BIOINFORMATICS TOOLS (AND WEB SERVER) FOR CANCER BIOMARKER DEVELOPMENT, VOLUME II

EDITED BY: Xiangqian Guo, Longxiang Xie, Liuyang Wang, Jing Zhao and Wan Zhu

PUBLISHED IN: Frontiers in Genetics



frontiers Research Topics



frontiers

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 978-2-88976-383-2
DOI 10.3389/978-2-88976-383-2

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

BIOINFORMATICS TOOLS (AND WEB SERVER) FOR CANCER BIOMARKER DEVELOPMENT, VOLUME II

Topic Editors:
Xiangqian Guo, Henan University, China
Longxiang Xie, Henan University, China
Liuyang Wang, Duke University, United States
Jing Zhao, Chongqing Medical University, China
Wan Zhu, Stanford University, United States

Citation: Guo, X., Xie, L., Wang, L., Zhao, J., Zhu, W., eds. (2022). Bioinformatics Tools (and Web Server) for Cancer Biomarker Development, Volume II. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-383-2

Table of Contents

05 ***Editorial: Bioinformatics Tools (and Web Server) for Cancer Biomarker Development, Volume II***
Longxiang Xie, Liuyang Wang, Wan Zhu, Jing Zhao and Xiangqian Guo

09 ***PER1 Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Ovarian Cancer***
Mali Chen, Lili Zhang, Xiaolong Liu, Zhen Ma and Ling Lv

22 ***Immune-Related Gene Expression Analysis Revealed Three lncRNAs as Prognostic Factors for Colon Cancer***
Xiao-Liang Xing, Ti Zhang, Zhi-Yong Yao, Chaoqun Xing, Chunxiao Wang, Yuan-Wu Liu and Minjiang Huang

30 ***Establishment and Validation of a Comprehensive Prognostic Model for Patients With HNSCC Metastasis***
Yajun Shen, Lingyu Li, Yunping Lu, Min Zhang, Xin Huang and Xiaofei Tang

49 ***Construction and Validation of an Immune-Based Prognostic Model for Pancreatic Adenocarcinoma Based on Public Databases***
Miaobin Mao, Hongjian Ling, Yuping Lin, Yanling Chen, Benhua Xu and Rong Zheng

68 ***Identification and Validation of a Ferroptosis-Related Long Non-coding RNA Signature for Predicting the Outcome of Lung Adenocarcinoma***
Zhiyuan Zheng, Qian Zhang, Wei Wu, Yan Xue, Shuhan Liu, Qiaoqian Chen and Donghong Lin

79 ***The CPNE Family and Their Role in Cancers***
Haicheng Tang, Pei Pang, Zhu Qin, Zhangyan Zhao, Qingguo Wu, Shu Song and Feng Li

86 ***A Prognostic Autophagy-Related Gene Pair Signature and Small-Molecule Drugs for Hepatocellular Carcinoma***
ZeBing Song, GuoPei Zhang, Yang Yu and ShaoQiang Li

99 ***Tumor Expression Profile Analysis Developed and Validated a Prognostic Model Based on Immune-Related Genes in Bladder Cancer***
Bingqi Dong, Jiaming Liang, Ding Li, Wenping Song, Shiming Zhao, Yongkang Ma, Jinbo Song, Mingkai Zhu and Tiejun Yang

111 ***A Robust Seven-Gene Signature Associated With Tumor Microenvironment to Predict Survival Outcomes of Patients With Stage III–IV Lung Adenocarcinoma***
Hao Zhao, Xuening Zhang, Lan Guo, Songhe Shi and Ciyong Lu

124 ***Identification of Signature Genes Associated With Invasiveness and the Construction of a Prognostic Model That Predicts the Overall Survival of Bladder Cancer***
Yang He, Yongxin Wu, Zhe Liu, Boping Li, Ning Jiang, Peng Xu and Abai Xu

137 ***TUBA1C is a Prognostic Marker in Low-grade Glioma and Correlates with Immune Cell Infiltration in the Tumor Microenvironment***
Hua Zhu, Xinyao Hu, Lijuan Gu, Zhihong Jian, Liqin Li, Siping Hu, Sheng Qiu and Xiaoxing Xiong

148 ***Analysis of Genome-Wide Alternative Splicing Profiling and Development of Potential Drugs in Lung Adenocarcinoma***
Jing Song, Jia Liu, Dekang Lv, Xuan Meng and Xiaodong Li

159 ***Identification of Pathway-Based Biomarkers with Crosstalk Analysis for Overall Survival Risk Prediction in Breast Cancer***
Xiaohua Liu, Lili Su, Jingcong Li and Guoping Ou

169 ***Molecular Subclassification Based on Crosstalk Analysis Improves Prediction of Prognosis in Colorectal Cancer***
Xiaohua Liu, Lili Su, Jingcong Li and Guoping Ou

179 ***High Expression of C1ORF112 Predicts a Poor Outcome: A Potential Target for the Treatment of Low-Grade Gliomas***
Zhe Zhang, Zilong Tan, Qiaoli Lv, Lichong Wang, Kai Yu, Huan Yang, Huaizhen Liang, Tianzhu Lu, Yulong Ji, Junjun Chen, Wei He, Zhen Chen, Shuhui Chen and Xiaoli Shen

191 ***Comprehensive Analysis of Inhibitor of Apoptosis Protein Expression and Prognostic Significance in Non–Small Cell Lung Cancer***
Jun Liu, Yi Lu, Wenan Huang and Zhibo He

198 ***High Expression of CEMIP Correlates Poor Prognosis and the Tumur Microenvironment in Breast Cancer as a Promisingly Prognostic Biomarker***
Xingxing Dong, Yalong Yang, Qianqian Yuan, Jinxuan Hou and Gaosong Wu

220 ***Identification and Validation of DEPDC1B as an Independent Early Diagnostic and Prognostic Biomarker in Liver Hepatocellular Carcinoma***
Xiaoyan Fan, Junye Wen, Lei Bao, Fei Gao, You Li and Dongwei He

233 ***A Prognostic Model for Predicting Tumor Mutation Burden and Tumor-Infiltrating Immune Cells in Bladder Urothelial Carcinoma***
Chengbang Wang, Shaohua Chen, Songheng Li and Hua Mi

245 ***DNTTIP1 is a Prognostic Biomarker Correlated With Immune Infiltrates in Hepatocellular Carcinoma: A Study Based on The Cancer Genome Atlas Data***
Changyu Zhu, Rongsheng Tong, Xiaolei Jiang, Hua Xiao, Jianmei Guan and Jianchen Shu

256 ***Establishment and Validation of a 5 m6A RNA Methylation Regulatory Gene Prognostic Model in Low-Grade Glioma***
Zhiqun Bai, Xuemei Wang and Zhen Zhang

268 ***Identification and Validation of a Pyroptosis-Related Prognostic Model for Gastric Cancer***
Chaowei Liang, Jiaxin Fan, Chaojie Liang and Jiansheng Guo

281 ***A Novel Inflammatory Response–Related Gene Signature Improves High-Risk Survival Prediction in Patients With Head and Neck Squamous Cell Carcinoma***
Yanxun Han, Zhao Ding, Bangjie Chen, Yuchen Liu and Yehai Liu

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 10 October 2022
DOI 10.3389/fgene.2022.959159

Check for updates

OPEN ACCESS

EDITED BY
Anton A. Buzdin,
European Organisation for Research and Treatment of Cancer, Belgium

*CORRESPONDENCE
Xiangqian Guo,
xqguo@henu.edu.cn

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

RECEIVED 01 June 2022
ACCEPTED 05 September 2022
PUBLISHED 10 October 2022

CITATION
Xie L, Wang L, Zhu W, Zhao J and Guo X (2022), Editorial: Bioinformatics tools (and web server) for cancer biomarker development, volume II. *Front. Genet.* 13:959159. doi: 10.3389/fgene.2022.959159

COPYRIGHT
© 2022 Xie, Wang, Zhu, Zhao 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.

Editorial: Bioinformatics tools (and web server) for cancer biomarker development, volume II

Longxiang Xie[1], Liuyang Wang[2], Wan Zhu[3], Jing Zhao[4] and Xiangqian Guo[1]*

[1]Department of Predictive Medicine, Institute of Biomedical Informatics, Bioinformatics Center, Henan Provincial Engineering Center for Tumor Molecular Medicine, School of Basic Medical Sciences, Academy for Advanced Interdisciplinary Studies, Henan University, Kaifeng, China, [2]Department of Molecular Genetics and Microbiology, School of Medicine, Duke University, Durham, NC, United States, [3]Department of Anesthesia, Stanford University, Stanford, CA, United States, [4]Department of Pathophysiology, Chongqing Medical University, Chongqing, China

KEYWORDS

bioinformatics, webserver, prognostic, biomarker, TCGA, GEO

Editorial on the Research Topic
Bioinformatics tools (and web server) for cancer biomarker development, volume II

Cancer is a major disease and a heavy burden to public health. With the rapid development of high-throughput (HTP) technologies, increasing multi-omics data have been produced to identify biomarkers to facilitate the risk assessment, early detection, prognosis, and treatment response prediction of tumors, and these biomarkers have helped to successfully decrease the mortality rate of certain types of cancer patients (Zou and Wang, 2019). In this Research Topic, we collected 23 articles representing recent advanced biomarkers studies for tumor diagnosis, prognosis, and treatment response.

Due to the wide application of next generation sequencing technologies in tumor research in recent decades, a few public omics data depositories, including The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and International Cancer Genome Consortium (ICGC), have been established and have aggregated a large number of clinical tumor tissue omics data and clinical data of tumor patients. These public databases provide reliable data support and additional opportunities for comprehensive biomarker identification (Tomczak et al., 2015; Ren et al., 2020). Song et al. investigated the TCGA and ICGC data of hepatocellular carcinoma (HCC), and performed the differential expression analysis and Cox regression analysis to identify core genes associated with HCC clinical outcomes, and a 6-autophagy-related-gene-pair (ARGP) prognostic signature was identified for overall survival (OS) of HCC patients. Based on the TCGA and GEO datasets, Shen et al. identified 108 differentially expressed genes, which were mainly involved in the proliferation and metastasis of head and neck squamous cell

carcinoma (HNSCC). Furthermore, they constructed a novel multi-factor prognostic model of HNSCC and verified the reliability of the prognostic model in an additional 36 HNSCC patients.

In recent years, more new multi-omics databases were available and provided more opportunities for cancer diagnosis, treatments, and prevention, such as Chinese Glioma Genome Atlas (CGGA), which contains a wide range of data derived from whole-exome sequencing (WES), mRNA sequencing and microarray, DNA methylation microarray, and microRNA microarray analyses, as well as comprehensive clinical data (Zhao et al., 2021). Bai et al. evaluated m6A methylation regulatory genes and constructed a prognostic model of low-grade glioma (LGG) based on data of 495 LGGs from TCGA and 172 LGGs from the CGGA. This model contained 5 m6A-methylation- related genes and might classify LGGs into high- or low-risk subgroups. High-throughput technologies have been extended to characterize genomic status including but not limited to DNA methylation modification, genetic alteration, and gene expression regulation. Dong et al. explored the expression profiling data of CEMIP in different kinds of human cancers, and found that CEMIP was a prognostic and metastatic biomarker of breast cancer (BC). Their finding also uncovered that gene expression of CEMIP was mediated by TP53 mutation and DNA hypomethylation.

By integrating public accessible datasets, several powerful bioinformatics webservers/tools including KM plotter, GEPIA (Gene Expression Profiling Interactive Analysis), Oncomine, and TIMER (Tumor Immune Estimation Resource), have been developed to analyze the association or differential expression of genes with (in) distinct clinical factors in tumors (Li et al., 2017; Li et al., 2021). These online webservers/tools would help researchers to discover novel prognostic biomarkers which might be related to tumorigenesis or tumor malignant progression, and can provide more data basis for prognosis and therapeutic target identification of tumors. Based on a series of online tools, such as TIMER, Oncomine, and GEPIA, Chen et al. systematically analyzed the expression and prognostic value of PER1 in ovarian cancer (OV), and showed that low PER1 expression was related with poor prognosis of OV patients and concluded that PER1 may be a novel prognostic biomarker for OV. Zhang et al. used GEPIA to evaluate key co-expressed genes with C1ORF112 in LGGs, and performed gene ontology (GO) and Kyoto Gene and Genome Encyclopedia (KEGG) pathway analyses using the DAVID tool and identified a relationship between C1ORF112 expression and immune cell infiltration by TIMER. Their data suggested that C1ORF112 was closely related to the OS of patients with LGG, and was a prognostic biomarker of LGG. Liu et al. explored the relationship between inhibitors of apoptosis proteins and non-small cell lung cancer (NSCLC) progression through online bioinformatics tools, and found that BIRC1 is a potential biomarker associated with OS and BIRC5 is a potential diagnostic and staging biomarker of NSCLC patients. By using multiple bioinformatics analyses tools, Zhu et al. identified and verified that high expression of TUBA1C is related to a poor prognosis of LGG patients. Based on the TCGA dataset, Fan et al. identified DEPDC1B as a potential biomarker for diagnosis and prognosis of liver hepatocellular carcinoma (LIHC). In addition, Zhu et al. found that DNTTIP1 is a prognostic biomarker associated with overall survival (OS) and disease-free survival (DFS) in LIHC as well by using the TCGA dataset. In a review of the CPNE family, Tang et al. summarized the expression pattern and clinical roles of CPNE family members in cancer development. As membrane-bound proteins, CPNEs can mediate cell dedifferentiation and immune microenvironment organization of tumors. High expression of CPNE1 and CPNE3 could predict adverse prognosis. In contrast, CPNE5 might act as a positive indicator for esophageal squamous cell carcinoma (ESCC) and multiple myeloma (MM).

Although many single-gene biomarkers have been reported, multi-gene signature is more beneficial and meaningful for cancer prognosis. Accumulated public microarray data and RNA-seq data offer the opportunities to develop multi-gene signatures (Xie et al., 2020). Using the LASSO and Cox regression models, Zhao et al. established and validated a seven-gene signature (AFAP1L2, CAMK1D, LOXL2, PIK3CG, PLEKHG1, RARRES2, and SPP1) which might serve as a prognosis stratification tool to predict survival outcomes of advanced lung adenocarcinoma (LUAD) patients. He et al. explored bladder cancer (BLCA) transcriptome from two GEO datasets and TCGA, and performed the differential expression analysis and weighted gene co-expression network analysis (WGCNA) to identify core genes, finally a three hub gene signature (VSIG2, PPFIBP2, and DENND2D) related to invasion was constructed to divide BLCA patients into high-risk and low-risk groups. The high-risk group showed a higher mortality rate than the low-risk group using a Kaplan-Meier curve. Using a gastric cancer (GC) TCGA dataset, Liang et al. established a ten gene pyroptosis-related prognostic model which was able to divide GC patients into high-risk and low-risk groups, and this pyroptosis-related model was further verified in two independent GSE84437 and GSE66229 datasets. Using univariate and multivariate Cox regression algorithms, Han et al. constructed a prognostic model consisting of 14 inflammatory-related genes that might estimate outcomes for HNSCC patients by RT-PCR.

Cancer immunotherapies have shown great benefits for multiple types of tumor, including LUAD, PAAD and BLAC

(Hayes, 2021; Zhao and Subramanian, 2021). Immune-related bioinformatics analysis would facilitate the improved treatment of tumor patients, and is expected to find new therapeutic targets. By analyzing TCGA and GEO BLCA datasets, Dong et al. constructed an immune-related eight-gene signature which was positively associated with immunotherapy response and prognosis of BLCA patients. By exploring immune-related genes (IRGs) obtained from the Immunology Database and Analysis Portal (ImmPort) database, Wang et al. constructed a prognostic model for BLCA patients and further validated this model by cross-validation. Mao et al. screened 446 significant immune-related genes in pancreatic adenocarcinoma and developed an immune-based prognostic model that was related to the status of the PAAD immune microenvironment. In order to investigate the effect of alternative splicing (AS) on the prognosis of LUAD, Song et al. integrated the prognostic AS genes and corresponding splicing factors to make a new prognostic signature with higher predictive ability than the mRNA signature.

Most cancer biomarker studies examine individual genetic variables, transcriptome alterations, and impaired protein function as distinct risk and prognostic factors. Increasing studies have demonstrated that the non-coding RNA (ncRNA) regulation network plays a significant role in cancer development (Anastasiadou et al., 2018). Xing et al. used Cox regression analysis to identify three immune-related lncRNAs (AC124067.4, LINC02604, and MIR4435-2HG), which were correlated with OS in COAD patients, and might play a role in anti-tumor immunotherapy. By analyzing 259 ferroptosis-associated genes from FerrDb, RNA-seq data and clinicopathological characteristics from TCGA, Zheng et al. established a novel prognostic signature including 10 ferroptosis-related lncRNAs, which was associated with the outcome of LUAD and tumor immune response.

Identification of high risk cancer patients is an important way to improve the clinical outcome and further facilitate understanding of the mechanisms of tumorigenesis and progression (Alifrangis et al., 2019). Based on crosstalk factorization, Liu et al. developed a new pathway activity score estimation method, and established a new prognostic classification system for colorectal cancer (CC), by which they separated the cancer patients in TCGA and three GEO datasets into aggressive (G2) and moderate (G1) subgroups, this prognostic system was shown to provide a complement to the current staging system. Using the same method, Liu et al. analyzed and classified breast cancer into G1 (moderate) and G2 (aggressive) subgroups with different OS risk. The survival outcome of the G1 subgroup was significantly better than the G2 subgroup. Most importantly, this BC risk classifier was validated in multiple BC datasets.

This Research Topic emphasises how prosperous public accessible profiling datasets and bioinformatics tools can be when widely used in oncology research. Although the tools/webservers presented here need to be improved, such as multi-omics network mapping and multi-gene signature assessment, these tools serve as a starting point for promoting the development of tumor diagnostic, prognostic, and prediction biomarker in the future.

Author contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Funding

This work was supported by the supporting program for "Central Plain Young Top Talents, grant number ZYQR201912176" and the "Program From Academy For Advanced Interdisciplinary Studies Of Henan University, grant number Y21008L".

Acknowledgments

We want to thank Meina Wang for working on and drafted this editorial, undertook part of the Editorial work.

Conflict of interest

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

Publisher's note

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

References

Alifrangis, C., Mcgovern, U., Freeman, A., Powles, T., and Linch, M. (2019). Molecular and histopathology directed therapy for advanced bladder cancer. *Nat. Rev. Urol.* 16, 465–483. doi:10.1038/s41585-019-0208-0

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18, 5–18. doi:10.1038/nrc.2017.99

Hayes, C. (2021). Cellular immunotherapies for cancer. *Ir. J. Med. Sci.* 190, 41–57. doi:10.1007/s11845-020-02264-w

Li, C., Tang, Z., Zhang, W., Ye, Z., and Liu, F. (2021). GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.* 49, W242–W246. doi:10.1093/nar/gkab418

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

Ren, N., Liang, B., and Li, Y. (2020). Identification of prognosis-related genes in the tumor microenvironment of stomach adenocarcinoma by TCGA and GEO datasets. *Biosci. Rep.* 40, BSR20200980. doi:10.1042/BSR20200980

Tomczak, K., Czerwinska, P., and Wiznerowicz, M. (2015). The cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* 19, A68–A77. doi:10.5114/wo.2014.47136

Xie, L., Cai, L., Wang, F., Zhang, L., Wang, Q., and Guo, X. (2020). Systematic review of prognostic gene signature in gastric cancer patients. *Front. Bioeng. Biotechnol.* 8, 805. doi:10.3389/fbioe.2020.00805

Zhao, X., and Subramanian, S. (2021). Cancer Immunology and immunotherapies: Mechanisms that affect antitumor immune response and treatment resistance. *Cancers (Basel)* 13, 5655. doi:10.3390/cancers13225655

Zhao, Z., Zhang, K. N., Wang, Q., Li, G., Zeng, F., Zhang, Y., et al. (2021). Chinese glioma Genome Atlas (CGGA): A comprehensive Resource with functional genomic data from Chinese glioma patients. *Genomics Proteomics Bioinforma.* 19, 1–12. doi:10.1016/j.gpb.2020.10.005

Zou, J., and Wang, E. (2019). Cancer biomarker discovery for precision medicine: New progress. *Curr. Med. Chem.* 26, 7655–7671. doi:10.2174/0929867325666180718164712

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fgene.2021.697471

PER1 Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Ovarian Cancer

Mali Chen[1†], Lili Zhang[1†], Xiaolong Liu[2], Zhen Ma[2] and Ling Lv[1]*

[1] Department of Obstetrics, Gansu Province Maternity and Child-Care Hospital, Lanzhou, China, [2] Department of Surgical Oncology, Lanzhou University Second Hospital, Lanzhou, China

Background: Period circadian protein homolog 1 (PER1) is an important component of the biorhythm molecular oscillation system and plays an important part in the development and progression of mammalian cancer. However, the correlations of PER1 with prognosis and tumor-infiltrating lymphocytes in ovarian cancer (OV) remain unclear.

Methods: The Oncomine and TIMER databases were used to examine the expression of PER1 in OV. Kaplan–Meier Plotter and PrognoScan were used to evaluate the relationship between PER1 and prognosis. Kaplan–Meier Plotter was used to analyze the relationships between PER1 and clinicopathological features of OV patients. The relationship between PER1 expression and immune infiltration in OV was investigated using the TIMER database and CIBERSORT algorithm. The STRING database was used to analyze PER1-related protein functional groups, the GeneMANIA online tool was used to analyze gene groups with similar functions to those of PER1, and Network Analyst was used to identify transcription factors that regulate PER1. The correlation between PER1 and immunoinvasion of OV was analyzed using TIMER. Finally, quantitative real-time polymerase chain reaction (qRT-PCR) was performed to detect PER1 expression.

Results: PER1 was differentially expressed in different cancer tissues, and its expression in various OV subtypes was lower than that in normal ovarian tissue. OV patients with low PER1 expression had a reduced overall survival rate. Decreased PER1 expression in stage 1 and stage 1+2 OV patients was related to poor prognosis, while increased PER1 expression in stage 3+4 patients and TP53 mutation were related to poor overall survival and progression-free survival. We identified eight genes whose expression was strongly correlated with that of PER1, as well as four transcription factors that regulate PER1. In OV, PER1 expression levels were positively correlated with infiltration levels of cells including neutrophils, regulatory T cells, and M2 macrophages, and closely related to a variety of immune markers. Reduced expression of PER1 was significantly associated with poor overall survival.

Conclusion: These findings suggest that PER1 could be used as a prognostic biomarker to determine prognosis and immune infiltration in OV patients.

Keywords: ovarian cancer, PER1, bioinformatics, diagnosis and prognosis biomarkers, immune infiltrates

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Gan Luan,
New Jersey Institute of Technology, United States
Peijie Wu,
Chengdu University of Traditional Chinese Medicine, China

****Correspondence:***
Ling Lv
1358638351@qq.com

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

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

Received: *19 April 2021*
Accepted: *19 May 2021*
Published: *17 June 2021*

Citation:
Chen M, Zhang L, Liu X, Ma Z and Lv L (2021) PER1 Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Ovarian Cancer. Front. Genet. 12:697471. doi: 10.3389/fgene.2021.697471

INTRODUCTION

Ovarian cancer (OV) is the fifth leading cause of cancer-related deaths in women worldwide (Dong et al., 2014). Despite advances in diagnosis, surgery, chemotherapy, and radiation therapy, OV continues to have a poor survival rate and the highest mortality rate among gynecological cancers. OV patients diagnosed at an advanced stage have poor prognosis and high recurrence rates. Determining effective markers for early diagnosis and prognosis is thus essential to reduce mortality and ensure effective therapy for OV (Bowtell et al., 2015).

Period circadian protein homolog 1 (PER1) is a core gene involved in circadian rhythm. It is located on the short arm of human chromosome 17 (17p12–p13.1) and has a total length of ∼16 kb (Badiu, 2003). PER1 also has an important effect on the occurrence and development of tumors. Its main functions involve regulating the body's circadian rhythm, regulating the cell cycle, and promoting DNA damage repair. Many studies have shown that PER1 directly regulates levels of reactive oxygen species in multiple organs and regulates key effectors of energy substrate utilization in response to daily behavioral rhythms and oxidative stress, thereby maintaining the daily rhythms of mitochondrial morphology and function (Sun et al., 2020). Changes in expression levels of PER1 have been found in some cancers, including pancreatic cancer (Guo et al., 2020), oral squamous cell carcinoma (Yang et al., 2020), endometrial cancer (Wang et al., 2020), colorectal cancer (Holipah and Kuroda, 2020), and non-small-cell carcinoma (Lin et al., 2020), but this has not been studied in OV. It has also been reported that the occurrence and development of cancer, as well as its prognosis and treatment outcomes, are closely related to the abnormal expression of certain circadian clock genes (Cao et al., 2009; Sahar and Sassone-Corsi, 2009; Yang et al., 2009). Based on these findings, we propose that PER1 could be used as a prognostic indicator in OV.

We also note that circadian rhythm is closely related to the human immune system (Lange et al., 2010). Abnormalities of the biological clock contribute to exhaustion of T cells and overall upregulation of immunosuppressive molecules (including programmed death ligand 1 [PD-L1] and cytotoxic T lymphocyte-associated antigen 4 [CTLA-4]); however, the impact of PER1 on the immune environment of OV remains unknown.

In this research, we analyzed the expression levels of PER1 in OV, using the online tools TIMER and Oncomine, and studied the relationships between PER1 expression and clinicopathological parameters. The prognostic value of PER1 expression in OV was determined using Kaplan–Meier Plotter and the PrognoScan database. Quantitative real-time PCR (qRT-PCR) was used to detect PER1 expression, and patients were divided into high and low PER1 expression groups for survival analysis. In addition, we used the TIMER database to determine the correlations between PER1 and immune cell infiltration. Our findings reveal an important role of PER1 in OV and also clarify the potential relationship of PER1 with tumor immunity and its underlying mechanism.

MATERIALS AND METHODS

Tissue Samples

We collected 60 OV cancerous and paracancerous tissue samples together with associated clinical information from Gansu Provincial Maternal and Child Health Hospital. None of the subjects underwent radiotherapy or chemotherapy prior to the surgery. All patient materials were obtained with informed consent, and this study was carried out with the approval of the Clinical Research Ethics Committees.

qRT-PCR

TRIzol reagent (Invitrogen, Carlsbad, CA, USA) was used for total RNA extraction from tissues. RNA was reverse transcribed into cDNA using a RevertAid First Strand cDNA Synthesis Kit (Thermo-Fisher Scientific, Waltham, MA, USA). Subsequently, with the cDNA as the template, qRT-PCR was performed using SYBR Premix Ex Taq^{TM} (TaKaRa, Otsu, Shiga, Japan). Relative gene expression was calculated by the 2–ΔΔCt method. GAPDH served as the internal reference. The primer sequences were as follows ("F" represents "forward"; "R" represents "reverse"). GAPDH: 5′-AGAAGGCTGGGGCTCATTTG-3′(F), 5′AGGGGCCATCCACAGTCTTC-3′(R); PER1: 5′-CTGCTACAGGCACGTTCAAG-3′(F), 5′-CTCAGGG ACCAAGGCTAGTG-3′(R).

Oncomine Database Analysis

Oncomine is currently the world's largest oncogene chip database and data mining platform, containing 715 gene expression datasets and 86,733 tumor and normal tissue samples. Here, the Oncomine database was used to detect the differential expression of PER1 between OV tissues and normal ovarian tissues. The data type was mRNA, and differential genes were identified by *t*-test with the following screening criteria: fold change >2, $P < 0.05$.

GEPIA Database Analysis

GEPIA is a newly developed bioinformatics platform for analysis and processing of transcriptome data from The Cancer Genome Atlas (TCGA) and GTEx databases. We analyzed the correlations between PER1 and other genes using the "Correlation Analysis" module of GEPIA with $P < 0.05$ as the screening criterion.

TIMER Database

TIMER (https://cistrome.shinyapps.io/timer/) is a database for comprehensive analysis of tumor immunity. We used the "Gene" module to analyze the correlations between PER1 gene expression and infiltration of immune cells, including CD4+ T cells, CD8+ T cells, B cells, neutrophils, macrophages, and dendritic cells. Immune cells included in the new version of TIMER showed increased infiltration levels, as presented in the results section. We used the "Diff Exp" module to detect the differential expression of the PER1 gene in different tumor tissues and normal tissues. Then, we used TIMER to detect the correlations between PER1 and immune markers of immune cells.

Survival Analysis and Prognosis Evaluation

We used Kaplan–Meier Plotter (http://kmplot.com/analysis/) and PrognoScan (http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html) for prognostic analysis. Based on the median expression of PER1, patient samples were divided into two groups for analysis with respect to overall survival (OS), disease-free survival (DFS), progression-free survival (PFS), and post-progression survival (PPS). The PrognoScan database was used to determine the prognostic value of PER1 in OV patients.

Univariate and Multivariate Cox Regression of PER1

Univariate and multivariate Cox regression were used for survival analyses. Multivariate Cox analysis was used to compare the effects of PER1 expression and other clinical characteristics on survival. Taking the patients were divided into high and low PER1 expression groups. The statistical significance level for the two-tail test was set to 0.05.

Protein–Protein Interaction Network Analysis

STRING (https://string-db.org/) was used to construct a PPI network, choosing the options "PER1" and "Human (Organism)."

GeneMANIA Analysis

GeneMANIA was used to identify gene groups with similar functions to PER1 and to construct a functional network.

Correlation Analysis of Transcription Factors and Genes

The TFs of PER1 and related genes were identified using Network Analyst (http://www.networkanalyst.ca), and their expression was analyzed using the GEPIA tool based on TCGA samples.

Statistical Analysis

The results generated from Oncomine are displayed as *P*-values and fold change values. Kaplan–Meier Plotter, PrognoScan, and GEPIA results are presented as hazard ratios, *P*-values, and Cox *P*-values. In addition, Pearson correlation, Spearman correction, and statistical significance were used to evaluate the correlations of gene expression, and the absolute value was used to determine the strength of the correlation.

RESULTS

PER1 Expression in Different Human Cancers

We first used the Oncomine database to explore the expression levels of PER1 in different tumor types. PER1 was upregulated in three tumor types and downregulated in 24 types (**Figure 1A**). Then, we analyzed the expression levels of PER1 in different tumor types using the TIMER database. Compared with normal tissues, PER1 showed low expression levels in bladder cancer, breast cancer, colon cancer, kidney cancer, lung adenocarcinoma, lung squamous cell cancer, pancreatic cancer, gastric cancer, and urethral epithelial cancer tissues, while there was no control samples were available for OV (**Figure 1B**). We further analyzed the Oncomine data and found low expression of PER1 in mucinous ovarian adenocarcinoma, clear cell ovarian adenocarcinoma, serous ovarian adenocarcinoma, ovarian endometrioid adenocarcinoma, and ovarian endometrioid cystadenocarcinoma compared with normal ovarian tissue (**Figure 1C**).

PER1 Is an Independent Prognostic Factor for OV

We analyzed the relationship between PER1 expression and the prognosis of OV patients. Kaplan–Meier analysis based on microarray data showed that low PER1 expression was associated with poor OS in OV patients ($P = 0.014$), whereas high PER1 expression was associated with poor PFS ($P = 0.029$) and PPS ($P = 0.0021$) (**Figure 2A**). We also assessed the prognostic potential of PER1 in OV using the PrognoScan database. In the GSE26712 cohort, high PER1 expression levels were associated with poor OS and DFS (**Figure 2B**). Univariate Cox analysis showed that age and PER1 were independent prognostic factors in patients with OV (**Table 1**); this was confirmed by subsequent multivariate Cox regression analysis (**Table 2**).

Confirmation of the Prognostic Value of PER1 in OV Based on Clinical Characteristics

To further understand the prognostic value of PER1 expression in OV, we used Kaplan–Meier analysis to determine the relationship between PER1 mRNA expression and clinical characteristics of OV patients. In the risk model for PER1 and OS, PER1 expression in OV stages 1, 1+2, 2+3, 2+3+4, 3, and 4, and grades 2, 2+3, and 3, PER1 low expression and OV patients. Poor prognosis (in terms of OS) was significantly related; in stage 3+4, TP53 mutation and high expression of PER1 were significantly related to poor prognosis (OS) in OV patients (**Figure 3A**).

In the risk model for PER1 and PFS, PER1 was expressed in ovarian serous tumors, stage 2+3, stage 2+3+4, stage 3, stage 4, and stage 3+4 OV, and TP53 mutant OV. High expression of PER1 was significantly correlated with poor prognosis (PFS) of OV patients (**Figure 3B**).

Identification of Key Candidate Gene Networks Based on PER1 Interactions

We constructed a gene–gene interaction network for PER1 to investigate its mechanism in OV and analyzed the function of these genes using the GeneMANIA database. The network contained 24 nodes representing genes that were highly correlated with PER1 (**Figure 4A**). In order to further explore the function of PER1, the STRING database was used to construct a PPI network. A total of 10 proteins with potential interactions with PER1 were filtered and screened, forming a very complex network (**Figure 4B**). We also found that the PER1 interaction network contained 10 common PER1-interacting genes across the GeneMANIA and STRING databases: CLOCK, TIMELESS, PER2, CRY2, CSNK1D, PER3, CSNK1E, NPAS2, CRY1, and ARNTL. We further evaluated the correlations between PER1

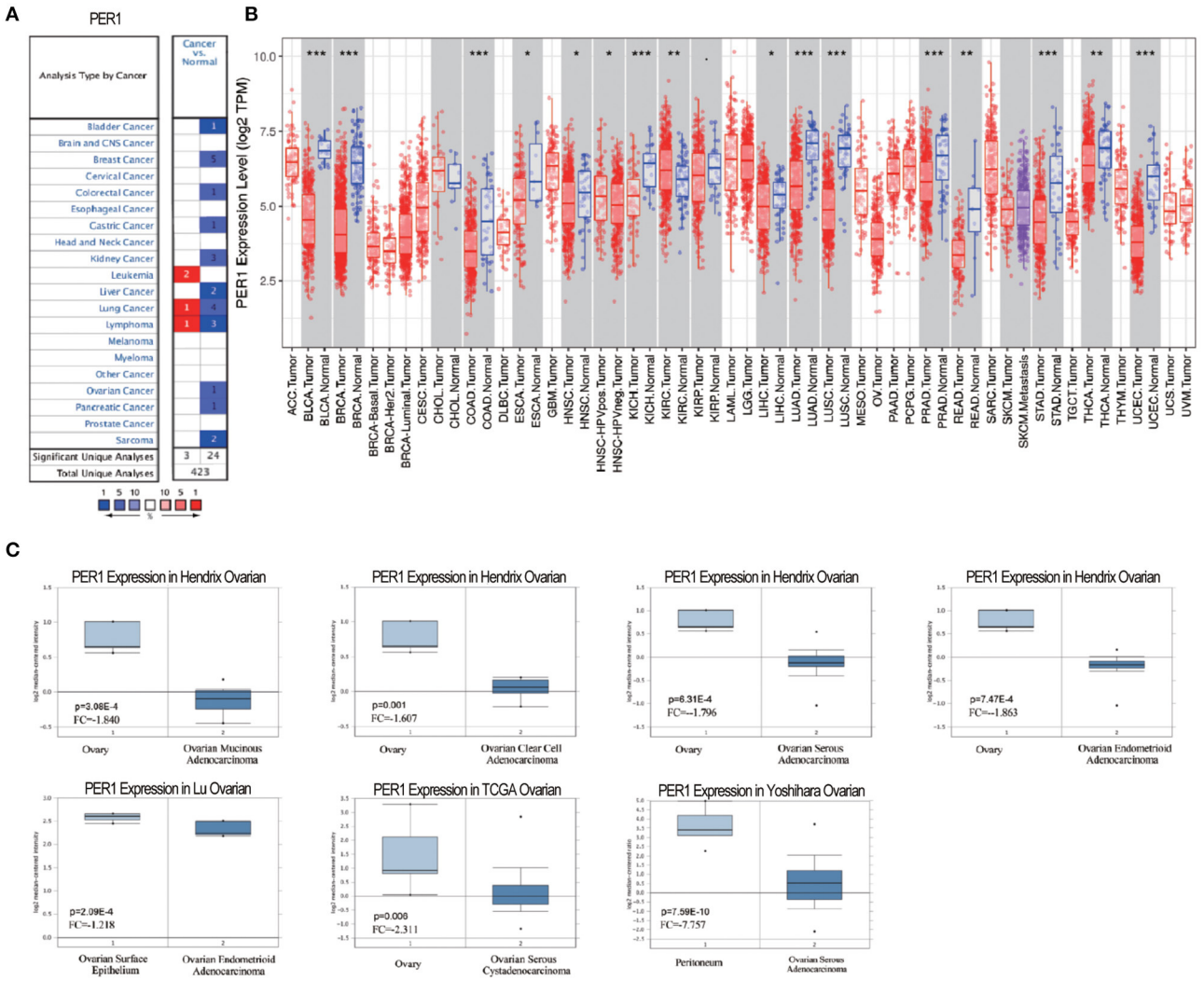


FIGURE 1 | Transcription levels of PER1 in different human cancers. **(A)** Increased or decreased expression of PER1 in various cancer tissues compared with normal tissues from the Oncomine database. **(B)** PER1 expression levels in various types of cancer from the TIMER database. **(C)** Expression of PER1 is downregulated in different types of OV. Box plots comparing PER1 expression in OV patients and normal individuals constructed based on data from the Oncomine database.

and these 10 interacting proteins based on the GEPIA database and found that the expression of PER1 in ovarian cancer was correlated with that of ARNTL, CLOCK, CRY1, CRY2, CSNK1D, CSNK1E, NPAS2, and PER3 (**Figure 4C**).

Analysis of TFs and PER1-Related Genes

To further study PER1, we explored the molecules that could regulate it and its related genes. We used the Network Analyst tool to predict TFs that could regulate PER1, ARNTL, CLOCK, CRY1, CRY2, CSNK1D, CSNK1E, NPAS2, and PER3, and constructed a PER1–TF network (**Figure 5A**). We found that four TFs (MAX, SREBF1, USF2, and YY1) could regulate the expression of PER1 and their expression was strongly correlated with that of PER1 (**Figure 5B**).

Correlations Between PER1 Expression and Immune Cell Infiltration in OV

Tumor-infiltrating immune cells are independent predictors of cancer prognosis. Therefore, it was of great significance to analyze the relationships between the expression of PER1 and the infiltration levels of immune cells. As shown in **Figure 6**, using the TIMER database, we found that PER1 expression was significantly negatively correlated with the infiltration of B cells, macrophages, and neutrophils. In order to further explore the effects of PER1 expression on immune cell infiltration in OV, we used the "Gene" module in TIMER2.0 to search the database, inputting the target gene as PER1 and selecting OV. This module displays immune infiltration results obtained by different methods, including TIMER, EPIC, MCP-COUNTER, CIBERSORT,

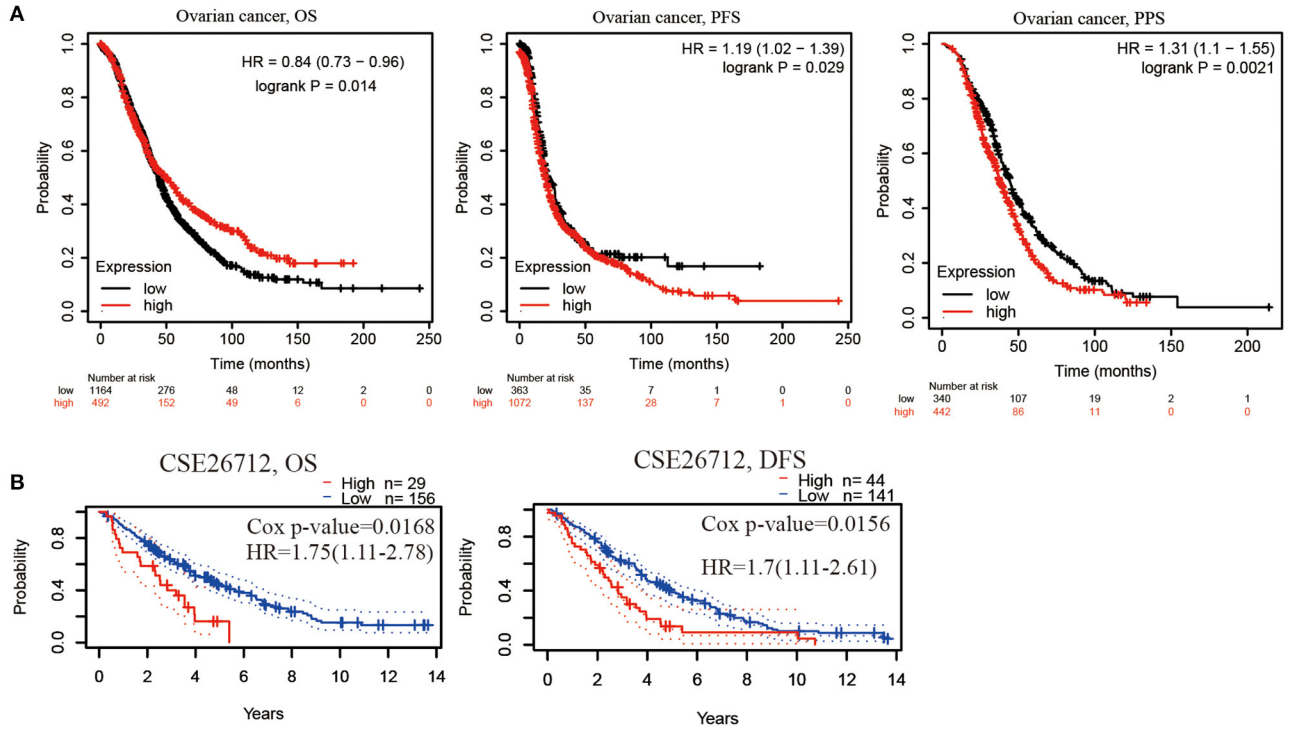


FIGURE 2 | Prognostic value of PER1 mRNA expression in OV patients. **(A)** Low PER1 expression was associated with poor OS in BC patients using Kaplan–Meier Plotter, and high expression of PER1 was associated with poor PFS and PPS. **(B)** Survival curve from PrognoScan analysis for OS and DFS of patients with OV. OS, Overall survival; PFS, Progression-free survival; DFS, disease-free survival; PPS, Post Progression survival.

TABLE 1 | Univariate COX regression analysis for PER1.

Id	HR	HR.95L	HR.95H	*p*-value
Age	1.02146581	1.008809015	1.0342814	0.000841869
Grade	1.236738772	0.850976786	1.79737311	0.265299859
Stage	1.26590041	0.944222869	1.697166952	0.114962158
PER1	1.137153358	1.011959892	1.277834991	0.03079268

HR, Hazard ratio.

TABLE 2 | Multivariate COX regression analysis for PER1.

Id	HR	HR.95L	HR.95H	*p*-value
Age	1.022023712	1.009170894	1.035040223	0.000741433
Grade	1.143213873	0.783174087	1.668770686	0.487968691
Stage	1.262130053	0.932519954	1.708244701	0.131671624
PER1	1.126446708	1.000510744	1.268234441	0.049021419

HR, Hazard ratio.

