and Chabner, 2018). ZX found that there was a notable decline in the mutation frequency of cfDNA, and during follow-up, the increased frequency of mutations in cfDNA often revealed the presence of residual lesions and, eventually, relapse (Cai et al., 2017). CtDNA methylation of a panel of genes has shown to be an important prognostic factor. The study of Wong IH showed that RTQ-MSP quantitative analysis can detect epigenetic changes (DNA methylation levels) in peripheral blood from patients with HCC. These researchers observed that the $p16^{INK4a}$ sequence of methylation is present in 80% of HCC patients, and a 12-fold decline was observed in the patients after surgery (Wong et al., 2003). Another study indicated that co-evaluation of *LINE-1* hypomethylation and *RASSF1A* promoter hypermethylation was effective in predicting early recurrence of HCC after curative resection (Liu et al., 2017). Furthermore, the level of demethylation was also significantly related to the presence of residual lesions after surgery (Chan et al., 2013). Overall, methylation patterns of the patients after surgery could provide an early warning of residual lesions and early relapse.

Challenges and Perspectives

In general, cfDNA may be an informative, inherently specific and highly sensitive biomarker, and can be performed in clinical and research purposes in many types of cancer. A blood sample from peripheral venous blood is not only convenient and easy to implement, but cfDNA in the plasma is relatively stable and easy to extract. In addition, with the characteristic of short half-life, cfDNA in this context is a more accurate and timely reflection of the tumor staging and recurrence than serum biomarkers. With the accumulation of liquid biopsy data, detection panels involving mutations at multiple loci and multiple methylation patterns using cfDNA may be employed for therapeutic monitoring, prognostic evaluation and early diagnosis in HCC.

The descriptions of molecular aberrations of the cfDNA in this review only focus on HCC, while mutations of the same keeper genes, such as TP53, KRAS, and RAF also occur in many other types of cancers. Besides, the cfDNA in plasma mainly derives from rupture and release of blood cells, and the concentration of plasma cfDNA can be affected by the different physiological and pathological conditions. There is an emerging problem which cannot be neglected: how could cfDNA be cancer-specific? The epigenetic biomarkers of cfDNA may be one of the most promising directions. DNA methylation is favored epigenetic biomarker because of cancer-type-specific and tissue-type-specific for the cancer diagnosis and tracing the tissue origin of cfDNA (Gai and Sun, 2019). Moreover, in combination with machine learning, methylation biomarkers in cfDNA provides a new clue to discriminate specific cancer type. Zhou and colleagues modeled the plasma cfDNA as a mixture of DNA derived from tumor and normal tissues, then they used a probabilistic model to sensitively identify a trace amount of tumor cfDNAs in plasma (Li et al., 2018a). Epigenetic biomarkers in cfDNA provide generalizable solutions for early detection and trace the origin of ctDNA.

Although cfDNA performed as a powerful detection and analysis tool along with promising potential, there are still many challenges for this technology to become clinical reality. First, the experimental data described in this review are diverse and vulnerable, which may attribute to a variety of individual differences, experimental designs and detection methods for cfDNA across different studies. In addition, there is a lack of an effective medical therapy and universal actionable mutations for HCC patients, resulting in rare medical applications of cfDNA in the plasma based on HCC-associated gene mutations (Li et al., 2018b). Altogether, cfDNA is an exciting product of precise medicine, and in the near future it may realize the purpose of real-time surveillance and tailored treatment for patients with HCC.

AUTHOR CONTRIBUTIONS

YZ and ZAL made conceptualization; YZ and CW consulted and summarized the literature; YZ, ZHL, and KJ prepared original draft, reviewed and edited the manuscript; KJ, ZAL, XC, YL, XH, LM, and XL supervised the study; LL, XC, and ZR contributed to submission. The authors read and approved the final manuscript.

FUNDING

Project was supported by the National Natural Science Foundation of China (Grant No. U1904143); National Science and Technology Major Project, Prevention and Treatment of Major Infectious Diseases such as AIDS and Viral Hepatitis (2018ZX10303502); The National Natural Science Foundation of China (Grant No. 82002433); Science and Technology Project of Henan Provincial Department of Education (Grant No. 21A320036); Henan Province Medical Science and Technology Research Project Joint Construction Project (Grant No. LHGJ20190003, LHGJ20190055).

REFERENCES

Adalsteinsson, V. A., Ha, G., Freeman, S. S., Choudhury, A. D., Stover, D. G., Parsons, H. A., et al. (2017). Scalable Whole-Exome Sequencing of Cell-Free DNA Reveals High Concordance with Metastatic Tumors. *Nat. Commun.* 8 (1), 1324. doi:10.1038/s41467-017-00965-y

Agostini, M., Pucciarelli, S., Enzo, M. V., Del Bianco, P., Briarava, M., Bedin, C., et al. (2011). Circulating Cell-free DNA: A Promising Marker of Pathologic Tumor Response in Rectal Cancer Patients Receiving Preoperative Chemoradiotherapy. *Ann. Surg. Oncol.* 18 (9), 2461–2468. doi:10.1245/s10434-011-1638-y

Aucamp, J., Bronkhorst, A. J., Badenhorst, C. P. S., and Pretorius, P. J. (2018). The Diverse Origins of Circulating Cell-Free DNA in the Human Body: A Critical Re-Evaluation of the Literature. *Biol. Rev.* 93 (3), 1649–1683. doi:10.1111/brv.12413

Benesova, L., Belsanova, B., Suchanek, S., Kopeckova, M., Minarikova, P., Lipska, L., et al. (2013). Mutation-Based Detection and Monitoring of Cell-free Tumor DNA in Peripheral Blood of Cancer Patients. *Anal. Biochem.* 433 (2), 227–234. doi:10.1016/j.ab.2012.06.018

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., et al. (2014). Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Translational Med.* 6 (224), 224ra24. doi:10.1126/scitranslmed.3007094

Björkman, L., Reich, C. F., and Pisetsky, D. S. (2003). The Use of Fluorometric Assays to Assess the Immune Response to DNA in Murine Systemic Lupus Erythematosus. *Scand. J. Immunol.* 57 (6), 525–533. doi:10.1046/j.1365-3083.2003.01261.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 (6), 394–424. doi:10.3322/caac.21492

Cai, Z.-X., Chen, G., Zeng, Y.-Y., Dong, X.-Q., Lin, M.-J., Huang, X.-H., et al. (2017). Circulating Tumor DNA Profiling Reveals Clonal Evolution and Real-Time Disease Progression in Advanced Hepatocellular Carcinoma. *Int. J. Cancer* 141 (5), 977–985. doi:10.1002/ijc.30798

Cai, J., Chen, L., Zhang, Z., Zhang, X., Lu, X., Liu, W., et al. (2019). Genome-wide Mapping of 5-Hydroxymethylcytosines in Circulating Cell-Free DNA as a Non-Invasive Approach for Early Detection of Hepatocellular Carcinoma. *Gut* 68 (12), 2195–2205. doi:10.1136/gutjnl-2019-318882

Cai, Z., Chen, G., Zeng, Y., Dong, X., Li, Z., Huang, Y., et al. (2019). Comprehensive Liquid Profiling of Circulating Tumor DNA and Protein Biomarkers in Long-Term Follow-Up Patients with Hepatocellular Carcinoma. *Clin. Cancer Res.* 25 (17), 5284–5294. doi:10.1158/1078-0432.Ccr-18-3477

Chan, K. C. A., Jiang, P., Chan, C. W. M., Sun, K., Wong, J., Hui, E. P., et al. (2013). Noninvasive Detection of Cancer-Associated Genome-Wide Hypomethylation and Copy Number Aberrations by Plasma DNA Bisulfite Sequencing. *Proc. Natl. Acad. Sci.* 110 (47), 18761–18768. doi:10.1073/pnas.1313995110

Chen, H., Sun, L.-y., Zheng, H.-q., Zhang, Q.-f., and Jin, X.-m. (2012). Total Serum DNA and DNA Integrity: Diagnostic Value in Patients with Hepatitis B Virus-Related Hepatocellular Carcinoma. *Pathology* 44 (4), 318–324. doi:10.1097/PAT.0b013e328353a24c

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 (2), 115–132. doi:10.3322/caac.21338

Cohen, J. D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., et al. (2018). Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test. *Science* 359 (6378), 926–930. doi:10.1126/science.aar3247

Corcoran, R. B., and Chabner, B. A. (2018). Application of Cell-Free DNA Analysis to Cancer Treatment. *N. Engl. J. Med.* 379 (18), 1754–1765. doi:10.1056/NEJMra1706174

Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. (2013). Liquid Biopsy: Monitoring Cancer-Genetics in the Blood. *Nat. Rev. Clin. Oncol.* 10 (8), 472–484. doi:10.1038/nrclinonc.2013.110

Diaz, L. A., Jr., and Bardelli, A. (2014). Liquid Biopsies: Genotyping Circulating Tumor DNA. *J. Clin. Oncol.* 32 (6), 579–586. doi:10.1200/jco.2012.45.2011

Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., et al. (2005). Detection and Quantification of Mutations in the Plasma of Patients with Colorectal Tumors. *Proc. Natl. Acad. Sci.* 102 (45), 16368–16373. doi:10.1073/pnas.0507904102

Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., et al. (2008). Circulating Mutant DNA to Assess Tumor Dynamics. *Nat. Med.* 14 (9), 985–990. doi:10.1038/nm.1789

Dor, Y., and Cedar, H. (2018). Principles of DNA Methylation and Their Implications for Biology and Medicine. *The Lancet* 392 (10149), 777–786. doi:10.1016/s0140-6736(18)31268-6

Eisenberg, E., Konopniki, M., Veitsman, E., Kramskay, R., Gaitini, D., and Baruch, Y. (2003). Prevalence and Characteristics of Pain Induced by Percutaneous Liver Biopsy. *Anesth. Analgesia* 96 (5), 1392–1396. table of contents. doi:10.1213/01.ane.0000060453.74744.17

El-Shazly, S. F., Eid, M. A., El-Sourogy, H. A., Attia, G. F., and Ezzat, S. A. (2010). Evaluation of Serum DNA Integrity as a Screening and Prognostic Tool in Patients with Hepatitis C Virus-Related Hepatocellular Carcinoma. *Int. J. Biol. Markers* 25 (2), 79–86. doi:10.1177/172460081002500204

Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., and Quake, S. R. (2010). Analysis of the Size Distributions of Fetal and Maternal Cell-free DNA by Paired-End Sequencing. *Clin. Chem.* 56 (8), 1279–1286. doi:10.1373/clinchem.2010.144188

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular Carcinoma. *Lancet* 391 (10127), 1301–1314. doi:10.1016/s0140-6736(18)30010-2

Gai, W., and Sun, K. (2019). Epigenetic Biomarkers in Cell-free DNA and Applications in Liquid Biopsy. *Genes* 10 (1), 32. doi:10.3390/genes10010032

Gao, Y.-J., He, Y.-J., Yang, Z.-L., Shao, H.-Y., Zuo, Y., Bai, Y., et al. (2010). Increased Integrity of Circulating Cell-Free DNA in Plasma of Patients with Acute Leukemia. *Clin. Chem. Lab. Med.* 48 (11), 1651–1656. doi:10.1515/cclm.2010.311

Gao, Y., Zhang, K., Xi, H., Cai, A., Wu, X., Cui, J., et al. (2017). Diagnostic and Prognostic Value of Circulating Tumor DNA in Gastric Cancer: A Meta-Analysis. *Oncotarget* 8 (4), 6330–6340. doi:10.18632/oncotarget.14064

García-Fernández, N., Macher, H. C., Rubio, A., Jiménez-Arriscado, P., Bernal-Bellido, C., Bellido-Díaz, M. L., et al. (2016). Detection of P53 Mutations in Circulating DNA of Transplanted Hepatocellular Carcinoma Patients as a Biomarker of Tumor Recurrence. *Adv. Exp. Med. Biol.* 924, 25–28. doi:10.1007/978-3-319-42044-8_5

Han, L.-Y., Fan, Y.-C., Mu, N.-N., Gao, S., Li, F., Ji, X.-F., et al. (2014). Aberrant DNA Methylation of G-Protein-Coupled Bile Acid Receptor Gpbar1 (TGR5) Is a Potential Biomarker for Hepatitis B Virus Associated Hepatocellular Carcinoma. *Int. J. Med. Sci.* 11 (2), 164–171. doi:10.7150/ijms.6745

Harding, J. J., Nandakumar, S., Armenia, J., Khalil, D. N., Albano, M., Ly, M., et al. (2019). Prospective Genotyping of Hepatocellular Carcinoma: Clinical Implications of Next-Generation Sequencing for Matching Patients to Targeted and Immune Therapies. *Clin. Cancer Res.* 25 (7), 2116–2126. doi:10.1158/1078-0432.Ccr-18-2293

Hlady, R., and Robertson, K. (2018). Genetic and Epigenetic Heterogeneity in Normal Liver Homeostasis and its Implications for Liver Disease and Hepatocellular Cancer. *Semin. Liver Dis.* 38 (1), 041–050. doi:10.1055/s-0037-1621712

Huang, Z.-H., Hu, Y., Hua, D., Wu, Y.-Y., Song, M.-X., and Cheng, Z.-H. (2011). Quantitative Analysis of Multiple Methylated Genes in Plasma for the Diagnosis and Prognosis of Hepatocellular Carcinoma. *Exp. Mol. Pathol.* 91 (3), 702–707. doi:10.1016/j.yexmp.2011.08.004

Huang, Z., Hua, D., Hu, Y., Cheng, Z., Zhou, X., Xie, Q., et al. (2012). Quantitation of Plasma Circulating DNA Using Quantitative PCR for the Detection of Hepatocellular Carcinoma. *Pathol. Oncol. Res.* 18 (2), 271–276. doi:10.1007/s12253-011-9438-z

Huang, A., Zhang, X., Zhou, S.-L., Cao, Y., Huang, X.-W., Fan, J., et al. (2016). Plasma Circulating Cell-free DNA Integrity as a Promising Biomarker for Diagnosis and Surveillance in Patients with Hepatocellular Carcinoma. *J. Cancer* 7 (13), 1798–1803. doi:10.7150/jca.15618

Huang, A., Zhang, X., Zhou, S.-L., Cao, Y., Huang, X.-W., Fan, J., et al. (2016). Detecting Circulating Tumor DNA in Hepatocellular Carcinoma Patients Using Droplet Digital PCR Is Feasible and Reflects Intratumoral Heterogeneity. *J. Cancer* 7 (13), 1907–1914. doi:10.7150/jca.15823

Ikeda, S., Tsigelny, I. F., Skjevik, Å. A., Kono, Y., Mendler, M., Kuo, A., et al. (2018). Next-Generation Sequencing of Circulating Tumor DNA Reveals Frequent

Alterations in Advanced Hepatocellular Carcinoma. *Oncol.* 23 (5), 586–593. doi:10.1634/theoncologist.2017-0479

Jen, J., Wu, L., and Sidransky, D. (2000). An Overview on the Isolation and Analysis of Circulating Tumor DNA in Plasma and Serum. *Ann. N. Y Acad. Sci.* 906, 8–12. doi:10.1111/j.1749-6632.2000.tb06581.x

Ji, X.-F., Fan, Y. C., Gao, S., Yang, Y., Zhang, J. J., and Wang, K. (2014). MT1MandMT1Gpromoter Methylation as Biomarkers for Hepatocellular Carcinoma. *World J. Gastroenterol.* 20 (16), 4723–4729. doi:10.3748/wjg.v20.i16.4723

Jiang, W.-W., Zahurak, M., Goldenberg, D., Milman, Y., Park, H. L., Westra, W. H., et al. (2006). Increased Plasma DNA Integrity index in Head and Neck Cancer Patients. *Int. J. Cancer* 119 (11), 2673–2676. doi:10.1002/ijc.22250

Jiang, P., Chan, C. W. M., Chan, K. C. A., Cheng, S. H., Wong, J., Wong, V. W.-S., et al. (2015). Lengthening and Shortening of Plasma DNA in Hepatocellular Carcinoma Patients. *Proc. Natl. Acad. Sci. USA* 112 (11), E1317–E1325. doi:10.1073/pnas.1500076112

Jiao, J., Watt, G. P., Stevenson, H. L., Calderone, T. L., Fisher-Hoch, S. P., Ye, Y., et al. (2018). Telomerase Reverse Transcriptase Mutations in Plasma DNA in Patients with Hepatocellular Carcinoma or Cirrhosis: Prevalence and Risk Factors. *Hepatol. Commun.* 2 (6), 718–731. doi:10.1002/hep4.1187

Jiao, J., Niu, W., Wang, Y., Baggerly, K., Ye, Y., Wu, X., et al. (2018). Prevalence of Aflatoxin-Associated TP53R249S Mutation in Hepatocellular Carcinoma in Hispanics in South Texas. *Cancer Prev. Res.* 11 (2), 103–112. doi:10.1158/1940-6207.CAPR-17-0235

Kaseb, A. O., Sánchez, N. S., Sen, S., Kelley, R. K., Tan, B., Bocobo, A. G., et al. (2019). Molecular Profiling of Hepatocellular Carcinoma Using Circulating Cell-free DNA. *Clin. Cancer Res.* 25 (20), 6107–6118. doi:10.1158/1078-0432.Ccr-18-3341

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W., and Vogelstein, B. (2011). Detection and Quantification of Rare Mutations with Massively Parallel Sequencing. *Proc. Natl. Acad. Sci.* 108 (23), 9530–9535. doi:10.1073/pnas.1105422108

Labgaa, I., Villacorta-Martin, C., D'Avola, D., Craig, A. J., von Felden, J., Martins-Filho, S. N., et al. (2018). A Pilot Study of Ultra-Deep Targeted Sequencing of Plasma DNA Identifies Driver Mutations in Hepatocellular Carcinoma. *Oncogene* 37 (27), 3740–3752. doi:10.1038/s41388-018-0206-3

Lehmann-Werman, R., Neiman, D., Zemmour, H., Moss, J., Magenheim, J., Vaknin-Dembinsky, A., et al. (2016). Identification of Tissue-specific Cell Death Using Methylation Patterns of Circulating DNA. *Proc. Natl. Acad. Sci. USA* 113 (13), E1826–E1834. doi:10.1073/pnas.1519286113

Leon, S. A., Shapiro, B., Sklaroff, D. M., and Yaros, M. J. (1977). Free DNA in the Serum of Cancer Patients and the Effect of Therapy. *Cancer Res.* 37 (3), 646–650.

Li, W., Li, Q., Kang, S., Same, M., Zhou, Y., Sun, C., et al. (2018). CancerDetector: Ultrasensitive and Non-invasive Cancer Detection at the Resolution of Individual Reads Using Cell-free DNA Methylation Sequencing Data. *Nucleic Acids Res.* 46 (15), e89. doi:10.1093/nar/gky423

Li, J., Han, X., Yu, X., Xu, Z., Yang, G., Liu, B., et al. (2018). Clinical Applications of Liquid Biopsy as Prognostic and Predictive Biomarkers in Hepatocellular Carcinoma: Circulating Tumor Cells and Circulating Tumor DNA. *J. Exp. Clin. Cancer Res.* 37 (1), 213. doi:10.1186/s13046-018-0893-1

Liao, W., Yang, H., Xu, H., Wang, Y., Ge, P., Ren, J., et al. (2016). Noninvasive Detection of Tumor-Associated Mutations from Circulating Cell-free DNA in Hepatocellular Carcinoma Patients by Targeted Deep Sequencing. *Oncotarget* 7 (26), 40481–40490. doi:10.18632/oncotarget.9629

Lim, C., Bhangui, P., Salloum, C., Gómez-Gavara, C., Lahat, E., Luciani, A., et al. (2017). Impact of Time to Surgery in the Outcome of Patients with Liver Resection for BCLC 0-A Stage Hepatocellular Carcinoma. *J. Hepatol.* 68, 100–108. doi:10.1016/j.jhep.2017.09.017

Liu, Z. J., Huang, Y., Wei, L., He, J. Y., Liu, Q. Y., Yu, X. Q., et al. (2017). Combination of LINE-1 Hypomethylation and RASSF1A Promoter Hypermethylation in Serum DNA Is a Non-invasion Prognostic Biomarker for Early Recurrence of Hepatocellular Carcinoma after Curative Resection. *Neoplasma* 64 (5), 795–802. doi:10.4149/neo_2017_519

Liu, Z., Zhang, Y., Shi, C., Zhou, X., Xu, K., Jiao, D., et al. (2021). A Novel Immune Classification Reveals Distinct Immune Escape Mechanism and Genomic Alterations: Implications for Immunotherapy in Hepatocellular Carcinoma. *J. Transl Med.* 19 (1), 5. doi:10.1186/s12967-020-02697-y

Liu, Z., Liu, L., Lu, T., Wang, L., Li, Z., Jiao, D., et al. (2021). Hypoxia Molecular Characterization in Hepatocellular Carcinoma Identifies One Risk Signature and Two Nomograms for Clinical Management. *J. Oncol.* 2021, 1–20. doi:10.1155/2021/6664386

Liu, Z., Wang, L., Liu, L., Lu, T., Jiao, D., Sun, Y., et al. (2021). The Identification and Validation of Two Heterogenous Subtypes and a Risk Signature Based on Ferroptosis in Hepatocellular Carcinoma. *Front. Oncol.* 11, 619242. doi:10.3389/fonc.2021.619242

Llovet, J. M., Montal, R., Sia, D., and Finn, R. S. (2018). Molecular Therapies and Precision Medicine for Hepatocellular Carcinoma. *Nat. Rev. Clin. Oncol.* 15 (10), 599–616. doi:10.1038/s41571-018-0073-4

Lo, Y. M. D., Zhang, J., Leung, T. N., Lau, T. K., Chang, A. M. Z., and Hjelm, N. M. (1999). Rapid Clearance of Fetal DNA from Maternal Plasma. *Am. J. Hum. Genet.* 64 (1), 218–224. doi:10.1086/302205

Mandel, P., and Metais, P. (1948). Nuclear Acids in Human Blood Plasma. *C R. Seances Soc. Biol. Fil* 142 (3-4), 241–243. doi:10.1159/000300537

Marchio, A., Amougou Atsama, M., Béré, A., Komas, N.-P., Noah Noah, D., Atangana, P. J. A., et al. (2018). Droplet Digital PCR Detects High Rate of TP53 R249S Mutants in Cell-free DNA of Middle African Patients with Hepatocellular Carcinoma. *Clin. Exp. Med.* 18 (3), 421–431. doi:10.1007/s10238-018-0502-9

Marzese, D. M., Hirose, H., and Hoon, D. S. B. (2013). Diagnostic and Prognostic Value of Circulating Tumor-Related DNA in Cancer Patients. *Expert Rev. Mol. Diagn.* 13 (8), 827–844. doi:10.1586/14737159.2013.845088

Mehra, N., Dolling, D., Sumanasuriya, S., Christova, R., Pope, L., Carreira, S., et al. (2018). Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-Resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). *Eur. Urol.* 74 (3), 283–291. doi:10.1016/j.eururo.2018.02.013

Méndez-Blanco, C., Fondevila, F., García-Palomo, A., González-Gallego, J., and Mauriz, J. L. (2018). Sorafenib Resistance in Hepatocarcinoma: Role of Hypoxia-Inducible Factors. *Exp. Mol. Med.* 50 (10), 1–9. doi:10.1038/s12276-018-0159-1

Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., et al. (2011). High Fragmentation Characterizes Tumour-Derived Circulating DNA. *PLoS One* 6 (9), e23418. doi:10.1371/journal.pone.0023418

Ng, C. K. Y., Di Costanzo, G. G., Tosti, N., Paradiso, V., Coto-Llerena, M., Roscigno, G., et al. (2018). Genetic Profiling Using Plasma-Derived Cell-free DNA in Therapy-Naïve Hepatocellular Carcinoma Patients: a Pilot Study. *Ann. Oncol.* 29 (5), 1286–1291. doi:10.1093/annonc/mdy083

Oh, J. E., Lim, H. S., An, C. H., Jeong, E. G., Han, J. Y., Lee, S. H., et al. (2010). Detection of Low-Level KRAS Mutations Using PNA-Mediated Asymmetric PCR Clamping and Melting Curve Analysis with Unlabeled Probes. *J. Mol. Diagn.* 12 (4), 418–424. doi:10.2353/jmoldx.2010.090146

Pezzuto, F., Buonaguro, L., Buonaguro, F., and Tornesello, M. (2018). The Role of Circulating Free DNA and MicroRNA in Non-invasive Diagnosis of HBV- and HCV-Related Hepatocellular Carcinoma. *Int. J. Mol. Sci.* 19 (4), 1007. doi:10.3390/ijms19041007

Qu, C., Wang, Y., Wang, P., Chen, K., Wang, M., Zeng, H., et al. (2019). Detection of Early-Stage Hepatocellular Carcinoma in Asymptomatic HBsAg-Seropositive Individuals by Liquid Biopsy. *Proc. Natl. Acad. Sci. USA* 116 (13), 6308–6312. doi:10.1073/pnas.1819799116

Russo, F. P., Imondi, A., Lynch, E. N., and Farinati, F. (2018). When and How Should We Perform a Biopsy for HCC in Patients with Liver Cirrhosis in 2018? A Review. *Dig. Liver Dis.* 50 (7), 640–646. doi:10.1016/j.dld.2018.03.014

Silva, M. A., Hegab, B., Hyde, C., Guo, B., Buckels, J. A. C., and Mirza, D. F. (2008). Needle Track Seeding Following Biopsy of Liver Lesions in the Diagnosis of Hepatocellular Cancer: a Systematic Review and Meta-Analysis. *Gut* 57 (11), 1592–1596. doi:10.1136/gut.2008.149062

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* 122 (1), e59. doi:10.1002/cpmb.59

Strickler, J. H., Loree, J. M., Ahronian, L. G., Parikh, A. R., Niedzwiecki, D., Pereira, A. A. L., et al. (2018). Genomic Landscape of Cell-Free DNA in Patients with Colorectal Cancer. *Cancer Discov.* 8 (2), 164–173. doi:10.1158/2159-8290.Cd-17-1009

Swanton, C. (2012). Intratumor Heterogeneity: Evolution through Space and Time. *Cancer Res.* 72 (19), 4875–4882. doi:10.1158/0008-5472.Can-12-2217

Tangkijvanich, P., Hourpai, N., Rattanatanyong, P., Wisedopas, N., Mahachai, V., and Mutirangura, A. (2007). Serum LINE-1 Hypomethylation as a Potential Prognostic Marker for Hepatocellular Carcinoma. *Clinica Chim. Acta* 379 (1-2), 127–133. doi:10.1016/j.cca.2006.12.029

Tokuhisa, Y., Iizuka, N., Sakaida, I., Moribe, T., Fujita, N., Miura, T., et al. (2007). Circulating Cell-free DNA as a Predictive Marker for Distant Metastasis of Hepatitis C Virus-Related Hepatocellular Carcinoma. *Br. J. Cancer* 97 (10), 1399–1403. doi:10.1038/sj.bjc.6604034

Tuaeva, N. O., Abramova, Z. I., and Sofronov, V. V. (2008). The Origin of Elevated Levels of Circulating DNA in Blood Plasma of Premature Neonates. *Ann. N. Y Acad. Sci.* 1137, 27–30. doi:10.1196/annals.1448.043

Umetani, N., Giuliano, A. E., Hiramatsu, S. H., Amersi, F., Nakagawa, T., Martino, S., et al. (2006). Prediction of Breast Tumor Progression by Integrity of Free Circulating DNA in Serum. *J. Clin. Oncol.* 24 (26), 4270–4276. doi:10.1200/jco.2006.05.9493

Umetani, N., Kim, J., Hiramatsu, S., Reber, H. A., Hines, O. J., Bilchik, A. J., et al. (2006). Increased Integrity of Free Circulating DNA in Sera of Patients with Colorectal or Periampullary Cancer: Direct Quantitative PCR for ALU Repeats. *Clin. Chem.* 52 (6), 1062–1069. doi:10.1373/clinchem.2006.068577

Valpione, S., Gremel, G., Mundra, P., Middlehurst, P., Galvani, E., Girotti, M. R., et al. (2018). Plasma Total Cell-free DNA (cfDNA) Is a Surrogate Biomarker for Tumour burden and a Prognostic Biomarker for Survival in Metastatic Melanoma Patients. *Eur. J. Cancer* 88, 1–9. doi:10.1016/j.ejca.2017.10.029

Villanueva, A., Portela, A., Sayols, S., Battiston, C., Hoshida, Y., Méndez-González, J., et al. (2015). DNA Methylation-Based Prognosis and Epidrivers in Hepatocellular Carcinoma. *Hepatology* 61 (6), 1945–1956. doi:10.1002/hep.27732

von Felden, J., Craig, A. J., Garcia-Lezana, T., Labgaa, I., Haber, P. K., D'Avola, D., et al. (2021). Mutations in Circulating Tumor DNA Predict Primary Resistance to Systemic Therapies in Advanced Hepatocellular Carcinoma. *Oncogene* 40 (1), 140–151. doi:10.1038/s41388-020-01519-1

Wang, B. G., Huang, H. Y., Chen, Y. C., Bristow, R. E., Kassauei, K., Cheng, C. C., et al. (2003). Increased Plasma DNA Integrity in Cancer Patients. *Cancer Res.* 63 (14), 3966–3968.

Wong, I. H., Lo, Y. M., Yeo, W., Lau, W. Y., and Johnson, P. J. (2000). Frequent P15 Promoter Methylation in Tumor and Peripheral Blood from Hepatocellular Carcinoma Patients. *Clin. Cancer Res.* 6 (9), 3516–3521.

Wong, I. H., Zhang, J., Lai, P. B., Lau, W. Y., and Lo, Y. M. (2003). Quantitative Analysis of Tumor-Derived Methylated p16INK4a Sequences in Plasma, Serum, and Blood Cells of Hepatocellular Carcinoma Patients. *Clin. Cancer Res.* 9 (3), 1047–1052.

Wu, J., Tang, W., Huang, L., Hou, N., Wu, J., Cheng, X., et al. (2019). The Analysis of Cell-free DNA Concentrations and Integrity in Serum of Initial and Treated of Lymphoma Patients. *Clin. Biochem.* 63, 59–65. doi:10.1016/j.clinbiochem.2018.10.002

Xu, R.-h., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., et al. (2017). Circulating Tumour DNA Methylation Markers for Diagnosis and Prognosis of Hepatocellular Carcinoma. *Nat. Mater* 16 (11), 1155–1161. doi:10.1038/nmat4997

Yan, L., Chen, Y., Zhou, J., Zhao, H., Zhang, H., and Wang, G. (2018). Diagnostic Value of Circulating Cell-free DNA Levels for Hepatocellular Carcinoma. *Int. J. Infect. Dis.* 67, 92–97. doi:10.1016/j.ijid.2017.12.002

Yang, Y.-J., Chen, H., Huang, P., Li, C.-H., Dong, Z.-H., and Hou, Y.-L. (2011). Quantification of Plasma hTERT DNA in Hepatocellular Carcinoma Patients by Quantitative Fluorescent Polymerase Chain Reaction. *Clin. Invest. Med.* 34 (4), 238. doi:10.25011/cim.v34i4.15366

Yang, J. D., Mohamed, E. A., Aziz, A. O. A., Shousha, H. I., Hashem, M. B., Nabeel, M. M., et al. (2017). Characteristics, Management, and Outcomes of Patients with Hepatocellular Carcinoma in Africa: A Multicountry Observational Study from the Africa Liver Cancer Consortium. *Lancet Gastroenterol. Hepatol.* 2 (2), 103–111. doi:10.1016/s2468-1253(16)30161-3

Yang, J. D., Liu, M. C., and Kisiel, J. B. (2019). Circulating Tumor DNA and Hepatocellular Carcinoma. *Semin. Liver Dis.* 39 (4), 452–462. doi:10.1055/s-0039-1688503

Ye, Q., Ling, S., Zheng, S., and Xu, X. (2019). Liquid Biopsy in Hepatocellular Carcinoma: Circulating Tumor Cells and Circulating Tumor DNA. *Mol. Cancer* 18 (1), 114. doi:10.1186/s12943-019-1043-x

Zhong, S., Yeo, W., Tang, M. W., Wong, N., Lai, P. B., and Johnson, P. J. (2003). Intensive Hypermethylation of the CpG Island of Ras Association Domain Family 1A in Hepatitis B Virus-Associated Hepatocellular Carcinomas. *Clin. Cancer Res.* 9 (9), 3376–3382.

Zill, O. A., Greene, C., Sebisanovic, D., Siew, L. M., Leng, J., Vu, M., et al. (2015). Cell-Free DNA Next-Generation Sequencing in Pancreatobiliary Carcinomas. *Cancer Discov.* 5 (10), 1040–1048. doi:10.1158/2159-8290.Cd-15-0274

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 Zhang, Liu, Ji, Li, Wang, Ren, Liu, Chen, Han, Meng, Li 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fmolb.2021.715728

Development and Verification of an Immune-Related Gene Pairs Prognostic Signature in Hepatocellular Carcinoma

Xiaofei Feng[1†], *Shanshan Mu*[2†], *Yao Ma*[3†] *and Wenji Wang*[1*]

[1]Department of Orthopedics, Lanzhou University First Affiliated Hospital, Lanzhou, China, [2]Pediatric Rheumatism Immunology Department, Lanzhou University Second Hospital, Lanzhou, China, [3]Clinical Laboratory Center, Gansu Provincial Maternity and Child-care Hospital, Lanzhou, China

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Giuliana Cavalloni,
Fondazione del Piemonte per l'Oncologia (IRCCS), Italy
Udhaya Kumar S,
Vellore Institute of Technology, India
Sanjay Mishra,
The Ohio State University, United States
Harpreet Kaur,
Pine Biotech, United States

***Correspondence:**
Wenji Wang
ldyyjzwwj@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 27 May 2021
Accepted: 21 September 2021
Published: 01 October 2021

Citation:
Feng X, Mu S, Ma Y and Wang W (2021) Development and Verification of an Immune-Related Gene Pairs Prognostic Signature in Hepatocellular Carcinoma. Front. Mol. Biosci. 8:715728. doi: 10.3389/fmolb.2021.715728

With the increasing prevalence of Hepatocellular carcinoma (HCC) and the poor prognosis of immunotherapy, reliable immune-related gene pairs (IRGPs) prognostic signature is required for personalized management and treatment of patients. Gene expression profiles and clinical information of HCC patients were obtained from the TCGA and ICGC databases. The IRGPs are constructed using immune-related genes (IRGs) with large variations. The least absolute shrinkage and selection operator (LASSO) regression analysis was used to construct IRGPs signature. The IRGPs signature was verified through the ICGC cohort. 1,309 IRGPs were constructed from 90 IRGs with high variability. We obtained 50 IRGPs that were significantly connected to the prognosis and constructed a signature that included 17 IRGPs. In the TCGA and ICGC cohorts, patients were divided into high and low-risk patients by the IRGPs signature. The overall survival time of low-risk patients is longer than that of high-risk patients. After adjustment for clinical and pathological factors, multivariate analysis showed that the IRGPs signature is an independent prognostic factor. The Receiver operating characteristic (ROC) curve confirmed the accuracy of the signature. Besides, gene set enrichment analysis (GSEA) revealed that the signature is related to immune biological processes, and the immune microenvironment status is distinct in different risk patients. The proposed IRGPs signature can effectively assess the overall survival of HCC, and provide the relationship between the signature and the reactivity of immune checkpoint therapy and the sensitivity of targeted drugs, thereby providing new ideas for the diagnosis and treatment of the disease.

Keywords: hepatocellular carcinoma, immune-related gene pairs, prognosis, tumor immune environment, targeting therapy

Abbreviations: AUC, area under curve; GO, gene ontology; GSEA, gene set enrichment analysis; HCC, hepatocellular carcinoma; HR, hazard ratio; ICGC, international cancer genome consortium; IRG(s), immune-related gene(s); IRGP(s), immune-related gene pair(s); KEGG, kyoto encyclopedia of genes and genomes; LASSO, least absolute shrinkage and selection operator; PCA, principal component analysis; ROC, receiver operating characteristic; TCGA, the cancer genome atlas.

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common malignant tumors and the leading cause of cancer-related deaths, and its morbidity and mortality are increasing (Forner et al., 2018; Villanueva, 2019). According to the latest cancer statistics in 2019, approximately 42,030 people in the United States are diagnosed with liver cancer each year, and 31,780 people die from the disease (Siegel et al., 2019). Although breakthroughs in the diagnosis and treatment of HCC have been made in recent years, the prognostic outcome of patients remains poor (Yang et al., 2019). The current clinical application of immunotherapy in HCC has benefited some patients, but this approach has not effectively improved the prognosis, and the long-term survival rate is still very poor (Dong et al., 2020). Due to the complexity and heterogeneity of HCC, individualized decision-making plans are required in diagnosis and treatment. Therefore, it is necessary to identify the novel prognostic signature of HCC and use it to guide clinical treatment as an effective way to improve the prognosis of patients.

Understanding the characteristics of tumor immune cell infiltration can improve the responsiveness of immunotherapy and is of great help in understanding the mechanism of cancer occurrence (Ino et al., 2013; Binnewies et al., 2018). The landscape of immune cells in HCC mapped by single-cell sequencing and other evidence indicate that the tumor immune microenvironment plays an indispensable role in the progression of HCC (Nishida and Kudo, 2017; Zheng et al., 2017; Kurebayashi et al., 2018). Different immune cells infiltrating into HCC have different prognostic effects (Prieto et al., 2015; Cariani and Missale, 2019). It has been observed clinically that the increase of $PD1^{+}CD8^{+}$ T cells in HCC is associated with poor clinical outcomes (Chew et al., 2017). Although numerous studies have found the importance of immunology in HCC, its molecular mechanism is still unclear. Current studies have shown that tumor immune-related markers have commendable effects on the diagnosis and treatment of cancer (Peng et al., 2019; Hong et al., 2020). There have been many studies on the prognostic value of identifying key genes to build models to predict the prognosis of HCC patients (Li et al., 2017a; Kaur et al., 2019; Liu et al., 2020; Ouyang et al., 2020). However, there is no in-depth survey on the clinical relevance and prognostic significance of IRGPs in HCC.

In this study, we used the HCC gene expression datasets of the TCGA and ICGC databases to develop individualized prognostic signature based on IRGPs. Then, we evaluated its ability to predict prognosis in HCC patients and its responsiveness to immune checkpoint therapy and targeted therapy.

MATERIALS AND METHODS

Data Sources

The HCC level 3 RNA-seq data and clinical information were downloaded from the TCGA database, including normal tissues (n = 50) and tumor tissues (n = 374). 365 cases of tumor patients have survival time and survival status, of which 247 cases contain clinical information (age, gender, histologic grade, TNM stage, vascular invasion, and alpha fetoprotein). The RNA-seq data and clinical information of HCC were downloaded from the ICGC database, including normal tissues (n = 202) and tumor tissues (n = 243). 232 tumor patients have survival time, survival status, and clinical information (age, gender, and TNM stage). The list of IRGs was downloaded from the IMMPORT (Bhattacharya et al., 2018). Both the TCGA and ICGC data are publicly available. Therefore, this research does not require the approval of the local ethics committee.

Construction of an Individualized Prognostic Signature Based on IRGPs

In the TCGA and ICGC cohorts, the limma package (Ritchie et al., 2015) was used to identify the IRGs that are different between cancer tissues and normal tissues. The filtering criteria were $|\log FC| > 1$, FDR < 0.01. We identified the common differential IRGs in the two cohorts. The protein-protein interaction network was used to demonstrate differential genes (Udhaya Kumar et al., 2021), and the Cytoscape software (v3.8.2) was used for visualization. Genes with the median absolute deviation greater than 1 in tumor samples whose expression levels of these common differential IRGs were considered as candidate genes. The expression levels of these candidate genes were compared in pairs to generate a score for each IRGP. If IRG 1 is less than IRG 2, the IRGP score is 1, otherwise, the IRGP score is 0 (Li et al., 2017b). In the TCGA cohort, univariate cox was used to select prognostic IRGPs to assess the association between each IRGP and the overall survival rate of the patient. Prognostic IRGPs with a p value of less than 0.001 were candidates for establishing an IRGPs signature. From these IRGPs, the R language was used to perform Lasso Cox proportional hazard regression to construct the risk score, and finally 17 IRGPs were used to define the IRGPs signature. To classify patients into high- or low-risk patients, the optimal IRGPs signature cut-off value was determined by the ROC curve.

Verification of IRGPs Signature

In the TCGA and ICGC cohorts, the survival package and survminer package (https://CRAN.R-project.org/package = survminer) were used to establish the survival curve of the high and low-risk patients through the Kaplan-Meier diagram, and the Log-rank test was used to analyze the difference in survival curves. Cox proportional hazards analysis was used for univariate and multivariate analysis. The survivalROC package (https://CRAN.R-project.org/package = survivalROC) was used for ROC curve analysis to evaluate the predictive ability of the IRGPs signature.

Functional Annotation and Immune Cell Infiltration

To understand the underlying molecular mechanism of the IRGPs signature, we divided the TCGA cohort patients into high- and low-risk patients so that the software GSEA 4.1.0 (http://www.gsea-msigdb.org) can be used for gene ontology

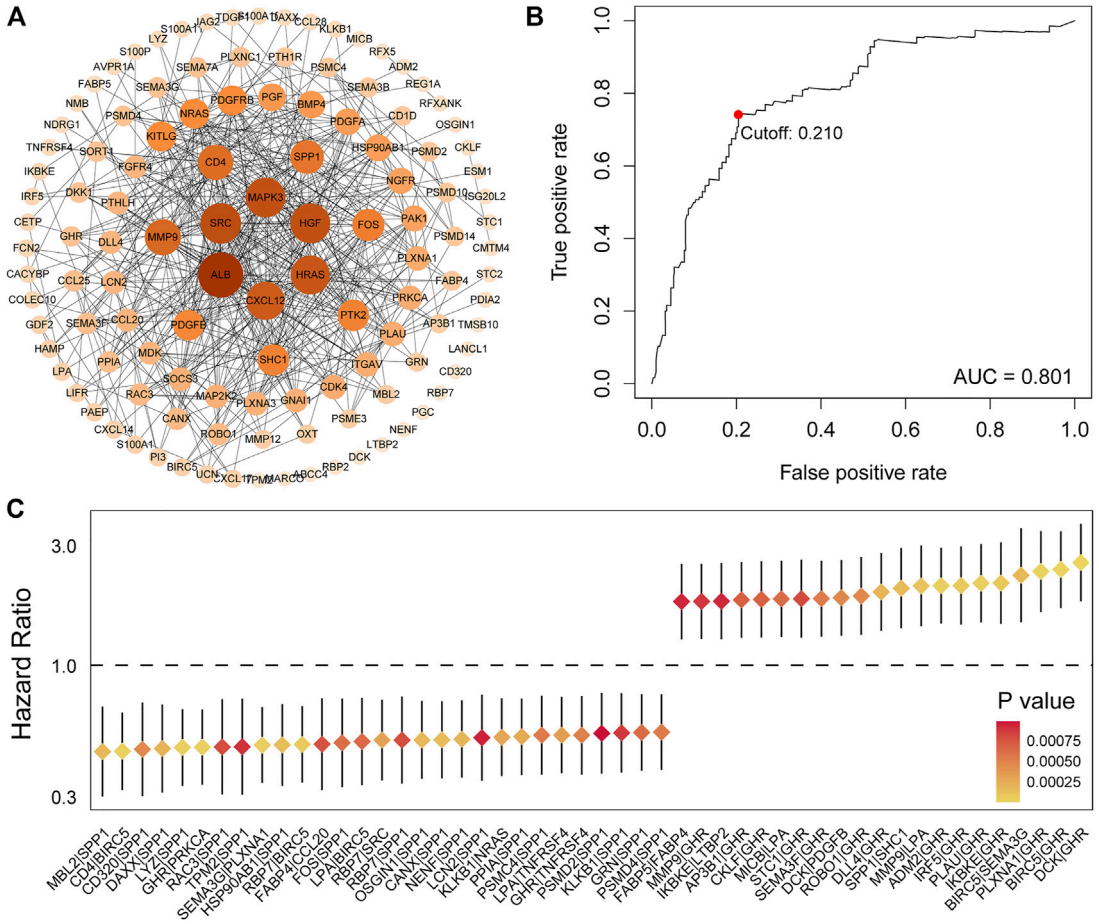


FIGURE 1 | Identification of IRGs and determination of risk score. **(A)** 120 shared differential immunity related genes were displayed on the protein-protein interaction network. **(B)** Univariate Cox analysis obtained 50 IRGPs related to prognosis. **(C)** ROC curve of the risk score of the IRGPs signature in the TCGA cohort. The risk score was 0.210, which served as a threshold for classifying patients into the high and low-risk patients.

(GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations. When $p < 0.05$ and FDR < 0.25, the enriched gene set is considered to be statistically significant. The gsva package (Hänzelmann et al., 2013) was used to quantify immune cell infiltrations, and the correlation between high- and low-risk patients and 16 immune cell infiltrations and 13 immune-related pathways was evaluated.

Explore the Relationship Between the IRGPs Signature and the Reactivity of Immune Checkpoint Therapy and the Sensitivity of Targeted Drugs

Immune checkpoint blockers of immunotherapy have obvious effects in the treatment of many human cancers. Therefore, we explored the expression and correlation between high and low-risk patients and the current common immune checkpoints. In addition, small molecule tyrosine kinase inhibitors, as a class of molecularly targeted drugs, have become one of the mainstream trends in current anti-liver cancer research. The pRRophetic package is used to analyze the sensitivity of tyrosine kinase inhibitors in high- and low-risk patients (Geeleher et al., 2014).

RESULTS

Construction and Definition of the IRGPs Signature

Principal component analysis (PCA) was first performed on the expression profiles of the two cohorts. The results showed that there were differences in the distribution of tumor tissue and normal tissue in the two data sets, which could be used for further analysis (**Supplementary Figures S1A,C**). 231 differential IRGs were identified in the TCGA cohort and 172 differential IRGs were identified in the ICGC cohort (**Supplementary Figures S1B,D**). We obtained 120 shared differential IRGs and displayed them in the protein-protein interaction network (**Figure 1A**). In order to obtain genes with high variability, we screened among 120 IRGs. The filter criterion was that the median absolute deviation of these genes in tumor samples is greater than 1. 90

TABLE 1 | Baseline characteristics of the patients in different risk groups.

Variables	TCGA cohort				ICGC cohort			
	High risk	low risk	$\chi2$	P	High risk	low risk	$\chi2$	P
Age (year)	—	—	0	1	—	—	150.09	<0.001
<60	50 (50.0%)	119 (46.3%)	—	—	5 (16.4%)	36 (23.6%)	—	—
≧60	58 (50.0%)	138 (53.7%)	—	—	58 (83.6%)	129 (76.4%)	—	—
Gender	—	—	0.1	0.75	—	—	5.13	0.02
female	37 (34.3%)	82 (31.9%)	—	—	25 (37.3%)	36 (21.8%)	—	—
male	71 (65.7%)	175 (68.1%)	—	—	42 (62.7%)	129 (78.2%)	—	—
Histologic grade	—	—	19.71	<0.001	—	—	—	—
G1 and G2	50 (46.3%)	180 (70.0%)	—	—	—	—	—	—
G3 and G4	57 (52.8%)	73 (28.4%)	—	—	—	—	—	—
unknow	1 (0.9%)	4 (1.6%)			—	—	—	—
Child–Pugh grade		—	6.49	0.04	—	—	—	—
A	53 (49.1%)	163 (63.4%)	—	—	—	—	—	—
B and C	8 (7.4%)	14 (5.4%)	—	—	—	—	—	—
unknow	47 (43.5%)	80 (31.1%)	—	—	—	—	—	—
Alpha fetoprotein	—	—	15.69	<0.001	—	—	—	—
<200 ng/ml	43 (39.8%)	158 (61.5%)	—	—	—	—	—	—
≧200 ng/ml	33 (30.6%)	42 (16.3%)	—	—	—	—	—	—
unknow	32 (29.6%)	57 (22.2%)	—	—	—	—	—	—
Vascular invasion	—	—	20.46	<0.001	—	—	—	—
YES	40 (37.0%)	66 (25.7%)	—	—	—	—	—	—
NO	42 (38.9%)	163 (63.4%)	—	—	—	—	—	—
unknow	26 (24.1%)	28 (10.9%)	—	—	—	—		—
TNM stage	—	—	13.57	<0.001	—	—	9.76	<0.001
I and II	65 (60.2%)	189 (73.5%)	—	—	30 (44.8%)	112 (67.9%)	—	—
III and IV	39 (36.1%)	48 (18.7%)	—	—	37 (55.2%)	53 (32.1%)	—	—
unknow	4 (3.7%)	20 (7.8%)	—	—	—	—	—	—

IRGs were finally got for constructing 1,309 pairs of IRGPs. The correlation between these IRGPs and overall survival was evaluated in the TCGA cohort, and 50 IRGPs related to prognosis were screened out, with $p < 0.001$ as the cut-off criterion (**Figure 1C**). A signature composed of 17 IRGPs was constructed on the TCGA cohort through LASSO Cox proportional hazard regression. The IRGPs signature consists of 22 unique IRGs (**Supplementary Table S1**). Finally, the time-dependent ROC curve was used to determine the optimal cutoff value of 0.210 for the risk score, classified patients as high- or low-risk patients (**Figure 1B**). At this point, the sensitivity was 74.1% and the specificity was 75.5%. In the TCGA cohort, the high-risk patients were found to be significantly correlated with histologic grade, Child-Pugh grade, alpha-fetoprotein, vascular invasion, and TNM stage (**Table 1**). Overall survival in the TCGA cohort was worse in the high-risk patients than in the low-risk patients (Log-rank, $p < 0.001$) (**Figure 2A**). PCA analysis shows that the distribution patterns of patients in different risk patients are contrasting (**Figure 2B**). The ROC curve assessed the prognostic ability of the prognostic signature, and the area under the curve (AUC) reaches 0.809 at 1 year, 0.757 at 2 years, and 0.712 at 3 years (**Figure 2C**).

Validation of the Feasibility of the IRGPs Signature to Predict Survival

To determine whether the signature has prognostic value, the signature was applied to the ICGC cohort as independent external verification. Patients in the ICGC cohort were divided into high or low-risk patients based on the above risk score. The high-risk patients in the ICGC cohort were also correlated with the TNM stage (**Table 1**). Also, the overall survival rate of the high-risk patients in the ICGC cohort is lower than in the low-risk patients (Log-rank, $p < 0.001$) (**Figure 2D**). PCA analysis showed that the distribution pattern of the two groups of patients was distinct (**Figure 2E**). In addition, the ROC curve showed 0.772 at 1 year, 0.744 at 2 years, and 0.759 at 3 years (**Figure 2F**). These results are similar to those obtained in the TCGA cohort.

Validation of the IRGPs Signature as an Independent Prognostic Factor

In the TCGA cohort, univariate COX analysis showed that TNM stage ($p = 0.002$), vascular invasion ($p = 0.035$), and the IRGPs signature ($p < 0.001$) were significantly correlated with the prognosis of HCC. After adjusting for clinical and pathological factors such as age, gender, histologic grade, TNM stage, vascular invasion, and alpha-fetoprotein, the TNM stage (HR, 2.098; 95% CI, 1.226,3.590; $p = 0.007$) and the IRGPs signature (HR, 3.688; 95%CI, 2.222,6.119; $p < 0.001$) were independent risk factors in the multivariate COX analysis (**Figure 3A**). Similarly, in the ICGC cohort, univariate COX analysis showed that TNM stage ($p < 0.001$) and the IRGPs signature ($p < 0.001$) were also significantly correlated with the prognosis of hepatocellular carcinoma. Multivariate COX analysis revealed that the TNM stage (HR, 1.882; 95%CI, 1.319,2.685; $p < 0.001$) and the IRGPs signature (HR, 4.340; 95% CI, 2.272, 8.291; $p < 0.001$) were also independent risk factors of overall survival (**Figure 3B**). In the

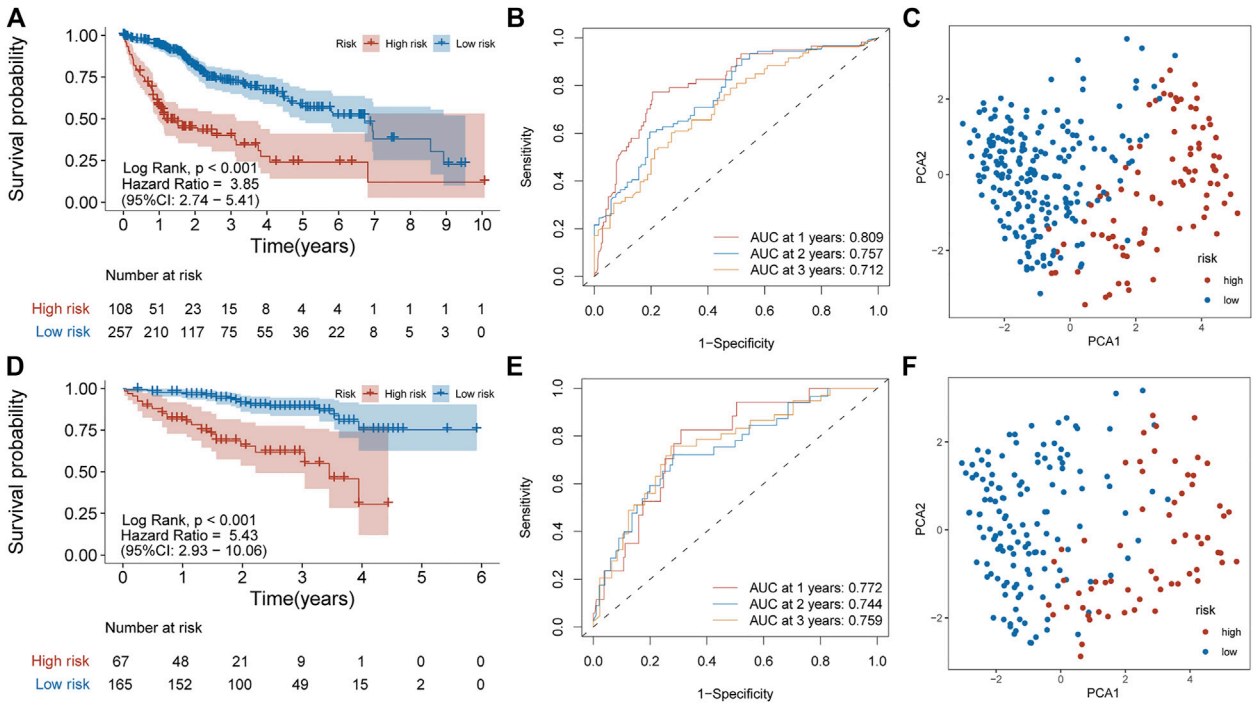


FIGURE 2 | Prognostic analysis of the IRGPs signature in TCGA and ICGC cohorts. **(A)** Kaplan-Meier curve of overall survival in the high- and low-risk patients in the TCGA cohort. **(B)** ROC curve in TCGA cohort. **(C)** PCA plot of TCGA cohort. **(D)** Kaplan-Meier curve of overall survival in the high- and low-risk patients in the ICGC cohort. **(E)** ROC curve in ICGC cohort **(F)** PCA plot of ICGC cohort.

TCGA and ICGC cohorts, high and low-risk patients have different distributions in the TNM stage (**Supplementary Figures S2A,C**). In addition, the IRGPs signature divided patients with early (I and II) and late (III and IV) HCC into different prognostic groups. Also, for patients with stage IandII disease, low-risk patients have a good prognosis in both the TCGA cohort (Log-rank, $p < 0.001$) and the ICGC cohort (Log-rank, $p < 0.001$) (**Supplementary Figures S2B,D**). Similarly, for patients with advanced stages III and IV, low-risk patients also have a good prognosis in the TCGA cohort (Log-rank, $p < 0.001$) and ICGC cohort (Log-rank, $p = 0.01$) (**Supplementary Figures S2B,D**). Overall, the IRGPs signature seems to be able to independently assess the overall survival of HCC.

Functional Annotation and Immune Cell Infiltration Between High and Low-Risk Patients

GSEA results showed that some immune-related biological processes are involved in high-risk patients (**Supplementary Figure S3A**), such as activation of innate immune response, antigen processing and presentation of peptide antigen via MHC class I, positive regulation of activated T cell proliferation, regulation of type I interferon mediated signalling pathway. Interestingly, some immune-related KEGG pathways are enriched in high-risk patients (**Supplementary Figures S3B**), such as FC epsilon RI signalling pathway, MAPK signalling pathway, mTOR signalling pathway, NOD like receptor signalling pathway. These results indicate that immune-related biological processes may play an indispensable role in the development of HCC. We further explored the status of immune cells and immune-related functions in high-risk and low-risk populations. In the TCGA cohort, the high-risk patients were positively correlated with tumor-infiltrating immune cells (aDCs, Macrophages, TH2 cells, Treg), while negatively correlated with Mast cells, Neutrophils, and NK cells (**Figure 4A**). In the ICGC cohort, except for Mast cells, which were not statistically significant, the results were consistent with the results in TCGA (**Figure 4C**). In the TCGA cohort, the high-risk patients were positively correlated with MHC class I and negatively correlated with Type II IFN Response (**Figure 4B**). In the ICGC cohort, the same result was obtained (**Figure 4D**).

Reactivity of Immune Checkpoint Therapy and Sensitivity of Targeted Drugs

Immune checkpoint therapy has shown better results in the treatment of cancer, and it has made a major breakthrough in the field of HCC. Therefore, we investigated the expression of immune checkpoint markers in low and high-risk patients. In the TCGA and ICGC cohorts, immune checkpoint markers (CD276, HHLA2, TNFRSF18, TNFSF9, LGALS9) were expressed higher in high-risk patients, and there was a positive correlation between the IRGPs signature and these markers ($|R| > 0.3$, $p < 0.05$, **Figures 5A–E**). Tyrosine kinase inhibitors currently approved and in clinical trials have demonstrated efficacy in the treatment

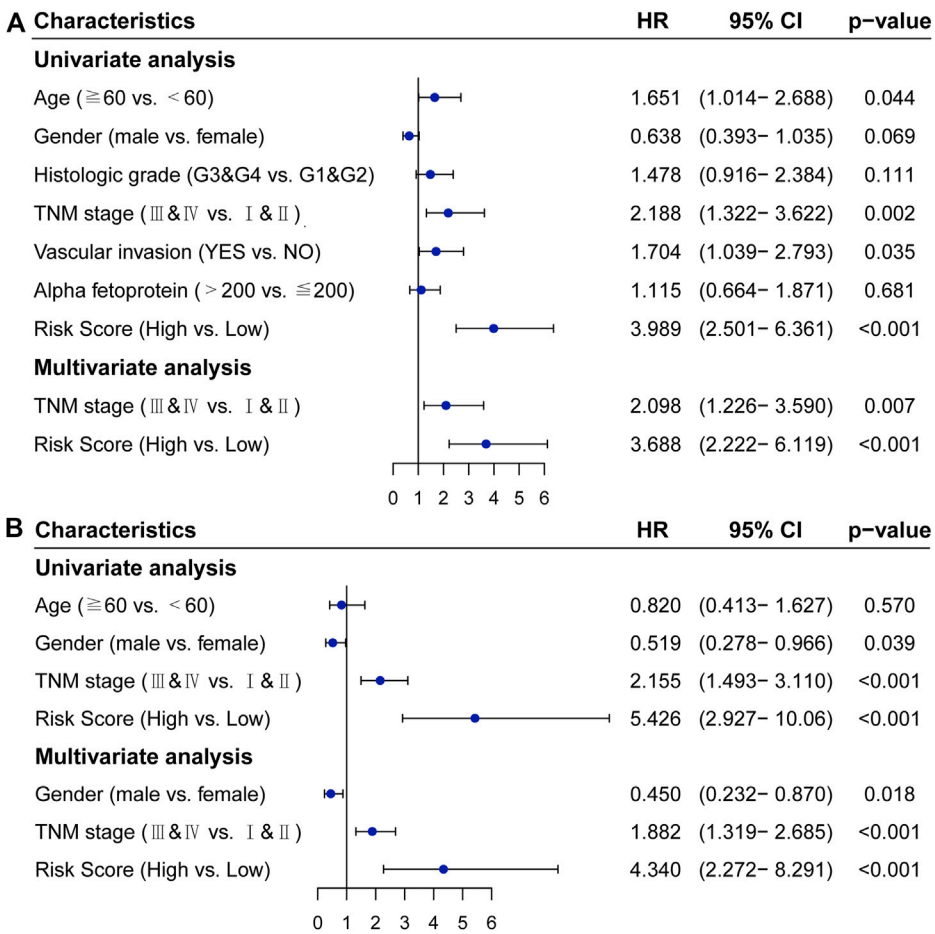


FIGURE 3 | Cox proportional hazards regression model analysis of overall survival in patients with hepatocellular carcinoma. **(A)** TCGA cohort. **(B)** ICGC cohort. HR: hazard ratio, CI: confidence interval.

of hepatocellular carcinoma, which can be helpful in the treatment and management of patients with HCC. Hence, we evaluated the IC50 of each sample and observed that the IC50 of the eight small molecule tyrosine kinase inhibitors was significantly different between the two groups. The results showed that Axitinib, Motesanib, Dasatinib, Nilotinib, Erlotinib, Pazopanib, Lapatinib, and Saracatinib are more sensitive to low-risk patients ($p < 0.001$, **Figures 6A–H**). This may provide an accurate strategy for the treatment of HCC patients.

DISCUSSION

Immune infiltration plays an important role in cancer progression. With the exploration in the field of immunotherapy, it is of great help in the treatment of tumor. There have been studies that have provided strong evidence for the treatment and diagnosis of some diseases through bioinformatics analysis (Kumar et al., 2019; Udhaya Kumar et al., 2020). Due to the high heterogeneity of HCC, some patients fail to achieve the expected curative effect on immunotherapy. Therefore, it is extremely crucial to determine the sensitivity of different patient subsets to treatment response. Individualized treatment for different patient subgroups will help to enhance the prognosis of patients (Gan et al., 2020; Zhou et al., 2020; Kaur et al., 2021). Novel signature related to tumor immune infiltration may be a sword for identifying new molecular targets and improving patient prognosis.

In this study, we developed a prognostic signature based on 17 IRGPs in HCC and validated them in two independent data sets on different platforms. Our prognostic immune signature can further divide clinically defined patients [for example, early stage (I and II) and late stage (III and IV)] into subgroups with different survival outcomes. Univariate COX analysis showed that TNM stage and the IRGPs signature were significantly correlated with the prognosis of HCC. By multivariate COX analysis, the TNM stage and the prognostic signature can be used as independent prognostic factors. Therefore, our prognostic signature can be used as a personalized prognosis and diagnosis and treatment of HCC patients and can be easily translated into clinical practice.