CIBERSORT-ABS, QUANTISEQ, XCELL, QUANTISE, and TIDE.

The results are shown in **Table 3**, where "+" indicates that PER1 is positively correlated with specific immune cells in OV, and "-" indicates that PER1 is negatively correlated with specific immune cells in OV. The result is the correlation index Rho (italics and black indicate a statistically significant positive correlation, black and underlined indicates a statistically significant negative correlation). PER1 expression in OV was positively correlated with neutrophils, regulatory T cells (Tregs), M2 macrophages, cancer-associated fibroblasts, naïve B cells, activated dendritic cells, and resting CD4+ memory T cells, and negatively correlated with B cells and M1 macrophages. The different methods used for CD4+ cells and macrophages led to inconsistent conclusions; this needs to be further explored.

Correlation Analysis Between PER1 and Immune Cell Markers

In order to further explore the relationship between PER1 and the infiltrating immune cell subsets in OV, we used the TIMER database again to explore the correlations between PER1 and different immune cell markers. As shown in **Table 4**, the expression of PER1 was significantly correlated with various immune cell markers. In addition, we studied the correlation between PER1 and different T cell subgroups of maker (**Table 5**). The subgroups included Th1, Th1-like, and Th2 cells; Tregs, resting Tregs, and effector Tregs; and effector, naive, effector memory, resistant memory, and exhausted T cells. These results indicate that PER1 is closely related to immune cell infiltration in OV and suggest that it has an important role in the tumor microenvironment of OV.

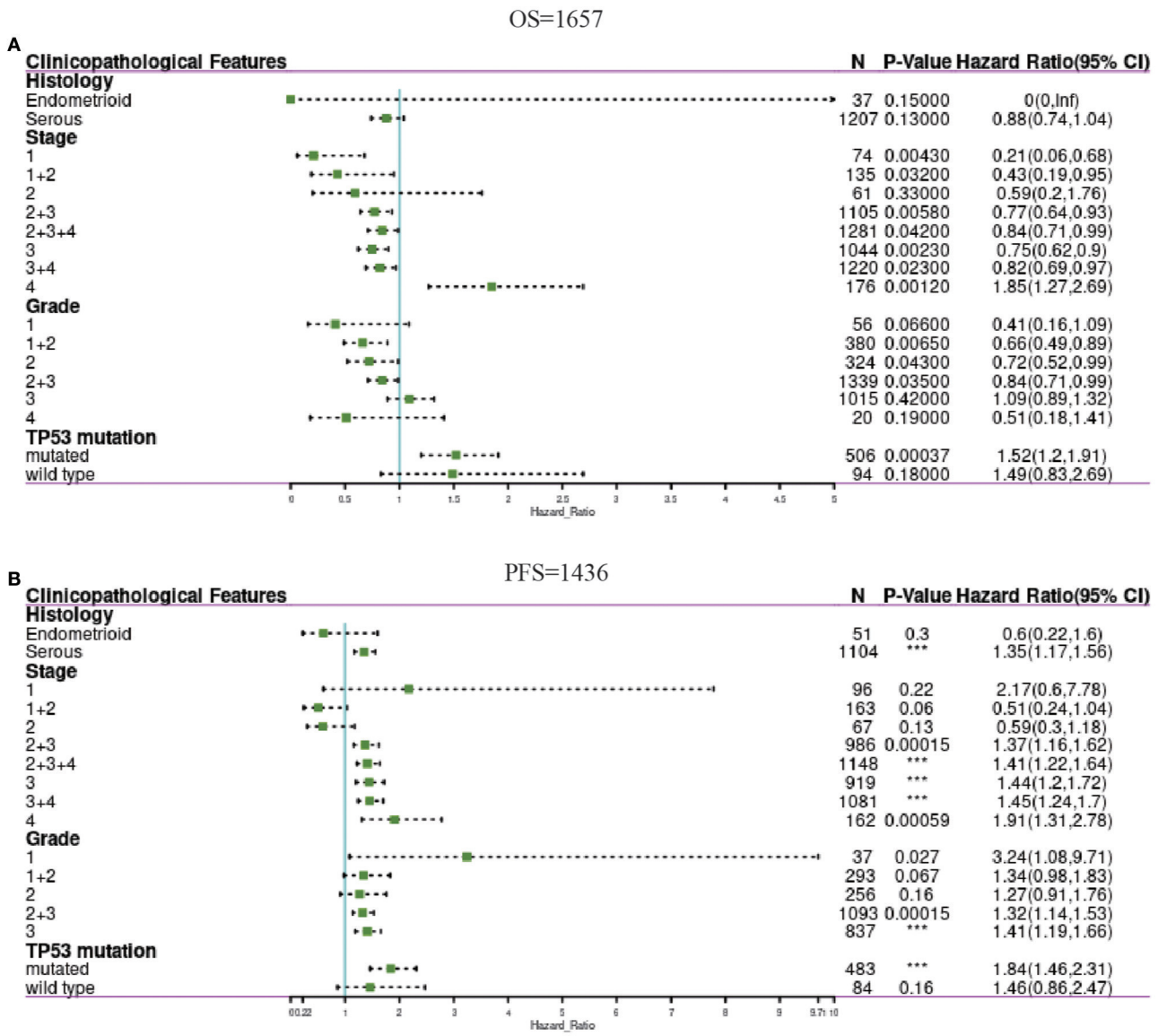


FIGURE 3 | Forest plots showing associations between PER1 expression and clinicopathological features in OV patients. The above table show correlation of PER1 expression and OS in OV with different clinicopathological factors **(A)**. The below table show correlation of PER1 expression and FPS in OV with different clinicopathological factors **(B)**. ***$P < 0.001$.

PER1 Affects the Prognosis of Patients via Effects on Immune Cells

Based on the close correlations between PER1 and immune cell infiltration, we investigated whether PER1 could affect the prognosis of patients via effects on immune cells. We stratified patients into lymphocyte-based subsets and examined the associations between PER1 and OV prognosis using Kaplan–Meier Plotter. As shown in **Figure 7**, decreased CD8+ memory T cells, decreased Treg cells, enriched B cells, enriched Treg cells, enriched macrophages, enriched NK T cells, enriched Th1 cells, decreased Th1 cells, decreased Th1 cells, decreased Th1 cells, decreased Th1 cells, and enriched Th2 cells, low expression of PER1 has a better prognosis. High expression of PER1 was significantly associated with poor prognosis in patients in the HTH2 cell subgroup but not in those in the CD4+ T cell subgroup (**Figure 7A**).

Validation of PER1 in Independent OV Cohorts

We used qRT-PCR to characterize the expression of PER1 in OV tissues and paracancerous tissues. PER1 expression in OV tissues was observably lower than that in paracancerous tissues (**Figure 8A**). Then, we explored the effects of PER1 on prognosis

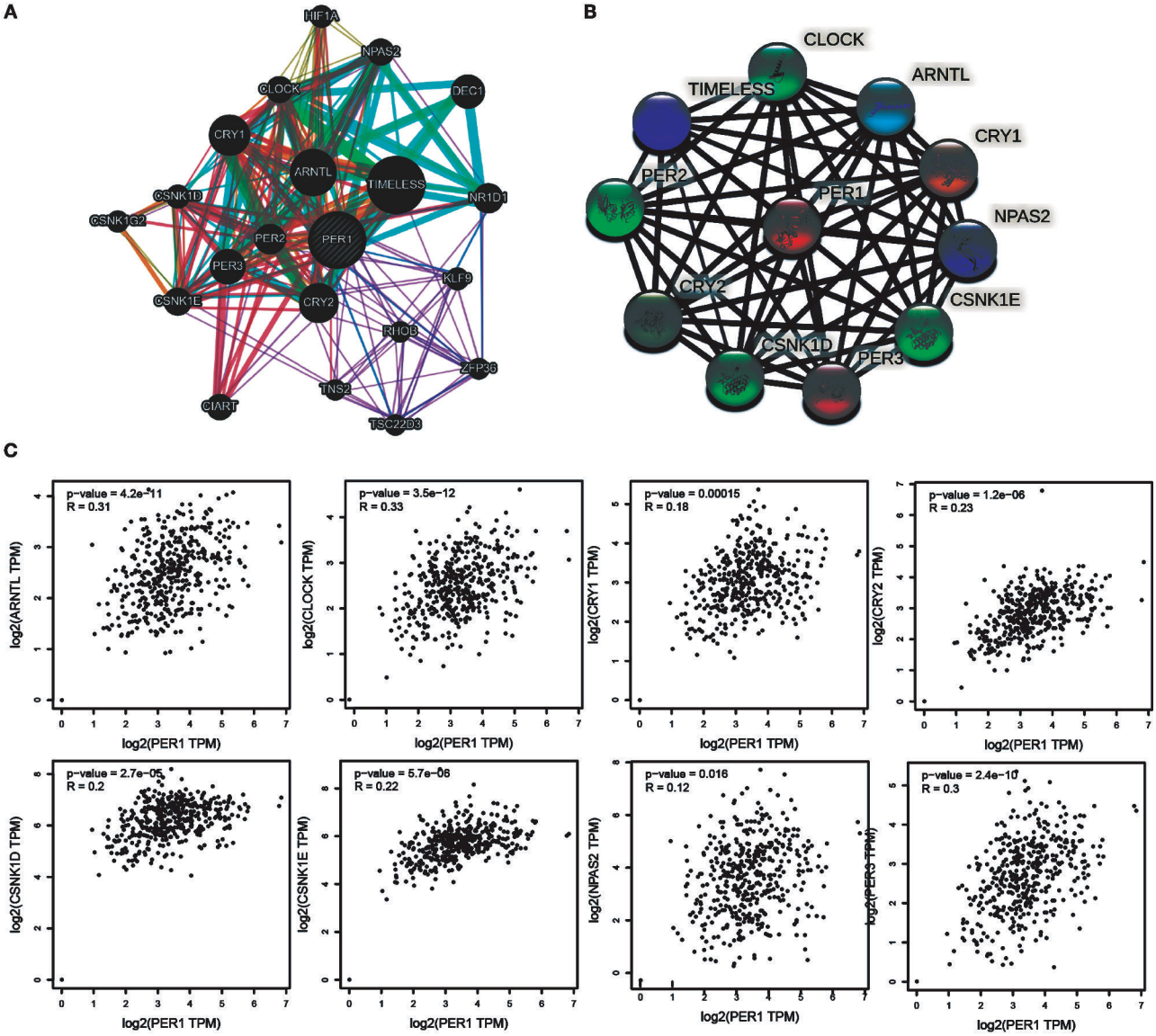


FIGURE 4 | Interaction analysis for PER1 at the gene and protein levels. **(A)** Gene–gene interaction network for PER1 constructed using the GeneMANIA. The 24 most frequently changed neighboring genes are shown. Each node represents a gene. The node color represents the possible functions of the respective gene. **(B)** PPI network containing 10 nodes, constructed using the STRING database. **(C)** Scatter plots of correlations between PER1 expression and ARNTL, CLOCK, CRY1, CRY2, CSNK1D, CSNK1E, NPAS2, and PER3 in OV.

of OV patients and found that lower expression of PER1 was significantly associated with poor survival (**Figure 8B**).

DISCUSSION

One of the characteristics of malignant tumors is uncontrolled proliferation. As an important circadian rhythm gene, PER1 controls the metabolism of normal cells in the body. When it is abnormal, the resulting changes to the normal rhythm of cell proliferation lead to the occurrence of tumors. PER1 has a role in feeding and active behavior in mammals and is controlled by circadian rhythms (Kim et al., 2020). From a non-tumor perspective, PER1 is related to conditions such as Parkinson's disease, obesity, sleep, drug resistance, and premature ovarian insufficiency (Zheng et al., 2019; Delgado-Lara et al., 2020; EmeklI et al., 2020; Mabrouk et al., 2020; Arellanes-Licea et al., 2021). PER1 enhances the activity of GPX through the interaction of PER1 and GPX1 in the cytoplasm, thereby improving the oxidative phosphorylation efficiency of mitochondria (Sun et al., 2020). Current studies indicate that the clock gene PER1 is downregulated in a variety of tumors and plays an important part in promoting tumor progression. However,

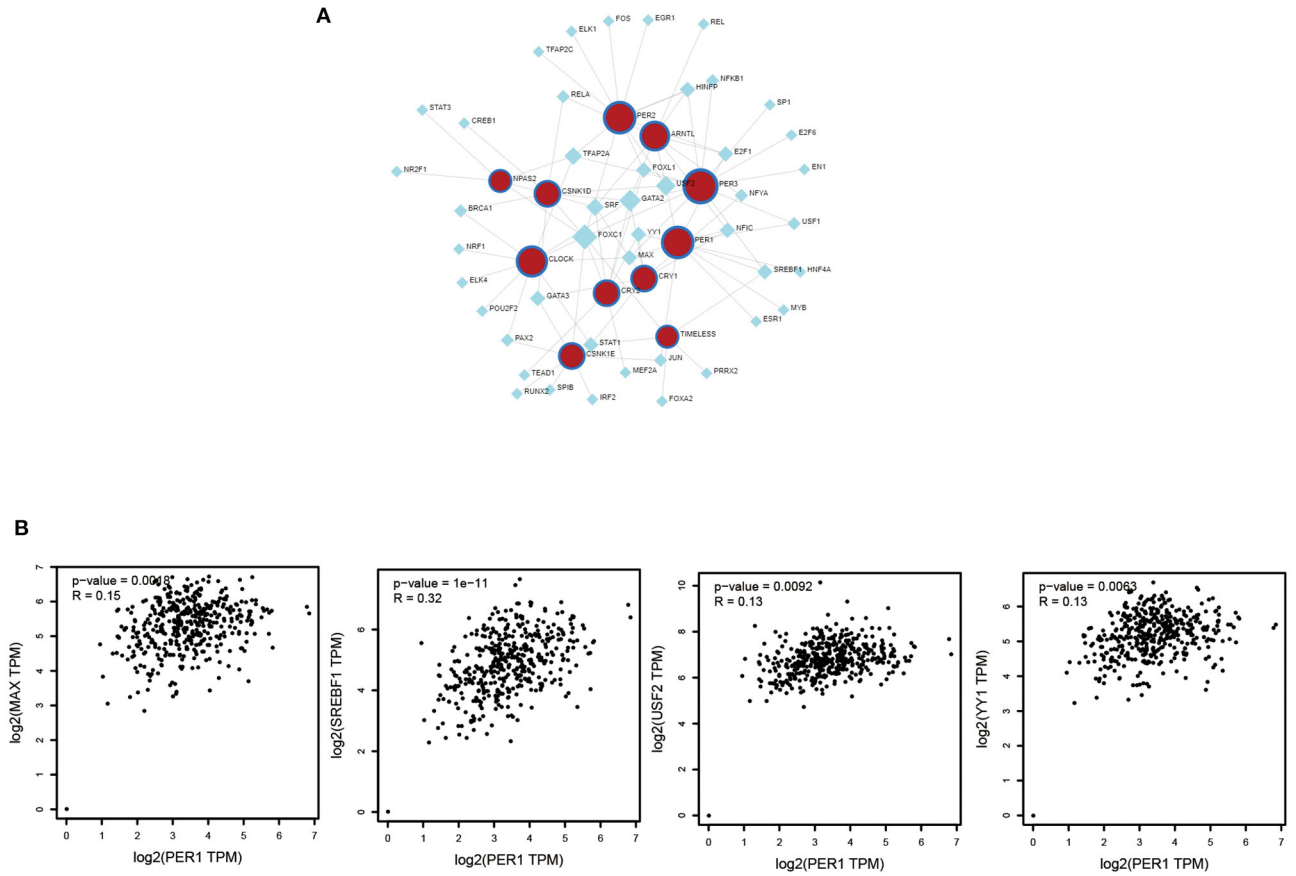


FIGURE 5 | Associations of TFs and PER1-interacting genes (ARNTL, CLOCK, CRY1, CRY2, CSNK1D, CSNK1E, NPAS2, and PER3). **(A)** Network of TFs and selected PER1-interacting genes. **(B)** Correlation between TFs and PER1 expression.

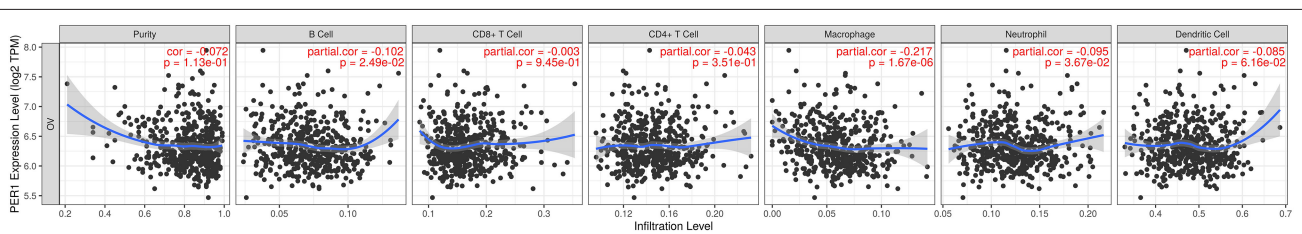


FIGURE 6 | Correlation of PER1 expression with immune infiltration levels in OV. PER1 expression was significantly negatively related to infiltrating levels of macrophages, neutrophils, and B cells.

the biological function and mechanism of PER1 in tumors remain unknown.

Recently, studies have shown that PER1 is silenced or inhibited in a variety of cancers, and that activation or upregulation of PER1 can effectively inhibit cancer cell growth and increase spontaneous cell death. ALKBH5–PER1 is associated with the development of pancreatic cancer. Mechanistically, ALKBH5 activates PER1 in a M6a-YTHDF2-dependent manner after transcription through M6a demethylation (Liu et al., 2020). Upregulation of PER1 leads to the reactivation of ATM-CHK2-p53/CDC25c signaling, thereby inhibiting the growth of cells and the occurrence, development, and invasion of tumors. PER1 promotes the progression of oral squamous cell carcinoma by inhibiting autophagy-mediated apoptosis and enhancing cell proliferation in an Akt/mTOR-pathway-dependent manner. Low expression of PER1 is associated with poor prognosis in patients, and PER1 may be an important therapeutic target for oral squamous

TABLE 3 | PER1 expression and infiltration levels of immune cells.

TIMER	EPIC	MCP-COUNTER	CIBERSORT	CIBERSORT-ABS	QUANTISEQ	XCELL	QUANTISE	TIDE
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
Cancer associated fibroblast	0.185(b)	0.074(g)	0.189(i)	0.18(c)				
Neutrophil	0.236(a)	0.405(c)	0.107(h)	0.04(g)	−0.054(d)	−0.031(e)		
Treg	0.25(f)	0.196(g)	−0.018(d)	0.059(e)				
Macrophage M2	0.148(h)	−0.029(g)	−0.093	−0.076(d)	0.091(e)			
B cell naïve	0.168(e)	0.044(g)	0.119(d)					
Dentritic cell activated	0.129(e)	−0.015(g)	0.118(d)					
T Cell CD4+ memory resting	0.208(e)	0.098(d)						
Macrophage	0.219(a)	−0.016(b)	–0.139(g)					
T Cell CD4+	0.298(b)	0.029(a)	0.065(d)	–0.143(h)				
Monocyte	0.128(c)	0.012(g)	0.014(d)	0.075(e)	–0.16(f)			
Macrophage M1	0.241(h)	−0.049(e)	–0.145(g)	–0.147(d)				
B Cell	0.188(h)	0.094(b)	0.002(g)	0.105(c)	−0.11(d)	–0.157(a)		
T Cell CD8+	0.116(a)	−0.033(b)	−0.065(c)	−0.108(d)	0.003(e)	−0.071(f)	−0.047(g)	
Dentritic cell	−0.011(a)	−0.031(g)	0.082(c)	0.032(f)				
T Cell CD4+ memory activated	−0.041(d)	−0.04(e)						
T Cell folicular helper	−0.002(d)	0.097(e)						

PER1 expression in OV was positively correlated with infiltration of neutrophils and Tregs, and negatively correlated with that of B cells and M1 macrophages.

cell carcinoma (Yang et al., 2020). PER1 can promote cell apoptosis and expression of TNF-α, IL-6, and programmed death 1 (PD-1)/PD-L1, and inhibit tumor invasion and TUBB2B gene expression.

The severity of endometrial cancer is related to night-shift work and dysrhythmia. Rhythm-related factors PER1, TUBB2B, and tumor immune factors can regulate the pathogenesis and progression of endometrial cancer (Wang et al., 2020). As a circadian rhythm gene, PER1 has an important role in the cell cycle and affects the occurrence of cancer, including colorectal cancer (Holipah and Kuroda, 2020). In non-small-cell lung cancer, mangiferin inhibits lipopolysaccharide-induced epithelial–mesenchymal transition and enhances the expression of PER1 (Lin et al., 2020).

In this study, we investigated the expression of PER1 in OV tissues and adjacent normal tissues based on the TIMER and Oncomine databases and found that its expression in cancer tissues was lower than that in normal tissues. We used Kaplan–Meier analysis to determine the relationship between PER1 mRNA expression and the clinical characteristics of OV patients. In our models for OS and PFS, we found that PER1 was related to the stage, grade, and type of OV, and the presence of T53 mutations. Then, we evaluated the expression of PER1 and the survival of OV patients using Kaplan–Meier Plotter and PrognoScan. The results showed that PER1 was correlated with prognosis. According to the PrognoScan data, low expression of PER1 was associated with better patient prognosis in both the OS and DFS models. According to the Kaplan–Meier analysis, however, low expression of PER1 was associated with better prognosis of patients in the PPS and PFS models, whereas in the OS model high expression was associated with good prognosis. This apparent discrepancy may be related to the limited sample size; further research is needed. Overall, these results indicate that PER1 may be a valuable biomarker and an independent predictor of BC.

The development of immunotherapy strategies to eliminate cancer cells by enhancing natural defenses represents a milestone in cancer treatment. Blockade of PD-1/PD-L1 and CTLA-4 has achieved significant therapeutic success in a variety of cancers. Studies of immune infiltration have shown that the tumor immune microenvironment has a key role in cancer progression and affects the clinical outcomes of cancer patients (Murciano-Goroff et al., 2020; Zhang and Zhang, 2020). As immune cells are the cellular basis of immunotherapy, an in-depth understanding of immune infiltration in the tumor immune microenvironment can reveal underlying molecular mechanisms and provide new strategies to improve the efficacy of immunotherapy (Wu et al., 2021). A key result of the present work was that PER1 expression was closely related to the infiltration levels of a variety of immune cells. Many functional parameters in the immune system are related to cycle, including lymphocyte proliferation, natural killer cell activity, and cytokine levels (Young et al., 1995; Arjona and Sarkar, 2005; Keller et al., 2009). Inconsistent results were obtained for macrophages: PER1 expression showed

TABLE 4 | Correlation analysis between PER1 and immune cell markers.

Description	Gene markers	OV			
		None		Purity	
		Cor	*P*	Cor	*P*
CD8+ T cell	CD8A	0.082	0.155	−0.006	0.924
	CD8B	0.011	0.843	−0.065	0.307
T cell (general)	CD3D	0.015	0.795	−0.095	0.136
	CD3E	0.1	0.0816	0.003	0.966
	CD2	0.067	0.244	−0.029	0.644
B cell	CD19	0.102	0.0777	0.139	*
	CD79A	0.094	0.103	0.036	0.57
Monocyte	CD86	0.13	*	0.044	0.491
	CSF1R	0.214	***	0.134	*
TAM	CCL2	0.054	0.35	−0.036	0.567
	CD68	0.159	**	0.068	0.285
	IL10	0.134	*	0.105	0.0975
M1	IRF5	0.128	*	0.075	0.239
	NOS2	0.18	**	0.119	0.0612
	PTGS2	0.21	***	0.165	**
M2	CD163	0.248	***	0.199	**
	VSIG4	0.128	*	0.054	0.4
	MS4A4A	0.147	*	0.07	0.269
Neutrophils	CEACAM8	0.252	***	0.296	***
	ITGAM	0.256	***	0.189	**
	CCR7	0.139	*	0.085	0.182
Natrual	KIR2DL1	0.103	0.0737	0.051	0.42
Killer cell	KIR2DL3	0.2	***	0.196	**
	KIR2DL4	0.021	0.718	−0.009	0.889
	KIR3DL1	0.09	0.117	−0.001	0.983
	KIR3DL2	0.073	0.203	0.003	0.965
	KIR3DL3	0.116	*	0.126	*

*TAM, tumor-associated macrophage; Cor, R value of Spearman's correlation; None, correlation without adjustment; Purity, correlation adjusted by purity. *P < 0.05; **P < 0.01; ***P < 0.001.*

TABLE 5 | Correlations between PER1 and different T cell subgroups of maker.

Description	Gene markers	OV			
		None		Purity	
		Cor	*P*	Cor	*P*
Th1	STAT4	0.173	**	0.137	*
	STAT1	0.096	0.0951	0.105	0.0978
	IFNG	−0.007	0.91	−0.061	0.339
	TNF	0.09	0.117	0.102	0.107
Th1-like	CXCL13	0.004	0.943	−0.044	0.493
	HAVCR2	0.124	*	0.038	0.548
	IFNG	−0.007	0.91	−0.061	0.339
	BHLHE40	0.252	***	0.227	***
	CD4	0.167	**	0.084	0.187
	IL21	−0.026	0.647	0.004	0.952
Th17	STAT3	0.275	***	0.28	***
	IL17A	0.095	0.1	0.085	0.18
Treg	FOXP3	0.152	**	0.126	*
	CCR8	0.094	0.103	0.052	0.414
	STAT5B	0.357	***	0.328	***
	TGFB1	0.331	***	0.255	***
Effector T-cell	FGFBP2	0.095	0.0978	0.067	0.291
	FCGR3A	0.119	*	0.032	0.619
Naïve T-cell	SELL	0.072	0.211	0.009	0.886
	TCF7	0.042	0.471	0.084	0.189
Effective memory T-cell	PDCD1	0.099	0.0869	0.06	0.343
	DUSP4	0.115	*	0.079	0.215
	GZMK	0.091	0.112	−0.021	0.746
	GZMA	0.042	0.47	−0.049	0.442
	IFNG	−0.007	0.91	−0.061	0.339
Resistant memory T-cell	CD69	0.195	***	0.183	**
	ITGAE	0.014	0.813	−0.027	0.67
	CXCR6	0.108	0.061	0.029	0.65
	MYADM	0.316	***	0.237	***
Exhausted T-cell	HAVCR2	0.124	*	0.038	0.548
	TIGIT	0.103	0.0736	0.041	0.515
	CXCL13	0.004	0.943	−0.044	0.493
	LAYN	0.067	0.246	0.077	0.229
Resting Treg T-cell	FOXP3	0.152	**	0.126	*
	IL2RA	0.197	***	0.172	**
Effective Treg T-cell	FOXP3	0.152	**	0.126	*
	CTLA4	0.135	*	0.072	0.26
	CCR8	0.094	0.103	0.052	0.414
	TNFRSF9	0.099	0.0851	0.065	0.309
General memory T-cell	SELL	0.072	0.211	0.009	0.886
	IL7R	0.229	***	0.182	**

*Th, T helper cell; Cor, R value of Spearman's correlation; None, correlation without adjustment; Purity, correlation adjusted by purity. *P < 0.05; **P < 0.01; ***P < 0.001.*

different correlations with gene markers of M1 macrophages (e.g., PTGS2 and IRF5) and those of M2 macrophages (e.g., CD163, VSIG4, and MS4A4A), which may explain the potential ability of PER1 to regulate tumor-associated macrophages. In OV, there were also significant correlations between PER1 expression and the regulation of several T helper cell markers (Th1, Th2, Tfh, and Th17). These correlations indicate that PER1 could potentially regulate T cell growth in OV. Together, these findings indicate that PER1 plays an important part in the recruitment and regulation of infiltrating immune cells in OV.

This study investigated the links between PER1 and OV; however, there were some limitations. First, although we verified the expression of PER1 based on data from multiple public databases, further experiments are needed to clarify the molecular mechanism and mode of action by which PER1 regulates tumor-infiltrating cells and thereby affects the prognosis of OV patients. Second, although PER1 expression showed a certain correlation with prognosis, different results

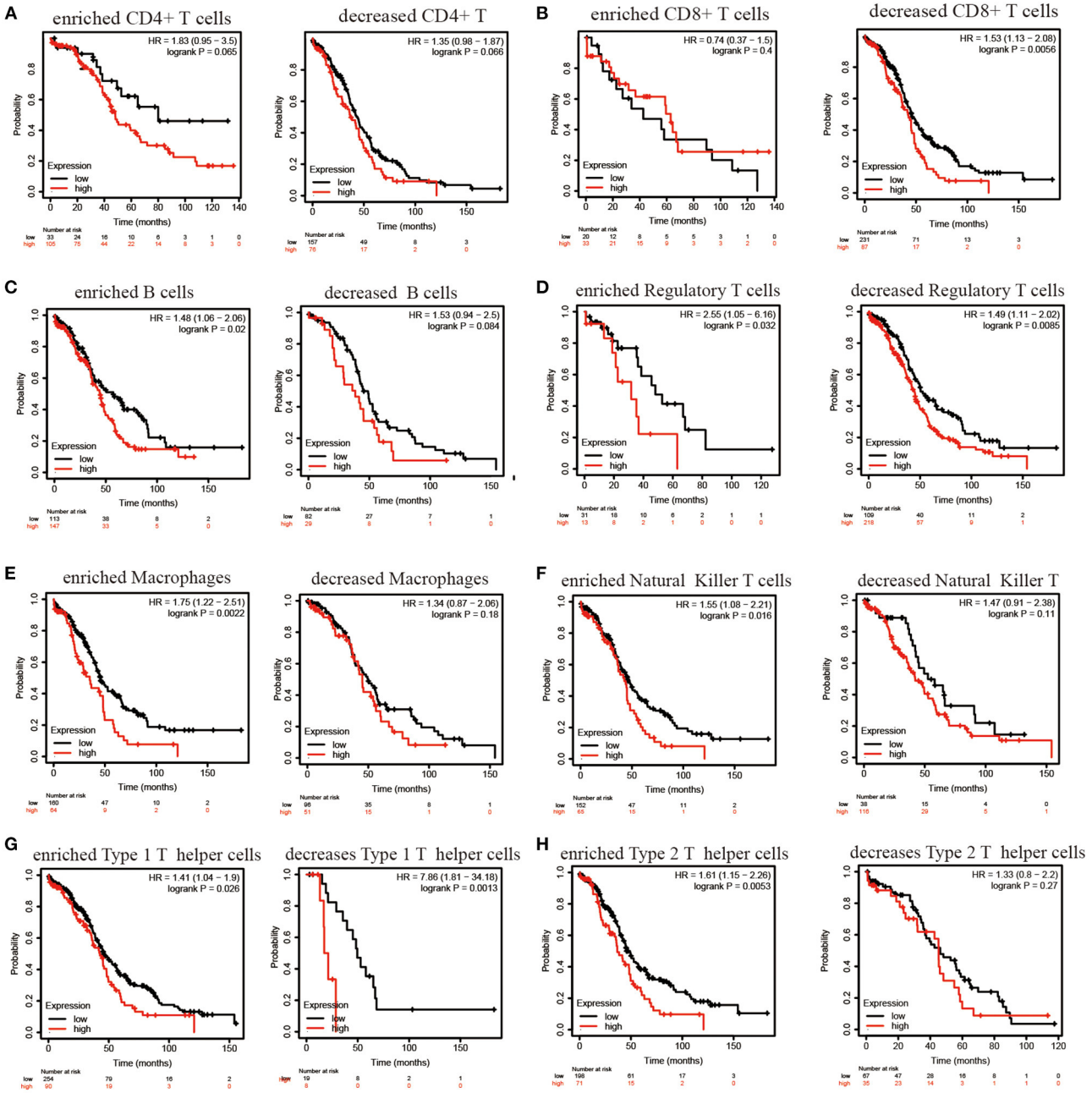


FIGURE 7 | Kaplan–Meier survival curves based on expression of PER1 related to immune cell subtypes in OV. **(A–H)** Correlations between PER1 and OS in OV patients in different immune cell subgroups as estimated by the CIBERSORT algorithm.

were found with different databases, possibly owing to differences in inclusion and exclusion criteria, sample size, human factors and random factors among studies; our research may be affected by the same problems and so further investigation is needed. In future experimental work, it will be necessary to address these problems. In conclusion, PER1 expression is downregulated in OV and significantly related to the pathological stage and prognosis of OV patients. PER1 and related genes have functions in the immune response. In addition, PER1 expression is related to levels of immune cell infiltration; therefore, PER1 may have a role in immunotherapy and is a potential prognostic indicator in OV.

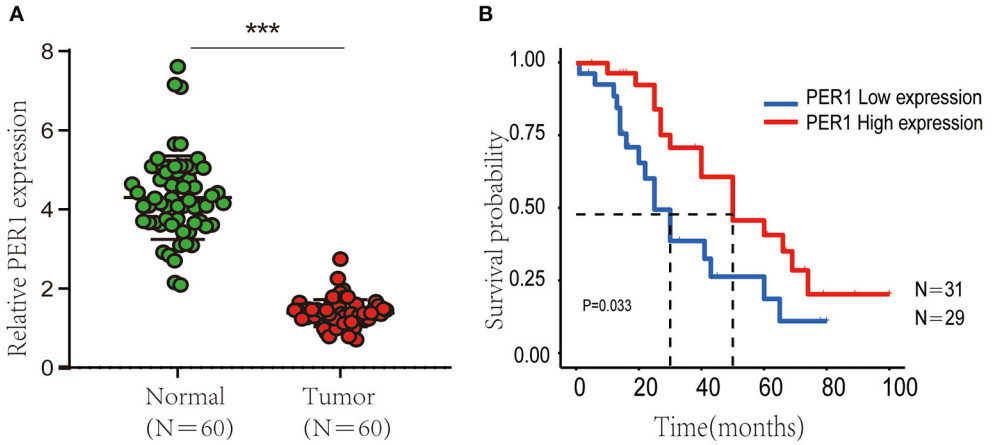


FIGURE 8 | Expression of PER1 in 60 normal and tumor samples. **(A)** Kaplan–Meier OS analysis of Chinese OV patients based on expression of PER1 **(B)**. ***$p < 0.001$.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Gansu Provincial Maternal and Child Health Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MC and LZ conceived of the study and participated in design and coordination. XL drafted the manuscript. ZM collected and analyzed immune related information. LL revised manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by Suggestions on the management of novel coronavirus infection in pregnant women in Gansu Province (2020-XG-10).

REFERENCES

Arellanes-Licea, E., Pérez-Mendoza, M., Carmona-Castro, A., Díaz-Muñoz, M., and Miranda-Anaya, M. (2021). Obese *Neotomodon alstoni* mice exhibit sexual dimorphism in the daily profile of circulating melatonin and clock proteins PER1 and BMAL1 in the hypothalamus and peripheral oscillators. *Chronobiol. Int.* 38, 584–597. doi: 10.1080/07420528.2020.1860999

Arjona, A., and Sarkar, D. K. (2005). Circadian oscillations of clock genes, cytolytic factors, and cytokines in rat NK cells. *J. Immunol.* 174, 7618–7624. doi: 10.4049/jimmunol.174.12.7618

Badiu, C. (2003). Genetic clock of biologic rhythms. *J. Cell. Mol. Med.* 7, 408–416. doi: 10.1111/j.1582-4934.2003.tb00243.x

Bowtell, D. D., Böhm, S., Ahmed, A. A., Aspuria, P. J., Bast, R. C., Beral, V., et al. (2015). Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat. Rev. Cancer* 15, 668–679. doi: 10.1038/nrc4019

Cao, Q., Gery, S., Dashti, A., Yin, D., Zhou, Y., Gu, J., et al. (2009). A role for the clock gene per1 in prostate cancer. *Cancer Res.* 69, 7619–7625. doi: 10.1158/0008-5472.CAN-08-4199

Delgado-Lara, D. L., González-Enríquez, G. V., Torres-Mendoza, B. M., González-Usigli, H., Cárdenas-Bedoya, J., Macías-Islas, M. A., et al. (2020). Effect of melatonin administration on the PER1 and BMAL1 clock genes in patients with Parkinson's disease. *Biomed. Pharmacother.* 129:110485. doi: 10.1016/j.biopha.2020.110485

Dong, X., Men, X., Zhang, W., and Lei, P. (2014). Advances in tumor markers of ovarian cancer for early diagnosis. *Indian J. Cancer* 51(Suppl. 3), e72–e76. doi: 10.4103/0019-509X.154049

EmeklI, R., IsmaIlogullari, S., Bayram, A., Akalin, H., Tuncel, G., and Dündar, M. (2020). Comparing expression levels of PERIOD genes PER1, PER2 and PER3 in chronic insomnia patients and medical staff working in the night shift. *Sleep Med.* 73, 101–105. doi: 10.1016/j.sleep.2020.04.027

Guo, X., Li, K., Jiang, W., Hu, Y., Xiao, W., Huang, Y., et al. (2020). RNA demethylase ALKBH5 prevents pancreatic cancer progression by posttranscriptional activation of PER1 in an m6A-YTHDF2-dependent manner. *Mol. Cancer* 19:91. doi: 10.1186/s12943-020-01158-w

Holipah, H. T., and Kuroda, Y. (2020). PER1 rs3027188 polymorphism and its association with the risk of colorectal cancer in the Japanese population. *Gan To Kagaku Ryoho* 47, 941–946.

Keller, M., Mazuch, J., Abraham, U., Eom, G. D., Herzog, E. D., Volk, H. D., et al. (2009). A circadian clock in macrophages controls inflammatory immune responses. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21407–21412. doi: 10.1073/pnas.0906361106

Kim, T. J., Sung, J. H., Shin, J. C., and Kim, D. Y. (2020). CRISPR/Cas-mediated Fubp1 silencing disrupts circadian oscillation of Per1 protein via downregulating Syncrip expression. *Cell Biol. Int.* 44, 424–432. doi: 10.1002/cbin.11242

Lange, T., Dimitrov, S., and Born, J. (2010). Effects of sleep and circadian rhythm on the human immune system. *Ann. N. Y. Acad. Sci.* 1193, 48–59. doi: 10.1111/j.1749-6632.2009.05300.x

Lin, Y. S., Tsai, K. L., Chen, J. N., and Wu, C. S. (2020). Mangiferin inhibits lipopolysaccharide-induced epithelial-mesenchymal transition (EMT) and enhances the expression of tumor suppressor gene PER1 in non-small cell lung cancer cells. *Environ. Toxicol.* 35, 1070–1081. doi: 10.1002/tox.22943

Liu, T., Wei, Q., Jin, J., Luo, Q., Liu, Y., Yang, Y., et al. (2020). The m6A reader YTHDF1 promotes ovarian cancer progression via augmenting EIF3C translation. *Nucleic Acids Res.* 48, 3816–3831. doi: 10.1093/nar/gkaa048

Mabrouk, A., Chebbi, Y., Raddaoui, A., Krir, A., Messadi, A. A., Achour, W., et al. (2020). Clonal spread of PER-1 and OXA-23 producing extensively drug resistant Acinetobacter baumannii during an outbreak in a burn intensive care unit in Tunisia. *Acta Microbiol. Immunol. Hung.* 67, 222–227. doi: 10.1556/030.2020.01208

Murciano-Goroff, Y. R., Warner, A. B., and Wolchok, J. D. (2020). The future of cancer immunotherapy: microenvironment-targeting combinations. *Cell Res.* 30, 507–519. doi: 10.1038/s41422-020-0337-2

Sahar, S., and Sassone-Corsi, P. (2009). Metabolism and cancer: the circadian clock connection. *Nat. Rev. Cancer* 9, 886–896. doi: 10.1038/nrc2747

Sun, Q., Yang, Y., Wang, Z., Yang, X., Gao, Y., Zhao, Y., et al. (2020). PER1 interaction with GPX1 regulates metabolic homeostasis under oxidative stress. *Redox Biol.* 37:101694. doi: 10.1016/j.redox.2020.101694

Wang, Z., Wang, H., Wang, Z., He, S., Jiang, Z., Yan, C., et al. (2020). Associated analysis of PER1/TUBB2B with endometrial cancer development caused by circadian rhythm disorders. *Med. Oncol.* 37:90. doi: 10.1007/s12032-020-01415-4

Wu, P., Gao, W., Su, M., Nice, E. C., Zhang, W., Lin, J., et al. (2021). Adaptive mechanisms of tumor therapy resistance driven by tumor microenvironment. *Front. Cell Dev. Biol.* 9:641469. doi: 10.3389/fcell.2021.641469

Yang, G., Yang, Y., Tang, H., and Yang, K. (2020). Loss of the clock gene Per1 promotes oral squamous cell carcinoma progression via the AKT/mTOR pathway. *Cancer Sci.* 111, 1542–1554. doi: 10.1111/cas.14362

Yang, X., Wood, P. A., Ansell, C. M., Quiton, D. F., Oh, E. Y., Du-Quiton, J., et al. (2009). The circadian clock gene Per1 suppresses cancer cell proliferation and tumor growth at specific times of day. *Chronobiol. Int.* 26, 1323–1339. doi: 10.3109/07420520903431301

Young, M. R., Matthews, J. P., Kanabrocki, E. L., Sothern, R. B., Roitman-Johnson, B., and Scheving, L. E. (1995). Circadian rhythmometry of serum interleukin-2, interleukin-10, tumor necrosis factor-alpha, and granulocyte-macrophage colony-stimulating factor in men. *Chronobiol. Int.* 12, 19–27. doi: 10.3109/07420529509064496

Zhang, Y., and Zhang, Z. (2020). The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell. Mol. Immunol.* 17, 807–821. doi: 10.1038/s41423-020-0488-6

Zheng, Y., Liu, C., Li, Y., Jiang, H., Yang, P., Tang, J., et al. (2019). Loss-of-function mutations with circadian rhythm regulator Per1/Per2 lead to premature ovarian insufficiency†. *Biol. Reprod.* 100, 1066–1072. doi: 10.1093/biolre/ioy245

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fgene.2021.690053

Immune-Related Gene Expression Analysis Revealed Three lncRNAs as Prognostic Factors for Colon Cancer

*Xiao-Liang Xing[1], Ti Zhang[1], Zhi-Yong Yao[1], Chaoqun Xing[1], Chunxiao Wang[1], Yuan-Wu Liu[2] and Minjiang Huang[1]**

[1] School of Public Health and Laboratory Medicine, Hunan University of Medicine, Huaihua, China, [2] Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China

Colorectal cancer (CRC) is one of the most common cancers. Almost 80% of CRC cases are colon adenocarcinomas (COADs). Several studies have indicated the role of immunotherapy in the treatment of various cancers. Our study aimed to identify immune-related long non-coding RNAs (lncRNAs) and to use them to construct a risk assessment model for evaluating COAD prognosis. Using differential expression, correlation, and Cox regression analyses, we identified three immune-related differentially expressed lncRNAs (IR-DELs) and used them to construct a risk assessment model. The area under the curve (AUC) for each receiver operating characteristic (ROC) curve at 3-, 5-, and 10-years were greater than 0.6. In addition, the risk assessment model was correlated with several immune cells and factors. The three IR-DELs (AC124067.4, LINC02604, and MIR4435-2HG) identified in this study can be used to predict outcomes for patients with COAD and might help in identifying those who can benefit from anti-tumor immunotherapy.

Keywords: COAD, immune, lncRNAs, overall survival, risk model

OPEN ACCESS

Edited by:
Xiangqian Guo,
Henan University, China

Reviewed by:
Yinan Jiang,
University of Pittsburgh, United States
Andrey Zamyatnin,
I.M. Sechenov First Moscow State Medical University, Russia

***Correspondence:**
Minjiang Huang
whmj530@163.com

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

Received: *02 April 2021*
Accepted: *21 June 2021*
Published: *09 July 2021*

Citation:
Xing X-L, Zhang T, Yao Z-Y, Xing C, Wang C, Liu Y-W and Huang M (2021) Immune-Related Gene Expression Analysis Revealed Three lncRNAs as Prognostic Factors for Colon Cancer. Front. Genet. 12:690053. doi: 10.3389/fgene.2021.690053

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide, with nearly 1.8 million new cases and 861,000 deaths reported in 2018 (Bray et al., 2018). Depending on its anatomical location, CRC can be classified either as colon adenocarcinoma (COAD) or rectal adenocarcinoma (READ). COAD accounts for approximately 80% of CRC cases, and almost 50% patients with CRC will develop distant metastases (Arnold et al., 2017). The standard methods for CRC treatment are surgery, chemotherapy, and radiotherapy. These treatments can be combined, depending on the location and progression of the cancer (Johdi and Sukor, 2020). For example, approximately 66–61% of stage II and stage III patients with colon and rectal cancer undergo further treatments with adjuvant chemotherapy and/or radiotherapy, respectively (Miller et al., 2019). Approximately 50% of patients relapse even after additional treatment with neoadjuvant therapy (Ogura et al., 2019). Even with combination therapy, the overall survival (OS) rate of patients with advanced COAD remains low.

Cancer immunotherapy targets specific cancer antigens on the malignant cells, alerting the immune system to eradicate cancer through concerted immune responses (Johdi and Sukor, 2020).

Immune cells and factors have antitumor effects, including antitumor initiation and progression (Berraondo et al., 2016; Chen et al., 2017). Cancer immunotherapy has been successfully used in the treatment of many cancers, particularly hematological malignancies and solid tumors (Im and Pavletic, 2017; Nixon et al., 2018). Increasing evidence suggests that immune cells and factors play important roles in the initiation and progression of CRC. Higher proportions of activated $CD8^+$ TILs in the early stages of tumor development suggest that immune system surveillance recognizes CRC (Emambux et al., 2018).

Long non-coding RNAs (lncRNAs) are non-translated RNA transcripts and account for 68% of the human transcriptome (Iyer et al., 2015). They are involved in regulating the expression of genes at the epigenetic and transcriptional levels. LncRNAs participate in cell proliferation, differentiation, and apoptosis and play important roles in tumorigenesis and tumor suppression (Ogunwobi et al., 2020). For example, the lncRNAs HEIH, AK023391, and PAGBC can promote CRC tumorigenesis (Huang et al., 2017; Wu et al., 2017; Cui et al., 2018), while the lncRNAs RPPH1, FEZF1-AS1, and u50535 can promote invasion and metastasis of CRC (Bian et al., 2018; Yu et al., 2018; Liang et al., 2019). In this study, we aimed to investigate the relationship between lncRNA gene expression profiles and the immunity characteristics of patients with COAD, and to construct a risk assessment model which can be used to predict the outcome of COAD and identify patients likely to benefit from cancer immunotherapy.

MATERIALS AND METHODS

Data Acquisition and Analysis of Differentially Expressed Genes

We downloaded RNA sequence data and the corresponding clinical information for 497 individuals (41 healthy controls and 456 patients with COAD) from The Cancer Genome Atlas (TCGA) database. The DESeq2 package in R software (3.6.1) was used to identify differentially expressed genes (DEGs) based on the criteria: adj. $P < 0.05$, $|\text{logFC}| \geq 1.0$, and basemean ≥ 100. A Gene Transfer Format (GTF) file from Ensembl[1] was used to identify lncRNAs. A list of recognized immune-related genes (IR-Genes) from the ImmPort database[2] was used to screen immune-related lncRNAs (IR-lncRNAs) using Spearman correlation analysis. David 6.8 was used to carry out KEGG (KEGG: Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) analysis for those DEGs with the default parameters.

The ESTIMATE package in R software (3.6.1) was used to calculate the immune and stromal scores. The extent of immune cell infiltration was obtained from Tumor Immune Estimation Resource (TIMER) and used to estimate the immune cell infiltrates of patients in the TCGA database[3].

[1]http://asia.ensembl.org

[2]http://www.immport.org

[3]https://cistrome.shinyapps.io/timer/

Correlation Analysis

Spearman correlation analysis was used to identify the relationship between IR-DEGs and DELs based on the criteria: P-value < 0.05 and $|R| > 0.5$, and to identify the relationship between the risk score and immune infiltration based on the criteria: P-value < 0.05 and $|R| > 0$.

Overall Survival (OS) Analysis

To conduct OS analysis of IR-DELs, we first grouped the patients with COAD into low- and high-expression groups based on the expression of IR-DELs. Survival, Survminer, and RegParallel packages in R software were used for univariate and multivariate Cox regression analyses. To compute the risk value, we first calculated the cutoff value according to IBM SPSS 22, and then grouped the patients into low- and high-risk groups based on the cutoff value. Survival, Survminer, and RegParallel packages in R software were used for univariate and multivariate Cox regression analyses.

Risk Assessment Model Construction

We constructed a risk assessment model based on the IR-DELs identified through univariate and multivariate Cox regression analyses. The formula of the risk assessment model was set as follows: risk score = ExpIR-DEL1*βIR-DEL1 + ExpIR-DEL2*βIR-DEL2+.....+ ExpIR-DELn*βIR-DELn. Unpaired two-tailed Student's t-test was used to investigate the relationship between the risk assessment model and the clinical characteristics of COAD. Time-dependent receiver operating characteristic (ROC) curves were used to estimate the utility of this model as a prognostic tool for predicting survival status. All statistical analyses were performed in Prism 8.0.1.244.

RESULTS

Aberrant Immune and Stromal Scores for COAD

We analyzed the data of 497 individuals, 41 controls and 456 patients with COAD, downloaded from the TCGA-COAD database. The ESTIMATE package in R software (3.6.1) was used to calculate the immune and stromal scores. The immune and stromal scores for patients with COAD were significantly lower than those for the controls (**Figures 1A,C**). Immune and stromal scores of COAD showed a significant correlation of both immune and stromal scores with tumor purity predictions (**Figures 1B,D**). We identified the maximum inflection points of the immune and stromal scores as the cut-off points on the ROC curve. The cutoff values for immune and stromal scores were -292.07 (**Figure 1E**) and −1336.78 (**Figure 1F**), respectively. The tumor purity for patients with COAD in different groups was displayed in **Figure 1G**. After categorizing the patients into low- and high-score groups based on the cutoffs, we performed OS analysis of the immune and stromal scores. Patients with COAD with high immune scores displayed significantly better OS (**Figure 1H**). However, there were no significant differences in the OS between patients with low and high stromal scores (**Figure 1I**).

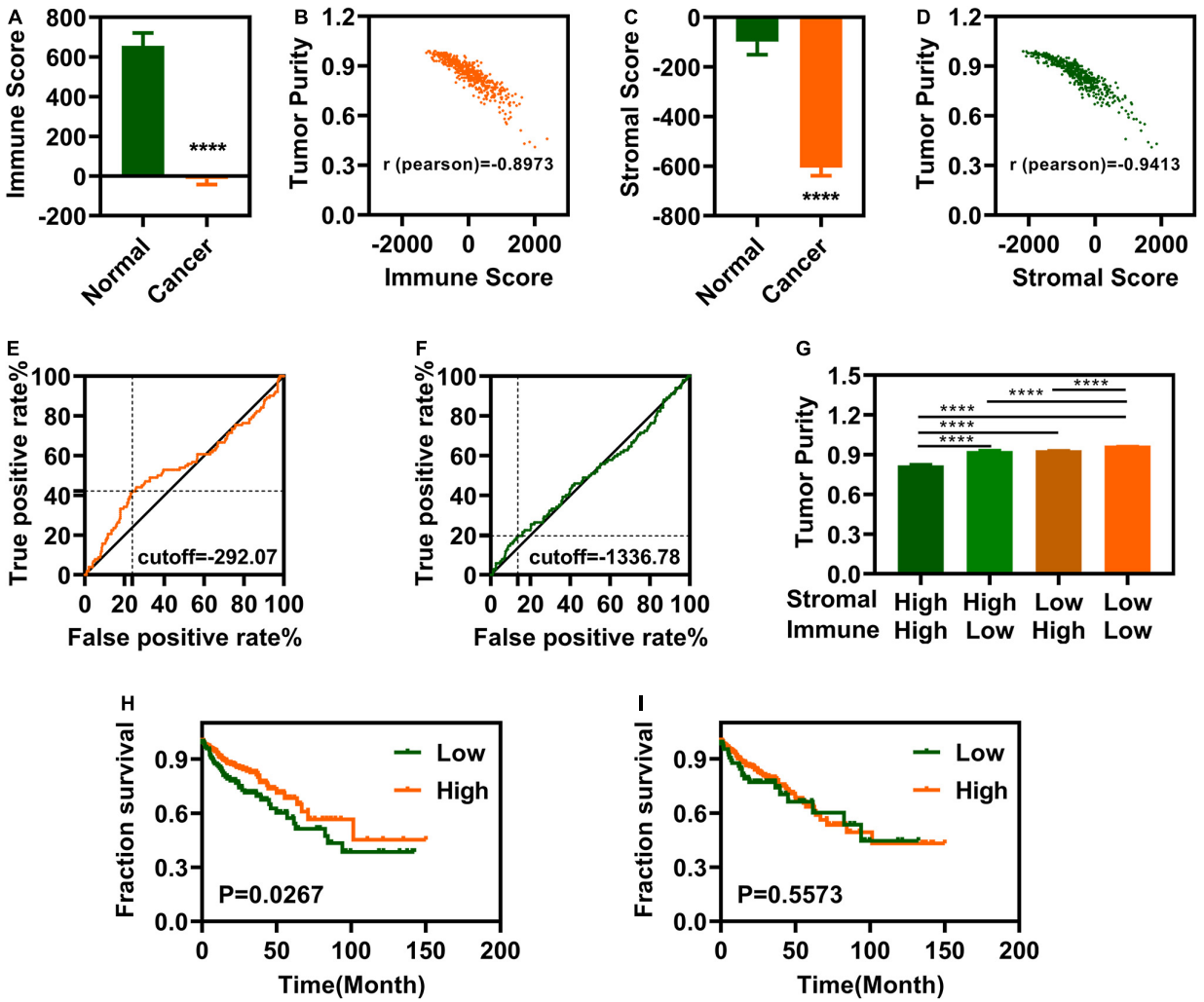


FIGURE 1 | Colon Adenocarcinoma (COAD) is associated with immune and stromal scores. **(A)** Immune scores are significantly associated with COAD. **(B)** Immune scores are significantly correlated with tumor purity. **(C)** Stromal scores are significantly associated with COAD. **(D)** Stromal scores are significantly correlated with tumor purity. **(E,F)** maximum inflection point is the cut-off point for immune scores **(E)** and stromal **(F)**. **(G)** The tumor purity is difference significantly in different groups. **(H)** Patients with COAD with high immune scores group have a longer survival time. **(I)** There is no significantly difference between low stromal scores and high stromal scores groups. $^{****}P < 0.0001$.

Identification of Immune-Related Differentially Expressed LncRNAs (IR-DELs)

Differential expression analysis was carried out using R software's DEseq2 package. We screened 2977 genes and identified 1502 upregulated and 1475 downregulated genes (**Figure 2A**). Cross analysis with the recognized IR-genes identified 343 IR-DEGs (118 upregulated and 225 downregulated) (**Figure 2B**) while cross analysis with GTF annotation data showed 130 DELs (104 upregulated and 26 downregulated) (**Figure 2C**). To determine the relationship between the 343 IR-DEGs and the 130 DELs, we performed Spearman correlation analysis and obtained 483 pairs of IR-DEGs-DELs, which included 187 IR-DEGs and 53 DELs (**Supplementary Table 1**). The expression profiles of these 53 DELs was shown in **Figure 2D**.

Subsequently, we performed univariate Cox regression analysis for the 53 DELs and found that three (AC124067.4, LINC02604, and MIR4435-2HG) were correlated with OS in patients with COAD (**Figure 2E**). Furthermore, multivariate Cox regression analysis also showed that these three DELs were correlated with OS (**Figure 2F**). The relationships between AC124067.4, LINC02604, and MIR4435-2HG expression and OS are shown in **Figures 2G–I**.

Establishment of the Risk Assessment Model

We constructed a risk assessment model using the three DELs (AC124067.4, LINC02604, and MIR4435-2HG). The expression profiles of these three IR-DELs in the normal group and cancer group were shown in **Figure 3A**. The risk score and

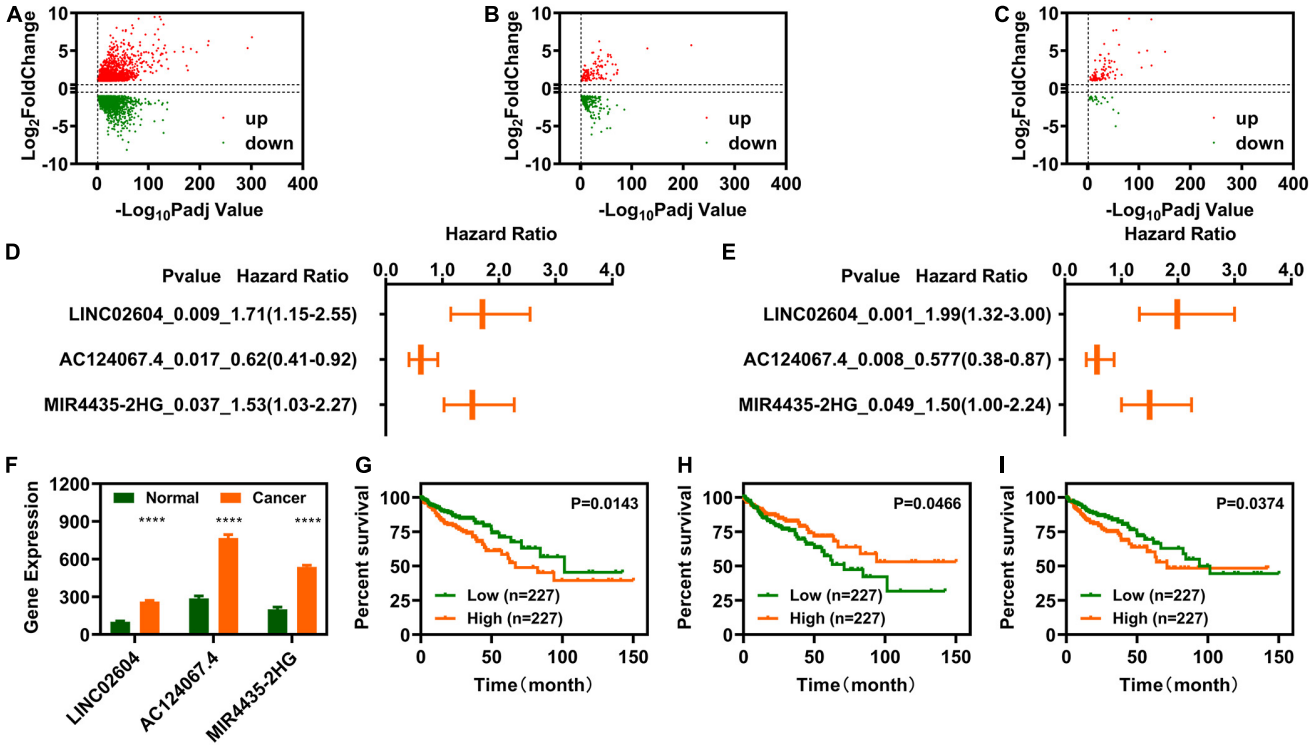


FIGURE 2 | Identification of immune-related differentially expressed long non-coding RNAs (IR-DELs). **(A)** Volcano plot of DEGs for COAD. **(B)** Volcano plot of immune-related DEGs in COAD. **(C)** Volcano plot of immune-related DELs for COAD. **(D)** Heat map of 53 IR-DEGs verified using correlation analysis. **(E)** IR-DELs identified using univariate Cox regression analysis. **(F)** IR-DELs identified through multivariate Cox regression analysis. **(G–I)** Patients with high AC124067.4 expression **(G)**, low LINC02604 expression **(H)**, and low MIR4435-2HG expression **(I)** have a longer survival time. ****$P < 0.0001$.

survival status of each case are displayed in **Figures 3B,C**. Next, we calculated the area under the curve (AUC) for each receiver operating characteristic (ROC) curve of the three DELs, generated the curved line, and found that the highest point corresponded to 0.6128 (**Figure 3D**). To validate the optimality, we further plotted the 3-, 5-, and 10-year ROC curves. All AUC values identified were greater than 0.60 (**Figure 3E**). The maximum inflection point occurred at a cut-off point of 4.115 (**Figure 3F**). Using this cutoff point, we re-distinguished the high- and low-risk groups, and then carried out OS analysis. The results indicate that patients with COAD with low-risk scores displayed better OS (**Figure 3G**). Then, we carried out differential expression analysis for patients with COAD between high-risk group and low-risk group, and found 49 DEGs (34 upregulated and 15 downregulated) (**Supplementary Figure 1**). We performed KEGG and GO analysis for those 49 DEGs. The results were displayed in **Supplementary Table 3**.

Clinical Evaluation Using the Risk Assessment Model

To investigate the relationship between risk scores and clinical characteristics, we first computed correlations between risk scores and pathologic TNM (tumor, node and metastasis), pathologic stage, and vital status (**Figures 4A–H**). Pathologic TNM was previously correlated with OS (Chansky et al., 2009; Wu et al., 2015). In this study, we verified the relationship between pathologic TNM, the pathologic stage, and OS and showed that pathologic TNM was closely associated with OS (**Figures 4I–K**). We plotted the ROC curves of pathologic TNM, pathologic stage, and vital status, which showed that the AUC values of pathologic TNM and pathologic stage were lower than those of vital status (**Figure 4L**). These results indicate that the risk model was more accurate at predicting survival status than predicting clinical characteristics. Then, we also analyzed the expression of those the IR-DELs in difference pathologic TNM. The results were displayed in **Supplementary Figure 2**.

Estimation of Tumor Immune Infiltration Using the Risk Assessment Model

This study aimed to screen out IR-DELs that could be used to predict COAD outcome by investigating whether the risk score was associated with the three IR-DELs (AC124067.4, LINC02604, and MIR4435-2HG). We found that the expression of one IR-DELs (AC124067.4) and two IR-DELs (LINC02604 and MIR4435-2HG) was negatively and positively correlated with the risk score respectively (**Figures 5A–C**).

Differential analysis of tumor immune infiltration showed that 73 and 20 immune cells and factors were significantly reduced and increased, respectively, between normal and patients with COAD (**Supplementary Table 2**). Of these, 41

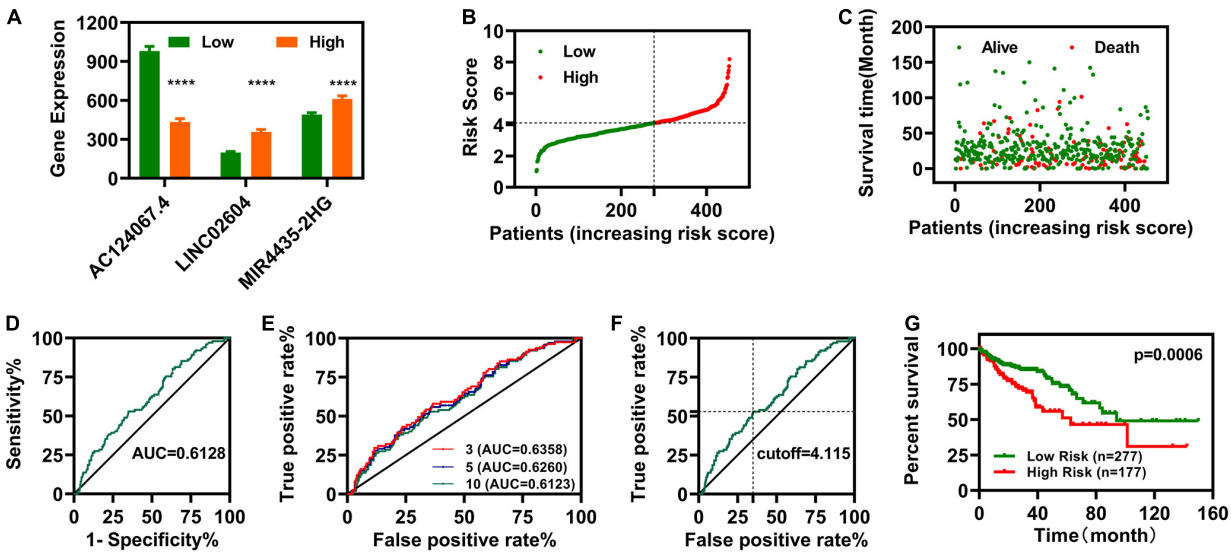


FIGURE 3 | Establishment of a Risk Assessment Model. **(A)** The expression of LINC02604, AC124067.4, and MIR4435-2HG in different groups. **(B,C)** Risk scores **(B)** and survival outcomes **(C)** for each COAD case. **(D)** The ROC curve for the three IR-DELs (AC124067.4, LINC02604, and MIR4435-2HG) models is related to the maximum AUC. **(E)** The 3-, 5-, and 10-year ROC of the optimal model show that all AUC values were over 0.60. **(F)** Risk Scores for 454 patients with COAD; the maximum inflection point is the cut-off point. **(G)** Patients with COAD with low-risk scores have a longer survival time. ****$P < 0.0001$.

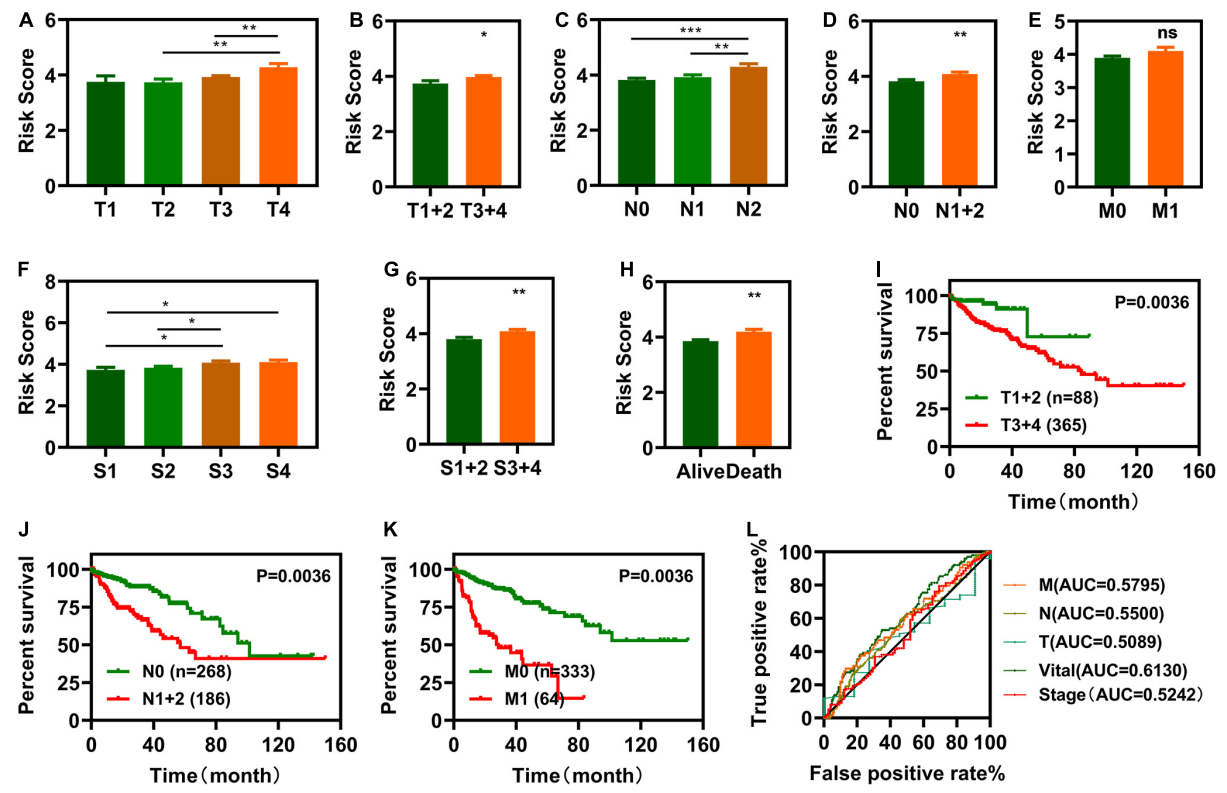


FIGURE 4 | Clinical evaluations with the risk model. **(A–I)** Histograms showing that pathologic tumor (T) **(A,B)**, pathologic node (N) **(C,D)**, pathologic metastasis (M) **(E)**, pathologic stage **(F,G)**, and vital status **(H)** are significantly associated with risk scores. **(I–K)** Patients with COAD in the T1 + 2, N0, and M0 groups have a longer survival time. I, The pathologic N, pathologic M, pathologic T, and pathologic stage ROC of the optimal model showed that all AUC values are lower than those of the vital status. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

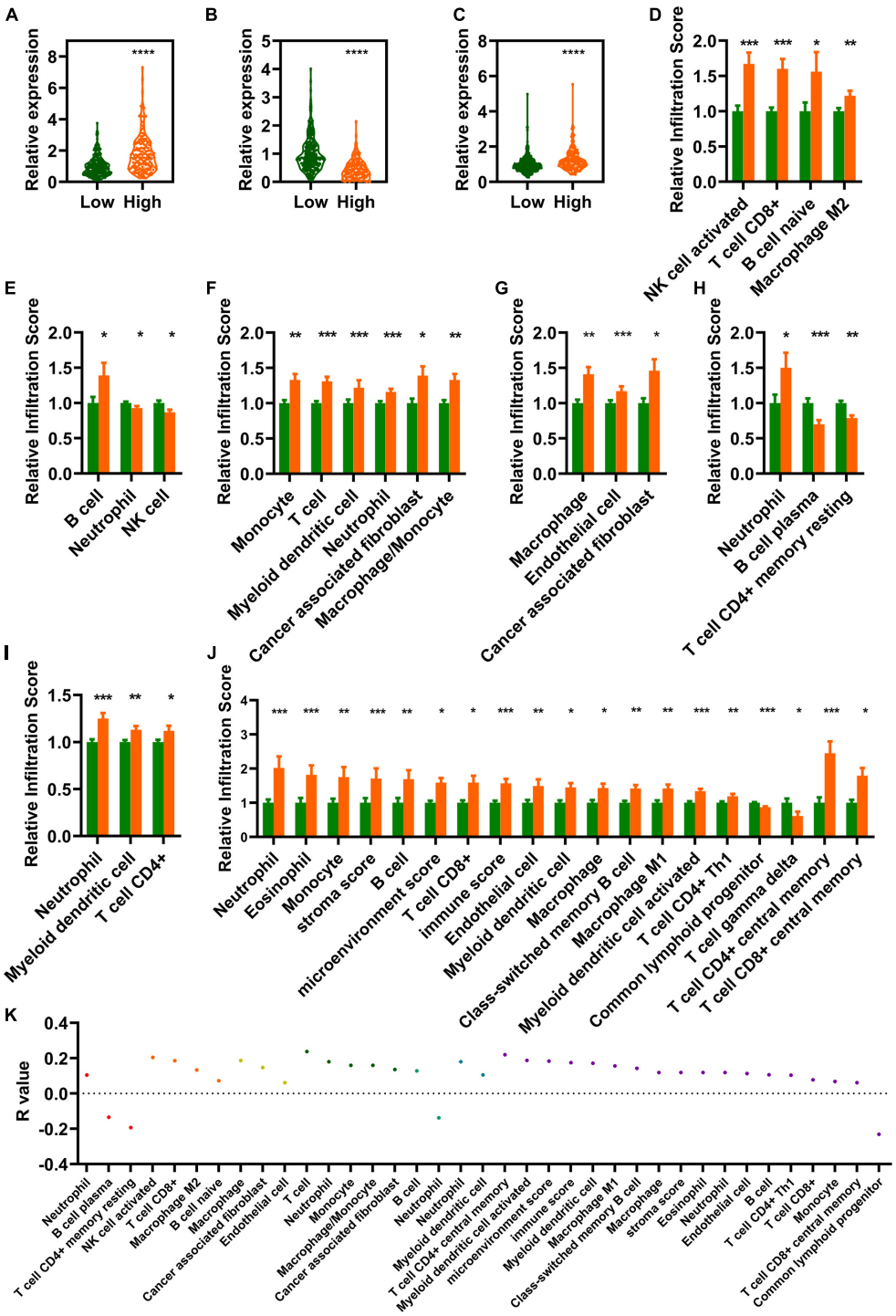


FIGURE 5 | Estimation of tumor immune infiltration using the risk assessment model. **(A–C)** AC124067.4 **(A)**, LINC02604 **(B)**, and MIR4435-2HG **(C)** expression between the low- and high-risk groups. **(D–J)** The infiltrating scores of different tumor immune infiltrations in the low- and high-risk groups are significantly different. **(D)** CIBERSORT-ABS. **(E)** QUANTISEQ. **(F)** MCPCOUNTER. **(G)** EPIC. **(H)** CIBERSORT. **(I)** TIMER. **(J)** XCELL). **(K)** Risk models are significantly correlated with several immune cells and factors ($|R| > 0$, $p < 0.05$). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

immune cells and factors were significantly different between the low- and high-risk groups (**Figures 5D–J**). Spearman correlation analysis of these 41 immune cells and factors showed that 33 and 4 immune cells and factors were positively and negatively correlated with the risk score value respectively (**Figure 5K**).

DISCUSSION

CRC is a common and aggressive cancer associated with high mortality. Although overall mortality rates continue to decline due to improvements in diagnosis and treatment, the survival rate for advanced disease remains low and approximately 50% of patients relapse even after additional treatment with neoadjuvant therapy (Ogura et al., 2019). Cancer immunotherapy is a new alternative for cancer treatment that overcomes the non-specific problems associated with radiotherapy and chemotherapy (Johdi and Sukor, 2020). Hence, it is important to find a suitable immunity-related prognostic biomarker for CRC and to identify patients who can benefit from anti-tumor immunotherapy.

In this study, we identified three IR-DEGs (AC124067.4, LINC02604, and MIR4435-2HG) as possible prognostic biomarkers. AC124067.4, also known as RP11-150O12.3, is located on the shorter arm of chromosome 8 (8p). AC124067.4 was previously demonstrated to be an independent predictor of gastric cancer prognosis (Ren et al., 2016) and its expression was associated with survival in patients with CRC (Poursheikhani et al., 2020). Our study found that the expression of AC124067.4 was significantly upregulated, and its expression level was also associated with the OS rate in patients with COAD. Our findings are consistent with previous reports, reinforcing the possibility of using AC124067.4 as a prognostic biomarker for COAD. LINC02604, also known as LncHERG, is located on the longer arm of chromosome 7 (7q). Shi et al. (2017) found that LINC02604 knockdown inhibited glioblastoma cell proliferation, migration, and invasion *in vitro* and *in vivo*. These inhibitory effects were achieved by regulating miR-940 (Shi et al., 2017). In addition, they also found that patients with glioblastoma with high LINC02604 expression had poor prognosis and low survival (Shi et al., 2017). In the present study, we found that LINC02604 expression was significantly increased in patients with COAD, and they showed poor OS rate, consistent with a previous study (Shi et al., 2017). MIR4435-2HG is located on human chromosome 2q13. Previous studies correlated MIR4435-2HG with several cancers, including gastric cancer, hepatocellular cancer, ovarian cancer, and CRC (Freitas-Andrade et al., 2019; Gong et al., 2019; Kong et al., 2019; Dong et al., 2020; Shen and Zhou, 2020; Zhu et al., 2020). Freitas-Andrade et al. (2019) found that MIR4435-2HG expression was significantly increased in gastric cancer and promoted the growth and metastasis of gastric cancer by activating the Wnt signaling pathway. Kong et al. (2019) found that MIR4435-2HG expression was significantly increased in hepatocellular cancer and could promote proliferation by upregulating miRNA-487a. MIR4435-2HG has been identified as an early diagnostic biomarker for ovarian cancer (Gong et al., 2019) and gastric cancer (Wang et al., 2019). Ke et al. (2017) found that MIR4435-2HG expression was increased in colorectal cancer, and patients with high MIR4435-2HG expression displayed poorer progression-free survival and overall survival. Our results further indicate that MIR4435-2HG can be used as a prognostic biomarker to predict COAD outcome.

Till now, there are many risk assessment models for COAD based on differential expression analysis, including differentially expressed genes, differentially expressed lncRNA, differentially expressed microRNA. Li et al. (2020) found that a risk signature constructed by 10 genes (*CEBPB*, *CXCL9*, *IRF8*, *ITGB1*, *LAG3*, *MCFD2*, *PSMD11*, *RNASE7*, *SPARC*, and *TAP2*) displayed an accuracy of predictions of survival (AUC = 0.6763/0.6465). Miao et al. (2020) constructed a prognosis model by using 12 immune genes, including *SLC10A2*, *CXCL3*, *NOX4*, *FABP4*, *ADIPOQ*, *IGKV1-33*, *IGLV6-57*, *INHBA*, *UCN*, *VIP*, *NGFR*, and *TRDC*. The OS was significantly lower in the high-risk group than in the low-risk group. The 1-, 3-, and 5-year AUC value of ROC was 0.625, 0.646, and 0.713, respectively (Miao et al., 2020). Dai et al. (2018) also constructed a risk signature based on 16 genes (*ATOH1*, *CDC6*, *CXCL10*, *EGR3*, *GALNT4*, *GZMB*, *HSD17B2*, *IFI6*, *INHBB*, *KLK11*, *OSER1-AS1*, *PLAT*, *PTPRR*, *RAB15*, *SPAG1*, and *SPINK1*) and 2 lncRNAs (*PRKAG2-AS1* and *SNHG17*). The 1-, 3-, and 5-year AUC value of ROC in the external validation set was 0.733, 0.667, and 0.673, respectively (Dai et al., 2018). In the present study, we constructed a risk assessment model using three IR-DELs. The AUC value of ROC was 0.6128 which was comparable with the risk model constructed by Li et al. and Miao et al., lower than the risk model constructed by Dai et al. (2018) Comparatively, there were fewer DEGs or DELs used in our present risk assessment model. In conclusion, we identified three IR-DELs that can be used as prognostic biomarkers for COAD and constructed a risk prediction model. However, whether the risk assessment models constructed by previous studies or our present investigation now are feasible in clinical prediction needs to be further verified.

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/s.

AUTHOR CONTRIBUTIONS

MH and X-LX conceived and designed the experiments. X-LX performed the analysis. Z-YY, TZ, CX, and CW helped to analyze the data. Y-WL and X-LX wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was financially supported by the Doctor Foundation of Hunan University of Medicine (2020122004), the Hunan Provincial Science and Technology Department (2020SK4004), and the Hunan Provincial Education Department (20C1328).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.690053/full#supplementary-material

REFERENCES

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66, 683–691. doi: 10.1136/gutjnl-2015-310912

Berraondo, P., Minute, L., Ajona, D., Corrales, L., Melero, I., and Pio, R. (2016). Innate immune mediators in cancer: between defense and resistance. *Immunol. Rev.* 274, 290–306. doi: 10.1111/imr.12464

Bian, Z., Zhang, J., Li, M., Feng, Y., Wang, X., Zhang, J., et al. (2018). LncRNA-FEZF1-AS1 promotes tumor proliferation and metastasis in colorectal cancer by regulating PKM2 signaling. *Clin. Cancer Res.* 24, 4808–4819. doi: 10.1158/1078-0432.ccr-17-2967

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Chansky, K., Sculier, J. P., Crowley, J. J., Giroux, D., Van Meerbeeck, J., Goldstraw, P., et al. (2009). The international association for the study of lung cancer staging project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *J. Thorac. Oncol.* 4, 792–801. doi: 10.1097/jto.0b013e3181a7716e

Chen, Y. P., Zhang, Y., Lv, J. W., Li, Y. Q., Wang, Y. Q., He, Q. M., et al. (2017). Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. *Theranostics* 7, 3585–3594. doi: 10.7150/thno.21471

Cui, C., Zhai, D., Cai, L., Duan, Q., Xie, L., and Yu, J. (2018). Long noncoding RNA HEIH promotes colorectal cancer tumorigenesis via counteracting miR-939mediated transcriptional repression of Bcl-xL. *Cancer Res. Treat.* 50, 992–1008. doi: 10.4143/crt.2017.226

Dai, W., Feng, Y., Mo, S., Xiang, W., Li, Q., Wang, R., et al. (2018). Transcriptome profiling reveals an integrated mRNA-lncRNA signature with predictive value of early relapse in colon cancer. *Carcinogenesis* 39, 1235–1244. doi: 10.1093/carcin/bgy087

Dong, X., Yang, Z., Yang, H., Li, D., and Qiu, X. (2020). Long non-coding RNA MIR4435-2HG promotes colorectal cancer proliferation and metastasis through miR-206/YAP1 Axis. *Front. Oncol.* 10:160. doi: 10.3389/fonc.2020.00160

Emambux, S., Tachon, G., Junca, A., and Tougeron, D. (2018). Results and challenges of immune checkpoint inhibitors in colorectal cancer. *Expert Opin. Biol. Ther.* 18, 561–573. doi: 10.1080/14712598.2018.1445222

Freitas-Andrade, M., Wang, N., Bechberger, J. F., De Bock, M., Lampe, P. D., Leybaert, L., et al. (2019). Targeting MAPK phosphorylation of Connexin43 provides neuroprotection in stroke. *J. Exp. Med.* 216, 916–935. doi: 10.1084/jem.20171452

Gong, J., Xu, X., Zhang, X., and Zhou, Y. (2019). LncRNA MIR4435-2HG is a potential early diagnostic marker for ovarian carcinoma. *Acta Biochim. Biophys. Sin.* 51, 953–959. doi: 10.1093/abbs/gmz085

Huang, Y., Zhang, J., Hou, L., Wang, G., Liu, H., Zhang, R., et al. (2017). LncRNA AK023391 promotes tumorigenesis and invasion of gastric cancer through activation of the PI3K/Akt signaling pathway. *J. Exp. Clin. Cancer Res.* 36:194.

Im, A., and Pavletic, S. Z. (2017). Immunotherapy in hematologic malignancies: past, present, and future. *J. Hematol. Oncol.* 10:94.

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208.

Johdi, N. A., and Sukor, N. F. (2020). Colorectal cancer immunotherapy: options and strategies. *Front. Immunol.* 11:1624. doi: 10.3389/fimmu.2020.01624

Ke, D., Li, H., Zhang, Y., An, Y., Fu, H., Fang, X., et al. (2017). The combination of circulating long noncoding RNAs AK001058, INHBA-AS1, MIR4435-2HG, and CEBPA-AS1 fragments in plasma serve as diagnostic markers for gastric cancer. *Oncotarget* 8, 21516–21525. doi: 10.18632/oncotarget.15628

Kong, Q., Liang, C., Jin, Y., Pan, Y., Tong, D., Kong, Q., et al. (2019). The lncRNA MIR4435-2HG is upregulated in hepatocellular carcinoma and promotes cancer cell proliferation by upregulating miRNA-487a. *Cell Mol. Biol. Lett.* 24:26.

Li, X., Wen, D., Li, X., Yao, C., Chong, W., and Chen, H. (2020). Identification of an immune signature predicting prognosis risk and lymphocyte infiltration in colon cancer. *Front. Immunol.* 11:1678. doi: 10.3389/fimmu.2020.01678

Liang, Z. X., Liu, H. S., Wang, F. W., Xiong, L., Zhou, C., Hu, T., et al. (2019). LncRNA RPPH1 promotes colorectal cancer metastasis by interacting with TUBB3 and by promoting exosomes-mediated macrophage M2 polarization. *Cell Death Dis.* 10:829.

Miao, Y., Wang, J., Ma, X., Yang, Y., and Mi, D. (2020). Identification prognosis-associated immune genes in colon adenocarcinoma. *Biosci. Rep.* 40:BSR20201734.

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., et al. (2019). Cancer treatment and survivorship statistics, 2019. *CA Cancer J. Clin.* 69, 363–385.

Nixon, N. A., Blais, N., Ernst, S., Kollmannsberger, C., Bebb, G., Butler, M., et al. (2018). Current landscape of immunotherapy in the treatment of solid tumours, with future opportunities and challenges. *Curr. Oncol.* 25, e373–e384.

Ogunwobi, O. O., Mahmood, F., and Akingboye, A. (2020). Biomarkers in colorectal cancer: current research and future prospects. *Int. J. Mol. Sci.* 21:5311. doi: 10.3390/ijms21155311

Ogura, A., Konishi, T., Cunningham, C., Garcia-Aguilar, J., Iversen, H., Toda, S., et al. (2019). Neoadjuvant (Chemo)radiotherapy with total mesorectal excision only is not sufficient to prevent lateral local recurrence in enlarged nodes: results of the multicenter lateral node study of patients with low cT3/4 rectal cancer. *J. Clin. Oncol.* 37, 33–43. doi: 10.1200/jco.18.00032

Poursheikhani, A., Abbaszadegan, M. R., Nokhandani, N., and Kerachian, M. A. (2020). Integration analysis of long non-coding RNA (lncRNA) role in tumorigenesis of colon adenocarcinoma. *BMC Med. Genomics* 13:108. doi: 10.1186/s12920-020-00757-2

Ren, W., Zhang, J., Li, W., Li, Z., Hu, S., Suo, J., et al. (2016). A tumor-specific prognostic long non-coding RNA signature in gastric cancer. *Med. Sci. Monit.* 22, 3647–3657. doi: 10.12659/msm.901190

Shen, M. Y., and Zhou, G. R. (2020). Z YZ: LncRNA MIR4435-2HG contributes into colorectal cancer development and predicts poor prognosis. *Eur. Rev. Med. Pharmacol. Sci.* 24, 1771–1777.