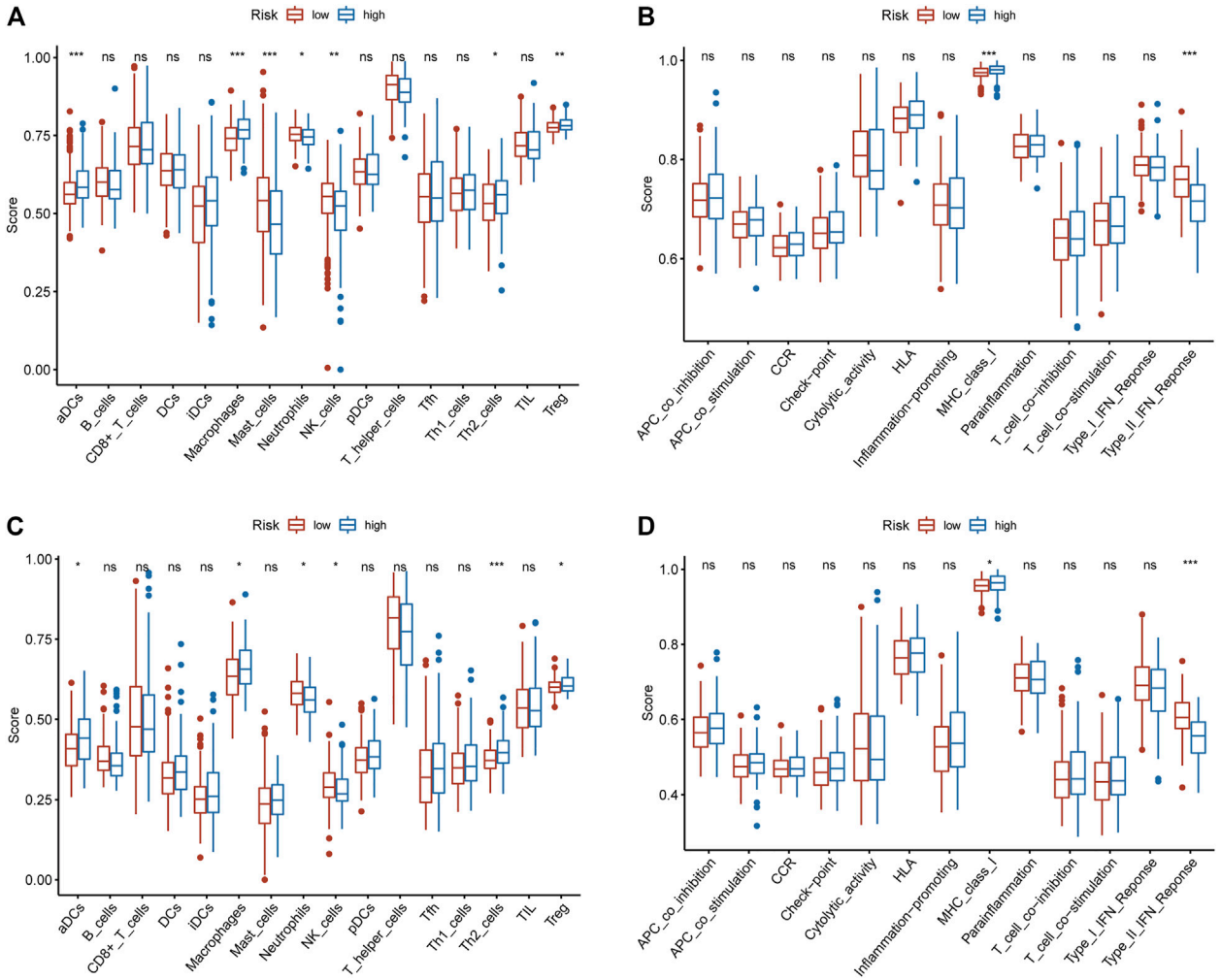


FIGURE 4 | The relationship between the IRGPs signature and immune infiltrating cells and immune-related functions. **(A)** The relationship between the IRGPs signature and immune cells in the TCGA cohort. **(B)** The relationship between the IRGPs signature and immune-related functions in the TCGA cohort. **(C)** The relationship between the IRGPs signature and immune cells in the ICGC cohort. **(D)** The relationship between the IRGPs signature and immune-related functions in the ICGC cohort. The adjusted *p* value is ns, which is not significant. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The IRGPs signature we constructed included 22 IRGs, and the results of GSEA indicated that some immune-related biological processes and signal pathways were enriched in high-risk patients. Such as, activation of innate immune response, MAPK signalling pathway and mTOR signalling pathway, etc. Innate immunity is a part of the HCC tumor microenvironment, which can suppress and promote cancer. For example, dendritic cells, neutrophils, and macrophages can promote the occurrence of HCC, while natural killer cells and natural killer T cells can inhibit the development of HCC (Ruf et al., 2021). The activation of MAPK signalling pathway is closely related to the development of tumors, and it is activated in about 50% of patients with early HCC and almost all patients with advanced HCC (Neuzillet et al., 2014). Excessive activation of mTOR promotes the development of tumors, and affects the immune regulation involved in the differentiation of immune cells, and plays an important role in tumor metabolism (Zou et al., 2020). There may be a link between aberrant activation of these pathways for the low overall survival of high-risk patients. We further studied the relationship between tumor immune infiltration and the IRGPs signature. In the high-risk patients of the TCGA and ICGC cohorts, aDCs, Macrophages, TH2 cells, and Treg infiltration increased. Studies have shown that the increase of regulatory DCs can promote the increase of Treg in liver cancer (Cheng et al., 2016). Treg and Macrophages, which are less infiltration in HCC, have a good prognosis (Zhu et al., 2008; Zhou et al., 2016). The increase of TH2 cells is associated with poor prognosis of HCC (Zhu et al., 2008; Zhou et al., 2010). In this study, in the TCGA and ICGC cohorts, the infiltration of Neutrophils and NK cells in high-risk patients was reduced, while the prognosis of the high-risk patients was worse than that of the low-risk patients. Studies have pointed out that neutrophils can inhibit the development of cancer (Souto et al., 2011). NK cells infiltrate less in advanced HCC and are associated with poor

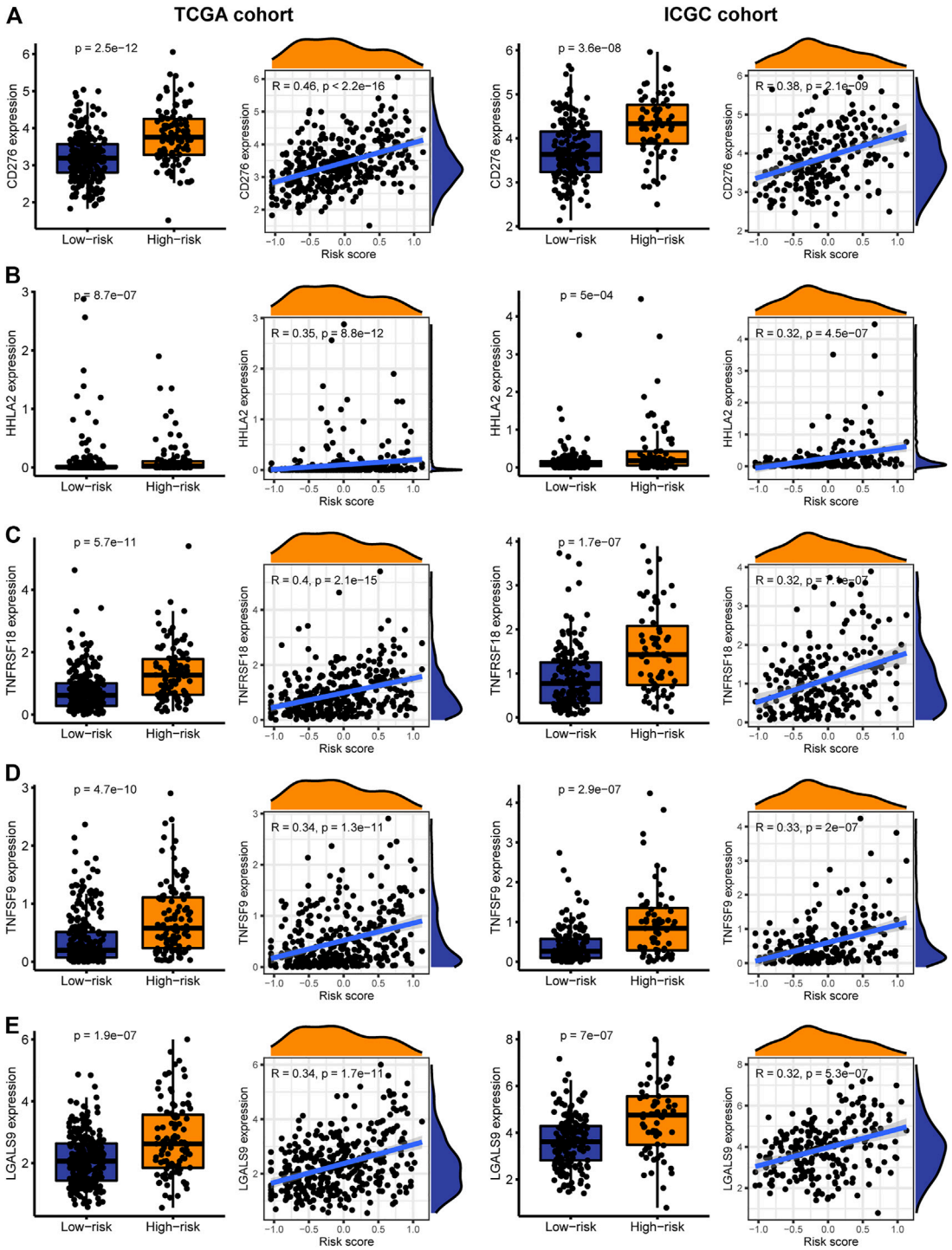


FIGURE 5 | The expression and correlation between immune checkpoints in high and low-risk patients in the TARGET and ICGC cohorts. **(A)** CD276 **(B)** HHLA2 **(C)** TNFRSF18 **(D)** TNFSF9 **(E)** LGALS9.

prognosis (Wu et al., 2013). These results are consistent with our research results. In the high-risk patients in the TCGA and ICGC cohorts, the type II IFN reactivity decreased, while the MHC class I activity increased. Type II IFN is mainly produced by activated NK cells (Perussia, 1991), and it plays an important role in regulating the tumor immune environment. In our study, the NK cell invasion and type II IFN reactivity in the high-risk patients were reduced, and the prognosis of the high-risk patients was poor, which was consistent with the above study results. The level of MHC class I is elevated in patients with

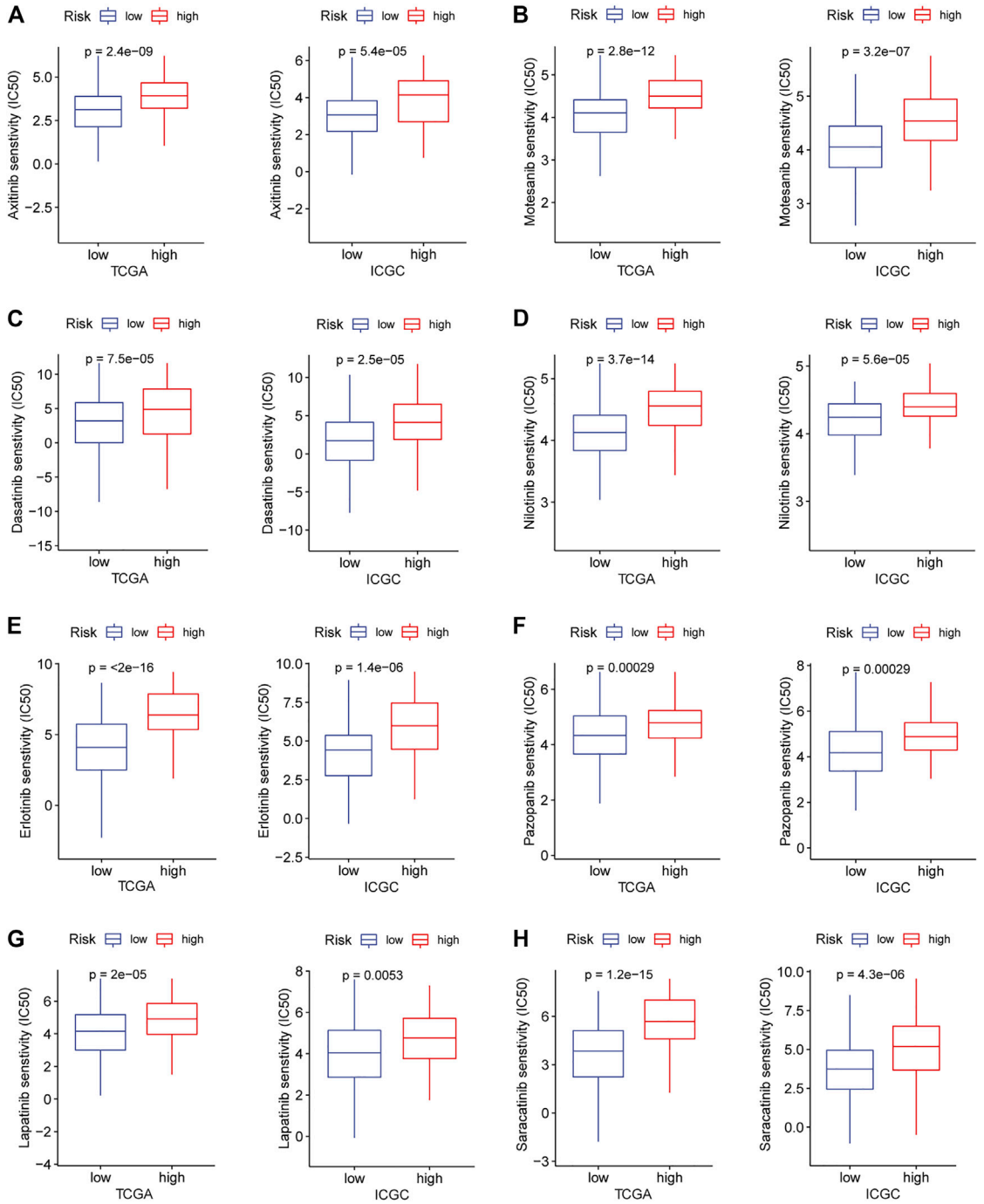


FIGURE 6 | In the TARGET and ICGC cohorts, high and low-risk patients and targeted drug sensitivity. **(A)** Axitinib **(B)** Motesanib **(C)** Dasatinib **(D)** Nilotinib **(E)** Erlotinib **(F)** Pazopanib **(G)** Lapatinib **(H)** Saracatinib.

advanced HCC, and it may negatively regulate innate immunity and adaptive immunity to cause tumor escape mechanisms to occur (Jinushi et al., 2005). According to the above findings, dysregulation of the immune microenvironment may account for the survival differences between the IRGPs signature subgroups.

Among the immune checkpoint markers, CD276, HHLA2, TNFRSF18, TNFSF9, and LGALS9 were highly expressed in high-risk patients, indicating differences in responsiveness to these immune checkpoint treatments among patients grouped by this signature. CD276 plays an important role in innate immunity and T cell-mediated adaptive immunity, which is highly expressed in HCC and other cancers and is associated with poor patient prognosis. And it has great potential for immunotherapy (Picarda et al., 2016). HHLA2 can inhibit the function of CD4

and CD8 T cells, and blocking HHLA2 can enhance the proliferation and activation of T cells, which is helpful for cancer immunotherapy (Zhao et al., 2013; Rieder et al., 2021). TNFRSF18, also known as glucocorticoid-induced TNFR-related protein (GITR), is a costimulatory receptor in malignant tumors. Agonistic targeting of GITR can enhance the anti-tumor response of TIL derived from HCC patients (van Beek et al., 2019). TNFSF9, also known as 4-1BBL, is expressed on active T cells and antigen cells. 4-1BBL targeted immunotherapy has shown anti-tumor effects in the treatment of HCC (Xiao et al., 2007; Li et al., 2011). The high expression of LGALS9 is associated with the poor prognosis of many human cancers. LGALS9 preferentially kills T cells rather than cancer cells may assist cancer immune escape (Yang et al., 2021).

Several tyrosine kinase inhibitors have been used as first or second-line agents in the treatment of liver cancer, but one challenge is the lack of reliable biomarkers to identify patients who benefit from these treatments (Qin et al., 2019). Interestingly, we found that low-risk patients are more sensitive to Axitinib, Motesanib, Dasatinib, Nilotinib, Erlotinib, Pazopanib, Lapatinib, and Saracatinib. These findings provide effective treatment strategies for patients stratified by the IRGPs signature.

Although we used two independent data sets to rigorously verify the signature we proposed, our research has certain limitations. First of all, our study only carried out retrospective research and lacked prospective research. Second, our study only includes immune-related genes, and important prognostic genes in HCC may have been deleted. Finally, the relationship between the IRGPs signature and immune infiltration will be verified in follow-up studies.

In conclusion, the proposed IRGPs signature can accurately predict the prognosis of HCC patients and guide clinicians to make specific treatment decisions. It also provides the relationship between this signature and the responsiveness of immune checkpoints and targeted drugs. These results will be beneficial to the effective treatment of HCC patients. Prospective studies are required to further verify its accuracy.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/; https://daco.icgc.org/; https://www.immport.org/home.

AUTHOR CONTRIBUTIONS

SM and YM: conceptualization. XF and SM: software and methodology. XF and YM: writing—original draft preparation. WW: writing—review and editing. All authors have read and agreed to the published version of the article.

ACKNOWLEDGMENTS

We would like to acknowledge the TCGA and the ICGC for providing data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.715728/full#supplementary-material

REFERENCES

Beek, A. A., Zhou, G., Doukas, M., Boor, P. P. C., Noordam, L., Mancham, S., et al. (2019). GITR Ligation Enhances Functionality of Tumor-Infiltrating T Cells in Hepatocellular Carcinoma. *Int. J. Cancer* 145 (4), 1111–1124. doi:10.1002/ijc.32181

Bhattacharya, S., Dunn, P., Thomas, C. G., Smith, B., Schaefer, H., Chen, J., et al. (2018). ImmPort, toward Repurposing of Open Access Immunological Assay Data for Translational and Clinical Research. *Sci. Data* 5, 180015. doi:10.1038/sdata.2018.15

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

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

Cheng, J.-T., Deng, Y.-N., Yi, H.-M., Wang, G.-Y., Fu, B.-S., Chen, W.-J., et al. (2016). Hepatic Carcinoma-Associated Fibroblasts Induce Ido-producing Regulatory Dendritic Cells through IL-6-mediated STAT3 Activation. *Oncogenesis* 5 (2), e198. doi:10.1038/oncsis.2016.7

Chew, V., Lai, L., Pan, L., Lim, C. J., Li, J., Ong, R., et al. (2017). Delineation of an Immunosuppressive Gradient in Hepatocellular Carcinoma Using High-Dimensional Proteomic and Transcriptomic Analyses. *Proc. Natl. Acad. Sci. USA* 114 (29), E5900–E5909. doi:10.1073/pnas.1706559114

Dong, L.-Q., Peng, L.-H., Ma, L.-J., Liu, D.-B., Zhang, S., Luo, S.-Z., et al. (2020). Heterogeneous Immunogenomic Features and Distinct Escape Mechanisms in Multifocal Hepatocellular Carcinoma. *J. Hepatol.* 72 (5), 896–908. doi:10.1016/j.jhep.2019.12.014

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular Carcinoma. *The Lancet* 391 (10127), 1301–1314. doi:10.1016/S0140-6736(18)30010-2

Gan, X., Luo, Y., Dai, G., Lin, J., Liu, X., Zhang, X., et al. (2020). Identification of Gene Signatures for Diagnosis and Prognosis of Hepatocellular Carcinomas Patients at Early Stage. *Front. Genet.* 11, 857. doi:10.3389/fgene.2020.00857

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

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

Hong, W., Liang, L., Gu, Y., Qi, Z., Qiu, H., Yang, X., et al. (2020). Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. *Mol. Ther. - Nucleic Acids* 22, 937–947. doi:10.1016/j.omtn.2020.10.002

Ino, Y., Yamazaki-Itoh, R., Shimada, K., Iwasaki, M., Kosuge, T., Kanai, Y., et al. (2013). Immune Cell Infiltration as an Indicator of the Immune Microenvironment of Pancreatic Cancer. *Br. J. Cancer* 108 (4), 914–923. doi:10.1038/bjc.2013.32

Jinushi, M., Takehara, T., Tatsumi, T., Hiramatsu, N., Sakamori, R., Yamaguchi, S., et al. (2005). Impairment of Natural Killer Cell and Dendritic Cell Functions by the Soluble Form of MHC Class I-Related Chain A in Advanced Human Hepatocellular Carcinomas. *J. Hepatol.* 43 (6), 1013–1020. doi:10.1016/j.jhep.2005.05.026

Kaur, H., Dhall, A., Kumar, R., and Raghava, G. P. S. (2019). Identification of Platform-Independent Diagnostic Biomarker Panel for Hepatocellular

Carcinoma Using Large-Scale Transcriptomics Data. *Front. Genet.* 10, 1306. doi:10.3389/fgene.2019.01306

Kaur, H., Lathwal, A., and Raghava, G. P. S. (2021). Integrative Multi-Omics Approach for Stratification of Tumor Recurrence Risk Groups of Hepatocellular Carcinoma Patients. *bioRxiv*. doi:10.1101/2021.03.03.433841

Kumar, S. U., Kumar, D. T., Siva, R., Doss, C. G. P., and Zayed, H. (2019). Integrative Bioinformatics Approaches to Map Potential Novel Genes and Pathways Involved in Ovarian Cancer. *Front. Bioeng. Biotechnol.* 7, 391. doi:10.3389/fbioe.2019.00391

Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., et al. (2018). Landscape of Immune Microenvironment in Hepatocellular Carcinoma and its Additional Impact on Histological and Molecular Classification. *Hepatology* 68 (3), 1025–1041. doi:10.1002/hep.29904

Li, B., Cui, Y., Diehn, M., and Li, R. (2017a). Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. *JAMA Oncol.* 3 (11), 1529–1537. doi:10.1001/jamaoncol.2017.1609

Li, B., Feng, W., Luo, O., Xu, T., Cao, Y., Wu, H., et al. (2017b). Development and Validation of a Three-Gene Prognostic Signature for Patients with Hepatocellular Carcinoma. *Sci. Rep.* 7 (1), 5517. doi:10.1038/s41598-017-04811-5

Li, G., Wu, X., Zhang, F., Li, X., Sun, B., Yu, Y., et al. (2011). Triple Expression of B7-1, B7-2 and 4-1BBL Enhanced Antitumor Immune Response against Mouse H22 Hepatocellular Carcinoma. *J. Cancer Res. Clin. Oncol.* 137 (4), 695–703. doi:10.1007/s00432-010-0905-9

Liu, M., Liu, X., Liu, S., Xiao, F., Guo, E., Qin, X., et al. (2020). Big Data-Based Identification of Multi-Gene Prognostic Signatures in Liver Cancer. *Front. Oncol.* 10, 847. doi:10.3389/fonc.2020.00847

Neuzillet, C., Tijeras-Raballand, A., de Mestier, L., Cros, J., Faivre, S., and Raymond, E. (2014). MEK in Cancer and Cancer Therapy. *Pharmacol. Ther.* 141 (2), 160–171. doi:10.1016/j.pharmthera.2013.10.001

Nishida, N., and Kudo, M. (2017). Immunological Microenvironment of Hepatocellular Carcinoma and its Clinical Implication. *Oncology* 92 (Suppl. 1), 40–49. doi:10.1159/000451015

Ouyang, G., Yi, B., Pan, G., and Chen, X. (2020). A Robust Twelve-Gene Signature for Prognosis Prediction of Hepatocellular Carcinoma. *Cancer Cel Int* 20, 207. doi:10.1186/s12935-020-01294-9

Peng, D., Wang, L., Li, H., Cai, C., Tan, Y., Xu, B., et al. (2019). An Immune Infiltration Signature to Predict the Overall Survival of Patients with colon Cancer. *IUBMB life* 71 (11), 1760–1770. doi:10.1002/iub.2124

Perussia, B. (1991). Lymphokine-Activated Killer Cells, Natural Killer Cells and Cytokines. *Curr. Opin. Immunol.* 3 (1), 49–55. doi:10.1016/0952-7915(91)90076-d

Picarda, E., Ohaegbulam, K. C., and Zang, X. (2016). Molecular Pathways: Targeting B7-H3 (CD276) for Human Cancer Immunotherapy. *Clin. Cancer Res.* 22 (14), 3425–3431. doi:10.1158/1078-0432.CCR-15-2428

Prieto, J., Melero, I., and Sangro, B. (2015). Immunological Landscape and Immunotherapy of Hepatocellular Carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 12 (12), 681–700. doi:10.1038/nrgastro.2015.173

Qin, S., Li, A., Yi, M., Yu, S., Zhang, M., and Wu, K. (2019). Recent Advances on Anti-Angiogenesis Receptor Tyrosine Kinase Inhibitors in Cancer Therapy. *J. Hematol. Oncol.* 12 (1), 27. doi:10.1186/s13045-019-0718-5

Rieder, S. A., Wang, J., White, N., Qadri, A., Menard, C., Stephens, G., et al. (2021). B7-H7 (HHLA2) Inhibits T-Cell Activation and Proliferation in the Presence of TCR and CD28 Signaling. *Cell Mol Immunol* 18 (6), 1503–1511. doi:10.1038/s41423-020-0361-7

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

Ruf, B., Heinrich, B., and Greten, T. F. (2021). Immunobiology and Immunotherapy of HCC: Spotlight on Innate and Innate-Like Immune Cells. *Cel Mol Immunol* 18 (1), 112–127. doi:10.1038/s41423-020-00572-w

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

Souto, J. C., Vila, L., and Brú, A. (2011). Polymorphonuclear Neutrophils and Cancer: Intense and Sustained Neutrophilia as a Treatment against Solid Tumors. *Med. Res. Rev.* 31 (3), 311–363. doi:10.1002/med.20185

Udhaya Kumar, S., Madhana Priya, N., Thirumal Kumar, D., Anu Preethi, V., Kumar, V., Nagarajan, D., et al. (2021). An Integrative Analysis to Distinguish between Emphysema (EML) and Alpha-1 Antitrypsin Deficiency-Related Emphysema (ADL)-A Systems Biology Approach. *Adv. Protein Chem. Struct. Biol.* 127, 315–342. doi:10.1016/bs.apcsb.2021.02.004

Udhaya Kumar, S., Thirumal Kumar, D., Bithia, R., Sankar, S., Magesh, R., Sidenna, M., et al. (2020). Analysis of Differentially Expressed Genes and Molecular Pathways in Familial Hypercholesterolemia Involved in Atherosclerosis: A Systematic and Bioinformatics Approach. *Front. Genet.* 11, 734. doi:10.3389/fgene.2020.00734

Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380 (15), 1450–1462. doi:10.1056/NEJMra1713263

Wu, Y., Kuang, D.-M., Pan, W.-D., Wan, Y.-L., Lao, X.-M., Wang, D., et al. (2013). Monocyte/macrophage-elicited Natural Killer Cell Dysfunction in Hepatocellular Carcinoma Is Mediated by CD48/2B4 Interactions. *Hepatology* 57 (3), 1107–1116. doi:10.1002/hep.26192

Xiao, H., Huang, B., Yuan, Y., Li, D., Han, L.-F., Liu, Y., et al. (2007). Soluble PD-1 Facilitates 4-1bbl-Triggered Antitumor Immunity against Murine H22 Hepatocarcinoma *In Vivo*. *Clin. Cancer Res.* 13 (6), 1823–1830. doi:10.1158/1078-0432.CCR-06-2154

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019). A Global View of Hepatocellular Carcinoma: Trends, Risk, Prevention and Management. *Nat. Rev. Gastroenterol. Hepatol.* 16 (10), 589–604. doi:10.1038/s41575-019-0186-y

Yang, R., Sun, L., Li, C.-F., Wang, Y.-H., Yao, J., Li, H., et al. (2021). Galectin-9 Interacts with PD-1 and TIM-3 to Regulate T Cell Death and Is a Target for Cancer Immunotherapy. *Nat. Commun.* 12 (1), 832. doi:10.1038/s41467-021-21099-2

Zhao, R., Chinai, J. M., Buhl, S., Scandiuzzi, L., Ray, A., Jeon, H., et al. (2013). HHLA2 Is a Member of the B7 Family and Inhibits Human CD4 and CD8 T-Cell Function. *Proc. Natl. Acad. Sci.* 110 (24), 9879–9884. doi:10.1073/pnas.1303524110

Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., et al. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* 169 (7), 1342–1356. doi:10.1016/j.cell.2017.05.035

Zhou, H., Huang, H., Shi, J., Zhao, Y., Dong, Q., Jia, H., et al. (2010). Prognostic Value of Interleukin 2 and Interleukin 15 in Peritumoral Hepatic Tissues for Patients with Hepatitis B-Related Hepatocellular Carcinoma after Curative Resection. *Gut* 59 (12), 1699–1708. doi:10.1136/gut.2010.218404

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, T., Cai, Z., Ma, N., Xie, W., Gao, C., Huang, M., et al. (2020). A Novel Ten-Gene Signature Predicting Prognosis in Hepatocellular Carcinoma. *Front. Cel Dev. Biol.* 8, 629. doi:10.3389/fcell.2020.00629

Zhu, X.-D., Zhang, J.-B., Zhuang, P.-Y., Zhu, H.-G., Zhang, W., Xiong, Y.-Q., et al. (2008). High Expression of Macrophage Colony-Stimulating Factor in Peritumoral Liver Tissue Is Associated with Poor Survival after Curative Resection of Hepatocellular Carcinoma. *J. Clin. Oncol.* 26 (16), 2707–2716. doi:10.1200/jco.2007.15.6521

Zou, Z., Tao, T., Li, H., and Zhu, X. (2020). mTOR Signaling Pathway and mTOR Inhibitors in Cancer: Progress and Challenges. *Cell Biosci* 10, 31. doi:10.1186/s13578-020-00396-1

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 Feng, Mu, Ma 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fmolb.2021.745649

Pan-Cancer Analysis of the Associations of TGFBI Expression With Prognosis and Immune Characteristics

Yun Chen[1†], Han Zhao[2,3,4†], Yao Feng[1], Qin Ye[1], Jing Hu[1], Yue Guo[1] and Yunzhi Feng[1]**

[1]Department of Stomatology, The Second Xiangya Hospital, Central South University, Changsha, China, [2]Department of Ophthalmology, Eye, Ear, Nose, and Throat Hospital of Fudan University, Shanghai, China, [3]Laboratory of Myopia, NHC Key Laboratory of Myopia (Fudan University), Chinese Academy of Medical Sciences, Shanghai, China, [4]Shanghai Key Laboratory of Visual Impairment and Restoration, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong, SAR China

Reviewed by:
Pallavi Shrivastava, Universidad Católica de Santa María, Peru
Ashok Kumar, All India Institute of Medical Sciences Bhopal, India

***Correspondence:**
Yue Guo
ggyue74@126.com
Yunzhi Feng
fengyunzhi001@csu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *22 July 2021*
Accepted: *17 September 2021*
Published: *04 October 2021*

Citation:
Chen Y, Zhao H, Feng Y, Ye Q, Hu J, Guo Y and Feng Y (2021) Pan-Cancer Analysis of the Associations of TGFBI Expression With Prognosis and Immune Characteristics. Front. Mol. Biosci. 8:745649. doi: 10.3389/fmolb.2021.745649

Transforming growth factor-beta-induced (TGFBI) protein has important roles in tumor growth, metastasis, and immunity. However, there is currently no pan-cancer evidence regarding TGFBI. In this study, we conducted a pan-cancer analysis of TGFBI mRNA and protein expression and prognoses of various cancer types using public databases. We also investigated the associations of *TGFBI* expression with tumor microenvironment (TME) components, immune cell infiltration, tumor mutational burden (TMB), and microsatellite instability (MSI), along with the *TGFBI* genetic alteration types. The results showed that *TGFBI* expression varied among different cancer types, and it was positively or negatively related to prognosis in various cancers. *TGFBI* expression was also significantly correlated with TME components, TMB, MSI, immune cell infiltration, and immunoinhibitory and immunostimulatory gene subsets. These findings indicate that *TGFBI* participates in various immune responses and it may function as a prognostic marker in various cancers. The findings may be useful for developing immunotherapies that target *TGFBI*.

Keywords: TGFBI, pan-cancer, immune subtype, tumor microenvironment, prognoses, drug responses

INTRODUCTION

Transforming growth factor-beta-induced (TGFBI) protein, also known as keratoepithelin or βig-H3, is involved in extracellular matrix (ECM) formation. It can help to bind integrin to ECM proteins such as fibronectin, laminin, and collagen. It was first found in human lung adenocarcinoma cell line. TGFBI protein (68 kDa) has a secretory signal peptide sequence at the N-terminus, a cysteine-rich domain (known as EMI), an Arg-Gly-Asp (RGD) motif, and four internal homologous repeat domains (known as fasciclin 1 [FAS1]) at the C-terminus. The RGD motif and FAS1 domain play key roles in tumorigenesis and development (Skonier et al., 1992). TGFBI is a conserved connective protein that has important roles in cell proliferation, differentiation, adhesion, migration, embryonic development, and inflammation (Kim et al., 2002; Thapa et al., 2007).

In recent years, an increasing number of cancer research has found that TGFBI plays a crucial role in tumor growth, metastasis, and immunosuppression (Wang et al., 2019; Fico and Santamaria-Martínez, 2020; Steitz et al., 2020). The FAS1 domain and RGD motif can bind to integrins $\alpha 3\beta 1$, $\alpha 5\beta 1$, and $\alpha 6\beta 1$ to regulate the PI3K-AKT-mTOR signaling pathway, reduce cell death or promote cell survival, and increase metastasis and angiogenesis (Razumilava and Gores, 2014). TGFBI can also activate the

FAK-MAPK-ERK signaling pathway to increase $Ca2^+$ and regulate calpain. It thereby stimulates matrix-metalloproteinase (MMP) secretion and alterations in the tumor microenvironment (TME) and cell adhesion, ultimately leading to tumorigenesis, tumor invasion, and metastasis (Ma et al., 2012). Interestingly, TGFBI exhibits dual functions in ovarian cancer (OV) by acting as both a tumor promoter and suppressor. Some studies have shown that upregulated TGFBI can act as a tumor promoter in esophageal squamous cell carcinoma (Ozawa et al., 2016), colon cancer (Ma et al., 2008), gastric cancer (Han et al., 2015), and bladder cancer (Bhagirath et al., 2012). For example, TGFBI is highly expressed in glioblastoma multiforme (GBM) and is related to DC cell infiltration, which is an adverse biomarker of GBM (Yin et al., 2020). In addition, downregulated TGFBI, acting as a tumor suppresser, can cause tumor growth in breast cancer (Li et al., 2012), lung cancer (Wen et al., 2011a), and mesothelioma (Wen et al., 2011b). However, there is still no pan-cancer information about the roles of TGFBI in various types of cancer based on large clinical datasets.

With the rapid development of various public databases, pan-cancer analysis can be used to obtain a profile of any gene of interest, including its associations with cancer (based on analyses comparing tumor and matched normal tissues) and with prognosis, and its potential molecular mechanisms. Here, for the first time, we conducted a pan-cancer analysis of TGFBI mRNA and protein expression and prognoses across various cancer types using several public databases such as, Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), Oncomine, and Human Protein Atlas (HPA) databases. We also assessed the associations between *TGFBI* expression and immune cell infiltration, immunomodulatory genes, TME components, tumor mutational burden (TMB), tumor microsatellite instability (MSI), and molecular pathways in various types of cancer, along with the types of *TGFBI* genetic alterations.

MATERIALS AND METHODS

TCGA Data and Processing

TCGA database (http://cancergenome.nih.gov) is a landmark public cancer genomics program that, as of 2021, had analyzed molecular characteristics of more than 20,000 primary cancer and normal samples covering 33 cancer types. We used The University of California Santa Cruz (UCSC) Xena website (https://xenabrowser.net/) to collect *TGFBI* data from the TCGA database including RNA-Seq data, clinical data, DNA methylation data, and stemness scores (Goldman et al., 2020). Strawberry Perl (http://strawberryperl.com/; version 5.32.0) was used to obtain the *TGFBI* expression data from the TCGA database and construct a data matrix for further analysis.

TGFBI Expression Analysis

A comprehensive website TIMER2 (Tumor Immune Estimation Resource, http://timer.cistrome.org/; version 2) was used to systematic analysis the differential gene expression between different cancer types and normal tissues (Li et al., 2017). The "Gene_DE" module was used to compare *TGFBI* expression between various cancer types with adjacent normal tissues using TCGA data.

A comprehensive website UALCAN (http://ualcan.path.uab.edu/analysis.html) can analyze protein expression using data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) database and the TCGA database (Chandrashekar et al., 2017). The "CPTAC" module of UALCAN website was used to investigate TGFBI protein expression in various tumors and adjacent normal tissues. $p < 0.05$ was considered significant.

An interactive website GEPIA2 (Gene Expression Profiling Interactive Analysis, http://gepia2.cancer-pku.cn/#analysis; version 2) was used to analyze RNA-Seq expression data of cancer tissues and normal samples from the Genotype–Tissue Expression (GTEx) Project and the TCGA database (Tang et al., 2017). The "Expression Analysis-Box Plots" module of GEPIA2 was used to get box plots comparing *TGFBI* expression between cancer and normal tissues. Setting p value cutoff <0.01, $\log_2$ (fold change) cutoff >1, and "Match TCGA normal and GTEx data." In addition, we used the "Pathological Stage Plot" module to analyze *TGFBI* expression in different pathological stages of various tumors using TCGA data.

A publicly accessible platform Oncomine platform (www.oncomine.org) was used to analyze genome-wide expression, which, as of April 2021, contained 715 datasets and 86,733 samples (Rhodes et al., 2007). We used this platform to compare *TGFBI* expression between various cancer types and adjacent normal tissues using Student's t-test. We set p value cutoff <0.05 and fold change cutoff >2, and selected the top 10% genes.

CCLE (Cancer Cell Line Encyclopedia, portals.broadinstitute.org/ccle/) database provides public access to genomic data, visualization and analysis for over 1,100 cancer cell lines.

TGFBI Immunohistochemical Analysis

The HPA database (https://www.proteinatlas.org) was used to map all human proteins at the cell, tissue, and organ levels by integrating various omics technologies. We used this database to explore TGFBI mRNA and protein expression data from various cancer types. We also obtained immunohistochemistry images of TGFBI protein expression in cancer tissues.

TGFBI Expression and Cancer Survival

We used the "Survival Map" module in GEPIA2 to gain the overall survival (OS) and disease-free survival (DFS) data correlated with *TGFBI* expression across different cancer types from the TCGA database. The cases were split into high- and low-expression subgroups based on the median expression. The survival data were assessed basing on the Kaplan–Meier (KM) method, with the results being presented as hazard ratios, 95% confidence intervals, and p values of log-rank tests.

The PrognoScan database (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) and Cox regression analysis (with $p < 0.05$ indicating significance) were used to verify the relationships between *TGFBI* expression and various survival outcomes in a pan-cancer analysis, including OS, DFS, relapse-free survival (RFS), disease-specific survival (DSS), distant metastasis-free survival (DMFS) and distant recurrence-free survival (DRFS) (Mizuno et al., 2009).

The "survival" and "survminer" R package were used to analyze the correlation of the *TGFBI* expression and OS, DSS, disease-free interval (DFI), and progression-free interval (PFI)

across all TCGA tumors. We computed log-rank *p*-values and hazard ratios (HR) with 95% confidence intervals (95% CI). Data were visualized as forest plots (using the "forestplot" R package) and survival curves.

We used KM plotter (http://kmplot.com/analysis/), to assess the relationships between *TGFBI* expression and prognosis in various cancers from the TCGA and GEO databases. Breast, gastric, lung, ovarian, and liver cancer datasets were each split into high- and low-expression subgroups using the "autoselect best cutoff" option, and $p < 0.05$ was considered significant.

TGFBI Mutation Profiles

A comprehensive website cBioPortal (www.cbioportal.org) was used to explore, analyze, and visualize polydimensional cancer genomics data (Cerami et al., 2012). The *TGFBI* alteration frequencies, mutation types, and copy number alterations across the TCGA database were obtained using the "Cancer Types Summary" module of cBioPortal. To query the *TGFBI* genetic alteration characteristics, we set the "Quick select" field to "TCGA Pan Cancer Atlas Studies."

The COSMIC (Catalog of Somatic Mutations in Cancer) website (https://cancer.sanger.ac.uk/cosmic/) is the largest and most comprehensive resource for exploring the impact of somatic mutations in human cancers (Tate et al., 2019). We used COSMIC to investigate the specific distribution of various *TGFBI* mutation types.

Correlation Analyses

The Cancer Regulome Explorer (http://explorer.cancerregulome.org/) is a website that enables researchers to search, filter, and visualize analytical results obtained from the TCGA database. We used this website to investigate and visualize, at the chromosome level, the *TGFBI* mutation types in various cancer types, and the results were depicted in Circos plots. We set the pairwise correlation cutoff ≥0.4 and $-\log_{10}$ (p value) cutoff ≥10.

Pearson's correlation analysis was performed to analyze the correlations between *TGFBI* expression and immunoinhibitory and immunostimulatory gene subsets, TMB, and MSI. The results were displayed as heatmaps basing on the "pheatmap" package in R.

TGFBI Expression and Immune Cell Infiltration

We used the "Immune_Gene" module of TIMER2 to evaluate the correlations of *TGFBI* expression with immune cell infiltration across diverse cancer types in the TCGA database. The immune cells comprised $CD4^+$ T cells, $CD8^+$ T cells, regulatory T cells (Tregs), B cells, myeloid-derived suppressor cells (MDSCs), neutrophils, dendritic cells (DCs), monocytes, macrophages, mast cells, cancer-associated fibroblasts (CAFs), natural killer (NK) cells, endothelial cells, and T follicular helper (Tfh) cells.

TGFBI Expression and TME Components

Tumor purity was assessed in 33 human cancers in the TCGA database basing on the "estimate" R package. Specifically, the ESTIMATE score is the sum of the immune and stromal scores, which represent the abundance of immune and stromal components, respectively (Yoshihara et al., 2013). Higher ESTIMATE scores correspond to lower tumor purity. Spearman's correlation analysis was used to reveal the relationship between *TGFBI* expression and the immune, and stromal scores.

We also explored tumor RNA and DNA stemness scores (RNAss and DNAss) based on epigenetic and transcriptome data from the TCGA database. Specifically, RNAss is based on RNA-Seq data, and DNAss is based on DNA methylation data. Spearman's correlation analysis was used to investigate the correlations of TGFBI expression with RNAss and DNAss.

TGFBI Expression and Drug Responses

The CellMiner tool (https://discover.nci.nih.gov/cellminer/home.do) was employed to obtain National Cancer Institute (NCI)-60 data on *TGFBI* expression (i.e., transcript data) and drug responses (i.e., drug sensitivity based on data on the drug concentration that reduces total cell growth by 50% [GI50]). The NCI-60 data comprises molecular and pharmacological data on 60 diverse human cancer cell lines. The drug response data includes data on drugs approved by the US Food and Drug Administration and those assessed in clinical trials. Pearson's correlation analysis was employed to evaluate the correlations between *TGFBI* expression and drug responses.

Protein–Protein Interactions Network and Enrichment Analyses of TGFBI

The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (https://string-db.org/) contains known and predicted protein-protein interactions (PPIs) (Szklarczyk et al., 2019). We obtained the top 50 TGFBI-binding proteins using this database to construct a PPIs network. The parameters were set as follows: meaning of network edges: "evidence"; minimum required interaction score: "low confidence" (i.e., the line color represents the type of interaction evidence); and the maximum number of interactors to show: "no more than 50 interactors". The PPIs network was visualized using STRING. Then the "Similar Gene Detection" module of GEPIA2 was used to obtain the top 100 TGFBI-correlated genes using TCGA data. Lastly, we conducted a Venn diagram analysis (http://jvenn.toulouse.inra.fr) to identify the common members of these two groups.

DAVID (Database for Annotation, Visualization, and Integrated Discovery, https://david.ncifcrf.gov/home.jsp; version 6.8) supplies a comprehensive, functional annotation tool to identify genes' biological functions (Huang et al., 2009). It was used to conduct Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses. The KEGG results were analyzed and visualized employing the "ggplot2" and "clusterProfiler" R packages. The GO biological process (BP), cellular component (CC), and molecular function (MF) results were obtained using the "cnetplots" R package.

Statistical Analyses

TGFBI expression was compared between tumor and adjacent normal tissues employing the Oncomine database, with the results presented as p values, fold changes, and gene ranks. The survival results were presented as hazard ratios, 95% CI,

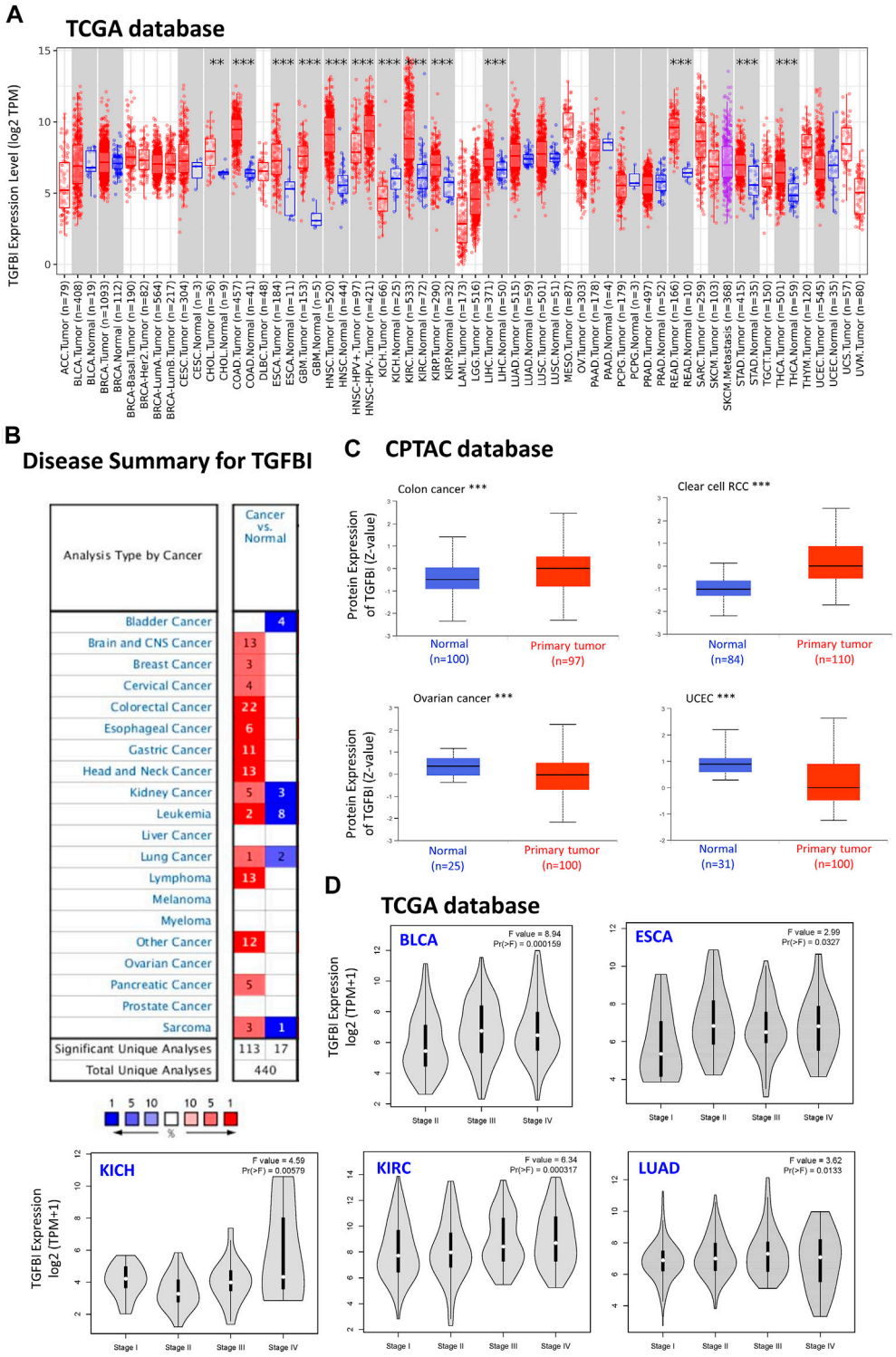


FIGURE 1 | TGFBI expression in various types of human cancers and pathological stages. **(A)** Comparison of *TGFBI* expression between various cancers/cancer subtypes and normal tissues using TIMER2 based on TCGA data. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. **(B)** Comparison of *TGFBI* expression between various cancer tissues and normal tissues in the Oncomine database. Blue represents low expression of TGFBI in the cancer tissues, red represents high expression, and gray represents no data. The number represents the number of studies that meet the filter criteria. **(C)** Comparison of TGFBI protein expression between primary colon cancer, clear cell renal cell carcinoma, uterine corpus endometrial carcinoma, and ovarian cancer tissues and normal tissues, based on CPTAC data. ***$p < 0.001$. **(D)** *TGFBI* expression in pathological stages I, II, III, and IV in BLCA, CESC, KICH, KIRC, and LUAD, based on TCGA data. The data were transformed using log2 (transcripts per million [TPM] + 1). F represents the statistical value of F test, Pr(>F) < 0.05 was considered significant.

and p values of log-rank tests. Using R version 4.0.4 (64-bit; https://www.r-project.org/) for the analyses. For all statistical analyses, $p < 0.05$ was considered statistically significant.

RESULTS

TGFBI Expression Analysis in Pan-Cancer

To compare *TGFBI* expression between tumor and normal tissues, we applied TIMER2 to analyze *TGFBI* expression in various cancer types of TCGA. Compared to matched normal tissues, *TGFBI* was upregulated in cholangiocarcinoma (CHOL; $p < 0.01$), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), GBM, head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), and thyroid carcinoma (THCA; $p < 0.001$). However, *TGFBI* was downregulated in kidney chromophobe (KICH; $p < 0.001$; **Figure 1A**).

Furthermore, we applied the Oncomine database to compare *TGFBI* expression between tumor and matched normal tissues. *TGFBI* was upregulated in colorectal cancer, esophageal cancer, gastric cancer, head and neck cancer, and lymphoma (**Figure 1B**). In addition, we explored *TGFBI* expression across different tumor cell lines in the CCLE database. As shown in **Supplementary Figure S1**, RNAseq showed that *TGFBI* was upregulated in several cell lines, including kidney cancer, chondrosarcoma, upper aerodigestive tract cancer, glioma, osteosarcoma, thyroid cancer, liver cancer and mesothelioma cell lines.

We further evaluated the difference in TGFBI protein expression between the tumor tissues and normal tissues in the CPTAC database using UALCAN tools. TGFBI protein was upregulated in colon cancer and clear cell renal cell carcinoma compared to normal tissues ($p < 0.001$) but downregulated in OV and uterine corpus endometrial carcinoma (UCEC; $p < 0.001$; **Figure 1C**).

We also applied GEPIA2 to analyze *TGFBI* expression in various pathological stages of multiple cancer types. *TGFBI* expression was associated with bladder urothelial carcinoma (BLCA), ESCA, KICH, KIRC and lung adenocarcinoma (LUAD; all Pr(>F) < 0.05; **Figure 1D**). We also compared *TGFBI* expression between these TCGA tumor tissues and the corresponding normal tissues in the GTEx database, which showed that *TGFBI* was upregulated in 19 types of cancer tissues and downregulated in three types compared to normal tissues (**Supplementary Figure S2**).

Using HPA data, we analyzed TGFBI protein expression. Analysis revealed that aberrant TGFBI protein expression was detected in 20 cancer types. The immunohistochemical results on TGFBI protein expression are shown in **Figure 2**.

Correlation Between the Expression of *TGFBI* and Prognosis

We assessed the associations between *TGFBI* expression and cancer survival outcomes in the TCGA database using GEPIA2. High *TGFBI* expression was associated with poor OS for cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC, $p = 0.0024$), GBM ($p = 0.0082$), HNSC ($p = 0.013$), KIRC ($p = 0.0019$), pancreatic adenocarcinoma (PAAD; $p = 0.047$), and uveal melanoma (UVM; $p < 0.001$; **Figure 3A**). In addition, high *TGFBI* expression was associated with poor DFS for CHOL ($p = 0.03$), COAD ($p = 0.0095$), KIRC ($p = 0.0011$), PAAD ($p = 0.033$), and UVM ($p < 0.0029$; **Figure 3B**).

We also applied the PrognoScan database to analyze the associations between *TGFBI* expression and cancer survival outcomes, mainly based on GEO datasets (GSE13507, GSE4475, GSE4271, GSE7849, GSE9893, GSE2034, GSE7390, GSE11595, GSE22138, GSE31210, GSE8894, and GSE30929). High *TGFBI* expression was associated in Cox regression analyses with poor DSS in bladder cancer ($p = 0.031$), OS in brain cancer ($p = 0.047$), DFS, DMFS, and RFS in breast cancer (BRCA; all $p < 0.05$), DMFS in eye cancer ($p = 0.010$), OS and RFS in lung cancer ($p < 0.001$), and DRFS in soft tissue cancer ($p < 0.001$). In contrast, high *TGFBI* expression was associated with better OS in blood cancer ($p = 0.016$), breast cancer ($p = 0.045$), and esophageal cancer ($p = 0.047$; **Figure 3C**).

We assessed the correlation between the expression of *TGFBI* and OS, DSS, PFI, and DFI in different types of cancer using a single-variate Cox regression analysis based on TCGA database. The results are summarized in forest plot (**Figure 4**). KM analysis showed that high *TGFBI* expression predicted poor prognosis of CESC, GBM, KIRC, LGG, testicular germ cell tumors (TGCT), and UVM (all $p < 0.05$) but good prognosis of adrenocortical carcinoma (ACC) and HNSC (all $p < 0.05$; **Figure 5A**). In addition, increased expression of *TGFBI* was associated with poor DSS in CESC, GBM, KIRC, LGG, PAAD, and UVM, whereas, increased *TGFBI* expression predicted good DSS in ACC (all $p < 0.05$, **Figure 5B**). The same method is used to analyze PFI in 33 TCGA tumors. *TGFBI* had a protective role in ACC. On the other hand, *TGFBI* had a detrimental role in BRCA, CESC, KIRC, PAAD, and UVM (all $p < 0.05$, **Figure 5C**). Finally, we also analyzed the DFI in 33 TCGA tumors. *TGFBI* expression had a detrimental role in BRCA, CHOL, and COAD (all $p < 0.05$, **Figure 5D**).

Then we studied the prognostic value of *TGFBI* expression was further evaluated using KM plotter. Notably, high *TGFBI* expression was significantly associated with poor DMFS, OS, post-progression survival (PPS), and RFS in breast cancer ($p < 0.01$; **Supplementary Figure S3A**). However, high *TGFBI* expression was associated with good DSS, OS, progression-free survival (PFS), and RFS in liver cancer ($p < 0.05$; **Supplementary Figure S3B**). High *TGFBI* expression was associated with poor first progression (FP), OS, and PPS in gastric cancer ($p < 0.001$; **Supplementary Figure S3C**). High *TGFBI* expression was associated with poor FP and OS but good PPS in lung cancer ($p < 0.001$; **Supplementary Figure S3D**). Moreover, high *TGFBI* expression was associated with poor OS, PFS, and PPS in OV ($p < 0.05$; **Supplementary Figure S3E**).

The Mutation Profiles of *TGFBI* in Pan-Cancer

Using TCGA data in cBioPortal (10,967 samples from 32 studies), we assessed the *TGFBI* alteration frequency and

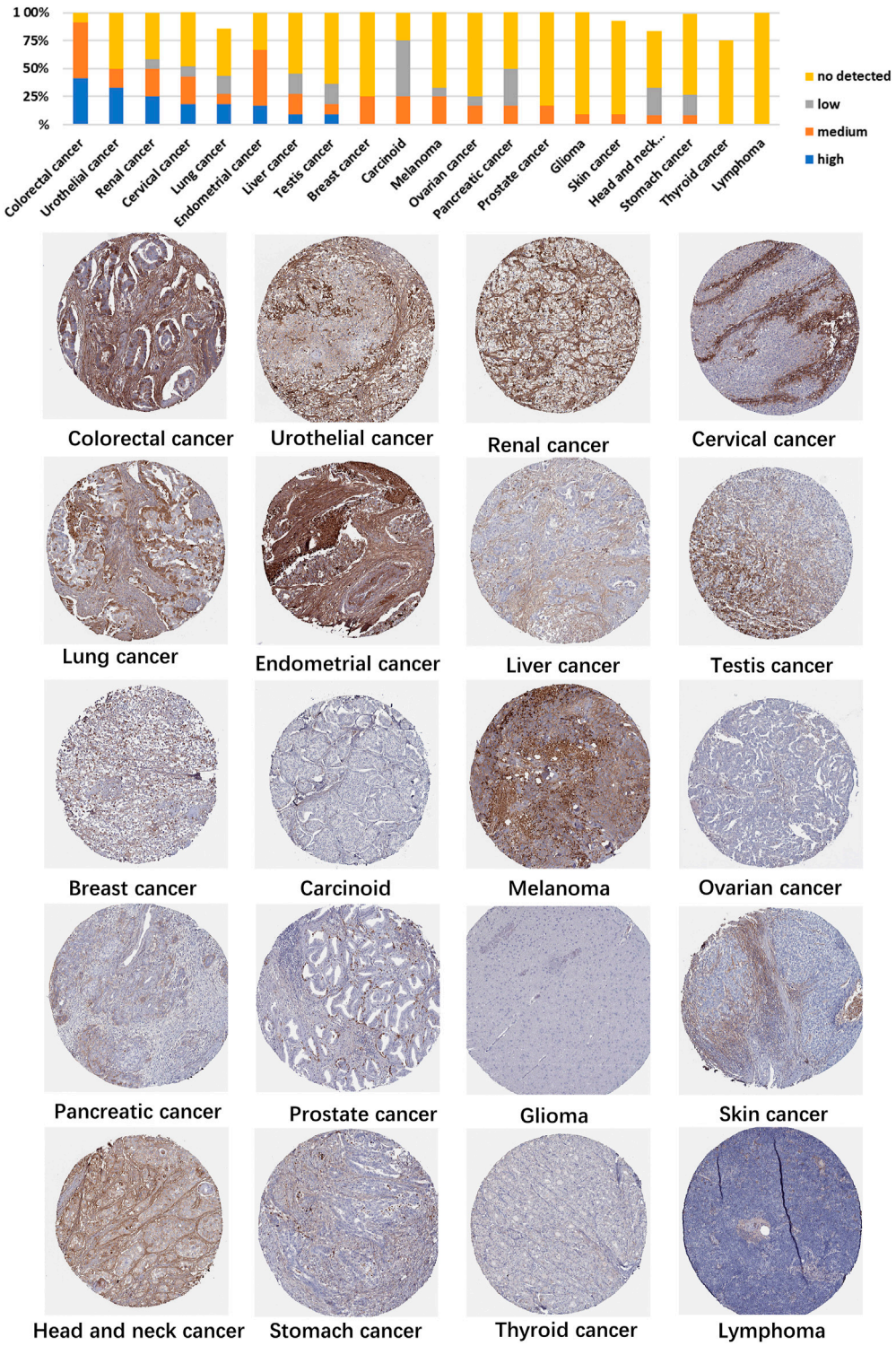


FIGURE 2 | Representative immunohistochemical staining of TGFBI protein in various cancer tissues.

mutation count in tumor samples. Melanoma had the highest frequent *TGFBI* alteration (>6%), with "mutation" as the primary type of alteration (**Figures 6A,B**). Data on the *TGFBI* alteration types, sites, and numbers of cases, including data on missense, nonsense (truncation), inframe, and fusion mutations, are shown in **Figure 6C**. The *TGFBI* mutation hotspot was R469C/H/S (missense mutation) in the FAS1 domain, which occurred in four cancers (UCEC, CESC,

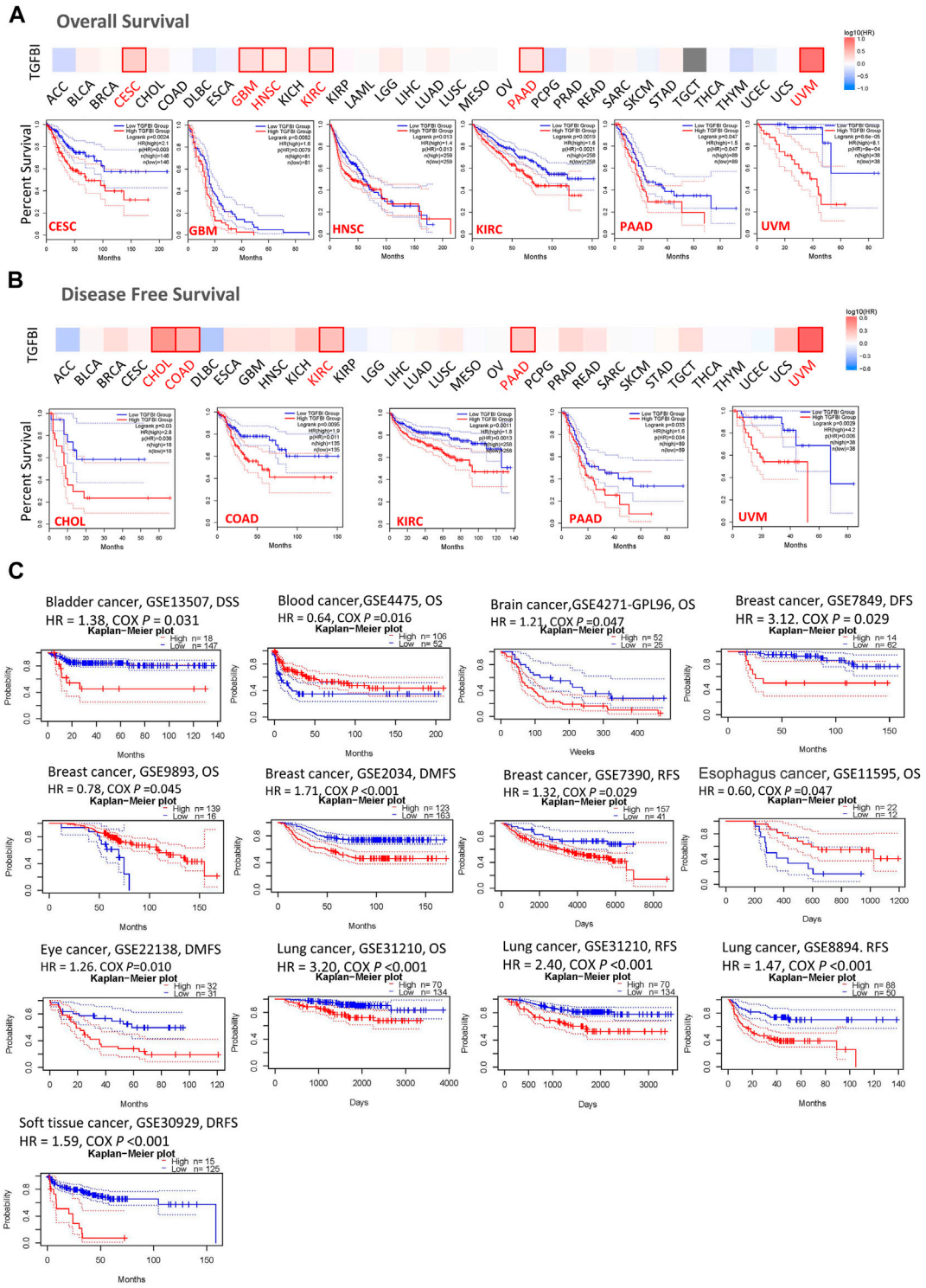


FIGURE 3 | Associations of *TGFBI* expression with survival among cancer patients in the TCGA and GEO databases. *TGFBI* expression and **(A)** overall survival rate and **(B)** disease-free survival (DFS) rate in various tumors in TCGA were analyzed using GEPIA2. The Kaplan–Meier survival curves with significant results are shown. **(C)** Kaplan–Meier survival curves of TGFBI expression in 12 GEO cohorts (GSE13507, GSE4475, GSE4271, GSE7849, GSE9893, GSE 2034, GSE7390, GSE11595, GSE22138, GSE31210, GSE8894, and GSE30929), with significant differences shown. HR represents hazard ratios.