Shi, J., Wang, Y. J., Sun, C. R., Qin, B., Zhang, Y., and Chen, G. (2017). Long noncoding RNA lncHERG promotes cell proliferation, migration and invasion in glioblastoma. *Oncotarget* 8, 108031–108041. doi: 10.18632/oncotarget.22446

Wang, H., Wu, M., Lu, Y., He, K., Cai, X., Yu, X., et al. (2019). LncRNA MIR4435-2HG targets desmoplakin and promotes growth and metastasis of gastric cancer by activating Wnt/beta-catenin signaling. *Aging (Albany N.Y.)* 11, 6657–6673. doi: 10.18632/aging.102164

Wu, C. Y., Fu, J. Y., Wu, C. F., Hsieh, M. J., Liu, Y. H., Wu, Y. C., et al. (2015). Survival prediction model using clinico-pathologic characteristics for nonsmall cell lung cancer patients after curative resection. *Medicine* 94:e2013. doi: 10.1097/md.0000000000002013

Wu, X. S., Wang, F., Li, H. F., Hu, Y. P., Jiang, L., Zhang, F., et al. (2017). LncRNA-PAGBC acts as a microRNA sponge and promotes gallbladder tumorigenesis. *EMBO Rep.* 18, 1837–1853. doi: 10.15252/embr.201744147

Yu, X., Yuan, Z., Yang, Z., Chen, D., Kim, T., Cui, Y., et al. (2018). The novel long noncoding RNA u50535 promotes colorectal cancer growth and metastasis by regulating CCL20. *Cell Death Dis.* 9:751.

Zhu, L., Wang, A., Gao, M., Duan, X., and Li, Z. (2020). LncRNA MIR4435-2HG triggers ovarian cancer progression by regulating miR-128-3p/CKD14 axis. *Cancer Cell Int.* 20:145.

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 Xing, Zhang, Yao, Xing, Wang, Liu 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 Genetics

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fgene.2021.685104

Establishment and Validation of a Comprehensive Prognostic Model for Patients With HNSCC Metastasis

Yajun Shen[1], Lingyu Li[1], Yunping Lu[1], Min Zhang[1], Xin Huang[2] and Xiaofei Tang[1]**

[1] Division of Oral Pathology, Beijing Institute of Dental Research, Beijing Stomatological Hospital, Capital Medical University, Beijing, China, [2] Department of Oral and Maxillofacial Surgery, Beijing Stomatological Hospital, Capital Medical University, Beijing, China

Objective: To identify biomarkers related to head and neck squamous cell carcinoma (HNSCC) metastasis and establish a prognostic model for patients with HNSCC.

Methods: HNSCC mRNA expression data of metastasis and non-metastatic samples were downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. After screening the differentially expressed genes (DEGs) in the two datasets, a prognostic model, including clinical factors and biomarkers, was established, and verified in 36 samples of HNSCC by quantitative real-time transcription (qRT)-PCR. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene sets enrichment analysis (GSEA) were consulted to explore the functions of the DEGs.

Results: In total, 108 DEGs were identified. GSEA, GO, and KEGG analyses showed that these DEGs were mainly involved in the proliferation and metastasis of HNSCC. Six genes that were significantly related to metastasis, immune cell infiltration and prognosis were further identified to construct a prognostic gene signature. The reliability of the gene signature was verified in 36 samples of HNSCC. A prognostic model, including tumor stage, risk level, and a nomogram for prediction were further established. Receiver operating characteristic (ROC) analysis, decision curve analysis (DCA), C-index, and calibration plots showed that the model and nomogram perform well.

Conclusion: We constructed a six-gene signature and a nomogram with high performance in predicting the prognosis of patients with HNSCC metastasis.

Keywords: head and neck squamous cell carcinoma, metastasis, overall survival, prognostic signature, nomogram

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Jinhui Liu,
Nanjing Medical University, China
Qian Chen,
Guangxi Medical University Cancer Hospital, China

****Correspondence:***
Xin Huang
huangxin@ccmu.edu.cn
Xiaofei Tang
xftang10@ccmu.edu.cn

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

Received: *24 March 2021*
Accepted: *21 June 2021*
Published: *12 July 2021*

Citation:
Shen Y, Li L, Lu Y, Zhang M, Huang X and Tang X (2021) Establishment and Validation of a Comprehensive Prognostic Model for Patients With HNSCC Metastasis. Front. Genet. 12:685104. doi: 10.3389/fgene.2021.685104

INTRODUCTION

Head and neck cancers (HNCs) rank sixth among the most common cancers worldwide. They arise anywhere in the head and neck, including the tongue, palate, buccal mucosa, throat, and pharynx (Marur and Forastiere, 2016). According to their pathological classification, the most common type is head and neck squamous cell carcinoma (HNSCC), accounting for 95% of HNCs. Continuous exposure to tobacco or alcohol are the main risk factors for HNSCC, and HPV infection is the most important risk factor for oropharyngeal tumors (Hatcher et al., 2016; Lydiatt et al., 2017). Although

there are some ways to treat HNSCC, the survival rate of patients is still very low and the 5-year survival rate has remained below 60% in the past few decades (Miller et al., 2016). The main reasons for death are invasion and metastasis (Zhang et al., 2018). Due to the rich lymphatic system in the head and neck, lymphatic metastasis often occurs in HNSCC, and distant metastasis is also likely (Cho et al., 2015; Duprez et al., 2017). Therefore, the clinical staging of HNC is mainly based on the primary site of the tumor (T), the number of lymph nodes involved (N), and the presence of distant metastasis (M). TNM classification plays an important role in diagnosis, clinical treatment, and cancer registry activities (Huang and O'Sullivan, 2017). However, the tumor staging system is only based on clinicopathological data, and the key biomarkers and exact targets for predicting the development and prognosis of HNSCC remain unavailable. Currently, second-generation gene sequencing has brought a promising future for identifying valuable prognostic factors in HNSCC (Kamps et al., 2017). Although some biomarkers for HNSCC have been developed based on second-generation sequencing, there seems to be no perfect biomarkers that can predict the prognosis of HNSCC patients with metastasis.

The latest progress in the development of whole-genome sequencing and bioinformatics technology has provided new highlights for cancer genomes. The common open tumor database includes The Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), which use innovative genome analysis technologies to accelerate the comprehensive understanding of cancer genetics, thereby helping to develop new strategies for cancer diagnosis, prevention and treatment (Zhang Z. et al., 2019).

In this study, the mRNA expression profiles and metastasis-related clinical data of 499 and 270 HNSCC samples were obtained from the TCGA database and GEO database, respectively. Six key genes related to metastasis of HNSCC were screened out, namely, SYT14, METTL7B, FOXA2, GNG8, TNFRSF13B, and MYO1H, and the expression of these six genes was further verified in 36 samples of HNSCC patients. A novel multi-factor prognostic model, which included tumor stage and risk level was established and could predict the prognosis of patients with HNSCC effectively. The main design and process of this study are shown in **Figure 1**.

MATERIALS AND METHODS

Data Source

The mRNA expression data and clinical data of 528 HNSCC samples were obtained from the TCGA dataset[1] and 270 HNSCC samples from GEO datasets[2], respectively, the accession number of GEO datasets is GSE65858. The following inclusion criteria were used: (1) Patients pathologically diagnosed with HNSCC, with no history of tumors in any other part; (2) Patients with complete clinical data and follow-up data; (3) Patients whose survival time was more than 1 month. Since these databases are public, ethical reviews were exempt. In this study, 499 HNSCC samples from TCGA and 270 HNSCC samples from GSE65858 were enrolled for further analysis. In addition, genes with lower expression levels were deleted. As R package SVA can estimate surrogate variables, directly adjust the known batch effects, and adjust the batch and latent variables in the prediction problem, SVA was used to eliminate batch effects and bias between different data sets (Leek et al., 2012). Moreover, the RNA-seq data of the two data sets were converted into transcripts per million (TPM) values and then transformed to log2 format, the "scale" function in the R package limma was further used for normalization. these samples were divided into two groups according to N stage and M stage. N0 and M0 samples were all pooled in the non-metastatic group, while the others were classified as the metastatic group.

Gene Set Enrichment Analysis (GSEA)

To identify the KEGG pathway enrichment between non-metastatic group and metastatic group, the RNA expression data of the metastatic group and the non-metastatic group were extracted from the TCGA data set, GSEA[3] was performed using GSEA software (version 4.1.0). the predefined gene set "c2.cp.kegg.v7.2.symbols.gmt" was downloaded from the Molecular Signatures Database (MSigDB). Pathways with $FDR < 0.05$ after performing 1,000 permutations were considered to be significantly enriched.

Analysis of Differentially Expressed Genes

R package edgeR was used to identify differentially expressed genes (DEGs) between non-metastasis group and metastasis group with a cutoff of $FDR < 0.05$, and $|\log_2\text{FoldChange}| < 0.5$ in TCGA and GSE65858. The intersection function in R Studio was applied to determine the common DEGs between GSE65858 and TCGA. In order to explore the functions of these DEGs, David was used to perform GO and KEGG pathway enrichment analysis. $P < 0.05$ was defined as significant enrichment.

Kaplan-Meier (KM) Survival Analysis of DEGs

X-tile software (Version 3.6.1) was used to define the best cut-off value for each DEGs, the 499 HNSCC samples in TCGA were divided into high-expression group and low-expression group according to the cut-off value. The KM method in the survival R package was used to found the DEGs that had a significant impact on the overall survival rate, two-stage method were applied to test the significance of the survival analysis. In the first stage, the log-rank test was performed, if $P < 0.05$ was obtained, the entire testing procedure ended, and the two hazard rates were considered to be significantly different. If the two hazard rates crossed, the stage two would be performed by landmark test to calculate the significance before and after the cross point (Qiu and Sheng, 2008).

[1] https://portal.gdc.ancer.gov/

[2] http://www.ncbi.nlm.nih.gov/geo/

[3] https://www.gsea-msigdb.org/gsea/index.jsp

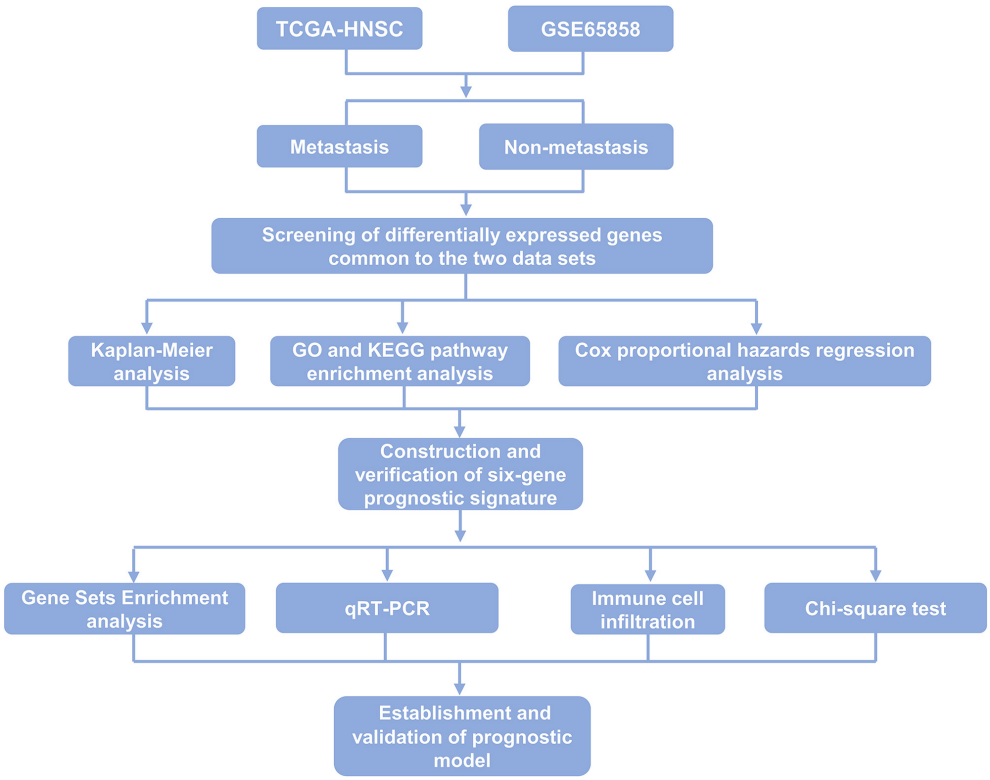


FIGURE 1 | Flow chart of the main procedures of this study.

Identification of DEGs Related to Prognosis

To further identify the DEGs related to prognosis, univariate Cox proportional hazards regression analysis was utilized to analyze the degree of risk of the DEGs selected above. Then, the genes with $P < 0.05$ were included to multivariate Cox proportional hazards regression analysis to find a panel of key candidate genes significantly related to metastasis and prognosis. The best panel was identified by adopting a selection strategy based on the Akaike information criterion (AIC); the less information lost, the higher the quality of the model. GSEA and predefined gene set "c2.cp.kegg.v7.2.symbols.gmt" were used again to explore the functional pathways in which the high and low expression of these candidate genes were significantly enriched ($FDR < 0.05$).

Construction and Validation of the Multi-Gene Prognostic Signature

Next, a multi-gene prognostic signature was constructed with the candidate genes for HNSCC. All the HNSCC samples were scored using the following equation: Risk score = Σexp (RNAi) $\times$ coef (RNAi); where exp (RNA) is the expression level of RNA, and the coefficient (RNA) is the regression coefficient calculated by the multivariate Cox proportional hazards regression model. Then, according to the risk score, the 499 HNSCC samples in TCGA were divided into two groups by X-tile: low-risk and high-risk groups. KM survival analysis and log-rank test were performed between the two groups. In addition, the accuracy of the risk score model for predicting prognosis was assessed using the ROC curve and the area under the curve (AUC) was calculated. GSE65858 was used as an external validation of the risk model.

Patients and Tissue Samples

A total of 36 HNSCC samples were obtained from Beijing Stomatological Hospital of Capital Medical University. This study complied with the Declaration of Helsinki and was approved by the Research Ethics Committee of Beijing Stomatological Hospital of Capital Medical University. Patients were selected according to the following criteria: (1) Patients with HNSCC pathologically diagnosed without other tumors and tumor history; (2) Patients with HNSCC who underwent primary tumor resection and neck lymphatic dissection, but did not receive radiotherapy and/or chemotherapy; (3) Patients with complete follow-up data. The HNSCC tissues obtained during the operation was immediately frozen in liquid nitrogen for storage. As all patients were in M0 stage, patients in N0 stage were defined as non-metastatic group, and patients with N1, N2, and N3 stages were defined as metastatic group.

Quantitative Real-Time Transcription (qRT)-PCR

To further verify the accuracy of the multi-gene signature, total RNA was extracted from all the HNSCC tissues using TRIzol

(Invitrogen Life Technologies, United States) reagent according to the manufacturer's instructions. cDNA was synthesized using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, United States). GAPDH was used as an internal control, and mRNA expression levels were determined by qRT-PCR using SYBR Green (Qiagen, Germany). Gene expression was calculated using the $2^{-\Delta\Delta CT}$ method. All primers, listed in **Table 1**, were designed and compounded by Sangon Biotech (Shanghai, China).

Chi-Square Test

Chi-square test was performed to further explore which clinical factors were significantly related to risk level and survival status in TCGA, such as age, sex, smoking and drinking habits, location of tumor, HPV status, Pathological grade, TNM stage, $P < 0.05$ was considered to be significantly correlated.

The Correlation Between the Risk Level and Immune Cell Infiltration

The R package CIBERSORT was applied in TCGA and GSE65858 to quantify the proportions of 22 immune cell types in the tumor microenvironment, and to find immune cells that are significantly related to the risk level ($P < 0.05$). The immune scores of 499 samples in TCGA were calculated by the ESTIMATE algorithm.

Establishment and Validation of the Multi-Factor Prognostic Model

To establish an efficient multi-factor prognostic model, the 499 samples in TCGA were randomly assigned to two cohorts: the training cohort ($n = 250$) for the establishment of the model and the testing cohort ($n = 249$) for internal validation, GSE65858 was used as an external validation again. The risk level and clinical factors of the training cohort, including age, gender, smoking, drinking, location of tumor, HPV status, pathological grade and TNM stage, were included in the univariate Cox proportional hazards regression analysis. Next, we conducted multivariate Cox proportional hazards regression analysis on the factors with $P < 0.05$. In the output results, $P < 0.05$ was considered as an independent predictor of prognosis. Then, a multi-factor prognostic model was constructed based on the results of the Cox proportional hazards regression analysis. A nomogram was used to estimate the 1-, 3-, and 5-year survival rates of patients with HNSCC. Calibration plots and ROC curves were constructed to assess the predictive accuracy of this nomogram in predicting prognosis. Decision curve analysis (DCA) was performed to further evaluate the clinical net benefit of the nomogram. All the analyses were also applied to the internal validation cohort and external validation cohort. The C-index values of the nomogram in the three cohorts were calculated.

TABLE 1 | The primers used in qRT-PCR.

Gene name	Primer	Sequence (5′–3′)
GNG8	Forward	GAACATCGACCGCATGAAGGT
	Reverse	AGAACACAAAAGAGGCGCTTG
MYO1H	Forward	TAGCCCGTGACAGACTGCT
	Reverse	TTGGTAGACAACTCGGGACTT
TNFRSF13B	Forward	GAGCAAGGCAAGTTCTATGACC
	Reverse	CCTTCCCGAGTTGTCTGAATTG
METTL7B	Forward	TGCTCTTTTTCTGGGAGCAT
	Reverse	GCTGTCGTTCCATTTGGATT
SYT14	Forward	GGTGGAGAGAGAACCTGTGG
	Reverse	ATCTGGAAACCCGCCAACAT
FOXA2	Forward	CGTCGCTGGCTGGCATGTC
	Reverse	GGCTCAGACTCGGACTCAGGTG
GAPDH	Forward	GGAGAAACCTGCCAAGTATGA
	Reverse	CAACCTGGTCCTCAGTGTAGC

RESULTS

DEGs Identification

In the TCGA dataset, the 499 HNSCC samples were divided into a non-metastatic group of 163 samples and a metastatic group of 336 samples. In the GSE65858 dataset, the 270 HNSCC samples were divided into a non-metastatic group of 93 samples and a metastatic group of 177 samples. To explore the potential difference between the two groups in regard to the expression of genes involved in the KEGG pathway, GSEA analysis was conducted. The results showed that the metastasis group was mainly enriched in DNA replication, cell cycle, spliceosome, the P53 signaling pathway, and colorectal cancer, while the non-metastasis group was mainly enriched in arachidonic acid metabolism, the PPAR signaling pathway, and epithelial cell signaling in Helicobacter pylori infection (**Figure 2A**). We then found 462 DEGs in the TCGA dataset (**Figure 2B**), and 370 DEGs in the GSE65858 dataset (**Figure 2C**). There were 108 common DEGs, of which 69 were down-regulated and 39 were up-regulated (**Figure 2D**).

Functional Enrichment Analysis of DEGs

In order to further explore the functions and pathways of these 108 DEGs, we performed GO and KEGG enrichment analysis. A number of GO terms and pathways related to tumors were enriched, mainly extracellular exosome, muscle filament sliding, keratinization, metabolic pathway, the Ras signaling pathway, arachidonic acid metabolism, and the VEGF signaling pathway (**Figure 3**).

KM Survival Analysis of DEGs

Next, we organized the clinical data of HNSCC samples in the TCGA database and used the KM method to further obtain the DEGs that may influence survival outcomes. We found that ACTL8, BCO1, CDHR4, CEBPE, FOXA2, GNG8, METTL7B, MYO1H, SGK2, SLC13A4, SYT14, and TNFRSF13B had a significant impact on the overall survival rate ($P < 0.05$) (**Figure 4**).

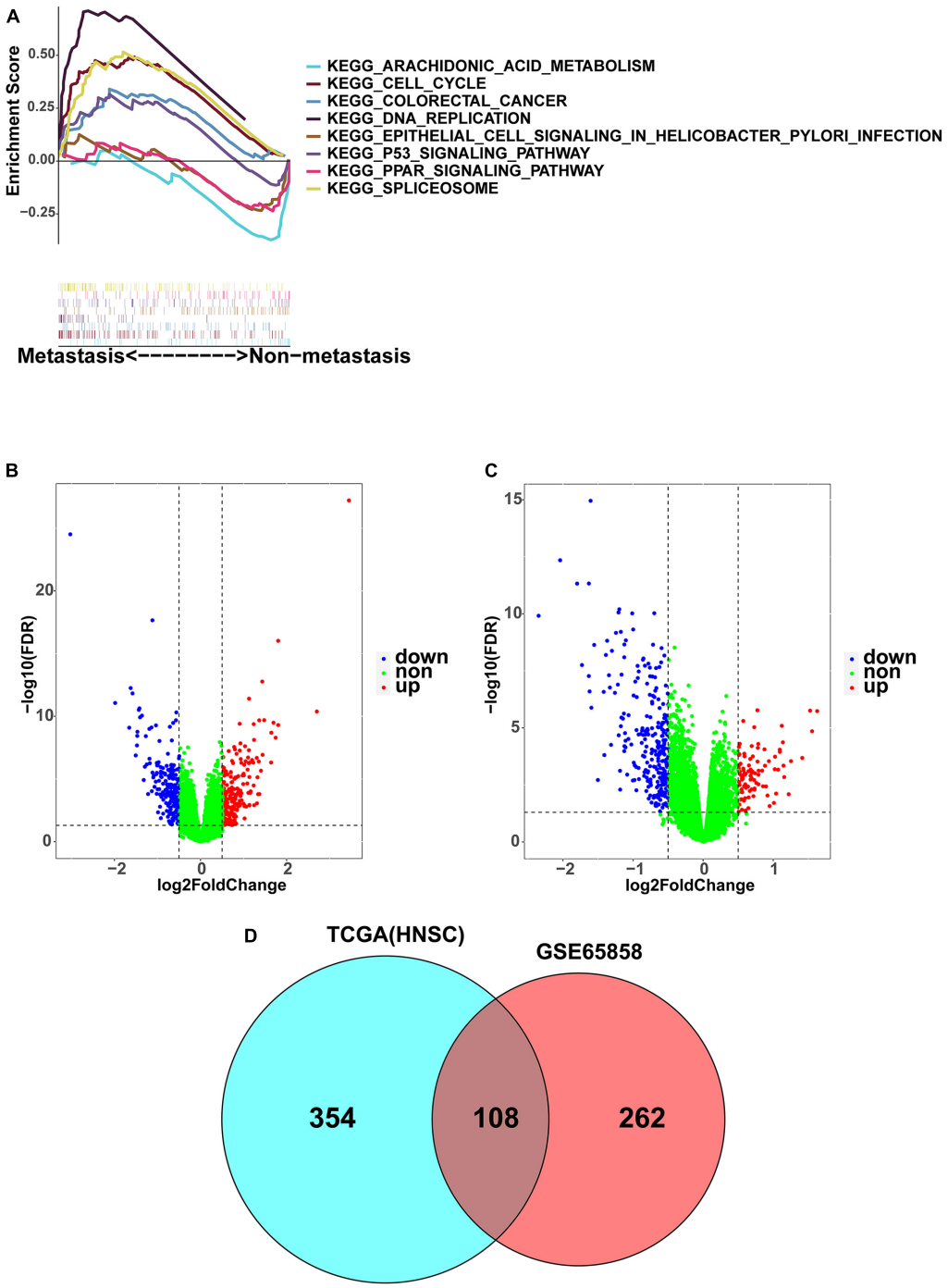


FIGURE 2 | Identification of DEGs associated with HNSCC metastasis. **(A)** The GSEA enrichment analysis between metastasis group and non-metastasis group. **(B)** The volcano map presenting the DEGs involved in metastasis in TCGA-HNSC. **(C)** The volcano map presenting the DEGs involved in metastasis in GSE65858. Blue: downregulated; Red: upregulated. **(D)** As shown in Venn diagram, there are 108 DEGs shared by GSE65858 (red) and TCGA-HNSC (blue).

Construction and Validation of Prognostic Gene Signature

Univariate Cox proportional hazards regression analysis was performed with the 12 candidate genes, and seven genes among them were obtained that were significantly associated with the outcome of each patient ($P < 0.05$), including four low-risk genes (HR < 1): SGK2, MYO1H, TNFRSF13B, and GNG8, as well as three high-risk genes (HR > 1): FOXA2, METTL7B, and SYT14 (**Figure 5A**). The above seven genes were included into multivariate Cox proportional hazards regression analysis.

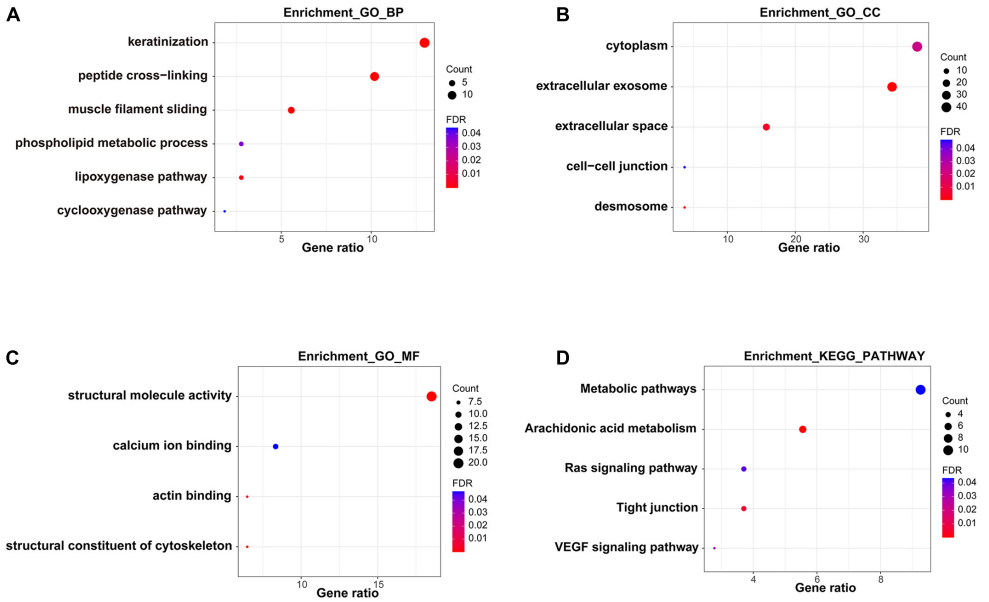


FIGURE 3 | Functional enrichment analysis of DEGs. **(A–C)** GO analysis of DEGs. **(D)** KEGG pathway analysis of DEGs.

Finally, we obtained a panel of genes with the lowest AIC values, including MYO1H, TNFRSF13B, GNG8, FOXA2, METTL7B, and SYT14. The *P*-values of TNFRSF13B, FOXA2, and METTL7B were greater than 0.05, indicating that TNFRSF13B, FOXA2, and METTL7B could not be used as independent prognostic factors, but could be treated as auxiliary prognostic factors, so they were also retained (**Figure 5B**). It was worth noting that GNG8 and TNFRSF13B were highly expressed in the metastasis group, but played an opposite role in the prognosis (**Figures 5C,D**). A prognostic signature containing these six genes was established to predict the risk level of each patient as follows: Risk score = $(-0.757) \times \exp(\text{MYO1H}) + (-0.31) \times \exp(\text{TNFRSF13B}) + (-0.235) \times \exp(\text{GNG8}) + 0.133 \times \exp(\text{FOXA2}) + 0.117 \times \exp(\text{METTL7B}) + 0.186 \times \exp(\text{SYT14})$ (**Figures 5E–H**).

KM survival analysis showed that patients with high-risk levels had significantly lower overall survival rates (**Figures 6A,B**), and the AUC of ROC was 0.7565 ($P < 0.0001$, 95%CI of HR: 0.7113–0.8016) in TCGA, and 0.7766 ($P < 0.0001$, 95%CI of HR: 0.7152–0.8380) in GSE65858 (**Figures 6C,D**), demonstrating that this risk score model is of a certain value. The chi-square test showed that, Age, location of tumor, tumor stage, T stage, N stage M stage and risk level are all significantly correlated with survival status (**Table 2**), among them, tumor stage, T stage, and N stage were also significantly correlated with risk level, it was worth noting that pathological grade is related to risk level, but not to clinical outcome (**Table 3**).

Functional Enrichment Analysis of the Prognostic Gene Signature

GSEA analysis was performed to explore the related functional pathways involving the six-gene signature in HNSCC. It turned out that highly expressed FOXA2, SYT14, and METTL7B were mainly enriched in the ECM receptor interaction, TGFβ signaling pathway, WNT signaling pathway, and MAPK signaling pathway, respectively, while lowly expressed FOXA2, SYT14, and METTL7B were mainly correlated with oxidative phosphorylation, peroxisome, fatty acid metabolism, and Natural killer cell mediated cytotoxicity. Highly expressed MYO1H was mainly enriched in linoleic acid metabolism, metabolism of xenobiotics by cytochrome P450, and retinol metabolism, while lowly expressed MYO1H was mainly enriched in glycosaminoglycan biosynthesis chondroitin sulfate, P53 signaling pathway, and small cell lung cancer. Interestingly, the high expression of GNG8 and TNFRSF13B is not only associated with the VEGF signaling pathway and JAK-STAT signaling pathway, but also related to immune-related pathways, including antigen processing and presentation, B and T cell receptor signaling pathway, and toll-like receptor signaling pathway, while lowly expressed GNG8 and TNFRSF13B were enriched in the adherens junction and proteasome (**Table 4**). These results also explain, to a certain extent, why GNG8 and TNFRSF13B were highly expressed in the metastasis group, but correlated with poor prognosis.

Validation of the Prognostic Value of the Six-Gene Signature

Next, the mRNA expression of GNG8, MYO1H, TNFRSF13B, METTL7B, SYT14, and FOXA2 were detected in 36 HNSCC tissues by qRT-PCR. The results showed that the expression of MYO1H and TNFRSF13B was significantly down-regulated, while the expression of METTL7B, SYT14, and FOXA2 was significantly up-regulated in the HNSCC tissues with lymphatic metastasis (**Figure 7A**). Moreover, the expression

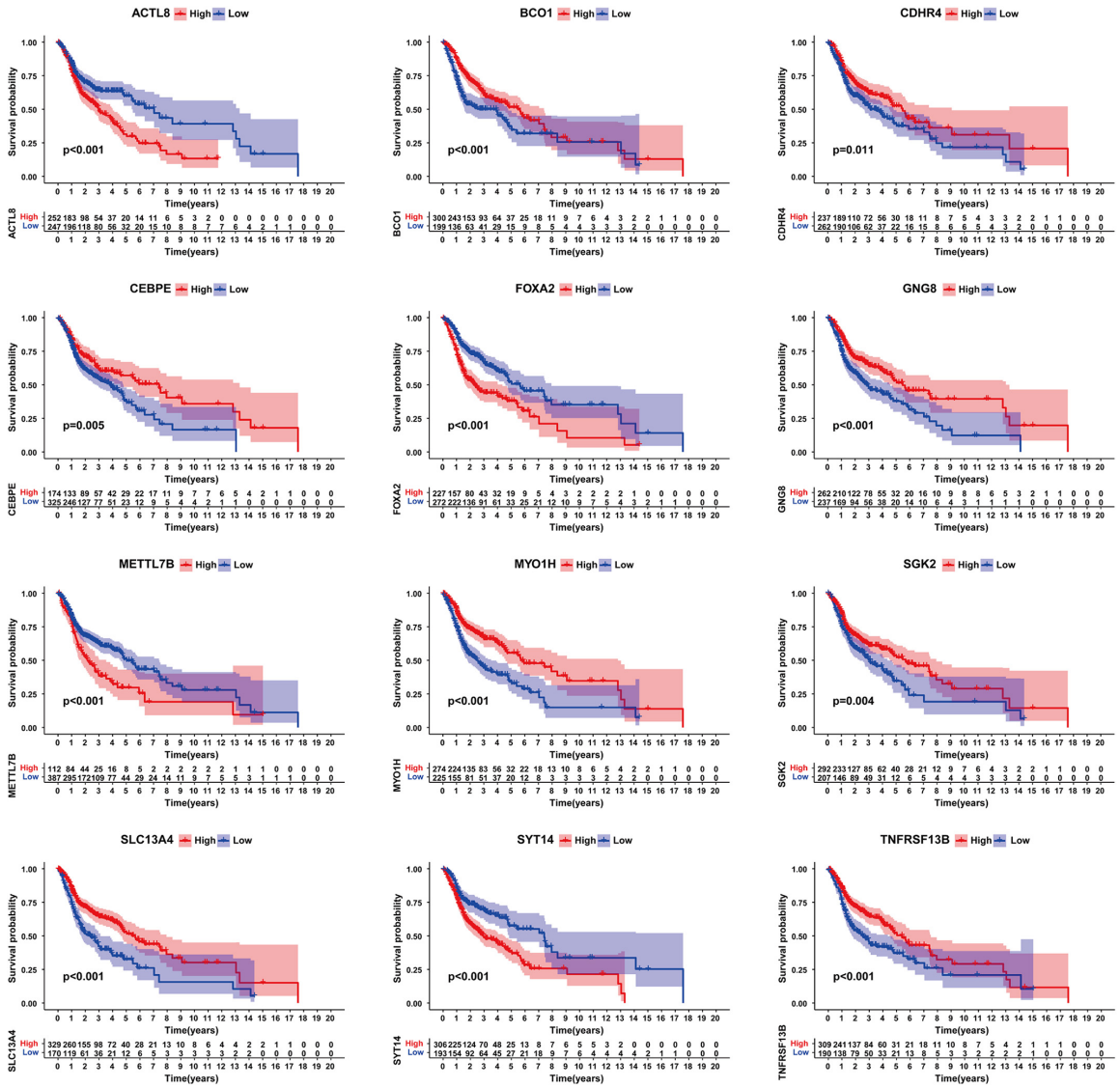


FIGURE 4 | Survival analysis of genes significantly related to survival. Including ACTL8, BCO1, CDHR4, CEBPE, FOXA2, GNG8, METTL7B, MYO1H, SGK2, SLC13A4, SYT14, and TNFRSF13B.

of GNG8, MYO1H, and TNFRSF13B was significantly down-regulated, while the expression of SYT14 and FOXA2 was significantly up-regulated in the samples with higher tumor stage (**Figure 7B**). The risk scores of 36 patients were also calculated, and the high-risk group had a lower overall survival rate (**Figure 7C**). The AUC value of the ROC curve was 0.8515, indicating high accuracy of the six-gene signature in HNSCC samples (**Figure 7D**).

The Correlation Between the Risk Level and Immune Cell Infiltration

The CIBERSORT analysis indicated that the infiltration levels of plasma cells, T cells CD8, T cells CD4 memory activated and T cells follicular helper were significantly lower in the high-risk group than those in the low-risk group, and were negatively associated with the risk level. The infiltration levels of macrophages M0 and mast cells activated in the low-risk group were significantly higher than those in the high-risk group, and were positively correlated with the risk level (**Figures 8A–H**). ESTIMATE algorithm showed that the immune score of the high-risk group was significantly lower than that of the low-risk group (**Figure 8I**). In addition, the immune score also significantly affected the overall survival rate of patients in TCGA (**Figure 8J**).

Establishment and Validation of the Multi-Factor Prognostic Model

The clinical data of the training, testing cohorts and GSE65858 are shown in **Table 5**. Through univariate Cox proportional

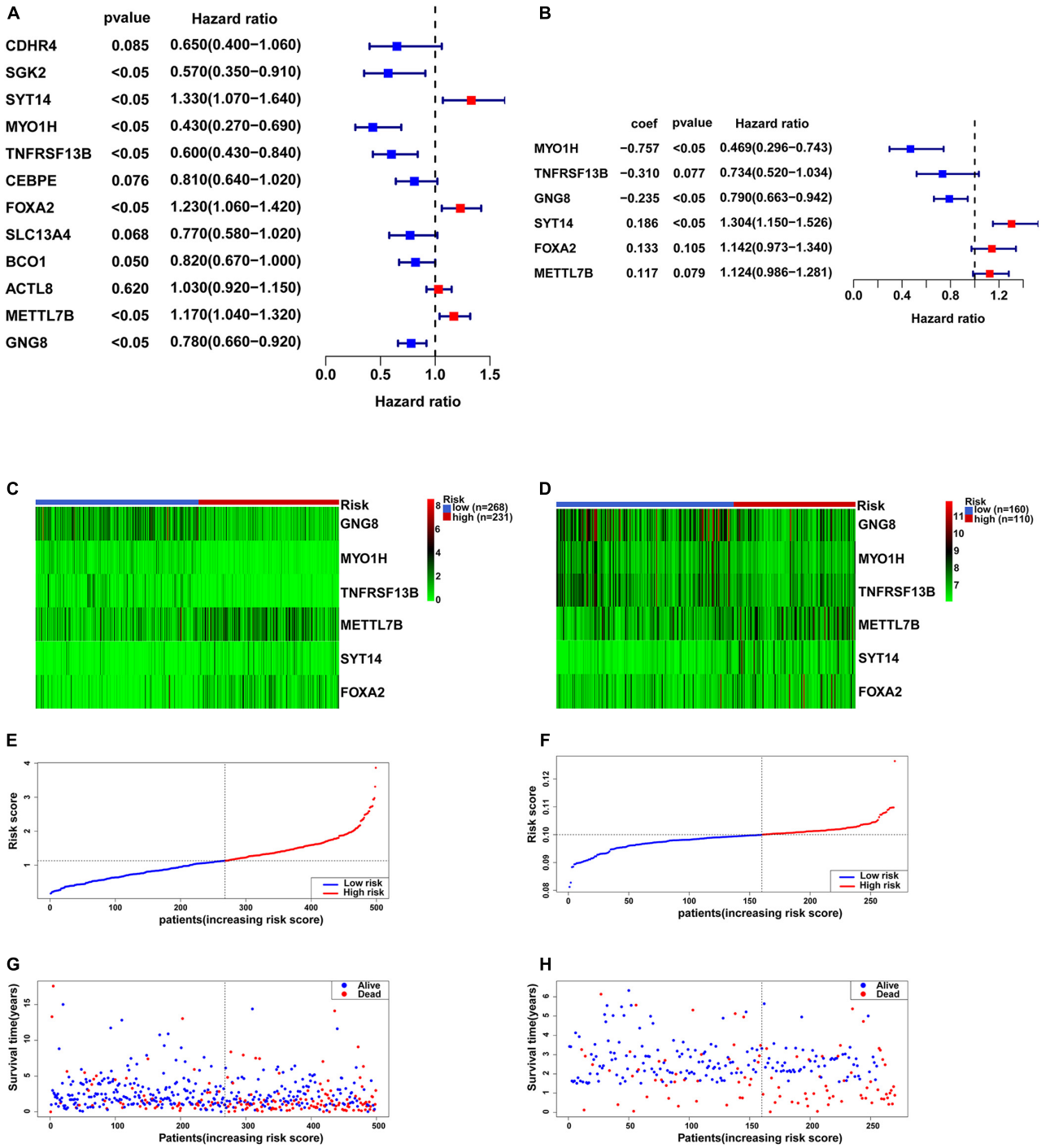


FIGURE 5 | Construction of the six-gene prognostic signature. **(A)** Univariate Cox proportional hazards regression analysis was used to assess whether genes have significant prognostic value. **(B)** Multivariate Cox proportional hazards regression analysis was used to construct a risk biological prognostic signature. coef is the regression coefficient. **(C)** Heatmap of the six genes in the signature in TCGA. **(D)** Heatmap of the six genes in the signature in GSE65858. **(E)** The risk score of each HNSCC patient in TCGA. **(F)** The risk score of each HNSCC patient in GSE65858. **(G)** The survival status and overall survival time of each HNSCC patient in TCGA. **(H)** The survival status and overall survival time of each HNSCC patient in GSE65858.

hazards regression analysis, we found that tumor stage, N stage, and risk level were all high-risk factors for prognosis in the training cohort ($P < 0.05$) (**Figure 9A**), showing that the tumor stage, N stage, and risk level all had prognostic value. The three factors were then included in multivariate Cox proportional regression analysis. We found that the tumor

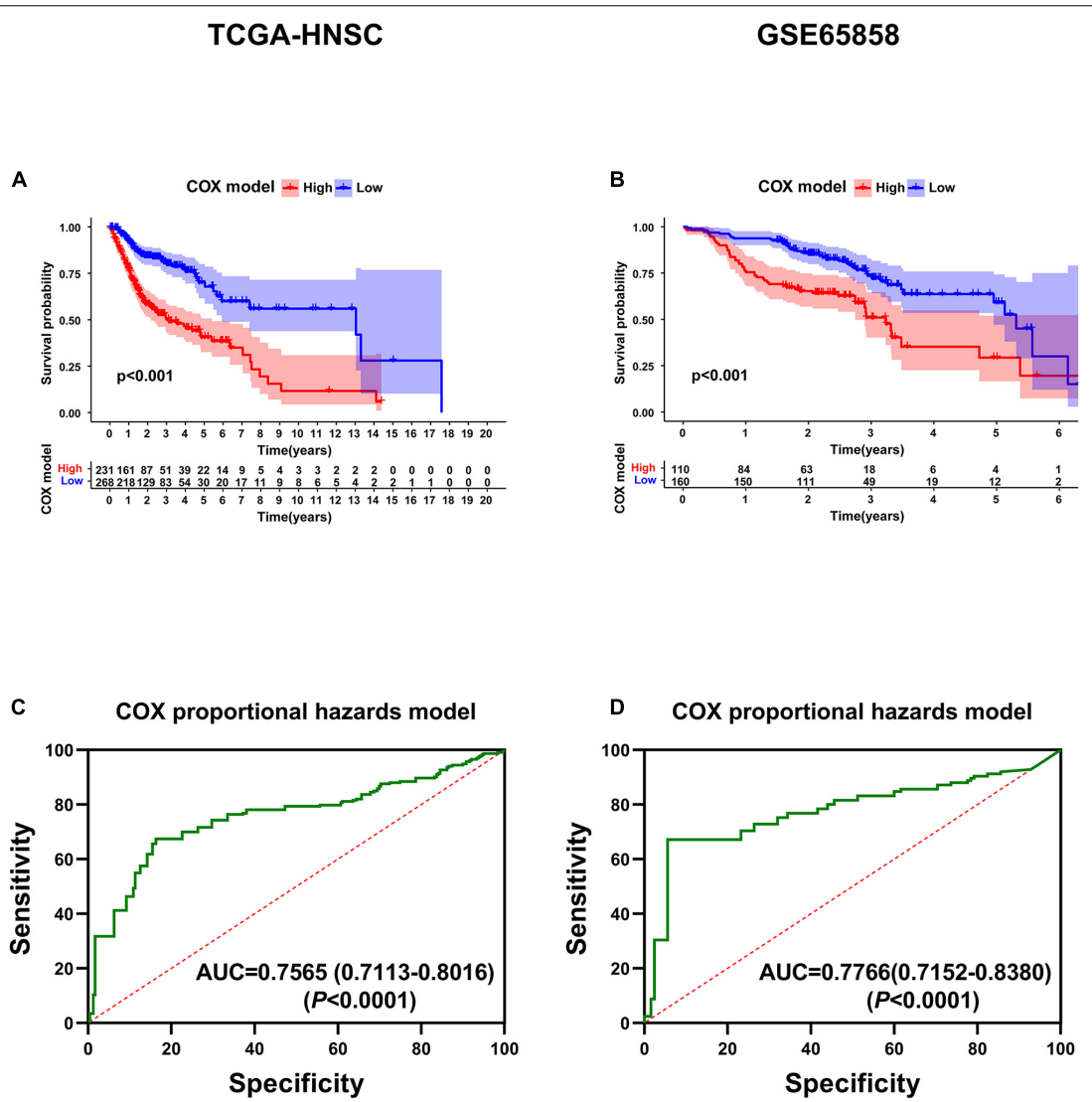


FIGURE 6 | Validation of the six-gene prognostic signature. **(A)** Survival analysis of the prognostic signature in TCGA. **(B)** Survival analysis of the prognostic signature in GSE65858. **(C)** AUC in ROC analysis for the signature at overall survival in TCGA. **(D)** AUC in ROC analysis for the signature at overall survival in GSE65858.

stage and risk level were finally selected, indicating that they could be utilized as independent factors to predict the clinical outcome of patients with HNSCC ($P < 0.05$) (**Figure 9B**). Further, a nomogram was built with the two factors to predict the 1-, 3-, and 5-year survival rates (**Figure 9C**). In the training cohort, the AUC values of the ROC curve for the nomogram were 0.657, 0.700, and 0.745, at 1-, 3-, and 5-year survival, respectively, and 0.724, 0.751, and 0.682 in the testing cohort, 0.747, 0.732, and 0.790 in GSE65858 (**Figures 10A–C**). DCA demonstrated that the combined model consisting of risk level and tumor stage showed the best clinical net benefit (**Figures 10D–F**). The calibration plot also showed the accuracy of this nomogram in the three cohorts, and the C-index values of the nomogram were 0.656, 0.7065, and 0.7055 in the training, testing cohorts and GSE65858, respectively (**Figures 10G–I**).

DISCUSSION

Because of the special anatomical features of the head and neck, HNSCC is very prone to early metastasis, especially lymphatic metastasis, which seriously affects the patient's prognosis (Evans et al., 2019). Exploring molecular biological markers is of great significance to an early diagnosis, prognosis prediction, and treatment strategies of HNSCC. Studies based on deep sequencing have revealed various biomarkers for HNSCC diagnosis. A previous study has developed a 16-gene prognostic signature that can predict the prognosis of patients with tongue squamous cell carcinoma, including CD96, HNF1B, and SMG1 (Qiu et al., 2017). VEGF overexpression is closely related to advanced disease and poor survival in *in vitro* experiments and bioinformatics, revealing a novel HDGF/HIF-1α/VEGF axis in oral cancer prognosis (Lin et al., 2019).

TABLE 2 | The relationship between clinical outcome and clinical factors.

Category	Alive (*n* = 333)	Dead (*n* = 166)	Total (*n* = 499)	*P*-value
Age	Mean 59.96 ± 11.13	Mean 63.31 ± 13.05		**0.020**
< 60	159 (31.9%)	61 (12.1%)	220 (44.0%)	
≥ 60	174 (34.9%)	105 (21.1%)	279 (56.0%)	
Gender				0.060
Male	253 (50.7%)	113 (22.7%)	366 (73.4%)	
Female	80 (16.0%)	53 (10.6%)	133 (26.6%)	
Smoking				0.114
Yes	248 (49.7%)	130 (26.1%)	378 (75.8%)	
No	81 (16.2%)	29 (5.8%)	110 (22.0%)	
Unknown	4 (0.8%)	7 (1.4%)	11 (2.2%)	
Alcohol				0.095
Yes	227 (45.5%)	103 (20.7%)	330 (66.2%)	
No	96 (19.2%)	61 (12.2%)	157 (31.4%)	
Unknown	10 (2.0%)	2 (0.4%)	12 (2.4%)	
HPV status				Invalid
Negative	70 (14.0%)	10 (2.0%)	80 (16.0%)	
Positive	31 (6.2%)	1 (0.2%)	32 (6.4%)	
Unknown	232 (46.5%)	155 (31.1%)	387 (77.6%)	
Location of tumor				**0.022**
Alveolar ridge	14 (2.8%)	4 (0.8%)	18 (3.6%)	
Base of tongue	18 (3.6%)	5 (1.0%)	23 (4.6%)	
Buccal mucosa	17 (3.4%)	5 (1.0%)	22 (4.4%)	
Floor of mouth	37 (7.4%)	23 (4.6%)	60 (12.0%)	
Hard palate	6 (1.2%)	1 (0.2%)	7 (1.4%)	
Hypopharynx	7 (1.4%)	3 (0.6%)	10 (2.0%)	
Larynx	72 (14.5%)	39 (7.8%)	111 (22.3%)	
Lip	2 (0.4%)	1 (0.2%)	3 (0.6%)	
Oral cavity	35 (7.0%)	37 (7.4%)	72 (14.4%)	
Oral tongue	85 (17.1%)	40 (8.0%)	125 (25.1%)	
Oropharynx	8 (1.6%)	1 (0.2%)	9 (1.8%)	
Tonsil	32 (6.4%)	7 (1.4%)	39 (7.8%)	
Pathological grade				0.578
G1	44 (8.8%)	17 (3.4%)	61 (12.2%)	
G2	198 (39.7%)	100 (20.0%)	298 (59.7%)	
G3	76 (15.2%)	42 (8.5%)	118 (23.7%)	
Unknown	15 (3.0%)	7 (1.4%)	22 (4.4%)	
Tumor stage				**0.004**
I + II	81 (16.2%)	22 (4.4%)	103 (20.6%)	
III + IV	252 (50.5%)	144 (28.9%)	396 (79.4%)	
T stage				**0.001**
T1 + T2	149 (29.9%)	48 (9.6%)	197 (39.5%)	
T3 + T4	184 (36.9%)	118 (23.6%)	302 (60.5%)	
N stage				**0.018**
N0	130 (26.1%)	47 (9.4%)	177 (35.5%)	
NX + N1 + N2 + N3	203 (40.7%)	119 (23.8%)	322 (64.5%)	
M stage				**0.015**
M0	281 (56.3%)	153 (30.7%)	434 (87.0%)	
MX + M1	52 (10.4%)	13 (2.6%)	65 (13.0%)	
Risk level				**<0.001**
Low	213 (42.7%)	55 (11.0%)	268 (53.7%)	
High	120 (24.0%)	111 (22.3%)	231 (46.3%)	

P < 0.05 is indicated in bold.

TABLE 3 | The relationship between risk level and clinical factors.

Category	Low risk (*n* = 268)	High risk (*n* = 231)	Total (*n* = 499)	*P*-value
Age	Mean 60.97 ± 11.58	Mean 61.20 ± 12.32		0.487
< 60	122 (24.4%)	98 (19.6%)	220 (44.0%)	
≥ 60	146 (29.3%)	133 (26.7%)	279 (56.0%)	
Gender				0.908
Male	196 (39.3%)	170 (34.1%)	366 (73.4%)	
Female	72 (14.4%)	61 (12.2%)	133 (26.6%)	
Smoking				0.058
Yes	195 (39.1%)	183(36.7%)	378 (75.8%)	
No	68 (13.6%)	42 (8.4%)	110 (22.0%)	
Unknown	5 (1.0%)	6 (1.2%)	11 (2.2%)	
Alcohol				0.677
Yes	179 (35.9%)	151 (30.3%)	330 (66.2%)	
No	82 (16.4%)	75 (15.0%)	157 (31.4%)	
Unknown	7 (1.4%)	5 (1.0%)	12 (2.4%)	
HPV status				Invalid
Negative	48 (9.6%)	32 (6.4%)	80 (16.0%)	
Positive	27 (5.4%)	5 (1.0%)	32 (6.4%)	
Unknown	193 (38.7%)	194 (38.9%)	387 (77.6%)	
Location of tumor				0.124
Alveolar ridge	11 (2.2%)	7 (1.4%)	18 (3.6%)	
Base of tongue	13 (2.6%)	10 (2.0%)	23 (4.6%)	
Buccal mucosa	8 (1.6%)	14 (2.8%)	22 (4.4%)	
Floor of mouth	31 (6.2%)	29 (5.8%)	60 (12.0%)	
Hard palate	5 (1.0%)	2 (0.4%)	7 (1.4%)	
Hypopharynx	5 (1.0%)	5 (1.0%)	10 (2.0%)	
Larynx	59 (11.8%)	52 (10.5%)	111 (22.3%)	
Lip	2 (0.4%)	1 (0.2%)	3 (0.6%)	
Oral cavity	38 (7.6%)	34 (6.8%)	72 (14.4%)	
Oral tongue	60 (12.0%)	65 (13.1%)	125 (25.1%)	
Oropharynx	5 (1.0%)	4 (0.8%)	9 (1.8%)	
Tonsil	31 (6.2%)	8 (1.6%)	39 (7.8%)	
Pathological grade				**0.016**
G1	41 (8.2%)	20 (6.6%)	61 (12.2%)	
G2	147 (29.5%)	151 (26.3%)	298 (59.7%)	
G3	70 (14.0%)	48 (11.1%)	118 (23.7%)	
Unknown	10 (2.0%)	12 (2.4%)	22 (4.4%)	
Tumor stage				**0.01**
I + II	67 (13.4%)	36 (7.2%)	103 (20.6%)	
III + IV	201 (40.3%)	195 (39.1%)	396 (79.4%)	
T stage				**0.005**
T1 + T2	121 (24.2%)	76 (15.2%)	197 (39.5%)	
T3 + T4	147 (29.5%)	155 (31.1%)	302 (60.5%)	
N stage				**0.005**
N0	110 (2.2%)	67 (13.4%)	177 (35.5%)	
NX + N1 + N2 + N3	158 (31.7%)	164 (32.9%)	322 (64.5%)	
M stage				0.577
M0	231 (46.3%)	203 (40.7%)	434 (87.0%)	
MX + M1	37 (74.1%)	28 (5.6%)	65 (13.0%)	

P < 0.05 is indicated in bold.

TABLE 4 | Single-gene GSEA in FOXA2, METTL7B, SYT14, GNG8, MYO1H, and TNFRSF13B.

Gene	High expression	Low expression
FOXA2	ECM receptor interaction	Fatty acid metabolism
	WNT signaling pathway	T cell receptor signaling pathway
	Focal adhesion	Oxidative phosphorylation
	TGFβ signaling pathway	Peroxisome
	Glycosaminoglycan biosynthesis chondroitin sulfata	Arachidonic acid metabolism
METTL7B	Basal cell carcinoma	Linoleic acid metabolism
	ECM receptor interaction	
	Lysosome	
	MAPK signaling pathway	
	WNT signaling pathway	
SYT14	Basal cell carcinoma	Fatty acid metabolism
	Pathways in cancer	Natural killer cell mediated cytotoxicity
	TGFβ signaling pathway	
	WNT signaling pathway	
GNG8	Antigen processing and presentation	Adherens junction
	B cell receptor signaling pathway	Tight junction
	Cytokine cytokine receptor interaction	
	JAK stat signaling pathway	
	Toll like receptor signaling pathway	
	VEGF signaling pathway	
MYO1H	Retinol metabolism	Cell cycle
	Glycosaminoglycan biosynthesis chondroitin sulfata	Linoleic acid metabolism
	Metabolism of xenobiotics by cytochrome P450	P53 signaling pathway
		Small cell lung cancer
		Cytokine cytokine receptor interaction
TNFRSF13B	Natural killer cell mediated cytotoxicity	Proteasome
	Primary immunodeficiency	
	T cell receptor signaling pathway	
	Toll like receptor signaling pathway	
	VEGF signaling pathway	

TGFBI, SPP1, and LAMB3 have been identified as potential biomarkers and survival-influencing factors of HNSCC (Shen et al., 2019). Using the TCGA data, Zhou et al. (2018) found that FAM135B methylation is a favorable independent prognostic biomarker for the overall survival of patients with HNSCC. In addition, Zhang et al. (2020) have identified 14 genes related to immune cell infiltration and can predict the prognosis of HNSCC. However, most studies are limited to comparing non-tumor and tumor tissues, whether these gene signatures are related to the metastasis-prone characteristics of HNSCC is little available. Therefore, exploring biomarkers closely related to HNSCC metastasis and constructing a comprehensive prognostic prediction model are extremely important for improving the survival rate of HNSCC patients.

In this study, we explored DEGs between the metastatic samples and the non-metastatic samples in order to obtain biomarkers related to metastasis. For more reliability of the results, we obtained 108 common DEGs from different data sets, which were likely to be changed in the majority of HNSCC samples. GO analysis showed that keratinization, muscle filament sliding and actin binding were mainly enriched in the "biological process" and "molecular function," which is consistent with a previous study which found that highly invasive tumor cells exhibit enhanced actin polymerization activity and abnormal expression of actin regulatory proteins (Bravo-Cordero et al., 2012). In the "cellular component," extracellular exosomes and extracellular spaces were significantly enriched, which showed that the communication between cells might be essential for cancer progression. In KEGG and GSEA analyses, a variety of signaling pathways were enriched. The P53 signaling pathway, cell cycle, Ras signaling pathway, and VEGF signaling pathway have been proven to play key roles in the progression of many tumors (Diez et al., 2011; Sohn et al., 2018; Meireles Da Costa et al., 2020). In general, the functions and pathways enriched in this study might also promote HNSCC metastasis.

As shown in the KM analysis of the 108 common DEGs, 12 DEGs significantly affected the overall survival rate. Next, we conducted Cox proportional hazards regression analysis on these 12 DEGs, and a six-gene prognostic signature was constructed, including GNG8, MYO1H, TNFRSF13B, SYT14, METTL7B, and FOXA2. The risk score of each patient was calculated, KM analysis showed that the high-risk group had significantly worse overall survival than the low-risk group. To further verify the reliability of this six-gene signature, we collected 36 HNSCC tissue samples and detected the mRNA expression of the six genes by qRT-PCR, and got similar results. The results showed that the expression of MYO1H and TNFRSF13B was significantly down-regulated, while the expression of METTL7B, SYT14, and FOXA2 was significantly up-regulated in the HNSCC tissues with lymphatic metastasis. The expression of GNG8, MYO1H, and TNFRSF13B was significantly down-regulated, while the expression of SYT14 and FOXA2 was significantly up-regulated in the samples with higher tumor stage, the samples with high risk had lower overall survival rates.

Current research shows that MYO1H is significantly related to mandibular deformities (Tassopoulou-Fishell et al., 2012). Interestingly, as the entire myosin superfamily, its internal members have different roles in tumors; for example, MYO1A can inhibit gastrointestinal tumors (Mazzolini et al., 2013), while MYO1E can promote breast cancer invasion (Hallett et al., 2012). It is worth noting that compared with the non-metastasis group, the expression of MYO1H was lower in the metastasis group. The GSEA results also showed that the low expression of MYO1H was significantly related to the cell cycle and P53 signaling pathway.

TNFRSF13B codes for the transmembrane activator, calcium modulator, and cyclophilin ligand interactor (TACI), which is

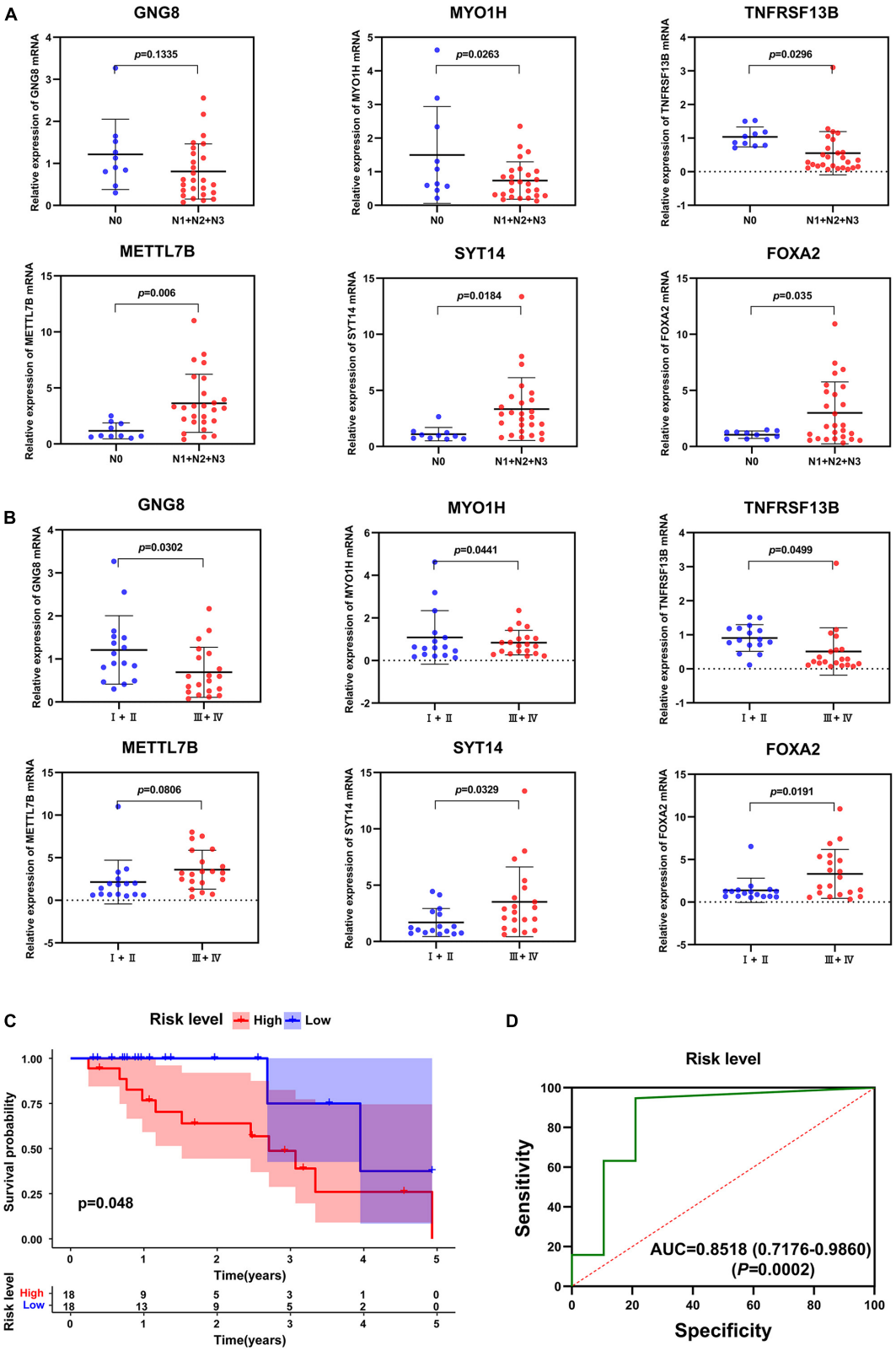


FIGURE 7 | Validation of six-gene prognostic signature by qRT-PCT. **(A)** The expression of these 6 genes in different N stages. **(B)** The expression of these 6 genes in different tumor stages. **(C)** Survival analysis of the prognostic signature in clinical samples. **(D)** AUC in ROC analysis for the signature at overall survival in clinical samples.

TABLE 5 | Clinical data of TCGA-HNSC, GSE65858, and clinical samples.

Category	TCGA training	TCGA testing	GSE65858	Clinical samples
Age	Mean = 61.5 ± 11.0	Mean = 60.7 ± 12.8	Mean = 60.1 ± 10.3	Mean = 63.4 ± 12.2
< 60	105 (21.0%)	115 (23.0%)	153 (56.7%)	21 (58.3%)
≥ 60	145 (29.1%)	134 (26.9%)	117 (43.3%)	15 (41.7%)
Gender				
Male	189 (37.9%)	177 (35.5%)	223 (82.6%)	27 (75.0%)
Female	61 (12.2%)	72 (14.4%)	47 (17.4%)	9 (25.0%)
Smoking				
Yes	46 (9.2%)	64 (12.8%)	48 (17.8%)	0
No	198 (6.6%)	180 (7.8%)	222 (82.2%)	0
Unknown	6 (1.2%)	5 (1.0%)	0 (0%)	36 (100%)
Alcohol				
Yes	169 (33.9%)	161 (32.3%)	239 (88.5%)	0
No	75 (15.0%)	82 (16.4%)	31 (11.5%)	0
Unknown	6 (1.2%)	6 (1.2%)	0 (0%)	36 (100%)
HPV status				
Negative	40 (8.0%)	40 (8.0%)	196 (72.6%)	0
Positive	16 (3.2%)	16 (3.2%)	74 (27.4%)	0
Unknown	194 (38.9%)	193 (38.7%)	0	36 (100%)
Location of tumor				
Alveolar ridge	8 (1.6%)	10 (2.0%)	0	0
Base of tongue	14 (2.8%)	9 (1.8%)	0	2 (5.6%)
Buccal mucosa	8 (1.6%)	14 (2.8%)	0	2 (5.6%)
Floor of mouth	29 (5.8%)	31 (6.2%)	0	0
Hard palate	0	7 (1.4%)	0	0
Hypopharynx	7 (1.4%)	3 (0.6%)	33 (12.2%)	0
Larynx	59 (11.9%)	52 (10.4%)	48 (17.8%)	0
Lip	1 (0.2%)	2 (0.4%)	0	0
Oral cavity	43 (8.6%)	29 (5.8%)	87 (32.2%)	22 (61.1%)
Oral tongue	60 (12.0%)	65 (13.1%)	0	10 (27.7%)
Oropharynx	4 (0.8%)	5 (1.0%)	102 (37.8%)	0
Tonsil	17 (3.4%)	22 (4.4%)	0	0
Pathological grade				
G1	30 (6.0%)	31 (6.2%)	0	11 (30.6%)
G2	153 (30.7%)	145 (29.1%)	0	14 (38.8%)
G3	55 (11.0%)	63 (12.6%)	0	11 (30.6%)
Unknown	12 (2.4%)	10 (2.0%)	270 (100%)	0
Tumor stage				
I + II	49 (9.8%)	54 (10.8%)	55 (20.4%)	16 (44.4%)
III + IV	201 (40.3%)	195 (39.1%)	215 (79.6%)	20 (55.6%)
T stage				
T1 + T2	96 (19.2%)	101 (20.2%)	115 (42.6%)	16 (44.4%)
T3 + T4	154 (30.9%)	148 (29.7%)	155 (57.4%)	20 (55.6%)
N stage				
N0	86 (17.2%)	91 (18.2%)	94 (34.8%)	10 (27.8%)
NX + N1 + N2 + N3	164 (32.9%)	158 (31.7%)	176 (65.2%)	26 (72.2%)
M stage				
M0	217 (43.5%)	217 (43.5%)	263 (97.4%)	36 (100%)
MX + M1	33 (6.6%)	32 (6.4%)	7 (2.6%)	0
Status				
Alive	160 (32.1%)	173 (34.7%)	176 (65.2%)	23 (63.9%)
Dead	90 (18.0%)	76 (15.2%)	94 (34.8%)	13 (36.1%)

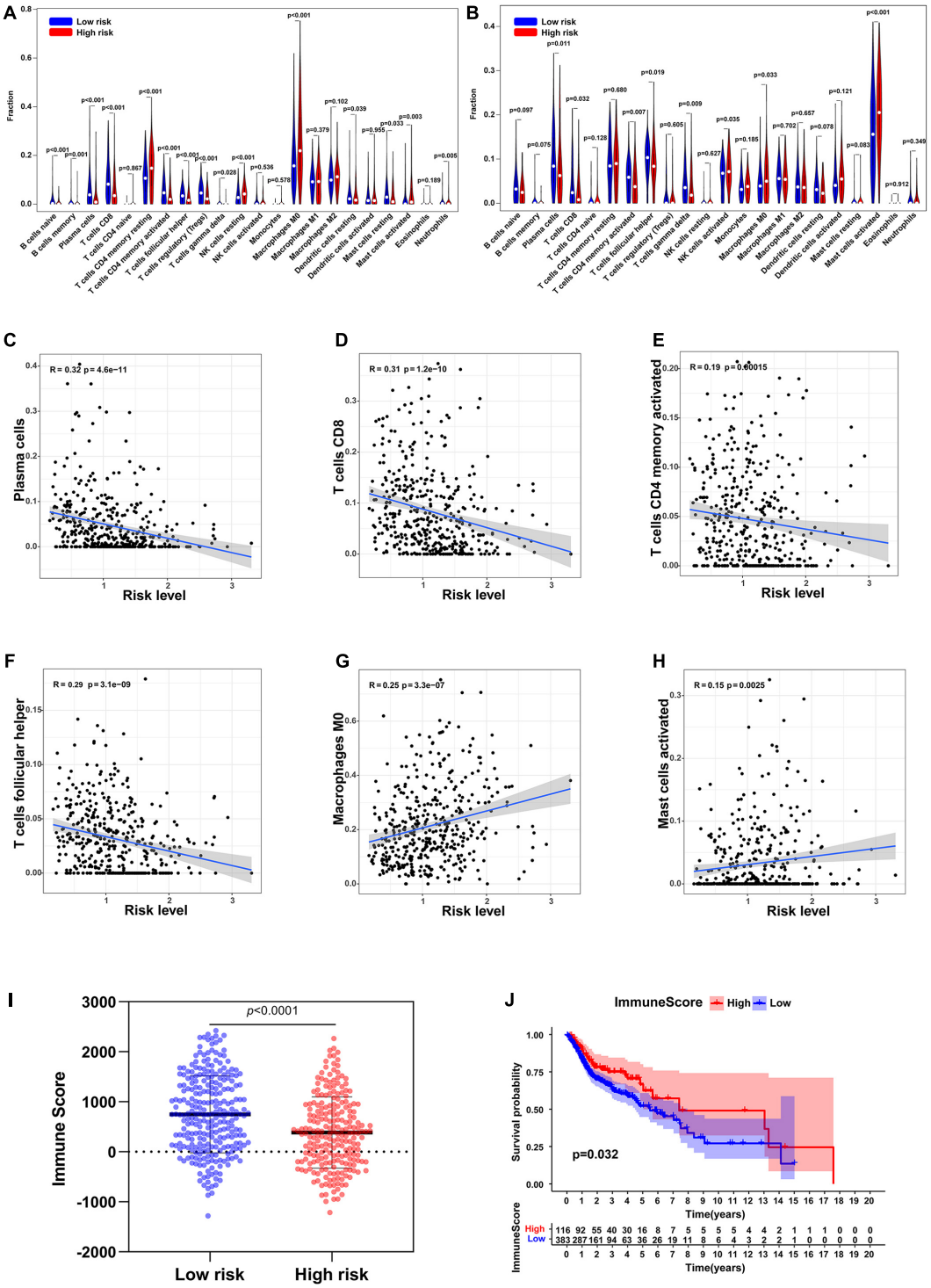


FIGURE 8 | Immune cell infiltration in low-risk and high-risk group. **(A)** Infiltration of 22 types of immune cells in TCGA. **(B)** Infiltration of 22 types of immune cells in GSE65858. **(C–H)** The relationship between the risk level and the infiltration levels of plasma cells, T cells CD8, T cells CD4 memory activated, T cells follicular helper, Macrophages M0, and Mast cells activated. **(I)** Immune scores in low-risk and high-risk groups. **(J)** Survival analysis of the immune score.

mainly expressed on the surface of several B cells and in the marrow of humans (Sakurai et al., 2007). As the main receptor of B cell activation factor (BAFF), TACI can be combined with BAFF to trigger the NFκB typical signal pathway, thereby activating the NFκB anti-apoptotic cascade (Ouyang et al., 2012). TCAI is also a proliferation-inducing ligand (APRIL), and plays an

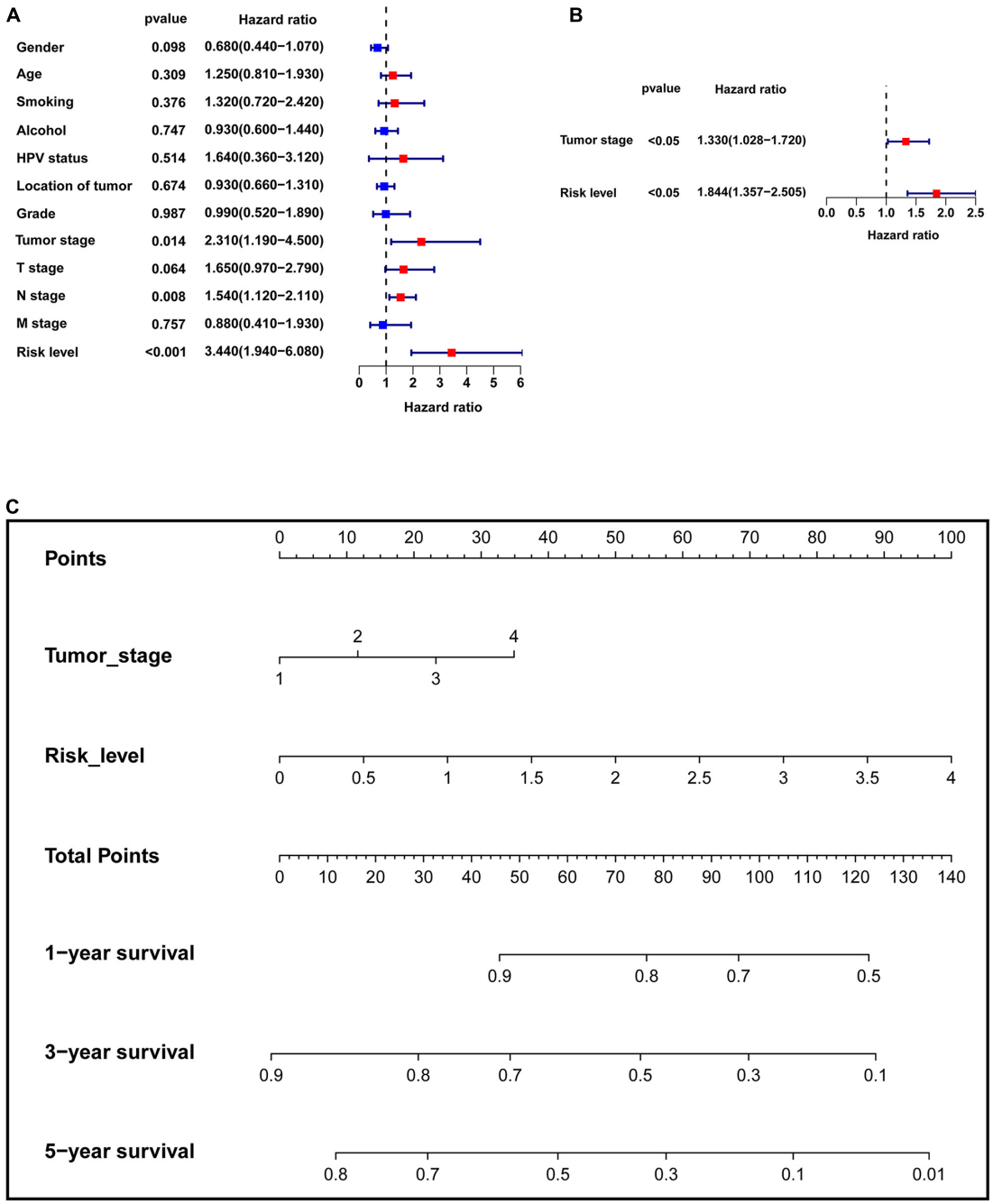


FIGURE 9 | Establishment of the multi-factor prognostic model. **(A)** Univariate Cox proportional hazards regression analysis was performed to assess which factors have significant prognostic value. **(B)** Multivariate Cox proportional hazards regression analysis was used to establish the multivariate prognostic model. **(C)** Nomogram was built to predict 1-, 3-, and 5-year survival rate.

important role in tumor growth and metastasis, including the promotion of cell cycle proliferation and anti-apoptosis (Garcia-Castro et al., 2015). Such a mechanism was verified in a study of breast cancer (Abo-Elfadl et al., 2020). Our results also showed that TNFRSF13B was highly expressed in the metastasis group; unexpectedly, it was low in the high-risk group. GSEA results showed that high TNFRSF13B expression is significantly related to the VEGF signaling pathway, which promotes metastasis, but is also associated with immune responses related to tumors, promoting a good prognosis for patients with HNSCC.

A recent study showed that GNG8 could regulate cognitive function by regulating cholinergic activity (Lee et al., 2020). However, there are few reports about this gene in tumors. The recurrence of sporadic chronic lymphocytic leukemia (CLL)

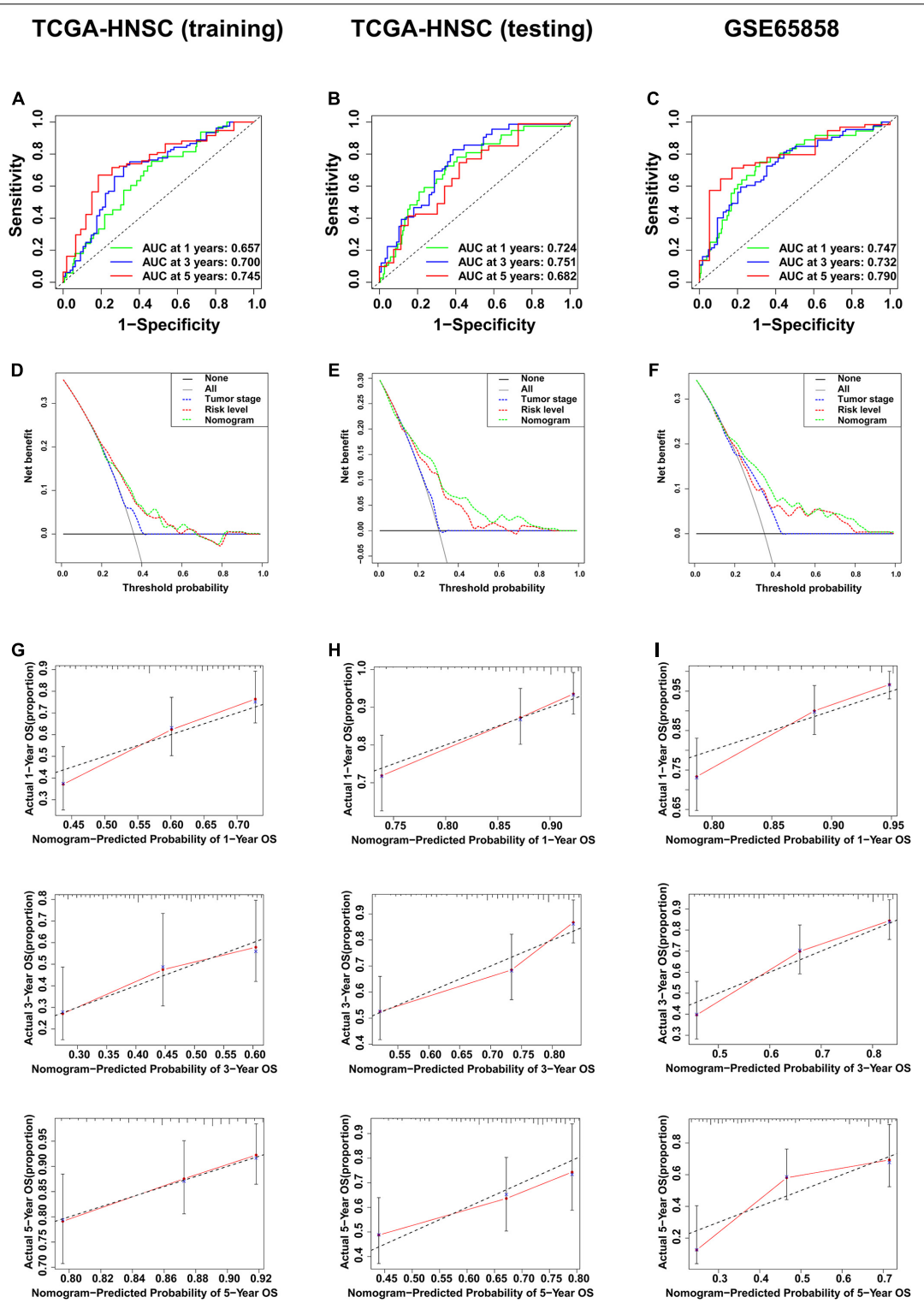


FIGURE 10 | Validation of the multi-factor prognostic model. **(A)** AUC in ROC analysis for the prognostic model in TCGA training cohort. **(B)** AUC in ROC analysis for the prognostic model in TCGA testing cohort. **(C)** AUC in ROC analysis for the prognostic model in GSE65858. **(D)** DCA for the prognostic model in TCGA training cohort. **(E)** DCA for the prognostic model in TCGA testing cohort. **(F)** DCA for the prognostic model in GSE65858. **(G)** Calibration plot for 1-, 3-, and 5-year overall survival in TCGA training cohort. **(H)** Calibration plot for 1-, 3-, and 5-year overall survival in TCGA testing cohort. **(I)** Calibration plot for 1-, 3-, and 5-year overall survival in GSE65858.

and small lymphocytic lymphoma (SLL) may be related to the signaling pathways of certain G proteins (including GNG8) and G protein-coupled receptors (Nuckel et al., 2003). Moreover, there is also a certain relationship between GNG8 and the migration of CLL and SLL cells. In this study, we found that the high expression of GNG8 is significantly related to the VEGF and JAK-STAT signaling pathways, but similarly, it may also cause corresponding immune responses.

It has been reported that FOXA2 is highly expressed in colorectal cancer, participates in the epithelial–mesenchymal transition (EMT) process, and is closely related to the metastasis and clinical staging of colorectal cancer (Wang et al., 2018). In addition, FOXA2 has also been shown to be a target gene of MiR-942, which is involved in the proliferation, migration, and invasion of breast cancer cells (Zhang J. et al., 2019). Through the analysis of our data, we found that FOXA2 was highly expressed in the metastasis and high-risk groups, and was significantly related to the TGFβ signaling pathway and WNT signaling pathway.

METTL7B, encoded by the gene with the same name, is a member of the METTL protein family. The members of this protein family are DNA, RNA, and protein methyltransferases (Ignatova et al., 2019). In the study of non-small cell lung cancer (NSCLC), METTL7B is the target of NSCLC treatment and is involved in the regulation of the tumor cell cycle (Liu et al., 2020). In addition, METTL7B may also activate TGFβ1 and induce EMT in thyroid cancer (Ye et al., 2019). Our results also showed that METTL7B is highly expressed in both the metastasis group and the high-risk group, and is related to signaling pathways, such as the WNT and MAPK signaling pathways.

The function of SYT14 in human cancer is unclear, it has been reported that RNAi-mediated SYT14 knockdown inhibits the growth of human glioma cell line (Sheng et al., 2018). Our results showed thatSYT14 is also highly expressed in both the metastasis group and the high-risk group, and is highly related to the pathways in cancer, TGFβ signaling pathway and WNT signaling pathway.

Immune cell infiltration is an important part of the tumor microenvironment and has important value in predicting tumor prognosis (Gentles et al., 2015). In this study, we found that patients with low immune scores had worse prognosis, and our six-gene signature was significantly related to immune cell infiltration, plasma cells. T cells CD8, T cells CD4 memory activated, and T cells follicular helper were negatively correlated with risk level, while macrophages M0 and mast cells activated were positively correlated with risk level. plasma cells, T cells CD8, T cells CD4 memory activated, and T cells follicular helper were reported to play important roles in inhibiting tumors (Kim and Cantor, 2014; Wouters and Nelson, 2018; Panneton et al., 2019), while macrophages and mast cells activated have been proved to be associated with tumor metastasis and poor prognosis (Komohara et al., 2016; Zhang et al., 2020).

Moreover, we found that tumor stage, T stage, and N stage were all significantly related to survival status and risk level based on the six-gene signature in the chi-square test. As shown in the univariate Cox analysis, tumor stage, N stage, and risk level are all poor prognostic indicators of HNSCC. As the traditional tumor stage has certain limitations in predicting the prognosis of HNSCC (Beltz et al., 2018), we established a prognostic model that combines six-gene signatures related to metastasis with tumor stage in HNSCC using multivariate Cox analysis. Next, the nomogram constructed based on this model provided a more intuitive evaluation tool. In the training cohort, the internal verification cohort and external verification cohort, the ROC curve, C-index, and calibration plot all revealed that the nomogram had ideal predictive performance in terms of the prognosis of 1-, 3-, and 5-year survival. As expected, the DCA also confirmed that the nomogram had higher net benefit than the traditional tumor stage. In addition, this prognostic model was developed based on metastasis, which was more in line with the characteristics of easy metastasis of HNSCC.

However, because of the nature of the data source, clinical data relative to treatment, epidemiology, etc. is limited or not available, the information about impact of tumor location and HPV status on the prognosis of HNSCC, and whether the effects of immunotherapy and chemotherapy are related to risk scores are limited in the current study. In addition, tumor is a very complex disease, including genetic and epigenetic changes which may all lead to inconsistencies in the prediction results. Therefore, it is necessary to combine with a large number of clinical samples for further verification.

In conclusion, we identified 108 DEGs related to HNSCC metastasis, and constructed a biological signature composed of GNG8, MYO1H, TNFRSF13B, METTL7B, SYT14, and FOXA2 for predicting the prognosis of patients with HNSCC. This biological signature was not only related to metastasis and prognosis, but also related to immune cell infiltration. The combined application of these biomarkers can divide HNSCC patients into low-risk or high-risk groups, which can provide useful guidance for individualized and precise treatment. Moreover, a multi-factor prognostic model integrating tumor stage and molecular biomarkers has been established, which can be an effective and convenient tool for the clinical prediction of HNSCC prognosis and selection of treatment strategies.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the Beijing Stomatological Hospital of Capital Medical University (Approval No. CMUSH-IRB-KJ-PJ-2018-01). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS conducted the most of the data mining and data analysis, and drafted the most original manuscript. LL contributed to the design of the research and revised the manuscript. YL participated in data interpretation and manuscript revision. MZ contributed to the revision of the manuscript. XH contributed to the conception and analysis of the study. XT received research funding and participated in the design and supervision of the research. All authors reviewed the manuscript and allowed publication.

FUNDING

This study was funded by the Beijing Municipal Natural Science Foundation of China (Grant No. 7192075).

ACKNOWLEDGMENTS

We are grateful to the GEO database created and maintained by the National Center for Biotechnology Information, and the TCGA database launched by the National Cancer Institute and the National Human Genome Research Institute.