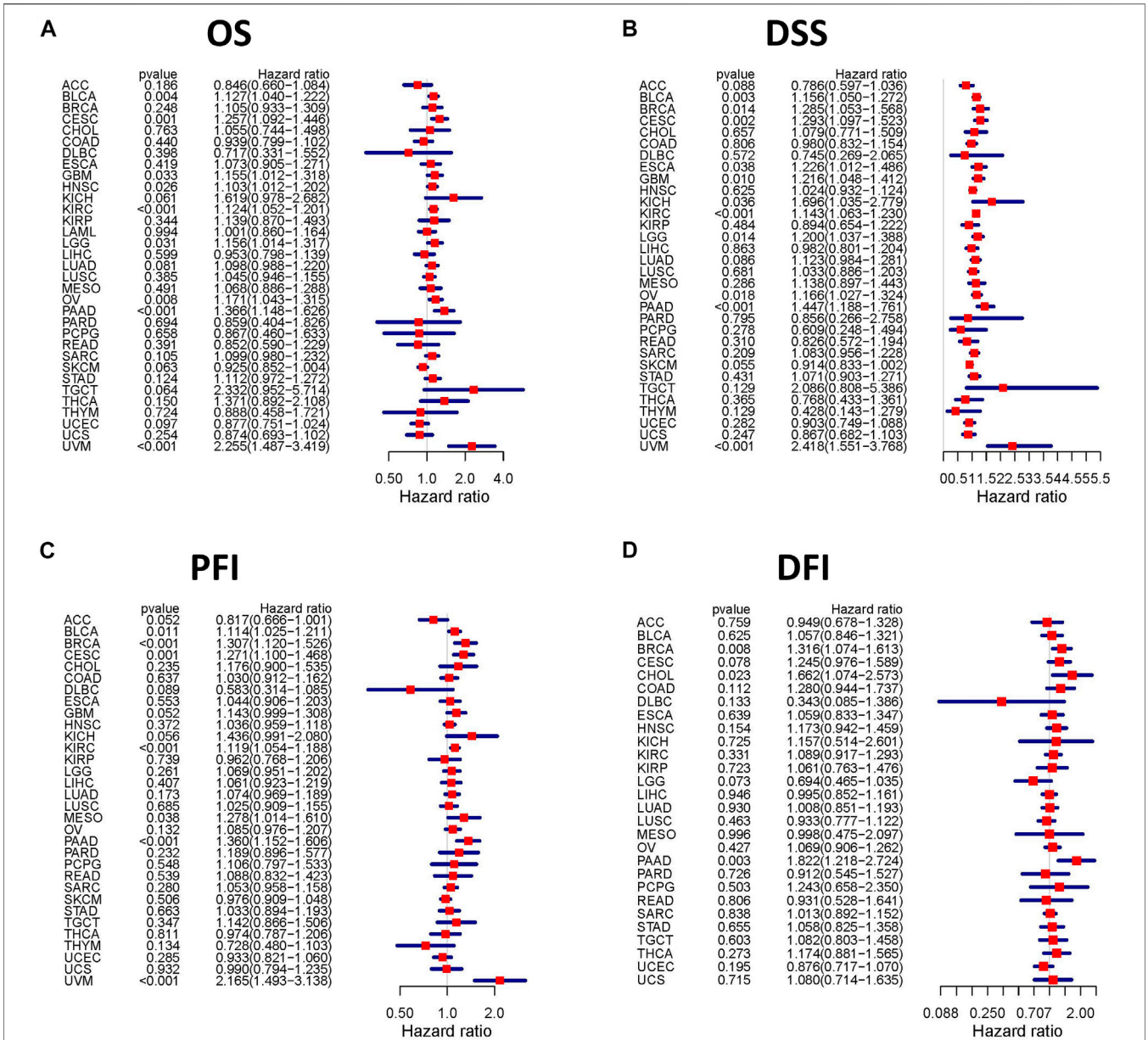


FIGURE 4 | Correlation analysis of *TGFBI* expression and patient survival by the Cox regression analysis in different cancer types. HR < 1 represents low risk and HR > 1 represents high risk. Univariate Cox proportional hazard regression models were applied for the association tests.

COAD, and skin cutaneous melanoma [SKCM]) in four patients.

Moreover, we explored the impact of somatic *TGFBI* mutations in human cancer. Using the COSMIC website, nonsense (truncation) mutations were found in lung cancer (10%); missense mutations were found in endometrial (7.14%), hematopoietic and lymphoid (40%), large intestine (15%), lung (20%), and skin (61.54%) cancer; and synonymous substitutions were found in breast (28.57%), endometrial (28.57%), and large intestine (10%) cancer. G > A was the primary mutation type in these cancer cases (**Supplementary Figure S4**).

Genomic Analysis of *TGFBI* in Cancer

Using the Cancer Regulome Explorer website, we investigated the relevant genomic correlation between *TGFBI* gene and certain signatures. DNA methylation, somatic copy number, somatic mutation, microRNA expression, and protein expression were showed to reveal the interrelation in different tumors using data from the TCGA database. The *TGFBI* was associated with above signatures in ACC, BLCA, BRCA, ESCA, STAD, GBM, HNSC, KIRC, LUAD, OV, SKCM, THCA, UCEC, lung squamous cell carcinoma (LUSC), and prostate adenocarcinoma (PRAD) (**Supplementary Figure S5**).

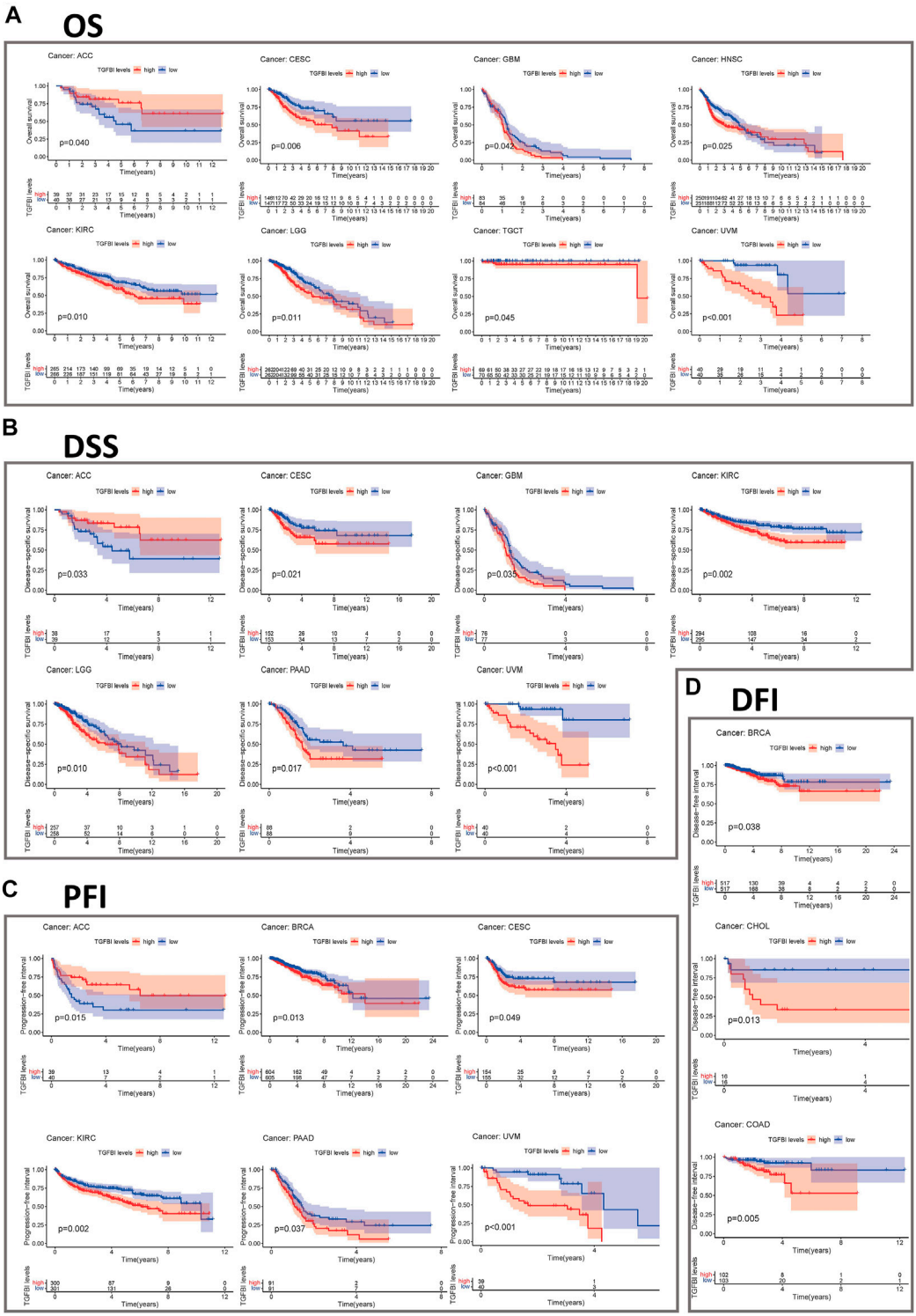


FIGURE 5 | Kaplan–Meier survival curves comparing pan-cancer high and low expression of *TGFBI*. **(A)** OS survival curves for *TGFBI* in different cancers; **(B)** DSS survival curves for *TGFBI* in different cancers; **(C)** PFI survival curves for *TGFBI* in different cancers; **(D)** DFI survival curves for *TGFBI* in different cancers.

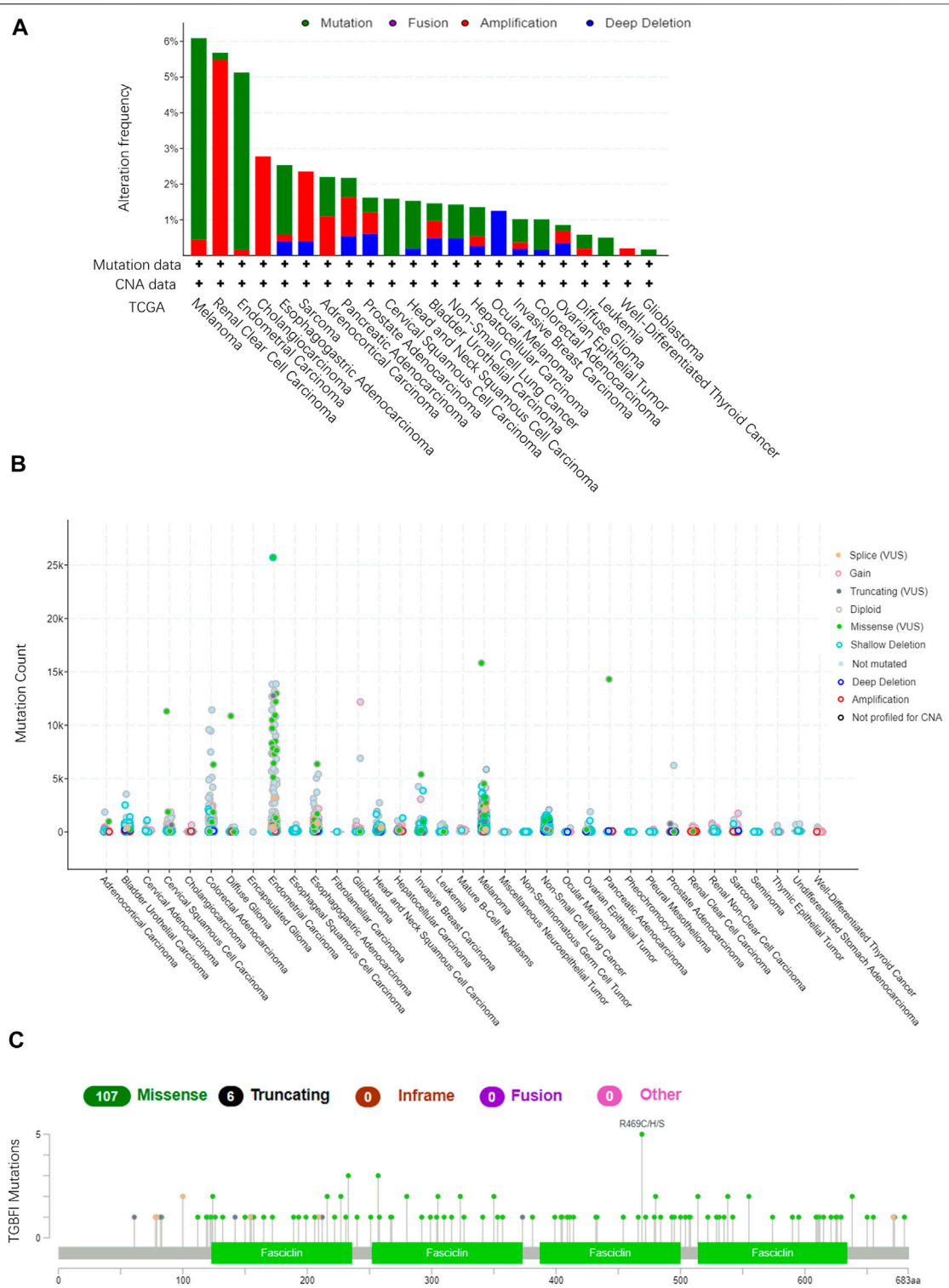


FIGURE 6 | *TGFBI* mutation landscape in various cancer types. **(A)** *TGFBI* mutation frequency in various TCGA pan-cancer studies according to cBioPortal. **(B)** *TGFBI* mutation count in various TCGA cancer types according to cBioPortal. **(C)** TGFBI mutations across protein domains in various cancer types.

TGFBI Expression and Immune Cell Infiltration

TGFBI is involved in immune cell infiltration and inflammatory responses, which play key roles in cancer initiation, progression, and metastasis. To investigate immune cell infiltration at the pan-cancer level, we used TIMER2.0 to explore the associations between *TGFBI* expression in human cancer and the infiltration of various types of immune cells (based on CIBERSORT, CIBERSORT-ABS, XCELL, MCPCOUNTER, QUANTISEQ, and EPIC algorithms). Overall, *TGFBI* expression was positively correlated with neutrophils, monocytes, macrophages, CAFs and MDSCs, and negatively correlated with B cells, T follicular helper (Tfh) cells, and $CD8^+$ T cells. Thus, *TGFBI* expression may play vital roles in the immune cell infiltration process. The immune cell infiltration was quite different in OV, PRAD, UVM, and THCA due to the distinct TME components in the central nervous system (**Supplementary Figure S6**).

Correlations of *TGFBI* Expression with Immunomodulatory Genes, TMB, and MSI

Tumor immunotherapy is a novel therapeutic strategy which been proven efficacious in multiple types of cancers. Thus, we explore whether TGFBI could be used as a novel target for tumor immunotherapy. As shown in **Figure 7A**, in most cancers, except CHOL, ESCA, sarcoma (SARC), SKCM and UCS, *TGFBI* expression was significantly correlated with immunoinhibitory genes. Additionally, except CESC, CHOL, mesothelioma (MESO), READ, SARC, UCEC and UCS, *TGFBI* expression was significantly correlated with immunostimulatory genes (**Figure 7B**).

We also explored the associations of *TGFBI* expression with TMB and MSI in various types of cancer. The *TGFBI* expression was negatively correlated with TMB in COAD, HNSC, LUSC and BLCA (all $p < 0.05$; **Figure 7C**). In addition, *TGFBI* expression was negatively correlated with MSI in LUSC, LUAD, COAD, and CHOL (all $p < 0.05$; **Figure 7D**).

TGFBI Expression and TME Components

Since our results have confirmed the immunoregulation role of the TGFBI in various types of cancer, it is vary needed to explore further the correlation between TGFBI expression and TME. We explored the correlations of *TGFBI* expression with TME components, using the ESTIMATE algorithm to calculate the ESTIMATE, stromal, and immune scores in diverse tumor types in the TCGA database. *TGFBI* expression was positively associated with stromal and immune scores in the pan-cancer analysis (all $p < 0.05$; **Figures 8A,B**). Furthermore, we assessed the correlations between *TGFBI* expression and tumor stemness in a pan-cancer analysis using TCGA data. *TGFBI* expression was negatively correlated with RNAss, and positively correlated with DNAss, in ACC, SARC, THCA, thymoma (THYM) and UVM (**Figure 8A**).

The Drug Responses Analysis of *TGFBI* Expression

To investigate the potential relationships between *TGFBI* expression and drug responses in various types of human cancer, we performed a correlation analysis to identify potential drug candidates using CellMiner. *TGFBI* expression was positively correlated with staurosporine, dasatinib, midostaurin, and bleomycin sensitivity (all $p < 0.05$; **Figures 8C,D,G,H**), but negatively associated with lapachone, CUDC-305 byproduct, and fulvestrant sensitivity (all $p < 0.05$; **Figures 8E,F,I**).

Protein-Protein Interactions Network and Enrichment Analyses of TGFBI

To clarify the molecular mechanisms of TGFBI in tumorigenesis, we conducted a PPIs network analysis of TGFBI and enrichment analyses of TGFBI-binding proteins (based on STRING) that were also TGFBI-correlated genes (based on GEPIA2). The PPIs network, constructed using STRING, was in the base of experimental evidence and had 22 nodes and 34 edges (**Figure 9A**). We obtained 50 TGFBI-binding proteins supported by experimental evidence in STRING. Using GEPIA2, we then obtained the top 100 genes associated with *TGFBI* expression. Next, a Venn diagram analysis of these two groups showed one common member, namely, *COL4A1* (**Figure 9B**). The heatmap in **Figure 9C** shows the positive correlation between *TGFBI* and *COL4A1* expression.

To conduct KEGG and GO functional enrichment analyses of TGFBI, we used DAVID 6.8. The KEGG analysis showed that the "PI3K-Akt signaling pathway", "Focal adhesion", and "ECM–receptor interaction" might be involved in the effects of TGFBI on tumor pathogenesis (**Figure 9D**). The GO analysis showed similar results, as TGFBI was closely involved in a variety of terms, such as "extracellular matrix organization" in the BP category, "collagen-containing extracellular matrix" in the CC category, and "extracellular matrix structural constituent" in the MF category (**Figure 9E**).

DISCUSSION

In recent years, a number of studies have shown that TGFBI is closely related to tumor growth, invasion, metastasis, and drug responses. However, there are no previous pan-cancer studies on the relationship between *TGFBI* expression and various tumors. In this study, we investigated the *TGFBI* expression, survival status, *TGFBI* mutations, immune cell infiltration, and associated molecular pathways, to explore the potential cellular mechanisms of TGFBI in various types of cancers.

Based on the Oncomine and TIMER2.0 databases, *TGFBI* was upregulated in CHOL, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LIHC, READ, STAD, and THCA, and downregulated in KICH, compared to adjacent normal tissues. A previous study also found that *TGFBI* was upregulated in colon cancer, promoting metastasis (Ma et al., 2008). Another study showed that the proliferation and metastasis ability of oral squamous cell carcinoma cells was also enhanced by *TGFBI* upregulation (Wang et al., 2019).

We also identified associations between *TGFBI* expression and prognosis of lung squamous cell, gastric, colorectal, and bladder

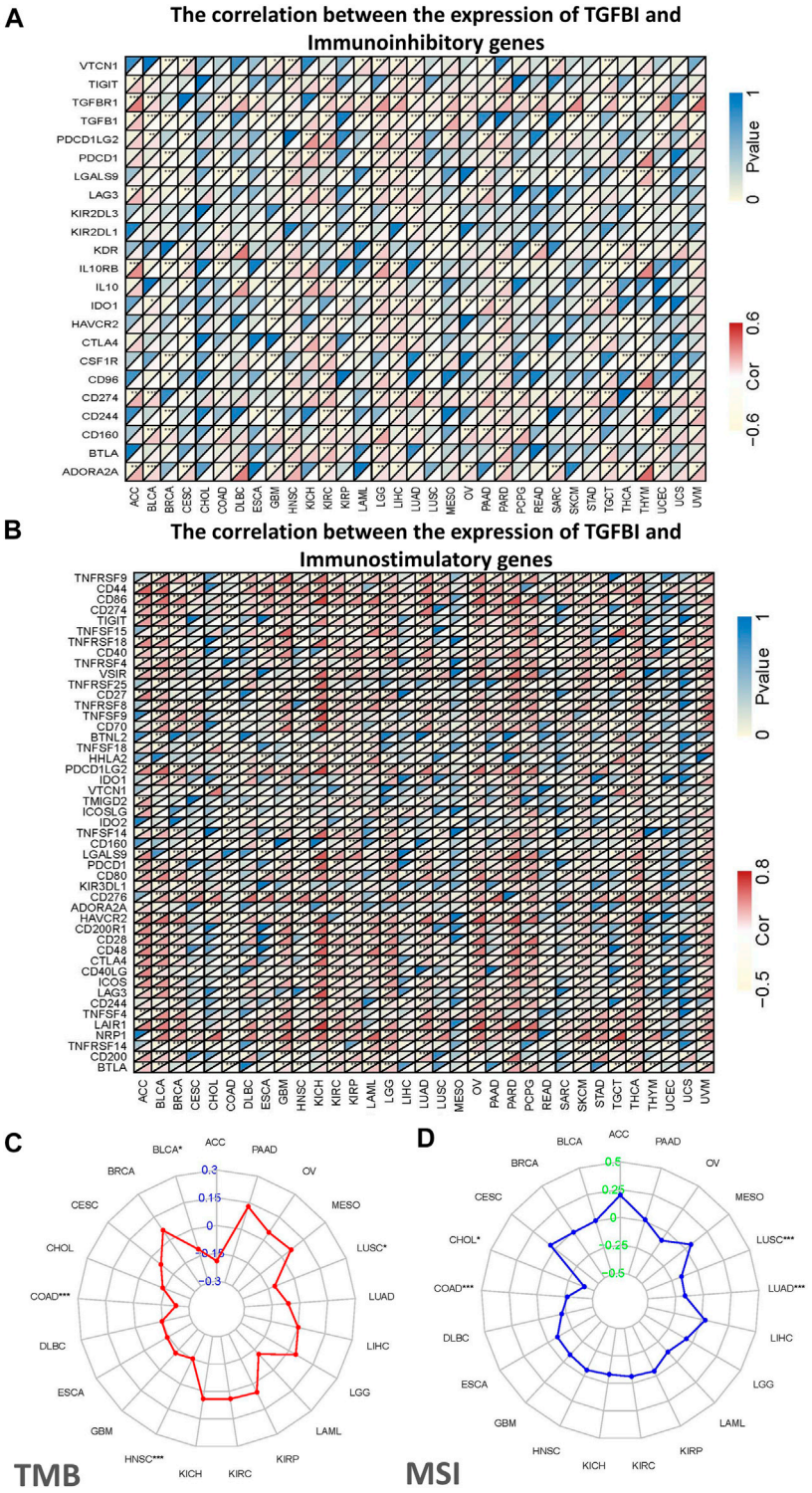


FIGURE 7 | Correlations of *TGFBI* expression with immunomodulatory genes, TMB, and MSI in various cancer types. Correlations of *TGFBI* expression with **(A)** immunoinhibitory and **(B)** immunostimulatory genes. For each pair, the upper left triangle is colored to represent the *p* value, and the lower right one is colored to indicate Spearman's correlation coefficient. **(C)** Radar chart of the overlap between TMB and *TGFBI* expression. The blue number represents Spearman's correlation coefficient. **(D)** Radar chart of the overlap between MSI and *TGFBI* expression. The green number represents Spearman's correlation coefficient. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

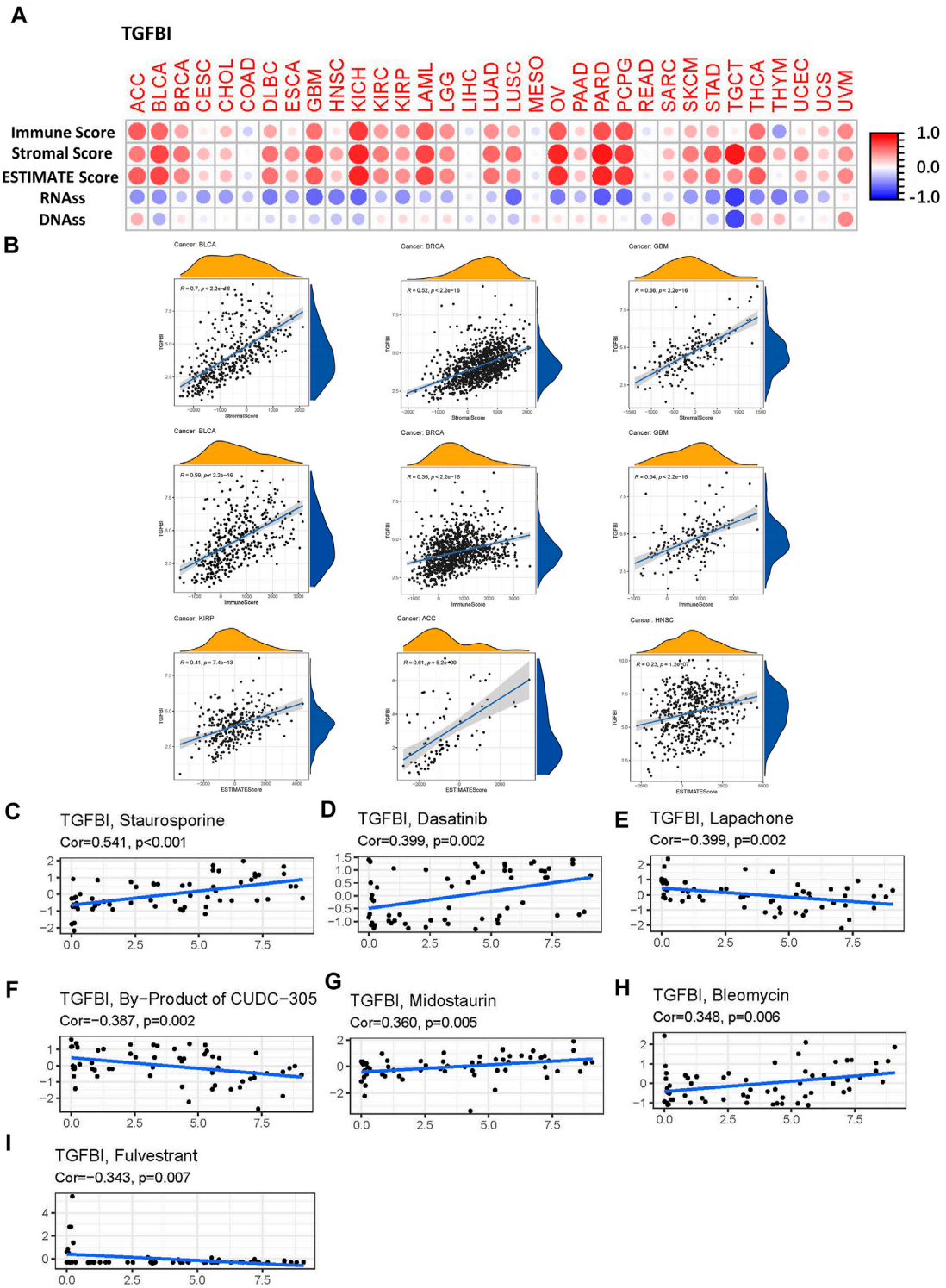


FIGURE 8 | Correlation matrix of *TGFBI* expression with tumor microenvironment (TME) components, stemness scores, and drug sensitivity in a pan-cancer analysis. **(A)** *TGFBI* expression was correlated with ESTIMATE score, immune score, stromal score, RNAss, and DNAss in various cancers. Red and blue dots indicate a positive and negative correlation, respectively, between *TGFBI* expression in cancer and immune/stromal score. **(B)** Top three scatter plots of correlation between *TGFBI* and stromal score, immune score, ESTIMATE score in various cancers. *TGFBI* expression was positively associated with **(C)** staurosporine, **(D)** dasatinib, **(G)** midostaurin, and **(H)** bleomycin sensitivity, and negatively associated with **(E)** lapachone, **(F)** CUDC-305 byproduct, and **(I)** fulvestrant sensitivity. R represents the correlation coefficient, Cor represents correlation.

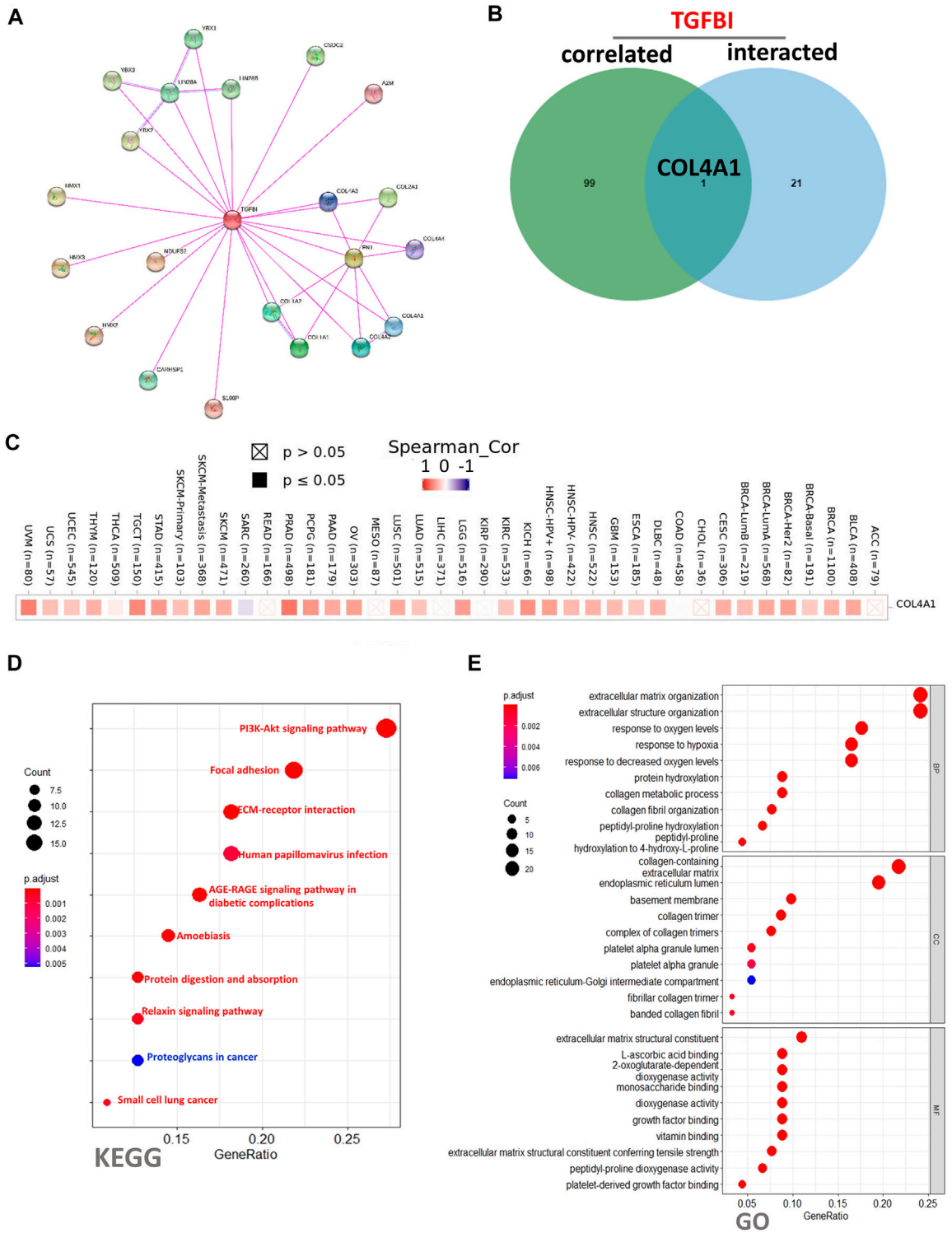


FIGURE 9 | TGFBI-related gene enrichment analysis. **(A)** Experimentally determined TGFBI-binding proteins, based on STRING. **(B)** Venn diagram analysis of TGFBI-binding proteins (based on STRING) and TGFBI-correlated genes (based on GEPIA2). **(C)** Corresponding heatmap in various cancer types. **(D)** KEGG pathway enrichment analysis and **(E)** Gene Ontology (GO) biological process, cellular component, and molecular function enrichment analysis of TGFBI-binding proteins (based on STRING) and TGFBI-correlated genes (based on GEPIA2).

cancer, which concurs with previous studies (Pajares et al., 2014; Suzuki et al., 2018; Zhang et al., 2019; Zou et al., 2019; Yang et al., 2020). Overall, *TGFBI* was upregulated in most tumors, though the *TGFBI* expression and survival analysis suggested different conclusions for different tumors. For example, *TGFBI* upregulation had a detrimental effect on UVM, which was consistent with a previous study showing that human metastatic melanoma express and released a significantly higher amount of TGFBI (Lauden et al., 2014), rendering TGFBI a potential target for therapeutic interventions. We also found some contradictory results about the prognostic value of *TGFBI* expression in LUAD and STAD. High *TGFBI* expression was significantly associated with poor OS in LUAD and STAD in the KM plotter analysis but not in the GEPIA2 analysis. These inconsistent results may be attributable to the different data collection methods, types of cancer patients, or biological characteristics of each sample. These finding suggest that TGFBI might be a novel prognostic biomarker for some types of cancer.

Another essential finding in this study was that *TGFBI* expression was correlated with different levels of immune cell infiltration among diverse types of tumors (Du et al., 2020; Nakazawa et al., 2020). *TGFBI* expression was positively correlated with CAFs, macrophages, monocytes, MDSCs, and neutrophils, and negatively correlated with B cells, Tfh cells, and $CD8^+$ T cells. In this study, *TGFBI* expression was an indicator of the level of infiltration of CAFs, which represent the main component of the tumor stroma and are strongly associated with epithelial–mesenchymal transition, massive stromal cell infiltration, and poor cancer prognosis. CAF infiltration can be used to predict survival in patients with oral squamous cell carcinoma (Ko et al., 2020). Tumor-associated macrophages promote OV cell migration, adhesion, and invasion by secreting TGFBI, and they have been associated with short PFS in high-grade serous OV patients (Steitz et al., 2020). Although we found that *TGFBI* expression was negatively correlated with $CD8^+$ T cells, another study showed that both high stromal *TGFBI* expression and intratumoral $CD8^+$ T cells infiltration were associated with poor prognosis and drug resistance in lung cancer patients (Nakazawa et al., 2020). These contradictory findings necessitate further research.

In this study, we presented evidence regarding the correlations between *TGFBI* expression and MSI and TMB across the TCGA pan-cancer. MSI is a molecular marker of deficient mismatch repair (MMR), which leads to errors in DNA replication, the accumulation of DNA mutations, and a high TMB in many cancer types (Baretti and Le, 2018). TMB is an emerging marker for identifying potential responders to immunoinhibitory and immunostimulatory factors across cancer types (Goodman et al., 2017). The *TGFBI* expression was negatively correlated with TMB in COAD, HNSC, BLCA, and LUSC. In addition, *TGFBI* expression was negatively correlated with MSI in LUSC, LUAD, COAD, and CHOL. Our results suggest that the relationships of *TGFBI* expression with TMB and MSI are diverse in those types of cancer. *TGFBI* might influence tumorigenesis in these cancer types by participating in the process of genetic alterations. expression might also independently predict responses to immunoinhibitory and immunostimulatory factors.

In addition, we explored the relationships of *TGFBI* expression with TME components and tumor stemness. The TME plays an important role in tumorigenesis and metastasis (Binnewies et al., 2018; Yan et al., 2019; Baghban et al., 2020). Based on the ESTIMATE algorithm and TCGA data, *TGFBI* expression was correlated with the levels of immune and stromal cell infiltration. Moreover, we analyzed the correlations between *TGFBI* and tumor stemness scores (RNAss, and DNAss), which are associated with tumor pathology and tumor dedifferentiation (Malta et al., 2018). The tumor stemness analysis of the pan-cancer TCGA data showed that RNAss and DNAss was negatively and positively correlated with prognosis. Similarly, a previous study indicated that TGFBI may support cancer stem cell growth and tumor progression to metastasis in breast cancer (Fico and Santamaria-Martínez, 2020). Besides, we analyze the correlation between *TGFBI* expression with various immune cell infiltration in different types of human cancers. Overall, *TGFBI* expression showed a significant was positively correlated correlation with immune cell infiltrating levels of multiple infiltrates including $CD8^+$ T cells, dendritic cells, macrophages, monocytes, NK cells, neutrophils, Tregs, and Tfh. Otherwise, MDSCs abundance was negatively correlated with *TGFBI* expression. The profile indicated that *TGFBI* play an important role in the recruitment and regulation of immune infiltrating cells in tumors. It is worth noting that in tumors such as ACC, BLCA, BRCA, HNSC, DLBC, GBM, LGG, THCA, THYM, and UVM (**Supplementary Figure S7**), the correlation between *TGFBI* expression and immune cell infiltration was subtly different, which may be caused by the various immune cell infiltration ratios in different types of cancers. Our results emphasize that *TGFBI* expression is closely related to tumor cells, immune cell infiltration, and TME components, affecting cancer prognosis. These results provide new insights for developing more effective treatment.

Additionally, we assessed the correlations between *TGFBI* expression and drug response in various human cancer cell lines from CellMiner database, which facilitates the integration and study of molecular and pharmacological data on the 60 different human cancer cell lines. *TGFBI* expression was positively correlated with staurosporine, dasatinib, midostaurin, and bleomycin sensitivity, but negatively associated with lapachone, CUDC-305 byproduct, and fulvestrant sensitivity. A previous study reported that *TGFBI* is frequently methylated, the loss of *TGFBI* is associated with paclitaxel resistance in OV (Wang et al., 2012), but the overexpression of *TGFBI* makes nasopharyngeal carcinoma cells sensitive to cisplatin (Bissey et al., 2018), and increase the sensitivity of human non-small cell lung cancer cell lines to etoposide, paclitaxel, cisplatin and gemcitabine (Irigoyen et al., 2010). These studies suggest that *TGFBI* might be used as a predictive factor of chemotherapy in some tumors (Yin et al., 2020).

Furthermore, we predicted the PPIs network and KEGG pathways associated with TGFBI. TGFBI was involved in "PI3K/Akt signaling pathway," "ECM–receptor interaction"

and "Focal adhesion," which is consistent with the current research on TGFBI. Previous studies have reported that TGFBI play as irreplaceable role in inducing suppressive mesothelioma tumorigenesis and progression through the PI3K/Akt signaling pathway (Wen et al., 2011b). The ECM has essential roles in tumorigenesis and cancer progression, and different ECM components depend on different signaling mechanisms in various cancer cells. In addition, TGFBI preferentially interacts via a ß3 integrin-receptor-mediated mechanism to modulate paclitaxel-induced OV cell death (Tumbarello et al., 2012). In addition, TGFBI activates the focal adhesion kinase signaling pathway by binding to integrin $\alpha V\beta 5$, significantly enhancing the invasion of pancreatic ductal adenocarcinoma (Costanza et al., 2019). Our results indicated that TGFBI is an oncogenic ECM formation protein that may serve as a valuable therapeutic target for new anti-cancer treatment strategies.

There are several limitations in this study. First, although the study involved a bioinformatic analysis of *TGFBI*, including its expression, prognostic value, associations with immune cell infiltration, and mutation status in various human cancer types, there were no *in vivo* or *in vitro* experiments to validate the results. Therefore, future studies should focus on the mechanisms of TGFBI in various human cancer types. Second, we analyzed prognostic data from TCGA, KM plotter, and PrognoScan, and there might be heterogeneity among these datasets. So, higher-resolution analysis such as single-cell RNA sequencing should be performed to verify our claims.

In conclusion, we investigated *TGFBI* expression characteristics, prognostic value, mutation profiles, associations with tumor-infiltrating immune cells, and associated molecular pathways in various types of cancer. We have provided new clues for improving cancer diagnosis and developing cancer immunotherapies that target *TGFBI*.

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

YC and HZ: Contributed to conception, design, data acquisition and interpretation, drafted and critically revised the manuscript. YG and YZF: Contributed to design and critically revised the manuscript. YF, QY and JH: Contributed to analysis and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant nos. 81773339 and 81800788), the Science and Technology Department of Hunan Province, China (Grant nos. 2017WK2041 and 2018SK52511) and the Open Sharing Fund for the Large-scale Instruments and Equipment of Central South University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.745649/full#supplementary-material

REFERENCES

Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., et al. (2020). Tumor Microenvironment Complexity and Therapeutic Implications at a Glance. *Cell Commun Signal* 18, 59. doi:10.1186/s12964-020-0530-4

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

Bhagirath, D., Abrol, N., Khan, R., Sharma, M., Seth, A., and Sharma, A. (2012). Expression of CD147, BIGH3 and Stathmin and Their Potential Role as Diagnostic Marker in Patients with Urothelial Carcinoma of the Bladder. *Clinica Chim. Acta* 413, 1641–1646. doi:10.1016/j.cca.2012.05.005

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

Bissey, P.-A., Law, J. H., Bruce, J. P., Shi, W., Renoult, A., Chua, M. L. K., et al. (2018). Dysregulation of the MiR-449b Target TGFBI Alters the TGFβ Pathway to Induce Cisplatin Resistance in Nasopharyngeal Carcinoma. *Oncogenesis* 7, 40. doi:10.1038/s41389-018-0050-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

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

Costanza, B., Rademaker, G., Tiamiou, A., De Tullio, P., Leenders, J., Blomme, A., et al. (2019). Transforming Growth Factor Beta-induced, an Extracellular Matrix Interacting Protein, Enhances Glycolysis and Promotes Pancreatic Cancer Cell Migration. *Int. J. Cancer* 145, 1570–1584. doi:10.1002/ijc.32247

Du, G.-W., Yan, X., Chen, Z., Zhang, R.-J., Tuoheti, K., Bai, X.-J., et al. (2020). Identification of Transforming Growth Factor Beta Induced (TGFBI) as an Immune-Related Prognostic Factor in clear Cell Renal Cell Carcinoma (ccRCC). *Aging* 12, 8484–8505. doi:10.18632/aging.103153

Fico, F., and Santamaria-Martínez, A. (2020). TGFBI Modulates Tumour Hypoxia and Promotes Breast Cancer Metastasis. *Mol. Oncol.* 14, 3198–3210. doi:10.1002/1878-0261.12828

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

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol. Cancer Ther.* 16, 2598–2608. doi:10.1158/1535-7163.mct-17-0386

Han, B., Cai, H., Chen, Y., Hu, B., Luo, H., Wu, Y., et al. (2015). The Role of TGFBI (βig-H3) in Gastrointestinal Tract Tumorigenesis. *Mol. Cancer* 14, 64. doi:10.1186/s12943-015-0335-z

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

Irigoyen, M., Pajares, M. J., Agorreta, J., Ponz-Sarvisé, M., Salvo, E., Lozano, M. D., et al. (2010). TGFBI Expression Is Associated with a Better Response to Chemotherapy in NSCLC. *Mol. Cancer* 9, 130. doi:10.1186/1476-4598-9-130

Kim, J.-E., Jeong, H.-W., Nam, J.-O., Lee, B.-H., Choi, J.-Y., Park, R.-W., et al. (2002). Identification of Motifs in the Fasciclin Domains of the Transforming Growth Factor-β-Induced Matrix Protein βig-h3 that Interact with the αvβ5 Integrin. *J. Biol. Chem.* 277, 46159–46165. doi:10.1074/jbc.m207055200

Ko, Y. C., Lai, T. Y., Hsu, S. C., Wang, F. H., Su, S. Y., Chen, Y. L., et al. (2020). Index of Cancer-Associated Fibroblasts Is Superior to the Epithelial-Mesenchymal Transition Score in Prognosis Prediction. *Cancers (Basel)* 12, 1718. doi:10.3390/cancers12071718

Lauden, L., Siewiera, J., Boukouaci, W., Ramgolam, K., Mourah, S., Lebbe, C., et al. (2014). TGF-β-Induced (TGFBI) Protein in Melanoma: A Signature of High Metastatic Potential. *J. Invest. Dermatol.* 134, 1675–1685. doi:10.1038/jid.2014.20

Li, B., Wen, G., Zhao, Y., Tong, J., and Hei, T. K. (2012). The Role of TGFBI in Mesothelioma and Breast Cancer: Association with Tumor Suppression. *BMC cancer* 12, 239. doi:10.1186/1471-2407-12-239

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, e108–e110. doi:10.1158/0008-5472.can-17-0307

Ma, C., Rong, Y., Radiloff, D. R., Datto, M. B., Centeno, B., Bao, S., et al. (2008). Extracellular Matrix Protein Ig-H3/TGFBI Promotes Metastasis of colon Cancer by Enhancing Cell Extravasation. *Genes Develop.* 22, 308–321. doi:10.1101/gad.1632008

Ma, J., Cui, W., He, S.-m., Duan, Y.-h., Heng, L.-j., Wang, L., et al. (2012). Human U87 Astrocytoma Cell Invasion Induced by Interaction of βig-h3 with Integrin α5β1 Involves Calpain-2. *PloS one* 7, e37297. doi:10.1371/journal.pone.0037297

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, 338–354. doi:10.1016/j.cell.2018.03.034

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

Nakazawa, N., Yokobori, T., Kaira, K., Turtoi, A., Baatar, S., Gombodorj, N., et al. (2020). High Stromal TGFBI in Lung Cancer and Intratumoral CD8-Positive T Cells Were Associated with Poor Prognosis and Therapeutic Resistance to Immune Checkpoint Inhibitors. *Ann. Surg. Oncol.* 27, 933–942. doi:10.1245/s10434-019-07878-8

Ozawa, D., Yokobori, T., Sohda, M., Sakai, M., Hara, K., Honjo, H., et al. (2016). TGFBI Expression in Cancer Stromal Cells Is Associated with Poor Prognosis and Hematogenous Recurrence in Esophageal Squamous Cell Carcinoma. *Ann. Surg. Oncol.* 23, 282–289. doi:10.1245/s10434-014-4259-4

Pajares, M. J., Agorreta, J., Salvo, E., Behrens, C., WistubaII, Montuenga, L. M., et al. (2014). TGFBI Expression Is an Independent Predictor of Survival in Adjuvant-Treated Lung Squamous Cell Carcinoma Patients. *Br. J. Cancer* 110, 1545–1551. doi:10.1038/bjc.2014.33

Razumilava, N., and Gores, G. J. (2014). Cholangiocarcinoma. *The Lancet* 383, 2168–2179. doi:10.1016/s0140-6736(13)61903-0

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, 166–180. doi:10.1593/neo.07112

Skonier, J., Neubauer, M., Madisen, L., Bennett, K., Plowman, G. D., and Purchio, A. F. (1992). cDNA Cloning and Sequence Analysis of βig-h3, a Novel Gene Induced in a Human Adenocarcinoma Cell Line after Treatment with Transforming Growth Factor-β. *DNA Cel. Biol.* 11, 511–522. doi:10.1089/dna.1992.11.511

Steitz, A. M., Steffes, A., Finkernagel, F., Unger, A., Sommerfeld, L., Jansen, J. M., et al. (2020). Tumor-associated Macrophages Promote Ovarian Cancer Cell Migration by Secreting Transforming Growth Factor Beta Induced (TGFBI) and Tenascin C. *Cell Death Dis* 11, 249. doi:10.1038/s41419-020-2438-8

Suzuki, M., Yokobori, T., Gombodorj, N., Yashiro, M., Turtoi, A., Handa, T., et al. (2018). High Stromal Transforming Growth Factor β-induced Expression Is a Novel Marker of Progression and Poor Prognosis in Gastric Cancer. *J. Surg. Oncol.* 118, 966–974. doi:10.1002/jso.25217

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

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

Thapa, N., Lee, B.-H., and Kim, I.-S. (2007). TGFBIp/βig-h3 Protein: A Versatile Matrix Molecule Induced by TGF-β. *Int. J. Biochem. Cel Biol.* 39, 2183–2194. doi:10.1016/j.biocel.2007.06.004

Tumbarello, D. A., Temple, J., and Brenton, J. D. (2012). β3 Integrin Modulates Transforming Growth Factor Beta Induced (TGFBI) Function and Paclitaxel Response in Ovarian Cancer Cells. *Mol. Cancer* 11, 36. doi:10.1186/1476-4598-11-36

Wang, B.-j., Chi, K.-p., Shen, R.-l., Zheng, S.-w., Guo, Y., Li, J.-f., et al. (2019). TGFBI Promotes Tumor Growth and Is Associated with Poor Prognosis in Oral Squamous Cell Carcinoma. *J. Cancer* 10, 4902–4912. doi:10.7150/jca.29958

Wang, N., Zhang, H., Yao, Q., Wang, Y., Dai, S., and Yang, X. (2012). TGFBI Promoter Hypermethylation Correlating with Paclitaxel Chemoresistance in Ovarian Cancer. *J. Exp. Clin. Cancer Res.* 31, 6. doi:10.1186/1756-9966-31-6

Wen, G., Hong, M., Li, B., Liao, W., Cheng, S. K., Hu, B., et al. (2011). Transforming Growth Factor-β-Induced Protein (TGFBI) Suppresses Mesothelioma Progression through the Akt/mTOR Pathway. *Int. J. Oncol.* 39, 1001–1009. doi:10.3892/ijo.2011.1097

Wen, G., Partridge, M. A., Li, B., Hong, M., Liao, W., Cheng, S. K., et al. (2011). TGFBI Expression Reduces *In Vitro* and *In Vivo* Metastatic Potential of Lung and Breast Tumor Cells. *Cancer Lett.* 308, 23–32. doi:10.1016/j.canlet.2011.04.010

Yan, H., Qu, J., Cao, W., Liu, Y., Zheng, G., Zhang, E., et al. (2019). Identification of Prognostic Genes in the Acute Myeloid Leukemia Immune Microenvironment Based on TCGA Data Analysis. *Cancer Immunol. Immunother.* 68, 1971–1978. doi:10.1007/s00262-019-02408-7

Yang, L., Cui, R., Li, Y., Liang, K., Ni, M., and Gu, Y. (2020). Hypoxia-Induced TGFBI as a Serum Biomarker for Laboratory Diagnosis and Prognosis in Patients with Pancreatic Ductal Adenocarcinoma. *Lab. Med.* 51, 352–361. doi:10.1093/labmed/lmz063

Yin, J., Liu, J.-S., Feng, M., Li, J.-M., Lu, S., Yang, M., et al. (2020). Comprehensively Investigating the Expression Levels and the Prognostic Role of Transforming Growth Factor Beta-Induced (TGFBI) in Glioblastoma Multiforme. *Translational Cancer Res.* 9. doi:10.21037/tcr-20-2906

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, H., Dong, S., and Feng, J. (2019). Epigenetic Profiling and mRNA Expression Reveal Candidate Genes as Biomarkers for Colorectal Cancer. *J. Cel Biochem* 120, 10767–10776. doi:10.1002/jcb.28368

Zou, J., Huang, R., Li, H., Wang, B., Chen, Y., Chen, S., et al. (2019). Secreted TGF-Beta-Induced Protein Promotes Aggressive Progression in Bladder Cancer Cells. *Cancer. Manage. Res* 11, 6995–7006. doi:10.2147/cmar.s208984

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 Chen, Zhao, Feng, Ye, Hu, Guo and Feng. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/fmolb.2021.724889

Prognostic Value of N6-Methyladenosine-Related lncRNAs in Early-Stage Colorectal Cancer: Association With Immune Cell Infiltration and Chemotherapeutic Drug Sensitivity

Zhizhong Xiong[1,2†], *Xianzhe Li*[1,2†], *Shi Yin*[1,2†], *Minghao Xie*[1,2], *Chaobin Mao*[2], *Fengxiang Zhang*[1], *Huaxian Chen*[1,2], *Longyang Jin*[1] and *Lei Lian*[1,2*]

[1]Department of Gastrointestinal Surgery, The Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China, [2]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

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Siva Koganti,
Stony Brook Medicine, United States
Tabish H. Khan,
University of Virginia, United States

***Correspondence:**
Lei Lian
lianlei2@mail.sysu.edu.cn

[†]These authors contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 14 June 2021
Accepted: 28 September 2021
Published: 12 October 2021

Citation:
Xiong Z, Li X, Yin S, Xie M, Mao C, Zhang F, Chen H, Jin L and Lian L (2021) Prognostic Value of N6-Methyladenosine-Related lncRNAs in Early-Stage Colorectal Cancer: Association With Immune Cell Infiltration and Chemotherapeutic Drug Sensitivity. Front. Mol. Biosci. 8:724889. doi: 10.3389/fmolb.2021.724889

Purpose: Accumulating evidence indicates that N6-methyladenosine-related long non-coding RNAs (m6A-related lncRNAs) play a crucial role in the occurrence and development of several cancers. We aimed to explore the potential role of m6A-related lncRNA signatures in predicting prognosis for early-stage (stages I and II) colorectal cancer (CRC).

Methods: m6A-related lncRNA data were obtained from The Cancer Genome Atlas. Univariate Cox regression analysis was used to screen for prognostic m6A-related lncRNAs. Immune characteristics were analyzed in different subgroups created via unsupervised clustering analysis. Next, patients were randomly divided into training and test cohorts. In the training cohort, least absolute shrinkage and selection operator (LASSO) regression was performed to establish a prognostic model. The predictive value of the signature was evaluated in the training and test cohorts. Drug sensitivity was also examined.

Results: A total of 1,478 m6A-related lncRNAs were identified. Two subgroups were created based on the expression of seven prognostic m6A-related lncRNAs. Prognosis was worse for cluster 1 than for cluster 2, and cluster 1 was characterized by increased numbers of M2 macrophages, decreased numbers of memory B cells, and higher expression of checkpoint genes when compared with cluster 2. Five m6A-related lncRNAs were selected to establish a risk prediction signature via LASSO regression. The 3 years overall survival (OS) was higher in the low-risk group than in the high-risk group. The area under the curve at 1, 2, and 3 years was 0.929, 0.954, and 0.841 in the

Abbreviations: CRC, colorectal cancer; GSEA, gene set enrichment analysis; IC50, half-maximal inhibitory concentration; LASSO, least absolute shrinkage and selection operator; lncRNA, non-coding RNA; m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs; m6A, N6-methyladenosine; miRNA, microRNAs; mRNAs, messenger RNAs; MSI, microsatellite instability; OS, overall survival; PD1, programmed cell death protein 1; ROC, receiver operating characteristic curve; TCGA, The Cancer Genome Atlas.

training cohort and 0.664, 0.760, and 0.754 in the test cohort, respectively. Multivariate Cox regression analysis suggests that the risk score was an independent predictor of OS in both the training and test cohorts. A prognostic nomogram based on the five m6A-related lncRNAs and their clinical features was built and verified. The high-risk group was more sensitive to chemotherapeutic drugs (camptothecin and cisplatin) than the low-risk group.

Conclusion: We identified two molecular subgroups of early-stage CRC with unique immune features based on seven prognostic m6A-related lncRNAs. Subsequent analyses demonstrated the usefulness of a five m6A-related lncRNA signature as a potential indicator of prognosis in patients with early-stage CRC.

Keywords: early-stage colorectal cancer, long non-coding RNA, N6-methyladenosine, nomogram, prognosis

INTRODUCTION

Colorectal cancer (CRC) is one of the most common gastrointestinal malignancies and is among the main causes of cancer-related deaths. Although the 5 years survival rate for early-stage CRC (stage I and stage II) can reach 90%, these patients are still at risk of recurrence, which is often fatal (Siegel et al., 2019; Walker et al., 2020). Whether chemotherapy is necessary for patients with early-stage CRC depends mainly on known clinical and pathological risk factors, such as microsatellite instability (MSI) status and bowel obstruction (Manfredi et al., 2006; Kannarkatt et al., 2017). However, these high-risk factors cannot distinguish between patients with poor prognosis and those who benefit from chemotherapy (Dienstmann et al., 2015; Kopetz et al., 2015). Therefore, a reliable molecular marker is urgently needed to identify high-risk groups of patients with early-stage CRC and to optimize treatment strategies.

N6-Methyladenosine (m6A)—which is present in messenger RNAs (mRNAs), non-coding RNAs (lncRNAs), and microRNAs (miRNAs) in most eukaryotes—is the most common epigenetic methylation modification of mammalian RNA (Desrosiers et al., 1974; Molinie and Giallourakis, 2017). M6A modification is regulated by a series of protein factors, including methyltransferases (writers), signal transducers (readers), and demethylases (erasers) (Yang et al., 2018a). Abnormal modification of m6A plays an important role in the occurrence and development of many tumors, such as hepatocellular carcinoma, breast cancer, glioblastoma, and lung cancer (Lin et al., 2016; Cui et al., 2017; Vu et al., 2017; Cai et al., 2018; Liu et al., 2019). Moreover, Zhang et al. demonstrated the effect of mutation of the m6A regulator on CRC prognosis (Zhang et al., 2020). Similar to mRNA, lncRNA is also regulated by m6A. Moreover, m6A-related lncRNAs also regulate a series of biological and pathological processes. Recent studies have demonstrated that m6A-related lncRNAs can reliably predict the prognosis of low-grade glioma, lung adenocarcinoma, and gastric cancer (Tu et al., 2020; Wang et al., 2021a; Wang et al., 2021b). Luo et al.(Zuo et al., 2021) further observed that m6A-related lncRNAs were associated with the occurrence and development of CRC, indicating that they may be an accurate prognostic factor for early-stage CRC.

The tumor microenvironment (TME) also plays an important role in tumor initiation and progression, and is comprised of tumor cells, stromal cells and innate and adaptive immune cells (Colangelo et al., 2017). Immunotherapy has quickly become the main treatment modality for CRC, such as programmed cell death1 (PD1)-blocking antibodies (Ganesh et al., 2019). Recently, more studies have focused on comprehensive analysis of specific m6A regulatory factors to strengthen the in-depth understanding of the heterogeneity and complexity of TME (Yang et al., 2019; Li et al., 2020). For example, Na Li et al. (Li et al., 2020) have found a deletion in Alkbh5 that sensitizes tumors to cancer immunotherapy by regulating myeloid-derived suppressor cells and suppressive lymphocyte Treg accumulation. However, at present, there are no studies on this aspect of early-stage CRC.

In the present study, we identified two molecular subgroups of early-stage CRC with unique immune characteristics and showed the potential role of m6A-related lncRNA signatures in predicting prognosis in patients with early-stage CRC.

MATERIALS AND METHODS

Study Approval and Consent

This study was approved by the Medical Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China (no. 2021ZSLYEC-006).

Data Acquisition and Preprocessing

Initially, we downloaded RNA sequencing data (FPKM value) related to gene expression and the corresponding clinical information for CRC from The Cancer Genome Atlas (TCGA) website (https://portal.gdc.cancer. gov). According to the 8th edition of the American Joint Committee on Cancer, we identified patients with early-stage CRC. The clinical information included age, sex, TNM stage, and survival status. In addition, patients with no survival information or a survival time of less than 3 months were excluded from further evaluation to reduce statistical bias. To distinguish the mRNAs and lncRNAs, we downloaded the GTF files from Ensembl (http://asia.ensembl.org) for further analysis. Data were normalized, processed, and analyzed using R software 4.0.3.

Identification of m6A-Related lncRNAs

We extracted expression data for 23 m6A-related genes identified in previous studies, including methyltransferases (*METTL3*,

METTL14, *METTL16*, *WTAP*, *VIRMA*, *ZC3H13*, *RBM15*, and *RBM15B*), demethylases (*ALKBH5* and *FTO*), and recognition proteins (*YTHDC1*, *YTHDC2*, *YTHDF1*, *YTHDF2*, *YTHDF3*, *HNRNPC*, *FMR1*, *LRPPRC*, *HNRNPA2B1*, *IGFBP1*, *IGFBP2*, *IGFBP3*, and *RBMX*) (Arguello et al., 2017; Huang et al., 2021a). Pearson correlation analysis was performed between the m6A genes and all lncRNAs. Absolute correlation coefficients >0.4 and p values <0.001 were used to define m6A-related lncRNAs. Univariate Cox regression analysis was used to screen for m6A-related lncRNAs significantly associated with overall survival (OS) ($p < 0.01$).

Unsupervised Clustering of Seven m6A-Related lncRNAs Associated With Prognosis

Unsupervised clustering analysis was used to classify patients with early-stage CRC into different subgroups based on the expression of the seven m6A-related lncRNAs associated with prognosis. The ConsensusClusterPlus package was used to perform the above steps and was repeated 1,000 times to ensure the stability of the classification (Wilkerson and Hayes, 2010). Then, we used the Euclican disease to compute the similarity disease between patients with early-stage CRC. The optimal number of clusters was identified by CDF and consensus matrices.

GSEA (version 4.1.0) software was applied to determine the gene expression in cluster1 and cluster2 in the Molecular Signatures database (MSigDB) Collection (c2.cp.kegg. v7.2symbols.gmt) to further analyze the difference in KEGG pathway enrichment. The threshold of statistical significance was defined by a nominal p-value <0.05 and a false discovery rate (FDR) q value <0.25 (Subramanian et al., 2005).

Immune Characteristics for Molecular Subtypes of Early-Stage CRC

The ESTIMATE algorithm was used to calculate the ESTIMATE score, stromal score, and immune score between cluster 1 and cluster 2 (Gentles et al., 2015). The CIBERSORT algorithm, which can sensitively and specifically distinguish 22 human immune cell phenotypes, was used to quantify the immune cells in early-stage CRC samples. The Wilcoxon signed-rank test was used to analyze the differences in immune infiltrating cells between the different molecular subtypes using the above method. In addition, levels of checkpoint gene expression were also compared. Furthermore, the relationship between PD1 and m6A-related lncRNAs associated with prognosis was quantified using Pearson correlation analysis. The thresholds were set as p-value < 0.05.

Establishment and Validation of the m6A-Related lncRNA Risk Prognosis Model

First, patients with early-stage CRC were randomly divided into training and test cohorts at a ratio of 1:1 using the caret package. Second, least absolute shrinkage and selection operator (LASSO) Cox regression analysis was performed to analyze the best candidates and multiple m6A-related lncRNA characteristics for constructing the prognostic signature. Ten-fold cross-validation was used to prevent overfitting. Based on the above model, the risk score for each patient was calculated as follows: risk score = $\sum_{i=1}^{k} \beta i Si$ (k: the number of m6A-related lncRNAs incorporated into the signature; βi: the coefficient for each m6A-related lncRNA; Si: the level of m6A-related lncRNA expression). Then, patients with early-stage CRC were divided into high-risk and low-risk groups based on the median score of the training cohort. The Kaplan–Meier method and log-rank test were used to evaluate the survival difference between the two risk groups in the training and test cohorts. The area under the curve (AUC) and time-dependent receiver operating characteristic (ROC) curves were examined to assess the predictive power of the signal for survival using the "survival" and "survivalROC" packages in R software. The "survcomp" package is used to calculate the Consistency index (C-index) to compare the prediction accuracy of prognostic features. Univariate and multivariate Cox regression analyses were conducted to verify the independence of the risk score by comparing common clinical features including age, sex, and stage. We used a forest map to demonstrate the results. The R package used in the above operation is "survival", "pheatmap", and "ggpupr".

Construction of a Predictive Nomogram for OS in Patients With Early-Stage CRC

Nomograms are widely used to predict the outcomes of patients with cancer (Iasonos et al., 2008). Risk scores and clinical indicators were incorporated into a nomogram to evaluate the probability of 1, 2, and 3 years OS in patients with early-stage CRC using "rms" R package. The C-index, ROC curves, and calibration curves (by a bootstrap method with 1,000 peplicates) were used to evaluate the predictive ability and discriminative value of the nomogram.

Chemosensitivity Prediction

The Cancer Drug Sensitivity Genomics (GDSC) database (https://cancerrxgene.org) can be used for large-scale drug screening. Combined with genomic analysis, the response of tumors to chemotherapy drugs can be systematically identified. Using the GDSCdatabase, the half-maximal inhibitory concentration (IC_{50}) of common chemotherapeutic drugs for gastrointestinal tumors (camptothecin, cisplatin, rapamycin, bryostatin, and methotrexate) was calculated to evaluate the clinical application of this model in the treatment of early-stage CRC. The R package used in the above operation is "pRRophetic". Then, we compared the difference in the IC_{50} between the high-risk group and low-risk group using the Wilcoxon signed-rank test. To visualize the data, box drawings were created using "pRRophetic" and "ggplot2" in R (Slavin, 1989).