REFERENCES

Abo-Elfadl, M. T., Gamal-Eldeen, A. M., Ismail, M. F., and Shahin, N. N. (2020). Silencing of the cytokine receptor TNFRSF13B: A new therapeutic target for triple-negative breast cancer. *Cytokine* 125:154790. doi: 10.1016/j.cyto.2019.154790

Beltz, A., Gosswein, D., Zimmer, S., Stauber, R. H., Hagemann, J., Strieth, S., et al. (2018). [Staging of oropharyngeal carcinomas : New TNM classification as a challenge for head and neck cancer centers]. *HNO* 66, 375–382. doi: 10.1007/s00106-018-0499-0

Bravo-Cordero, J. J., Hodgson, L., and Condeelis, J. (2012). Directed cell invasion and migration during metastasis. *Curr. Opin. Cell. Biol.* 24, 277–283. doi: 10.1016/j.ceb.2011.12.004

Cho, J. K., Hyun, S. H., Choi, N., Kim, M. J., Padera, T. P., Choi, J. Y., et al. (2015). Significance of lymph node metastasis in cancer dissemination of head and neck cancer. *Transl. Oncol.* 8, 119–125. doi: 10.1016/j.tranon.2015.03.001

Diez, D., Sanchez-Jimenez, F., and Ranea, J. A. (2011). Evolutionary expansion of the Ras switch regulatory module in eukaryotes. *Nucl. Acids Res.* 39, 5526–5537. doi: 10.1093/nar/gkr154

Duprez, F., Berwouts, D., De Neve, W., Bonte, K., Boterberg, T., Deron, P., et al. (2017). Distant metastases in head and neck cancer. *Head Neck* 39, 1733–1743. doi: 10.1002/hed.24687

Evans, M., Baddour, H. M. Jr., Magliocca, K. R., Muller, S., Nannapaneni, S., Chen, A. Y., et al. (2019). Prognostic implications of peritumoral vasculature in head and neck cancer. *Cancer Med.* 8, 147–154. doi: 10.1002/cam4.1910

Garcia-Castro, A., Zonca, M., Florindo-Pinheiro, D., Carvalho-Pinto, C. E., Cordero, A., del Fernando, B., et al. (2015). APRIL promotes breast tumor growth and metastasis and is associated with aggressive basal breast cancer. *Carcinogenesis* 36, 574–584. doi: 10.1093/carcin/bgv020

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, 938–945. doi: 10.1038/nm.3909

Hallett, R. M., Dvorkin-Gheva, A., Bane, A., and Hassell, J. A. (2012). A gene signature for predicting outcome in patients with basal-like breast cancer. *Sci. Rep.* 2:227. doi: 10.1038/srep00227

Hatcher, J. L., Sterba, K. R., Tooze, J. A., Day, T. A., Carpenter, M. J., Alberg, A. J., et al. (2016). Tobacco use and surgical outcomes in patients with head and neck cancer. *Head Neck* 38, 700–706. doi: 10.1002/hed.23944

Huang, S. H., and O'Sullivan, B. (2017). Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr. Treat. Opt. Oncol.* 18:40. doi: 10.1007/s11864-017-0484-y

Ignatova, V. V., Jansen, P., Baltissen, M. P., Vermeulen, M., and Schneider, R. (2019). The interactome of a family of potential methyltransferases in HeLa cells. *Sci. Rep.* 9:6584. doi: 10.1038/s41598-019-43010-2

Kamps, R., Brandao, R. D., Bosch, B. J., Paulussen, A. D., Xanthoulea, S., Blok, M. J., et al. (2017). Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* 18:308. doi: 10.3390/ijms18020308

Kim, H. J., and Cantor, H. (2014). CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. *Cancer Immunol. Res.* 2, 91–98. doi: 10.1158/2326-6066.CIR-13-0216

Komohara, Y., Fujiwara, Y., Ohnishi, K., and Takeya, M. (2016). Tumor-associated macrophages: Potential therapeutic targets for anti-cancer therapy. *Adv. Drug Deliv. Rev.* 99, 180–185. doi: 10.1016/j.addr.2015.11.009

Lee, H. J., Choi, T. I., Kim, Y. M., Lee, S., Han, B., Bak, I. S., et al. (2020). Regulation of habenular G-protein gamma 8 on learning and memory via modulation of the central acetylcholine system. *Mol. Psychiatr.* 516:1. doi: 10.1038/s41380-020-00893-2

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. doi: 10.1093/bioinformatics/bts034

Lin, Y. W., Huang, S. T., Wu, J. C., Chu, T. H., Huang, S. C., Lee, C. C., et al. (2019). Novel HDGF/HIF-1alpha/VEGF axis in oral cancer impacts disease prognosis. *BMC Cancer* 19:1083. doi: 10.1186/s12885-019-6229-5

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

Lydiatt, W. M., Patel, S. G., O'Sullivan, B., Brandwein, M. S., Ridge, J. A., Migliacci, J. C., et al. (2017). Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J. Clin.* 67, 122–137. doi: 10.3322/caac.21389

Marur, S., and Forastiere, A. A. (2016). Head and Neck Squamous Cell Carcinoma: Update on Epidemiology. Diagnosis, Treatment. *Mayo. Clin. Proc.* 91, 386–396. doi: 10.1016/j.mayocp.2015.12.017

Mazzolini, R., Rodrigues, P., Bazzocco, S., Dopeso, H., Ferreira, A. M., Mateo-Lozano, S., et al. (2013). Brush border myosin Ia inactivation in gastric but not endometrial tumors. *Int Cancer* 132, 1790–1799. doi: 10.1002/ijc.27856

Meireles Da Costa, N., Palumbo, A. Jr., De Martino, M., Fusco, A., Ribeiro Pinto, L. F., and Nasciutti, L. E. (2020). Interplay between HMGA and TP53 in cell cycle control along tumor progression. *Cell. Mol. Life Sci.* doi: 10.1007/s00018-020-03634-4

Miller, K. D., Siegel, R. L., Lin, C. C., Mariotto, A. B., Kramer, J. L., Rowland, J. H., et al. (2016). Cancer treatment and survivorship statistics, 2016. *CA Cancer J. Clin.* 66, 271–289. doi: 10.3322/caac.21349

Nuckel, H., Frey, U., Aralh, N., Durig, J., Duhrsen, U., and Siffert, W. (2003). The CC genotype of the C825T polymorphism of the G protein beta3 gene (GNB3) is associated with a high relapse rate in patients with chronic lymphocytic leukaemia. *Leuk Lymphoma* 44, 1739–1743. doi: 10.1080/1042819031000111017

Ouyang, L., Shi, Z., Zhao, S., Wang, F. T., Zhou, T. T., Liu, B., et al. (2012). Programmed cell death pathways in cancer: a review of apoptosis, autophagy and programmed necrosis. *Cell. Prolif.* 45, 487–498. doi: 10.1111/j.1365-2184.2012.00845.x

Panneton, V., Chang, J., Witalis, M., Li, J., and Suh, W. K. (2019). Inducible T-cell co-stimulator: Signaling mechanisms in T follicular helper cells and beyond. *Immunol. Rev.* 291, 91–103. doi: 10.1111/imr.12771

Qiu, P., and Sheng, J. (2008). A two-stage procedure for comparing hazard rate functions. *J. R. Statist.* 70, 191–208. doi: 10.1111/j.1467-9868.2007.00622.x

Qiu, Z., Sun, W., Gao, S., Zhou, H., Tan, W., Cao, M., et al. (2017). A 16-gene signature predicting prognosis of patients with oral tongue squamous cell carcinoma. *PeerJ.* 5:e4062. doi: 10.7717/peerj.4062

Sakurai, D., Kanno, Y., Hase, H., Kojima, H., Okumura, K., and Kobata, T. (2007). TACI attenuates antibody production costimulated by BAFF-R and CD40. *Eur. J. Immunol.* 37, 110–118. doi: 10.1002/eji.200636623

Shen, Y., Liu, J., Zhang, L., Dong, S., Zhang, J., Liu, Y., et al. (2019). Identification of Potential Biomarkers and Survival Analysis for Head and Neck Squamous Cell Carcinoma Using Bioinformatics Strategy: A Study Based on TCGA and GEO Datasets. *Biomed. Res. Int.* 2019:7376034. doi: 10.1155/2019/7376034

Sheng, B., Jiang, Y., Wu, D., Lai, N., Ye, Z., Zhang, B., et al. (2018). RNAi-mediated SYT14 knockdown inhibits the growth of human glioma cell line U87MG. *Brain Res. Bull.* 140, 60–64. doi: 10.1016/j.brainresbull.2018.04.002

Sohn, E. J., Jung, D. B., Lee, H., Han, I., Lee, J., Lee, H., et al. (2018). CNOT2 promotes proliferation and angiogenesis via VEGF signaling in MDA-MB-231 breast cancer cells. *Cancer Lett.* 412, 88–98. doi: 10.1016/j.canlet.2017.09.052

Tassopoulou-Fishell, M., Deeley, K., Harvey, E. M., Sciote, J., and Vieira, A. R. (2012). Genetic variation in myosin 1H contributes to mandibular prognathism. *Am. J. Orthod. Dentofacial. Orthop.* 141, 51–59. doi: 10.1016/j.ajodo.2011.06.033

Wang, B., Liu, G., Ding, L., Zhao, J., and Lu, Y. (2018). FOXA2 promotes the proliferation, migration and invasion, and epithelial mesenchymal transition in colon cancer. *Exp. Ther. Med.* 16, 133–140. doi: 10.3892/etm.2018.6157

Wouters, M. C. A., and Nelson, B. H. (2018). Prognostic Significance of Tumor-Infiltrating B Cells and Plasma Cells in Human Cancer. *Clin. Cancer Res.* 24, 6125–6135. doi: 10.1158/1078-0432.CCR-18-1481

Ye, D., Jiang, Y., Sun, Y., Li, Y., Cai, Y., Wang, Q., et al. (2019). METTL7B promotes migration and invasion in thyroid cancer through epithelial-mesenchymal transition. *J. Mol. Endocrinol.* 63, 51–61. doi: 10.1530/JME-18-0261

Zhang, F., Liu, Y., Yang, Y., and Yang, K. (2020). Development and validation of a fourteen- innate immunity-related gene pairs signature for predicting prognosis head and neck squamous cell carcinoma. *BMC Cancer* 20:1015. doi: 10.1186/s12885-020-07489-7

Zhang, J., Zhang, Z., Sun, J., Ma, Q., Zhao, W., Chen, X., et al. (2019). MiR-942 regulates the function of breast cancer cell by targeting FOXA2. *Biosci. Rep.* 39:BSR20192298. doi: 10.1042/BSR20192298

Zhang, Z., Li, H., Jiang, S., Li, R., Li, W., Chen, H., et al. (2019). A survey and evaluation of Web-based tools/databases for variant analysis of TCGA data. *Brief Bioinform.* 20, 1524–1541. doi: 10.1093/bib/bby023

Zhang, X., Zhang, L., Tan, X., Lin, Y., Han, X., Wang, H., et al. (2018). Systematic analysis of genes involved in oral cancer metastasis to lymph nodes. *Cell. Mol. Biol. Lett.* 23:53. doi: 10.1186/s11658-018-0120-2

Zhou, C., Ye, M., Ni, S., Li, Q., Ye, D., Li, J., et al. (2018). DNA methylation biomarkers for head and neck squamous cell carcinoma. *Epigenetics* 13, 398–409. doi: 10.1080/15592294.2018.1465790

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 July 2021
doi: 10.3389/fgene.2021.702102

Construction and Validation of an Immune-Based Prognostic Model for Pancreatic Adenocarcinoma Based on Public Databases

Miaobin Mao[1,2,3†], Hongjian Ling[1,3†], Yuping Lin[1,2,3], Yanling Chen[4], Benhua Xu[2,3,5,6*] and Rong Zheng[2,3,5,6*]*

[1] The Graduate School, Fujian Medical University, Fuzhou, China, [2] Department of Radiation Oncology, Fujian Medical University Union Hospital, Fuzhou, China, [3] Union Clinical Medicine College, Fujian Medical University, Fuzhou, China, [4] Department of Hepatobiliary Surgery, Fujian Medical University Union Hospital, Fuzhou, China, [5] College of Medical Technology and Engineering, Fujian Medical University, Fuzhou, China, [6] School of Clinical Medicine, Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:
Liuyang Wang,
Duke University, United States

Reviewed by:
Emil Bulatov,
Kazan Federal University, Russia
Antonella Iuliano,
Telethon Institute of Genetics and Medicine (TIGEM), Italy

***Correspondence:**
Rong Zheng
zhengrrong@outlook.com
Benhua Xu
benhuaxu@163.com
Yanling Chen
ylchen@medmail.com.cn

[†]*These authors share first authorship*

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

Received: *29 April 2021*
Accepted: *21 June 2021*
Published: *14 July 2021*

Citation:
Mao M, Ling H, Lin Y, Chen Y, Xu B and Zheng R (2021) Construction and Validation of an Immune-Based Prognostic Model for Pancreatic Adenocarcinoma Based on Public Databases. Front. Genet. 12:702102. doi: 10.3389/fgene.2021.702102

Background: Pancreatic adenocarcinoma (PAAD) is a highly lethal and aggressive tumor with poor prognoses. The predictive capability of immune-related genes (IRGs) in PAAD has yet to be explored. We aimed to explore prognostic-related immune genes and develop a prediction model for indicating prognosis in PAAD.

Methods: The messenger (m)RNA expression profiles acquired from public databases were comprehensively integrated and differentially expressed genes were identified. Univariate analysis was utilized to identify IRGs that related to overall survival. Whereafter, a multigene signature in the Cancer Genome Atlas cohort was established based on the least absolute shrinkage and selection operator (LASSO) Cox regression analysis. Moreover, a transcription factors regulatory network was constructed to reveal potential molecular processes in PAAD. PAAD datasets downloaded from the Gene Expression Omnibus database were applied for the validations. Finally, correlation analysis between the prognostic model and immunocyte infiltration was investigated.

Results: Totally, 446 differentially expressed immune-related genes were screened in PAAD tissues and normal tissues, of which 43 IRGs were significantly related to the overall survival of PAAD patients. An immune-based prognostic model was developed, which contained eight IRGs. Univariate and multivariate Cox regression revealed that the risk score model was an independent prognostic indicator in PAAD (HR $>$ 1, $P < 0.001$). Besides, the sensitivity of the model was evaluated by the receiver operating characteristic curve analysis. Finally, immunocyte infiltration analysis revealed that the eight-gene signature possibly played a pivotal role in the status of the PAAD immune microenvironment.

Conclusion: A novel prognostic model based on immune genes may serve to characterize the immune microenvironment and provide a basis for PAAD immunotherapy.

Keywords: pancreatic adenocarcinoma, immune-related genes, transcription factors, prognostic model, tumor immune microenvironment

INTRODUCTION

Pancreatic cancer (PC) is one of the deadliest and most aggressive malignant neoplasms worldwide (Ilic and Ilic, 2016). In the next decade, PC is estimated to be the second leading cause of death among malignant cancer-related diseases (Rahib et al., 2014; Ferlay et al., 2016). Pancreatic adenocarcinoma (PAAD) occurs in approximately 85% of all PC cases and is associated with a less favorable prognosis (Higuera et al., 2016).

Pancreatic cancer treatment comprises surgery, chemotherapy, radiotherapy, neoadjuvant therapy, targeted molecular therapy, and immunotherapy. Nevertheless, the therapeutic effect of these strategies for PAAD is limited. Therefore, accurate prediction of the prognosis can determine if the patient will benefit from more radical treatment, thereby providing the patient with "individualized" systemic treatment to improve the prognosis.

Pancreatic adenocarcinoma is characterized by the high complexity of stomal tissue, which includes immune cells, various growth factors, the extracellular matrix, and fibroblasts. The tumor microenvironment (TME) accounts for about 15–85% of the whole tumor component in PAAD (Erkan et al., 2012; Liang et al., 2017). The complex and heterogeneous TME induced by interactions between pancreatic epithelial/cancer cells and stromal cells is responsible for PC progression and has been implicated in resistance to chemotherapy and immunotherapy (Markowitz et al., 2015; Incio et al., 2016; Ren et al., 2018). Besides, components of the PAAD microenvironment that contribute to immunosuppression correlate with a poor prognosis of patients (Tang et al., 2014; Whatcott et al., 2015; Wang et al., 2017). With deepening of the understanding of the microenvironment of PC, TME-based clinical and translational therapies could be a breakthrough hotspot in PC treatment in the future.

With the remarkable progress of bioinformatics analysis, in many studies, the mining of public databases has been used increasingly to predict cancer prognosis. Among them, immune-related genes (IRGs) have shown an increasingly prominent role in cancer development and immunotherapy (Ge et al., 2019; Huang et al., 2020; Kong et al., 2020; Yang et al., 2020). Predictive biomarkers related to the tumor immune microenvironment are expected to identify additional target molecules and to enhance immunotherapy efficacy (Taube et al., 2018; Bianco et al., 2019; Jiang et al., 2019; Liu et al., 2020; Zhao B. et al., 2020; Zwing et al., 2020). Currently, PC still lacks prognostic biomarkers related to the tumor immune microenvironment. Therefore, it is necessary to explore important biomarkers in PAAD to guide appropriate treatment options to improve the therapeutic efficacy of patients.

In our research, we investigated the messenger (m)RNA expression and corresponding clinical information of PAAD patients from public databases. Next, we constructed an IRGs-based prognostic model in The Cancer Genome Atlas (TCGA) cohort and validated it in the Gene Expression Omnibus (GEO) dataset. The regulatory network structured by differentially expressed transcription factors (DETFs) and prognosis-related IRGs may provide a theoretical basis to reveal the potential mechanisms at the molecular level. Finally, analyses of prognostic "gene signatures" and infiltration of immune cells may provide new ideas for the role of IRGs in predicting PAAD prognosis.

MATERIALS AND METHODS

Data Acquisition

The transcriptome sequencing data and corresponding clinical data of 176 PAAD patients were extracted from TCGA (172 PAAD specimens and four normal tissue specimens). The RNA-sequencing data of normal pancreatic tissue were acquired from the Genotype-Tissue Expression (GTEx) Project[1] as well (Carithers et al., 2015). It contains the RNA-expression profile of 167 normal pancreatic tissues. Meanwhile, RNA sequencing fragments per kilobase of exon model per million reads mapped (FPKM) data were also obtained for further analyses. For validation cohort, gene expression matrix files and clinical data of 125 patients with PAAD in the GSE71729 dataset were downloaded from the GEO[2]. Match the gene symbols corresponding to the probes according to the annotation file provided by the manufacturer. If a single gene matches multiple probes, the median ranking value accounts for the expression value. We normalized gene expression value using the robust multiarray average (RMA) algorithm, and the normalized data were log2-transformed for further analyses. Publishing guidelines provided by the GEO database were observed, Therefore, there was no requirement for additional ethical approval. Furthermore, a list of IRGs was acquired from the Immunology Database and Analysis Portal (ImmPort) database that shares resources for immunology-related research[3] (Bhattacharya et al., 2014). Then, a list of transcription factors (TFs) was obtained from the Cistrome Project[4], including 318 TFs (Mei et al., 2017).

Analyses of Differentially Expressed Genes in PAAD

The "limma" R package[5] (Ritchie et al., 2015) was used for analyses of differential expression. Differential gene expression was defined with adjusted-$P < 0.01$ and $|\log2 \text{ fold change}| > 2$ as the cutoff criteria. Then, we extracted differentially expressed immune-related genes (DEIRGs) and DETFs from all DEGs based on the lists obtained from ImmPort and Cistrome Cancer databases.

Analyses of DEIRGs in PAAD Using the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Databases

The functions and pathway enrichment of candidate DEIRGs were analyzed using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8[6] (Dennis et al., 2003).

[1] www.gtexportal.org/

[2] https://www.ncbi.nlm.nih.gov/geo/

[3] www.immport.org/

[4] www.cistrome.org/

[5] www.bioconductor.org/packages/release/bioc/html/limma.html

[6] https://david.ncifcrf.gov/

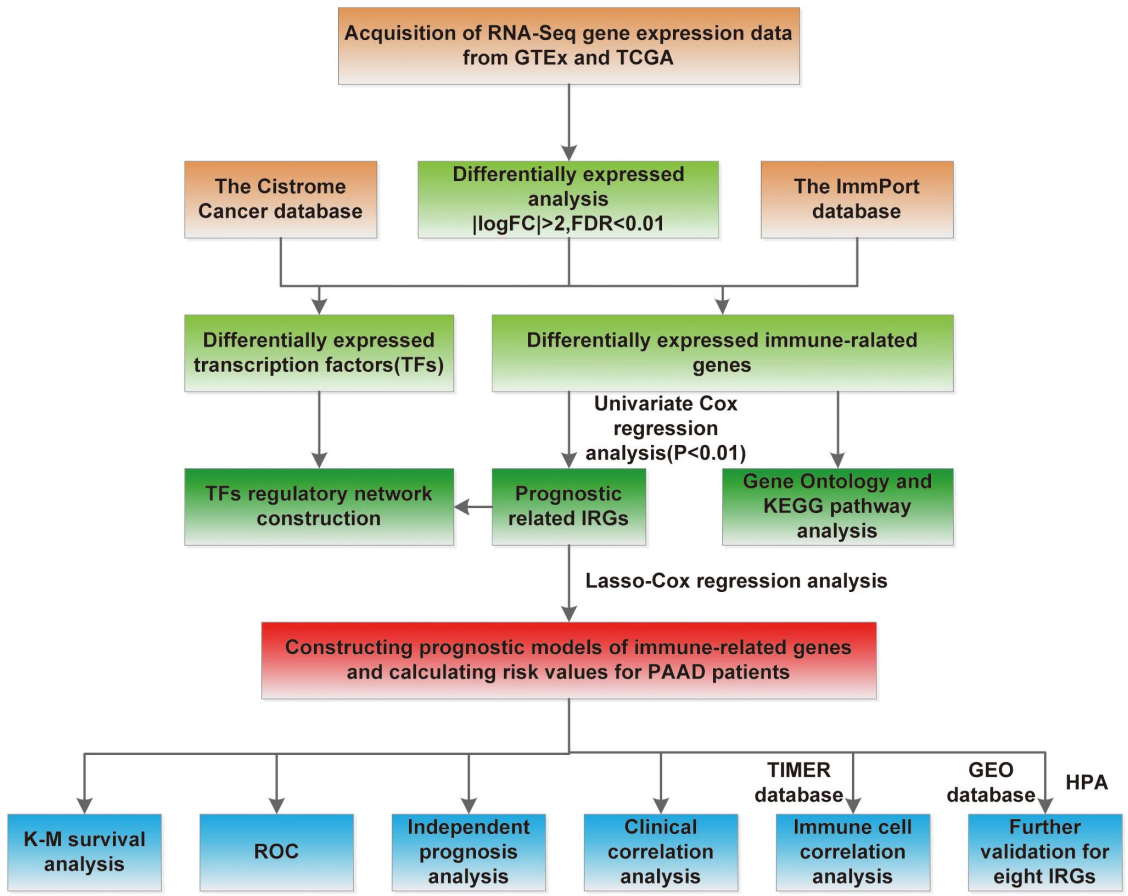


FIGURE 1 | Study flowchart.

To explore the underlying biological functions of DEIRGs, the GO and KEGG databases were searched using the R packages "GOplot[7]" (Walter et al., 2015) and "clusterProfiler[8]" (Kanehisa et al., 2017), respectively. Moreover, the cutoff value for pathway screening and significant functionality was placed at $P < 0.05$. To explain the correlation between enriched pathways and prognostic IRGs, an interaction network was constructed for visual representations.

Transcription Factors-Mediated Prognosis-Related IRGs Modulation Network

A short duration of follow-up usually limits the accuracy of survival analyses. Hence, we selected patients whose duration of follow-up was ≥60 days. To investigate the prognosis-related DEIRGs in PAAD patients, the "survival" R package[9] was applied to implement the univariate Cox regression analysis ($P < 0.01$). To explore the interactions between DETFs and prognosis-related DEIRGs, the correlation test function was employed (set thresholds: $P < 0.001$ and correlation coefficient > 0.4).

Construction of the IRGs-Based Prognostic Model for PAAD

An IRG prognostic model was developed based on the LASSO Cox regression analysis. To minimize the risk of overfitting, the Lasso method used 10-fold cross validation based on the "glmnet" package[10] in R (R Project for Statistical Computing, Vienna, Austria) (Tibshirani, 1997; Simon et al., 2011). Then, we used β coefficients of the LASSO Cox regression analysis to establish the DEIRGs-based prognostic model for PAAD. We used it to establish a formula to predict the risk score of each patient. The receiver operating characteristic (ROC) curve was used to judge the discrimination ability of various statistical methods on the basis of the binary gold standard (Hanley and McNeil, 1982). The ROC curve was created by the "survival ROC" R package[11] to evaluate the sensitivity of the model. Finally, principal component analysis (PCA) was done based on the "prcomp" function from the "stats" R library.

[7]https://cran.r-project.org/web/packages/GOplot/citation.html

[8]www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html

[9]https://cran.r-project.org/web/views/Survival.html

[10]https://cran.r-project.org/web/packages/glmnet/index.html

[11]https://cran.r-project.org/web/packages/survivalROC/index.html

Correlation Between the Immune-Related Signature and Clinical Features in a Prognostic Model of PAAD

The relevance between clinical characteristics (age, gender, histology grade, tumor stage, T staging, N staging, M staging, residual tumor, and outcomes) and expression of eight prognosis signatures in the prognostic model was analyzed using the "beeswarm" R package.

Further Verification of a Prognostic IRG Signature

To verify the prognostic value of the immune-related signature risk score model, we used the GSE71729 dataset as the validation cohort. Samples in the GSE71729 cohort were then divided into high-risk and low-risk groups based on the optimal cut-off point. Kaplan–Meier and ROC curve analysis of the eight-gene signature were performed as mentioned above. In addition, the Human Protein Atlas[12] (Pontén et al., 2011) was used to extract the protein expression of prognostic-related immune genes in tumor samples and normal samples.

Analysis of Immunocyte Infiltration

The Tumor Immune Estimation Resource (TIMER) was employed to analyze and visualize the abundance of tumor-infiltrating immunocytes[13] (Li et al., 2017). It detailed the abundance of six subsets of tumor-infiltrating immunocytes: B cells, $CD8^+$T cells, $CD4^+$T cells, macrophages, neutrophils, and dendritic cells (DCs). The online "Immune Estimation" file was retrieved, and the potential correlation between the prediction model and tumor-infiltrating immunocytes was conducted in R.

Statistical Analysis

Statistical analysis was undertaken with R v3.6.3. Unless specified otherwise, $P < 0.05$ was considered significant.

RESULTS

The flowchart of our study was displayed in **Figure 1**. The clinical features of the 185 PAAD patients enrolled in the TCGA–PAAD cohort were presented in **Table 1**.

Identification of DEIRGs and TFs in PAAD

A total of 343 tissues were analyzed [172 PAAD tissues and 171 normal tissues (167 from the GTEx database)]. Compared with normal tissue specimens, 4,194 genes (expression of 2,313 was upregulated and expression of 1,881 was downregulated; **Supplementary Table 1**), 446 IRGs (expression of 387 was upregulated and expression of 59 was downregulated; **Supplementary Table 2**) and 36 TFs (expression of 29 was upregulated and expression of seven was downregulated; **Table 2**) were identified as differentially expressed in PAAD tissues (set threshold: $P < 0.01$, fold change > 2). The results mentioned above were shown as a heatmap and volcano map (**Figure 2**).

[12]www.proteinatlas.org/

[13]https://cistrome.shinyapps.io/timer/

Functional and Pathway Analyses Using GO and KEGG Databases

We wished to elucidate the biological properties and pathways of DEIRGs in PAAD patients. Hence, the GO and KEGG databases were employed. Inevitably, the DEIRGs were enriched in several immune-related molecular functions. The correlation between the top-five most important GO terms and their related DEIRGs was displayed (adjusted-$P < 0.05$; **Figures 3A–C**). Among them, "GO: 0019814 immunoglobulin complex" was the most prominent GO term. **Figure 3D** displays the top-20 significant pathways. The "pathway-DEIRGs" network (**Figure 3E**) was used for visualizing the reciprocity between the top-10 significant pathways and DEIRGs. **Supplementary Table 4** shows 57 significant pathways according to the KEGG database. Adjusted-$P < 0.05$ was considered indicative of significance. Based on visualized data mining, hsa04060 ("cytokine–cytokine receptor interaction"), hsa04061 ("viral protein interaction with cytokine and cytokine receptor"), and hsa04062 ("chemokine signaling pathway") were used more often to validate our findings using the KEGG database.

TABLE 1 | Clinical features of patients with pancreatic adenocarcinoma (PAAD).

Clinical characteristics	Patients (n = 185)	Percentage (%)
Age (years)		
≤65	96	51.9
>65	89	48.1
Sex		
Female	83	44.9
Male	102	55.1
Grade		
G1	32	17.3
G2	97	52.4
G3	51	27.6
G4	2	1.1
GX	3	1.6
Stage		
I	21	11.4
II	152	82.2
III	4	2.2
IV	5	2.7
NA	3	1.6
Residual tumor		
R0	111	60.0
R1	53	28.6
R2	5	2.7
NA	16	8.6
Outcome		
CR[1]	70	37.8
PR[2]	9	4.9
SD[3]	9	4.9
PD[4]	57	30.8
NA	40	21.6

[1]CR, complete response; [2]PR, partial response; [3]SD, stable disease; [4]PD, progressive disease.

TABLE 2 | Differentially expressed transcription factors (TFs).

TF	Non-tumor (mean)	Tumor (mean)	Log FC[1]	*P* value	FDR[2]
HOXB7	0.624741	7.039385	3.494119	7.91×10^{-55}	2.71×10^{-53}
LMNB1	1.356783	5.614074	2.048858	2.02×10^{-54}	5.34×10^{-53}
E2F1	0.47699	3.251045	2.768873	2.93×10^{-54}	7.12×10^{-53}
BATF	0.696809	10.0601	3.851738	3.84×10^{-54}	8.76×10^{-53}
FOXP3	0.154277	1.493511	3.275111	1.25×10^{-53}	2.13×10^{-52}
SMAD2	6.780617	1.690707	−2.00379	2.19×10^{-53}	3.27×10^{-52}
MEIS1	14.091	3.360955	−2.06783	2.96×10^{-53}	4.07×10^{-52}
BRF1	6.292508	1.1787	−2.41644	4.50×10^{-53}	5.60×10^{-52}
VDR	0.980843	6.460891	2.719638	4.74×10^{-53}	5.80×10^{-52}
SIX5	9.490325	2.016525	−2.23459	2.24×10^{-52}	1.97×10^{-51}
GTF2I	18.92405	3.002807	−2.65584	3.28×10^{-52}	2.72×10^{-51}
GATA3	0.180424	2.441417	3.758254	3.72×10^{-52}	3.02×10^{-51}
PRDM1	0.660779	3.259162	2.302262	3.62×10^{-51}	2.26×10^{-50}
EGR2	0.578491	5.060113	3.128803	7.21×10^{-51}	4.26×10^{-50}
EPO	2.29224	0.230069	−3.31662	8.16×10^{-50}	4.07×10^{-49}
FOSL1	0.866588	9.11024	3.394071	4.07×10^{-49}	1.84×10^{-48}
SPDEF	0.307047	9.991052	5.024106	4.98×10^{-49}	2.23×10^{-48}
CENPA	0.182789	1.443159	2.98098	1.09×10^{-48}	4.68×10^{-48}
LEF1	0.609301	3.669304	2.590279	1.50×10^{-48}	6.35×10^{-48}
MYB	0.105902	0.743503	2.811615	2.54×10^{-48}	1.05×10^{-47}
NCAPG	0.208888	1.467141	2.812209	2.88×10^{-48}	1.19×10^{-47}
KLF5	3.126404	28.38247	3.182424	1.52×10^{-47}	5.88×10^{-47}
FOXP2	1.660682	0.392566	−2.08077	2.35×10^{-47}	8.93×10^{-47}
PPARG	1.176273	6.645312	2.498113	2.19E-46	7.75×10^{-46}
BHLHE40	14.81576	66.10077	2.157535	6.77×10^{-45}	2.15×10^{-44}
FOXM1	0.749049	3.172979	2.082705	9.94×10^{-45}	3.13×10^{-44}
TFAP2A	0.241558	1.740172	2.848786	2.09×10^{-42}	5.82×10^{-42}
FOXA1	0.314465	2.174509	2.789721	1.14×10^{-40}	2.95×10^{-40}
HOXC9	0.104638	1.134456	3.438516	2.26×10^{-39}	5.55×10^{-39}
E2F7	0.12603	0.548239	2.121042	1.47×10^{-33}	3.02×10^{-33}
IRF4	0.180677	0.839005	2.215269	1.41×10^{-32}	2.81×10^{-32}
TP63	0.0931	0.689169	2.88801	2.86×10^{-28}	5.14×10^{-28}
HOXC11	0.014602	0.74272	5.668628	2.76×10^{-25}	4.67×10^{-25}
SOX2	0.106565	0.65707	2.624311	1.33×10^{-22}	2.13×10^{-22}
HOXB13	0.016004	0.577841	5.174176	1.80×10^{-14}	2.47×10^{-14}
MYH11	3.82786	16.02631	2.065833	3.11×10^{-12}	4.09×10^{-12}

[1]*logFC, log fold change (tumor tissues vs. non-tumor tissues);* [2]*FDR, false discovery rate.*

Regulatory Network of TFs

Univariate Cox regression analysis revealed that 43 DEIRGs were associated with overall survival (OS) ($P < 0.01$): 37 high-risk IRGs and six low-risk IRGs (**Figure 4A**). We constructed a regulatory network based on 54 DEIRGs and 36 DETFs (set threshold: $P < 0.001$; correlation coefficient > 0.4). According to the cutoff criteria, 29 prognostic-related DEIRGs and 14 DETFs (**Figure 4B**) participated in the establishment of the network. Finally, the regulatory network was constructed and visualized using Cytoscape software[14]. The TFs lamin B1 (LMNB1) and lymphoid enhancer-binding factor 1 (LEF1) act as negative regulators of IRG SHC adaptor protein 2 (SHC2) (**Figure 4B**). Besides, the TF vitamin-D receptor (VDR) had a negative relationship with IRG fibroblast growth factor 17 (FGF17) and neuregulin 2 (NRG2). The specific regulatory relationship between TFs and OS-related IRGs in PAAD was listed in **Supplementary Table 5**.

[14]www.cytoscape.org/

Construction of an Eight-IRGs Prognostic Model

Least absolute shrinkage and selection operator Cox regression analysis was applied to build a prognostic model based on the expression profile of the 43 prognostic DEIRGs mentioned above. Finally, eight genes were selected to construct a prognostic model based on the optimal value of λ (**Supplementary Figure 1**). The specific formula for the calculation was:

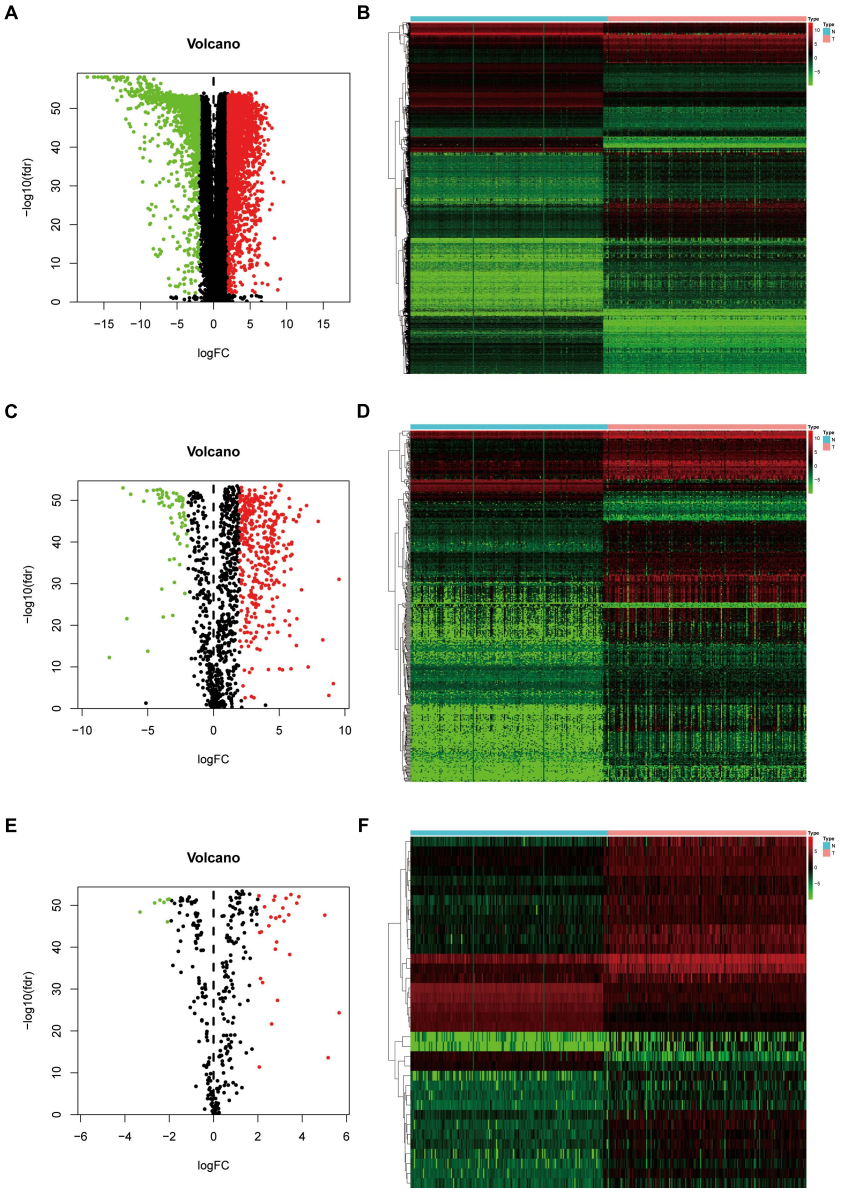


FIGURE 2 | Identification of differentially expressed genes (DEGs), immune-related genes (IRGs), and transcription factors (TFs) in Pancreatic adenocarcinoma (PAAD) *vs*. normal tissues. **(A)** Volcano plot revealing clusters of DEGs with upregulated and downregulated expression. **(B)** The distinction between DEG expression in tumor tissues and normal tissues revealed by a hierarchical clustering heatmap. **(C)** Volcano plot demonstrates clusters of differentially expressed immune-related genes (DEIRGs) with upregulated and downregulated expression. **(D)** Heatmap showing the distinction between expression of DEIRGs in tumor tissues and normal tissues. **(E)** Volcano plot showing clusters of differentially expressed transcription factors (DETFs) with upregulated and downregulated expression. **(F)** Discrimination between DETFs expression in tumor tissues and normal tissues revealed by a heatmap.

$$\begin{aligned}\text{Risk score} = e^{} & [(-0.1301) \times \text{expression of WFIKKN1} \\ & + 0.0016 \times \text{expression of PLAU} + 0.0004 \\ & \times \text{expression of OASL} + (-0.2278) \\ & \times \text{expression of FGF17} + (-0.3925) \\ & \times \text{expression of NPPA} + 0.0258 \\ & \times \text{expression of IL20RB} + 0.0251 \\ & \times \text{expression of MET} + (-0.0081) \\ & \times \text{expression of SHC2}]\end{aligned}$$

Pancreatic adenocarcinoma patients were separated into a high-risk group (n = 81) and a low-risk group (n = 81) based on the median value of the risk score (**Figure 5C**). PCA was undertaken to study the differences between low- and high-risk populations using the expression profiles of all genes, IRGs, and risk-related genes (**Figure 6**). We discovered that low- and high-risk groups were distributed in different directions (**Figure 6C**). Patients with high risk were more likely to die sooner than those with a low risk (**Figure 5D**). The Kaplan–Meier curve

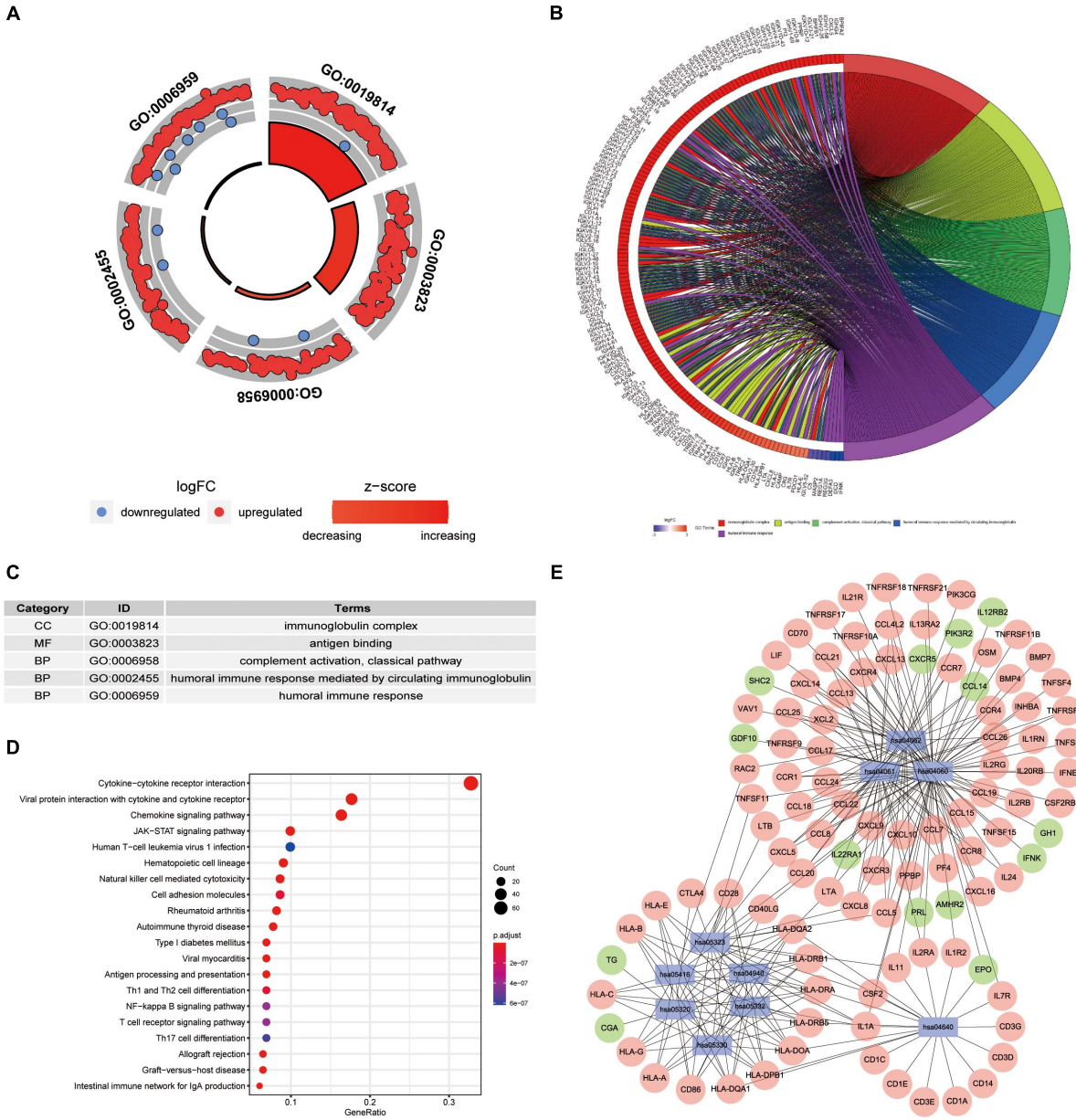


FIGURE 3 | Functional-enrichment analyses of DEIRGs in PAAD. **(A)** The outer circle shows expression (log FC) of DEIRGs in each enriched Gene Ontology (GO) term: red dots which are on each GO term indicate upregulation of DEIRGs. Blue dots indicate downregulated DEIRGs. The inner circle shows the prominence of GO terms (log10-adjusted *P*-values). **(B)** The circle represents the relationship between the top-five most significant GO terms and their related DEIRGs. **(C)** The top-five most significant GO terms and their annotations. **(D)** The top-20 pathways enriched in DEIRGs are shown in the bubble plot. **(E)** The top-10 pathways and the corresponding DEIRGs. The blue rectangles represent Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The red ellipses indicate upregulated DEIRGs. The green ellipses indicate downregulated DEIRGs.

demonstrated that patients with high risk showed markedly worse OS than those with low risk ($P < 0.001$; **Figure 5A**). The area under the time-dependent ROC curves for 1-, 2-, and 3-years OS reached 0.750, 0.697, and 0.707 respectively. Hence, the predictive performance of the prognostic model exhibited good sensitivity and specificity (**Figure 5B**). Also, the immune-based prognostic model was relatively consistent. **Figure 5E** shows the expression of eight IRGs in the form of a heatmap.