Statistical Analysis

R software (v 4.0.3) was used for statistical analysis (R packages: "limma," "pheatmap," "survival," "survminer," "glmnet," "reshape2," "ggpubr," "ConsensusClusterPlus," "ggplot2,"

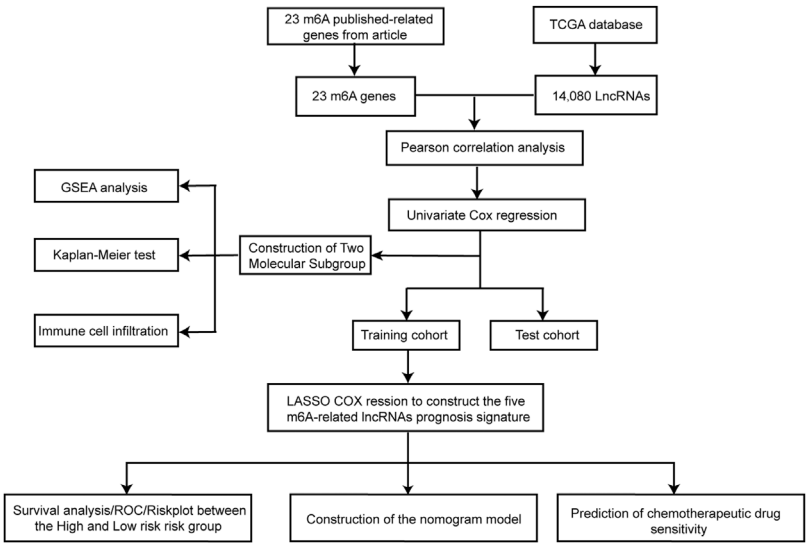


FIGURE 1 | Flow diagram of the study.

TABLE 1 | Clinical characteristics of early-stage CRC in TCGA.

Clinical characteristics	Cases
Patients, n	251
Survival status, n (%)	
Alive	228(90.8%)
Dead	23(9.2%)
Age, n (%)	
≤60	69(27.4%)
>60	182(72.6%)
Sex, n (%)	
Male	144(57.4%)
Female	107(42.6%)
TNM stage, n (%)	
Stage I	83(33.1%)
Stage II	168(66.9%)

CRC, colorectal cancer; TCGA, The Cancer Genome Atlas.

"corrplot," "utils," "vioplot," "plyr," "grid," "gridExtra (multi-GSEA)," "pRRophetic," "caret," "glmnet," "timeROC," "ggExtra", "survcomp"). In the above statistical tests, a two-sided *p*-value <0.05 was considered statistically significant.

RESULTS

Identification of m6A-Related lncRNAs Associated With Prognosis in Early-Stage CRC

The bioinformatic analysis of this study was performed as depicted in **Figure 1**. First, a total of 14,081 lncRNA expression and 23 m6A-related gene expression profiles were obtained from 299 early-stage CRC and 31 normal samples. The corresponding clinical data of 251 patients were also extracted from TCGA (**Table 1**). Then, 1,478 m6A-related lncRNAs were identified and seven survival-associated m6A-related lncRNAs were obtained (**Figure 2A**). The expression levels of these lncRNAs differed significantly between normal and tumor tissues ($p < 0.05$, **Figures 2B,C**). Expression levels of the three m6A-related lncRNAs (LINC00562, AC007991.4, and AL121583.1) in tumor tissue were significantly higher than those in normal tissues, whereas the expression levels of the remaining four lncRNAs (EPS15-AS1, AC087277.2, AC008494.3, and AC244629.1) were higher in normal tissues.

Construction of Two Molecular Subgroups of Early-Stage CRC Using Seven Survival-Associated m6A-Related lncRNAs

Unsupervised clustering identified two molecular subgroups as the optimal number of clusters, including 171 cases in cluster one and 80 cases in cluster two (**Figure 3A**, **Supplementary Figure S1**). Principal component analysis (PCA) was conducted based on the above classification methods, and there was a significant difference in the distributions of cluster one and cluster two (**Figure 3B**). The KEGG results revealed that "aminoacyl tRNA biosynthesis," "antigen processing and presentation," "cell cycle," "cysteine and methionine metabolism," "natural killer cell-mediated cytotoxicity," "protein export," and "ubiquitin-mediated proteolysis" were markedly enriched in cluster 1 (**Figure 3C**). Furthermore, the prognosis of cluster 1 was poorer than that of cluster 2 (**Figure 3D**). These results indicate that there are molecular subgroups of early-stage CRC with different characteristics.

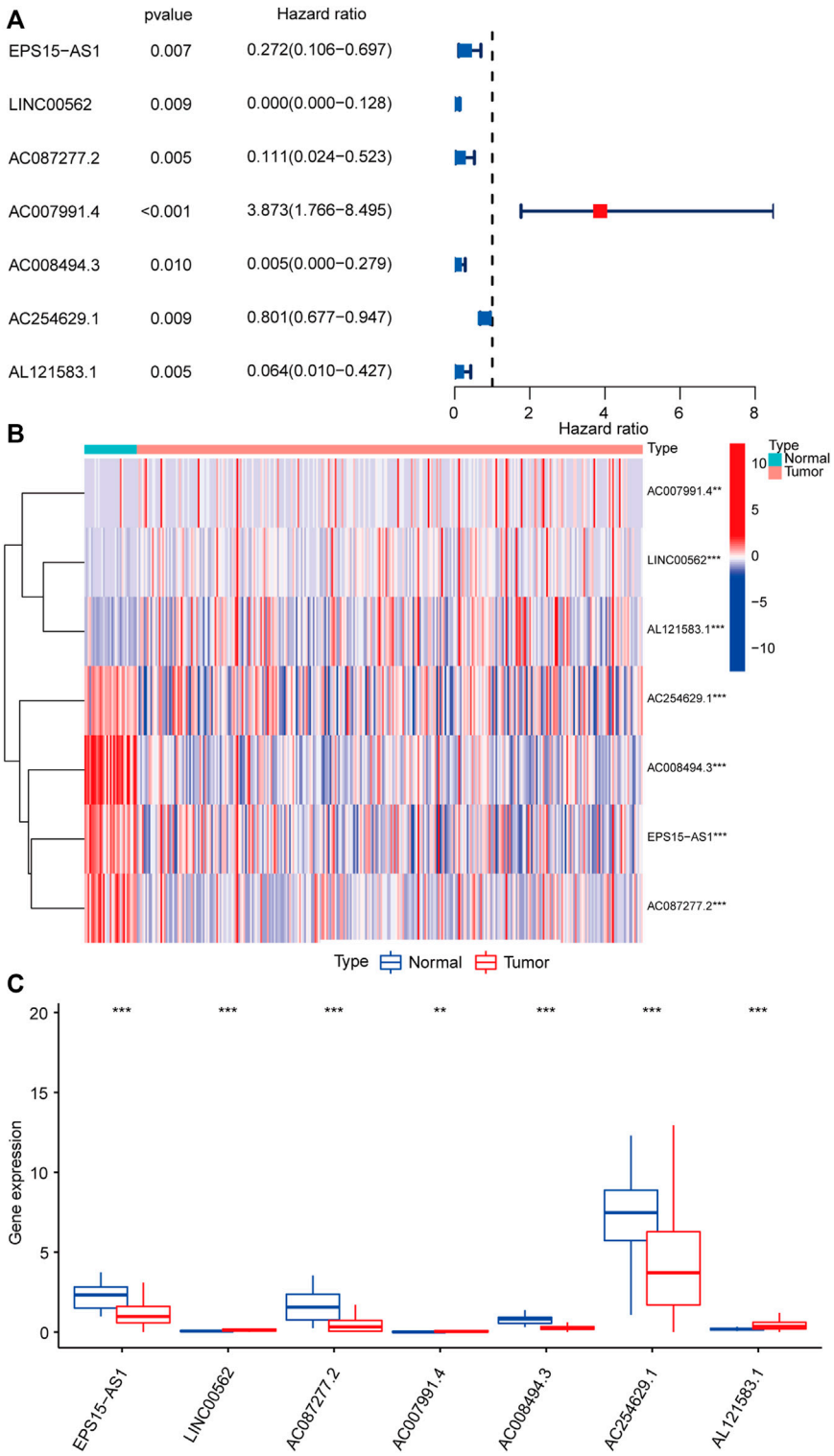


FIGURE 2 | Identification of m6A-related lncRNAs associated with the prognosis of early-stage CRC. Forest plot of seven m6A-related lncRNAs associated with prognosis based on univariate Cox regression analysis. Red indicated that the m6a-related lncRNA is a risk factor for the prognosis of patients with early-stage CRC, and blue indicated that m6a-related lncRNA is a protective factor for the prognosis of patients with early-stage CRC **(A)**. Heatmap of seven differentially expressed m6A-related lncRNAs between normal colorectal tissue (marked in green) and early-stage CRC (marked in red) **(B)**. Expression patterns of seven m6A-related lncRNAs in normal tissue (marked in blue) and tumor tissue (marked in red) **(C)**. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. Abbreviation: m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs; CRC, colorectal cancer.

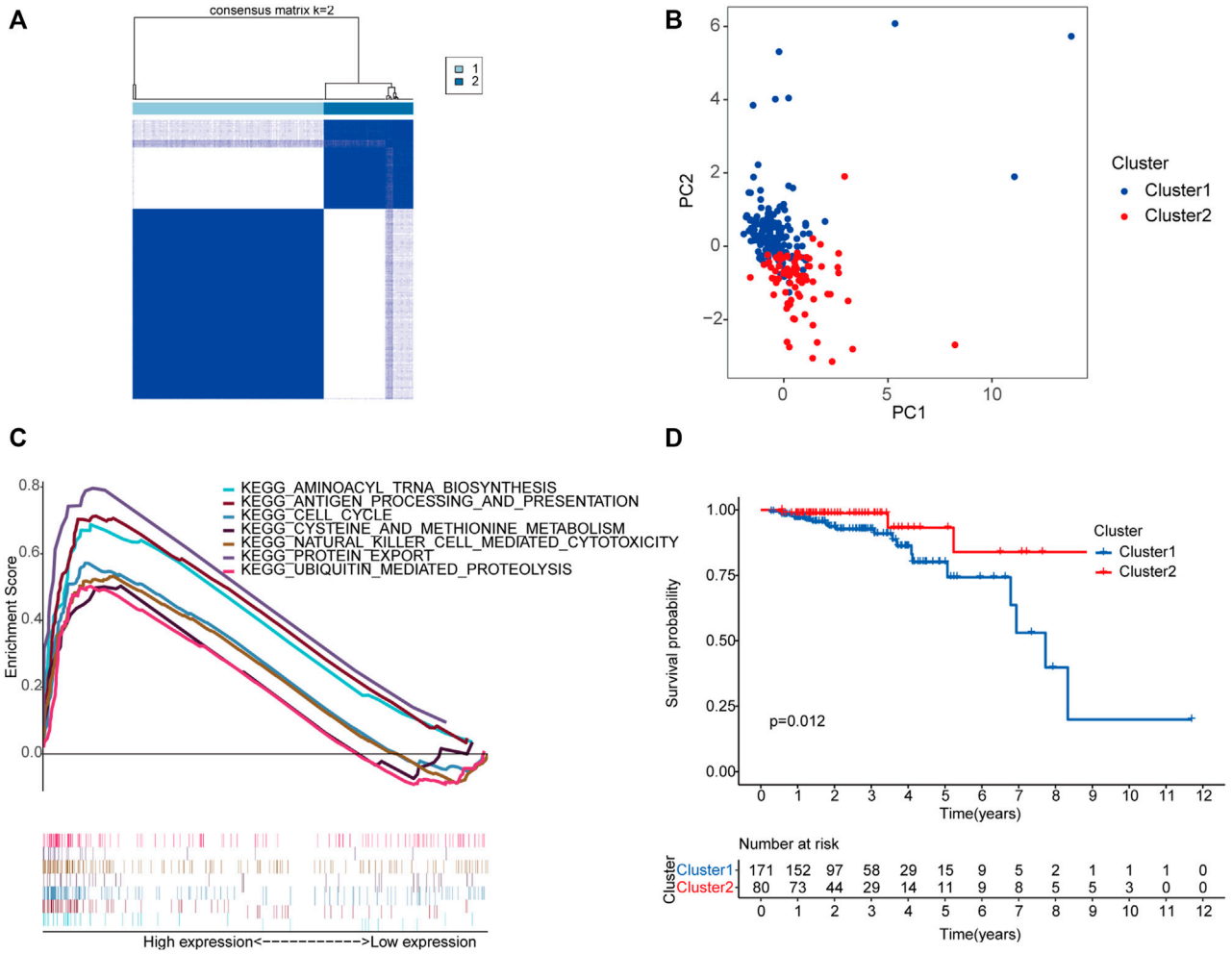


FIGURE 3 | Construction of two molecular subgroups of early-stage CRC using seven m6A-related lncRNAs associated with prognosis. The consensus matrices of the TCGA database for $k = 2$ **(A)**. PCA between cluster one (marked in blue) and cluster two (marked in red) **(B)**. GSEA analysis of the different activation status of the biological pathways between cluster one and cluster 2 **(C)**. Survival analysis of two subgroups (blue represented cluster 1, and red represented cluster 2) based on data from 251 patients with early-stage CRC in the TCGA cohort **(D)**. Abbreviation: CRC, colorectal cancer; m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs; PCA, principal component analysis; GSEA, gene set enrichment analysis; TCGA, The Cancer Genome Atlas.

Immune Characteristics of the Two Molecular Subgroups of Early-Stage CRC

In addition to tumor cells, the tumor microenvironment includes immune cells and stromal cells, which provide protection and support for the occurrence and development of tumors. As shown in **Figure 4A**, no significant differences in ESTIMATE or stromal scores were observed between cluster 1 and cluster 2; however, immune scores were higher in cluster 1 than in cluster two (p = 0.024), indicating that levels of immune cells differ based on molecular subgroup. Our analysis revealed enrichment of the proportion of M2 macrophages in cluster one, as well as a higher proportion of memory B cells in cluster two (**Figure 4B**). We also examined differences in the expression profiles of immune checkpoint genes between clusters 1 and 2. Levels of most immune checkpoint genes were higher in cluster 1 than in cluster two, suggesting that patients in cluster 1 were more likely to exhibit immune escape and poor prognosis (**Figure 4C**). We also observed that the expression of PD1 was positively correlated with AC087277.2 and AC007991.4 yet negatively correlated with AC254629.1 (**Figure 4D**).

Construction and Validation of the m6A-Related lncRNA Prognostic Signature

The 251 patients were randomly divided into a training cohort (n = 127) and a test cohort (n = 124) in a 1:1 ratio. To build the m6A-related lncRNA signature for forecasting the prognosis of patients with early-stage CRC, LASSO Cox analysis was performed based on the seven m6A-related lncRNAs in the training cohort. The m6A-related lncRNA signature included five m6A-related lncRNAs and the coefficients of each (**Figures 5A,B**). Based on the coefficients, the risk scores of each patient were calculated using the following formula: risk score = –0.828 ×

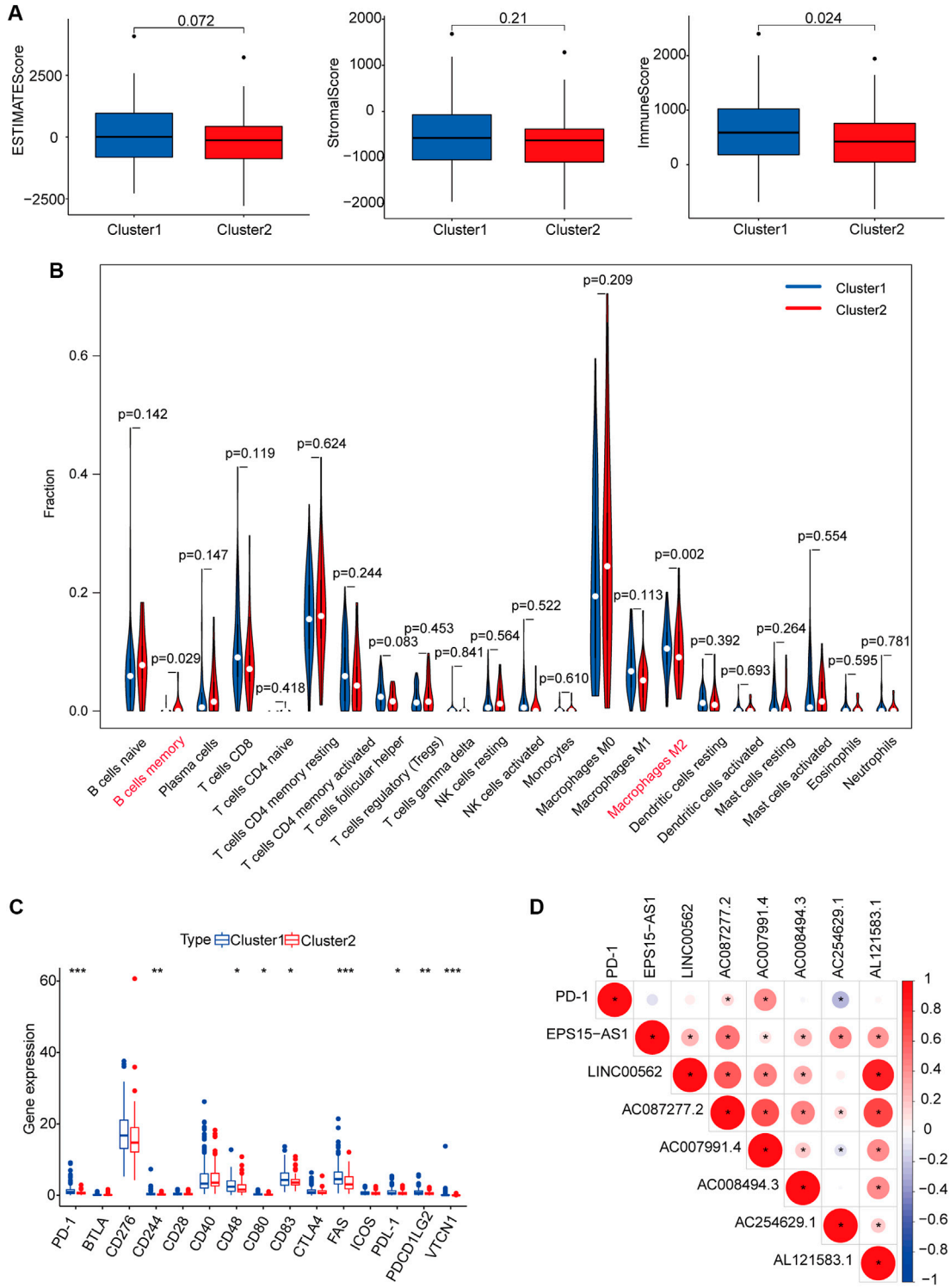


FIGURE 4 | Immune cell infiltration and biological functions in the two subgroups. ESTIMATE, stromal, and immune scores in cluster 1 (marked in blue) and cluster 2 (marked in red) **(A)**. Abundance of 22 immune cells in the two subgroups (blue represented cluster 1, and red represented cluster 2) **(B)**. Expression of checkpoint genes in the two subgroups (Blue represented cluster 1, and red represented cluster 2). $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ **(C)**. Heatmap for the correlation between PDL1 gene and seven prognostic m6A-related lncRNAs. Blue represented a positive correlation between the PD-1 gene and the m6A-related lncRNA, and red represented a negative correlation between the PD-1 gene and the m6A-related lncRNA. The darker the color, the stronger the relevance **(D)**. $^{*}p < 0.05$. Abbreviation: m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs.

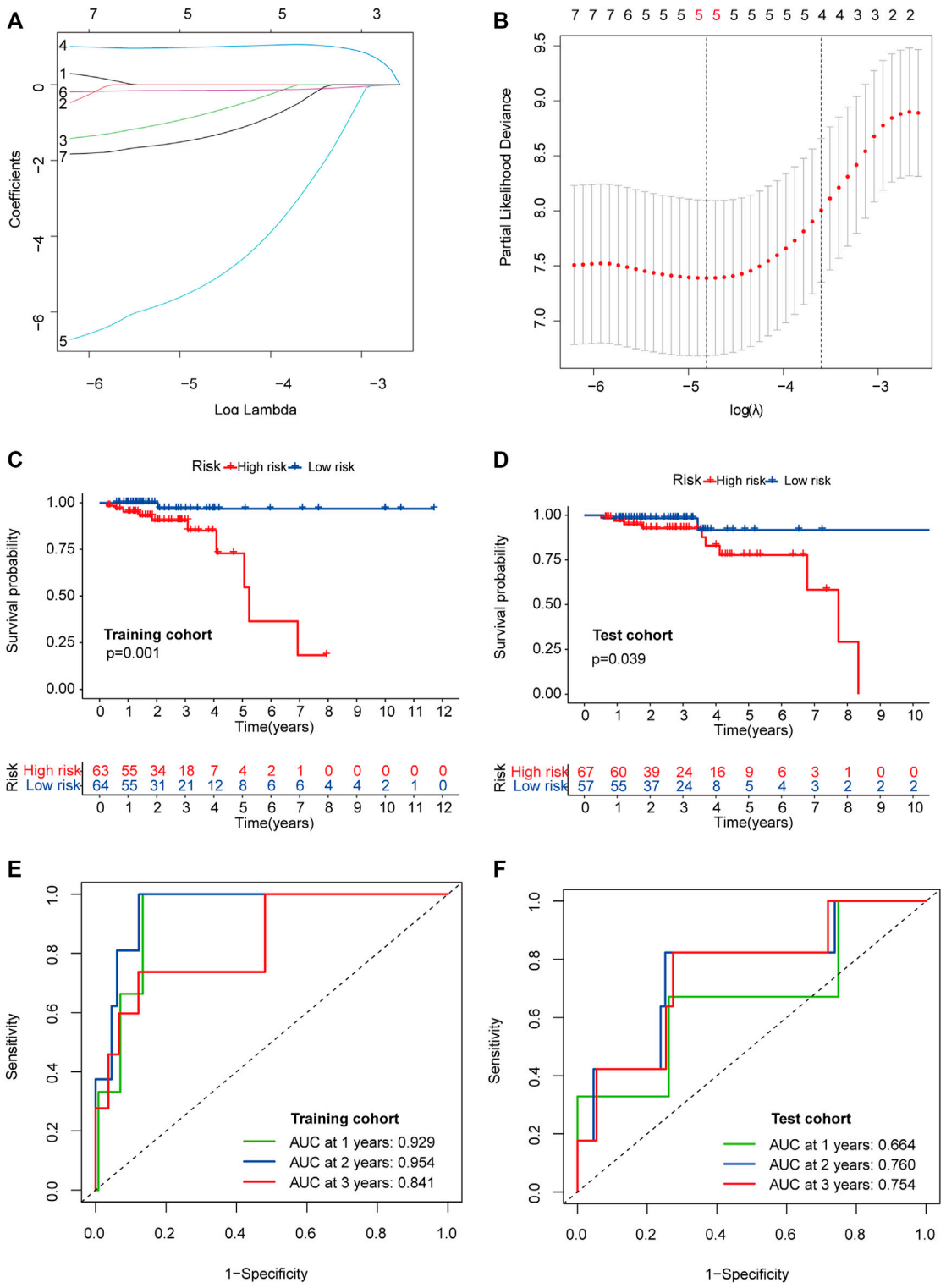


FIGURE 5 | Construction and validation of the prognostic risk model. LASSO coefficient profiles of seven candidates in the training cohort **(A)**. Selection of the optimal parameter (lambda) in the LASSO model **(B)**. Survival analysis in the high- (marked in red) and low-risk groups (marked in blue) in the training cohort **(C)** and test cohort **(D)**. Time-dependent ROC analysis for 1, 2, and 3-years OS prediction among patients with early-stage CRC in the training **(E)** and test cohorts **(F)**. Abbreviation: LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; OS, overall survival; CRC, colorectal cancer.

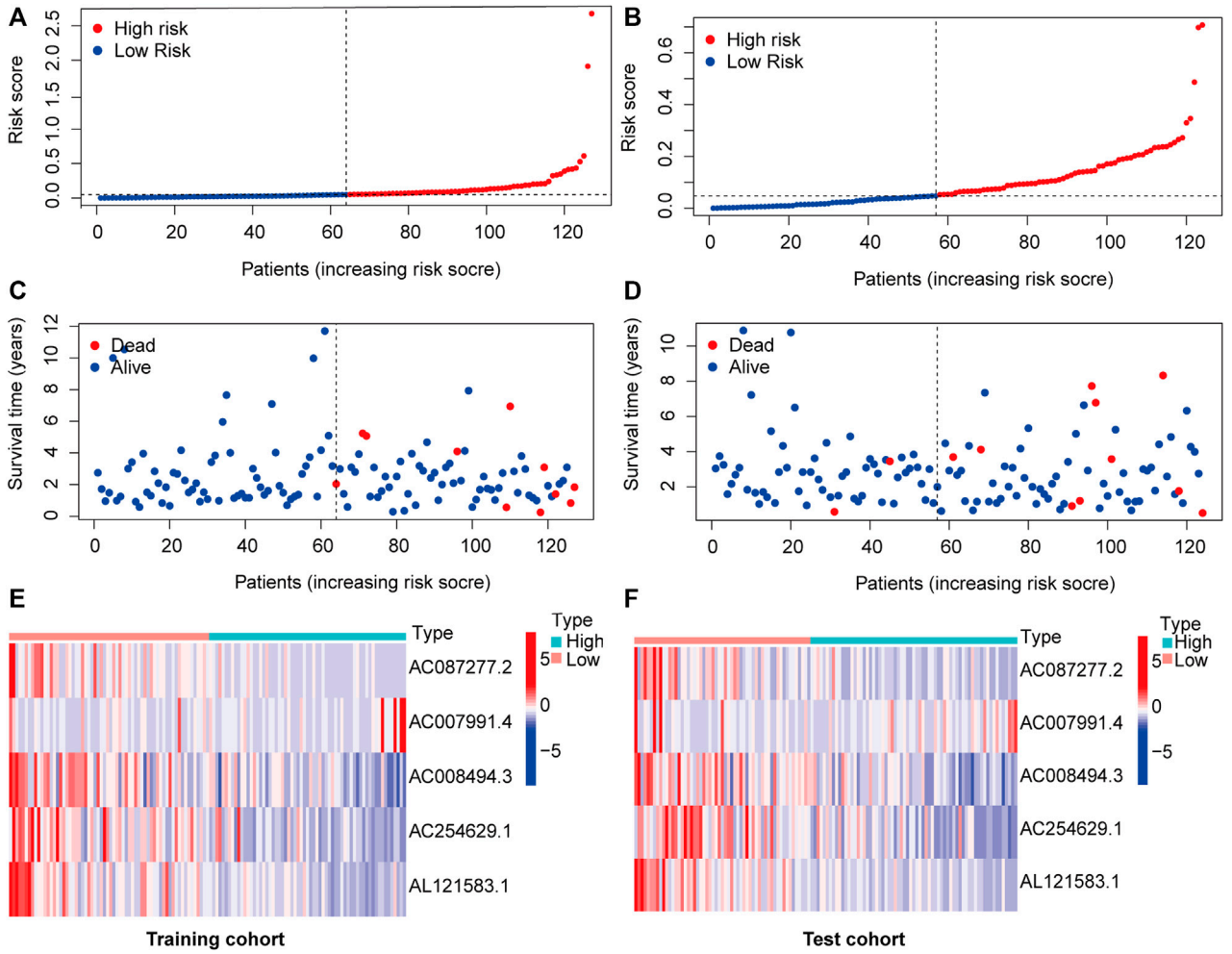


FIGURE 6 | m6A-related lncRNA signature risk score analysis. Risk curves of each sample reordered by risk scores in the training cohort **(A)** and test cohort **(B)**. Living status of patients with early-stage CRC in the training cohort **(C)** and test cohort **(D)**. Heatmap of five m6A-related lncRNA expression profiles in the low- (marked in red) and high-risk groups (marked in green) in the training cohort **(D)** and test cohort **(F)**. Abbreviation: m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs; CRC, colorectal cancer.

AC087277.2 + 0.989 × AC007991.4–5.396 × AC008494.3–0.151 × AC254629.1–1.435 × AL121583.1. Patients in the training cohort were grouped into low-risk and high-risk groups according to the median value of the risk scores. The Kaplan–Meier survival curve showed that patients with early-stage CRC with higher risk scores had poorer clinical outcomes than those with relatively lower risk scores (**Figure 5C**). The ROC curve showed that the m6A-related lncRNA signature had a good ability to predict OS in the training cohort (1 year AUC = 0.929, 2-years AUC = 0.954, 3-years AUC = 0.841; **Figure 5E**). Consistent with the results in the training cohort, higher risk scores were associated with shorter OS times and lower OS rates in the test cohort (**Figure 5D**). The ROC analysis also revealed that the m6A-related lncRNA signature had a strong prognostic value in the test cohort (1 year AUC = 0.664, 2 years AUC = 0.760, 3 years AUC = 0.754; **Figure 5F**). The distributions of risk, survival status, and expression of five m6A-related lncRNAs in the training and test cohorts are shown in **Figure 6**, which suggests that patients with higher risk scores exhibited shorter OS. Univariate and multivariate Cox regression analyses show that risk scores were significantly related to OS, independent of clinical parameters (**Figures 7A–D**).

Construction of the Nomogram Based on the m6A-Related lncRNA Signature

To develop a clinically applicable method for predicting OS in patients with early-stage CRC, we built a nomogram using age, sex, stage, and risk scores (**Figure 8A**). The C-index of the nomogram for predicting the OS rate was 0.821. The nomogram had excellent accuracy regarding the 1, 2, and 3-years OS rates (AUC = 0.860, 0.873, and 0.842, respectively; **Figure 8B**). Moreover, the calibration curve demonstrated that the nomogram was suitable for clinical practice (**Figure 8C**).

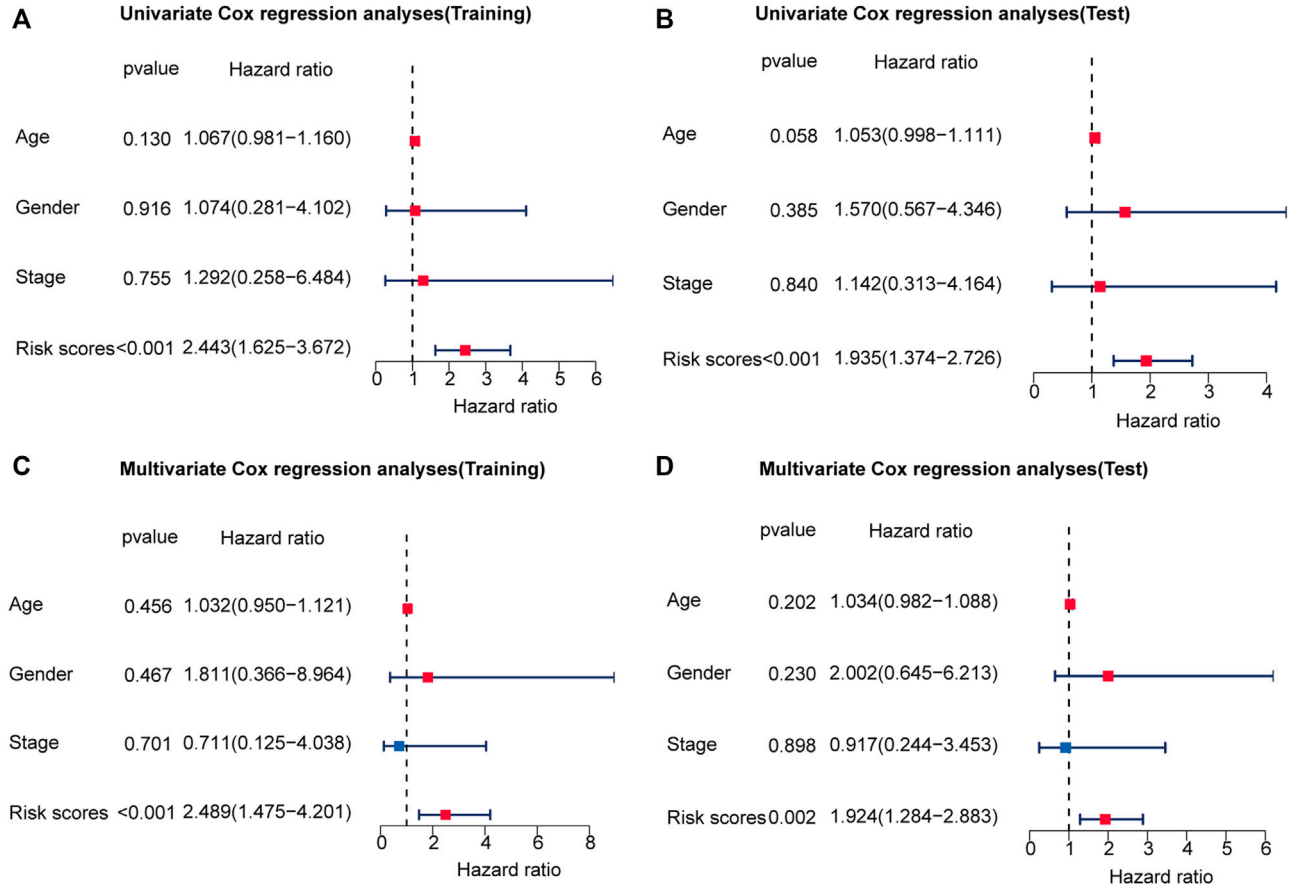


FIGURE 7 | The risk score is an independent prognostic factor for early-stage CRC. Univariate regression analysis of early-stage CRC in the training cohort **(A)** and test cohort **(B)**. Multivariate Cox regression analysis of early-stage CRC in the training cohort **(C)** and test cohort **(D)**. Red indicated that the indicator is a risk factor for the prognosis of patients with early-stage CRC, and blue indicated that the indicator is a protective factor for patients with early-stage CRC. And *p*-value < 0.05 was considered statistically significant.

Response to Chemotherapy in the High and Low-Risk Groups

To identify potential chemotherapeutic drugs targeting the m6A-related lncRNA signature for treating patients with early-stage CRC, the pRRophetic algorithm was used to estimate the therapeutic response based on the IC_{50} for each sample. We found that the high-risk group was more sensitive to camptothecin and cisplatin, which may provide insight into new treatment options for patients with early-stage CRC (**Figures 9A–D**).

DISCUSSION

Even after radical surgical treatment, the risk of recurrence in patients with stage I and stage II CRC is still 5 and 20%, respectively, (Dotan and Cohen, 2011). Conventional high-risk characteristics, such as MSI status and bowel obstruction, cannot meet the demands for predicting prognosis in patients with early-stage CRC. However, numerous studies have highlighted the relevance of lncRNAs. Specific lncRNAs can be modified by m6A regulators to participate in tumorigenesis, development, and metastasis (Tu et al., 2020). In the present study, we extracted data regarding the expression of lncRNA and m6A in these patients from TCGA. Seven m6A-associated lncRNAs related to prognosis were identified using univariate Cox regression analysis. Based on these lncRNAs, we identified two molecular subgroups of early-stage CRC with unique immune characteristics. Furthermore, we constructed and validated a novel m6A-related lncRNA model for predicting OS in patients with early-stage CRC, following which we developed a nomogram for clinical application in this patient population. Finally, we assessed responses to common chemotherapeutic drugs in the high- and low-risk groups.

Here, based on seven m6A-related lncRNAs associated with prognosis, we identified two molecular subgroups of early-stage CRC with significantly distinct features. Prognosis was worse for cluster 1 than for cluster two, and cluster 1 was characterized by increased numbers of M2 macrophages, decreased numbers of memory B cells, and increased expression of checkpoint genes when compared with cluster two. Tumor-associated macrophages

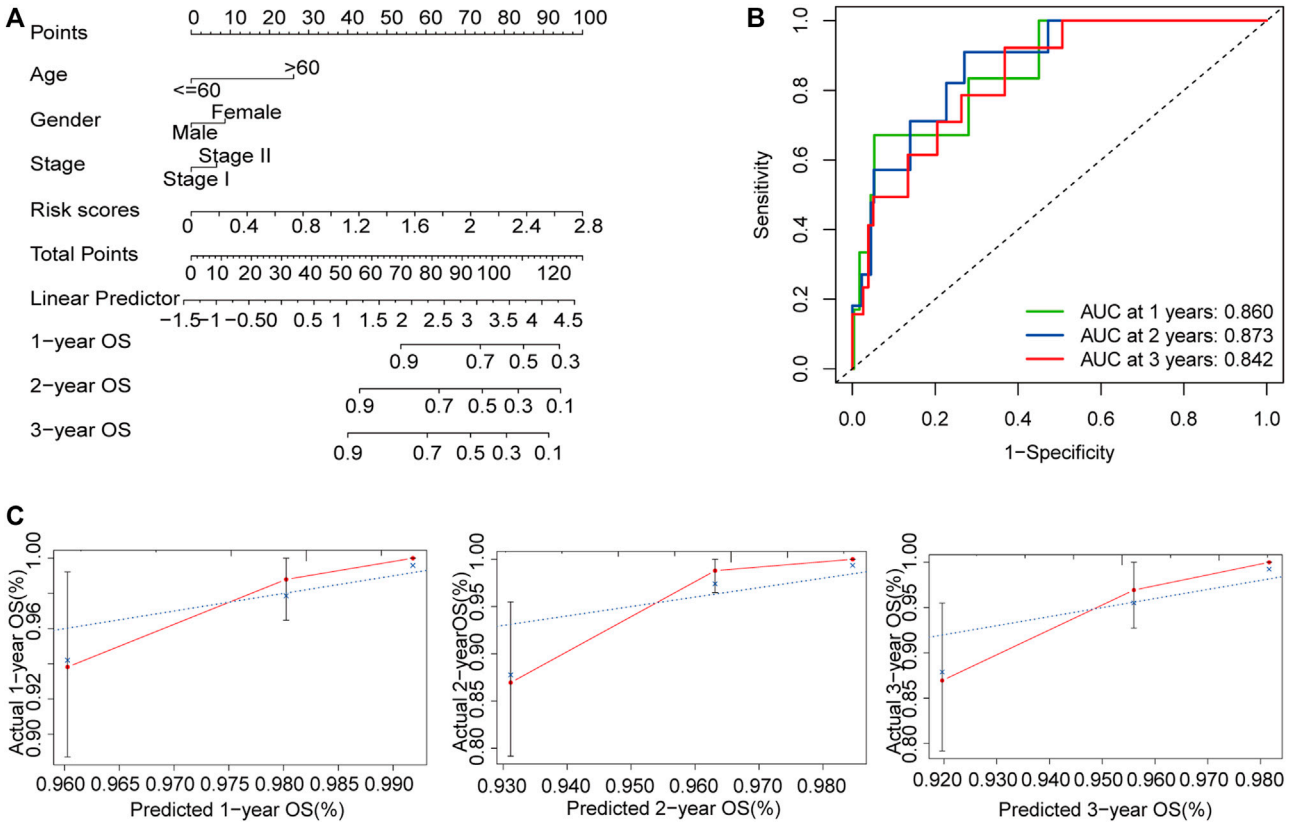


FIGURE 8 | Construction of a nomogram for the prediction of outcomes in patients with early-stage CRC. The nomogram combining clinical features and the m6A-related lncRNA signature **(A)**. ROC curves for the nomogram for predicting 1, 2, and 3-years OS **(B)**. Calibration plot showing that the predicted probability and the actual probability were consistent **(C)**. Abbreviation: m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs; CRC, colorectal cancer; ROC, receiver operating characteristic; OS, overall survival.

are comprised of M1 and M2 macrophages (Mantovani et al., 2017). It is well known that M1 (anti-tumor) and M2 (pro-tumor) phenotypes are related to different immunomodulatory functions (Yang et al., 2018b). In a previous study, Catherine et al. inferred that B lymphocytes may exert anti-tumor immune functions by collaborating with T lymphocytes (Sautès-Fridman et al., 2016). Edin et al. also found that tumor-infiltrating $CD20^{+}$ B lymphocytes are associated with a favorable prognosis in CRC (Edin et al., 2019). Thus, cluster 1 may be associated with immunosuppression. Furthermore, expression levels of immune checkpoint genes were higher in cluster 1 than in cluster two, which is a characteristic of immunosuppression. This finding suggests that patients in cluster 1 were more likely to exhibit immune escape and poor prognosis than those in cluster two.

It has been reported that m6A-related lncRNAs are involved in the occurrence and development of a variety of tumors. For example, m6A on lncRNA NEAT1-1 may be a novel specific marker for bone metastasis and is correlated with poor prognosis (Wen et al., 2020). METL3 and METTL14 on LNCAROD have tumor-promoting functions in the development of head and neck squamous cell carcinoma (Ban et al., 2020). Moreover, m6A-related lncRNAs can reliably predict the prognosis in several tumors such as metastatic skin cutaneous melanoma and adrenocortical carcinoma (Huang et al., 2021b; Jin et al., 2021). Similar to these studies, we conducted a series of bioinformatics analyses to generate a new m6A-related lncRNA signature for early-stage CRC prognosis. The signature included AC087277.2, AC007991.4, AC008494.3, AC254629.1, and AL121583.1. The construction of a novel signature allowed us to distinguish samples at different risks. In the training cohort, the OS time of early-stage CRC was shorter in the high-risk group than in the low-risk group, and consistent results were obtained in the test cohort. In addition, univariate and multivariate Cox regression analyses confirmed that the risk score was an independent predictor of OS in patients with early-stage CRC. Taken together, these results highlight the ability of our m6A-related lncRNA prognostic signature to accurately predict OS in this population.

High-risk patients with early-stage CRC are usually treated via a combination of surgery and chemotherapy. Using the GDSC database, we found that high-risk patients were more sensitive to commonly used chemotherapy drugs (including camptothecin and cisplatin) than low-risk patients, which may provide new treatment options for patients with early-stage CRC. Camptothecin is an alkaloid extracted from *Camptotheca acuminata* that can inhibit the catalytic activity of topoisomerase (Hsiang et al., 1985; Mollica et al., 2012). In agreement with previous reports (Cross-Knorr et al.,

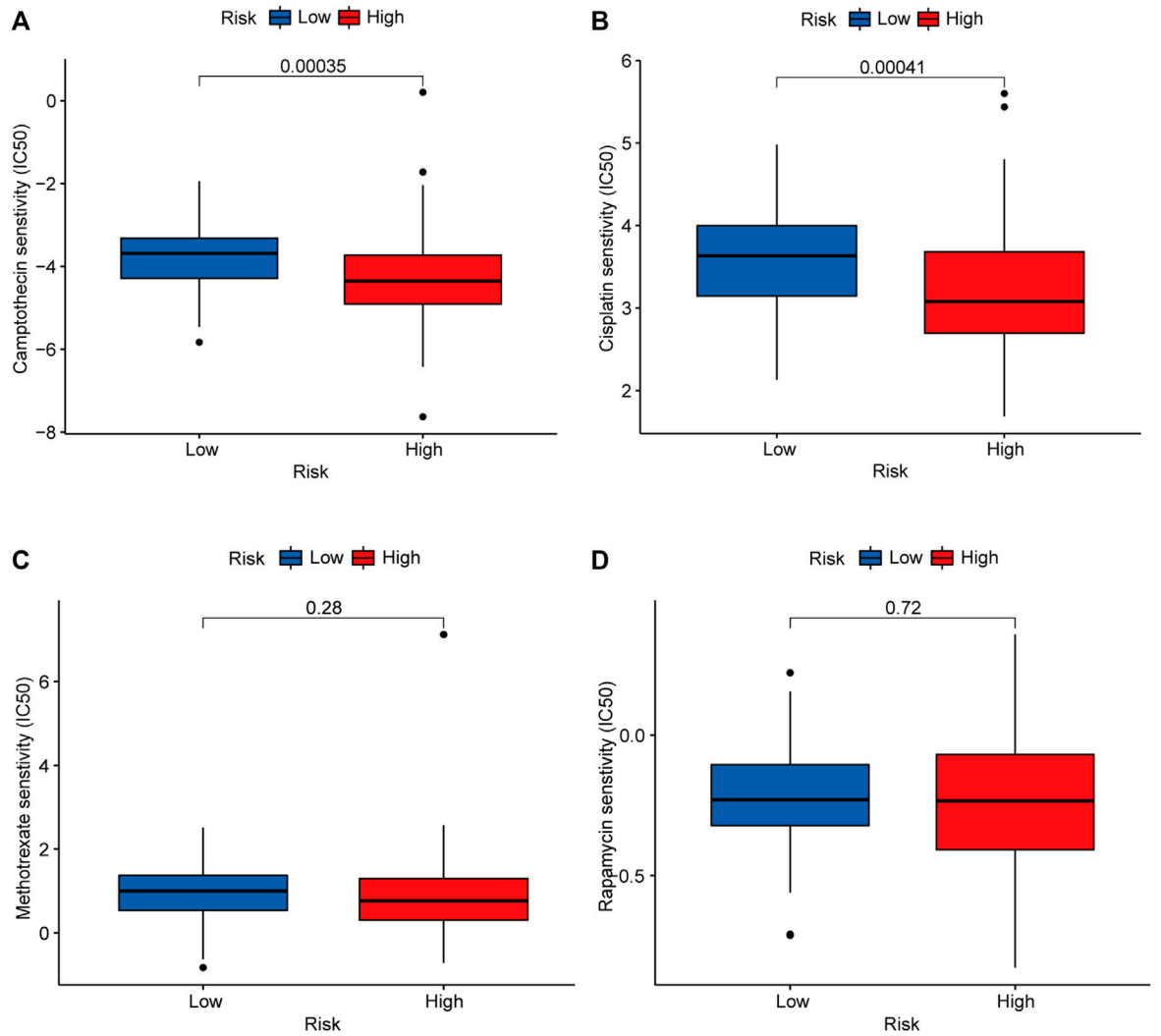


FIGURE 9 | Candidate chemotherapeutic drugs targeting the m6A-related lncRNA signature. IC50s of camptothecin **(A)**, cisplatin **(B)**, methotrexate **(C)**, and rapamycin **(D)** between the low- (marked in blue) and high-risk groups (marked in red). Abbreviation: m6A-related lncRNAs, N6-methyladenosine-related long non-coding RNAs.

2013), camptithecin can improve patients with early-stage CRC prognosis by inhibiting PKIP phosphorylation and STAT3 activation. Cisplatin is one of the main chemotherapy drugs used by oncologists to treat CRC (AlShamaileh et al., 2017). It has antitumor effects by binding to DNA and inducing DNA damage (Huang et al., 2019). More clinical trials are needed to explore the applicability of these two chemotherapeutic agents in patients with early-stage CRC.

However, our study had several limitations. First, this study was based on the TCGA database, and there was no external data set for verification. When searching for other data sets, such as the Gene Expression Omnibus, we were unable to locate one with both clinical information and the corresponding lncRNA and mRNA expression data. Future studies should use other data sets to verify the performance of our prognostic signature. Second, our findings must be verified *in vivo* and *in vitro*. Lastly, we did not explore the underlying mechanism of the studied m6A-related lncRNAs. Therefore, functional studies should examine these five lncRNAs individually and in combination to further verify the prediction accuracy of the signature and discover potential regulatory mechanisms. Despite these limitations, to the best of our knowledge, this is the first study to report an established external verification of m6A-related lncRNA signatures for early-stage CRC.

CONCLUSION

In summary, our analysis of seven m6A-related lncRNAs associated with prognosis identified two molecular subgroups of early-stage CRC with unique immune features. Subsequent analyses demonstrated the usefulness of a five m6A-related

lncRNA signature as a potential indicator of prognosis in patients with early-stage CRC. Further studies are required to explore the mechanisms underlying the association between the significant m6A-related lncRNAs and CRC prognosis and to verify the applicability of our prognostic signature.

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 the Sixth Affiliated Hospital of Sun Yat-Sen University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZX analyzed the data and conducted the experiments, XL designed the study, and SY wrote the articles. MX conducted the experiments and critically revised the articles. CM, FZ, HC, and LJ analyzed the data and critically revised the articles. LL participated in the conception of the study, designed the study, and critically revised the articles All authors contributed to the articles and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (grant nos. 81770557 and 82070684) and the Guangdong Natural Science Fund for Outstanding Youth Scholars (grant no. 2020B151502067).

ACKNOWLEDGMENTS

We thank all staff members of the Department of Colorectal Surgery at the Sixth Affiliated Hospital of Sun Yat-sen University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.724889/full#supplementary-material

REFERENCES

AlShamaileh, H., Wang, T., Xiang, D., Yin, W., Tran, P. H.-L., Barrero, R. A., et al. (2017). Aptamer-mediated Survivin RNAi Enables 5-fluorouracil to Eliminate Colorectal Cancer Stem Cells. *Sci. Rep.* 7 (1), 5898. doi:10.1038/s41598-017-05859-z

Arguello, A. E., DeLiberto, A. N., and Kleiner, R. E. (2017). RNA Chemical Proteomics Reveals the N6-Methyladenosine (m6A)-Regulated Protein-RNA Interactome. *J. Am. Chem. Soc.* 139 (48), 17249–17252. doi:10.1021/jacs.7b09213

Ban, Y., Tan, P., Cai, J., Li, J., Hu, M., Zhou, Y., et al. (2020). LNCAROD Is Stabilized by m6A Methylation and Promotes Cancer Progression via Forming a Ternary Complex with HSPA1A and YBX1 in Head and Neck Squamous Cell Carcinoma. *Mol. Oncol.* 14 (6), 1282–1296. doi:10.1002/1878-0261.12676

Cai, X., Wang, X., Cao, C., Gao, Y., Zhang, S., Yang, Z., et al. (2018). HBXIP-elevated Methyltransferase METTL3 Promotes the Progression of Breast Cancer via Inhibiting Tumor Suppressor Let-7g. *Cancer Lett.* 415, 11–19. doi:10.1016/j.canlet.2017.11.018

Colangelo, T., Polcaro, G., Muccillo, L., D'Agostino, G., Rosato, V., Ziccardi, P., et al. (2017). Friend or Foe? *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1867, 1–18. doi:10.1016/j.bbcan.2016.11.001

Cross-Knorr, S., Lu, S., Perez, K., Guevara, S., Brilliant, K., Pisano, C., et al. (2013). RKIP Phosphorylation and STAT3 Activation Is Inhibited by Oxaliplatin and Camptothecin and Are Associated with Poor Prognosis in Stage II colon Cancer Patients. *BMC Cancer* 13, 463. doi:10.1186/1471-2407-13-463

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). m 6 A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cel Rep.* 18 (11), 2622–2634. doi:10.1016/j.celrep.2017.02.059

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci.* 71 (10), 3971–3975. doi:10.1073/pnas.71.10.3971

Dienstmann, R., Salazar, R., and Tabernero, J. (2015). Personalizing colon Cancer Adjuvant Therapy: Selecting Optimal Treatments for Individual Patients. *Jco* 33 (16), 1787–1796. doi:10.1200/JCO.2014.60.0213

Dotan, E., and Cohen, S. J. (2011). Challenges in the Management of Stage II colon Cancer. *Semin. Oncol.* 38 (4), 511–520. doi:10.1053/j.seminoncol.2011.05.005

Edin, S., Kaprio, T., Hagström, J., Larsson, P., Mustonen, H., Böckelman, C., et al. (2019). The Prognostic Importance of CD20+ B Lymphocytes in Colorectal Cancer and the Relation to Other Immune Cell Subsets. *Sci. Rep.* 9 (1), 19997. doi:10.1038/s41598-019-56441-8

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 (6), 361–375. doi:10.1038/s41575-019-0126-x

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

Hsiang, Y. H., Hertzberg, R., Hecht, S., and Liu, L. F. (1985). Camptothecin Induces Protein-Linked DNA Breaks via Mammalian DNA Topoisomerase I. *J. Biol. Chem.* 260 (27), 14873–14878. doi:10.1016/s0021-9258(17)38654-4

Huang, L., Zhu, J., Kong, W., Li, P., and Zhu, S. (2021). Expression and Prognostic Characteristics of m6A RNA Methylation Regulators in Colon Cancer. *Ijms* 22 (4), 2134. doi:10.3390/ijms22042134

Huang, S., Lyu, S., Gao, Z., Zha, W., Wang, P., Shan, Y., et al. (2021). m6A-Related lncRNAs Are Potential Biomarkers for the Prognosis of Metastatic Skin Cutaneous Melanoma. *Front. Mol. Biosci.* 8, 687760. doi:10.3389/fmolb.2021.687760

Huang, Z., Zhang, Y., Li, H., Zhou, Y., Zhang, Q., Chen, R., et al. (2019). Vitamin D Promotes the Cisplatin Sensitivity of Oral Squamous Cell Carcinoma by Inhibiting LCN2-Modulated NF-Kb Pathway Activation through RPS3. *Cell Death Dis* 10 (12), 936. doi:10.1038/s41419-019-2177-x

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to Build and Interpret a Nomogram for Cancer Prognosis. *Jco* 26 (8), 1364–1370. doi:10.1200/JCO.2007.12.9791

Jin, Y., Wang, Z., He, D., Zhu, Y., Hu, X., Gong, L., et al. (2021). Analysis of m6A-Related Signatures in the Tumor Immune Microenvironment and Identification of Clinical Prognostic Regulators in Adrenocortical Carcinoma. *Front. Immunol.* 12, 637933. doi:10.3389/fimmu.2021.637933

Kannarkatt, J., Joseph, J., Kurniali, P. C., Al-Janadi, A., and Hrinczenko, B. (2017). Adjuvant Chemotherapy for Stage II Colon Cancer: A Clinical Dilemma. *Jop* 13 (4), 233–241. doi:10.1200/JOP.2016.017210

Kopetz, S., Tabernero, J., Rosenberg, R., Jiang, Z. Q., Moreno, V., Bachleitner-Hofmann, T., et al. (2015). Genomic Classifier ColoPrint Predicts Recurrence in Stage II Colorectal Cancer Patients More Accurately Than Clinical Factors. *The Oncologist* 20 (2), 127–133. doi:10.1634/theoncologist.2014-0325

Li, 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. USA* 117 (33), 20159–20170. doi:10.1073/pnas.1918986117

Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R. I. (2016). The M 6 A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cel* 62 (3), 335–345. doi:10.1016/j.molcel.2016.03.021

Liu, S., Lin, H., Wang, D., Li, Q., Luo, H., Li, G., et al. (2019). PCDH17 Increases the Sensitivity of Colorectal Cancer to 5-fluorouracil Treatment by Inducing Apoptosis and Autophagic Cell Death. *Sig Transduct Target. Ther.* 4, 53. doi:10.1038/s41392-019-0087-0

Manfredi, S., Bouvier, A. M., Lepage, C., Hatem, C., Dancourt, V., and Faivre, J. (2006). Incidence and Patterns of Recurrence after Resection for Cure of Colonic Cancer in a Well Defined Population. *Br. J. Surg.* 93 (9), 1115–1122. doi:10.1002/bjs.5349

Mantovani, A., Marchesi, F., Malesci, A., Laghi, L., and Allavena, P. (2017). Tumour-associated Macrophages as Treatment Targets in Oncology. *Nat. Rev. Clin. Oncol.* 14 (7), 399–416. doi:10.1038/nrclinonc.2016.217

Molinie, B., and Giallourakis, C. C. (2017). Genome-Wide Location Analyses of N6-Methyladenosine Modifications (m6A-Seq). *Methods Mol. Biol.* 1562, 45–53. doi:10.1007/978-1-4939-6807-7_4

Mollica, A., Stefanucci, A., Feliciani, F., Cacciatore, I., Cornacchia, C., and Pinnen, F. (2012). Delivery Methods of Camptothecin and its Hydrosoluble Analogue Irinotecan for Treatment of Colorectal Cancer. *Cdd* 9 (2), 122–131. doi:10.2174/156720112800234558

Sautès-Fridman, C., Lawand, M., Giraldo, N. A., Kaplon, H., Germain, C., Fridman, W. H., et al. (2016). Tertiary Lymphoid Structures in Cancers: Prognostic Value, Regulation, and Manipulation for Therapeutic Intervention. *Front. Immunol.* 7, 407. doi:10.3389/fimmu.2016.00407

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

Slavin, M. L. (1989). Metastatic Malignant Meningioma. *J. Clin. Neuroophthalmol* 9 (1), 55–59. doi:10.3109/01658108909019508

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

Tu, Z., Wu, L., Wang, P., Hu, Q., Tao, C., Li, K., et al. (2020). N6-Methylandenosine-Related lncRNAs Are Potential Biomarkers for Predicting the Overall Survival of Lower-Grade Glioma Patients. *Front. Cel Dev. Biol.* 8, 642. doi:10.3389/fcell.2020.00642

Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., et al. (2017). The N6-Methyladenosine (m6A)-Forming Enzyme METTL3 Controls Myeloid Differentiation of normal Hematopoietic and Leukemia Cells. *Nat. Med.* 23 (11), 1369–1376. doi:10.1038/nm.4416

Walker, B. S., Zarour, L. R., Wieghard, N., Gallagher, A. C., Swain, J. R., Weinmann, S., et al. (2020). Stem Cell Marker Expression in Early Stage Colorectal Cancer Is Associated with Recurrent Intestinal Neoplasia. *World J. Surg.* 44 (10), 3501–3509. doi:10.1007/s00268-020-05586-z

Wang, H., Meng, Q., and Ma, B. (2021). Characterization of the Prognostic m6A-Related lncRNA Signature in Gastric Cancer. *Front. Oncol.* 11, 630260. doi:10.3389/fonc.2021.630260

Wang, W., Li, J., Lin, F., Guo, J., and Zhao, J. (2021). Identification of N6-Methyladenosine-Related lncRNAs for Patients with Primary Glioblastoma. *Neurosurg. Rev.* 44 (1), 463–470. doi:10.1007/s10143-020-01238-x

Wen, S., Wei, Y., Zen, C., Xiong, W., Niu, Y., and Zhao, Y. (2020). Long Non-coding RNA NEAT1 Promotes Bone Metastasis of Prostate Cancer through N6-Methyladenosine. *Mol. Cancer* 19 (1), 171. doi:10.1186/s12943-020-01293-4

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

Yang, M., McKay, D., Pollard, J. W., and Lewis, C. E. (2018). Diverse Functions of Macrophages in Different Tumor Microenvironments. *Cancer Res.* 78 (19), 5492–5503. doi:10.1158/0008-5472.CAN-18-1367

Yang, S., Wei, J., Cui, Y.-H., Park, G., Shah, P., Deng, Y., et al. (2019). m6A mRNA Demethylase FTO Regulates Melanoma Tumorigenicity and Response to Anti-PD-1 Blockade. *Nat. Commun.* 10 (1), 2782. doi:10.1038/s41467-019-10669-0

Yang, Y., Hsu, P. J., Chen, Y.-S., and Yang, Y.-G. (2018). Dynamic Transcriptomic m6A Decoration: Writers, Erasers, Readers and Functions in RNA Metabolism. *Cell Res* 28 (6), 616–624. doi:10.1038/s41422-018-0040-8

Zhang, Q., Cai, Y., Kurbatov, V., Khan, S. A., Lu, L., Zhang, Y., et al. (2020). Gene Alterations of N6-Methyladenosine (m6A) Regulators in Colorectal Cancer: A TCGA Database Study. *Biomed. Res. Int.* 2020, 1–13. doi:10.1155/2020/8826456

Zuo, L., Su, H., Zhang, Q., Wu, W.-Y., Zeng, Y., Li, X.-M., et al. (2021). Comprehensive Analysis of lncRNAs N6-Methyladenosine Modification in Colorectal Cancer. *Aging* 13 (3), 4182–4198. doi:10.18632/aging.202383

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 Xiong, Li, Yin, Xie, Mao, Zhang, Chen, Jin and Lian. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fmolb.2021.714203

SYDE1 Acts as an Oncogene in Glioma and has Diagnostic and Prognostic Values

Zhenyuan Han[1,2,3†], Xiaodong Zhuang[4†], Biao Yang[5†], Lihui Jin[6], Pengjie Hong[5], Junqing Xue[2], Shunjie Chen[7] and Zhen Tian[2]**

[1]Department of Oral and Maxillofacial Surgery, Peking University School and Hospital of Stomatology, Beijing, China, [2]Department of Oral Pathology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3]National Clinical Research Center for Oral Diseases, Shanghai, China, [4]Department of Stomatology, Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China, [5]Department of Neurosurgery, Huashan Hospital of Fudan University, Shanghai, China, [6]Department of Pediatric Cardiology, Xinhua Hospital, Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China, [7]Department of Nephrology, Shanghai Fourth People's Hospital Affiliated to Tongji Universtiy, Shanghai, China

Objectives: Gliomas remain one of serious public health problems worldwide which demand further and deeper investigation. The aim of this study was to explore the association between synapse defective protein 1 homolog 1 (*SYDE1*) and gliomas *via* public database analysis and *in vitro* validation to determine the potential diagnostic and prognostic values.

Methods and Results: Compared with healthy brain tissues, there was a significant increase in *SYDE1* expression in glioma tissues. Additionally, *SYDE1* exhibited higher expression levels in glioma patients with unfavorable clinicopathological factors. *In vitro* knockdown of *SYDE1* in glioma cell lines A172 inhibited their migrative and invasive ability but not the proliferative ability. GO and KEGG pathway analysis of the top 100 genes coexpressed with *SYDE1* showed enrichments of tumor-associated terms. Further bioinformatic analysis revealed that the *SNHG16/hsa-miR-520e/SYDE1* axis might be involved in glioma development.

Conclusions: *SYDE1* is expressed at higher levels in gliomas than in healthy brains, and can promote metastasis and invasion but not proliferation of gliomas. Furthermore, *SYDE1* has values in the diagnosis and prognosis prediction of gliomas.

Keywords: SYDE1, glioma, unfavorable clinicopathological factors, prognosis, diagnosis

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong, SAR China

Reviewed by:
Zijun Wang, The Rockefeller University, United States
Huaide Qiu, Nanjing Medical University, China
Cheng Wu, Huazhong University of Science and Technology, China
Xing Fan, Capital Medical University, China

****Correspondence:***
Zhen Tian
tianzhen_256@126.com
Shunjie Chen
Chenshunjie77csj@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *24 May 2021*
Accepted: *28 September 2021*
Published: *14 October 2021*

Citation:
Han Z, Zhuang X, Yang B, Jin L, Hong P, Xue J, Chen S and Tian Z (2021) SYDE1 *Acts as an Oncogene in Glioma and has Diagnostic and Prognostic Values. Front. Mol. Biosci. 8:714203. doi: 10.3389/fmolb.2021.714203*

INTRODUCTION

Gliomas are one of the most common invasive malignancies in the central nervous system, accounting for approximately 51.4% of all primary brain tumors (Xue et al., 2017). In particular, gliomas constitute the majority of primary brain tumors in adults (Goodenberger and Jenkins, 2012). The current therapeutic approach for gliomas, however, is limited to complete surgical resection combined with radiotherapy and alkylating chemotherapy (Louis et al., 2016). Of note, since oncogenes are capable of sustaining tumor growth and conventional treatment does not take into account special parameters of various subtypes, the expectation of novel treatments is still grave for the general majority of patients (Arko et al., 2010; Gilbert et al., 2013; Kamiya-Matsuoka and Gilbert, 2015).

According to the 2007 WHO classification of central nervous system tumors, gliomas can be divided into low-grade glioma (LGG, grade I and II) and high-grade glioma (HGG, grade III and IV) (Louis et al., 2007). This classification is mainly based on the clinical and histopathological features of the glioma. The latest version of the WHO glioma category released in 2016 has included the evaluation of molecular characteristics, such as isocitrate dehydrogenase (IDH) mutations, chromosome 1p/19q codeletion and RELA fusion positivity (Louis et al., 2016). IDH enzymes catalyze the oxidative decarboxylation of isocitrate and are essential in maintaining cellular homoeostasis. IDH mutations have been reported by Yan et al. to occur in >80% of HGG cases (Yan et al., 2009). Chromosome 1p/19q codeletion is recognized as another hallmark of gliomas, which presents in approximately 60–90% of oligodendrogliomas and 30–50% of oligoastrocytomas (Hofer and Lassman, 2010). Of note, 1p/19q codeletion is associated with a better response to radiotherapy and chemotherapy and can predict longer progression-free and overall survival (Dunbar, 2009). Given that IDH mutations and 1p/19q codeletions occur mostly in different subtypes of gliomas, the 2016 edition of the WHO glioma classification faces great challenges in clinical practice. Therefore, it is of great necessity to acquire an in-depth understanding of gliomas and identify other causal genes.