Independent Prognostic Value of the Eight-Gene Signature

The independent predictive value of the prognostic signature was assessed by univariate and multivariate Cox regression analyses. Univariate Cox prognostic analyses demonstrated that the risk score was correlated significantly with OS [hazard ratio (HR) = 4.910, 95% confidence interval (CI) = 3.021–7.980, $P < 0.001$] (**Figure 7A**). After the multivariate analysis, the risk

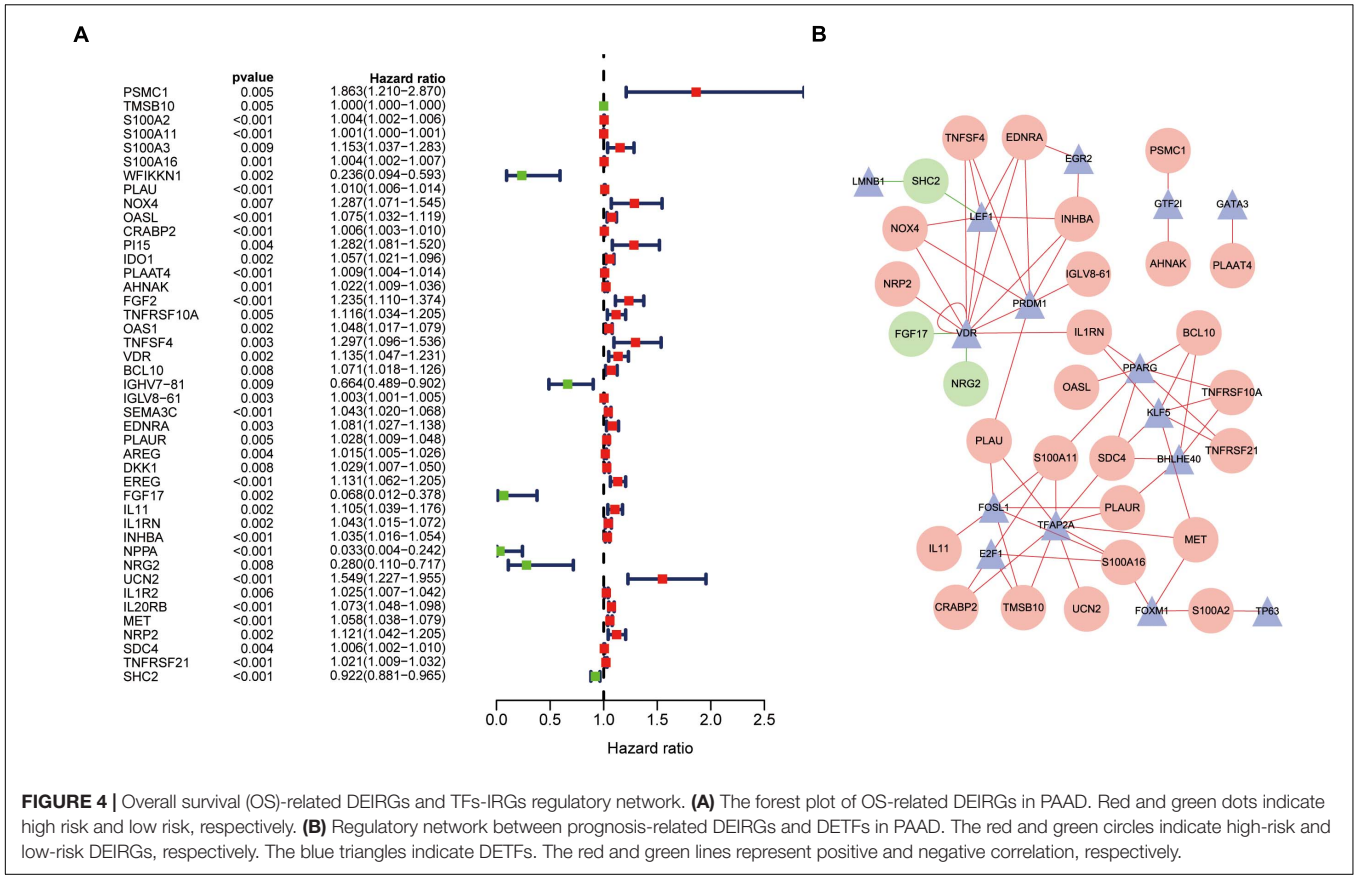


FIGURE 4 | Overall survival (OS)-related DEIRGs and TFs-IRGs regulatory network. **(A)** The forest plot of OS-related DEIRGs in PAAD. Red and green dots indicate high risk and low risk, respectively. **(B)** Regulatory network between prognosis-related DEIRGs and DETFs in PAAD. The red and green circles indicate high-risk and low-risk DEIRGs, respectively. The blue triangles indicate DETFs. The red and green lines represent positive and negative correlation, respectively.

score remained an independent prognostic factor correlated with OS (HR = 4.868, 95% CI = 2.899–8.175, $P < 0.001$; **Figure 7B**). Moreover, univariate and multivariate independent prognostic analyses (**Figures 7A,B**) showed that the residual tumor and outcomes were also significant independent prognostic factors for survival ($P < 0.05$; **Table 3**).

Eight-IRG Prognostic Model and Clinical Characteristics

Relationships between eight IRGs in the risk-score model and clinical features (age, gender, pathological TNM stage, histology grade, residual tumor, and outcomes) were assessed via the beeswarm packages in R ($P < 0.05$; **Table 4**). The cutoff value was determined by the median of the expression of the selected genes. As observed from **Figure 8**, the median values in the age $\leq$ 65 group were higher than those in the age $>$ 65 group between mesenchymal epithelial transition factor (MET) expression and riskscore (**Figures 8A,B**). The median value of the SHC2 and interleukin 20 receptor subunit beta (IL20RB) expression in pathological stage I-II was higher than that in stage III-IV (**Figures 8C,D**). With regard to histology grade, the median value of MET expression and riskscore in grades 1 and 2 was lower than that in grade 3 and 4, and the trend was exactly the opposite for Fibroblast Growth Factor (FGF)17 (**Figures 8E–G**). The median values in T1–2 staging were lower than those in T3–4 staging among 2′–5′-oligoadenylate synthetase like (OASL) expression, MET expression, and riskscore (**Figures 8H–J**). Moreover, the median value of MET expression was lower in residual tumor R0 than that in R1 and 2 (**Figure 8K**). Additionally, the median values of SHC2, plasminogen activator, urokinase (PLAU) expression, MET expression, IL20RB expression, and riskscore were notably different in PAAD at outcomes CR relative to those at PR + PD + SD ($P < 0.05$, **Figures 8L–P**).

Immunocyte Infiltration

We wished to ascertain if the eight-IRG prognostic model reflected the status of the PAAD immune microenvironment precisely. Hence, correlation analysis was done to explore the relationship between prognostic IRGs and infiltration of immune cells (**Figure 9**). The number of DCs, neutrophils, and $CD8^+$ T cells was positively correlated with the risk-score prediction model ($P < 0.05$; **Figures 9C,D,F**) but the trend of $CD4^+$ T cells was opposite (**Figure 9B**).

External Verification of the Eight-IRG Prognostic Model

Out of the eight prognostic IRGs in our model, the expression of four IRGs was upregulated and that of the remaining four IRGs was downregulated in the TCGA–PAAD cohort. In addition, a GEO dataset (GSE71729) was used to externally verify the difference in expression of eight IRGs between tumor tissues

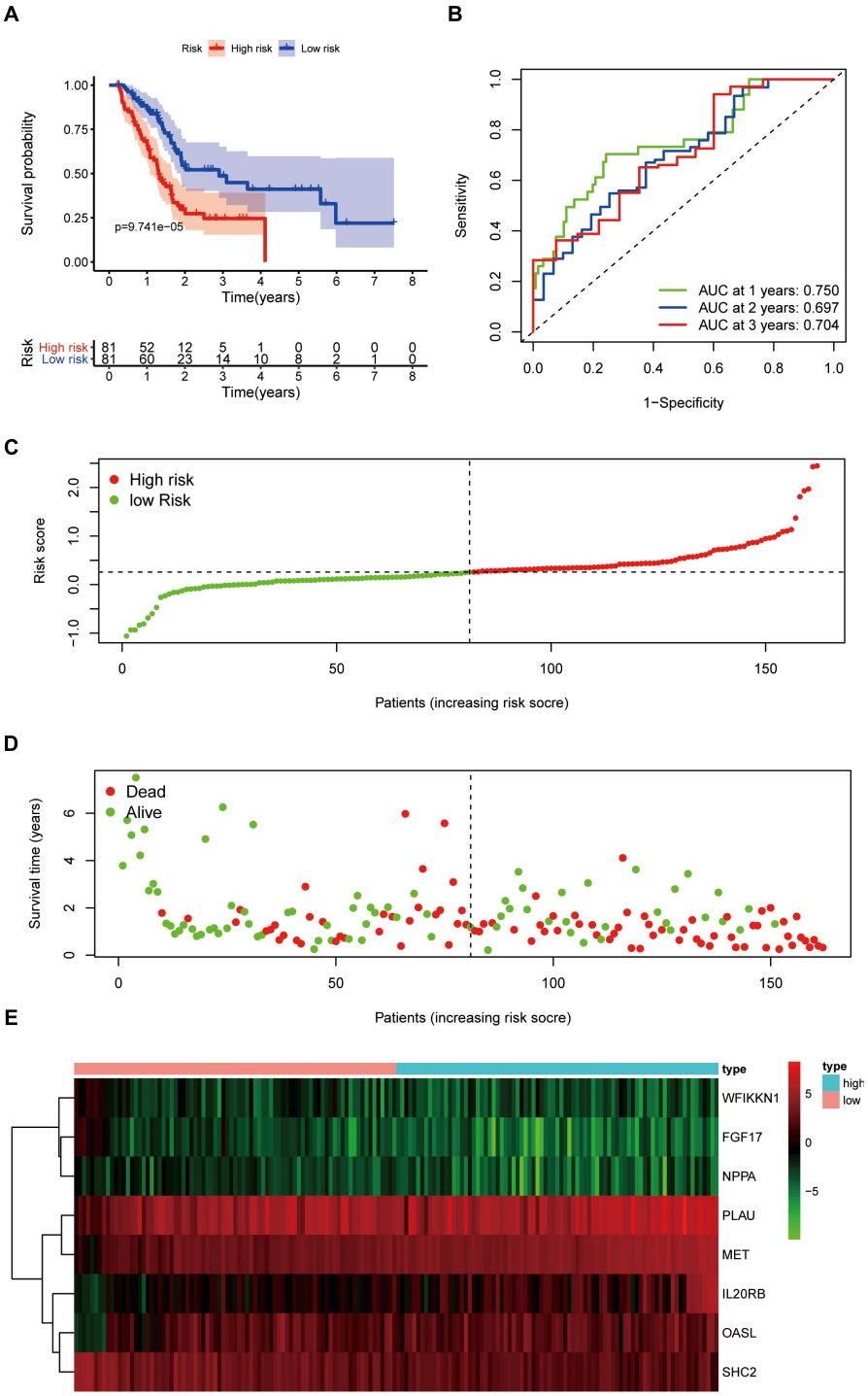


FIGURE 5 | Prognostic value of eight DEIRGs in PAAD patients. **(A)** Analyses of Kaplan–Meier curves for OS in PAAD patients using the signature of eight DEIRGs. **(B)** Receiver operating characteristic (ROC) curve suggesting the feasibility of our prognostic model. **(C)** Patients in high-risk (red dots) and low-risk (green dots) groups and the distribution of their corresponding risk score. **(D)** Patients in high-risk (red dots) and low-risk (green dots) groups, and their corresponding survival status. **(E)** Discrimination of expression of eight prognosis-related IRGs between high-risk and low-risk groups as revealed by a heatmap.

and normal tissues. As expected, the expression of IL20RB, MET, OASL, and PLAU in tumor tissues was significantly higher than that in normal tissues. FGF17, natriuretic peptide A (NPPA), SHC2, and WAP, follistatin/kazal, immunoglobulin, kunitz, and netrin domain containing 1 (WFIKKN1) was not expressed or at a minimal level in tumor tissues (**Figure 10A**). The

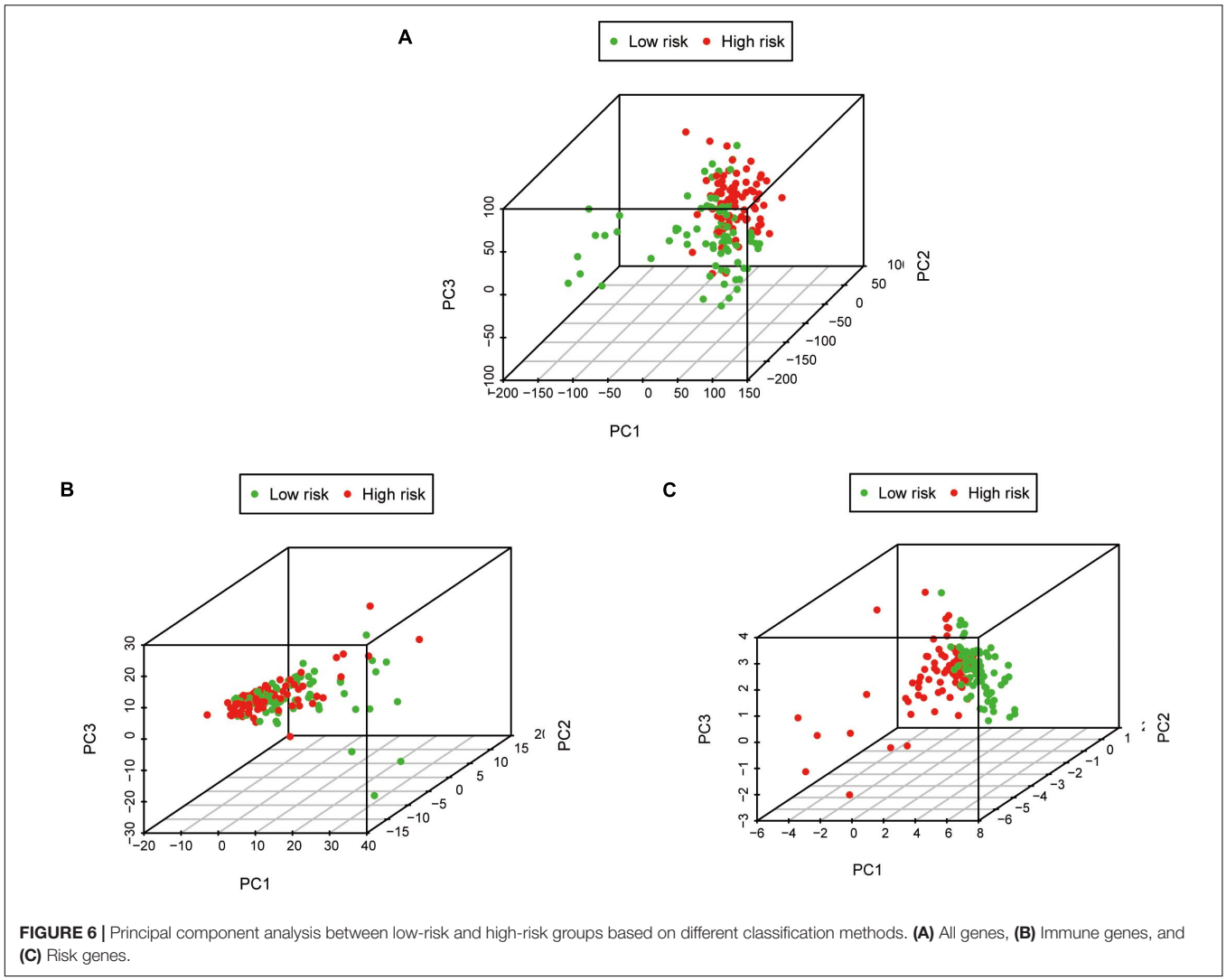


FIGURE 6 | Principal component analysis between low-risk and high-risk groups based on different classification methods. **(A)** All genes, **(B)** Immune genes, and **(C)** Risk genes.

protein distribution and expression of FGF17, MET, and SHC2 are displayed in **Figures 11A–F**, whereas the other five IRGs remained inaccessible in the Human Protein Atlas. Verification using the GEO database further confirmed that PAAD patients in the low-risk group showed a significant OS benefit compared with that in PAAD patients in the high-risk group. The Kaplan–Meier estimator effectively distinguished different groups of various risk ($P < 0.01$; **Figure 10B**). The predictive capacity of the signature was confirmed by analyses of the ROC curve. Our results showed that the prognostic signatures of the GEO dataset also performed well in forecasting 1-, 2-, and 3-years survival (**Figure 10C**).

DISCUSSION

Pancreatic cancer remains a lethal type of cancer due to its poor prognosis and lack of efficacious therapeutic approaches. Precise prediction of OS after contracting PAAD is very important for the choice of therapeutic method and improving the prognosis.

Pancreatic cancer lacks reliable and effective prognostic biomarkers related to the tumor immune microenvironment. An effective prediction model to accurately assess the prognosis of PAAD is long overdue. We intended to explore DEIRGs and establish a model of PAAD based on IRGs to uncover the biomarkers that predict the diagnosis and prognosis of PAAD.

In our study, 446 DEIRGs of PC were identified by comprehensive analyses. Analyses of pathway enrichment revealed that these DEIRGs correlated with the inflammatory response and typical tumor-related pathways shown in **Supplementary Table 4**. Most of them were related to the progression and treatment of PAAD. Cytokines and their correlated pathways may play a relevant part in PAAD progression and immune evasion (Padoan et al., 2019; Dey et al., 2020). As a vital component of the signaling between cancer cells and surrounding stromal cells, chemokine signaling participates in the development of the supportive TME of PAAD (Sleightholm et al., 2017). The Janus kinase family/signal transducer and activator of transcription (JAK/STAT) signaling

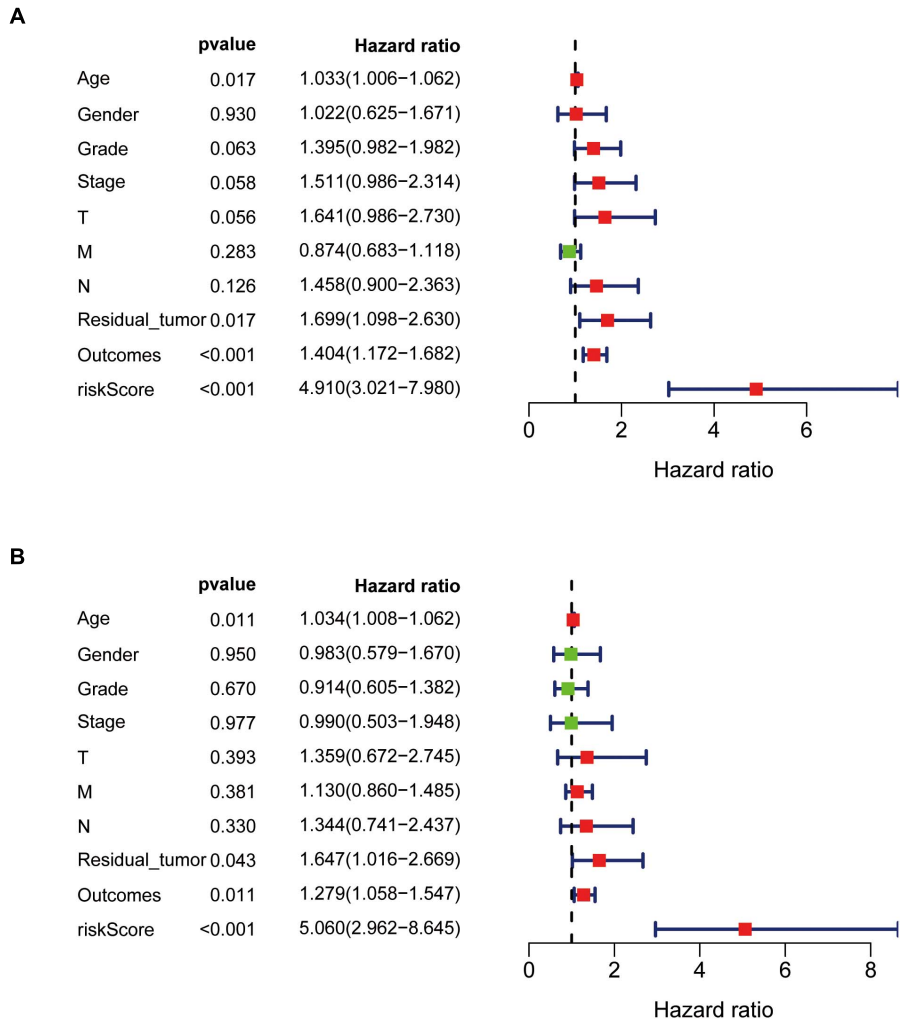


FIGURE 7 | Univariate and multivariate independent prognostic analysis in PAAD. **(A,B)** Forest plots of univariate and multivariate independent prognostic analysis.

pathway were central to tumor growth, tumor survival, and systemic inflammation, particularly in PC (Quintás-Cardama and Verstovsek, 2013; von Ahrens et al., 2017). In addition, two studies (Hurwitz et al., 2015; Beatty et al., 2019) showed that inhibitors of the JAK/STAT pathway may have clinical benefit. A follow-up study of a general population indicated that the high cytotoxic activity of natural killer (NK) cells is linked to a reduced risk of cancer (Imai et al., 2000). Lee et al. (2020) stated that the activity of NK cells decreased as cancer progressed, and that decreased activity of NK cells was associated with poor clinical outcomes. NF-κB is a pro-inflammatory signaling pathway in pancreatitis and PAAD. Increased basal levels and/or inducible levels of NF-κB activation are strongly linked to several aspects of treatment resistance, as well as the proliferation and metastasis of tumor cells in PAAD (Arlt et al., 2012; Kabacaoglu et al., 2019). In addition, NF-κB-mediated chemokine signaling plays a crucial part in the therapy resistance of PC (Geismann et al., 2019). Signaling by T-helper (Th1) and Th2 cytokines is complex in the microenvironment of pancreatic tumors (Andrianifahanana et al., 2006). The presence of tumor-infiltrating lymphocytes with high Th2:Th1 ratios demonstrates a poor prognosis in PAAD (Seicean et al., 2009). He et al. (2011) revealed that the accumulation of Th17 cells and their relevant cytokine levels in PC tissues may manifest engagement in the invasion and metastasis of PC, which may thereby have an impact on the prognosis. We conducted a comprehensive investigation of the biological functions of DEIRGs in PAAD populations to provide a basis for elucidating their possible molecular regulatory mechanisms.

More and more studies have found that abnormally expressed TFs in tumor tissues were related to aggressive diseases and poor prognosis. The research on new drugs that target specific TFs had great potential in developing clinically relevant strategies for the treatment of malignant tumors (Sankpal et al., 2012). To further investigate the possible molecular regulatory mechanisms, a TFs-mediated prognosis-related IRGs network was structured to find the significant TFs regulating DEIRGs in this network. DETFs such as basic helix-loop-helix family member E40 (BHLHE40),

E2F transcription factor 1 (E2F1), early growth response 2 (EGR2), FOS like 1, AP-1 transcription factor subunit (FOSL1), forkhead box M1 (FOXM1), kruppel like factor 5 (KLF5), LEF1, LMNB1, peroxisome proliferator activated receptor gamma (PPARG), PR/SET domain 1 (PRDM1), transcription factor AP-2 alpha (TFAP2A), tumor protein P63 (TP63), and VDR might regulate the DEIRGs in PAAD. BHLHE40 expression was upregulated by transforming growth factor (TGF)-β, and affected the morphology, migration, and invasion of PC cells by changing the expression of factors related to epithelial-to-mesenchymal (EMT) transition (Wu et al., 2012). E2F1-mediated overexpression of long non-coding (lnc)RNA-pancreatic cancer associated transcript 1 (PLACT1) promotes the growth of PAAD by continuously activating the NF-κB pathway and forming a positive feedback loop with IκBα in PC (Ren et al., 2020). Vallejo et al. (2017) revealed FOSL1 to be an oncogene in KRAS-driven lung cancer and PC, which partially factors through transcriptional regulation of a subset of genes involved in the mitotic machinery. Zhou et al. (2019) revealed an important epigenetic modification to FOXM1, and increased expression of FOXM1 suppressed the maturation of bone marrow−derived DCs via direct activation of Wnt5a signaling pathway and weakened the promotion of T−cell proliferation. He et al. (2018) demonstrated that KLF5 depletion in oncogenic Kras-expressing mouse PC cells reduced proliferation of tumor cells and PC progression. TP63 is a member of the p53 family and is transcribed from two promoters to produce two subtypes: TAp63 and ΔNp63 (Gonfloni et al., 2015). TP63 reprograms enhancers to drive squamous transdifferentiation in PC (Somerville et al., 2020). Sherman and collaborators discovered that the VDR is expressed in the stroma from PC cells and acts as a "master" transcriptional regulator of pancreatic stellate cells, thereby resulting in induced transcriptional reprogramming of tumor stroma in PAAD.

We innovatively established a TFs-mediated prognosis-related IRGs regulatory network in PAAD by bioinformatics analysis. This network showed that TFs regulated IRGs positively and negatively, which supplied a novel method to explore the IRGs underlying regulatory mechanisms in PAAD at the molecular level.

Eight IRGs involved in the prognostic model were considered to be potential biomarkers in PAAD. Among the eight genes,

TABLE 3 | Univariate and multivariate independent prognostic analyses.

Variable	Univariate analysis		Multivariate analysis	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Age	1.033 (1.006–1.062)	0.017	1.038 (1.012–1.065)	0.005
Sex	1.022 (0.625–1.671)	0.930	1.008 (0.592–1.715)	0.978
Grade	1.395 (0.982–1.982)	0.063	0.908 (0.601–1.371)	0.647
Stage	1.511 (0.986–2.314)	0.058	1.252 (0.788–1.989)	0.341
Residual tumor	1.699 (1.098–2.630)	0.017	1.678 (1.047–2.690)	0.032
Outcomes	1.404 (1.172–1.682)	<0.001	1.244 (1.030–1.503)	0.023
Risk score	4.910 (3.021–7.980)	<0.001	4.868 (2.899–8.175)	<0.001

TABLE 4 | Correlation between clinical features.

Gene	Age (≤65/>65)		Sex (male/female)		Grade (G1 and 2/G3 and 4)		Stage (I and II/III and IV)		Residual tumor (R0/R1 and 2)		Outcome (CR/PR and PD and SD)	
	t	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*	*t*	*P*
WFIKKN1	0.69	0.492	−0.963	0.337	1.713	0.09	0.893	0.41	−0.401	0.69	−0.142	0.887
PLAU	0.201	0.841	0.023	0.982	−1.287	0.204	0.56	0.592	−0.385	0.701	−2.356	0.02
OASL	−1.759	0.081	−1.355	0.178	−0.929	0.357	0.116	0.913	−1.946	0.056	−1.208	0.23
FGF17	1.194	0.236	−0.53	0.597	2.541	0.013	1.81	0.1	−0.688	0.494	1.472	0.145
NPPA	0.404	0.687	−1.054	0.294	1.31	0.194	1.382	0.226	0.463	0.644	1.933	0.056
IL20RB	0.51	0.611	−1.585	0.116	−1.715	0.094	1.982	0.067	−1.894	0.064	−2.584	0.012
MET	2.01	0.047	−0.561	0.576	−3.355	0.001	0.27	0.799	−2.783	0.007	−3.222	0.002
SHC2	0.586	0.559	0.304	0.761	1.598	0.114	3.112	0.012	0.163	0.871	2.184	0.031
Risk score	0.736	0.463	−0.662	0.51	−3.006	0.004	0.106	0.919	−1.966	0.054	−3.456	7.81×10^{-4}

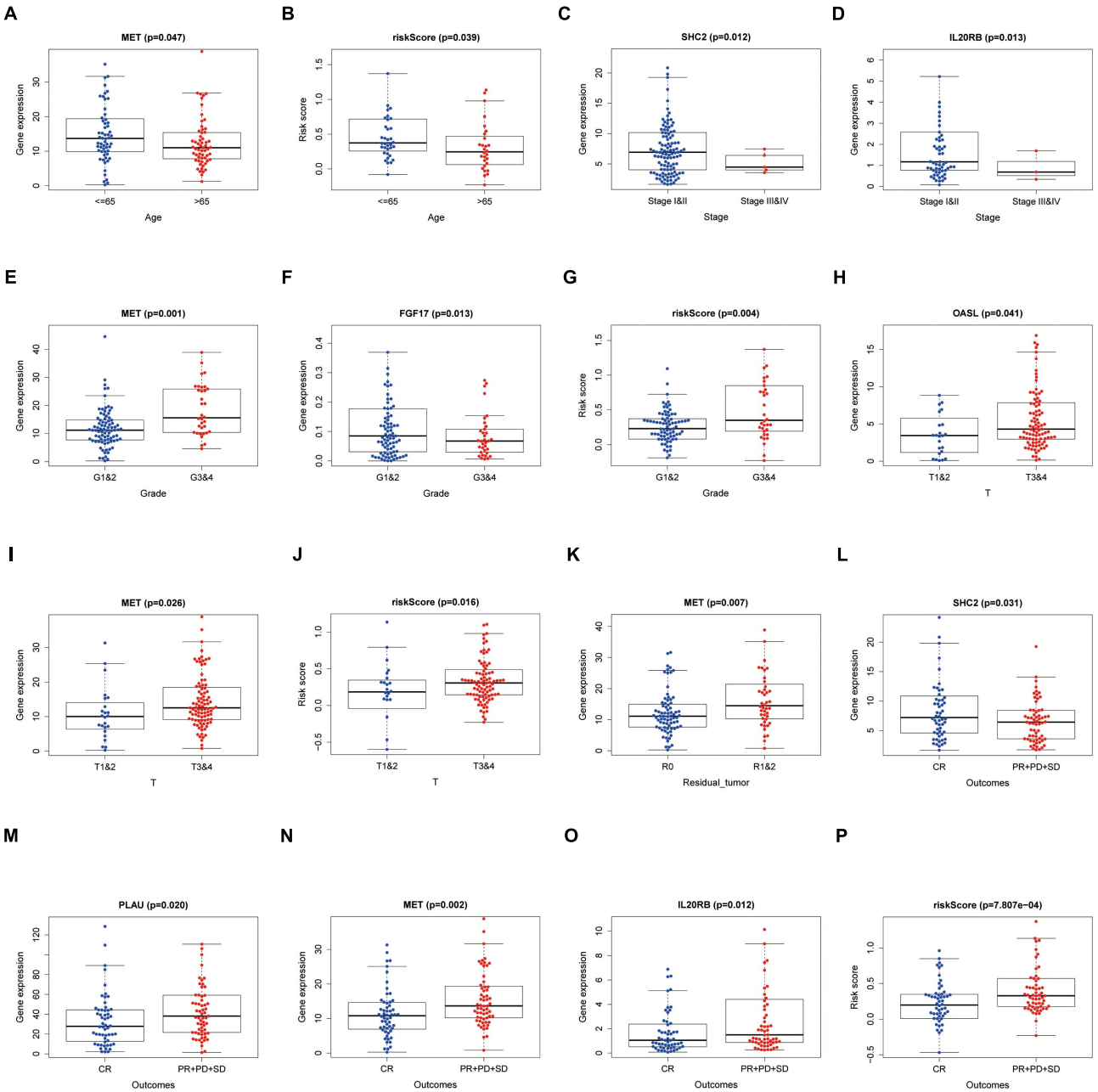


FIGURE 8 | Relationships between the clinical-pathological characteristics and expression of DEIRGs in PAAD. **(A)** Differences in expression of DEIRGs between the pathological TNM stages I and II/III and IV in PAAD. **(B,C)** Differences in expression of DEIRGs between the histology G1 and 2/G3 and 4 grades in PAAD. **(D,E)** Differences in expression of DEIRGs between the T stages T1 and 2/T3 and 4 in PAAD. **(F–K)** Differences in expression of DEIRGs between the M stages M0/M1/MX in PAAD. **(L)** Differences in expression of DEIRGs between the residual tumor R0/R1 and R2 in PAAD. **(M–P)** Differences in expression of DEIRGs between the outcomes (CR/PR + PD + SD) in PAAD.

MET, OASL, SHC2, and PLAU have been well studied in PAAD compared with other IRGs. Nan and coworkers found that hepatocyte growth factor (HGF) promotes the invasion and migration of PC cells by activating the HGF/c-Met pathway (Nan et al., 2019). Besides, MET/HGF co-targeting may represent a treatment option for patients with PC (Modica et al., 2018). As a member of the OAS protein family, OASL is associated with the innate immune defense against viral infections. Glaß et al. (2020) identified OASL to be a candidate oncogenic RNA-binding protein with partially validated target potential in PC. Recently, PLAU has been reported to be an oncogene that activates EMT progression in PAAD (Zhao X. et al., 2020). SHC2 was a proverbial adaptor molecule that binds to receptor tyrosine kinases via its SH2 domain.

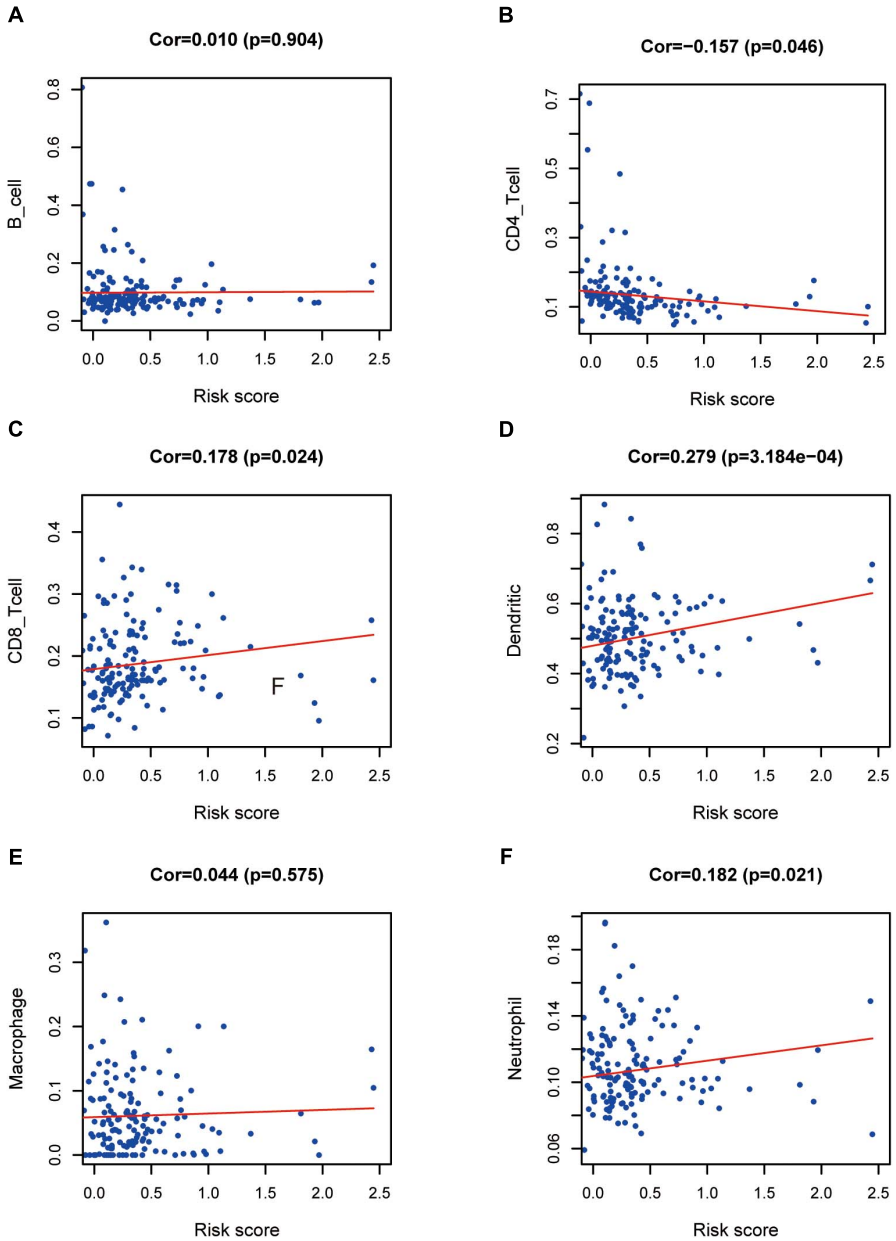


FIGURE 9 | Relationships between prognostic value and degree of infiltration of six types of immune cells. The relationship of the eight-IRG prognostic model with **(A)** B cells, **(B)** CD4 T cells, **(C)** CD8 T cells, **(D)** dendritic cells, **(E)** macrophages, and **(F)** neutrophils is revealed by scatter diagrams.

Teodorczyk and coworkers reported that CD95L could induce SCK recruitment and activation of the phosphoinositide 3-kinase/extracellular signal-regulated kinase (PI3K/ERK) pathway by stimulating CD95 receptors and, ultimately, lead to PC cell-cycle progression (Teodorczyk et al., 2015). One review stated that high expression of IL20RB was related to poor survival, thereby suggesting its oncogenic potential in PAAD (Haider et al., 2014). FGF17 was a member of the FGF8 subfamily, which promotes the development and progression of hepatocellular carcinoma (Gauglhofer et al., 2011). In addition, FGF17 was overexpressed in human prostate cancer, and involved in the progression of prostate cancer to high−grade disease (Heer et al., 2004). Both studies have reported that FGF17 may be a novel tumor-promoting gene whose expression is upregulated in neoplasms, data which contradicted our findings. The exact role of FGF17 in PAAD is not known. Few related studies have reported NPPA or WFIKKN1 being involved in PAAD.

Our TFs-IRGs-mediated network contained five of the eight modeling genes: PLAU, OASL, FGF17, MET, and SHC2. Their interactions with tumor-associated TFs can provide a certain theoretical direction/basis for mechanistic studies. Therefore,

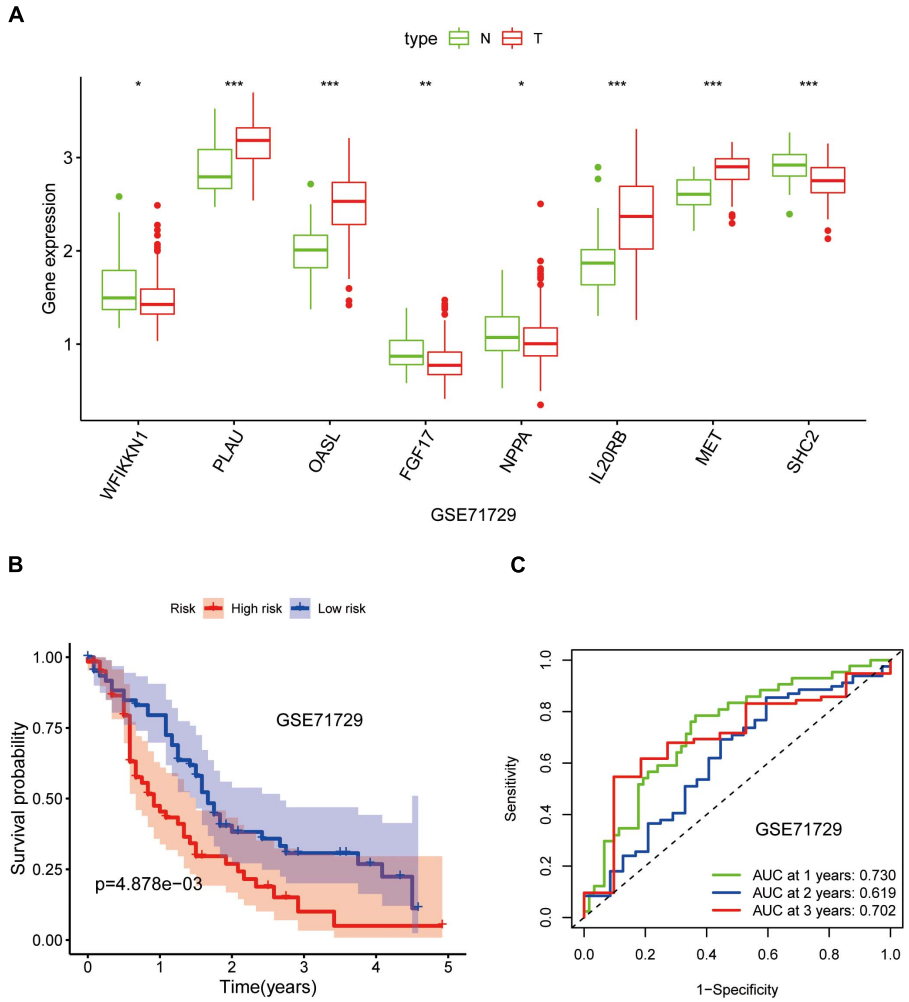


FIGURE 10 | Validation of the eight-gene signature in the Gene Expression Omnibus (GEO) database. **(A)** Expression of eight IRGs in tumor tissues and normal tissues in the GSE71729 database. **(B)** Kaplan–Meier curves for low- and high-risk groups in the GSE71729 database ($P < 0.01$). **(C)** ROC curve for predicting survival from PAAD based on the risk score of the GSE71729 database. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

further study of the potential regulatory mechanisms of these prognostic immune genes in PAAD is needed.

To clarify the immune microenvironment in PAAD, a correlation analysis on immunocytes was done based on the TIMER database. Results indicated that lower infiltration of DCs, $CD8^{+}$ T cells, and neutrophils may be observed in low-risk patients, whereas the tendency of $CD4^{+}$ T cells was the opposite. DCs, neutrophils, and $CD8^{+}$ T cells exhibited a significantly positive regulatory relationship with the prognostic model. Thus, our model may act as a predictive factor for increased infiltration of immune cells. One study reported that higher numbers of $CD4^{+}$ T lymphocytes were significantly associated with longer survival, which echoed our findings (Ino et al., 2013). A recent study showed that intratumoral infiltration by $CD8^{+}$ T lymphocytes and neutrophils and a favorable prognosis in PAAD patients were tightly linked (Miksch et al., 2019), which is the reverse of our results. Thus, our results must be validated by further investigations. Whether the infiltration level of DCs in tumors indicates the clinical prognosis of PAAD patients has not been reported. Studies on other tumors have yielded inconsistent or even conflicting results that doubt the value of infiltrating DCs (Karthaus et al., 2012). The exact role of immunocytes in PAAD has not been clarified. Considering the different levels of immunocyte infiltration between high-risk and low-risk PAAD groups, suitable immunotherapy strategies can be selected based on the basis of the immune microenvironment in PAAD.