Synapse defective protein 1 homolog 1 (*SYDE1*) is a 79-kDa Rho GTPase-activating protein that is encoded by the *SYDE1* gene located at chromosome 19p13.12. *SYDE1* is highly expressed in the placenta, bone marrow and brain and is involved in the positive regulation of placental trophoblast cell migration and bone marrow cell differentiation (Lo et al., 2017). Moreover, *SYDE1* has been revealed to be differentially expressed between cervical cancer and normal controls and is recognized as a potential causal gene related to cervical cancer (Zhang et al., 2021). Buchner et al. deciphered the role of *SYDE1* in renal cell carcinoma (RCC) via analysis of metastatic RCC mRNA expression profiles, which revealed that *SYDE1* could discriminate well between RCC patients with favorable prognosis or poor prognosis (Buchner et al., 2010). Given the high expression level of *SYDE1* in brain tissues, it is of great necessity to determine the function of *SYDE1* in the pathogenesis of cerebral tumors, especially gliomas.

In this study, we utilized publicly available data from the Oncomine, GEPIA2 and Human Protein Atlas (HPA) databases to examine SYDE1 expression in glioma tissues and normal control tissues, and we identified an increased level of SYDE1 in gliomas. Furthermore, *SYDE1* expression was higher in HGGs than in LGGs, as revealed by bioinformatic analysis of the Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA) and Chinese Glioma Genome Atlas (CGGA) datasets. Immunohistochemistry (IHC) staining of SYDE1 human glioma samples from grades I to IV draws the same conclusion. Then, we performed RNAi-mediated *SYDE1* knockdown in glioma cell lines *in vitro*, which revealed that *SYDE1* knockdown abolished the migration and invasion of glioma cells. To decipher the mechanisms of *SYDE1* in glioma pathogenesis, we further constructed a coexpression network of genes in gliomas from the cBioPortal database whose Spearman's correlation index values with *SYDE1* rank in the top 100 and found that these genes were enriched in tumor-associated GO terms.

MATERIALS AND METHODS

Application of Public Web-Based Databases

Gene expression profiling interactive analysis 2 (GEPIA2; http://gepia.cancer-pku.cn/index.html) is a Chinese online tool for bioinformatics and visualized analysis based on the TCGA and GTEx databases. It is widely used because of its convenient operation and high efficiency although users do not have access to the primary data (Tang et al., 2019). The present study used survival analysis of *SYDE1*, hsa-miR-136, hsa-miR-302a, hsa-miR-424, hsa-miR-497, hsa-miR-529e, FGD5-AS1, MIR17HG, and SNHG16 to explore the relationships between *SYDE1* expression and glioma prognosis and pathology.

To explore the expression level of *SYDE1* in CNS/brain tumors (especially gliomas) and normal brain tissues, Oncomine (https://www.oncomine.org) was utilized to analyze the web-based data on the expression of *SYDE1* in different types of gliomas and corresponding normal samples (Rhodes et al., 2004). This study compared the pattern of the expression of *SYDE1* in four major types of gliomas, including GBMs, anaplastic astrocytomas, diffuse astrocytomas, and oligodendrogliomas, with the inclusion threshold described as a fold-change (FC) > 1.5, $p < 0.01$, and gene rank = all.

The HPA (https://www.proteinatlas.org) is a widely used online protein analysis database that contains 26,000 human proteins, and it was used to develop immunoassay technologies to analyze the expression levels of proteins in cell lines, human normal tissues and tumor tissues (Uhlén et al., 2015). The present study assessed the translational expression levels of *SYDE1* in gliomas and normal brain tissues.

The top 100 coexpressed genes of *SYDE1* were acquired from the cBioPortal database (http://www.cbioportal.org/), and were visualized *via* Cytoscape (https://cytoscape.org/) (Cerami et al., 2012; Gao et al., 2013). The mRNA/miRNA/lncRNA interactions were forecasted using starBase 3.0 (http://starbase.sysu.edu.cn/), which includes seven supported algorithms (PITA, miRmap, microT, miRanda, PicTar, RNA22, and TargetScan) (Li et al., 2014). Target miRNAs predicted by at least six algorithms were selected for further functional analysis.

Data Obtaining and Preprocessing

To evaluate the relations between *SYDE1* expression and cancer prognosis and other clinicopathological characteristics, *SYDE1* mRNA expression profiles in glioma samples and normal samples were downloaded from TCGA, CGGA and GEO. These mRNA profiles were quality controlled1 using RSeQC, and were normalized using the trimmed mean of M-values. Within the TCGA database, we subgrouped the glioma samples into three major cohorts, glioma, LGG, and GBM samples, which were

TABLE 1 | Clinicopathological characteristics of patient samples and expression of *SYDE1* in glioma and normal tissues.

Characteristics	Category	Number of cases (%)
Age (years)	< 45	20 (44.4%)
	≥45	25 (55.6%)
Gender	Female	24 (53.3%)
	Male	21 (46.7%)
WHO grade	Normal	5 (11.1%)
	I	5 (11.1%)
	II	10 (22.2%)
	III	10 (22.2%)
	IV	15 (33.3%)
Histology	Normal	5 (11.1%)
	Pilocytic astrocytoma	5 (11.1%)
	Oligoastrocytoma	3 (6.7%)
	Astrocytoma	7 (15.6%)
	Anaplastic astrocytoma	9 (20.0%)
	Anaplastic oligodendroglioma	1 (2.2%)
	Glioblastoma	15 (33.3%)
Expression of *SYDE1*	Low expression	22 (48.8%)
	High expression	23 (51.1%)

named TCGA_glioma, TCGA_LGG, and TCGA_GBM, respectively. For the Chinese cohorts, *SYDE1* mRNA expression and its corresponding clinicopathological characteristics were downloaded from CGGA (http://www.cgga.org.cn/), and these cohorts included three datasets, mRNA-array_301, mRNAseq_325, and mRNAseq_693 datasets, with 301, 325, and 693 glioma tissue samples, respectively. Several microarray datasets from the GEO database were also selected for our present study, including GSE4271 (generated from GPL96), GSE4290 (generated from GPL570), GSE4412 (generated from GPL96), GSE68848 (generated from GPL570) and GSE13041 (generated from GPL96, GPL570 or GPL8300). The prognostic values between miRNAs and lncRNAs were assessed by CGGA microRNA-array_198 and CGGA mRNA-seq 325, respectively.

Gene Set Enrichment Analysis

698 glioma cases were divided into two expression level groups based on the median expression value of *SYDE1*. GSEA was then conducted to identify hallmark gene sets that were enriched in the gene rank in the two groups. The h.all.v7.4.symbols.gmt in the Molecular Signatures Database (MSigDB) was selected in GSEA version 4.1 to annotate gene sets. The cutoff criteria were set to nominal $p < 0.05$, normalized enrichment scores (NES) > 1.0 and false discovery rate (FDR) q > 0.25. Finally, hallmark gene sets with significant enrichment were chosen, and gene set enrichment plots were made.

Tissue Samples

This study included 40 human glioma samples (grade I: 5, grade II: 10, grade III: 10, and grade IV: 15) and five normal brain samples. Specifically, human glioma samples were paraffin-embedded and retrieved from Shanghai Ninth People's Hospital. Normal brain samples were obtained from patients diagnosed with cerebral trauma surgery but without other brain diseases, which were collected during surgery after informed consent was obtained from patients who needed brain trauma surgery. The clinicopathological characteristics of 45 glioma and control cases is summarized in **Table 1**. This study was approved by the ethical committee of Shanghai Ninth People's Hospital.

All mouse strains used in this study were on the C57BL/6 background and were approved by the ethical committee of Shanghai Ninth People's Hospital. Cerebrums were dissected and dissolved in TRIzol for RNA extraction.

Immunohistochemistry

Immunohistochemistry (IHC) was performed on paraffin-embedded human glioma and normal brain tissues collected from Shanghai Ninth People's Hospital. The sections were deparaffinized in a xylene gradient and rehydrated in an ethanol gradient. Antigen retrieval was performed in sodium citrate buffer (10 mM sodium citrate pH 6.0) at 100 C for 20 min. Then, endogenous peroxidase was deactivated by applying 3% H_2O_2 in methanol. IHC of SYDE1 was performed by the Dako EnvisionTM method. Briefly, the sections (3 μm) were sequentially incubated with the anti-SYDE1 antibody (NBP1-89350, Novus Biologicals) and the HRP-conjugated secondary antibody (ab6721, Abcam). Then, the sections were color-developed with a DAB Immunohistochemistry Color Development Kit (E670033, Sangon Biotech) and counterstained with hematoxylin.

A Leica LF200 microscope was used to image the stained sections. The intensity of SYDE1 signals was scored as negative (0), weak (1) or strong (2). The staining extent of SYDE1 was evaluated according to the immunoreactive tumor cell percentage, which was scored as I (0%, score = 0), II (1–5%, score = 1), III (6–25%, score = 2), IV (26–75%, score = 3) and V (76–100%, score = 4). SYDE1 signals (ranging from 0 to 8) were calculated by multiplying the intensity score with the staining extent score and were classified into low (0–4) or high (5–8) groups for Fisher's exact tests.

Cell Culture

Human astroglial A172 cells and mouse glioma GL261 cells were gifts from the Shanghai Institutes for Biological Sciences. All cell lines were confirmed to be free of microorganism contamination. Glioma cell lines were grown in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS) and 1% (100×) streptomycin/penicillin.

RNA Extraction and qPCR

Total RNA was extracted from cerebrums or cultured glioma cell lines using TRIzol and then reverse transcribed into cDNA using a Superscript II reverse transcriptase Kit. qPCR was carried out using Power SYBR Green PCR Master Mix on the Applied Biosystems 7900HT Fast Real-Time PCR System. qPCR primers used to detect the expression levels of *SYDE1* mRNA in human or mouse tissues were as follows: human forward 5′-CATCATCCAGAAGTGCGTTG-3′ and reverse: 5′-AATCCT TGAGGATGCCAGTG-3′; mouse forward: 5′-CCTACCAAA ACCTCCCGTACC-3′ and reverse: 5′-GGGGCGGTCCTCTCT CTATC-3′. The relative expression of each gene was normalized to ACTB and calculated using the $2^{-\Delta\Delta CT}$ method.

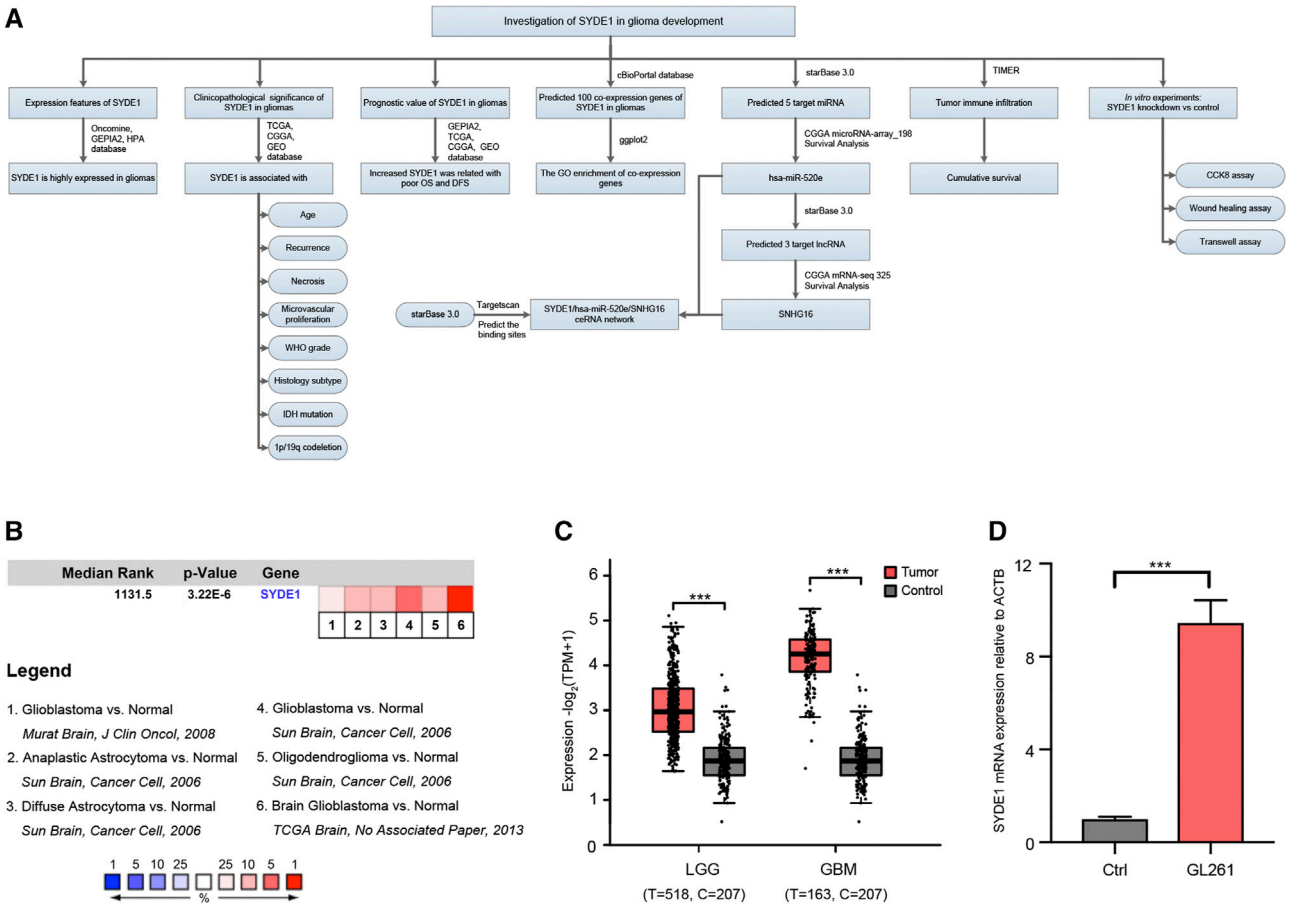


FIGURE 1 | The comparison of *SYDE1* expression between normal brain and glioma tissues. **(A)** A schematic overview of different steps taken to examine the association between *SYDE1* and glioma. **(B)** *SYDE1* expression in normal human brains and different glioma subtypes. The rank for a gene is the median rank for that gene across each of the analysis and the p.value for a gene is its p.value for the median-ranked analysis. **(C)** *SYDE1* expression in normal human brains, LGGs and GBMs. **(D)** *SYDE1* expression in normal mouse brains and mouse glioma GL261 cells. ***, p.value <0.001; Ctrl, normal C57BL6 mouse brain.

Transfection of siRNA

The siRNA targeting human *SYDE1* (siSYDE1) and the non-targeting control siRNA (siNC) were obtained from Shanghai Zorin Biological Technology. The sequence of siSYDE1 was 5′-UAGUGGGACUGUACCGUCUUUdTdT-3′ (sense) and 5′-AAAGACGGUACAGUCCCACUAdTdT-3′ (antisense). The sequence of siNC was 5′-UUCUCCGAACGUGUCACGUdTdT-3′ (sense) and 5′-ACGUGACACGUUCGGAGAAdTdT-3′ (antisense). LipoRNAiMAX transfection reagent was applied to deliver the siSYDE1 or siNC into glioma cell line A172. Transfection was performed according to the manufacturer's instructions when the cell monolayer reached 70–80% confluency.

Cell Proliferation Assay, Wound Scratch Assay and Transwell Assay

A172 cell lines were used for cell proliferation, wound scratch and Transwell assays. SYDE1-knockdown and control cells were seeded into 96-well plates (1,500 cells/well) for the cell proliferation assay. Cell Counting Kit-8 (CCK-8; Beyotime, C0038) was used to measure cell proliferation according to the manufacturer's instructions. Briefly, 10 μl of CCK-8 reagent was added to each well at the indicated detection times. The plates were incubated at 37°C for 1 h and then measured in a microplate reader at a wavelength of 450 nm to determine the optical density.

For the wound scratch assay, SYDE1-knockdown and control cells were placed into 6-well plates (1,500 cells/well). After confluence, a sterile pipette tip was used to scratch a straight line in the glioma monolayer cell. Images were recorded at 0 and 48 h after scratching, and ImageJ software was used to measure the scratch area.

Transwell assays were carried out using a 6-well culture insert (GIBCO, 140,640) according to the manufacturer's instructions. SYDE1-knockdown and control cells were seeded in each well (20,000 cells/well). After culturing at 37°C for 24 h, the culture insert was removed. Photographs were recorded at 0 and 48 h after culture-insert removal to investigate and analyze the healed wound area.

Statistical Analysis

Statistical details are available in the figure legends. Unpaired/paired Student's t-test (two-tailed) was used to analyze differences between two groups. One-way ANOVA was applied to analyze differences among three or more groups. Fisher's exact test was used for the IHC quantification of *SYDE1* expression in glioma patient samples and normal brain tissues. All data are presented as the mean ± standard deviation (SD) if not further implicated and considered statistically significant with a p value <0.05. All analyses were conducted in GraphPad Prism 7 or Microsoft Excel.

For glioma survival analysis, Kaplan-Meier analysis and Cox regression analysis were combined to analyze the prognostic value of SYDE1. Kaplan–Meier survival plots were used to evaluate patient OS and DFS. Log-rank tests were used to determine significant differences between two groups. The overall design of our study is presented in the flow chart in **Figure 1A**.

RESULTS

The Expression of *SYDE1* Significantly Correlated With the Clinical Features of Glioma Tissues

To explore expression levels of *SYDE1* in gliomas and their relation to clinical features and molecular subtypes, *SYDE1* expression levels in pairs of gliomas and adjacent normal samples were compared using the Oncomine and GEPIA2 databases. We identified four analyses of upregulated *SYDE1* between brain or CNS cancer and adjacent normal tissues in the Oncomine database that met the established threshold for $|\log_2 FC| > 1.5$, $p < 0.01$ and gene rank = all (**Figure 1B**). The GEPIA2 databases indicated that *SYDE1* was expressed at higher levels in glioma subclasses LGG and GBMs than in the corresponding normal tissues ($p < 0.05$, **Figure 1C**). Both databases support the hypothesis that *SYDE1* is upregulated in CNS tumors, such as gliomas, versus normal tissues. On the other hand, *SYDE1* expression is higher in mouse glioma GL261 cells than normal C57BL/6 brains (**Figure 1D**).

We assessed the relationship between the clinicopathological parameters and *SYDE1* expression. As shown in **Supplementary Tables S1–S6**, *SYDE1* expression significantly correlated with age in CGGA mRNA-array_325, CGGA mRNA-array_693, GSE4271, and TCGA_glioma ($p < 0.05$). For the histological subclasses, *SYDE1* expression was highly associated with this category in the CGGA mRNA-array_301, CGGA mRNA-array_325, CGGA mRNA-array_693, TCGA_LGG, and TCGA_glioma databases ($p < 0.05$). The analytical statistics also suggested that *SYDE1* was involved in WHO grade ($p < 0.05$), and the updated WHO classification-related diagnostic molecular characteristics IDH mutation and 1p19q_codeletion were also associated with *SYDE1* expression in CGGA mRNA-array_325 and CGGA mRNA-array_693 ($p < 0.001$). Our data also revealed that *SYDE1* was significantly associated with the PRS type in CGGA mRNA-array_325 and CGGA mRNA-array_693 ($p < 0.05$) and chemotherapy and radiotherapy in CGGA mRNA-array_325 and CGGA mRNA-array_693 ($p < 0.001$). Notably, the *SYDE1* marker was also related to microvascular proliferation and necrosis in GSE4271 ($p < 0.05$). Overall, these findings confirm that *SYDE1* expression is associated with different clinical outcomes and previous diagnostic biomarkers, which may benefit the diagnosis and treatment of gliomas.

Increased *SYDE1* Expression Was Positively Related to Older Age, Recurrence, Necrosis, and Microvascular Proliferation in Gliomas

To ensure the functionality of *SYDE1* in glioma development and clinical parameters, the *SYDE1* gene expression levels were divided into two groups according to age, with a cutoff of 45 years. There was a distinguished difference between the two age groups with differential *SYDE1* expression. Notably, the expression of *SYDE1* in older patients aged ≥45 years was higher than that in patients aged <45 years in the CGGA mRNA-array_301 ($p < 0.05$, **Supplementary Figure S1A**), CGGA mRNA-array_325 ($p < 0.05$, **Supplementary Figure S1B**), GSE4271 ($p < 0.05$, **Supplementary Figure S1C**), and TCGA_glioma datasets ($p < 0.05$, **Figure 2A**). There was no significant difference between the glioma patient groups regarding sex or race. *SYDE1* was chosen for further analysis because it closely correlated with the clinical features and because there were a small number of reports of its involvement in tumorigenesis. We also noted that the expression levels of *SYDE1* were significantly upregulated in the recurrent groups compared to the primary groups in CGGA mRNA-array_325 ($p < 0.05$, **Supplementary Figure S1D**) and CGGA mRNA-array_693 ($p < 0.05$, **Figure 2B**). One of the GBM diagnostic criteria, microvascular proliferation, and necrosis, was also related to *SYDE1* expression. Higher microvascular proliferation and grade IV necrosis occurred with higher *SYDE1* expression in gliomas ($p < 0.05$, **Figures 2C,D**). We evaluated these findings in the GSE4271 dataset.

High Expression of *SYDE1* Indicated Higher WHO Grade and More Malignant Histological Subtypes in Gliomas

We first observed that *SYDE1* was highly expressed in gliomas, and an important, but unsolved, question is whether increased expression of *SYDE1* is related to increasing glioma grade. The mRNA expression patterns of *SYDE1* in different grades were further evaluated in the GEO, CGGA, and TCGA databases, which are based on clinical outcomes and gene expression. As shown in **Figure 3A** and **Supplementary Figures S2A–H**, the expression level of *SYDE1* also increased with increasing WHO grade ($p < 0.05$). The protein expression pattern of *SYDE1* in gliomas described by the Human Protein Atlas also helped account for this characteristic. As shown in **Figures 4A–C**, SYDE1 protein expression was moderate in LGG, and strong expression was detected in HGG tissues. This result is consistent with the mRNA microarrays. In the GSE4290 and GSE4412_GPL96 datasets, the expression level of *SYDE1* increased in the order of control of oligodendrogliomas, astrocytomas, and GBMs. The expression of *SYDE1* in GBMs was also highest in the other histologies in the GSE68848 dataset ($p < 0.05$, **Figure 3B**,

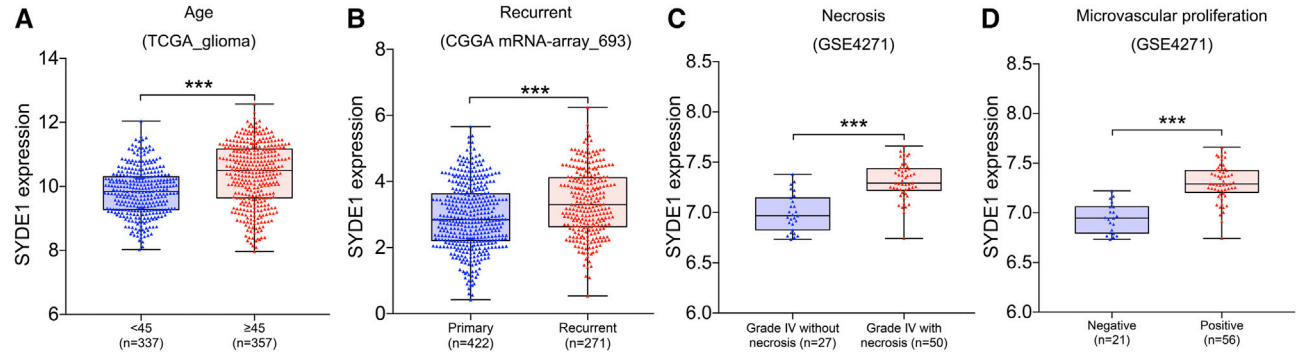


FIGURE 2 | The association between *SYDE1* expression and patient age, tumor recurrence, tumor necrosis or tumor microvascular proliferation. **(A)** The association between *SYDE1* expression and patient age. **(B)** The association between *SYDE1* expression and glioma recurrence. **(C)** The association between *SYDE1* expression and glioma necrosis. **(D)** The association between *SYDE1* expression and glioma microvascular proliferation. *, p.value <0.05; ***, p.value <0.001.

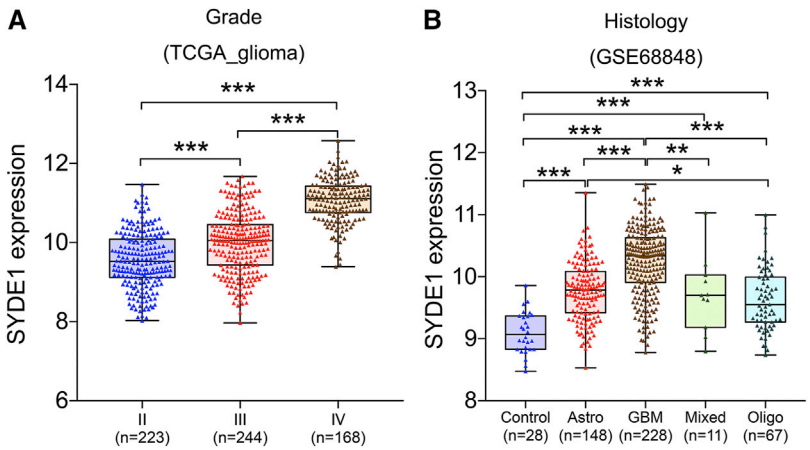


FIGURE 3 | The association between *SYDE1* expression and glioma grade or histology. **(A)** The association between *SYDE1* expression and glioma grades. **(B)** The association between *SYDE1* expression and glioma histology. *, p.value <0.05; **, p.value <0.01; ***, p.value <0.001.

Supplementary Figures S2I, SJ). Furthermore, we also performed IHC to investigate SYDE1 expression in glioma tissue. Intriguingly, our results also exhibited such characteristics (**Table 2**). As shown in **Figures 5A–F**, positive staining of SYDE1 was predominantly found in grade IV glioma compared with grade I glioma. Therefore, SYDE1 strongly correlated with WHO glioma grade at the mRNA and protein levels.

SYDE1 Expression Significantly Differentiated Glioma Subtypes and Was Statistically Associated With Molecular Genetic Features

Initially, *SYDE1* correlated with clinical parameters, and the expression levels of *SYDE1* were significantly associated with glioma WHO grade. Therefore, it is tempting to speculate that *SYDE1* expression correlates with glioma subtype. Several molecular biomarkers, including the IDH mutation and chromosome 1p/19q codeletion, are used for classification of gliomas. These biomarkers support the basis for individual glioma clinical therapy and molecular targeted therapy. Based on these markers, different molecular classifications were proposed. Therefore, we further analyzed the relationship between *SYDE1* expression and different classification subtypes. Our results showed that *SYDE1* was more highly expressed in Mes subtypes than neural, PN, and Prolif subtypes based on CGGA mRNA-array_301, GSE4271, GSE13041_GPL96, and TCGA_glioma datasets ($p < 0.05$, **Figures 6A–C,E**), but this association was not pronounced in GSE13041_GPL8300 (**Figure 6D**). Another part of the data (CGGA mRNA-array_301, GSE4271, GSE13041_GPL96, and TCGA_glioma dataset) suggested that *SYDE1* expression was significantly lower in the neural or PN subtypes ($p < 0.05$, **Figures 6A–C,E**). However, the ProMes subtypes had the lowest *SYDE1* expression of all subtypes in the GSE13041_GPL8300 dataset ($p < 0.05$, **Figure 6D**). We

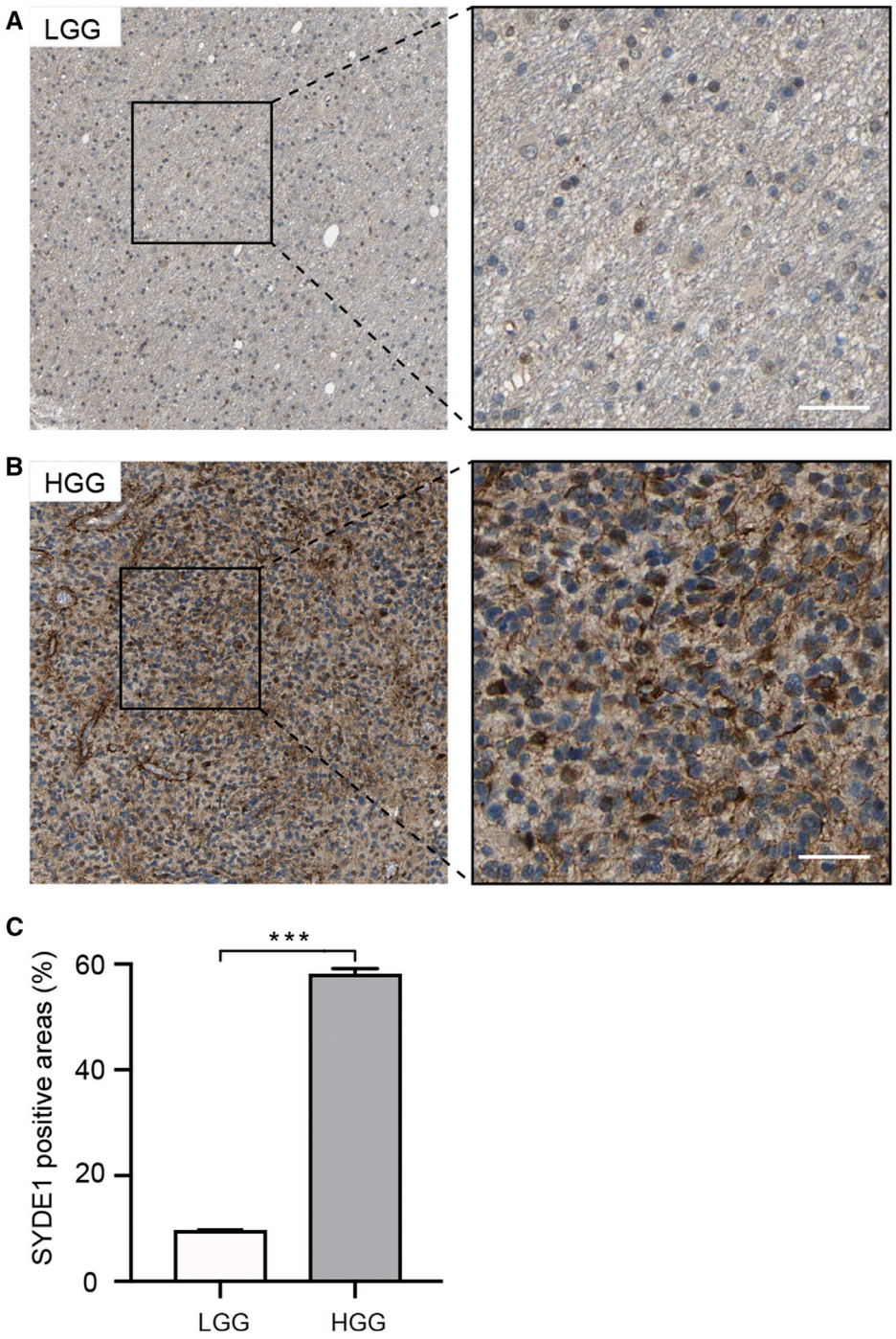


FIGURE 4 | Immunohistochemical analyses of SYDE1 expression in LGG and HGG (The HPA database). **(A)** Weak SYDE1 staining in LGG. **(B)** Strong SYDE1 staining in HGG. **(C)** Quantification of SYDE1 positive areas in LGG and HGG. ***, p.value <0.001; Scale bars = 50 μm.

also noted that *SYDE1* was differentially expressed in various subtypes between the CGGA, TCGA, and GEO databases ($p < 0.05$, **Figures 6A–E**).

Notably, *SYDE1* expression strongly correlated with favorable prognostic factors (i.e., IDH mutation and chromosome 1p/19q codeletion). The analyzed data (CGGA mRNA_301, CGGA mRNA_325, and CGGA mRNA_693) indicated that glioma patients with the IDH_mutation attribute exhibited lower *SYDE1* expression than patients with IDH_wildtype ($p < 0.001$, **Figures 6F–H**). Similarly, the 1p/19q codeletion demonstrated the same correlation, which means that *SYDE1* expression was decreased in the 1p/19q_codeletion glioma

TABLE 2 | Association between *SYDE1* expression and WHO grade in glioma.

Group	WHO grade					Fisher's exact test	*p*-value
	Normal	I	II	III	IV		
Low expression of *SYDE1*	4	4	7	6	3	10.615	0.021*
High expression of *SYDE1*	1	1	3	4	12		

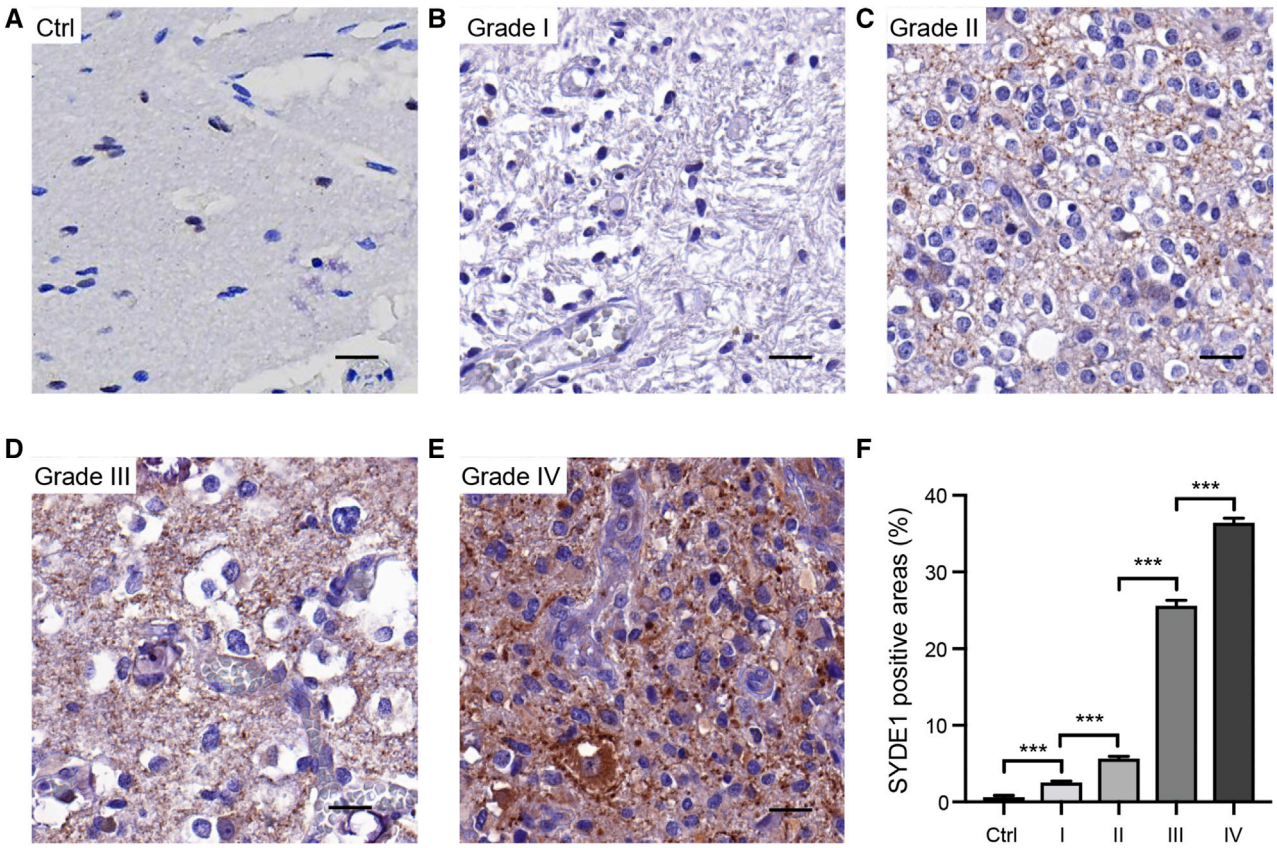


FIGURE 5 | Immunohistochemical analyses of SYDE1 expression in resection specimens of normal human brain and grade I to IV glioma. **(A)** Negative SYDE1 staining in normal human brain. **(B–C)** Weak SYDE1 staining in grade I glioma **(B)** and grade II glioma **(C)**. **(D–E)** Strong SYDE1 staining in grade III glioma **(D)** and grade IV glioma **(E)**. **(F)** Quantification of SYDE1 positive areas at A-E. ***, p.value <0.001; Scale bars = 25 μm.

samples compared with the corresponding 1p/19q_noncodeletion samples ($p < 0.05$, **Figures 6I–K**).

The Unfavorable Prognostic Role of *SYDE1* Expression in Glioma Patients

We investigated the prognostic ability of *SYDE1* for overall survival (OS) and disease-free survival (DFS) using the GEPIA2 online database. As expected, individuals with increased *SYDE1* in gliomas were associated with poor OS and DFS ($p < 0.0001$, **Figures 7A,B**). The TCGA, CGGA, and GEO cohorts also supported the result that high *SYDE1* expression in gliomas is associated with shorter OS time, and decreased levels of *SYDE1* expression are associated with longer OS time in glioma patients ($p < 0.05$, **Figures 7C–I**).

Cox Regression Analysis Revealed That *SYDE1* May Be an Independent Survival Factor in Gliomas

After validation of the prognostic function of *SYDE1* in gliomas, we further analyzed the existing data in the CGGA and TCGA databases to confirm whether it was an independent survival predictor in individuals with gliomas. The results of univariate

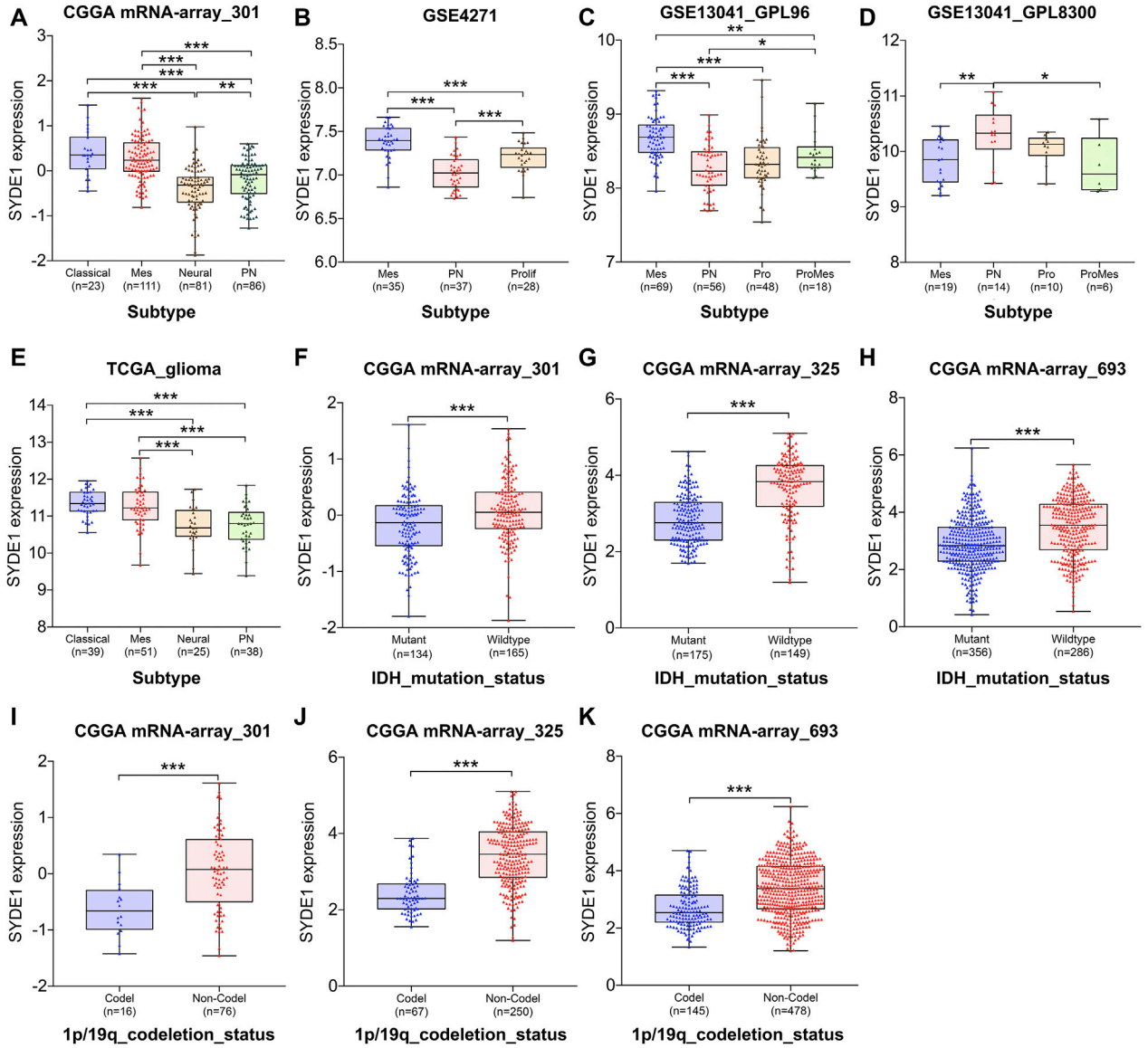


FIGURE 6 | The association between *SYDE1* expression and glioma subtypes, IDH mutation or 1p/19q co-deletion. **(A–E)** The association between *SYDE1* expression and glioma subtypes. **(F–H)** The association between *SYDE1* expression and IDH mutation. **(I–K)** The association between *SYDE1* expression and 1p/19q co-deletion. *, p.value <0.05; **, p.value <0.01; ***, p.value <0.001.

Cox regression analysis demonstrated that some parameters, including *SYDE1* expression, age, WHO grade, primary/recurrent/secondary (PRS) type, histology, radiotherapy, chemotherapy, IDH mutation, and 1p/19q codeletion status, were independent risk factors for the prognosis prediction in glioma patients based on the CGGA mRNA-array_325. Multivariate Cox regression analysis demonstrated that *SYDE1* expression, age, WHO grade, PRS type, chemotherapy, and IDH mutation were independent risk factors for OS ($p < 0.05$, **Supplementary Table S7**). Notably, a similar result showed these features in the CGGA mRNA-array_693, TCGA_glioma, and TCGA_LGG datasets ($p < 0.05$, **Supplementary Tables S8–S10**). Based on these statistical results, *SYDE1* can be considered an independent prognostic risk factor in gliomas, at least in part.

Knockdown of *SYDE1* Suppressed Migratory and Invasive, but Not Proliferative Abilities of Glioma Cells *in vitro*

To date, the role of *SYDE1* in glioma development remains largely undetermined. Based on the correlation between *SYDE1* expression and glioma revealed by bioinformatic analysis and IHC, we performed *in vitro* experiments to further verify this association. To achieve this, we first tested three different siRNA oligonucleotide sequences for their efficiency in suppressing

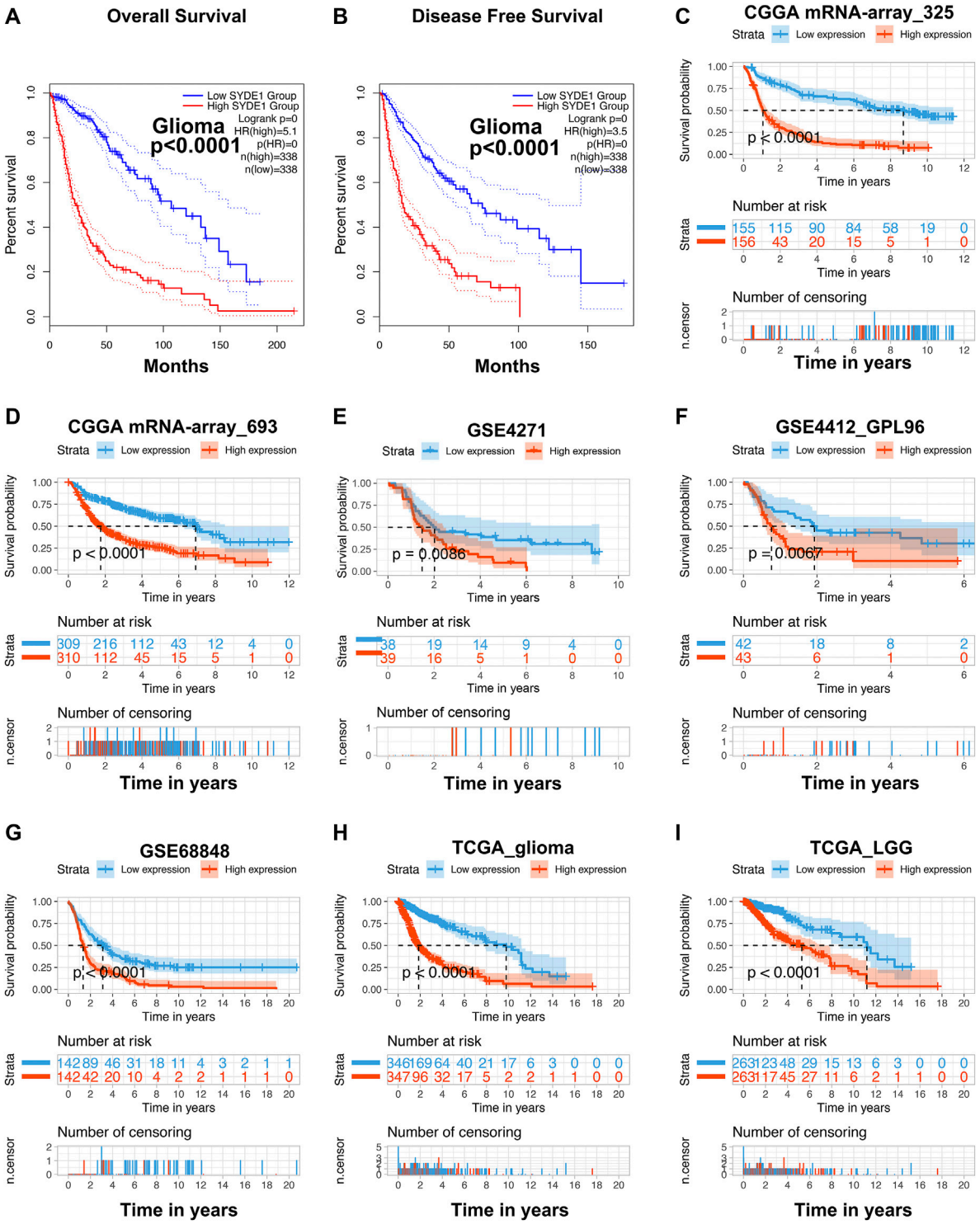


FIGURE 7 | Evaluation of the prognostic value of *SYDE1* in different datasets. **(A–B)** Patients with low *SYDE1* expression groups had a long median OS and DFS than high *SYDE1* expression groups, respectively (GEPIA2). **(C–I)** The high expression of *SYDE1* was positively related to inferior prognosis regarding OS by data mining in different datasets, including mRNA-array_325, mRNA-array_693 of CGGA, GSE4271, GSE4412_GPL96, GSE68848, TCGA_glioma, and TCGA_LGG, respectively.

SYDE1 in A172 cells. After 24 h of siRNA transfection into human astroglial A172 cells, analysis of *SYDE1* expression by qPCR indicated that siRNA#1 was the most effective, mediating 75.6 ± 0.9% mRNA knockdown ($p < 0.01$) compared with nontargeting siRNA (**Figure 8A**). The knockdown of *SYDE1* failed to affect *SYDE1* cell proliferation in the CCK-8 viability

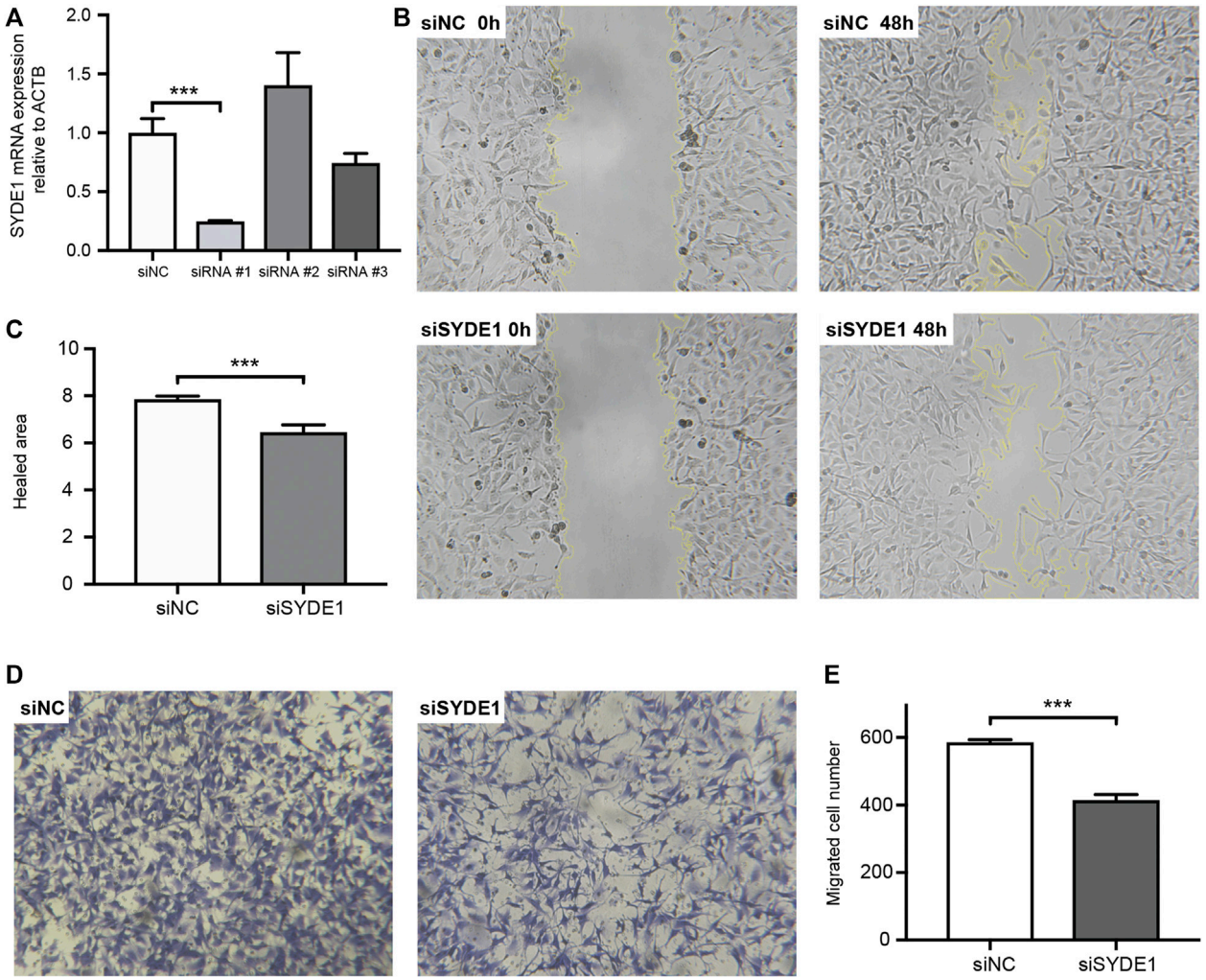


FIGURE 8 | Wound scratch assay and transwell assay. **(A)** qPCR results indicate that knockdown of *SYDE1* with siSYDE1 #1 achieves the highest efficiency. Data are presented as mean ± SD, n = 3 per group, ***, p.value <0.001 for siSYDE1 #1 group versus control group. **(B)** Representative images of A172 migration in the siNC and siSYDE1 groups at 0 and 48 h in the scratch assay. **(C)** Bar graph illustrating the number of cells per high field. Data are presented as mean ± SD, n = 3 per group, ***, p.value <0.001 for siSYDE1 group versus control group. (D) Representative images of invasive cells in the lower chamber stained with crystal violet in the siNC and siSYDE1 groups in the transwell assay. **(E)** Bar graph illustrating the percentage of the scratched area covered. Data are presented as mean ± SD, n = 3 per group, ***, p.value <0.001 for siSYDE1 group versus control group.

assay (**Supplementary Figure S3**). Then, wound healing and transwell assays were performed, which revealed that *SYDE1* knockdown significantly reduced the healed area at 24 h (**Figures 8B,C**) ($p < 0.01$) and the number of migrated cells at 24 h (**Figures 8D,E**) ($p < 0.01$).

Prediction of Coexpression Genes for *SYDE1* Using cBioPortal

GSEA has served as an effective strategy for determining the biological function of a novel gene implicated in tumor development (Wang et al., 2019). To preliminarily decipher the biological function of *SYDE1* in glioma, GSEA was performed to identify gene sets enriched in the high- and low *SYDE1* expression groups. In the high *SYDE1* expression group, significantly enriched hallmark gene sets were associated with tumorigenesis and metastasis, including "Epithelial Mesenchymal Transition", "p53 pathway", "Apoptosis" and "Angiogenesis" (**Figure 9A** i-9A iv). In the low *SYDE1* expression group, tumor-associated terms "Hedgehog signaling" and "Kras signaling" were the enriched gene sets (**Figure 9A** v-9A vi). As shown in **Figures 9A,B** comprehensive coexpression network was built according to the top 100 Spearman's correlation index values from the cBioPortal databases. The top-ranking genes were related to the TRIP10, NR2F6, ACTN4, SBNO2, and TGFB1L1 genes ($p < 0.05$), and a link between ACTN4 and TGFB1L1 expression and tumorigenesis was shown in glioma samples (Ji et al., 2019). GO enrichment analysis was conducted for the coexpressed genes of *SYDE1* ($p < 0.05$, **Figure 9C**), and the

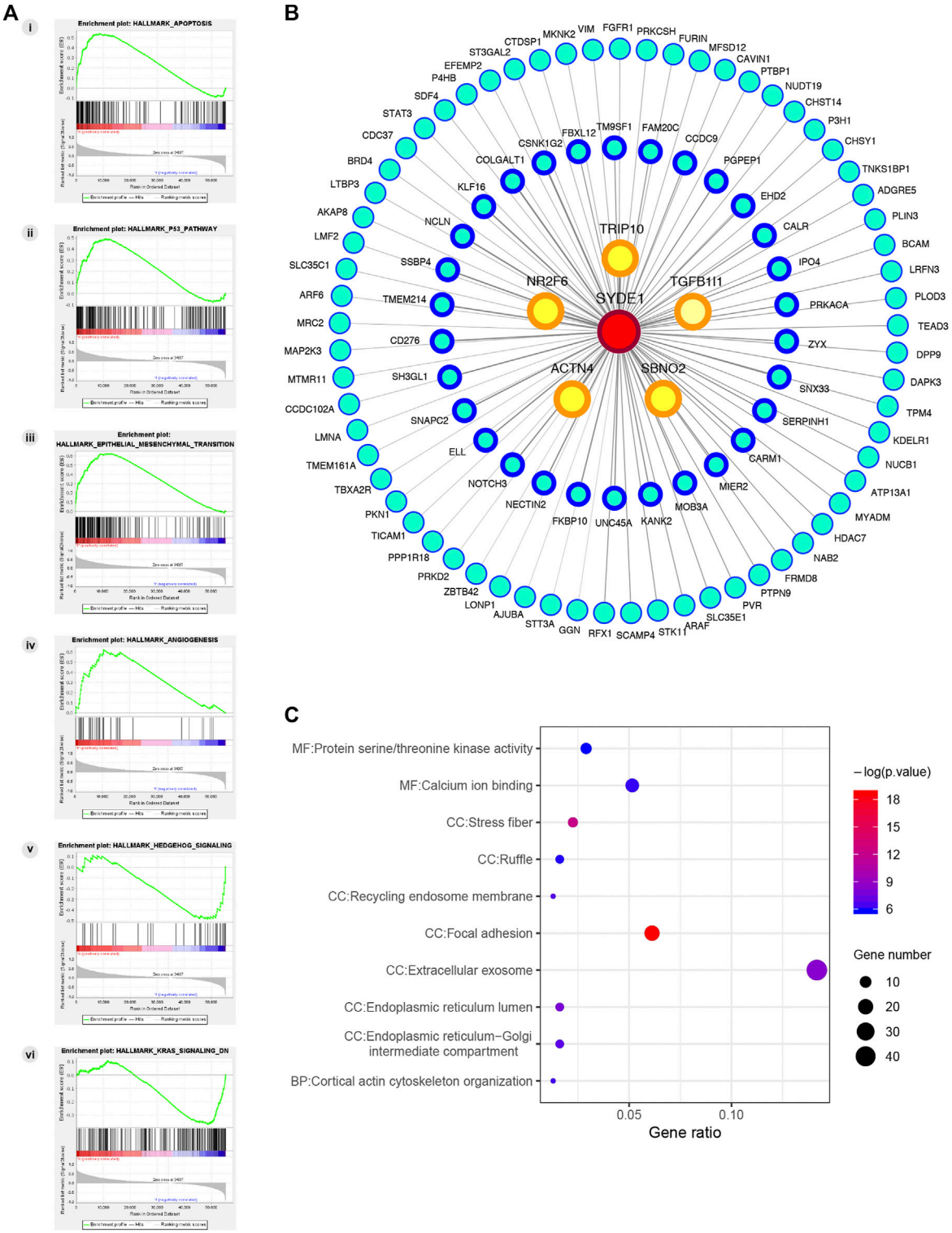


FIGURE 9 | Potential molecular mechanism of *SYDE1* in glioma. **(A)** GSEA pathway using single-gene method of *SYDE1*. (i-iv) Enriched gene sets in the group of high *SYDE1* expression. (v-vi) Enriched gene sets in the group of low *SYDE1* expression. **(B)** The PPI network of the top 100 genes co-expressed with *SYDE1* in glioma tissues. **(C)** GO enrichment results of 100 genes involved in PPI network.

TABLE 3 | Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for predicted co-expression genes of *SYDE1* in gliomas.

Term	ID	Count	*p*-Value	FDR	Fold enrichment	Genes
Insulin signaling pathway	ptr04910	4	0.041691	38.844556	5.033333333	ARAF, MKNK2, PRKACA, TRIP10
Adherens junction	ptr04520	3	0.062905	52.77482	7.13630137	FGFR1, ACTN4, NECTIN2
Protein processing in endoplasmic reticulum	ptr04141	4	0.066436	54.789246	4.159281437	P4HB, STT3A, CALR, PRKCSH
Dilated cardiomyopathy	ptr05414	3	0.086996	65.040685	5.919886364	LMNA, PRKACA, TPM4

TABLE 4 | Reactome pathways analysis for predicted co-expression genes of *SYDE1* in gliomas.

Pathway identifier	Pathway name	Entities *p*-Value	Entities FDR	Submitted entities found
R-HSA-1474244	Extracellular matrix organization	1.97E–05	0.010111351	EFEMP2;SERPINH1;P3H1;PLOD3;FURIN;LTBP3;P4HB;PRKACA; COLGALT1
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	2.50E–05	0.010111351	SERPINH1;P3H1;PLOD3;P4HB;COLGALT1
R-HSA-1474290	Collagen formation	1.74E–04	0.035258571	SERPINH1;P3H1;PLOD3;P4HB;COLGALT1
R-HSA-1566948	Elastic fibre formation	1.75E–04	0.035258571	EFEMP2;FURIN;LTBP3
R-HSA-5655302	Signaling by FGFR1 in disease	5.00E–04	0.080577744	PTBP1;STAT3;FGFR1
R-HSA-2129379	Molecules associated with elastic fibres	8.93E–04	0.119642788	EFEMP2;LTBP3
R-HSA-1839124	FGFR1 mutant receptor activation	0.002088233	0.223660696	STAT3;FGFR1
R-HSA-4419969	Depolymerisation of the Nuclear Lamina	0.002214462	0.223660696	LMNA;PRKACA
R-HSA-420597	Nectin/Necl trans heterodimerization	0.002812629	0.232344096	PVR;NECTIN2
R-HSA-5663202	Diseases of signal transduction by growth factor receptors and second messengers	0.002904301	0.232344096	PTBP1;NOTCH3;ATP13A1;CDC37;STAT3;LMNA;ARAF;FGFR1;HDAC7

coexpressed genes primarily participated in the molecular functions of protein serine kinase activity and calcium ion binding. The coexpressed genes were involved in various cellular components, including stress fiber, ruffle, recycling endosome membrane, focal adhesion, extracellular exosome, and endoplasmic reticulum lumen. Enriched signaling pathways for the top 100 coexpressed genes of *SYDE1* identified in the KEGG pathway analysis were ranked according to *p* values. As shown in **Table 3**, the insulin signaling pathway, adherens junctions, and protein processing in the endoplasmic reticulum remained particularly significant. Reactome pathway analysis was further used to identify the metabolic pathways in which the top 100 coexpressed genes were related to *SYDE1* with $p < 0.05$. The top 10 critically enriched Reactome pathways were obtained and are presented in **Table 4**.

Construction of the mRNA/miRNA/lncRNA Network

We used a comprehensive strategy for starBase 3.0, CGGA, and TargetScan databases to investigate the potential miRNA targets and the indirect lncRNAs targets of *SYDE1* in gliomas. Recapitulation of this part of the work is shown in **Figure 10A** as a flow chart. A total of 143 target miRNAs were acquired from the PITA, miRmap, microT, miRanda, PicTar, and TargetScan databases. By intersecting the predicted target miRNAs, we created a Venn diagram to identify five individual miRNAs that were predicted by all six databases from the 143 miRNA assemblies. There were 14 target miRNAs of *SYDE1* that were supported by at least five databases, and six miRNAs were predicted by the four most common databases (**Figure 10B**). The lncRNAs predicted by target miRNAs in starBase 3.0 were identified for the construction and integration of the mRNA-miRNA-lncRNA network, which may offer intuitive insight into the competing endogenous RNA (ceRNA) mechanism in gliomas (**Figure 10C**). The five optimal target miRNAs screened by the six databases were tested for relevance between the expression level and survival. Taking into consideration the clinical features from the CGGA dataset (miRNA-array_198 and CGGA mRNA-seq 325), the survival curves showed that patients with upregulation of hsa-miR-520e had a favorable OS in primary gliomas ($p < 0.05$, **Figure 10H**), but this generalization was not true for the other four miRNAs (hsa-miR-136, hsa-miR-302a, hsa-miR-424, and hsa-miR-497, **Figures 10D–G**). Considering that hsa-miR-520e is the target miRNA of *SYDE1* with favorable outcomes in gliomas, the expression of the target lncRNAs (FGD5-AS1, MIR17HG, and SNHG16) of hsa-miR-520e was measured to calculate the prognostic ability of these

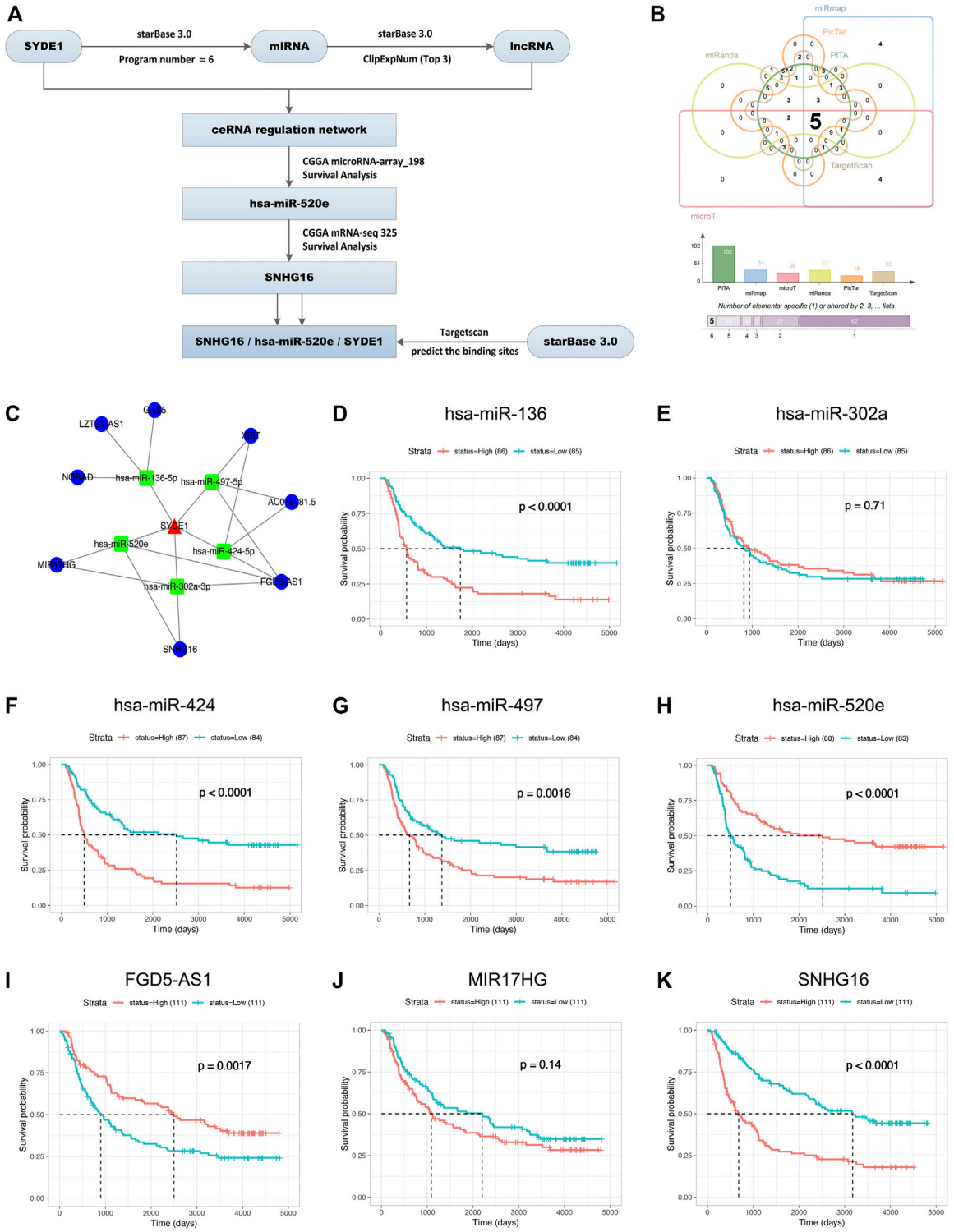


FIGURE 10 | Prediction of *SYDE1*-targeted miRNAs and lncRNAs. **(A)** A schematic overview of different steps taken to predict *SYDE1*-targeted miRNA and lncRNA. **(B)** Upset plot showing the number of predicted *SYDE1*-targeted miRNAs. **(C)** Network of *SYDE1*-targeted miRNAs and lncRNAs. **(D–H)** Kaplan-Meier survival curves for 5-miRNA signature. has-miR-136 **(D)**, has-miR-424 **(F)**, has-miR-497 **(G)** and has-miR-520e **(H)** are positively associated with OS, whereas has-miR-302a **(E)** is not associated with OS. (I–K) Kaplan-Meier survival curves for 3-lncRNA signature. FGD5-AS1 **(I)**, MIR17HG **(J)** and SNHG16 **(K)** are positively associated with OS.

signatures. Expression of SNHG16 was higher in LGGs or HGGs than normal brain tissues (**Supplementary Figure S4**). The results demonstrated that the high expression of SNHG16 in primary glioma patients had a lower OS ($p < 0.05$, **Figure 10K**), and in the other two lncRNA (FGD5-AS1, MIR17HG) groups, there was no apparent defined association described for SNHG16 (**Figures 10I,J**). These results provided initial evidence for the *SYDE1/hsa-miR-520e/SNHG16* network in gliomas.

Identification of *SYDE1/Hsa-miR-520e/SNHG16* Interactions

Hsa-miR-520e and SNHG16 were selected for further evaluation. A growing body of evidence demonstrates that lncRNAs play a key role in tumorigenesis by functioning as competing endogenous RNAs (ceRNAs) to mRNA. To delineate the precise *SYDE1*/hsa-miR-520e/SNHG16 interaction, the potential targets were collected from the PITA, miRmap, microT, miRanda, PicTar, TargetScan, and starBase 3.0 databases. Six databases identified *SYDE1*/hsa-miR-520e interactions, and one database identified an hsa-miR-520e/SNHG16 interaction (**Supplementary Figure S5**).

DISCUSSION

Gliomas are the most common primary malignant brain tumors in adults, and their exact molecular mechanisms have not yet been completely elucidated. In this study, we report the gene *SYDE1* as a novel regulator of glioma tumors. First, we analyzed publicly published mRNA expression data of human glioma tissues and normal control tissues and found that *SYDE1* expression was higher in gliomas than in healthy cerebral tissues. Then, normal human brain and glioma samples from grades I to IV were collected for IHC staining of SYDE1, which revealed that SYDE1 expression is positively correlated with the clinical malignancies of glioma. In addition, *SYDE1* is more highly expressed in recurrent or necrotic gliomas or gliomas that occur in elderly patients.

A difference in the expression of a gene between normal brain tissue and glioma tissues indicates a potential modulatory mechanism of glioma development. There are many well-studied pathogenic genes and regulators of gliomas, including IDH1/2, 1p/19q codeletion, integrin β1 (ITGB1) and v-raf murine sarcoma viral oncogene homolog B1 (BRAF) V600E mutation (Chintala et al., 1996; Schiffman et al., 2010; Yip et al., 2012). In particular, ITGB1 encodes the β1 subunit of extracellular matrix (ECM) integrins and shows reduced expression in gliomas compared with normal controls. Treating a human glioblastoma cell line with anti-β1 antibodies *in vitro* diminishes integrin synthesis on the cell surface, which can result in an increase in matrix metalloprotease-2 activity and invasiveness of the cell (Chintala et al., 1996). In a similar manner, our data show that the expression of *SYDE1* is higher in human glioma tissues, so it is reasonable to explore the biological function and molecular mechanism of *SYDE1* in gliomas.

With regard to the molecular mechanism, it is noteworthy that GSEA of the high *SYDE1* expression group showed enrichments in epithelial mesenchymal transition and the p53 pathway, which are highly related to tumorigenesis and metastasis. Moreover, GO and KEGG pathway analysis of *SYDE1* coexpressed genes revealed an enrichment of tumor-associated terms, including protein serine kinase activity and focal adhesion. Of note, adhesion to and migration through the extracellular matrix (ECM) is recognized as an important part of the metastatic process and is necessary for the invasion of a variety of tumors (Pauli et al., 1983). The role of SYDE1 in modulating cell migration has already been reported in recent years. As revealed by Lo et al., SYDE1 can promote cytoskeletal remodeling as well as migration and invasion of placental trophoblast cells, which is crucial for maintaining the maternal-trophoblast interface (Lo et al., 2017). In our study, knockdown of *SYDE1 in vitro* significantly abolished the migration and invasion of glioma cell lines A172. Taken together, increased expression of *SYDE1* in gliomas may lead to an overactivated transcriptional network that facilitates tumor invasion.

We next report the SNHG16/hsa-miR-520e axis as a downstream target of *SYDE1* in gliomas. SNHG16 is a novel cancer-related lncRNA and has been demonstrated to function as an oncogene in human breast cancer, gastric cancer, or hepatocellular cancer (Yang and Wei, 2019). For instance, increased SNHG16 expression in human gastric cancer can promote *in vitro* proliferation and *in vivo* growth of gastric tumors (Lian et al., 2017). SNHG16 is also reported to be downregulated in several other malignant tumors, such as hepatocellular carcinoma, and to inhibit tumor growth (Lin et al., 2019). Considering the versatile role of SNHG16 in tumor development, we examined SNHG16 expression in our study and found that the SNHG16 level was significantly increased in gliomas. That is, upregulated *SYDE1* in gliomas can potentially activate SNHG16 expression to facilitate the onset and progression of gliomas. microRNAs (miRNAs), such as hsa-miR-93, miR-338-3p, miR-124-3p and miR-128, are common targets of SNHG16 in human cancers (Xu et al., 2018; Chen et al., 2020; Wu et al., 2020). In our study, hsa-miR-520e was identified to interact with SNHG16 in gliomas via RMBase 2.0, and further *in vitro* experiments to validate the SNHG16/hsa-miR-520e axis are in progress.

In summary, this study identifies *SYDE1* as an oncogene in gliomas that can regulate the proliferation and migration of glioma cells and is predicted to interact with the SNHG16/hsa-miR-520e axis. In-depth elucidation of the functions of *SYDE1* in gliomas, such as tumor formation in nude mice and population mutation screening, will provide a deeper understanding of the molecular pathology of gliomas. With these advances, *SYDE1* can be a promising future biomarker for glioma clinical practices, including serving as a reference for the WHO classification and predicting the prognosis of glioma 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 The Ethical Committee of the Ninth People's Hospital of Shanghai Jiao Tong University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the Ethics Committee of Shanghai Ninth People's Hospital affiliated to Shanghai Jiao Tong University, School of Medicine.

AUTHOR CONTRIBUTIONS

ZT and SC conceived and developed the outline of the study. ZH, XZ, and BY downloaded the data and performed the statistical studies. ZH generated figures and tables and wrote the manuscript. JX participated in sample collection and performed *in vitro* assays. LJ and PH revised the manuscript. All authors read and approved the final manuscript.

FUNDING

The work was supported by the Natural Science Foundation of Shanghai (grant no. 18ZR1422200).

ACKNOWLEDGMENTS

The authors would like to thank Yun Sun for her valuable assistance to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.714203/full#supplementary-material

REFERENCES

Arko, L., Katsyv, I., Park, G. E., Luan, W. P., and Park, J. K. (2010). Experimental Approaches for the Treatment of Malignant Gliomas. *Pharmacol. Ther.* 128 (1), 1–36. doi:10.1016/j.pharmthera.2010.04.015

Buchner, A., Castro, M., Hennig, A., Popp, T., Assmann, G., Stief, C. G., et al. (2010). Downregulation of HNF-1B in Renal Cell Carcinoma Is Associated with Tumor Progression and Poor Prognosis. *Urology* 76 (2), e6–507. doi:10.1016/j.urology.2010.03.042

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

Chen, Z.-Y., Wang, X.-Y., Yang, Y.-M., Wu, M.-H., Yang, L., Jiang, D.-T., et al. (2020). LncRNA SNHG16 Promotes Colorectal Cancer Cell Proliferation, Migration, and Epithelial-Mesenchymal Transition through miR-124-3p/MCP-1. *Gene Ther.* 2020, 1–13. doi:10.1038/s41434-020-0176-2

Chintala, S. K., Sawaya, R., Gokaslan, Z. L., and Rao, J. S. (1996). Modulation of Matrix Metalloprotease-2 and Invasion in Human Glioma Cells by α3β1 Integrin. *Cancer Lett.* 103 (2), 201–208. doi:10.1016/0304-3835(96)04215-2

Dunbar, E. M. (2009). The Role of Chemotherapy for Pure and Mixed Anaplastic Oligodendroglial Tumors. *Curr. Treat. Options. Oncol.* 10 (3-4), 216–230. doi:10.1007/s11864-009-0091-7

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. Signaling* 6 (269), pl1. doi:10.1126/scisignal.2004088

Gilbert, M. R., Wang, M., Aldape, K. D., Stupp, R., Hegi, M. E., Jaeckle, K. A., et al. (2013). Dose-dense Temozolomide for Newly Diagnosed Glioblastoma: a Randomized Phase III Clinical Trial. *Jco* 31 (32), 4085–4091. doi:10.1200/jco.2013.49.6968

Goodenberger, M. L., and Jenkins, R. B. (2012). Genetics of Adult Glioma. *Cancer Genet.* 205 (12), 613–621. doi:10.1016/j.cancergen.2012.10.009

Hofer, S., and Lassman, A. B. (2010). Molecular Markers in Gliomas: Impact for the Clinician. *Targ Oncol.* 5 (3), 201–210. doi:10.1007/s11523-010-0157-2

Ji, J., Xu, R., Ding, K., Bao, G., Zhang, X., Huang, B., et al. (2019). Long Noncoding RNA SChLAP1 Forms a Growth-Promoting Complex with HNRNPL in Human Glioblastoma through Stabilization of ACTN4 and Activation of NF-κB Signaling. *Clin. Cancer Res.* 25 (22), 6868–6881. doi:10.1158/1078-0432.Ccr-19-0747

Kamiya-Matsuoka, C., and Gilbert, M. R. (2015). Treating Recurrent Glioblastoma: an Update. *CNS Oncol.* 4 (2), 91–104. doi:10.2217/cns.14.55

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. *Nucl. Acids Res.* 42, D92–D97. Database issue. doi:10.1093/nar/gkt1248

Lian, D., Amin, B., Du, D., and Yan, W. (2017). Enhanced Expression of the Long Non-coding RNA SNHG16 Contributes to Gastric Cancer Progression and Metastasis. *Cbm* 21 (1), 151–160. doi:10.3233/cbm-170462

Lin, Q., Zheng, H., Xu, J., Zhang, F., and Pan, H. (2019). LncRNA SNHG16 Aggravates Tumorigenesis and Development of Hepatocellular Carcinoma by Sponging miR-4500 and Targeting STAT3. *J. Cel Biochem* 120, 11604–11615. doi:10.1002/jcb.28440

Lo, H.-F., Tsai, C.-Y., Chen, C.-P., Wang, L.-J., Lee, Y.-S., Chen, C.-Y., et al. (2017). Association of Dysfunctional Synapse Defective 1 (SYDE1) with Restricted Fetal Growth - SYDE1 Regulates Placental Cell Migration and Invasion. *J. Pathol.* 241 (3), 324–336. doi:10.1002/path.4835

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., et al. (2007). The 2007 WHO Classification of Tumours of the central Nervous System. *Acta Neuropathol.* 114 (2), 97–109. doi:10.1007/s00401-007-0243-4

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Pauli, B. U., Schwartz, D. E., Thonar, E. J.-M., and Kuettner, K. E. (1983). Tumor Invasion and Host Extracellular Matrix. *Cancer Metast Rev.* 2 (2), 129–152. doi:10.1007/bf00048966

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

Schiffman, J. D., Hodgson, J. G., VandenBerg, S. R., Flaherty, P., Polley, M.-Y. C., Yu, M., et al. (2010). Oncogenic BRAF Mutation with CDKN2A Inactivation Is Characteristic of a Subset of Pediatric Malignant Astrocytomas. *Cancer Res.* 70 (2), 512–519. doi:10.1158/0008-5472.Can-09-1851

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

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Tissue-based Map of the Human Proteome. *Science* 347 (6220), 1260419. doi:10.1126/science.1260419

Wang, Z., Wang, Y., Peng, M., and Yi, L. (2019). UBASH3B Is a Novel Prognostic Biomarker and Correlated with Immune Infiltrates in Prostate Cancer. *Front. Oncol.* 9, 1517. doi:10.3389/fonc.2019.01517

Wu, W., Guo, L., Liang, Z., Liu, Y., and Yao, Z. (2020). Lnc-SNHG16/miR-128 axis Modulates Malignant Phenotype through WNT/β-catenin Pathway in Cervical Cancer Cells. *J. Cancer* 11 (8), 2201–2212. doi:10.7150/jca.40319

Xu, F., Zha, G., Wu, Y., Cai, W., and Ao, J. (2018). Overexpressing lncRNA SNHG16 Inhibited HCC Proliferation and Chemoresistance by Functionally Sponging Hsa-miR-93. *OncoTargets Ther.* 11, 8855–8863. doi:10.2147/ott.S182005

Xue, S., Hu, M., Iyer, V., and Yu, J. (2017). Blocking the PD-1/pd-L1 Pathway in Glioma: a Potential New Treatment Strategy. *J. Hematol. Oncol.* 10 (1), 81. doi:10.1186/s13045-017-0455-6

Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., et al. (2009). IDH1andIDH2Mutations in Gliomas. *N. Engl. J. Med.* 360 (8), 765–773. doi:10.1056/NEJMoa0808710

Yang, M., and Wei, W. (2019). SNHG16: A Novel Long-Non Coding RNA in Human Cancers. *OncoTargets Ther* 12, 11679–11690. doi:10.2147/ott.S231630

Yip, S., Butterfield, Y. S., Morozova, O., Chittaranjan, S., Blough, M. D., An, J., et al. (2012). Concurrent CIC Mutations, IDH Mutations, and 1p/19q Loss Distinguish Oligodendrogliomas from Other Cancers. *J. Pathol.* 226 (1), 7–16. doi:10.1002/path.2995

Zhang, J., Lu, H., Zhang, S., Wang, T., Zhao, H., Guan, F., et al. (2021). Leveraging Methylation Alterations to Discover Potential Causal Genes Associated with the Survival Risk of Cervical Cancer in TCGA through a Two-Stage Inference Approach. *Front. Genet.* 12, 667877. doi:10.3389/fgene.2021.667877

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 Han, Zhuang, Yang, Jin, Hong, Xue, Chen and Tian. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fmolb.2021.720645

ARHGAP11A Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric Cancer

*Biao Fan†, Ke Ji†, Zhaode Bu, Ji Zhang, Heli Yang, Jialin Li and Xiaojiang Wu**

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute, Beijing, China

Background: ARHGAP11A, belongs to RhoGAPs family, is vital for cell motility. However, the role of ARHGAP11A in gastric cancer is obscure.

Methods: The expression level of ARHGAP11A was analyzed by Oncomine database. The correlation of ARHGAP11A expression with immune infiltrates and associated gene markers was clarified by Tumor IMmune Estimation Resource and Gene Expression Profiling Interactive Analysis database. The correlation between ARHGAP11A expression and the patient prognosis was identified by Kaplan-Meier plotter and PrognoScan. Genetic changes of ARHGAP11A were analyzed by cBioPortal. The protein-protein interaction network and gene functional enrichment analysis were constructed and performed by GeneMANIA and Metascape.

Results: We found that the expression levels of ARHGAP11A were elevated in various cancers including gastric cancer when compared with normal tissues. High expression of ARHGAP11A was significantly correlated with a better prognosis in gastric cancer. We revealed that the expression of ARHGAP11A was negatively associated with infiltration levels of $CD8^+$ T cells, $CD4^+$ T cells, macrophages and dendritic cells. In addition, ARHGAP11A expression was significantly correlated with gene markers of these immune cells. Lastly, gene functional enrichment analysis indicated that ARHGAP11A involved in regulating lymphocyte activation, cell division, cell killing, myeloid leukocyte differentiation and leukocyte apoptosis.

Conclusion: Our findings demonstrated that ARHGAP11A was a valuable prognostic biomarker in gastric cancer. Further work is needed to validate its role and underlying mechanisms in regulating immune infiltrates.

Keywords: ARHGAP11A, immune infiltrates, gastric cancer, prognosis, biomarker

OPEN ACCESS

Edited by:
Ismail Hosen,
University of Dhaka, Bangladesh

Reviewed by:
Quan Zhou,
China Three Gorges University, China
Mahdi Sheikh,
International Agency For Research On Cancer (IARC), France

****Correspondence:***
Xiaojiang Wu
drwuxiaojiang@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *04 June 2021*
Accepted: *29 September 2021*
Published: *18 October 2021*

Citation:
Fan B, Ji K, Bu Z, Zhang J, Yang H, Li J and Wu X (2021) ARHGAP11A Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric Cancer. Front. Mol. Biosci. 8:720645. doi: 10.3389/fmolb.2021.720645

INTRODUCTION

Gastric cancer is a common cancer, especially in Asian countries (Sun et al., 2020; Yang et al., 2020; Cao et al., 2021). It is the fourth leading cause of cancer-related mortality worldwide (Sung et al., 2021). In recent years, the immunotherapy has bought revolutionary changes to the treatment of cancer (Ribas and Wolchok, 2018). However, the progress of immunotherapy in gastric cancer was hampered by a limited understanding of the immune microenvironment (Balkwill et al., 2012). The tumor immune microenvironment is crucial for tumor initiation and progression (Salmon et al.,

2019). It is composed of all populations of tumor infiltrating cells including macrophages, T cells and dendritic cells (Bruni et al., 2020). Studies have shown that the tumor immune infiltrates correlated with prognosis and response to therapy (Zeng et al., 2019; Refolo et al., 2020). Therefore, there is a clear need to investigate the immune infiltrates related markers and to reveal the molecular mechanisms in gastric cancer.

Rho GTPases are a subfamily of the Ras superfamily proteins which play central roles in multiple biological processes, such as cell motility, cell polarity, cell cycle progression, cell adhesion, migration and invasion. Rho GTPase-activating proteins (RhoGAPs), upstream regulators of Rho GTPases, are frequently dysregulated in various cancers (Porter et al., 2016; Muller et al., 2020). Previous studies showed that Rho GTPases played a role in immune homeostasis, which involved in key processes for the T lymphocytes activation and differentiation (Saoudi et al., 2014). Tumors with abundant tumor-infiltrating lymphocytes (TILs) are associated with a better prognosis than tumors with scarce TILs in various cancers (Brambilla et al., 2016; Sudo et al., 2017). Low-grade TILs are associated with lymph node metastasis of early-stage cancer cells (Zhao et al., 2020). Whether RhoGAPs involve in the tumor immune microenvironment is still unknown. ARHGAP11A, a protein coding gene locates on chr15q13.3, encodes a member of RhoGAPs (ARHGAP11A). The role of ARHGAP11A in cancer is still controversial. In human glioma cells, ARHGAP11A binds to p53 and promotes its function eventually leading to cell-cycle arrest and apoptosis (Xu et al., 2013). ARHGAP11A is upregulated in liver cancer and proceeds the liver cell proliferation and migration via Rac1B (Dai et al., 2018). ARHGAP11A was found involved in the cell migration of breast cancer (Lawson et al., 2016). In our ongoing parallel study, using whole exon and whole genome sequencing, we characterized multiple metastases arising from gastric cancer in twelve patients. We found that high expression of ARHGAP11A, a representative gene with mutation characteristics in the clonal evolution of gastric cancer metastasis, appeared more frequently in gastric cancer with lymph node metastasis (unpublished data). ARHGAP11A might play a key role in lymph node metastasis of gastric cancer. Nevertheless, the prognostic significance of ARHGAP11A and its correlation with immune infiltrates including TILs in gastric cancer is obscure.

In this study, we analyzed the correlation between ARHGAP11A expression and patient prognosis using PrognoScan and Kaplan-Meier plotter. We next investigated the correlation of ARHGAP11A expression and tumor infiltrates in Tumor IMmune Estimation Resource (TIMER) and Gene Expression Profiling Interactive Analysis (GEPIA).

MATERIALS AND METHODS

Oncomine Database Analysis

Oncomine (https://www.oncomine.org/resource/login.html) is a publicly available tumor microarray database and data mining platform that includes 715 datasets and 86,733 samples (includes tumor and normal tissue samples). Gene expression analyses for a single gene can be performed across various types of cancer and include comparisons relative to normal control (Rhodes et al., 2007). The expression of ARHGAP11A in different cancer tissues were analyzed by using Oncomine. The threshold was set as: $p <$ 1.0E−04, fold change >2, gene rank: top 10% and data type: mRNA. One dataset (Cui Gastric Dataset: 80 gastric carcinoma and 80 normal paired gastric tissue samples were analyzed (Cui et al., 2011)) met the threshold when compared the ARHGAP11A expression in gastric cancer and normal tissue.

TIMER Database and GEPIA Database Analysis

TIMER is a public resource for systematic analysis of immune infiltrates across 32 cancer types (10,897 tumor samples) from The Cancer Genome Atlas (TCGA) (https://cistrome.shinyapps.io/timer/) (Li et al., 2017). The DiffExp module of TIMER was used to identify the expression of ARHGAP11A in all TCGA tumors. Gene expression levels were displayed using box plots, with statistical significance of differential expression level of ARHGAP11A evaluated using the Wilcoxon test and marked with asterisk. The Gene module of TIMER database was used to clarify the correlation of ARHGAP11A expression with immune infiltration level, in which the scatterplots was generated and displayed, showed the purity-corrected partial Spearman's rho value and statistical significance. The Survival module was used to explore the survival differences of patients with different immune infiltrates. The Correlation module of TIMER database was used to draw the expression scatterplots between ARHGAP11A and immune related markers together with the Spearman's rho value and estimated statistical significance (Correlation adjusted by tumor purity). GEPIA (http://gepia.cancer-pku.cn/index.html) is an online database that includes 9,736 tumors and 8,587 normal samples from TCGA and the GTEx projects (Tang et al., 2017). It was used to further validate the significantly correlated genes in TIMER. Gene expression correlation analysis was performed for given sets of TCGA expression data. The correlation coefficient was determined by the Spearman method.

Kaplan-Meier Plotter and PrognoScan Analysis

The Kaplan Meier plotter (https://kmplot.com/analysis/) is an online database capable to assess the effect of 54,675 genes (mRNA, miRNA, protein) on survival in 21 cancer types including gastric cancer (1,440 samples) (Szasz et al., 2016). Sources for the databases include GEO, EGA, and TCGA. It was used to analyze the relationship of ARHGAP11A expression with Overall Survival (OS) and Relapse Free Survival (RFS) in various cancer types, and hazard ratio (HR) values with 95% confidence intervals and log-rank p-values were calculated. Adjusted analyses were performed by using data from gastric cancer patients with different clinical parameters such as gender, tumor stage, Lauren classification, differentiation, treatment and HER2 status. PrognoScan (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) provides a powerful platform to

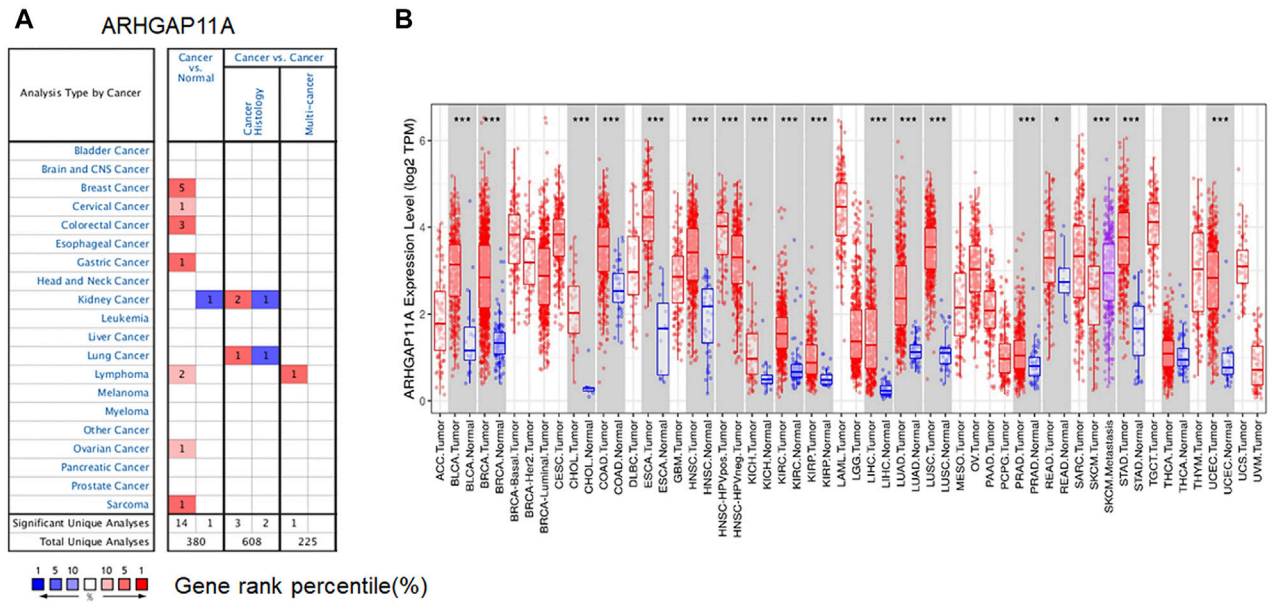


FIGURE 1 | ARHGAP11A expression in different tumor and normal tissues. **(A)** Oncomine. (*p*-value < 1.0E–04), Cell color is determined by the best gene rank percentile for the analyses within the cell. An analysis may be counted in more than one cancer type. **(B)** TIMER. (*p-value < 0.05, *** *p*-value < 1.0E–03).

explore the relationships between gene expression and patient prognosis across a large collection of publicly available cancer microarray datasets (Mizuno et al., 2009). The threshold was set as: Corrected *p*-value and Cox *p*-value both <0.05. The impact of both ARHGAP11A expression level and clinical parameters was analyzed.

cBioPortal Analysis

The cBio Cancer Genomics Portal (https://cbioportal.org) has multidimensional cancer genomics datasets (Gao et al., 2013). Data from 1,120 patients (TCGA datasets) was selected to analyze genetic changes of gastric cancer by using cBioPortal. Genetic alterations were shown in different colors.

GeneMANIA protein-protein interaction (PPI) Analysis and Metascape Gene Enrichment Analysis

GeneMANIA (http://genemania.org/) is an online tool uses bioinformatic methods to display a list of interacting genes, including gene co-expression, physical interactions, gene co-localization, gene enrichment analysis and website prediction. It always be used to construct a PPI network and analyze the function of interactive genes (Warde-Farley et al., 2010). Metascape (https://metascape.org/gp/index.html) is a gene function annotation website (Zhou et al., 2019). It integrates multiple authoritative data resources such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, UniProt and DrugBank to complete thorough pathway enrichment and biological process annotation. A PPI network which contained genes interacting with ARHGAP11A was constructed by using GeneMANIA. Genes identified by the GeneMANIA PPI network and TIMER analysis were included in the GO function analyses by using Metascape.

Statistical Analysis

Patient survival plots generated from the TIMER, GEPIA, PrognoScan and Kaplan-Meier plotter were displayed with HR, *p* or Cox *p*-value from a log-rank test. $p < 0.05$ was considered statistically significant.

RESULTS

Pan-Cancer Analysis of ARHGAP11A Expression Levels

We firstly analyzed the expression of ARHGAP11A in different tissues by using Oncomine. We revealed that expression levels of ARHGAP11A were elevated in breast, cervical, colorectal, gastric, ovarian cancers, lymphoma and sarcoma relative to normal tissues. In contrast, the ARHGAP11A expression was lower in kidney cancer when compared with normal kidney tissue (**Figure 1A**). Details were shown in **Supplementary Table S1**. Next, we assessed the transcriptional levels of ARHGAP11A by using RNA-sequencing data in TCGA and TIMER. Results showed significant differences in ARHGAP11A expression levels when compared tumor and normal tissues (**Figure 1B**). For example, the transcriptional expression of ARHGAP11A was significantly elevated relative to normal tissues in esophageal carcinoma (ESCA) and stomach adenocarcinoma (STAD). Moreover, the transcriptional expression level of ARHGAP11A in metastatic lesion of skin cutaneous melanoma (SKCM. Metastasis) was significantly higher than

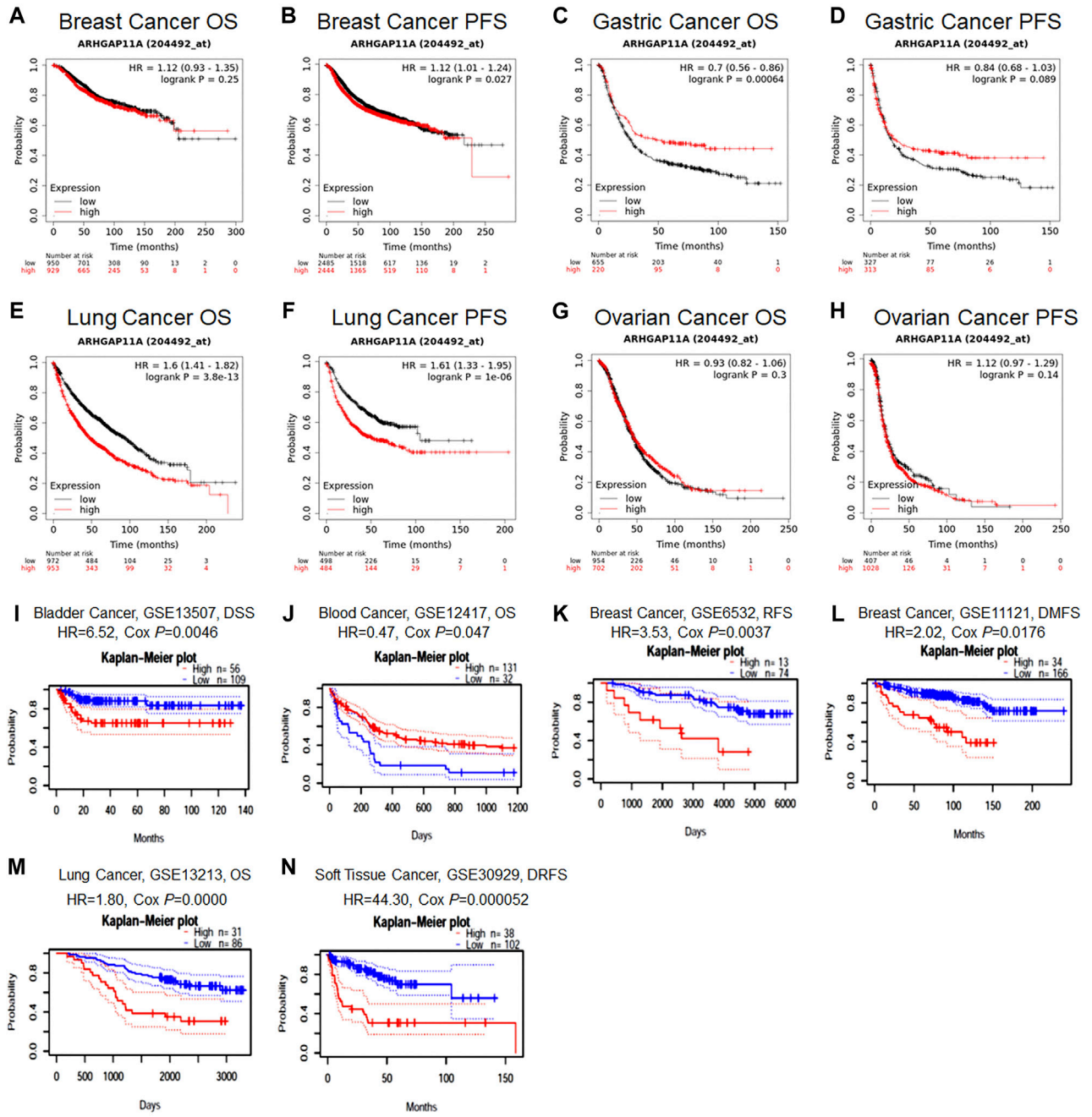


FIGURE 2 | The prognostic value of ARHGAP11A expression in various types of cancer. **(A–H)** Kaplan-Meier Plotter. **(I–N)** PrognoScan. The cohort GSE12417: expression profiling by array; cancer subtype: acute myeloid leukemia; n = 163; endpoint: overall survival. The cohort GSE13213: expression profiling by array; cancer subtype: lung adenocarcinoma; n = 117; patient age: 32–84 years old; endpoint: overall survival. The cohort GSE13507: expression profiling by array; cancer subtype: bladder transitional cell carcinoma; n = 165; patient age: 27–88 years old; sample type: frozen tissue; endpoint: disease specific survival. The cohort GSE6532: expression profiling by array; cancer subtype: breast cancer; n = 87; endpoint: distant metastasis free survival. The cohort GSE11121: expression profiling by array; cancer subtype: breast cancer; n = 200; patient age: 34–89 years old; sample type: frozen; endpoint: distant metastasis free survival. The cohort GSE30929: expression profiling by array; cancer subtype: liposarcoma; n = 140; endpoint: distant recurrence free survival.

in the primary lesion of skin cutaneous melanoma (SKCM. Tumor). In consistent with high transcription level of ARHGAP11A in gastric cancer tissues shown in **Figure 1**; Cytoplasmic expression of ARHGAP11A was also detected in most cancers including gastric cancer. As shown in **Supplementary Figure S1** 7 of 9 gastric cancer patients show high/median expression (data from The Human Protein Atlas: https://www.proteinatlas.org/).

TABLE 1 | Correlation of ARHGAP11A and clinical parameters in gastric cancer from Kaplan-Meier Plotter.

Clinical parameters	OS(n = 881)			PFS (n = 645)		
	N	HR	*p*	N	HR	*p*
Sex						
Female	236	0.5 (0.33–0.74)	0.00053	201	0.52 (0.34–0.81)	0.0028
Male	544	0.76 (0.6–0.97)	0.027	437	0.88 (0.68–1.13)	0.31
Tumor stage						
1	67	0 (0-Inf)	0.0012	60	0 (0-Inf)	0.0054
2	140	0.61 (0.3–1.24)	0.17	131	0.74 (0.41–1.36)	0.33
3	305	0.56 (0.4–0.78)	0.00048	186	0.55 (0.36–0.84)	0.0046
4	148	0.66 (0.44–0.98)	0.038	141	0.78 (0.52–1.16)	0.21
Lauren classification						
Intestinal	320	0.37 (0.25–0.55)	1.7e–07	263	0.56 (0.39–0.82)	0.0026
Diffuse	241	0.61 (0.43–0.86)	0.0046	231	0.66 (0.47–0.94)	0.019
Mixed	32	0.49 (0.14–1.77)	0.27	28	1.37 (0.51–3.72)	0.53
Differentiation						
Poor	165	1.29 (0.85–1.96)	0.23	121	0.78 (0.49–1.23)	0.28
Moderate	67	0.58 (0.29–1.14)	0.11	67	0.61 (0.32–1.16)	0.13
Well	32	1.43 (0.55–3.68)	0.46			
Treatment						
Surgery alone	380	0.5 (0.35–0.72)	9.3e–05	375	0.62 (0.44–0.87)	0.0052
5-Fu based adjuvant Chemotherapy	152	2.09 (1.39–3.15)	0.00029	152	2.06 (1.39–3.06)	0.00026
Other adjuvant Chemotherapy	76	0.59 (0.24–1.49)	0.26	80	1.53 (0.69–3.35)	0.29
HER2 status						
Negative	532	0.56 (0.43–0.72)	8.4e–06	408	0.68 (0.52–0.88)	0.0029
Positive	343	1.47 (1.12–1.93)	0.0054	232	1.69 (1.21–2.38)	0.0021

The Prognostic Value of ARHGAP11A Expression in Cancer Patients

We next identified the prognostic value of ARHGAP11A in various cancers by using the Prognoscan and Kaplan-Meier plotter (**Figure 2** and **Supplementary Tables 2–6**). Results revealed that high expression of ARHGAP11A was significantly associated with a better prognosis in gastric cancer (OS HR = 0.7, p = 6.4e–04) (**Figure 2C**) and blood cancer (The cohort GSE12417, OS HR = 0.47, Cox p = 0.047) (**Figure 2J**). In contrast, high expression of ARHGAP11A was correlated with poor prognosis in lung cancer (OS HR = 1.6, p = 3.8e–13; Progression free survival (PFS) HR = 1.61, P = 1e–06; The cohort GSE13213, OS HR = 1.80, Cox p = 0.0000) (**Figures 2E,F**, **2M**), bladder cancer (The cohort GSE13507, disease Specific Survival (DSS) HR = 6.52, Cox p = 0.0046) (**Figure 2I**), breast cancer (The cohort GSE6532, RFS HR = 3.53, Cox p = 0.0037; The cohort GSE11121, Distant Metastasis Free Survival (DMFS) HR = 2.02, Cox p = 0.0176) (**Figures 2K,L**) and soft tissue cancer (The cohort GSE30929, Distant relapse free survival (DRFS) HR = 44.30, Cox p = 0.000052) (**Figure 2N**). No significant relationship was identified between the expression of ARHGAP11A and prognosis of breast and ovarian cancer patients (**Figures 2A,B,G,H**).

Correlation of ARHGAP11A Expression and Clinical Parameters of Gastric Cancer

To further understand the role of ARHGAP11A in gastric cancer, we analyzed the correlation between the ARHGAP11A expression and clinical parameters by using the Kaplan-Meier plotter. High expression of ARHGAP11A was significantly correlated with better prognosis in patients with specific clinical parameters ($p < 0.05$) (**Table 1**). Interestingly, the therapeutic strategy and HER2 status influenced the prognostic value of ARHGAP11A. High expression level of ARHGAP11A was associated with better prognosis when the patient treated with surgery alone, while in contrast, associated with worse prognosis when treated with surgery and 5-Fu based adjuvant chemotherapy (**Table 1**). Moreover, High expression levels of ARHGAP11A indicated better prognosis in HER2 negative patients, while indicated worse prognosis in HER2 positive patients (**Table 1**).

Relationships Between ARHGAP11A Expression and Immune Infiltrates in Gastric Cancer

We next analyzed the correlation between ARHGAP11A expression and immune infiltrates in gastric cancer by using TIMER (**Figure 3**). We found that the expression of ARHGAP11A was negatively associated with infiltration levels of $CD8^+$ T cells (p = 1.38e–04), $CD4^+$ T cells (p = 1.64e–03), macrophages (p = 4.56e–09) and dendritic cells (p = 1.51e–04) (**Figure 3A**). Moreover, the macrophage and dendritic cell infiltration significantly correlate with prognosis of gastric cancer patients in KM survival analysis (**Figure 3B**). The upper results implied ARHGAP11A might affect patient prognosis via regulating immune infiltrates in gastric cancer.

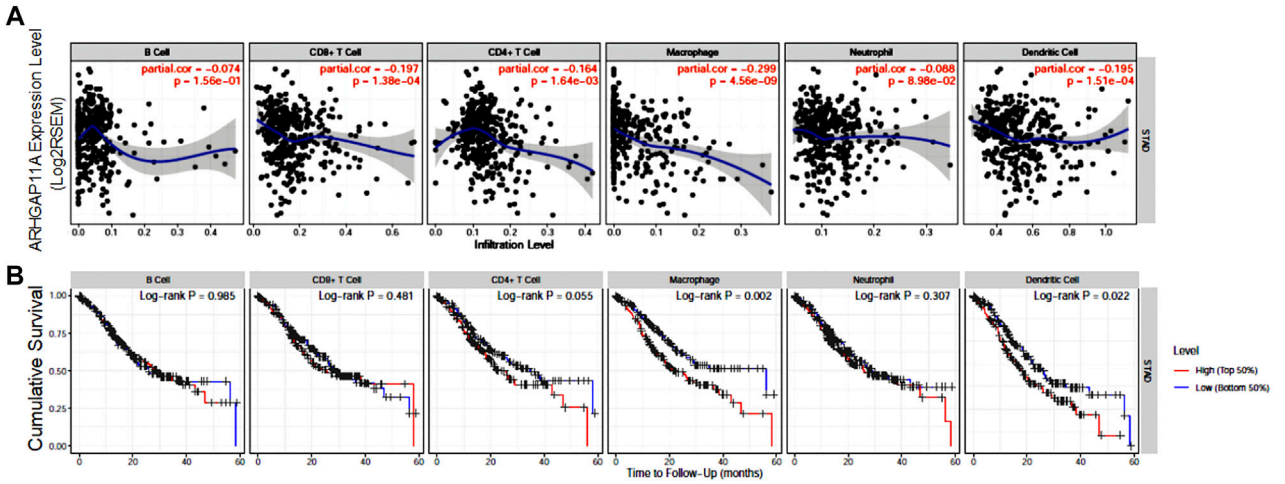


FIGURE 3 | ARHGAP11A expression is correlated with immune infiltrates in gastric cancer. **(A)** Correlation of ARHGAP11A expression with immune cell infiltration. **(B)** Prognostic value of immune cell infiltration in gastric cancer.

Relationships Between ARHGAP11A and Immune Markers Expression

We revealed the correlation between ARHGAP11A expression and gene markers of different types of immune cells by using the TIMER and GEPIA. As shown in **Table 2** and **Figures 4A–H**, ARHGAP11A expression was significantly correlated with multiple immune markers, in particular, including the macrophage/TAM marker (CCL2, NOS2, and MS4A4A), the neutrophil marker (CEACAM8 and CCR7), the natural killer cell marker (KIR2DL3, KIR2DL4, and KIR3DL3), the dendritic cell marker (HLA-DPB1 and CD1C), the Th1/Th2 marker (STAT1, IFNG and GATA3), the exhausted T cell marker (CTLA4 and GZMB). We further evaluated the relationship between ARHGAP11A expression and these immune markers in gastric cancer using the GEPIA. Similar results were shown in **Table 3**. For example, the M1 Macrophage marker NOS2 and the dendritic cell marker, HLA-DPB1 and CD1C, were significantly correlated with ARHGAP11A expression in gastric cancer. Taken together, ARHGAP11A might involve in infiltration of M1 Macrophages and dendritic cells.

Genetic Alteration, PPI Network and Enrichment Analyses of ARHGAP11A

We used the cBioPortal to analyze genetic changes in gastric cancer. Among 1,120 gastric cancer patients, ARHGAP11A was changed in 22 samples (2%), including 13 mutations, 5 amplifications and 4 deep deletions (**Figure 5A** and **Supplementary Figure S2**). In addition, ARHGAP11A mutations did not affect the prognosis of gastric cancer (**Supplementary Figures S3, 4**). The difference of prognosis between ARHGAP11A altered and unaltered group with no statistical significance might due to low frequency of ARHGAP11A alteration. In GeneMANIA analysis, molecular signals interacting with ARHGAP11A included MKI67, MEK2, DLGAP5, KIF14, KIF18B, AURKB, RHOBTB2, PLK4, KIFC1, CDC20, CENPF, SFN, PLK1, WDHD1, KIF2C, CCNB1, KIF20B, TTK, OIP5 and CCNA2 (**Figure 5B**). Genes from the PPI network and TIMER analysis were included in the GO function and KEGG pathway analyses by using Metascape. Results showed that ARHGAP11A and its interacting signals involved in regulating of lymphocyte activation, immune effector process, cell killing, myeloid leukocyte differentiation, antigen receptor-mediated signaling pathway and leukocyte apoptotic process (**Figure 5C**).

DISCUSSION

Dysregulation of Rho GTPases is identified in multiple cancers, and is associated with cancer development and malignant phenotypes. The activity of Rho GTPases (GDP/GTP cycling) is precisely controlled by regulators including GTPase-activating proteins (GAPs) (Lavanderos et al., 2020). Altered expression of GAPs is present in various cancers. ARHGAP10, a member of GAPs, is downregulated in ovarian and breast cancer (Luo et al., 2016; Li et al., 2019). ARHGAP5 is upregulated in metastatic colorectal cancers (Tian et al., 2020). In our study, we systematically analyzed the expression of ARHGAP11A in different tissue types. The expression of ARHGAP11A was found significantly higher in gastrointestinal cancers including stomach adenocarcinoma. Expression level of ARHGAP11A in the metastatic lesion of SKCM was higher than primary SKCM. In addition, the high expression level of ARHGAP11A was associated with better prognosis in gastric and blood cancer, while correlated with poor prognosis in lung, bladder, breast and soft tissue cancer. Taken together, ARHGAP11A revealed dual effects on different human cancers.

Interestingly, an interaction effect exists between ARHGAP11A expression and therapeutic strategy on the prognosis of gastric cancer. High expression level of

TABLE 2 | Correlation between ARHGAP11A and related immune markers in TIMER.

Cell types	Gene markers	Gastric cancer			
		None		Purity	
		Correlation	*p*	Correlation	*p*
$CD8^+$ T cell	CD8A	−0.0074	0.13	−0.0052	0.317
	CD8B	−0.0023	0.645	0.007	0.891
T cell general	CD3D	−0.0098	*	−0.042	0.415
	CD3E	−0.118	*	−0.064	0.214
	CD2	−0.056	0.252	−0.007	0.892
B cell	CD19	−0.164	***	−0.142	**
	CD79A	−0.252	***	−0.227	***
Monocyte	CD86	−0.02	0.68	0.023	0.655
	CSF1R	−0.112	*	−0.103	*
TAM	CCL2	−0.291	***	−0.282	***
	CD68	0.002	0.962	0.017	0.738
	IL10	−0.007	0.887	0.027	0.604
M1 Macrophage	NOS2	0.22	***	0.238	***
	IRF5	−0.051	0.297	−0.05	0.328
	PTGS2	0.035	0.479	0.046	0.367
M2 Macrophage	CD163	0.057	0.243	0.075	0.143
	VSIG4	−0.108	*	−0.096	0.0652
	MS4A4A	−0.112	*	−0.09	0.0786
Neutrophils	CEACAM8	0.217	***	0.224	***
	ITGAM	−0.064	0.19	−0.049	0.345
	CCR7	−0.233	***	−0.192	***
Natural killer cell	KIR2DL1	0.121	*	0.144	**
	KIR2DL3	0.127	**	0.145	**
	KIR2DL4	0.184	***	0.223	***
	KIR3DL1	0.06	0.224	0.05	0.328
	KIR3DL2	0.109	*	0.132	*
	KIR3DL3	0.155	*	0.158	**
	KIR2DS4	0.092	0.0621	0.108	*
Dendritic cell	HLA-DPB1	−0.169	***	−0.128	*
	HLA-DQB1	−0.083	0.0915	−0.03	0.564
	HLA-DRA	−0.06	0.226	−0.01	0.849
	HLA-DPA1	−0.108	*	−0.065	0.205
	CD1C	−0.347	***	−0.331	***
	NRP1	−0.065	0.187	−0.057	0.265
	ITGAX	0.067	0.172	0.113	*
Th1	TBX21	−0.009	0.85	0.026	0.611
	STAT4	−0.015	0.763	0.021	0.681
	STAT1	0.394	***	0.398	***
	IFNG	0.227	***	0.268	***
	TNF	0.01	0.845	0.045	0.388
Th2	GATA3	−0.207	***	−0.184	***
	STAT6	0.106	*	0.099	0.0548
	STAT5A	0.105	*	0.121	*
	IL13	0.03	0.543	0.047	0.359
Follicular helper T cell	BCL6	−0.171	**	−0.167	**
	IL21	0.161	*	0.2	***
T helper cell	STAT3	0.119	*	0.113	*
	IL17A	0.188	***	0.223	***
Regulatory T cell	FOXP3	0.059	0.228	0.102	*
	CCR8	0.082	0.0967	0.103	*
	STAT5B	0.036	0.466	0.037	0.470
	TGFB1	−0.169	**	−0.162	**
Exhausted T cell	PDCD1	0.011	0.819	0.051	0.325
	CTLA4	0.158	*	0.212	***
	LAG3	0.068	0.164	0.098	0.0561
	HAVCR2	0.061	0.212	0.095	0.0646
	GZMB	0.144	*	0.193	***

Purity: correlation adjusted by purity. *p < *0.05*, **p < *0.01*, ***p < *0.001*.

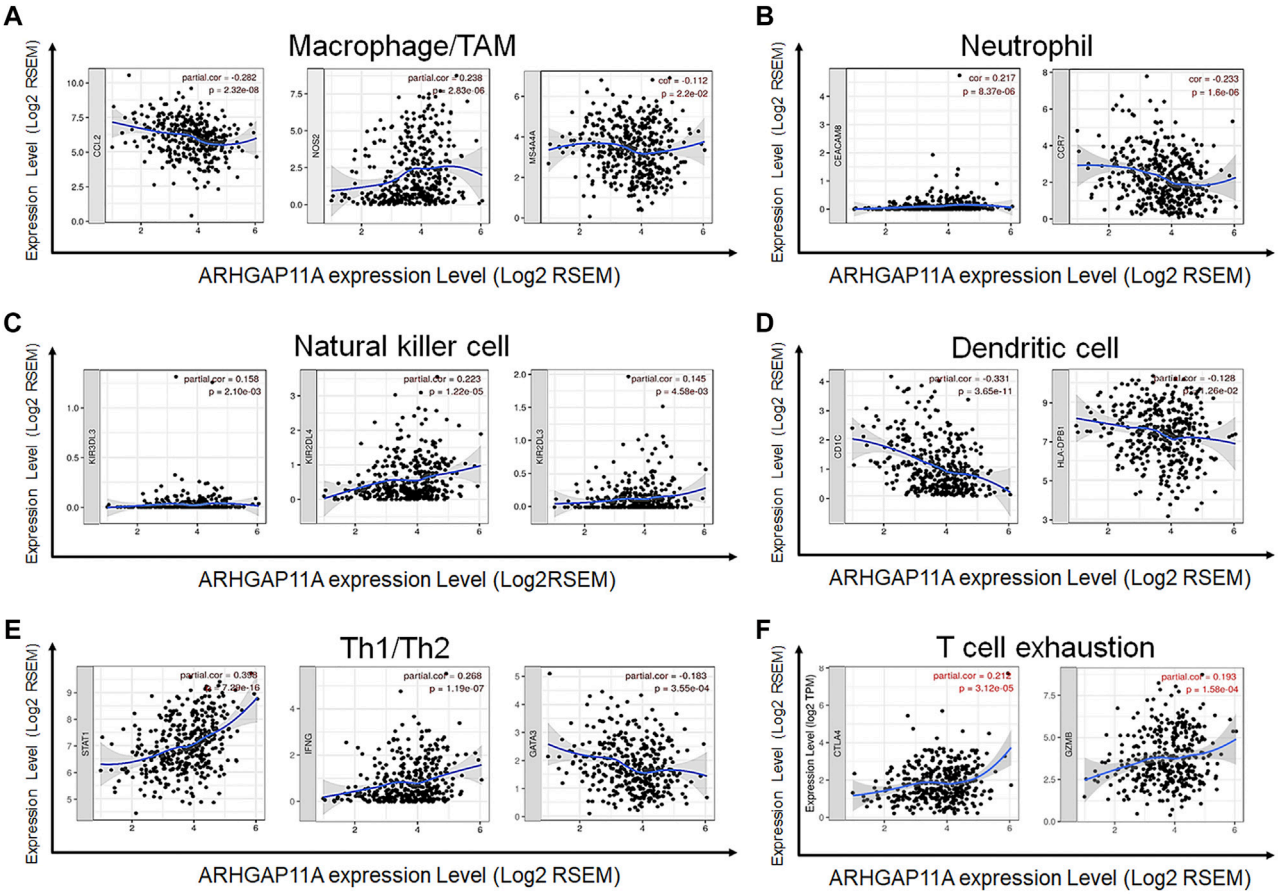


FIGURE 4 | Correlation between ARHGAP11A expression and immune markers in gastric cancer. **(A)** Markers of macrophages/TMAs. **(B)** Markers of neutrophils. **(C)** Markers of nature killer cells. **(D)** Markers of dendritic cells. **(E)** Markers of Th1/Th2 cells. **(F)** Markers of exhausted T cells.

TABLE 3 | Correlation between ARHGAP11A and related immune markers in GEPIA.

Cell types	Gene markers	Gastric tissues			
		Tumor		Non-tumor	
		R	*p*	R	*p*
T cell general	CD3E	−0.12	*	0.052	0.76
TAM	CCL2	−0.16	***	−0.37	*
M1 Macrophage	NOS2	0.15	**	0.2	0.24
Neutrophils	CCR7	−0.11	*	−0.035	0.84
Natural killer cell	KIR2DL3	0.114	**	0.140	0.41
	KIR2DL4	0.12	*	0.088	0.61
Dendritic cell	HLA-DPB1	−0.1	*	0.065	0.71
	CD1C	−0.24	***	−0.022	0.9
Th1	STAT1	0.47	***	0.21	0.23
	IFNG	0.13	**	−0.035	0.84
Th2	GATA3	−0.098	*	0.096	0.58
Tfh	IL21	0.11	*	0.011	0.95
T cell exhaustion	CTLA4	0.28	***	−0.028	0.87

*p < *0.05*, **p < *0.01*, ***p < *0.001*.

ARHGAP11A was associated with better prognosis when the patient treated with surgery alone, while in contrast, associated with worse prognosis when treated with surgery and 5-Fu based adjuvant chemotherapy. Microsatellite instability (MSI) is a mature biomarker for predicting the efficacy of immune checkpoint inhibitors. Similar effects were reported when MSI status was applied to predict the chemo-sensitivity in locally advanced colorectal cancer. Efficacy of 5-Fu based adjuvant chemotherapy was significantly different in MSI-H and MSI-L/MSS tumors (Ribic et al., 2003; Klingbiel et al., 2015). We consequently explored the correlation between ARHGAP11A expression and immune infiltrates in gastric cancer.

The lymphocyte in the immune microenvironment is a predictor of sentinel lymph node metastasis and patient survival (Azimi et al., 2012). The evaluation of immune infiltrates in gastric cancer showed that ARHGAP11A expression was significantly associated $CD8^+$ T cells, $CD4^+$ T cells, macrophages and dendritic cells. Immune genes correlated with ARHGAP11A expression included CCL2, NOS2, CCR7, KIR2DL3, KIR2DL4, HLA-DPB1, CD1C, STAT1, IFNG, and GATA3 and CTLA4. CCL2 is an inflammatory chemokine that promotes the recruitment of Tumor-associated macrophages (TAMs) to tumor sites (Nakatsumi et al., 2017). Chen C, et al. identified a long noncoding RNA, termed Lymph Node Metastasis Associated

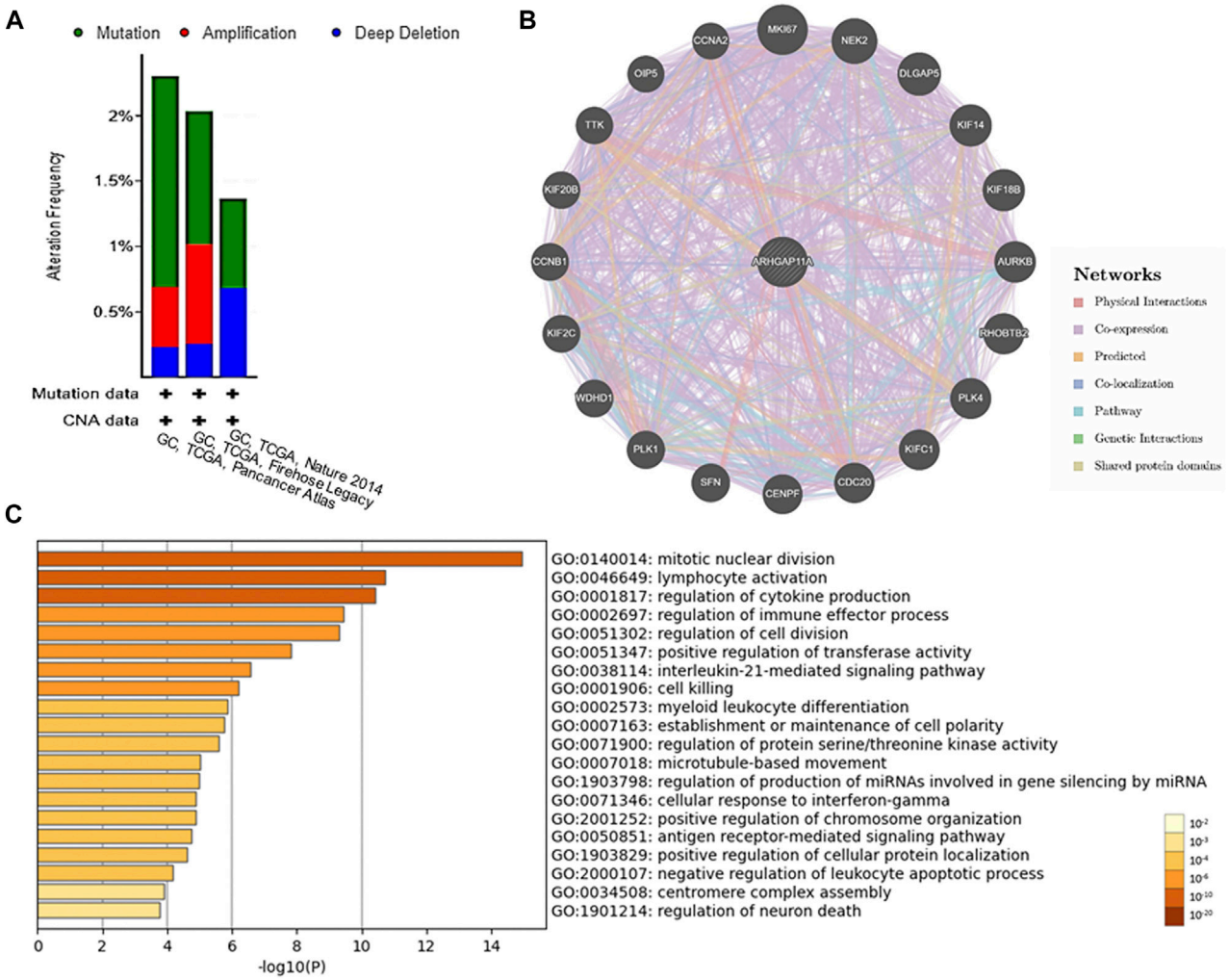


FIGURE 5 | Genetic alteration, PPI network and enrichment analyses of ARHGAP11A. **(A)** Genetic alteration of ARHGAP11A in gastric cancer. **(B)** PPI network of ARHGAP11A in GeneMANIA. **(C)** A heat map of GO function and KEGG pathway analysis of ARHGAP11A and its interacting proteins.

Transcript 1 (LNMAT1). Mechanistically, LNMAT1 epigenetically activates CCL2 expression and recruits macrophages into bladder cancer, which promotes lymphatic metastasis via VEGF-C excretion (Chen et al., 2018). In our ongoing parallel study, high expression of ARHGAP11A appeared more frequently in gastric cancer with lymph node metastasis. Silencing ARHGAP11A *in vitro*, resulting in the decrease of the invasive ability of gastric cancer cells to lymphatic endothelial cells (unpublished data). ARHGAP11A might play a key role in lymph node metastasis of gastric cancer. Whether LNMAT1 and VEGF-C involves in the biomolecular process needs to be verified in the future study. M1 macrophage marker NOS2 can activate macrophages and causes tumor cell death (Brune et al., 2017). KIR2DL3 and KIR2DL4 are transmembrane glycoproteins expressed by natural killer cells (Gomez-Luque et al., 2021). We found that expression of ARHGAP11A was correlated with NOS2, KIR2DL3 and KIR2DL4. Whether ARHGAP11A involves in regulating macrophages and natural kill cells is worth to be explored. Studies have shown that T cell infiltration defined immune-evasive environment in gastric cancer patients (Gu et al., 2020). In our study, the expression of ARHGAP11A was correlated with multiple T cell markers including STAT1, IFNG, GATA3, and CTLA4. ARHGAP11A might involve in the T cell infiltration of gastric cancer.

Mutations of Rho GTPases have been identified in various cancers. *Rac1* mutation was found in 4–9% of melanomas (Hodis et al., 2012). *RhoA* mutations were identified in over half of angioimmunoblastic T cell lymphomas (Sakata-Yanagimoto et al., 2014). In our study, genetic alterations of *ARHGAP11A* were identified in 2% of gastric cancer. Genes interacting with *ARHGAP11A* included *DLGAP5*, *KIF14*, *AURKB*, and *TTK*. Enhanced expression of DLGAP5 is observed in colorectal cancer. It defines a more aggressive type of colorectal cancer

(Branchi et al., 2019). *KIF14* is a potential oncogene, promotes gastric cancer progression and metastasis (Yang et al., 2019). *AURKB* and *TTK* participate in chromosomes segregation during mitosis via regulating kinetochore metaphase signaling (Su et al., 2021). Gene functional enrichment analysis showed ARHGAP11A and its interacting proteins involved in numerous processes including lymphocyte activation, cell division, cell killing, immune effector process regulating, and myeloid leukocyte differentiation.

There are still some limitations in our study. Kaplan Meier estimates are unadjusted in the Prognoscan and the results might be biased. Subgroup analysis is needed to confirm the prognostic value of ARHGAP11A expression in various types of cancer. More *in vivo* and *in vitro* experiments are needed to verify the abovementioned bioinformatic findings, especially the correlation between ARHGAP11A expression and immune infiltrates. On the other hand, the detailed mechanisms of ARHGAP11A in regulating gastric cancer metastasis needs further study.

CONCLUSION

In summary, ARHGAP11A might be a crucial regulator of immune infiltrates and a valuable prognostic marker in patients with gastric cancer. Additional studies are needed to validate its role both *in vitro* and *in vivo*.

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

BF and XW designed the analytical strategies. The data analyses was conducted by BF, KJ, ZB, JZ, HY, and JL. BF wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

This work is supported by the National Nature Science Foundation of China No. 81402308 and Science Foundation of Peking University Cancer Hospital No. 2021-24 and No.A001546.

ACKNOWLEDGMENTS

We would like to acknowledge the public databases including Oncomine, GEPIA, TIMER, Kaplan-Meier plotter and PrognoScan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.720645/full#supplementary-material

Supplementary Figure 1 | Cytoplasmic expression of ARHGAP11A in various types of cancer.

Supplementary Figure 2 | Genetic alterations of ARHGAP11A in gastric cancer from TCGA database.

Supplementary Figure 3 | The overall survival of gastric cancer patients with/ without *ARHGAP11A* alteration.

Supplementary Figure 4 | The disease-free survival of gastric cancer patients with/ without *ARHGAP11A* alteration.

REFERENCES

Azimi, F., Scolyer, R. A., Rumcheva, P., Moncrieff, M., Murali, R., McCarthy, S. W., et al. (2012). Tumor-infiltrating Lymphocyte Grade Is an Independent Predictor of sentinel Lymph Node Status and Survival in Patients with Cutaneous Melanoma. *Jco* 30, 2678–2683. doi:10.1200/jco.2011.37.8539

Balkwill, F. R., Capasso, M., and Hagemann, T. (2012). The Tumor Microenvironment at a Glance. *J. Cel Sci* 125, 5591–5596. doi:10.1242/jcs.116392

Brambilla, E., Le Teuff, G., Marguet, S., Lantuejoul, S., Dunant, A., Graziano, S., et al. (2016). Prognostic Effect of Tumor Lymphocytic Infiltration in Resectable Non-small-cell Lung Cancer. *Jco* 34, 1223–1230. doi:10.1200/jco.2015.63.0970

Branchi, V., García, S. A., Radhakrishnan, P., Győrffy, B., Hissa, B., Schneider, M., et al. (2019). Prognostic Value of DLGAP5 in Colorectal Cancer. *Int. J. Colorectal Dis.* 34, 1455–1465. doi:10.1007/s00384-019-03339-6

Brüne, B., Courtial, N., Dehne, N., Syed, S. N., and Weigert, A. (2017). Macrophage NOS2 in Tumor Leukocytes. *Antioxid. Redox Signaling* 26, 1023–1043. doi:10.1089/ars.2016.6811

Bruni, D., Angell, H. K., and Galon, J. (2020). The Immune Contexture and Immunoscore in Cancer Prognosis and Therapeutic Efficacy. *Nat. Rev. Cancer* 20, 662–680. doi:10.1038/s41568-020-0285-7

Cao, W., Chen, H.-D., Yu, Y.-W., Li, N., and Chen, W.-Q. (2021). Changing Profiles of Cancer burden Worldwide and in China: a Secondary Analysis of the Global Cancer Statistics 2020. *Chin. Med. J. (Engl)* 134, 783–791. doi:10.1097/cm9.0000000000001474

Chen, C., He, W., Huang, J., Wang, B., Li, H., Cai, Q., et al. (2018). LNMAT1 Promotes Lymphatic Metastasis of Bladder Cancer via CCL2 Dependent Macrophage Recruitment. *Nat. Commun.* 9, 3826. doi:10.1038/s41467-018-06152-x

Cui, J., Chen, Y., Chou, W.-C., Sun, L., Chen, L., Suo, J., et al. (2011). An Integrated Transcriptomic and Computational Analysis for Biomarker Identification in Gastric Cancer. *Nucleic Acids Res.* 39, 1197–1207. doi:10.1093/nar/gkq960

Dai, B., Zhang, X., Shang, R., Wang, J., Yang, X., Zhang, H., et al. (2018). Blockade of ARHGAP11A Reverses Malignant Progress via Inactivating Rac1B in Hepatocellular Carcinoma. *Cell Commun Signal* 16, 99. doi:10.1186/s12964-018-0312-4

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. Signaling* 6, pl1. doi:10.1126/scisignal.2004088

Gomez-Luque, J. M., Urrutia-Maldonado, E., Munoz de Rueda, P., Abril-Molina, A., and Ocete-Hita, E. (2021). Killer Immunoglobulin-like Receptor and Cancer. *Pediatr. (Barc)*.

Gu, Y., Chen, Y., Jin, K., Cao, Y., Liu, X., Lv, K., et al. (2020). Intratumoral CD103+CD4+ T Cell Infiltration Defines Immunoevasive Contexture and Poor Clinical Outcomes in Gastric Cancer Patients. *Oncoimmunology* 9, 1844402. doi:10.1080/2162402x.2020.1844402

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J.-P., et al. (2012). A Landscape of Driver Mutations in Melanoma. *Cell* 150, 251–263. doi:10.1016/j.cell.2012.06.024

Klingbiel, D., Saridaki, Z., Roth, A. D., Bosman, F. T., Delorenzi, M., and Tejpar, S. (2015). Prognosis of Stage II and III colon Cancer Treated with Adjuvant 5-fluorouracil or FOLFIRI in Relation to Microsatellite Status: Results of the PETACC-3 Trial. *Ann. Oncol.* 26, 126–132. doi:10.1093/annonc/mdu499

Lavanderos, B., Silva, I., Cruz, P., Orellana-Serradell, O., Saldías, M. P., and Cerda, O. (2020). TRP Channels Regulation of Rho GTPases in Brain Context and Diseases. *Front. Cel Dev. Biol.* 8, 582975. doi:10.3389/fcell.2020.582975

Lawson, C. D., Fan, C., Mitin, N., Baker, N. M., George, S. D., Graham, D. M., et al. (2016). Rho GTPase Transcriptome Analysis Reveals Oncogenic Roles for Rho GTPase-Activating Proteins in Basal-like Breast Cancers. *Cancer Res.* 76, 3826–3837. doi:10.1158/0008-5472.can-15-2923

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, e108–e110. doi:10.1158/0008-5472.can-17-0307

Li, Y., Zeng, B., Li, Y., Zhang, C., and Ren, G. (2019). Downregulated Expression of ARHGAP10 Correlates with Advanced Stage and High Ki-67 index in Breast Cancer. *PeerJ* 7, e7431. doi:10.7717/peerj.7431

Luo, N., Guo, J., Chen, L., Yang, W., Qu, X., and Cheng, Z. (2016). ARHGAP10, Downregulated in Ovarian Cancer, Suppresses Tumorigenicity of Ovarian Cancer Cells. *Cell Death Dis* 7, e2157. doi:10.1038/cddis.2015.401

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

Müller, P. M., Rademacher, J., Bagshaw, R. D., Wortmann, C., Barth, C., van Unen, J., et al. (2020). Systems Analysis of RhoGEF and RhoGAP Regulatory Proteins Reveals Spatially Organized RAC1 Signalling from Integrin Adhesions. *Nat. Cel Biol* 22, 498–511. doi:10.1038/s41556-020-0488-x

Nakatsumi, H., Matsumoto, M., and Nakayama, K. I. (2017). Noncanonical Pathway for Regulation of CCL2 Expression by an mTORC1-FOXK1 Axis Promotes Recruitment of Tumor-Associated Macrophages. *Cel Rep.* 21, 2471–2486. doi:10.1016/j.celrep.2017.11.014

Porter, A. P., Papaioannou, A., and Malliri, A. (2016). Deregulation of Rho GTPases in Cancer. *Small GTPases* 7, 123–138. doi:10.1080/21541248.2016.1173767

Refolo, M. G., Lotesoriere, C., Messa, C., Caruso, M. G., and D'Alessandro, R. (2020). Integrated Immune Gene Expression Signature and Molecular Classification in Gastric Cancer: New Insights. *J. Leukoc. Biol.* 108, 633–646. doi:10.1002/jlb.4mr0120-221r

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, 166–180. doi:10.1593/neo.07112

Ribas, A., and Wolchok, J. D. (2018). Cancer Immunotherapy Using Checkpoint Blockade. *Science* 359, 1350–1355. doi:10.1126/science.aar4060

Ribic, C. M., Sargent, D. J., Moore, M. J., Thibodeau, S. N., French, A. J., Goldberg, R. M., et al. (2003). Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for colon Cancer. *N. Engl. J. Med.* 349, 247–257. doi:10.1056/nejmoa022289

Sakata-Yanagimoto, M., Enami, T., Yoshida, K., Shiraishi, Y., Ishii, R., Miyake, Y., et al. (2014). Somatic RHOA Mutation in Angioimmunoblastic T Cell Lymphoma. *Nat. Genet.* 46, 171–175. doi:10.1038/ng.2872

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

Saoudi, A., Kassem, S., Dejean, A., and Gaud, G. (2014). Rho-GTPases as Key Regulators of T Lymphocyte Biology. *Small GTPases* 5 (4), e983862. doi:10.4161/sgtp.28208

Su, T., Qin, X. Y., Dohmae, N., Wei, F., Furutani, Y., Kojima, S., et al. (2021). Inhibition of Ganglioside Synthesis Suppressed Liver Cancer Cell Proliferation through Targeting Kinetochore Metaphase Signaling. *Metabolites* 11 (3), 167. doi:10.3390/metabo11030167

Sudo, T., Nishida, R., Kawahara, A., Saisho, K., Mimori, K., Yamada, A., et al. (2017). Clinical Impact of Tumor-Infiltrating Lymphocytes in Esophageal Squamous Cell Carcinoma. *Ann. Surg. Oncol.* 24, 3763–3770. doi:10.1245/s10434-017-5796-4

Sun, D., Cao, M., Cao, M., Li, H., He, S., and Chen, W. (2020). Cancer burden and Trends in China: A Review and Comparison with Japan and South Korea. *Chin. J. Cancer Res.* 32, 129–139. doi:10.21147/j.issn.1000-9604.2020.02.01

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

Szász, A. M., Lánczky, A., Nagy, Á., Förster, S., Hark, K., Green, J. E., et al. (2016). Cross-validation of Survival Associated Biomarkers in Gastric Cancer Using Transcriptomic Data of 1,065 Patients. *Oncotarget* 7, 49322–49333. doi:10.18632/oncotarget.10337

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

Tian, T., Chen, Z.-H., Zheng, Z., Liu, Y., Zhao, Q., Liu, Y., et al. (2020). Investigation of the Role and Mechanism of ARHGAP5-Mediated Colorectal Cancer Metastasis. *Theranostics* 10, 5998–6010. doi:10.7150/thno.43427

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

Xu, J., Zhou, X., Wang, J., Li, Z., Kong, X., Qian, J., et al. (2013). RhoGAPs Attenuate Cell Proliferation by Direct Interaction with P53 Tetramerization Domain. *Cel Rep.* 3, 1526–1538. doi:10.1016/j.celrep.2013.04.017

Yang, L., Ying, X., Ying, X., Liu, S., Lyu, G., Xu, Z., et al. (2020). Gastric Cancer: Epidemiology, Risk Factors and Prevention Strategies. *Chin. J. Cancer Res.* 32, 695–704. doi:10.21147/j.issn.1000-9604.2020.06.03

Yang, Z., Li, C., Yan, C., Li, J., Yan, M., Liu, B., et al. (2019). KIF14 Promotes Tumor Progression and Metastasis and Is an Independent Predictor of Poor Prognosis in Human Gastric Cancer. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1865, 181–192. doi:10.1016/j.bbadis.2018.10.039

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, 737–750. doi:10.1158/2326-6066.cir-18-0436

Zhao, Y., Xu, E., Yang, X., Zhang, Y., Chen, H., Wang, Y., et al. (2020). Tumor Infiltrative Growth Pattern Correlates with the Immune Microenvironment and Is an Independent Factor for Lymph Node Metastasis and Prognosis in Stage T1 Esophageal Squamous Cell Carcinoma. *Virchows Arch.* 477, 401–408. doi:10.1007/s00428-020-02801-z

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

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 Fan, Ji, Bu, Zhang, Yang, Li 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.

GLOSSARY

ACC Adrenocortical carcinoma
BLCA Bladder urothelial carcinoma
BRCA Breast invasive carcinoma
CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL Cholangiocarcinoma
COAD Colon adenocarcinoma
DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DMFS Distant Metastasis Free Survival
DRFS Distant relapse free survival
DSS Disease Specific Survival
ESCA Esophageal carcinoma
GAPs GTPase-activating proteins
GBM Glioblastoma multiforme
GEPIA Gene Expression Profiling Interactive Analysis
HNSC Head and neck squamous cell carcinoma
HR Hazard ratio
KICH Kidney chromophobe
KIRC Kidney renal clear cell carcinoma
KIRP Kidney renal papillary cell carcinoma
LAML Acute Myeloid Leukemia
LGG Brain Lower Grade Glioma
LIHC Liver hepatocellular carcinoma
LUAD Lung adenocarcinoma
LUSC Lung squamous cell carcinoma
MESO Mesothelioma Rectum adenocarcinoma
MSI Microsatellite instability
OS Overall survival
OV Ovarian serous cystadenocarcinoma
PFS Progression free survival
PPI protein-protein interaction
PRAD Prostate adenocarcinoma
READ Rectum adenocarcinoma
RFS Relapse Free Survival
RhoGAPs Rho GTPase-activating proteins
SARC Sarcoma
SKCM Skin cutaneous melanoma
STAD Stomach adenocarcinoma
TAMs Tumor-associated macrophages
TCGA The Cancer Genome Atlas
TGCT Testicular Germ Cell Tumors
THCA Thyroid carcinoma
THYM Thymoma
TILs tumor-infiltrating lymphocytes
TIMER Tumor IMmune Estimation Resource
UCEC Uterine corpus endometrial carcinoma
UCS Uterine Carcinosarcoma
UVM Uveal Melanoma

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 20 October 2021
doi: 10.3389/fmolb.2021.753563

Aberrant Methylation and Immune Microenvironment Are Associated With Overexpressed Fibronectin 1: A Diagnostic and Prognostic Target in Head and Neck Squamous Cell Carcinoma

Surui Sheng[1†], *Bing Guo*[2†], *Zhentao Wang*[3], *Zhihua Zhang*[3], *Jieyu Zhou*[3*] *and Zirong Huo*[3*]

[1]*Shanghai Key Laboratory of Stomatology, Department of Oral and Maxillofacial—Head and Neck Oncology, Shanghai Ninth People's Hospital, National Center for Stomatology, National Clinical Research Center for Oral Diseases, College of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai, China,* [2]*Shanghai Key Laboratory of Tissue Engineering, Department of Plastic and Reconstructive Surgery, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China,* [3]*Department of Otorhinolaryngology Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China*

Background: Fibronectin 1 (FN1) is involved in cell adhesion and migration processes such as metastasis, wound healing, embryogenesis, blood coagulation, and host defense. However, the role of FN1 in the diagnosis and prognosis of head and neck squamous cell carcinoma (HNSCC) is far from understood.

Methods: FN1 expression profiles and clinical parameters from multiple HNSCC datasets were applied to evaluate the association between FN1 expression and HNSCC survival. We also identified FN1 expression in the mRNA and protein levels in 20 pairs of clinical samples by quantitative polymerase chain reaction (qPCR) and immunohistochemistry. Receiver operator characteristic (ROC) analysis was used to demonstrate the potential diagnostic value of FN1 in HNSCC. Aberrant methylation PPI networks were established using multiple bioinformatic tools based on TCGA database. The immune microenvironment and levels of immune checkpoints were investigated between groups with high and low FN1 expression.

Results: FN1 was significantly upregulated in HNSCC compared with para-carcinoma tissues on the basis of TCGA database and our clinical samples. Univariate and multivariate Cox regression analysis revealed that FN1 could be an independent indicator for prognosis of HNSCC. GO enrichment and KEGG pathway analysis demonstrated that cell adhesion, focal adhesion, and the PI3K-Akt signaling pathway might be involved in the potential mechanisms of FN1's prognostic performance in HNSCC. Methylation of FN1 was also higher and closely associated with poorer survival in HNSCC. In addition, FN1 expression was positively correlated with three DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B). Furthermore, FN1 was positively associated with $CD4^+$ T cells, endothelial cells, macrophages, and NK cells and negatively correlated with $CD8^+$ T cells

OPEN ACCESS

Edited by:
Hossain Shekhar,
University of Dhaka, Bangladesh

Reviewed by:
Rupkatha Mukhopadhyay,
Johns Hopkins Medicine, United States
Yuanping Xiong,
First Affiliated Hospital of Nangchang university, China

***Correspondence:**
Jieyu Zhou
116215@sh9hospital.org.cn
Zirong Huo
zironghuo@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *05 August 2021*
Accepted: *20 September 2021*
Published: *20 October 2021*

Citation:
Sheng S, Guo B, Wang Z, Zhang Z, Zhou J and Huo Z (2021) Aberrant Methylation and Immune Microenvironment Are Associated With Overexpressed Fibronectin 1: A Diagnostic and Prognostic Target in Head and Neck Squamous Cell Carcinoma. Front. Mol. Biosci. 8:753563. doi: 10.3389/fmolb.2021.753563

Conclusion: FN1 might be an independent prognostic biomarker for HNSCC patients. Hypermethylation, the aberrant proportions of immune cells, and the PI3K/Akt signaling pathway might be involved in the mechanism of FN1's oncogene role in HNSCC.

Keywords: head and neck squamous cell carcinoma, fibronectin 1, prognosis, immune microenvironment, methylation

INTRODUCTION

Head and neck cancer is a worldwide deadly disease with an estimated annual incidence of more than 830,000 (Bray et al., 2018). The 5-year survival rate is only 40–50% and even lower among advanced cancer patients (Chauhan et al., 2015), although the examination and treatment has improved in recent decades. Head and neck squamous cell carcinoma (HNSCC) is the most common histological type, which accounts for about 4% of all new cancer diagnoses in the United States. It has been recognized as a highly heterogeneous malignant tumor, which can derive from various anatomical sites in the upper airway and digestive tract, including the mouth, pharynx, and larynx (Siegel et al., 2016). The heterogeneity renders clinical prognosis difficult to predict (Leemans et al., 2018). Clinical parameters such as TNM classification are commonly relied on for predicting the outcomes, which is far from accurate. Thus, there is an urgent necessity for a better understanding of the molecular alterations to make accurate early HNSCC diagnoses, improve the prognosis, and provide new therapeutic strategies.

Fibronectin 1 (FN1) is being increasingly considered as a part of tumor pathogenesis and contributes to various malignant behaviors in solid tumors (Kumra and Reinhardt, 2016; Kujawa et al., 2020). Some studies have revealed that overexpression of fibronectin in the pre-metastatic niche facilitates the adhesion of bone marrow–derived cells, which promotes tumor cell migration and cancer metastasis by providing support for tumor cells to escape from the primary site (Shinde et al., 2018; Kujawa et al., 2020). The deposition of fibronectin into the tumor extracellular matrix (ECM), followed by the formation of fibrin–fibronectin complexes, has been shown to facilitate tumor angiogenesis, proliferation, and metastasis (Malik et al., 2010). Therefore, the highly expressed fibronectin is a potential biomarker for the early diagnosis of malignant tumors and micrometastasis. Previous studies of HNSCC samples have revealed that FN1 is upregulated in the tumor stromal region and at the invasive front of the tumor (Kosmehl et al., 1999). However, whether FN1 expression could be used as a diagnostic or prognostic biomarker in HNSCC is not sufficiently understood.

In the present study, we assessed the expression and prognosis of FN1 in HNSCC for the first time. We identified FN1 as a reliable prognostic biomarker in HNSCC based on TCGA dataset and further validated its capability in our clinical samples. Furthermore, aberrant methylation, the immune microenvironment, and the protein–protein interaction (PPI) network were investigated, which could be interpreted with regard to the underlying mechanism of FN1's role in HNSCC patients.

MATERIALS AND METHODS

Microarray Data Collection and FN1 Filtering

We searched the gene expression profiles from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/) dataset (Modhukur et al., 2018) according to the following criteria: 1. the samples were verified by pathology as head and neck squamous cell carcinoma (including the oral cavity, oropharynx, nasopharynx, larynx, laryngeal pharynx, nasal cavity, and paranasal sinus). 2. The profiles included samples of squamous cell carcinoma and normal tissues of para-carcinoma. 3. The gene expression profiles of cell lines or experimental animals were excluded. As a result, two gene expression profiles, including GSE40290 and GSE55550, were selected. Then the database from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) was included as well.

Differentially expressed genes (DEGs) were filtered with the threshold of adjusted $p < 0.01$ and log (fold change) > 2 or log (fold change) <−2. There were 339, 251, and 285 DEGs in TCGA, GSE40290, and GSE55550, respectively. Two-crossing or three-crossing genes were investigated and visualized using Bioinformatics and Evolutionary Genomics (http://bioinformatics.psb.ugent.be/webtools/Venn/) among different datasets. Further screening was performed according to the following criteria: 1) there was a significant difference in the expression of DEGs between HNSCC and para-carcinoma tissues in all of the three datasets. 2) Kaplan–Meier (KM) survival analysis revealed an association between DEG expression and the prognosis of HNSCC. $p < 0.05$ was set as the cut-off value for filtering.

Expression Analysis of FN1

TCGA database was applied to investigate the FN1 expression at the mRNA level. A comprehensive investigation on the basis of the Human Protein Atlas (THPA, https://www.proteinatlas.org/) was performed to evaluate the protein expression level of FN1. Twenty patients who were diagnosed with HNSCC were retrospectively collected in Shanghai Ninth People's Hospital Affiliated to Shanghai Jiaotong University School of Medicine from April 2021 to June 2021, which was performed in accordance with the ethical principles described in the Declaration of Helsinki and approved by the bioethics committee of Shanghai Ninth People's Hospital Affiliated to Shanghai Jiaotong University School of Medicine (approval no. SH9H-2021-T300-1). Twenty pairs of samples were classified as the primary tumor and the para-carcinoma group. RNA was extracted and reversed to cDNA. Quantitative polymerase chain reaction (qPCR) was performed to explore the mRNA levels of

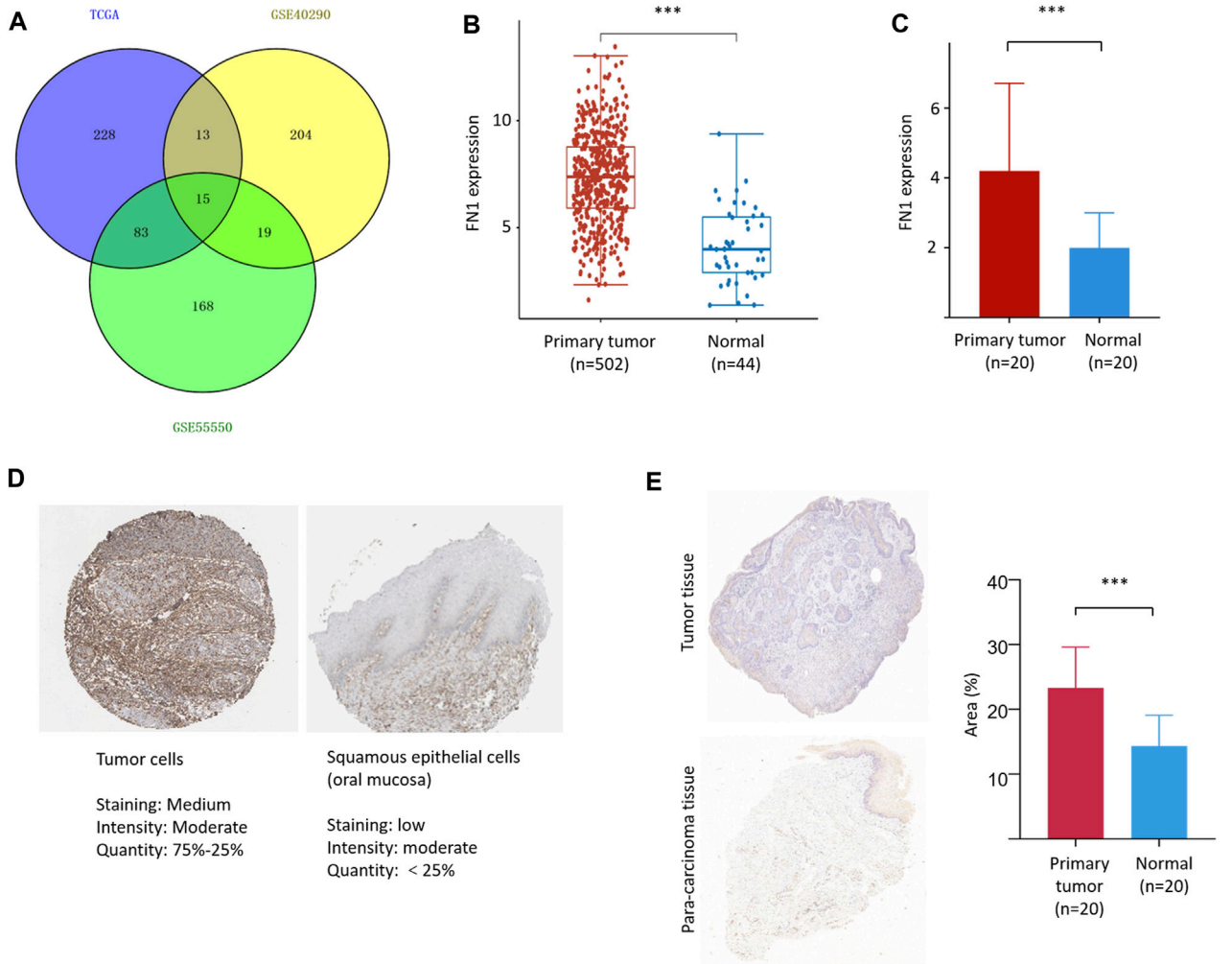


FIGURE 1 | FN1 filtering and its expression in HNSCC. **(A)** A total of 15 three-crossing and 115 two-crossing DEGs were filtered among three public datasets. The mRNA level of FN1 was significantly higher in HNSCCs than in the normal para-carcinoma tissues from TCGA database **(B)** and our patients **(C)**. **(D)** The protein level of FN1 was much higher in HNSCC **(left)** than in the normal tissues of oral mucosa **(right)**, both from THPA. **(E)** Representative pictures of FN1 expression between different groups examined by IHC were shown **(left)** and verified by density value **(right)** and our patients.

FN1. The following primers were used in this study. FN1: forward, CAAATGGTTCAGCCCCAGTCC, reverse, GTCCGC TCCCACTGTTGATTTATC. GAPDH: forward, CATGAGAAG TATGACAACAGCCT, reverse, AGTCCTTCCACGATACCA AAGT. Immunohistochemistry (IHC) was performed as previously described (Huo et al., 2017). Sections were incubated with anti-FN1 antibody (1:600, 66042-1-lg, Proteintech Group, Rosemont, PA, USA) overnight at 4°C. The density of FN1 expression was demonstrated by the percentage of staining cell infiltration.

Prognostic and Diagnostic Analysis of FN1

Raw count of RNA-sequencing data and clinical information of HNSCC were downloaded from TCGA dataset. The one-way ANOVA test was executed to explore the differences in FN1 expression between HNSCC groups classified by multiple clinical characteristics on the basis of TCGA database, including age, gender, tumor stage, lymph node metastasis, TNM classification, and smoking habit. The Kaplan–Meier (KM) survival analysis was applied to evaluate the correlation between FN1 expression and 5-year overall survival (OS). To further explore whether FN1 expression was independent of the other clinical variables in HNSCC, univariate and multivariate Cox regression analysis was performed. The receiver operating characteristic (ROC) curve was used to assess the diagnostic ability of FN1 in HNSCC. Further comparison of the diagnostic capability between FN1 and the epidermal growth factor receptor (EGFR), another common biomarker in malignant tumors, was also applied. Then, ROC analysis was applied to predict the performance of FN1 expression and other clinical parameters, such as TNM classification, in 5-year OS outcomes of HNSCC patients.

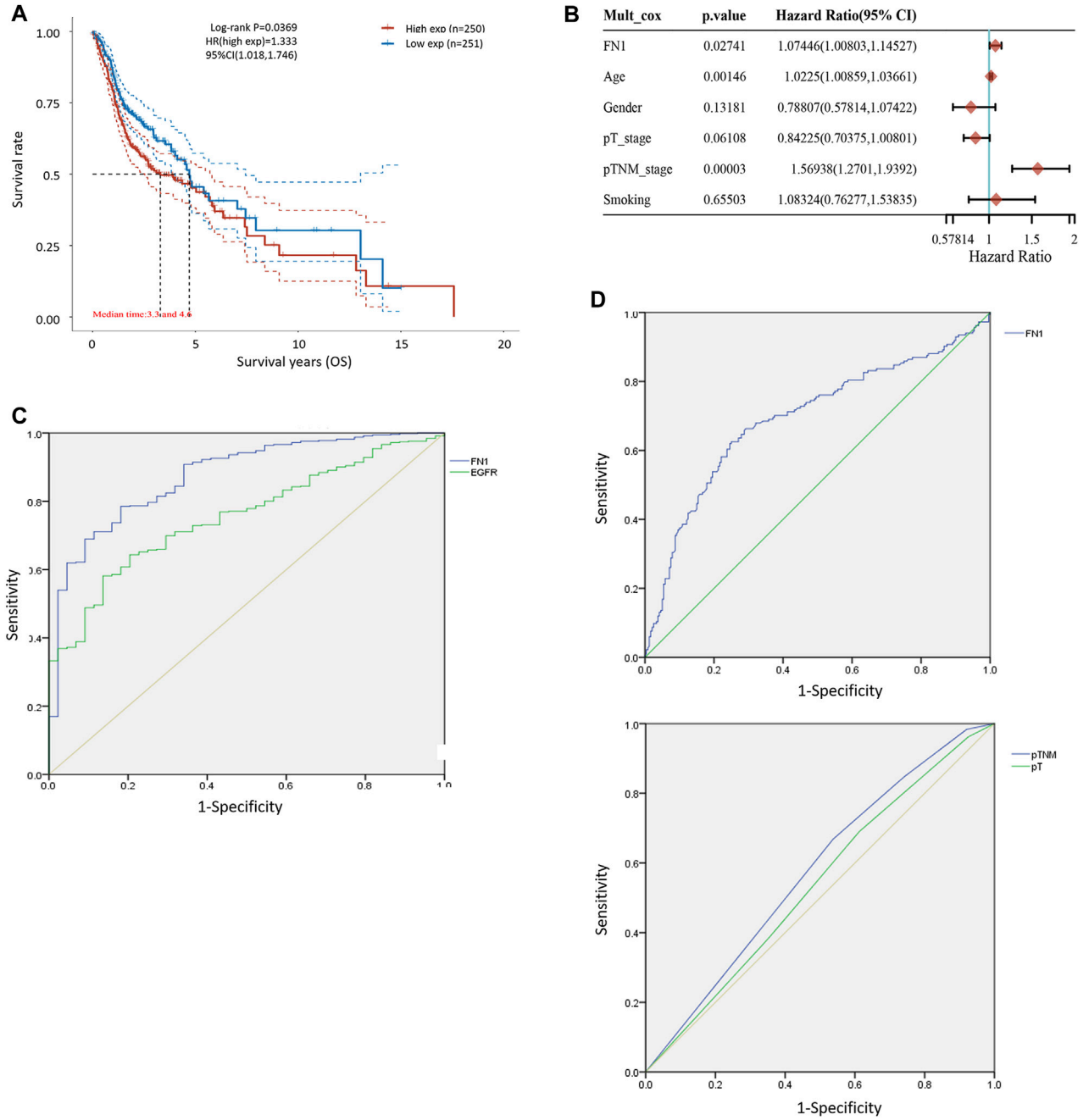


FIGURE 2 | Diagnostic and prognostic capability of FN1 in HNSCC was examined using the ROC curve. **(A)** KM survival analysis showed that higher expression of FN1 was associated with worse prognosis of HNSCC patients. **(B)** Age, TNM classification, and FN1 expression were important prognostic factors for HNSCC patients examined by multivariate Cox analysis. **(C)** AUC of FN1 expression was higher than that of EGFR. **(D)** AUC of FN1 expression **(upper)** was higher than that of TNM classification or the tumor stage **(lower)**.

Methylation Analysis of FN1

Methylation of FN1 was detected using UALCAN (http://ualcan.path.uab.edu/) (Chandrashekar et al., 2017), a web tool which could evaluate epigenetic regulation of gene expression by promoter methylation based on TCGA dataset. Furthermore, MethSurv (https://biit.cs.ut.ee/methsurv/) (Edgar et al., 2002), which could provide univariable and multivariable survival analysis based on DNA methylation biomarkers using TCGA (Edgar et al., 2002), was performed to evaluate the association between the position distribution of methylation around CpG islands and the prognosis of the HNSCC patient. Moreover, we also explored the correlation between three DNA

methyltransferases (DNMT1, DNMT3A, and DNMT1) and FN1 expression on the basis of TCGA database.

Immune Cell Environment Analysis

To evaluate immune infiltrations in different HNSCC patient groups divided by FN1 expression levels, we applied unsupervised clustering and CIBERSORT algorithms such as immunedeconv, an R package which integrates six state-of-the-art algorithms, including MCP-counter, TIMER, CIBERSORT, xCell, quanTIseq, and EPIC (Finotello et al., 2019). The scores or proportions of tumor-infiltrating cells and the expression values of immune-checkpoint–relevant genes were compared among groups with different FN1 expression levels. The TIDE algorithm, a computational method to model two primary mechanisms of tumor immune evasion (Jiang et al., 2018), was performed to predict the different responses of $FN1^{high}$ and $FN1^{low}$ after the treatment of the immune checkpoint blockade (ICB).

PPI Network and Pathway Analysis

The PPI network from STRING (https://string-db.org/) (Szklarczyk et al., 2015) was depicted and visualized using Cytoscape software v3.8.2 (Shannon et al., 2003). FN1-associated genes were presented *via* UALCAN (Chandrashekar et al., 2017). The molecular functions and signaling pathways involved in HNSCC were predicted using DAVID (Huang da et al., 2009), a web-based analysis platform for identifying enriched biological themes and visualizing these processes by Gene Ontology (GO) enrichment analysis and genes on Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway maps.

Statistical Analysis

All data were analyzed using the SPSS statistical package (version 24.0; SPSS Inc., Chicago, United States), and visualized using R software v4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). The Student's *t*-test was applied to compare the FN1 expression between HNSCC and para-carcinoma tissues. The Wilcoxon rank-sum test was performed to investigate the association between FN1 and clinical parameters. The correlation between FN1 expression and other genes was detected by Spearman's correlation analysis. Univariate and multivariate Cox regression models were used to predict the prognostic ability of FN1. $p < 0.05$ was regarded as a difference with statistical significance.

RESULTS

Gene Expression Profile Analysis and FN1 Filtering

A total of 15 three-crossing and 115 two-crossing DEGs (13 between TCGA and GSE40290, 83 between TCGA and GSE55550, and 19 between GSE55550 and GSE40290) were filtered (**Supplementary Table S1** and **Figure 1A**). Three DEGs (FN1, PLAU, and FAM3D) were filtered by KM survival analysis (**Supplementary Table S2** and **Figure 2A**). We selected FN1 in the present study for its potential role in tumor angiogenesis and metastasis reported in previous studies (Malik et al., 2010; Kujawa et al., 2020).

High Expression of FN1 in HNSCC

FN1 was significantly overexpressed in HNSCCs compared with normal para-carcinoma tissues on the basis of TCGA database (**Figure 1B**) and in our 20 matched pairs of clinical samples (**Figure 1C**). The protein expression level of FN1 was much higher in HNSCC than in the normal tissues of oral mucosa from THPA tool (**Figure 1D**). Similar results were found in our patients (**Figure 1E**). The clinical details of our patients are shown in **Supplementary Table S3**.

FN1 Was a Predictor of HNSCC Prognosis

The associations between FN1 expression and some clinical parameters, including gender, age, tumor stage, lymph mode metastasis, TNM classification, and smoking status, were investigated. As shown in **Table 1**, there were significant differences among the four groups according to TNM classification ($p = 0.027$), and higher expression levels of FN1 were found in patients classified as stage IV than in the combined groups of the other three stages ($p = 0.021$). The results of univariate and multivariate Cox regression analysis showed that age, TNM classification, and FN1 expression were important prognostic factors for HNSCC patients (**Table 2** and **Figure 2B**). Older patients and the patients with higher FN1 expression or a more advanced TNM stage were prone to having a worse survival outcome.

Diagnostic and Prognostic Capability of FN1 in HNSCC

The receiver operator characteristic (ROC) curve was applied to examine the diagnostic ability of FN1 in HNSCC patients. The area under the curve (AUC) of FN1 expression was 0.875 (**Figure 2C**), higher than that of epidermal growth factor receptor (EGFR) expression (AUC = 0.756, $p < 0.001$), which has been regarded as a biomarker in multiple malignant carcinomas (Yip et al., 2017; Serilmez et al., 2019; Park et al., 2020). Furthermore, the predictive performance of FN1 expression in 5-year OS outcomes of HNSCC patients was analyzed. As shown in **Figure 2D**, the AUC of FN1 expression was 0.694 (**Figure 2D** upper), higher than that of TNM classification (AUC: 0.556) or that of the tumor stage (AUC: 0.543) (**Figure 2D** lower), with a sensitivity of 62.5% and a specificity of 75.0%. Taken together, FN1 might facilitate the diagnosis and OS prediction for HNSCC patients.

Hypermethylation of FN1 and its Potential Prognostic Ability in HNSCC

To further clarify the mechanism of FN1 overexpression in HNSCC, we investigated the methylation status *via* multiple tools. The analysis of UALCAN demonstrated that FN1 was significantly hypermethylated in HNSCC tissues compared with

TABLE 1 | Association between FN1 expression and clinical variables in HNSCC patients.

Clinical variable	Number	FN1 expression	*p* value
Gender			
Male	368	7.356 ± 2.123	0.590
Female	134	7.239 ± 2.161	
Age			
<60	221	7.346 ± 2.167	0.943
≥60	256	7.332 ± 21.123	
Tumor stage			
T1	34	6.805 ± 1.861	0.456
T2	144	7.318 ± 2.132	
T3	133	7.377 ± 2.221	
T4	177	7.445 ± 2.216	
Lymph metastasis			
N0	241	7.271 ± 2.102	0.650
N1	81	7.332 ± 2.049	
N2	136	7.481 ± 2.176	
TNM classification			
I	25	7.472 ± 2.335	0.027[a]
II	81	7.273 ± 1.949	
III	90	6.734 ± 1.984	
IV	306	7.501 ± 2.180	
Smoking status			
Yes	381	7.367 ± 2.111	0.338
No	111	7.141 ± 2.206	

Complete data was unavailable in TCGA database.
[a]p<*0.05.*

TABLE 2 | Univariate Cox regression analysis of prognostic factors of HNSCC.

Variable	OS		
	Hazard ratio	95% CI	*p* value
Gender (male/female)	0.764	(0.574, 1.018)	0.066
Age, year	1.021	(1.008, 1.034)	0.001[a]
Smoking status (yes/no)	1.089	(0.778, 1.525)	0.618
Tumor stage (T1/T2/T3/T4)	1.081	(0.940, 1.243)	0.276
TNM classification (I/II/III/IV)	1.372	(1.162, 1.620)	<0.001[b]
FN1 expression	1.067	(1.000, 1.137)	0.049[c]

OS, overall survival; CI, confidence interval.
[a]p < *0.1.*
[b]p < *0.001.*
[c]p<*0.05.*

normal para-carcinoma tissues (**Figure 3A**). Besides, there was significant positive correlation between the expression levels of FN1 and three crucial DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) (**Figure 3B**). Survival analysis of different methylated regions was evaluated using the MethSurv tool. As shown in **Figure 3C**, higher methylation of FN1 predicted a shorter survival time than hypomethylation (TSS1500-S Shore-cg24877731, p = 0.002. Body-Open Sea-cg21494132, p = 0.019. 5′UTR; 1st Exon-Island-cg26910092, p = 0.009). FN1-associated differently methylated regions were compared and visualized as the heatmap (**Figure 3D**). The result revealed that more hypermethylated sites were located around the open sea regions, while more hypomethylated sites lied in CpG islands. In addition, relative position distributions in different locations of the gene were also depicted. More hypermethylated sites fell onto body regions, while more hypomethylated sites lied on TSS1500, 5′UTR, and 1st Exon regions.

Immune Microenvironment of HNSCC With Different FN1 Expressions

To interpret the role of FN1 expression in the immune microenvironment of HNSCC, the proportions of tumor-infiltrating immune cells were investigated in the TCGA-HNSCC cohort. The results showed that FN1 expression was positively correlated with $CD4^+$ T cells, endothelial cells, macrophages, and NK cells and negatively correlated with $CD8^+$ T cells (**Figure 4A**). The scores and percentage abundance of tumor-infiltrating cells in $FN1^{high}$ and $FN1^{low}$ samples are presented as **Figures 4B, C**. Further analysis revealed that the immune-checkpoint–relevant genes were differentially expressed in HNSCC patients classified according to FN1 expression (**Figure 4D**). Potential ICB response was predicted using the TIDE algorithm. We found that the TIDE score was much higher in HNSCC patients with higher FN1 expression levels (**Figure 4E**), indicating a potential positive effect of immune checkpoint inhibitors in these patients.

Network Establishment for FN1 Correlated Genes in HNSCC

The PPI network showed that FN1 was correlated with various genes. The top 10 related genes were predicted on STRING and visualized using Cytoscape software (**Figure 5A**). More genes positively or negatively correlated with FN1 were shown in a heatmap *via* UALCAN (**Figure 5B**). Gene Ontology (GO) analysis illuminated that the major biological processes (regulation of the actin cytoskeleton, cell junction, and cell adhesion), cellular components (ECM), and molecular functions (ECM–receptor interaction) might contribute to FN1-related biology. GO enrichment and KEGG pathway analysis showed that the P13K-Akt signaling pathway and focal adhesion were significantly enriched by the FN1 co-expressed genes (**Figures 5C, D**).

DISCUSSION

In the present study, we found that FN1 was upregulated in HNSCC, and its overexpression was correlated with a poorer prognosis. Furthermore, FN1 could be an independent prognostic factor for HNSCC patients, which was consistent with previous studies in other malignant tumors (Ruiz-Garcia et al., 2010; Kenny et al., 2014). As far as we know, this is the first study demonstrating that hypermethylation and an aberrant immune microenvironment might contribute to the overexpression of FN1 in HNSCC patients, which could provide a new perspective on the treatment of HNSCC.

Increased activation of FN1, a key component of the ECM, has been detected in metastasis and aggressiveness such as radio-resistance in various cancers (Van Obberghen-Schilling et al.,

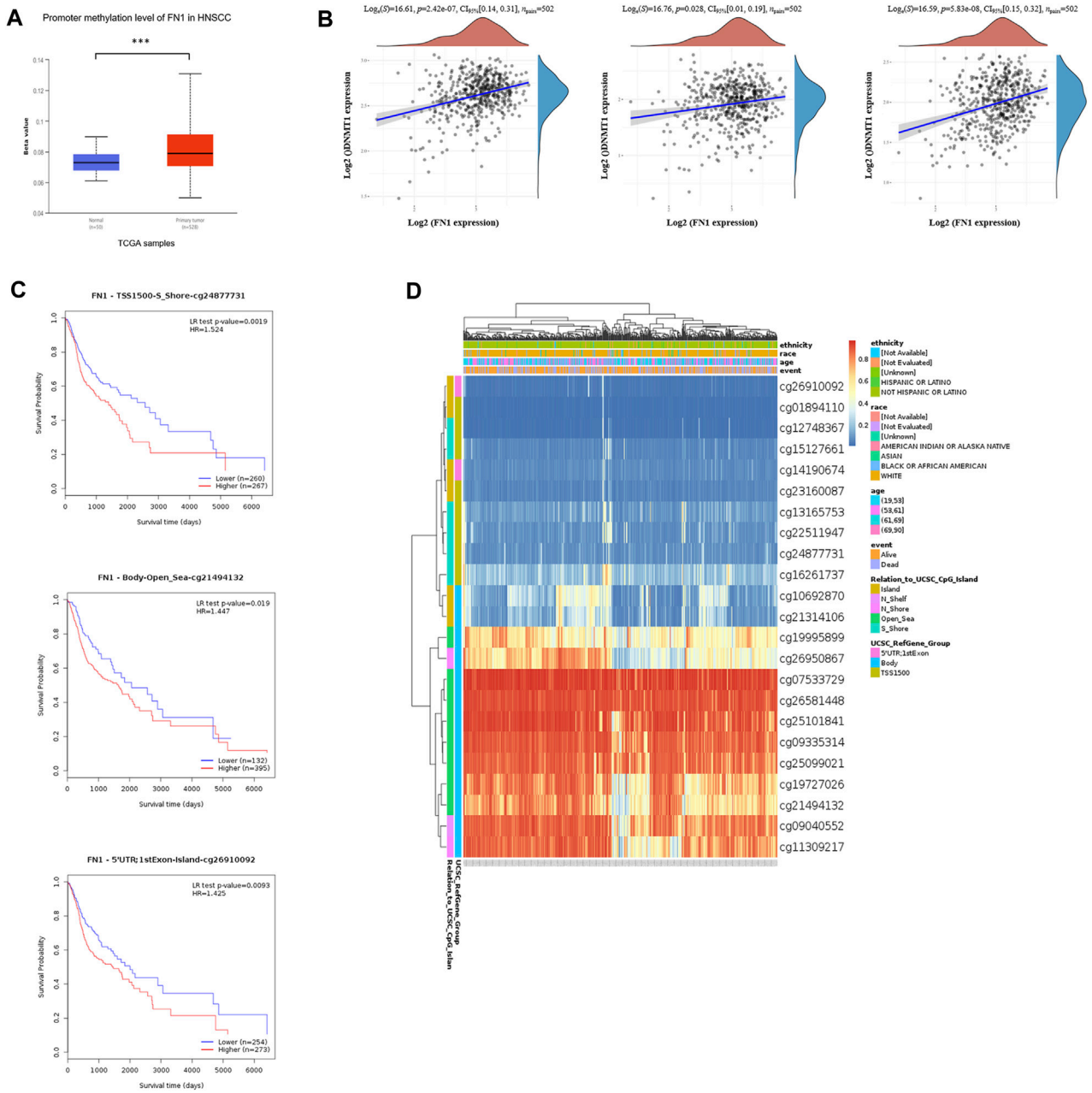


FIGURE 3 | Hypermethylation of FN1 and its prognostic value in HNSCC. **(A)** FN1 was significantly hypomethylated in HNSCC tissues compared with the normal group. **(B)** FN1 expression was positively correlated with three crucial DNA methyltransferases' (DNMT1, DNMT3A, and DNMT3B) levels. **(C)** Higher methylation of specific sites of FN1 predicted a shorter survival time compared to hypomethylation. **(D)** FN1-associated differently methylated regions were compared and visualized as the heatmap.

2011; Morita et al., 2015). While fibronectin (FN) is bound by multiple integrins at specific amino acid sequences, α5β1 integrin is selective for fibronectin (Mould et al., 1997). The binding of integrins to the ECM triggers integrin clustering and then promotes integrin-mediated intracellular signal transduction (Desgrosellier and Cheresh, 2010). At the same time, integrin receptors mediate matrix assembly of FN1 (Schwarzbauer and Sechler, 1999). Functionally, FN1 induces proliferation, adhesion, and invasion of tumor cells, contributing to the formation, adhesion, metastasis, and disaggregation of malignant tumors (Zand et al., 2003; Ritzenthaler et al., 2008; Mitra et al., 2011). On the other hand, it is confirmed that increased expression of FN1 is involved in the promotion of epithelial–mesenchymal transition (EMT), which has been implicated in tumor invasion and metastasis (Guarino et al., 2007; Margadant et al., 2012). EMT is characterized by disruption of intercellular contacts and

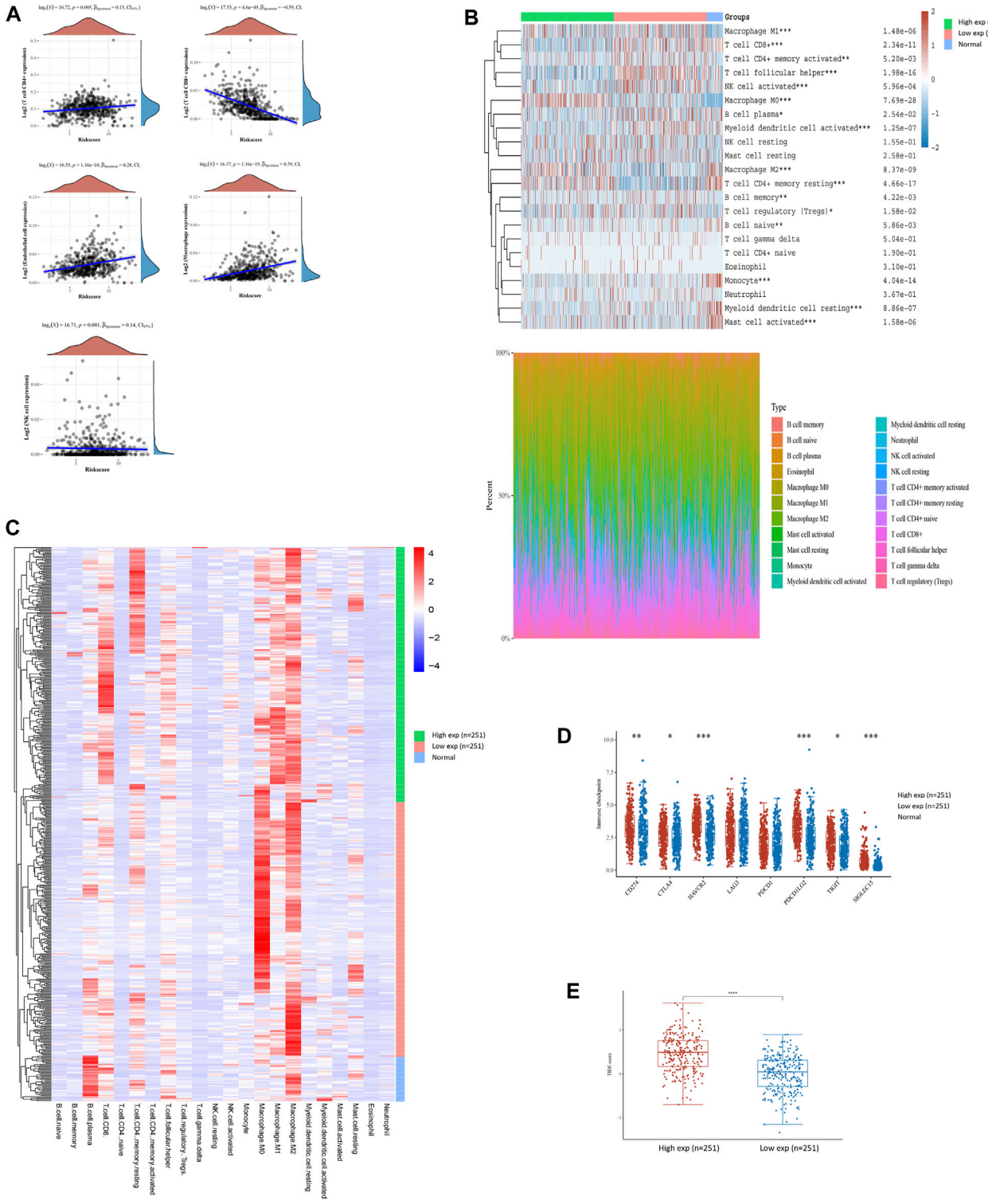


FIGURE 4 | FN1-related tumor immune microenvironment of HNSCC. **(A)** FN1 expression was positively correlated with $CD4^+$ T cells, endothelial cells, macrophages, and NK cells and negatively correlated with $CD8^+$ T cells. Scores and percentage abundance of tumor-infiltrating cells in $FN1^{high}$ and $FN1^{low}$ as well as normal samples performed using the CIRBSORT algorithm **(B)** and unsupervised clustering **(C)**. **(D)** Immune-checkpoint–relevant genes were differentially expressed in $FN1^{high}$ and $FN1^{low}$ samples. **(E)** Patients with higher expression level of FN1 had a higher TIDE score.

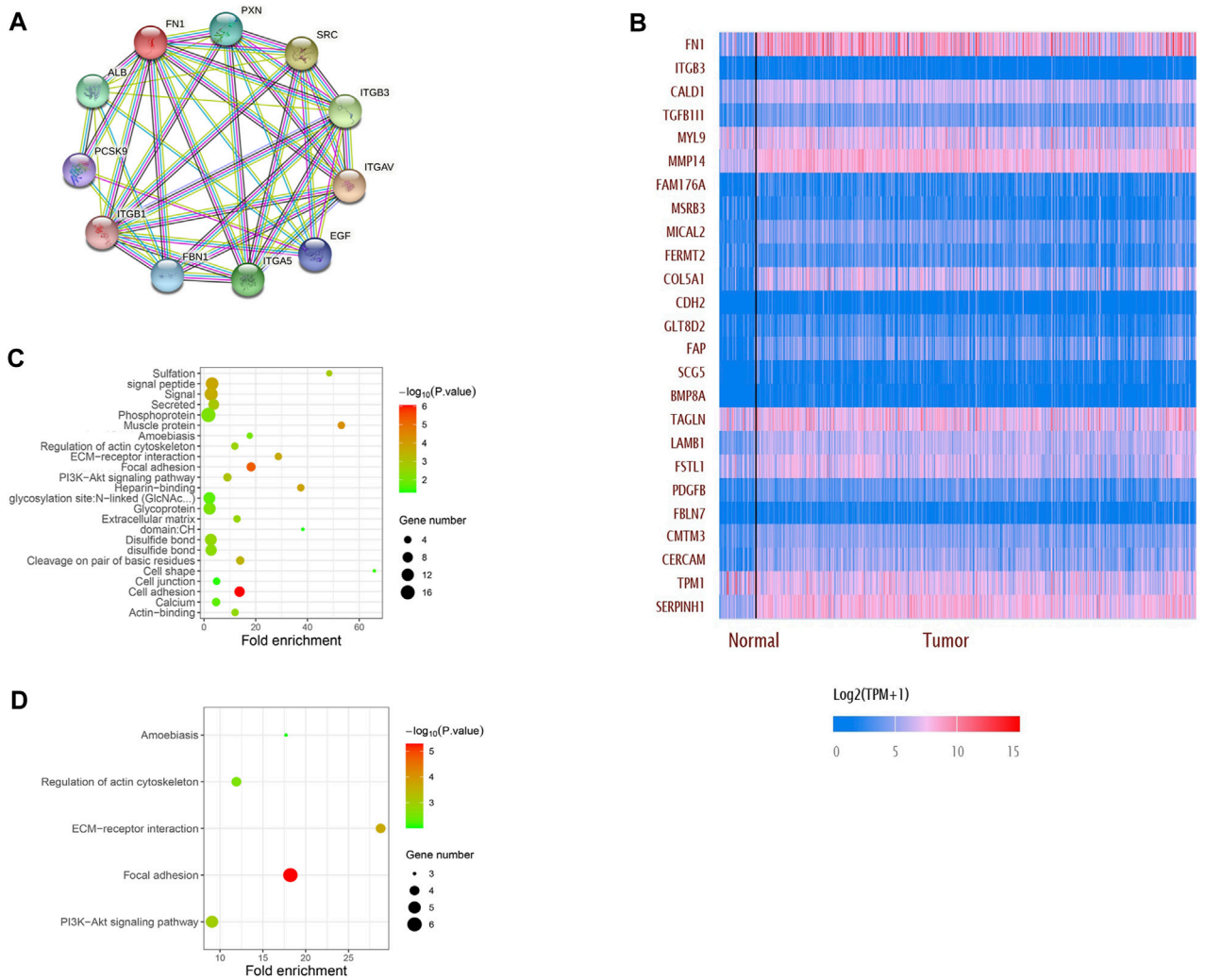


FIGURE 5 | Network establishment for FN1-correlated genes in HNSCC. **(A)** PPI networks of FN1 interaction partners generated using STRING and Cytoscape. **(B)** Top 25 genes positively or negatively correlated with FN1 were shown in the heatmap. Major biological processes, cellular components, molecular functions and signaling pathways of FN1 biology by GO enrichment **(C)** and KEGG pathway analysis **(D)**.

enhancement of cell motility, facilitating malignant cells to invade surrounding tissues and subsequently enter the circulation, thereby allowing distant metastasis. In our present study, FN1 was found to be upregulated in HNSCC patients in TCGA cohorts and in our clinical paired samples. In addition, FN1 overexpression was associated with poorer outcomes of 5-year OS, which could be an independent factor for the prognosis of HNSCC. However, FN1 expression levels were not correlated with lymph node metastasis in our study, indicating that FN1 might induce metastasis of HNSCC through other processes, such as hematogenous metastasis.

DNA methylation, as an important part of epigenetics, is presented by the faithful cross–cell division transmission of the gene transcription memory (Lujambio et al., 2007). The hypermethylated promoter and enhancer regions tightly correlate with the transcriptional silence of both protein-coding and non-coding RNA genes, subsequently regulating gene expression, especially for tumor suppressor genes (Shi et al., 2016). Thus, exploring the DNA methylation state of the promoter, rather than the levels of the corresponding mRNAs or proteins, promises a better way for both early diagnosis and personalized therapy of cancer. In the current study, we analyzed methylation of the FN1 gene *via* multiple bioinformatic tools based on TCGA database. We found that FN1 was hypermethylated in HNSCC tissues. Consistent results showed that the $FN1^{High}$ group co-occurred with higher expression of DNA methyltransferases (DNMT1, DNMT3A, and DMNT3B). Besides, the positive correlation between differently methylated sites and prognosis of HNSCC patients suggested that this epigenetic modification might be a potentially increased risk of HNSCC-related death. Although it has been known that inactivation of certain tumor-suppressor genes occurs as a consequence of hypermethylation within the promoter regions (Kulis and Esteller, 2010), our study observed a positive correlation between FN1 expression and

methylation, which could be explained by the fact that the most hypermethylated sites were located in body regions of FN1. In line with our findings was a reported mechanism that positive correlations were enriched in the 3′UTR and body regions, and only 20% were located in TSSs (Gyorffy et al., 2016).

Growing evidence suggests that the innate immune cells (neutrophils, macrophages, innate lymphoid cells, dendritic cells, natural killer cells, and myeloid-derived suppressor cells) and adaptive immune cells (B cells and T cells) contribute to tumor progression when present in the tumor microenvironment (TME) (Gajewski et al., 2013). Therefore, the investigation of the TME in HNSCC may allow for improved therapeutics that target multiple components of the TME simultaneously, improving the outcomes for HNSCC patients. Furthermore, the administrations of immune checkpoint modulators (such as anti-CTLA4 and anti-PD antibodies) and adoptive immune cells (such as CAR-T) have exhibited unexpected antitumor effect in various cancers (Lei et al., 2020). Recent research has demonstrated that exposure to immune checkpoint inhibitors (ICIs) promotes tumor sensitivity to chemotherapy in HNSCC (Saleh et al., 2019). In this study, we found that $FN1^{High}$ patients were prone to having higher levels of CD274 (PD-L1), CTLA4, and so on. Currently, the blockade of PD-1 signaling using the PD-1/PD-L1 antibody (such as pembrolizumab, nivolumab, and durvalumab) or the blockade of CTLA4 signaling using the CTLA4 antibody (such as ipilimumab and tremelimumab) has revealed the encouraging therapeutic effects in multiple cancers (Hodi et al., 2010; Topalian et al., 2012). Other checkpoint modulators, including agonistic antibodies (such as CD40 and GITR) and inhibitory antibodies (such as LAG-3), are still under clinical evaluation.

In our GO and KEGG pathway analysis, FN1 is involved in the synthesis of ECM components, and the PI3K-Akt pathway was enriched by FN1 co-expressed genes. Akt activation may be the downstream pathway of FN1 leading to tumor progression and poor prognosis in HNSCC. Previous reports have indicated that the PI3K/AKT/mTOR pathway is instrumental in FN transcription and alternative splicing, which modulates cell behavior (White et al., 2010). Matsuo et al. reported that the PI3K/Akt pathway was activated by FN with the aid of integrin αv-mediated a disintegrin and metalloprotease (ADAM) activity in hepatocellular carcinoma (Matsuo et al., 2006), providing us with a new understanding of the PI3K/Akt pathway in FN-related tumor progression.

Despite some merits of our study, limitations should be addressed. The analyses were performed based on the public datasets *via* multiple bioinformatic tools, which could provide a new perspective for our following research. Therefore, our findings need to be confirmed by more *in vitro* and *in vivo* experiments. In addition, the platforms applied in different cohorts are not the same, which may bring bias to the data analysis and difficulties for the deep integrated analysis.

In summary, we found that FN1 was overexpressed in HNSCC patients and upregulation of FN1 was correlated with a worse survival outcome. In addition, hypermethylation and an aberrant TME were strongly associated with FN1 overexpression *via* the PI3K/Akt signaling pathway. Therefore, FN1 could be considered as an independent diagnostic and prognostic biomarker in HNSCC.

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 Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SS: conceptualization, methodology, software, data curation, and writing—original draft. BG: methodology, software, data curation, and article revision. ZW: visualization, investigation, and software. ZZ: validation, data curation, and investigation. JZ: supervision, methodology, and writing—review and editing. ZH: supervision, analysis, and writing—review and editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.753563/full#supplementary-material

REFERENCES

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

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

Chauhan, S. S., Kaur, J., Kumar, M., Matta, A., Srivastava, G., Alyass, A., et al. (2015). Prediction of Recurrence-free Survival Using a Protein Expression-Based Risk Classifier for Head and Neck Cancer. *Oncogenesis* 4, e147. doi:10.1038/oncsis.2015.7

Desgrosellier, J. S., and Cheresh, D. A. (2010). Integrins in Cancer: Biological Implications and Therapeutic Opportunities. *Nat. Rev. Cancer* 10 (1), 9–22. doi:10.1038/nrc2748

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

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and Pharmacological Modulators of the Tumor Immune

Contexture Revealed by Deconvolution of RNA-Seq Data. *Genome Med.* 11 (1), 34. doi:10.1186/s13073-019-0638-6

Gajewski, T. F., Schreiber, H., and Fu, Y.-X. (2013). Innate and Adaptive Immune Cells in the Tumor Microenvironment. *Nat. Immunol.* 14 (10), 1014–1022. doi:10.1038/ni.2703

Guarino, M., Rubino, B., and Ballabio, G. (2007). The Role of Epithelial-Mesenchymal Transition in Cancer Pathology. *Pathology* 39 (3), 305–318. doi:10.1080/00313020701329914

Győrffy, B., Bottai, G., Fleischer, T., Munkácsy, G., Budczies, J., Paladini, L., et al. (2016). Aberrant DNA Methylation Impacts Gene Expression and Prognosis in Breast Cancer Subtypes. *Int. J. Cancer* 138 (1), 87–97. doi:10.1002/ijc.29684

Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* 363 (8), 711–723. doi:10.1056/NEJMoa1003466

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 (1), 44–57. doi:10.1038/nprot.2008.211

Huo, Z., Shi, J., Shu, Y., Xiang, M., Lu, J., and Wu, H. (2017). The Relationship between Allergic Status and Adenotonsillar Regrowth: a Retrospective Research on Children after Adenotonsillectomy. *Sci. Rep.* 7, 46615. doi:10.1038/srep46615

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

Kenny, H. A., Chiang, C.-Y., White, E. A., Schryver, E. M., Habis, M., Romero, I. L., et al. (2014). Mesothelial Cells Promote Early Ovarian Cancer Metastasis through Fibronectin Secretion. *J. Clin. Invest.* 124 (10), 4614–4628. doi:10.1172/JCI74778

Kosmehl, H., Berndt, A., Strassburger, S., Borsi, L., Rousselle, P., Mandel, U., et al. (1999). Distribution of Laminin and Fibronectin Isoforms in Oral Mucosa and Oral Squamous Cell Carcinoma. *Br. J. Cancer* 81 (6), 1071–1079. doi:10.1038/sj.bjc.6690809

Kujawa, K. A., Zembala-Nożyńska, E., Cortez, A. J., Kujawa, T., Kupryjańczyk, J., and Lisowska, K. M. (2020). Fibronectin and Periostin as Prognostic Markers in Ovarian Cancer. *Cells* 9 (1), 149. doi:10.3390/cells9010149

Kulis, M., and Esteller, M. (2010). DNA Methylation and Cancer. *Adv. Genet.* 70, 27–56. doi:10.1016/B978-0-12-380866-0.60002-2

Kumra, H., and Reinhardt, D. P. (2016). Fibronectin-targeted Drug Delivery in Cancer. *Adv. Drug Deliv. Rev.* 97, 101–110. doi:10.1016/j.addr.2015.11.014

Leemans, C. R., Snijders, P. J. F., and Brakenhoff, R. H. (2018). The Molecular Landscape of Head and Neck Cancer. *Nat. Rev. Cancer* 18 (5), 269–282. doi:10.1038/nrc.2018.11

Lei, X., Lei, Y., Li, J.-K., Du, W.-X., Li, R.-G., Yang, J., et al. (2020). Immune Cells within the Tumor Microenvironment: Biological Functions and Roles in Cancer Immunotherapy. *Cancer Lett.* 470, 126–133. doi:10.1016/j.canlet.2019.11.009

Lujambio, A., Ropero, S., Ballestar, E., Fraga, M. F., Cerrato, C., Setién, F., et al. (2007). Genetic Unmasking of an Epigenetically Silenced microRNA in Human Cancer Cells. *Cancer Res.* 67 (4), 1424–1429. doi:10.1158/0008-5472.CAN-06-4218

Malik, G., Knowles, L. M., Dhir, R., Xu, S., Yang, S., Ruoslahti, E., et al. (2010). Plasma Fibronectin Promotes Lung Metastasis by Contributions to Fibrin Clots and Tumor Cell Invasion. *Cancer Res.* 70 (11), 4327–4334. doi:10.1158/0008-5472.CAN-09-3312

Margadant, C., van den Bout, I., van Boxtel, A. L., Thijssen, V. L., and Sonnenberg, A. (2012). Epigenetic Regulation of Galectin-3 Expression by β1 Integrins Promotes Cell Adhesion and Migration. *J. Biol. Chem.* 287 (53), 44684–44693. doi:10.1074/jbc.M112.426445

Matsuo, M., Sakurai, H., Ueno, Y., Ohtani, O., and Saiki, I. (2006). Activation of MEK/ERK and PI3K/Akt Pathways by Fibronectin Requires Integrin Alphav-Mediated ADAM Activity in Hepatocellular Carcinoma: a Novel Functional Target for Gefitinib. *Cancer Sci.* 97 (2), 155–162. doi:10.1111/j.1349-7006.2006.00152.x

Mitra, A. K., Sawada, K., Tiwari, P., Mui, K., Gwin, K., and Lengyel, E. (2011). Ligand-independent Activation of C-Met by Fibronectin and α5β1-integrin Regulates Ovarian Cancer Invasion and Metastasis. *Oncogene* 30 (13), 1566–1576. doi:10.1038/onc.2010.532

Modhukur, V., Iljasenko, T., Metsalu, T., Lokk, K., Laisk-Podar, T., and Vilo, J. (2018). MethSurv: a Web Tool to Perform Multivariable Survival Analysis Using DNA Methylation Data. *Epigenomics* 10 (3), 277–288. doi:10.2217/epi-2017-0118

Morita, Y., Hata, K., Nakanishi, M., Omata, T., Morita, N., Yura, Y., et al. (2015). Cellular Fibronectin 1 Promotes VEGF-C Expression, Lymphangiogenesis and Lymph Node Metastasis Associated with Human Oral Squamous Cell Carcinoma. *Clin. Exp. Metastasis* 32 (7), 739–753. doi:10.1007/s10585-015-9741-2

Mould, A. P., Askari, J. A., Aota, S.-i., Yamada, K. M., Irie, A., Takada, Y., et al. (1997). Defining the Topology of Integrin α5β1-Fibronectin Interactions Using Inhibitory Anti-α5 and Anti-β1 Monoclonal Antibodies. *J. Biol. Chem.* 272 (28), 17283–17292. doi:10.1074/jbc.272.28.17283

Park, Y. W., Ahn, S. S., Park, C. J., Han, K., Kim, E. H., Kang, S.-G., et al. (2020). Diffusion and Perfusion MRI May Predict EGFR Amplification and the TERT Promoter Mutation Status of IDH-Wildtype Lower-Grade Gliomas. *Eur. Radiol.* 30 (12), 6475–6484. doi:10.1007/s00330-020-07090-3

Ritzenthaler, J. D., Han, S., and Roman, J. (2008). Stimulation of Lung Carcinoma Cell Growth by Fibronectin-Integrin Signalling. *Mol. Biosyst.* 4 (12), 1160–1169. doi:10.1039/b800533h

Ruiz-Garcia, E., Scott, V., Machavoine, C., Bidart, J. M., Lacroix, L., Delaloge, S., et al. (2010). Gene Expression Profiling Identifies Fibronectin 1 and CXCL9 as Candidate Biomarkers for Breast Cancer Screening. *Br. J. Cancer* 102 (3), 462–468. doi:10.1038/sj.bjc.6605511

Saleh, K., Daste, A., Martin, N., Pons-Tostivint, E., Auperin, A., Herrera-Gomez, R. G., et al. (2019). Response to Salvage Chemotherapy after Progression on Immune Checkpoint Inhibitors in Patients with Recurrent And/or Metastatic Squamous Cell Carcinoma of the Head and Neck. *Eur. J. Cancer* 121, 123–129. doi:10.1016/j.ejca.2019.08.026

Schwarzbauer, J. E., and Sechler, J. L. (1999). Fibronectin Fibrillogenesis: a Paradigm for Extracellular Matrix Assembly. *Curr. Opin. Cel Biol.* 11 (5), 622–627. doi:10.1016/s0955-0674(99)00017-4

Serilmez, M., Özgür, E., Karaman, S., Gezer, U., and Duranyıldız, D. (2019). Detection of Serum Protein and Circulating mRNA of cMET, HGF EGF and EGFR Levels in Lung Cancer Patients to Guide Individualized Therapy. *Cbm* 25 (2), 177–184. doi:10.3233/CBM-182231

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

Shi, J., Fu, H., Jia, Z., He, K., Fu, L., and Wang, W. (2016). High Expression of CPT1A Predicts Adverse Outcomes: A Potential Therapeutic Target for Acute Myeloid Leukemia. *EBioMedicine* 14, 55–64. doi:10.1016/j.ebiom.2016.11.025

Shinde, A., Libring, S., Alpsoy, A., Abdullah, A., Schaber, J. A., Solorio, L., et al. (2018). Autocrine Fibronectin Inhibits Breast Cancer Metastasis. *Mol. Cancer Res.* 16 (10), 1579–1589. doi:10.1158/1541-7786.MCR-18-0151

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer Statistics, 2016. *CA: A Cancer J. Clinicians* 66 (1), 7–30. doi:10.3322/caac.21332

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

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

Van Obberghen-Schilling, E., Tucker, R. P., Saupe, F., Gasser, I., Cseh, B., and Orend, G. (2011). Fibronectin and Tenascin-C: Accomplices in Vascular Morphogenesis during Development and Tumor Growth. *Int. J. Dev. Biol.* 55 (4-5), 511–525. doi:10.1387/ijdb.103243eo

White, E. S., Sagana, R. L., Booth, A. J., Yan, M., Cornett, A. M., Bloomheart, C. A., et al. (2010). Control of Fibroblast Fibronectin Expression and Alternative Splicing via the PI3K/Akt/mTOR Pathway. *Exp. Cel Res.* 316 (16), 2644–2653. doi:10.1016/j.yexcr.2010.06.028

Yip, S. S. F., Kim, J., Coroller, T. P., Parmar, C., Velazquez, E. R., Huynh, E., et al. (2017). Associations between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer. *J. Nucl. Med.* 58 (4), 569–576. doi:10.2967/jnumed.116.181826

Zand, L., Feng, Q., Roskelley, C. D., Leung, P. C. K., and Auersperg, N. (2003). Differential Effects of Cellular Fibronectin and Plasma Fibronectin on Ovarian Cancer Cell Adhesion, Migration, and Invasion. In Vitro *Cell.Dev.Biol.-Animal* 39 (3-4), 178–182. doi:10.1007/s11626-003-0013-0

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 Sheng, Guo, Wang, Zhang, Zhou and Huo. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fmolb.2021.736940

Molecular and Clinical Characteristics of Primary Pulmonary Lymphoepithelioma-Like Carcinoma

Ying Fan[1,2,3,4], *Qianyun Shan*[1,2,3,5], *Jiali Gong*[1,2,3,5], *Jing Qin*[1,2,3] *and Hongyang Lu*[1,2,3,4,5]*

[1]*Zhejiang Key Laboratory of Diagnosis and Treatment Technology on Thoracic Oncology (Lung and Esophagus), Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China,* [2]*Department of Thoracic Medical Oncology, Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China,* [3]*Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China,* [4]*The First Clinical Medical College, Wenzhou Medical University, Wenzhou, China,* [5]*The Second Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou, China*

Objectives: Primary pulmonary lymphoepithelioma-like carcinoma (PPLELC) is an extremely rare subtype of non-small cell lung cancer (NSCLC). Currently, there are no established treatment protocols due to rarity of the cancer. Thus, this study aimed to explore the molecular and clinical characteristics of PPLELC.

Study design and setting: Data from patients with PPLELC who were admitted to Zhejiang Cancer Hospital from August 2009 to September 2020 were retrospectively collected. Next-generation sequencing was performed to obtain a genomic profile and tumor mutation burden (TMB) value of patients with adequate tissue and divided them into two groups according to the expression level of PD-L1. The correlation of PD-L1 expression and the clinicopathological characteristics was evaluated by Pearson Chi-square test. Kaplan-Meier curves was applied to present the probability of survival between PD-L1 expression level and overall survival (OS). Moreover, the literature on the immunotherapy of advanced PPLELC published in PubMed between 2016 and 2020 were reviewed and the efficacy of immunotherapy were analyzed.

Results: A total of 18 patients pathologically diagnosed as PPLELC were included. After a follow-up period of 8.8–138 months, 14 patients survived, three patients died and one patient lost, the median OS was 45.3 months Seven samples (tissue-available) tested by NGS and the median TMB was 2.5 mutations/Mb. 19 somatic mutated genes were recognized and TP53 (43%) and CYLD (43%) were the two most commonly mutated genes. Only seven patients who underwent NGS were tested for PD-L1. Three patients with high PD-L1 expression (PD-L1≥ 50%) and four patients with low PD-L1 expression (PD-L1 <50%) were included. No significant correlation was observed between PD-L1 expression and clinical characteristics (age, gender, smoking status, tumor stage, lymph node metastasis) ($p > 0.05$) and OS ($p = 1$). What's more, 10 PPLELC patients involved in previous studies and one patient received nivolumab in the current study were collected retrospectively. 4/11 (36.4%) patients achieved PR, 6/11 (54.5%) patients achieved SD, and 1/11 (9.1%) patients achieved PD and the disease control rate (DCR) was 90.9%.

OPEN ACCESS

Edited by:
Xin Wang,
City University of Hong Kong, Hong Kong, SAR China

Reviewed by:
Stella Liberman-Aronov,
Ariel University, Israel
Sayan Chatterjee,
Guru Gobind Singh Indraprastha University, India
Zhongxu Zhu,
City University of Hong Kong, Hong Kong, SAR China

***Correspondence:**
Hongyang Lu
luhy@zjcc.org.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *06 July 2021*
Accepted: *07 October 2021*
Published: *25 October 2021*

Citation:
Fan Y, Shan Q, Gong J, Qin J and Lu H (2021) Molecular and Clinical Characteristics of Primary Pulmonary Lymphoepithelioma-Like Carcinoma. Front. Mol. Biosci. 8:736940. doi: 10.3389/fmolb.2021.736940

Conclusions: The prognosis of PPLELC is better than that of other NSCLC, and immunotherapy may be a promising treatment to prolong the survival of advanced PPLELC patients. Whether the immunotherapy efficacy of PPLELC can be predicted by PD-L1 and TMB needs further clinical investigation. CYLD genetic alterations may participate in Epstein–Barr virus-mediated tumorigenesis in PPLELC, providing a novel therapeutic target.

Keywords: PPLELC, clinical characteristics, molecular characteristics, immunotherapy, PD-L1, TMB

HIGHLIGHTS

This may be one of the first retrospective study based on the patient data of over a decade to understand the molecular characteristics and immunotherapy reaction of PPLELC. The strength of the work is complimented not only in the results obtained but also in the crisp methodology of the planning and designing of the experiments using simple analytical techniques. Statistical validation/significance of the data may not be adequate as the sample size is very limited.

INTRODUCTION

Lymphoepithelioma-like carcinoma (LELC) is a rare malignant tumor, which shares similar histology with undifferentiated nasopharyngeal carcinoma (NPC). It occurs in the submandibular gland, parotid gland, thymus, lung, stomach, uterus, bladder, and skin (Bégin et al., 1987). Primary pulmonary lymphoepithelioma-like carcinoma (PPLELC) is a rare lung tumor with specific clinicopathological characteristics (Liang et al., 2012), affecting younger patients (51–55-year-old), mostly non-smokers, and mainly Asian females and women from Southern China (Huang et al., 2007; Sun et al., 2014; Jiang et al., 2016). Moreover, a strong correlation has been established between Epstein–Barr virus (EBV) infection and the histological characteristics of PPLELC, which are similar to those of NPC (Huang et al., 2012; Liang et al., 2012).

The majority of the patients with PPLELC do not have obvious clinical manifestations at the time of diagnosis (Grimes et al., 2015), and the treatment modalities of PPLELC follow NSCLC regimens (Ho et al., 2000; Ho et al., 2009). In addition, patients with advanced PPLELC are less likely to benefit from targeted therapy due to the low mutation rates of epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), and Kirsten rat sarcoma viral oncogene (KRAS), v-RAF murine sarcoma viral oncogene homolog B1 (BRAF), repressor of silencing 1 (ROS1) and protein 53 (P53) genes (Chang et al., 2011; Liang et al., 2012; Chang et al., 2015). However, the high expression of programmed death receptor ligand-1 (PD-L1) in PPLELC indicates that PD-L1 inhibitors such as nivolumab and pembrolizumab may be suitable treatment options (Chang et al., 2015; Fang et al., 2015; Xie et al., 2018). The efficacy of nivolumab in the treatment of advanced NSCLC is significantly related to tumor mutation burden (TMB), and it is more effective in patients with high TMB compared to chemotherapy (Carbone et al., 2017; Peters et al., 2017). However, although the TMB of PPLELC is low, a large amount of gene copy number variations (CNVs), especially 11q13.3 amplification and 9p21.3 deletion, have been observed (Hong et al., 2019).

In the present study, we retrospectively collected 18 lung tissue samples of PPLELC diagnosed in our hospital. Next-generation sequencing (NGS) was performed in seven patients who have enough tissue and divided into two groups according to the expression level of PD-L1 to analyze the genetic signature and TMB. Furthermore, the literature on the immunotherapy of advanced PPLELC were reviewed to analyze the efficacy of immunotherapy. The flowchart of the sample collection and study design was shown in **Figure 1**.

MATERIALS AND METHODS

Sample Collection

Formalin-fixed paraffin-embedded (FFPE) blocks from patients with pathologically confirmed PPLELC in Zhejiang Cancer Hospital (Hangzhou, China) were enrolled. Those with the second primary malignant tumors (except cervical carcinoma *in situ* and skin basal cell carcinoma) and other primary LELCs outside the lung were excluded. Nasopharyngoscopy or Magnetic Resonance Imaging (MRI) was done to rule out lung metastasis from NPC in all the patients. Medical records were retrieved to collect clinicopathologic data, treatment history and survival outcomes. The follow-up deadline was December 30, 2020.

Immunohistochemistry Analysis

The specimens were fixed in 10% neutral formalin solution for 24 h and embedded in paraffin. Then the tissue block was cut into 4-μm thick serial sections and baked at 60°C for antigen retrieval. The slices were incubated with the primary antibody at 4°C overnight. Subsequently, the slices were labeled with horseradish peroxidase and stained with chromoplasma matrix to identify the target protein. Phosphate-buffered saline (PBS) was used as a negative control instead of a primary antibody. IHC was performed using the PD-L1 clone 22C3 pharmDx kit. PD-L1 tumor proportion score (TPS) was calculated as the percentage of viable tumor cells with complete or partial membrane staining. TPS was interpreted by a pathologist from commercial vendor. The expression level of PD-L1 was represented by TPS (0–1% as negative, 1–49% as low expression, and ≥50% as high expression) (Wang et al., 2018). The IHC data were scored by two pathologists independently. A total of 100 cells were counted

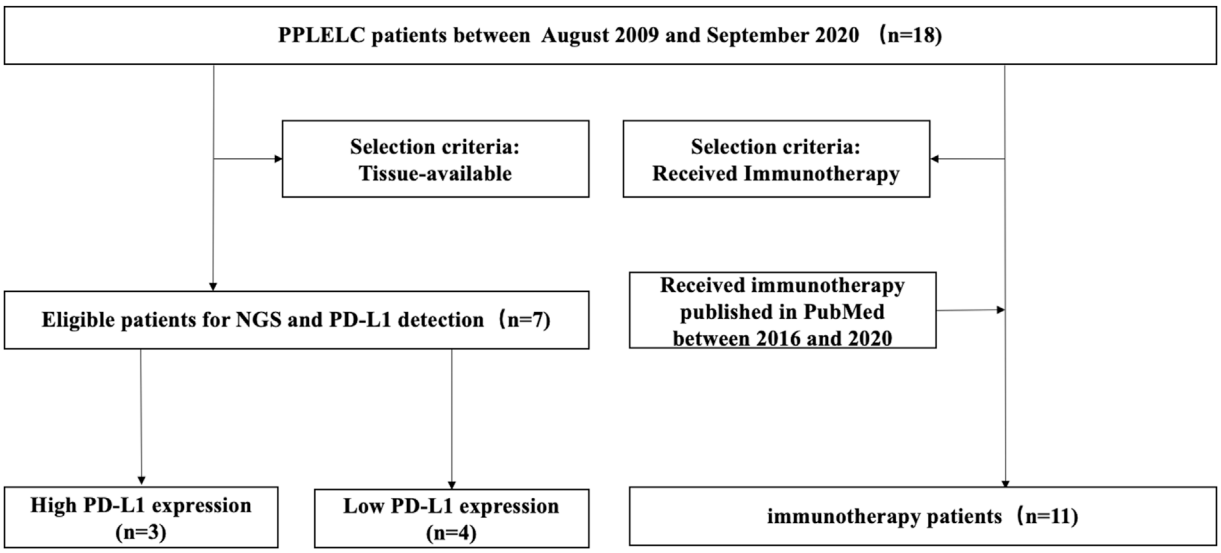


FIGURE 1 | Sample collection and study design of study.

to calculate the proportion of stained cells, and TPS ≥ 1% was defined as positive. PD-L1-positivity was defined when the tumor cells in tissue specimen showed at least 1% PD-L1 expression, while the absence of detection or a TPS of < 1% was considered negative.

In-Situ Hybridization Analysis

Patients with adequate tissue were tested for EBER using ISH according to manufacturer's instructions. The probe for EBER-1 was supplied by OriGene Technologies, Inc., Tumor nuclei stained with brown granules were interpreted as positive.

DNA Extraction

From each tumor-rich FFPE and matched normal lung tissue block, 4 μm of sections were cut, deparaffinized, and dissected to isolate 1 cm^2 of tumor tissue. DNA was isolated using the Cobas R DNA Sample Preparation Kit, according to the manufacturer protocol (Roche Molecular Systems, Pleasanton, CA, United States). The DsDNA concentration was determined using the Qubit R _ 2.0 Fluorometer and the Qubit R 2.0 dsDNA HS Assay Kit (ThermoFisherScientific, Waltham, MA, United States). The quality of the sample DNA was evaluated using a specimen control size ladder test (Invivoscribe Technologies, San Diego, CA, United States).

NGS

Clinical annotations were extracted from their medical records. Tumor and matched normal DNA were subjected to NGS, and the genomic landscape was explored for potential mutations and therapeutic targets. The genomic information was obtained from NGS-based YuanSu™ 450 gene panel (OrigiMed, Shanghai, China), which encompassed all coding exons of 450 cancer-related genes and 64 selected introns of 39 frequently rearranged solid tumor-related genes. The genes were captured and sequenced with a mean depth of 800X using Illumina NextSeq 500 (Illumina Inc.). Genomic alterations (GAs) were identified by the alignment of sequences from tissues with matched normal lung tissue, as described previously. The TMB was estimated by counting the somatic mutations containing single nucleotide variations (SNVs) and Indels per Mb of the sequence examined in each patient. The driver mutations and recorded germline alterations were not counted.

Statistical Analysis

Statistical analyses were performed with the R (version 4.1.1). The correlation of PD-L1 expression and prognosis and the clinicopathological characteristics (age, gender, smoking status, lymph node metastasis, and TNM stage) of patients was evaluated by Pearson Chi-square test. Survival curves were plotted using the Kaplan-Meier method and the differences in survival rates were assessed using the log-rank test. Univariate and multivariate analysis of prognostic factors was performed using the Cox proportional hazards model. A p value lower than 0.05 ($p < 0.05$) was considered to indicate a significant difference. Somatic mutations in p53 were retrieved from cbioPortal (http://www.cbioportal.org/). Graphs were prepared with "ggplot2" (Wickham, 2009).

Follow-Up

The follow-up deadline was December 30, 2020. 14 patients were still alive, two patients were deceased and one patient was lost to follow-up. The survival time was counted from the date of pathological diagnosis.

Search Strategy and Curative Effect Judgment

A comprehensive search was performed through PubMed using the literature retrieval strategy "[pulmonary lymphoepithelioma-

TABLE 1 | The clinical characteristics of PPLELC (n = 18).

Age at diagnosis	Years
Median	57
Range	43–79
Gender	No. of patients
Male	8
Female	10
Smoking situation	No. of patients
Smoker	6
Non-smoker	12
Specimen source	No. of patients
Surgical	13
Biopsies	5
EBER	No. of patients
Positive	16
Negative	0
Clinical stage	No. of patients
I	7
II	2
III	6
IV	3

like carcinoma (Title/Abstract)] AND [immunotherapy (Title/Abstract)] OR [pulmonary lymphoepithelioma-like carcinoma (Title/Abstract)] AND PD-L1 (Title/Abstract)]" in December 2020 (no year limit and all languages included). Relevant articles were obtained, and references from each of these articles were further searched for relevant articles. A total of 25 articles were reviewed (three were case reports or case series). The clinical efficacy, including complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD), was evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 standard to analyze the efficacy of immunotherapy in patients with advanced PPLELC (Eisenhauer et al., 2009). The objective response rate (ORR), overall survival (OS), and disease-free survival (DFS) were used as the observation indexes. The immune-related adverse reactions (ir-AEs) during the treatment were assessed based on the criteria of common AEs (Common Terminology Criteria for Adverse Events (CTCAE) version 4.0).

RESULTS

Clinicopathological Data

A total of 18 cases of PPLELC diagnosed in our hospital from August 2009 to September 2020 were enrolled in this study, and the basic clinical features were summarized in **Table 1**. Among them, eight patients were men and 10 were women. The patients were middle-aged (average 57 years, range 43–79 years), and 6/18 (33.3%) patients were smokers. 16 patients instead of 18 patients performed an EBER test due to the lack of sufficient tissue of two patients and the expression of Epstein–Barr virus-encoded RNA (EBER) was positive (16/16). 13 patients underwent surgery, of which six received adjuvant platinum-based chemotherapy, four received adjuvant radiotherapy while three did not received any treatment after surgery. Five patients did not receive surgery. Palliative therapy, including concurrent chemoradiotherapy, was undertaken in four patients while one patient did not receive any treatment after diagnosis. As of the follow-up deadline, 14 patients were still alive, two died, and one was lost to follow-up and the median OS was 45.3 months (**Supplementary Table S1**).

Moreover, the information of 11 immunotherapy patients (10 PPLELC patients involved in previous studies (Kim et al., 2016; Kumar et al., 2017; Narayanan et al., 2019; Qiu et al., 2019; Zhou et al., 2019; Tang et al., 2020; Xie et al., 2020) and one patient received nivolumab in the current study marked as "IP") were collected retrospectively (**Table 2**). 4/11 (36.4%) patients achieved PR, 6/11 (54.5%) patients achieved SD, and 1/11 (9.1%) patients achieved PD. The expression of PD-L1 and immunotherapy reaction of 11 patients was shown in **Figure 2**.

IHC Analysis

All 18 patients were diagnosed as PPLELC based on pathological detection with various lung-cancer-related IHC markers (**Table 3**). The expressions of IHC markers were highly correlated with the occurrence and progression of PPLELC. The majority of the patients showed positive for P40 (14/15) and CK5/6 (13/14), but negative for thyroid transcription factor-1 (TTF-1) (12/15) and CK7 (8/9). Moreover, all patients were positive for Ki-67 (6/6) and EMA (3/3), while negative for Napsin A (8/8) and CK5/6 (3/3).

Gene Mutation and TMB Analyses

Seven samples (tissue-available) tested by NGS and the mutation information (include the mutation type, position, functional changes, and so on) was shown in **Supplementary Table S2**. 19 somatic mutated genes were recognized, TP53 (43%) and CYLD (43%) were the two most commonly mutated genes. A lollipop chart for TP53 is shown in **Supplementary Figure S1**. Other mutations occurred in LRIG1 (14%), PTPRT (14%), PPP2R2A (14%), and other 17 genes (**Figure 3A**). The mutation information (include the mutation type, position, functional changes, and so on) was shown in **Supplementary Table S3**. The median TMB was 2.5 mutations/Mb. The differences of TMB between high and low PD-L1 expression groups were assessed via the Fisher's exact test and the p-value was 1 (**Figure 4B**).

The Association Between PD-L1 and Prognosis

Seven patients who underwent NGS were tested for PD-L1. 3/7 (42.8%) patients had ≥50% PD-L1 expression, 2/7 (28.6%) patients showed PD-L1 expression ≥1% and 2/7 (28.6%) patients had **<1%** PD-L1 expression (22C3) (**Figure 3B**). We defined 50% as the cut-off value and seven patients were divided into two groups. Higher than and/or equal to 50% was considered high (PD-L1≥ 50% as high expression) and lower than 50% low (PD-L1 <50%). No significant correlation was observed between PD-L1 expression and clinical characteristics (age, gender, smoking status, tumor stage, lymph node metastasis) (p > 0.05) (**Table 4**). PD-L1 expression was not associated with OS (p = 1) (**Figure 4A**).

TABLE 2 | The studies concerning immunotherapy of advanced PPLELC.

Study	No. of cases	Sex; age; smoking	Stage	The expression of PD-L1 (%)	The line of immunotherapy	Immunotherapy	iAEs	Response	Outcome	PFS (mo)	OS (mo)
Kim et al. (2016)	1	F; 37; No	IV (at recurrence)	0	3	Nivolumab	Yes	PD	Died	1	12
Kumar et al. (2017)	2	M; 56; Yes	IV	0	2	Nivolumab	No	PR	Alive	25	62
		F; 37; No	IIIA	5	3	Nivolumab	No	SD	Alive	27	56
Narayanan et al. (2019)	1	F; 76; No	IV	≥50	4	Atezolizumab	No	PR	Died	4	18
Qiu et al. (2019)	1	M; 56; No	IVA	about 80	2	Nivolumab	No	PR	Alive	5	NA
Zhou et al. (2019)	1	F; NM; NA	IV	90	4	Pembrolizumab	NA	SD	NA	12	NA
Tang et al. (2020)	1	F; 50; No	IVB	10	2	Nivolumab	No	PR	NA	5	NA
Xie et al. (2020)	3	F; 56; No	IV (at recurrence)	30	2	Nivolumab	NA	SD	Alive	1	14.9
		F; 49; No	IIIB	60	2	Nivolumab	NA	SD	Alive	6	14.9
		M; 48; No	IVA	15	3	Camrelizumab	NA	SD	Alive	8	13
The current study (P16)	1	M; 57; No	IVA	<1	2	Nivolumab	No	SD	Alive	21.9	37.5

M, male; F, female; iAE, immune-related adverse event; D, died; A, alive; NA, not available; PD, progressive disease; SD, stable disease; CR, complete response; OS, overall survival, mo, month.

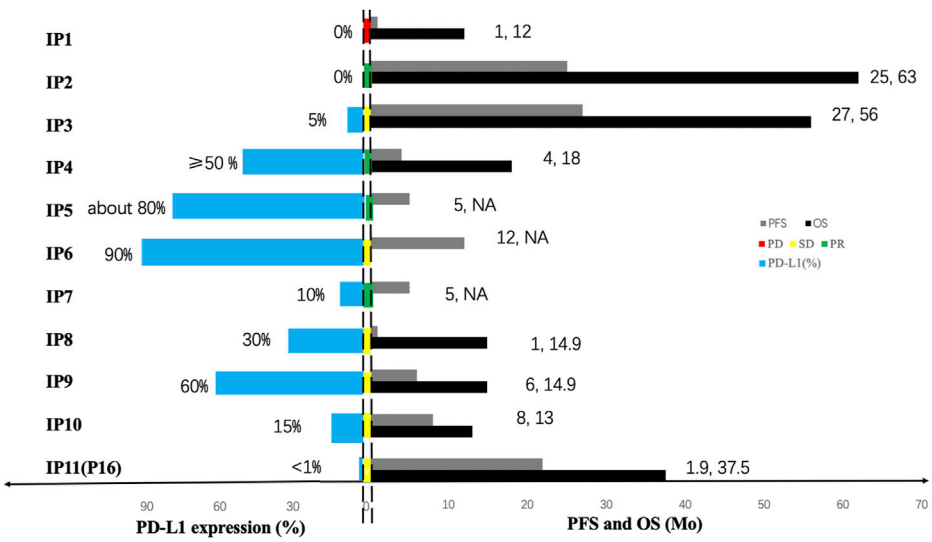


FIGURE 2 | The expression of PD-L1 and immunotherapy reaction of 11 patients. For each immunotherapy patient (marked as "IP"), the bars in the left side of two dotted line represent the expression of PD-L1 and the bars in the middle of two dotted line represent the treatment response to immunotherapy, the green one symbolizes partial response (PR), the yellow symbolizes stable disease (SD) and the yellow symbolizes progressive disease (PD). The bars in the right side of two dotted line represent the prognosis in each patient, black bars and grey bars indicate that the PFS and the OS, respectively. The OS time of two patient was not available and marked as "NA".

DISCUSSION

PPLELC has obvious racial and geographical distribution characteristics. Among the 18 patients in our study, the male to female ratio was 4:5, the median age was 57 years, and the ratio of non-smokers to smokers was 2:1, which was consistent with previous studies (Ho et al., 2000; Grimes et al., 2015). However, the epidemiological characteristics of patients in Western countries may differ from those of all Chinese cases. In He et al. (2015) inclusion of 62 cases of the United States' Open Database was performed; the median age of PPLELC patients was 65 (15~86) years, with slightly more male than female patients, accounting for 58.1% (36/62). The occurrence of PPLELC was related to EBV infection (Bégin et al., 1987), and EBER test results played a role in the diagnosis of PPLELC. 16 patients were all positive for EBER testing, similar to previous results (Liang et al., 2012; Ma et al., 2013;

Mo et al., 2014). EBV infection may be related to race and region. Almost all Asian patients have positive EBV detection, while most non-Asian patients showed negative results (Grimes et al., 2015).

IHC markers are significant in pathological diagnosis, especially for PPLELC, since they are not solely based on the morphologic features. The histopathological features of PPLELC are similar to those of undifferentiated nasopharyngeal carcinoma, and need to be distinguished from metastatic NPC, poorly differentiated primary lung squamous cell carcinoma and lung lymphoma (Da-yun et al., 2017; Anand et al., 2018; Qin et al., 2019). Wang et al. (2013) analyzed the pathological characteristics of 14 patients with PPLELC and pointed out that the high expressions of AE1/AE3, CK5/6, CK19, and LMP-1 could aid in the diagnosis. In the current study, a complete IHC detection was not performed on the patients due to retrospective analysis. However, in the tested patients, the positive rate of squamous cell carcinoma labeled with antibody p63 and CK5/6 was 91.7 and 92.6%, respectively. while the negative rate of TTF-1 and CK7 antibody-labeled adenocarcinoma was 80 and 88.9%, respectively, which were consistent with the above results. PPLELC had a low expression of glandular epithelial markers (TTF-1, CK20, and Napsin-A) and neuroendocrine differentiation markers (CgA, Syn, and CD56), but high expression of squamous epithelial

TABLE 3 | The expression of immunohistochemical markers of PPLELC patients.

IHC marker	Positive	Negative
Ki-67	6	0
EMA	3	0
P40	14	1
CK5/6	13	1
P63	11	1
CK	7	1
Napsin A	0	8
CD56	0	3
CK7	1	8
TTF-1	3	12

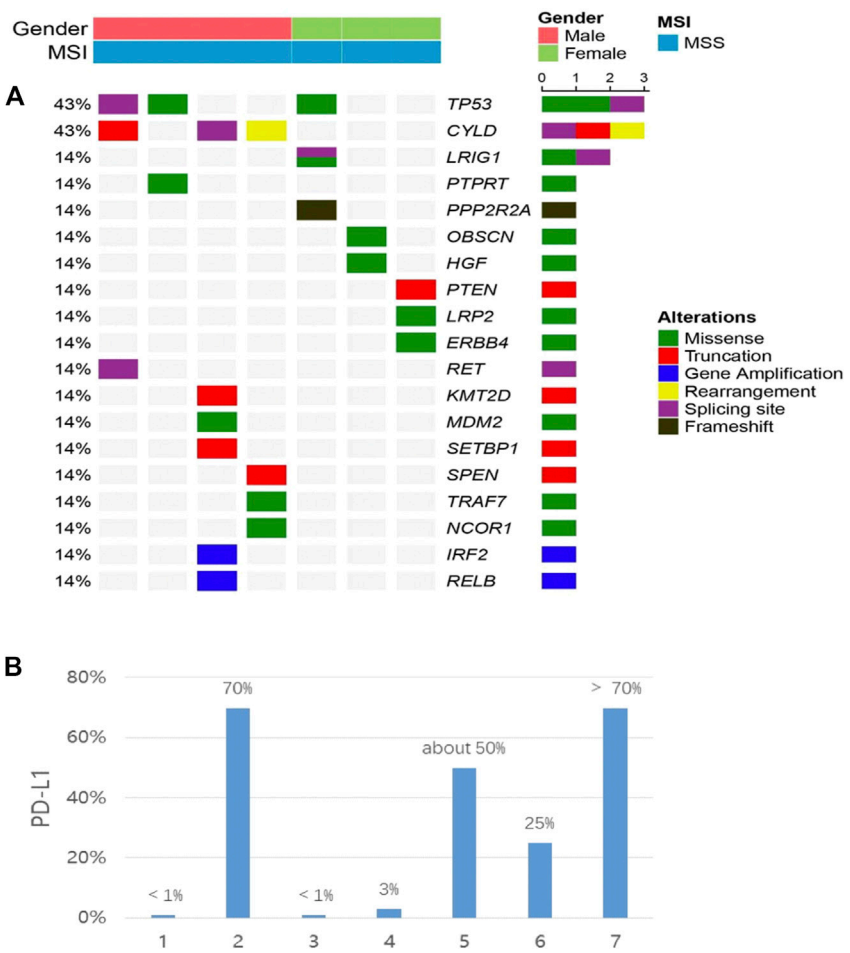


FIGURE 3 | (A) co-mutation plot of various types of mutations in all patients. Genes were grouped according to their functions. Each column represents one patient. The mutation rates of each gene were marked on the left in percentage and grouped according to their protein functions. Patient characteristics such as gender, disease stage and tumor type were shown at the top with different colors. **(B)** the expression of PD-L1 in each patient. All patients were placed in the same order in the two panels.

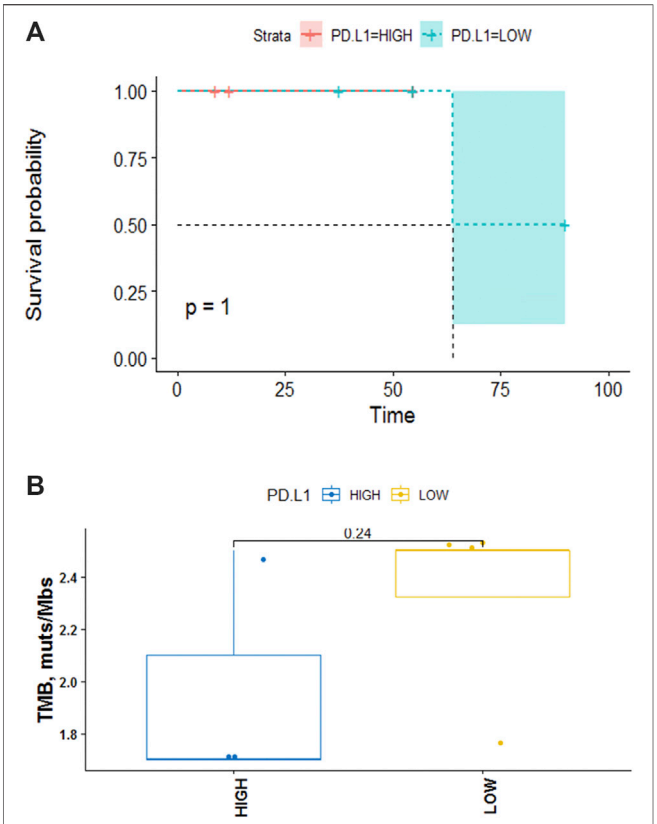


FIGURE 4 | (A) Kaplan-Meier analysis of the effect of PD-L1 on OS. The data indicated that no significant correlation between PD-L1 and disease prognosis (p = 1). **(B)** TMB in high PD-L1 expression group versus low. The data indicate no significant correlation between the expression level of PD-L1 and TMB in PPLCLC patients (p = 0.24).

markers (CK5/6, p63, and p40), indicating that PPLELC was derived from epithelial tissue with similar characteristics to squamous epithelial differentiation (Liang et al., 2012).

PPLELC had a better prognosis, and the two and 5-years survival rates were significantly higher than those of the non-LELC patients (both $p < 0.05$) (Han et al., 2001; Qin et al., 2019). In this study, the median DFS (mDFS) of 13 (72.2%) patients who underwent surgery supplemented by chemotherapy and radiotherapy was 45.5 (range 12.2–136) months. Therefore, early surgical resection of PPLELC is feasible, and no lymph node metastasis and complete resection of the tumor could improve the survival rate of the patients (Liang et al., 2012; Lin et al., 2016). Treatment at advanced tumor stages relies on multimodal therapy, including chemotherapy, and radiotherapy. In the present study, platinum-based regimens were our first choice owing to the similarity of PPLELC to NPC (Huang et al., 2007). However, patients with advanced PPLELC are less likely to benefit from targeted therapy. Several studies have explored the role of typical lung carcinogenic pathways in the development of PPLELC and found that advanced PPLELC is less likely to benefit from targeted therapy (Chang et al., 2011; Liu et al., 2014; Chang et al., 2015; Fang et al., 2015; Yeh et al., 2019). In the current study, instead of interrogating only the classic lung cancer

TABLE 4 | The association between PD-L1 and clinical characteristics.

	High PD-L1 expression	Low PD-L1 expression	*p*-value
	N = 3	N = 4	
Age			
<60	1 (33.3%)	2 (50%)	1
≥60	2 (66.7%)	2 (50%)	
Gender			
Male	1 (33.3%)	3 (75%)	0.741
Female	2 (66.7%)	1 (25%)	
Smoking			
Yes	0	3 (75%)	0.225
No	3 (100.0%)	1 (25%)	
Stage			
I-II	1 (33.3%)	4 (100%)	0.277
III-IV	2 (66.7%)	0	
Lymph node metastasis			
N0	1 (33.3%)	3 (75%)	0.741
N+	2 (66.7)	1 (25%)	

oncogenic drivers, we utilized NGS consisting of 450 cancer-related genes and 64 selected introns of 39 solid tumor-related genes that were frequently rearranged to obtain a comprehensive mutation profile of PPLELE. No frequently altered driver genes (e.g., EGFR, KRAS, and BRAF) in classic NSCLC was detected in our cohort which consistent with previous reports.

Instead, the results revealed that TP53 (43%) and CYLD (43%) were the two most commonly mutated genes, and mutations in other 17 genes including LRIG1 (14%), PTPRT (14%), and PPP2R2A (14%) gene were also noted (**Figure 2A**). A previous study also reported that TP53 mutations E298X, R273C, and G279R were detected in three PPLELC patients, resulting in a TP53 mutation rate of 6.5% (Chang et al., 2011). Notably, the frequency of the TP53 mutation rate was much lower than the result in this study. These findings could be attributed to the following reasons. Firstly, different detection methods produced different proportions of positive cells and different staining intensities. Secondly, increasing the sample size improved the accuracy of the TP53 mutation rate. TP53 mutation could lead to increase in tumor potential gene mutation and PD-L1 expression and may be served as a pair of potential predictive factors in guiding anti-PD-1/PD-L1 immunotherapy (Gibbons et al., 2014; Cortez et al., 2015). TP53 or KRAS mutation patients significantly prolonged PFS compared with wild-type patients who underwent pembrolizumab treatment (mPFS, TP53-mut vs. KRAS-mut vs. wild-type: 14.5 vs. 14.7 vs. 3.5 months p = 0.012) (Dong et al., 2017). Yin L et al. found that the presence of CYLD enhanced the chemosensitivity of bladder cancer to gemcitabine (Yin et al., 2016). Additionally, the curative effect of gemcitabine combined with cisplatin was significantly better compared to pemetrexed combined with cisplatin for PPLELC ($p < 0.001$) (Hong et al., 2019) indicating that the presence of CYLD in PPLELC may also enhance the chemosensitivity of gemcitabine. The efficacy of gemcitabine in the treatment of EBV-related tumors could be improved by ganciclovir (Feng et al., 2004), and the efficacy of ganciclovir combined with gemcitabine in the treatment of PPLELC needs further exploration.

PD-L1 expression is higher in PPLELC compared to conventional NSCLCs (Fang et al., 2014; Chang et al., 2015),

assuming that it is a potential biomarker and rational therapeutic target. All PPLELC patients expressed PD-L1, including 42.6% (3/7) high expression and 57.1% (4/7) moderate expression (**Figure 3B**). The high expression of PD-L1 in PPLELC sheds light on the possibility of using immunotherapy in this subtype of lung cancer. The up-regulation of PD-L1 expression may be related to EBV infection and EBV-related tumors may be more beneficial in the treatment of immune checkpoint inhibitors (Jiang et al., 2015). However, there are few study on the comparison of immunological efficacy differences between EBV-related and non-EBV-related malignancies. 11 cases of advanced-stage PPLELC progressed continually despite multiple lines of chemotherapy but responded favorably to a PD-L1 inhibitor and the disease control rate (DCR) was 90.9% (**Figure 2**). Noteblely, two patients with -negative PD-L1 (<1%) received immunotherapy, one patient achieved PD, while another patient achieved PR. The PD-L1-negative patient (PD-L1 < 1%) also received immunotherapy and responded to nivolumab for 48 cycles that lasted 21.9 months, indicating a long-term tumor response of PD-1/PD-L1 inhibitors in patients irrespective of PD-L1 status (Brahmer et al., 2015). Therefore, only a minority of patients acquired a good response to immunotherapy, although it could be highly effective. The effectiveness of immunotherapy may not only depend on the expression of PD-L1 in tumor tissue but also on whether there are sufficient immune effector cells in the tumor microenvironment (TME) (Zhang and Chen, 2016). A large number of lymphocytes with $CD8^+$ and TIA-1+ and cytotoxic T cells were detected, and plasma cell infiltration occurred in the stroma around the PPLELC tumor cells (Qin et al., 2019; Sathirareuangchai and Hirata, 2019). This might be one of the reasons why PPLELC patients benefit from immunotherapy despite the low expression of PD-L1.

The high nonsynonymous TMB was associated with improved objective response, durable clinical benefit, and progression-free survival (PFS) after immunotherapy (Rizvi et al., 2015; Devarakonda et al., 2018). Additionally, previous data showed that the median TMB of patients with PPLELC was 1.6 mutations/Mb, which was significantly lower than that of patients with lung adenocarcinoma in the TCGA data set of cancer genome map ($p < 0.01$) (Xie et al., 2020; Cancer Genome Atlas Research Network, 2014). Herein, the median TMB of our cohort was 2.5 (range 1.7–2.5 mutations/Mb, **Figure 4B**). Therefore, TMB could not fully reflect the immunogenicity of the tumor. Patients with low TMB could also respond to immunotherapy, while patients with high TMB may not have a good effect, which depends on human leukocyte antigen (HLA) classification (Chowell et al., 2018). The more diversity of HLA, especially the more super-subtypes of HLA-B44, the more kinds of new antigens can be presented, and the better the efficacy of immune drugs will be improved. However, Mcgranahan et al. (2017) confirmed that patients with clonal HLA heterozygote deletion had higher mutation levels and obvious subclonal expression of tumor cells in NSCLC patients compared to HLA patients without heterozygote mutation. Although their TMB levels were high, the immunotherapy was ineffective. Therefore, TMB and PD-L1 could complement each other to predict the efficacy of immunotherapy. Yet, no correlation between low/high PD-L1 expression and TMB value was observed (Fisher's exact test, $p = 0.24$) which need to be verified by additional prospective studies.

There are several limitations. The sample size was small and this was retrospective study. Due to its rare incidence, only 18 patients were diagnosed with PPLELC at our center over the last decade and only seven patients performed NGS. Statistical validation/significance of the data may not be adequate as the sample size is very limited. Second, immunotherapy is a relatively new treatment method for PPLELC. 11 advanced PPLELC patients who received immunotherapy were analyzed retrospectively with different regimens, aimed to demonstrate that the use of immune checkpoint may be promising beneficial treatments. Whether heterogeneity exists in the treatments used also requires further exploration. Multicenter studies with sufficiently long observation periods will be carried out in the future to provide more evidence.

In summary, PPLELC is a rare subtype of NSCLC which is closely related to EBV. Surgery is the first choice of treatment for early diagnosed patients, while and radiotherapy/chemotherapy can prolong the survival of advanced PPLELC patients. Although EGFR-sensitive mutations and other classical lung cancer gene mutations are rare, however, CYLD may be a new therapeutic target for PPLELC. Despite the low TMB, the PD-L1 positivity of a majority of the tumors raises the potential of utilizing checkpoint immunotherapy as a treatment regimen that could benefit PPLELC patients. However, the mechanism of immune checkpoint blockade in PPLELC remains yet unclear, and the activity of immune checkpoint inhibitors observed in other virus-associated cancers warrants further evaluation of this new class of cancer therapeutics in patients with PPLELC.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0002253.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Zhejiang Cancer Hospital (IRB-2020-208). 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

YF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing–original draft, Writing–review and editing. QS: Conceptualization, Data curation, Formal analysis, Investigation,

Methodology, Writing–review and editing. JG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing–review and editing. JQ: Conceptualization, Formal analysis, Investigation, Methodology, Writing–review and editing. HL: Formal analysis, Investigation, Methodology, Supervision, Funding acquisition, Project administration, Writing–original draft, Writing–review and editing.

FUNDING

This study was funded by the Zhejiang Province Medical Science and Technology Project (No. 2020ZH001).

ACKNOWLEDGMENTS

The abstract of this paper was presented at the 2021 ASCO Conference named "Molecular and clinical characteristics of primary pulmonary lymphoepithelioma-like carcinoma" (e20541) as an e-poster.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.736940/full#supplementary-material

REFERENCES

Anand, A., Zayac, A., Curtiss, C., and Graziano, S. (2018). Pulmonary Lymphoepithelioma-like Carcinoma Disguised as Squamous Cell Carcinoma. *J. Thorac. Oncol.* 13 (5), e75. doi:10.1016/j.jtho.2017.11.133

Bégin, L. R., Eskandari, J., Joncas, J., and Panasci, L. (1987). Epstein-Barr Virus Related Lymphoepithelioma-like Carcinoma of Lung. *J. Surg. Oncol.* 36 (4), 280–283. doi:10.1002/jso.2930360413

Brahmer, J., Reckamp, K. L., Baas, P., Crinò, L., Eberhardt, W. E. E., Poddubskaya, E., et al. (2015). Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-small-cell Lung Cancer. *N. Engl. J. Med.* 373 (2), 123–135. doi:10.1056/nejmoa1504627

Cancer Genome Atlas Research Network (2014). Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* 511 (7511), 543–550. doi:10.1038/nature13385

Carbone, D. P., Reck, M., Paz-Ares, L., Creelan, B., Horn, L., Steins, M., et al. (2017). First-Line Nivolumab in Stage IV or Recurrent Non-small-cell Lung Cancer. *N. Engl. J. Med.* 376 (25), 2415–2426. doi:10.1056/nejmoa1613493

Chang, Y.-L., Wu, C.-T., Shih, J.-Y., and Lee, Y.-C. (2011). Unique P53 and Epidermal Growth Factor Receptor Gene Mutation Status in 46 Pulmonary Lymphoepithelioma-like Carcinomas. *Cancer Sci.* 102 (1), 282–287. doi:10.1111/j.1349-7006.2010.01768.x

Chang, Y.-L., Yang, C.-Y., Lin, M.-W., Wu, C.-T., and Yang, P.-C. (2015). PD-L1 Is Highly Expressed in Lung Lymphoepithelioma-like Carcinoma: A Potential Rationale for Immunotherapy. *Lung Cancer* 88 (3), 254–259. doi:10.1016/j.lungcan.2015.03.017

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

Cortez, M. A., Ivan, C., Valdecanas, D., Wang, X., Peltier, H. J., Ye, Y., et al. (2015). PDL1 Regulation by P53 via miR-34. *J. Natl. Cancer Inst.* 108 (1), djv303. doi:10.1093/jnci/djv303

Da-yun, P., Qiu-lin, L., Chen, X.-d., Ri-quan, L., zhang, W., Han, L., et al. (2017). Clinicopathologic Study of Lung Primary Lymphoepithelial-like Carcinoma: an Analysis of 18 Cases. *Chin. J. Clin. Res.* 30 (4), 467–470. doi:10.13429/j.cnki.cjcr.2017.04.010

Devarakonda, S., Rotolo, F., Tsao, M.-S., Lanc, I., Brambilla, E., Masood, A., et al. (2018). Tumor Mutation Burden as a Biomarker in Resected Non-small-cell Lung Cancer. *J. Clin. Oncol.* 36 (30), 2995–3006. doi:10.1200/jco.2018.78.1963

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 (12), 3012–3024. doi:10.1158/1078-0432.ccr-16-2554

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

Fang, W., Hong, S., Chen, N., He, X., Zhan, J., Qin, T., et al. (2015). PD-L1 Is Remarkably Over-expressed in EBV-Associated Pulmonary Lymphoepithelioma-like Carcinoma and Related to Poor Disease-free Survival. *Oncotarget* 6 (32), 33019–33032. doi:10.18632/oncotarget.5028

Fang, W., Zhang, J., Hong, S., Zhan, J., Chen, N., Qin, T., et al. (2014). EBV-driven LMP1 and IFN-γ Up-Regulate PD-L1 in Nasopharyngeal Carcinoma: Implications for Oncotargeted Therapy. *Oncotarget* 5 (23), 12189–12202. doi:10.18632/oncotarget.2608

Feng, W.-h., Hong, G., Delecluse, H.-J., and Kenney, S. C. (2004). Lytic Induction Therapy for Epstein-Barr Virus-Positive B-Cell Lymphomas. *J. Virol.* 78 (4), 1893–1902. doi:10.1128/jvi.78.4.1893-1902.2004

Gibbons, D. L., Byers, L. A., and Kurie, J. M. (2014). Smoking, P53 Mutation, and Lung Cancer. *Mol. Cancer Res.* 12 (1), 3–13. doi:10.1158/1541-7786.mcr-13-0539

Grimes, B. S., Albores, J., and Barjaktarevic, I. (2015). A 65-Year-Old Man with Persistent Cough and Large Nodular Opacity. *Chest* 147 (1), e13. doi:10.1378/chest.14-1172

Han, A., Xiong, M., Gu, Y., Lin, S., and Xiong, M. (2001). Clinicopathologic Features and Prognosis of Lymphoepithelioma-like Carcinoma of the Lung. *Zhonghua Bing Li Xue Za Zhi* 30 (5), 328–331. doi:10.1016/S1579-2129(06)60325-X

He, J., Shen, J., Pan, H., Huang, J., Liang, W., and He, J. (2015). Pulmonary Lymphoepithelioma-like Carcinoma: a Surveillance, Epidemiology, and End Results Database Analysis. *J. Thorac. Dis.* 7 (12), 2330–2338. doi:10.3978/j.issn.2072-1439.2015.12.62

Ho, J. C., Lam, D. C., Wong, M. K., Lam, B., Ip, M. S., and Lam, W. K. (2009). Capecitabine as Salvage Treatment for Lymphoepithelioma-like Carcinoma of Lung. *J. Thorac. Oncol.* 4 (9), 1174–1177. doi:10.1097/jto.0b013e3181b28f15

Ho, J. C., Lam, W. K., Ooi, G. C., Lam, B., and Tsang, K. W. (2000). Chemoradiotherapy for Advanced Lymphoepithelioma-like Carcinoma of the Lung. *Respir. Med.* 94 (10), 943–947. doi:10.1053/rmed.2000.0856

Hong, S., Liu, D., Luo, S., Fang, W., Zhan, J., Fu, S., et al. (2019). The Genomic Landscape of Epstein-Barr Virus-Associated Pulmonary Lymphoepithelioma-like Carcinoma. *Nat. Commun.* 10 (1), 3108. doi:10.1038/s41467-019-10902-w

Huang, C.-J., Chan, K.-Y., Lee, M.-Y., Hsu, L.-H., Chu, N.-M., Feng, A.-C., et al. (2007). Computed Tomography Characteristics of Primary Pulmonary Lymphoepithelioma-like Carcinoma. *Br. J. Radiol.* 80 (958), 803–806. doi:10.1259/bjr/27788443

Huang, C. J., Feng, A. C., Fang, Y. F., Ku, W. H., Chu, N. M., Yu, C. T., et al. (2012). Multimodality Treatment and Long-Term Follow-Up of the Primary Pulmonary Lymphoepithelioma-like Carcinoma. *Clin. Lung Cancer* 13 (5), 359–362. doi:10.1016/j.cllc.2012.01.002

Jiang, L., Wang, L., Li, P. F., Zhang, X. K., Chen, J. W., Qiu, H. J., et al. (2015). Positive Expression of Programmed Death Ligand-1 Correlates with superior Outcomes and Might Be a Therapeutic Target in Primary Pulmonary Lymphoepithelioma-like Carcinoma. *Onco Targets Ther.* 8, 1451–1457. doi:10.2147/OTT.S84234

Jiang, W.-Y., Wang, R., Pan, X.-F., Shen, Y.-Z., Chen, T.-X., Yang, Y.-H., et al. (2016). Clinicopathological Features and Prognosis of Primary Pulmonary Lymphoepithelioma-like Carcinoma. *J. Thorac. Dis.* 8 (9), 2610–2616. doi:10.21037/jtd.2016.08.40

Kim, C., Rajan, A., DeBrito, P. A., and Giaccone, G. (2016). Metastatic Lymphoepithelioma-like Carcinoma of the Lung Treated with Nivolumab: a

Case Report and Focused Review of Literature. *Eur. J. Cancer* 5 (6), 720–726. doi:10.21037/tlcr.2016.11.06

Kumar, V., Dave, V., Harris, J., and Huang, Y. (2017). Response of Advanced Stage Recurrent Lymphoepithelioma-like Carcinoma to Nivolumab. *Immunotherapy* 9 (12), 955–961. doi:10.2217/imt-2017-0067

Liang, Y., Wang, L., Zhu, Y., Lin, Y., Liu, H., Rao, H., et al. (2012). Primary Pulmonary Lymphoepithelioma-like Carcinoma. *Cancer* 118 (19), 4748–4758. doi:10.1002/cncr.27452

Lin, Z., Situ, D., Chang, X., Liang, W., Zhao, M., Cai, C., et al. (2016). Surgical Treatment for Primary Pulmonary Lymphoepithelioma-like Carcinoma. *Interact Cardiovasc. Thorac. Surg.* 23 (1), 41–46. doi:10.1093/icvts/ivw064

Liu, Q., Ma, G., Yang, H., Wen, J., Li, M., Yang, H., et al. (2014). Lack of Epidermal Growth Factor Receptor Gene Mutations in Exons 19 and 21 in Primary Lymphoepithelioma-like Carcinoma of the Lung. *Thorac. Cancer* 5, 63–67. doi:10.1111/1759-7714.12060

Ma, H., Wu, Y., Lin, Y., Cai, Q., Ma, G., and Liang, Y. (2013). Computed Tomography Characteristics of Primary Pulmonary Lymphoepithelioma-like Carcinoma in 41 Patients. *Eur. J. Radiol.* 82 (8), 1343–1346. doi:10.1016/j.ejrad.2013.02.006

Mcgranahan, N., Rosenthal, R., Hiley, C. T., Rowan, A. J., Watkins, T. B. K., Wilson, G. A., et al. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171 (6), 1259–e11. doi:10.1016/j.cell.2017.10.001

Mo, Y., Shen, J., Zhang, Y., Zheng, L., Gao, F., Liu, L., et al. (2014). Primary Lymphoepithelioma-like Carcinoma of the Lung. *J. Thorac. Imaging* 29 (4), 246–251. doi:10.1097/rti.0000000000000070

Narayanan, A., Knollmann, F. D., Walby, J. A. S., Lim, S., Gandara, D. R., and Riess, J. W. (2019). EBV-positive Primary Pulmonary Lymphoepithelioma-like Carcinoma Response to PD-L1 Blockade. *Clin. Lung Cancer* 20 (3), e238–3. doi:10.1016/j.cllc.2018.12.015

Peters, S., Creelan, B., Hellmann, M. D., Socinski, M. A., and Carbone, D. (2017). Abstract CT082: Impact of Tumor Mutation Burden on the Efficacy of First-LineNiv in Stage IV or Recurrent Non-small Cell Lung Cancer: An Exploratory Analysis of CheckMate 026. *Cancer Res.* 77 (13 Suppl. l), S124–S129. doi:10.1158/1538-7445.AM2017-CT082

Qin, Y., Gao, G., Xie, X., Zhu, Z., Guan, W., Lin, X., et al. (2019). Clinical Features and Prognosis of Pulmonary Lymphoepithelioma-like Carcinoma: Summary of Eighty-Five Cases. *Clin. Lung Cancer* 20 (3), e329–e337. doi:10.1016/j.cllc.2018.12.014

Qiu, Z.-X., Zhou, P., and Wang, K. (2019). Primary Pulmonary Lymphoepithelioma-like Carcinoma Response Favorably to Nivolumab: A Case Report. *Ott* 12, 8595–8600. doi:10.2147/ott.s219512

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., et al. (2015). Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-small Cell Lung Cancer. *Science* 348 (6230), 124–128. doi:10.1126/science.aaa1348

Sathirareuangchai, S., and Hirata, K. (2019). Pulmonary Lymphoepithelioma-like Carcinoma. *Arch. Pathol. Lab. Med.* 143 (8), 1027–1030. doi:10.5858/arpa.2018-0149-rs

Sun, Y.-H., Lin, S.-W., Hsieh, C.-C., Yeh, Y.-C., Tu, C.-C., and Chen, K.-J. (2014). Treatment Outcomes of Patients with Different Subtypes of Large Cell Carcinoma of the Lung. *Ann. Thorac. Surg.* 98 (3), 1013–1019. doi:10.1016/j.athoracsur.2014.05.012

Tang, Z., Fang, R., Tong, G., Liu, P., Ou, Z. a., and Tang, Y. (2020). Overcoming Resistance to Anti-PD-1 Immunotherapy in Lymphoepithelioma-like Carcinoma: A Case Report and Review of the Literature. *Lung Cancer* 146, 335–340. doi:10.1016/j.lungcan.2020.06.027

Wang, C. Z., Ning, H., and Wang, Y. M. (2013). Clinicopathologic Analysis of Primary Pulmonary Lymphoepithelioma-like Carcinoma [J]. *Chongqing Med.* 42 (28), 3425–3426.

Wang, H., Agulnik, J., Kasymjanova, G., Wang, A., Jiménez, P., Cohen, V., et al. (2018). Cytology Cell Blocks Are Suitable for Immunohistochemical Testing for PD-L1 in Lung Cancer. *Ann. Oncol.* 29 (6), 1417–1422. doi:10.1093/annonc/mdy126

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. [Google Scholar].

Xie, M., Wu, X., Wang, F., Zhang, J., Ben, X., Zhang, J., et al. (2018). Clinical Significance of Plasma Epstein-Barr Virus Dna in Pulmonary Lymphoepithelioma-like Carcinoma (Lelc) Patients. *J. Thorac. Oncol.* 13 (2), 218–227. doi:10.1016/j.jtho.2017.10.031

Xie, Z., Liu, L., Lin, X., Xie, X., Gu, Y., Liu, M., et al. (2020). A Multicenter Analysis of Genomic Profiles and PD-L1 Expression of Primary Lymphoepithelioma-like Carcinoma of the Lung. *Mod. Pathol.* 33 (4), 626–638. doi:10.1038/s41379-019-0391-9

Yeh, Y.-C., Kao, H.-L., Lee, K.-L., Wu, M.-H., Ho, H.-L., and Chou, T.-Y. (2019). Epstein-Barr Virus-Associated Pulmonary Carcinoma. *Am. J. Surg. Pathol.* 43 (2), 211–219. doi:10.1097/pas.0000000000001173

Yin, L., Liu, S., Li, C., Ding, S., Bi, D., Niu, Z., et al. (2016). CYLD Downregulates Livin and Synergistically Improves Gemcitabine Chemosensitivity and Decreases Migratory/invasive Potential in Bladder Cancer: the Effect Is Autophagy-Associated. *Tumor Biol.* 37 (9), 12731–12742. doi:10.1007/s13277-016-5157-0

Zhang, Y., and Chen, L. (2016). Classification of Advanced Human Cancers Based on Tumor Immunity in the MicroEnvironment (TIME) for Cancer Immunotherapy. *JAMA Oncol.* 2 (11), 1403–1404. doi:10.1001/jamaoncol.2016.2450

Zhou, N., Lin, Y., Peng, X., Wang, Y., and Wang, Y. (2019). Thorough Survey and Analysis of Pulmonary Lymphoepithelioma-like Carcinoma in Macau and Multimodality Treatment for Advanced Disease. *Lung Cancer* 138, 116–123. doi:10.1016/j.lungcan.2019.10.004

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 Fan, Shan, Gong, Qin 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.

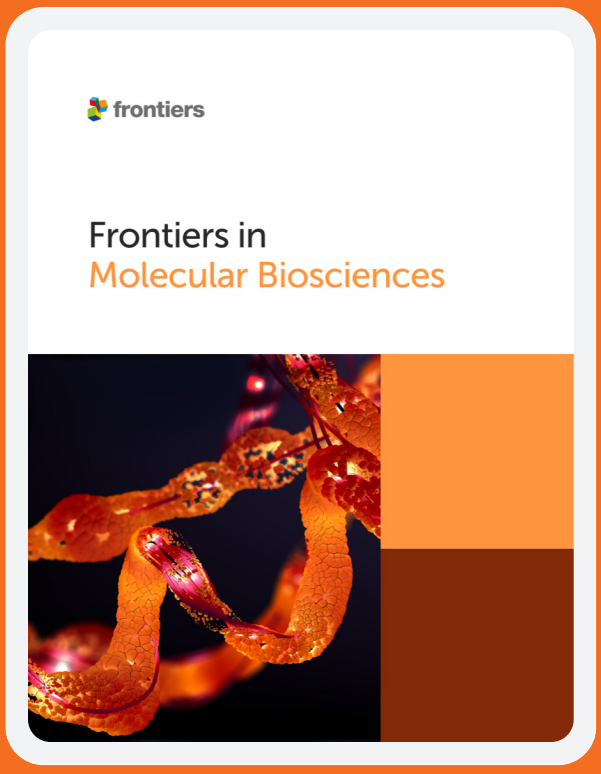
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

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