So far, several studies have proposed that prognostic gene signatures based on mRNA levels can predict the OS of PC prognosis. For instance, Birnbaum et al. (2017) built a 25-gene classifier that helped select patients with resectable disease for immediate surgery or neoadjuvant chemotherapy. Another study established a four-gene signature for prediction of OS from PC based on gene-expression data from the GEO database [1-, 2-, and 3-years survival area under the curve (AUC) reached 0.715, 0.654, and 0.715, respectively] (Yan et al., 2019).

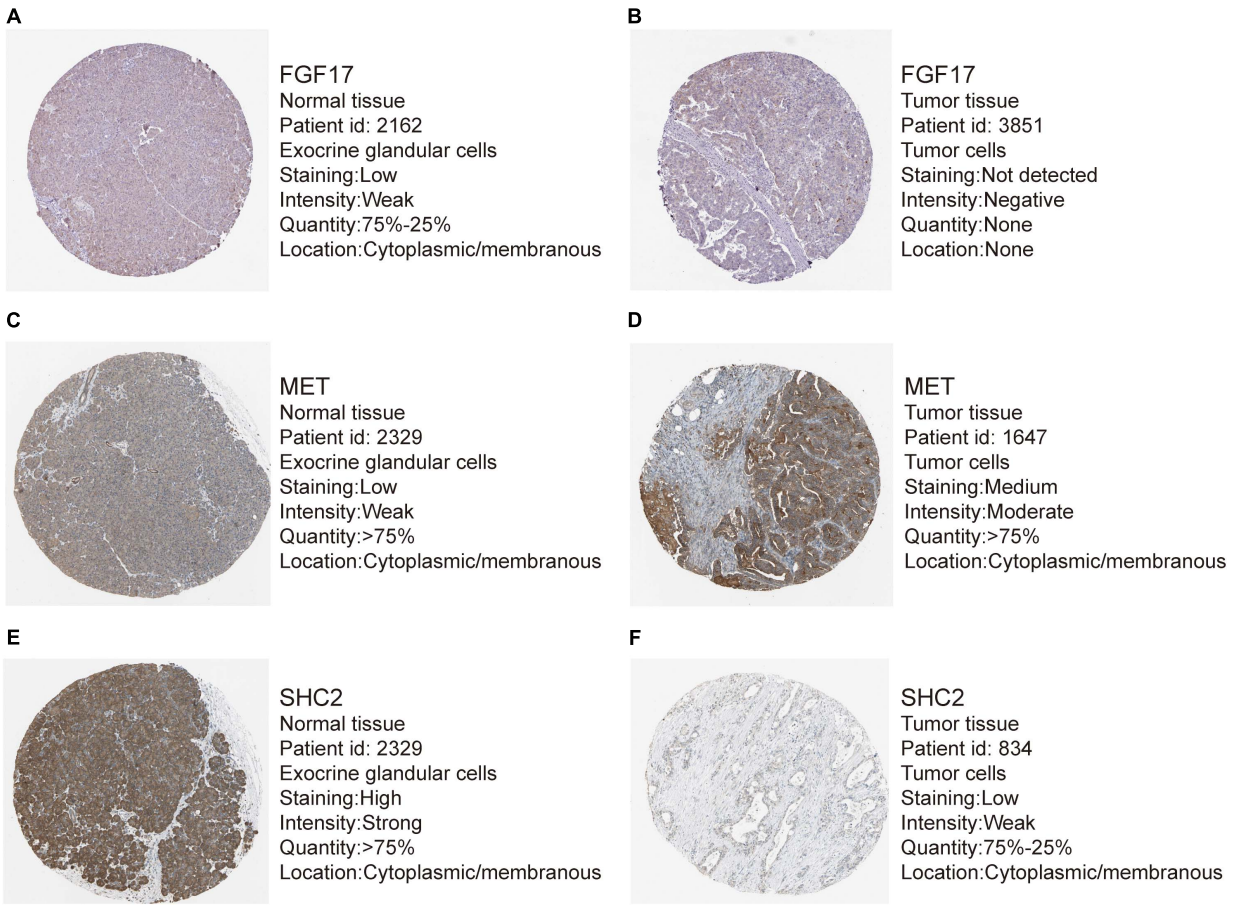


FIGURE 11 | Representative immunohistochemistry images for expression of FGF17, MET, and SHC2 in pancreatic cancer tissues and normal tissues were shown with the fraction of samples with antibody staining/protein expression level high, medium, low, or not detected. **(A–B)** Expression of FGF17 in PAAD tissues were lower than that in normal tissues. **(C–D)** Expression of MET in PAAD tissues were higher than that in normal tissues. **(E–F)** Expression of SHC2 in PAAD tissues were obviously lower than that in normal tissues.

A recent study investigated the survival-associated genes from the integrated analysis of multiple datasets, and established prognostic signatures in PAAD (1-, 2-, and 3-years survival AUC reached 0.699, 0.637, and 0.621, respectively; Wu et al., 2019). At present, there are few studies on the relationship between IRGs and the prognosis of PAAD. The latest research developed an immune prognostic model to identify low-risk patients who may benefit from immunotherapy (Gu et al., 2021). However, this predictive model still lacks an external cohort to verify the effectiveness of the model. We used a specialized immune database to explore the relationship between many IRGs and the prognosis of PAAD patients. Subsequently, we established a new immune-related prognostic signature. No overlap was found between the eight-gene signatures we developed and the one defined previously. Besides, the riskscore had a robust predictive performance with 1-, 2-, and 3-years survival AUC reached 0.750, 0.697, and 0.704, respectively. The predictive performance of our prognostic model was superior or comparable with that reported in other studies, and this model prediction is verified in an external validation cohort. These results suggest that an immune-related prognostic signature may be a valid marker for the prediction of the PC prognosis.

Nevertheless, our study still has perceived limitations. Firstly, we only used data from sections of public databases to build and validate our prediction model. Therefore, one must conduct more prospective studies to verify its clinical applicability. Secondly, we excluded many prominent prognostic genes in PAAD, so the potential weakness inherent in constructing a prognostic model with a single hallmark is inevitable. Moreover, the protein expression of IRGs related to the prognosis, and their potential molecular mechanisms in the pathogenesis and development of PAAD, must be confirmed by additional experimental studies.

CONCLUSION

We defined a novel eight-IRG model as an independent prognostic predictor for PAAD. The prognostic value of this model was verified by an external validation database. Moreover, the correlation between the eight-IRG prognostic model and

infiltrated immunocytes could demonstrate its pivotal role in the PAAD immune microenvironment, which could be utilized as a new prognostic and therapeutic biomarker in PAAD 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

MM, HL, and YL wrote the main manuscript text. MM and HL prepared figures and tables. YC, BX, and RZ reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Joint Funds for the Innovation of Science and Technology, Fujian province (2017Y9061).

ACKNOWLEDGMENTS

The authors would like to thank the TCGA, GEO, GTEx, ImmPort, Cistrome Cancer, HPA, and TIMER databases for the availability of the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.702102/full#supplementary-material

REFERENCES

Andrianifahanana, M., Chauhan, S. C., Choudhury, A., Moniaux, N., Brand, R. E., Sasson, A. A., et al. (2006). MUC4-expressing pancreatic adenocarcinomas show elevated levels of both T1 and T2 cytokines: potential pathobiologic implications. *Am. J. Gastroenterol.* 101, 2319–2329. doi: 10.1111/j.1572-0241.2006.00871.x

Arlt, A., Schäfer, H., and Kalthoff, H. (2012). The 'N-factors' in pancreatic cancer: functional relevance of NF-κB, NFAT and Nrf2 in pancreatic cancer. *Oncogenesis* 1:e35. doi: 10.1038/oncsis.2012.35

Beatty, G. L., Shahda, S., Beck, T., Uppal, N., Cohen, S. J., Donehower, R., et al. (2019). A Phase Ib/II Study of the JAK1 inhibitor, itacitinib, plus nab-paclitaxel and gemcitabine in advanced solid tumors. *Oncologist* 24, 14–e10. doi: 10.1634/theoncologist.2017-0665

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., et al. (2014). ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* 58, 234–239. doi: 10.1007/s12026-014-8516-1

Bianco, A., Perrotta, F., Barra, G., Malapelle, U., Rocco, D., and De Palma, R. (2019). Prognostic factors and biomarkers of responses to immune checkpoint inhibitors in lung cancer. *Int. J. Mol. Sci.* 20:4931. doi: 10.3390/ijms20194931

Birnbaum, D. J., Finetti, P., Lopresti, A., Gilabert, M., Poizat, F., Raoul, J. L., et al. (2017). A 25-gene classifier predicts overall survival in resectable pancreatic cancer. *BMC Med.* 15:170. doi: 10.1186/s12916-017-0936-z

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., et al. (2015). A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv. Biobank.* 13, 311–319. doi: 10.1089/bio.2015.0032

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:R60.

Dey, P., Li, J., Zhang, J., Chaurasiya, S., Strom, A., Wang, H., et al. (2020). Oncogenic KRAS-driven metabolic reprogramming in pancreatic cancer cells utilizes cytokines from the tumor microenvironment. *Cancer Discov.* 10, 608–625. doi: 10.1158/2159-8290.cd-19-0297

Erkan, M., Hausmann, S., Michalski, C. W., Fingerle, A. A., Dobritz, M., Kleeff, J., et al. (2012). The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. *Nat. Rev. Gastroenterol. Hepatol.* 9, 454–467. doi: 10.1038/nrgastro.2012.115

Ferlay, J., Partensky, C., and Bray, F. (2016). More deaths from pancreatic cancer than breast cancer in the EU by 2017. *Acta Oncol. (Stockholm, Sweden)* 55, 1158–1160. doi: 10.1080/0284186x.2016.1197419

Gauglhofer, C., Sagmeister, S., Schrottmaier, W., Fischer, C., Rodgarkia-Dara, C., Mohr, T., et al. (2011). Up-regulation of the fibroblast growth factor 8 subfamily in human hepatocellular carcinoma for cell survival and neoangiogenesis. *Hepatology (Baltimore, Md)* 53, 854–864. doi: 10.1002/hep.24099

Ge, P., Wang, W., Li, L., Zhang, G., Gao, Z., Tang, Z., et al. (2019). Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed. Pharmacother. = Biomed. Pharmacother.* 118:109228. doi: 10.1016/j.biopha.2019.109228

Geismann, C., Schäfer, H., Gundlach, J. P., Hauser, C., Egberts, J. H., Schneider, G., et al. (2019). NF-κB dependent chemokine signaling in pancreatic cancer. *Cancers* 11:1445. doi: 10.3390/cancers11101445

Glaß, M., Michl, P., and Hüttelmaier, A. S. (2020). RNA Binding proteins as drivers and therapeutic target candidates in pancreatic ductal adenocarcinoma. *Int. J. Mol. Sci.* 21:4190. doi: 10.3390/ijms21114190

Gonfloni, S., Caputo, V., and Iannizzotto, V. (2015). P63 in health and cancer. *Int. J. Dev. Biol.* 59, 87–93. doi: 10.1387/ijdb.150045sg

Gu, X., Zhang, Q., Wu, X., Fan, Y., and Qian, J. (2021). Gene coexpression network approach to develop an immune prognostic model for pancreatic adenocarcinoma. *World J. Surgical Oncol.* 19:112. doi: 10.1186/s12957-021-02201-w

Haider, S., Wang, J., Nagano, A., Desai, A., Arumugam, P., Dumartin, L., et al. (2014). A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. *Genome Med.* 6:105. doi: 10.1186/s13073-014-0105-3

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747

He, P., Yang, J. W., Yang, V. W., and Bialkowska, A. B. (2018). Krüppel-like Factor 5, increased in pancreatic ductal adenocarcinoma, promotes proliferation, acinar-to-ductal metaplasia, pancreatic intraepithelial neoplasia, and tumor growth in mice. *Gastroenterology* 154, 1494.e13–1508.e13. doi: 10.1053/j.gastro.2017.12.005

He, S., Fei, M., Wu, Y., Zheng, D., Wan, D., Wang, L., et al. (2011). Distribution and clinical significance of Th17 cells in the tumor microenvironment and peripheral blood of pancreatic cancer patients. *Int. J. Mol. Sci.* 12, 7424–7437. doi: 10.3390/ijms12117424

Heer, R., Douglas, D., Mathers, M. E., Robson, C. N., and Leung, H. Y. (2004). Fibroblast growth factor 17 is over-expressed in human prostate cancer. *J. Pathol.* 204, 578–586. doi: 10.1002/path.1668

Higuera, O., Ghanem, I., Nasimi, R., Prieto, I., Koren, L., and Feliu, J. (2016). Management of pancreatic cancer in the elderly. *World J. Gastroenterol.* 22, 764–775. doi: 10.3748/wjg.v22.i2.764

Huang, R., Mao, M., Lu, Y., Yu, Q., and Liao, L. (2020). A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. *Aging* 12, 6966–6980. doi: 10.18632/aging.103054

Hurwitz, H. I., Uppal, N., Wagner, S. A., Bendell, J. C., Beck, J. T., Wade, S. M. 3rd, et al. (2015). Randomized, double-blind, phase II study of ruxolitinib or placebo in combination with capecitabine in patients with metastatic pancreatic cancer for whom therapy with gemcitabine has failed. *J. Clin. Oncol. : Off. J. Am. Soc. Clin. Oncol.* 33, 4039–4047. doi: 10.1200/jco.2015.61.4578

Ilic, M., and Ilic, I. (2016). Epidemiology of pancreatic cancer. *World J. Gastroenterol.* 22, 9694–9705. doi: 10.3748/wjg.v22.i44.9694

Imai, K., Matsuyama, S., Miyake, S., Suga, K., and Nakachi, K. (2000). Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. *Lancet (London, England)* 356, 1795–1799. doi: 10.1016/s0140-6736(00)03231-1

Incio, J., Liu, H., Suboj, P., Chin, S. M., Chen, I. X., Pinter, M., et al. (2016). Obesity-induced inflammation and desmoplasia promote pancreatic cancer progression and resistance to chemotherapy. *Cancer Discov.* 6, 852–869. doi: 10.1158/2159-8290.cd-15-1177

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, 914–923. doi: 10.1038/bjc.2013.32

Jiang, Y., Xie, J., Huang, W., Chen, H., Xi, S., Han, Z., et al. (2019). Tumor immune microenvironment and chemosensitivity signature for predicting response to chemotherapy in gastric cancer. *Cancer Immunol. Res.* 7, 2065–2073. doi: 10.1158/2326-6066.cir-19-0311

Kabacaoglu, D., Ruess, D. A., Ai, J., and Algül, H. (2019). NF-κB/Rel transcription factors in pancreatic cancer: focusing on RelA, c-Rel, and RelB. *Cancers* 11:937. doi: 10.3390/cancers11070937

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Karthaus, N., Torensma, R., and Tel, J. (2012). Deciphering the message broadcast by tumor-infiltrating dendritic cells. *Am. J. Pathol.* 181, 733–742. doi: 10.1016/j.ajpath.2012.05.012

Kong, Y., Feng, Z. C., Zhang, Y. L., Liu, X. F., Ma, Y., Zhao, Z. M., et al. (2020). Identification of immune-related genes contributing to the development of glioblastoma using weighted gene co-expression network analysis. *Front. Immunol.* 11:1281. doi: 10.3389/fimmu.2020.01281

Lee, H. S., Leem, G., Kang, H., Jo, J. H., Chung, M. J., Jang, S. J., et al. (2020). Peripheral natural killer cell activity is associated with poor clinical outcomes in pancreatic ductal adenocarcinoma. *J. Gastroenterol. Hepatol.* 36, 516–522. doi: 10.1111/jgh.15265

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

Liang, C., Shi, S., Meng, Q., Liang, D., Ji, S., Zhang, B., et al. (2017). Complex roles of the stroma in the intrinsic resistance to gemcitabine in pancreatic cancer: where we are and where we are going. *Exp. Mol. Med.* 49:e406. doi: 10.1038/emm.2017.255

Liu, S. L., Bian, L. J., Liu, Z. X., Chen, Q. Y., Sun, X. S., Sun, R., et al. (2020). Development and validation of the immune signature to predict distant metastasis in patients with nasopharyngeal carcinoma. *J. Immunother. Cancer* 8:e000205. doi: 10.1136/jitc-2019-000205

Markowitz, J., Brooks, T. R., Duggan, M. C., Paul, B. K., Pan, X., Wei, L., et al. (2015). Patients with pancreatic adenocarcinoma exhibit elevated levels of myeloid-derived suppressor cells upon progression of disease. *Cancer Immunol. Immunother.: CII* 64, 149–159. doi: 10.1007/s00262-014-1618-8

Mei, S., Meyer, C. A., Zheng, R., Qin, Q., Wu, Q., Jiang, P., et al. (2017). Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res.* 77, e19–e22. doi: 10.1158/0008-5472.can-17-0327

Miksch, R. C., Schoenberg, M. B., Weniger, M., Bösch, F., Ormanns, S., Mayer, B., et al. (2019). Prognostic impact of tumor-infiltrating lymphocytes and neutrophils on survival of patients with upfront resection of pancreatic cancer. *Cancers* 11:39. doi: 10.3390/cancers11010039

Modica, C., Tortarolo, D., Comoglio, P. M., Basilico, C., and Vigna, E. M. E. T. (2018). /HGF Co-targeting in pancreatic cancer: a tool to provide insight into the tumor/stroma crosstalk. *Int. J. Mol. Sci.* 19:3920. doi: 10.3390/ijms19123920

Nan, L., Qin, T., Xiao, Y., Qian, W., Li, J., Wang, Z., et al. (2019). Pancreatic stellate cells facilitate perineural invasion of pancreatic cancer via HGF/c-Met pathway. *Cell Transplantation* 28, 1289–1298. doi: 10.1177/0963689719851772

Padoan, A., Plebani, M., and Basso, D. (2019). Inflammation and pancreatic cancer: focus on metabolism, cytokines, and immunity. *Int. J. Mol. Sci.* 20:676. doi: 10.3390/ijms20030676

Pontén, F., Schwenk, J. M., Asplund, A., and Edqvist, P. H. (2011). The Human Protein Atlas as a proteomic resource for biomarker discovery. *J. Internal Med.* 270, 428–446. doi: 10.1111/j.1365-2796.2011.02427.x

Quintás-Cardama, A., and Verstovsek, S. (2013). Molecular pathways: jak/STAT pathway: mutations, inhibitors, and resistance. *Clin. Cancer Res. : Off. J. Am. Assoc. Cancer Res.* 19, 1933–1940. doi: 10.1158/1078-0432.ccr-12-0284

Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., and Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 74, 2913–2921. doi: 10.1158/0008-5472.can-14-0155

Ren, B., Cui, M., Yang, G., Wang, H., Feng, M., You, L., et al. (2018). Tumor microenvironment participates in metastasis of pancreatic cancer. *Mol. Cancer* 17:108. doi: 10.1186/s12943-018-0858-1

Ren, X., Chen, C., Luo, Y., Liu, M., Li, Y., Zheng, S., et al. (2020). lncRNA-PLACT1 sustains activation of NF-κB pathway through a positive feedback loop with IκBα/E2F1 axis in pancreatic cancer. *Mol. Cancer* 19:35. doi: 10.1186/s12943-020-01153-1

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:e47. doi: 10.1093/nar/gkv007

Sankpal, U. T., Maliakal, P., Bose, D., Kayaleh, O., Buchholz, D., and Basha, R. (2012). Expression of specificity protein transcription factors in pancreatic cancer and their association in prognosis and therapy. *Curr. Med. Chem.* 19, 3779–3786. doi: 10.2174/092986712801661077

Seicean, A., Popa, D., Mocan, T., Cristea, V., and Berindan-Neagoe, I. (2009). Th1 and Th2 profiles in patients with pancreatic cancer compared with chronic pancreatitis. *Pancreas* 38, 594–595. doi: 10.1097/MPA.0b013e31819313d0

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39, 1–13. doi: 10.18637/jss.v039.i05

Sleightholm, R. L., Neilsen, B. K., Li, J., Steele, M. M., Singh, R. K., Hollingsworth, M. A., et al. (2017). Emerging roles of the CXCL12/CXCR4 axis in pancreatic cancer progression and therapy. *Pharmacol. Therapeutics* 179, 158–170. doi: 10.1016/j.pharmthera.2017.05.012

Somerville, T. D., Biffi, G., Daßler-Plenker, J., Hur, S. K., He, X. Y., Vance, K. E., et al. (2020). Squamous trans-differentiation of pancreatic cancer cells promotes stromal inflammation. *eLife* 9:e53381. doi: 10.7554/eLife.53381

Tang, Y., Xu, X., Guo, S., Zhang, C., Tang, Y., Tian, Y., et al. (2014). An increased abundance of tumor-infiltrating regulatory T cells is correlated with the progression and prognosis of pancreatic ductal adenocarcinoma. *PloS one* 9:e91551. doi: 10.1371/journal.pone.0091551

Taube, J. M., Galon, J., Sholl, L. M., Rodig, S. J., Cottrell, T. R., Giraldo, N. A., et al. (2018). Implications of the tumor immune microenvironment for staging and therapeutics. *Modern Pathol. : Off. J. U.S. Can. Acad. Pathol., Inc* 31, 214–234. doi: 10.1038/modpathol.2017.156

Teodorczyk, M., Kleber, S., Wollny, D., Sefrin, J. P., Aykut, B., Mateos, A., et al. (2015). CD95 promotes metastatic spread via Sck in pancreatic ductal adenocarcinoma. *Cell Death Differ.* 22, 1192–1202. doi: 10.1038/cdd.2014.217

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395. doi: 10.1002/(sici)1097-0258(19970228)16:4

Vallejo, A., Perurena, N., Guruceaga, E., Mazur, P. K., Martinez-Canarias, S., Zandueta, C., et al. (2017). An integrative approach unveils FOSL1 as an oncogene vulnerability in KRAS-driven lung and pancreatic cancer. *Nat. Commun.* 8:14294. doi: 10.1038/ncomms14294

von Ahrens, D., Bhagat, T. D., Nagrath, D., Maitra, A., and Verma, A. (2017). The role of stromal cancer-associated fibroblasts in pancreatic cancer. *J. Hematol. Oncol.* 10:76. doi: 10.1186/s13045-017-0448-5

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics (Oxford, England)* 31, 2912–2914. doi: 10.1093/bioinformatics/btv300

Wang, X., Lang, M., Zhao, T., Feng, X., Zheng, C., Huang, C., et al. (2017). Cancer-FOXP3 directly activated CCL5 to recruit FOXP3(+)Treg cells in pancreatic ductal adenocarcinoma. *Oncogene* 36, 3048–3058. doi: 10.1038/onc.2016.458

Whatcott, C. J., Diep, C. H., Jiang, P., Watanabe, A., LoBello, J., Sima, C., et al. (2015). Desmoplasia in primary tumors and metastatic lesions of pancreatic cancer. *Clin. Cancer Res. : Off. J. Am. Assoc. Cancer Res.* 21, 3561–3568. doi: 10.1158/1078-0432.ccr-14-1051

Wu, M., Li, X., Zhang, T., Liu, Z., and Zhao, Y. (2019). Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer. *Front. Oncol.* 9:996. doi: 10.3389/fonc.2019.00996

Wu, Y., Sato, F., Yamada, T., Bhawal, U. K., Kawamoto, T., Fujimoto, K., et al. (2012). The BHLH transcription factor DEC1 plays an important role in the epithelial-mesenchymal transition of pancreatic cancer. *Int. J. Oncol.* 41, 1337–1346. doi: 10.3892/ijo.2012.1559

Yan, X., Wan, H., Hao, X., Lan, T., Li, W., Xu, L., et al. (2019). Importance of gene expression signatures in pancreatic cancer prognosis and the establishment of a prediction model. *Cancer Manag. Res.* 11, 273–283. doi: 10.2147/cmar.s185205

Yang, S., Liu, T., Nan, H., Wang, Y., Chen, H., Zhang, X., et al. (2020). Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J. Cell. Physiol.* 235, 1025–1035. doi: 10.1002/jcp.29018

Zhao, B., Wang, Y., Wang, Y., Chen, W., Liu, P. H., Kong, Z., et al. (2020). Systematic identification, development, and validation of prognostic biomarkers involving the tumor-immune microenvironment for glioblastoma. *J. Cell. Physiol.* 236, 507–522. doi: 10.1002/jcp.29878

Zhao, X., Liu, Z., Ren, Z., Wang, H., Wang, Z., Zhai, J., et al. (2020). Triptolide inhibits pancreatic cancer cell proliferation and migration via down-regulating PLAU based on network pharmacology of Tripterygium wilfordii Hook F. *Eur. J. Pharmacol.* 880:173225. doi: 10.1016/j.ejphar.2020.173225

Zhou, Z., Chen, H., Xie, R., Wang, H., Li, S., Xu, Q., et al. (2019). Epigenetically modulated FOXM1 suppresses dendritic cell maturation in pancreatic cancer and colon cancer. *Mol. Oncol.* 13, 873–893. doi: 10.1002/1878-0261.12443

Zwing, N., Failmezger, H., Ooi, C. H., Hibar, D. P., Cañamero, M., Gomes, B., et al. (2020). Analysis of spatial organization of suppressive myeloid cells and effector T cells in colorectal cancer-a potential tool for discovering prognostic biomarkers in clinical research. *Front. Immunol.* 11:550250. doi: 10.3389/fimmu.2020.550250

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 Mao, Ling, Lin, Chen, Xu 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 Genetics

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fgene.2021.690509

Identification and Validation of a Ferroptosis-Related Long Non-coding RNA Signature for Predicting the Outcome of Lung Adenocarcinoma

*Zhiyuan Zheng†, Qian Zhang†, Wei Wu†, Yan Xue, Shuhan Liu, Qiaoqian Chen and Donghong Lin**

Medical Technology and Engineering College of Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Deli Liu,
Weill Cornell Medicine, United States
Ning Wang,
Arcus Biosciences, United States

***Correspondence:**
Donghong Lin
ldhfjmu@fjmu.edu.cn

†These authors have contributed equally to this work

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

Received: *03 April 2021*
Accepted: *21 June 2021*
Published: *22 July 2021*

Citation:
Zheng Z, Zhang Q, Wu W, Xue Y, Liu S, Chen Q and Lin D (2021) Identification and Validation of a Ferroptosis-Related Long Non-coding RNA Signature for Predicting the Outcome of Lung Adenocarcinoma. Front. Genet. 12:690509. doi: 10.3389/fgene.2021.690509

Background: Ferroptosis is a recently recognized type of programmed cell death that is involved in the biological processes of various cancers. However, the mechanism of ferroptosis in lung adenocarcinoma (LUAD) remains unclear. This study aimed to determine the role of ferroptosis-associated long non-coding RNAs (lncRNAs) in LUAD and to establish a prognostic model.

Methods: We downloaded ferroptosis-related genes from the FerrDb database and RNA sequencing data and clinicopathological characteristics from The Cancer Genome Atlas. We randomly divided the data into training and validation sets. Ferroptosis-associated lncRNA signatures with the lowest Akaike information criteria were determined using COX regression analysis and the least absolute shrinkage and selection operator. The risk scores of ferroptosis-related lncRNAs were calculated, and patients with LUAD were assigned to high- and low-risk groups based on the median risk score. The prognostic value of the risk scores was evaluated using Kaplan–Meier curves, Cox regression analyses, and nomograms. We then explored relationships between ferroptosis-related lncRNAs and the immune response using gene set enrichment analysis (GSEA).

Results: Ten ferroptosis-related lncRNA signatures were identified in the training group, and Kaplan–Meier and Cox regression analyses confirmed that the risk scores were independent predictors of LUAD outcome in the training and validation sets (all $P < 0.05$). The area under the curve confirmed that the signatures could determine the prognosis of LUAD. The predictive accuracy of the established nomogram model was verified using the concordance index and calibration curve. The GSEA showed that the 10 ferroptosis-related lncRNAs might be associated with tumor immune response.

Conclusion: We established a novel signature involving 10 ferroptosis-related lncRNAs (LINC01843, MIR193BHG, AC091185.1, AC027031.2, AL021707.2, AL031667.3, AL606834.1, AC026355.1, AC124045.1, and AC025048.4) that can accurately predict the outcome of LUAD and are associated with the immune response. This will provide new insights into the development of new therapies for LUAD.

Keywords: ferroptosis, lung adenocarcinoma, long non-coding RNA signature, prognosis, The Cancer Genome Atlas

INTRODUCTION

Lung cancer is a common type of malignancy worldwide (Sung et al., 2021) and is also the most lethal. It comprises small cell and non-small cell lung cancer types (Garon et al., 2021). Lung adenocarcinoma (LUAD) is the primary type of NSCLC; it is more frequent than squamous cell lung cancer, thus being the most common histological subtype of primary lung cancer (Lortet-Tieulent et al., 2014). Due to the lack of specific techniques that could diagnose cancer in the early stages (Goulart et al., 2021), more useful prognostic biomarkers and therapeutic targets are required.

Ferroptosis is a recently established type of programmed cell death that is iron dependent and characterized by the accumulation of intracellular reactive oxygen species (Dixon et al., 2012). It differs from apoptosis, necrosis, and autophagy in terms of morphology, biochemistry, and genetics (Chen X. et al., 2021). Ferroptosis plays a key role in killing tumor cells and inhibiting tumor growth (Jiang et al., 2021; Wang H. et al., 2021). Accumulating evidence indicates that ferroptosis is involved in the biological processes of LUAD (Wohlhieter et al., 2020; Li et al., 2021; Lou et al., 2021). However, the regulation of ferroptosis remains obscure, and it is far from being applied to cancer therapy. Therefore, identifying the key regulators of ferroptosis is a key step in broadening the options for treating cancer.

Long non-coding RNAs (lncRNAs) are transcripts of > 200 nucleotides that typically do not encode proteins (Kopp and Mendell, 2018). lncRNAs play important roles in many processes, such as epigenetic regulation, cell cycle regulation, and cell differentiation regulation by mediating transcriptional activation, interference, and chromosomal modification (Pang et al., 2019). In particular, the abnormal expression or function of lncRNAs might be associated with various diseases, including cancer. lncRNAs may affect ferroptosis; for example, the nuclear lncRNA LINC00618 accelerates ferroptosis in a manner dependent upon apoptosis in leukemia (Wang Z. et al., 2021). The lncRNA PVT1 regulates ferroptosis through miR-214-mediated TFR1 and p53 in patients with acute ischemic stroke (Lu et al., 2020). The lncRNAs GABPB1-AS1 and GABPB1 regulate oxidative stress during erastin-induced ferroptosis in HepG2 hepatocellular carcinoma cells (Qi et al., 2019). The lncRNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competitive endogenous RNA (Wang M. et al., 2019). The lncRNA MT1DP increases non-small cell lung cancer sensitivity to ferroptosis through regulating the miR-365a-3p/NRF2 axis (Meng et al., 2020). Therefore, ferroptosis-related lncRNAs involved in predicting the outcome of LUAD should be determined to provide a theoretical basis for novel strategies to treat patients with LUAD.

Previous studies have shown that one mechanism for non-coding RNA to participate in gene regulatory networks is through the direct binding of lncRNA to other RNA molecules to regulate their stability or translation. These interactions rely on the binding of lncRNA to its RNA target and either creating a substrate for protein function or inhibitory protein effectors (Anastasiadou et al., 2018). Therefore, we use the limma package in R, one of the high-throughput analysis methods, to conduct a comprehensive analysis of the genome of LUAD (Ge et al., 2020). Through the co-expression analysis, the least absolute shrinkage and selection operator (lasso) regression, and COX regression (Wang et al., 2013; Hu et al., 2014), we identified ferroptosis-related lncRNAs that are associated with the prognosis of LUAD, constructed a nomogram with which to determine the prognosis of LUAD, and herein discuss the relevance of the model to predicting the immune response. Our findings should help to improve the early diagnosis rate of LUAD and provide a theoretical basis for precise, individualized treatment.

MATERIALS AND METHODS

Acquisition of Gene Expression and Clinical Data

Gene expression profiles and clinical data from patients with LUAD were downloaded from The Cancer Genome Atlas (TCGA). mRNAs and lncRNAs were encoded according to GENCODE Release 29 (GRCh38.p12). We collected clinical data including sex, age, and TNM, pT, pN, and pM stages as well as follow-up data. We excluded patients with incomplete clinicopathological characteristics. Considering the possibility of non-cancer-death, we also excluded patients who survived ≤ 30 days. We finally included data from 443 patients and randomly assigned them to training ($n = 223$) and validation ($n = 220$) sets using the caret package in R.

Ferroptosis-Associated lncRNAs

We obtained 259 ferroptosis-associated gene sets from FerrDb, the first manually curated resource for regulators and markers of ferroptosis that was released in December 2019 (Zhou and Bao, 2020). Ferroptosis-associated lncRNAs detected using the limma package in R were identified by correlation analysis between ferroptosis-related genes and lncRNA expression levels in the LUAD samples. We determined the Pearson correlation

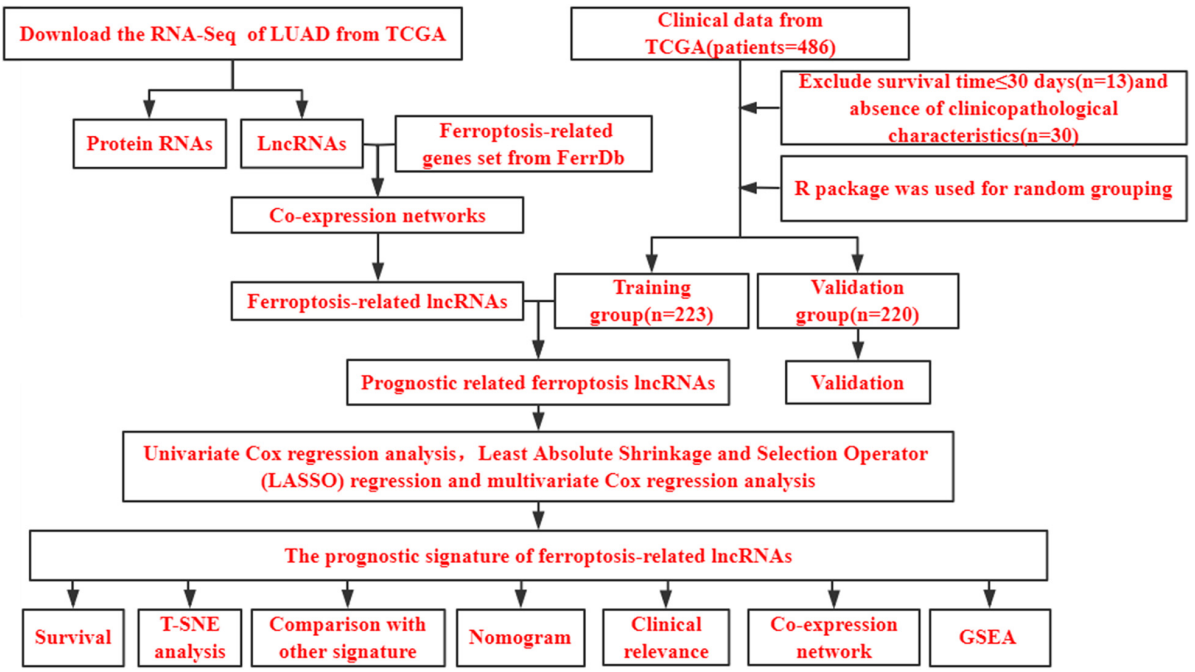


FIGURE 1 | Study flow chart. LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas database; lncRNAs, long non-coding RNAs; T-SNE, t-distributed stochastic neighbor embedding; GSEA, gene set enrichment analysis.

coefficients, and the threshold was established as a correlation coefficient > 0.4 and $P < 0.001$.

Establishment of Prognostic Ferroptosis-Associated lncRNA Signatures

We screened lncRNAs associated with the overall survival (OS) of patients with LUAD using univariate Cox regression analysis. Further screening was based on lasso regression to avoid overfitting, and we adjusted the L1 penalty parameter through 10-fold cross-validation to reduce the number of genes. Genes with a repetition frequency > 900 in 1,000 replicates were considered to be more closely associated with OS. We established prognostic lncRNA signatures using multivariate Cox regression and then constructed a lncRNA signature with the lowest Akaike information criterion (AIC) (Susko and Roger, 2020). The risk score for each patient with LUAD was calculated based on the amount of lncRNA expression and the corresponding coefficient as:

$$\beta\text{lncRNA1} \times \text{ExpressionlncRNA1} + \beta\text{lncRNA2} \times \text{ExpressionlncRNA2} + \ldots + \beta\text{lncRNAn} \times \text{ExpressionlncRNAn}$$

Evaluation of Risk Score Prediction Ability and Nomogram Construction

Patients in the training and validation sets were assigned to high- and low-risk groups using the median risk score of the training group. Kaplan–Meier survival curves were constructed, and predictive performance was evaluated using receiver operating characteristics (ROC) curves. We conducted t-distributed stochastic neighbor embedding (t-SNE) using the Rtsne package in R to reduce the dimensions and visualize ferroptosis status based on the high- and low-risk groups. We applied independent prognostic factors determined *via* multivariate Cox regression to construct a prognostic nomogram using the rms package in R. The accuracy of the nomogram was verified using the Harrell concordance (C) index and a calibration curve.

Construction of a Ferroptosis-Related lncRNA–mRNA Co-expression Network

Correlations between co-expressed lncRNAs and mRNAs associated with ferroptosis were analyzed using a co-expression network and a Sankey diagram visualized using Cytoscape and the ggalluvial package in R.

Gene Set Enrichment Analysis

Potential immune responses between the high- and low-risk patients were investigated using gene set enrichment analysis (GSEA) (Subramanian et al., 2005), and $P < 0.05$, with a false discovery rate (FDR q) < 0.05, was considered statistically significant.

Statistical Analysis

Clinicopathological characteristics were compared between the training and validation sets using chi-square tests, and

relationships between clinicopathological characteristics and lncRNA expression were compared using Mann–Whitney *U*-tests. All data were statistically analyzed using R (v. 4.0.1), and statistical significance was set at $P < 0.05$.

RESULTS

Acquisition of lncRNA Associated With Ferroptosis

Figure 1 shows a flow diagram of the study. **Table 1** shows that the clinicopathological characteristics of the 443 patients with LUAD did not significantly differ between the training and validation sets. We downloaded 259 ferroptosis-related genes from FerrDb (**Supplementary Table 1**). To obtain high-quality lncRNA, we excluded those with a low expression level using a cutoff criterion for the average expression in all samples > 0.5. We then used the limma package in R to analyze the correlation between the expression levels of ferroptosis-related genes and lncRNA in LUAD samples and set the threshold conditions to a correlation coefficient > 0.4 and $P < 0.001$, and finally 1,138 ferroptosis-associated lncRNAs were obtained.

Development and Validation of Prognostic Ferroptosis-Related lncRNA Signature

We constructed a ferroptosis-associated lncRNA signature based on the training group. We obtained 29 ferroptosis-associated lncRNAs associated with OS in a preliminary screen using univariate Cox analysis ($P < 0.05$; **Figure 2A**). We then analyzed these genes using lasso regression with 10-fold cross-validation and screened 19 ferroptosis-associated lncRNAs with a repetition rate > 900 in 1,000 replicates (**Figures 2B,C**). Thereafter, we conducted multivariate Cox regression analysis and constructed a lncRNA signature with the lowest AIC value. Therefore, we generated 10 ferroptosis-associated lncRNA prognostic signatures (**Table 2**). The median risk score classified the patients into high-risk ($n = 111$) or low-risk ($n = 112$) groups and was calculated as follows:

0.0717 × Expression LINC01843 + 0.3744 × Expression MIR193BHG–0.2992 × Expression AC026355.1–0.8269 × Expression AC124045.1 + 0.8407 × Expression AC091185.1 + 0.1046 × Expression AC027031.2 + 0.1700 × Expression AL021707.2 + 0.1748 × Expression AL031667.3 + 0.3496 × Expression AL606834.1–0.6308 × Expression AC025048.4.

Kaplan–Meier curves showed that patients with high-risk scores in the training set had a significantly higher probability of death ($P < 0.001$; **Figure 3A**). As risk scores increased, the risk of death increased, and survival duration decreased (**Figure 3B**). The risk heat map shows the expression of lncRNAs between the high- and low-risk groups (**Figure 3C**). We further used the Mann–Whitney *U*-test to explore the difference in lncRNA expression between the high- and low-risk groups (**Supplementary Figure 1A**). We then confirmed these results in the validation set, in which OS also significantly differed between the high- and low-risk groups ($P < 0.001$; **Figures 3D,E**). We also used the risk heat map to show the expression of lncRNA between the high- and low-risk groups in the validation set (**Figure 3F**) and used the Mann–Whitney *U*-test to explore the expression differences between them (**Supplementary Figure 1B**). The results showed that the expression of lncRNA including AC025048.4, AC026355.1, AC124045.1, AL031667.3, AL606834.1, and MIR193BHG was significantly different in the training and validation groups. We explored the distribution of the high- and low-risk groups using t-SNE (**Figures 4A,B**) and intuitively perceived that patients with LUAD can be better differentiated based on prognosis indicated by lncRNAs associated with ferroptosis.

TABLE 1 | Characteristics of patients with LUAD.

Characteristics	Patients (*N*, %)	(*N*, %) Training group (*n* = 223)	Validation group (*n* = 220)	*P*
Sex				0.823
Female	244 (55.1)	124 (55.6)	120 (54.5)	
Male	199 (44.9)	99 (44.4)	100 (45.5)	
Age, years				0.190
≤ 60	148 (33.4)	81 (36.3)	67 (30.5)	
>60	295 (66.6)	142 (63.7)	153 (69.5)	
TNM stage				0.867
I	239 (54.0)	120 (53.8)	119 (54.1)	
II	107 (24.1)	56 (25.1)	51 (23.2)	
III	73 (16.5)	34 (15.3)	39 (17.7)	
IV	24 (5.4)	13 (5.8)	11 (5.0)	
pT stage				0.504
T1	152 (34.3)	82 (36.8)	70 (31.8)	
T2	234 (52.8)	110 (49.3)	124 (56.4)	
T3	38 (8.6)	20 (9.0)	18 (8.2)	
T4	19 (4.3)	11 (4.9)	8 (3.6)	
pN stage				0.570
N0	291 (65.7)	145 (65.0)	146 (66.4)	
N1	87 (19.6)	44 (19.7)	43 (19.5)	
N2	63 (14.2)	32 (14.4)	31 (14.1)	
N3	2 (0.5)	2 (0.9)	0 (0)	
pM stage				0.700
M0	419 (94.6)	210 (94.2)	209 (95.0)	
M1	24 (5.4)	13 (5.8)	11 (5.0)	
Follow-up state				0.111
Alive	290 (65.5)	138 (61.9)	152 (69.1)	
Dead	153 (34.5)	85 (38.1)	68 (30.9)	

LUAD, lung adenocarcinoma; TNM, tumor node metastasis.

Independent Prognostic Analysis of OS

We assessed whether clinical characteristics (sex, age, TNM stage, pT, pN, and pM) and risk scores were independent prognostic factors for OS using univariate and multivariate Cox regression analyses. The results showed that TNM stage and risk scores were independent predictors of OS in the training and validation sets (**Figures 5A–D**). We further compared the

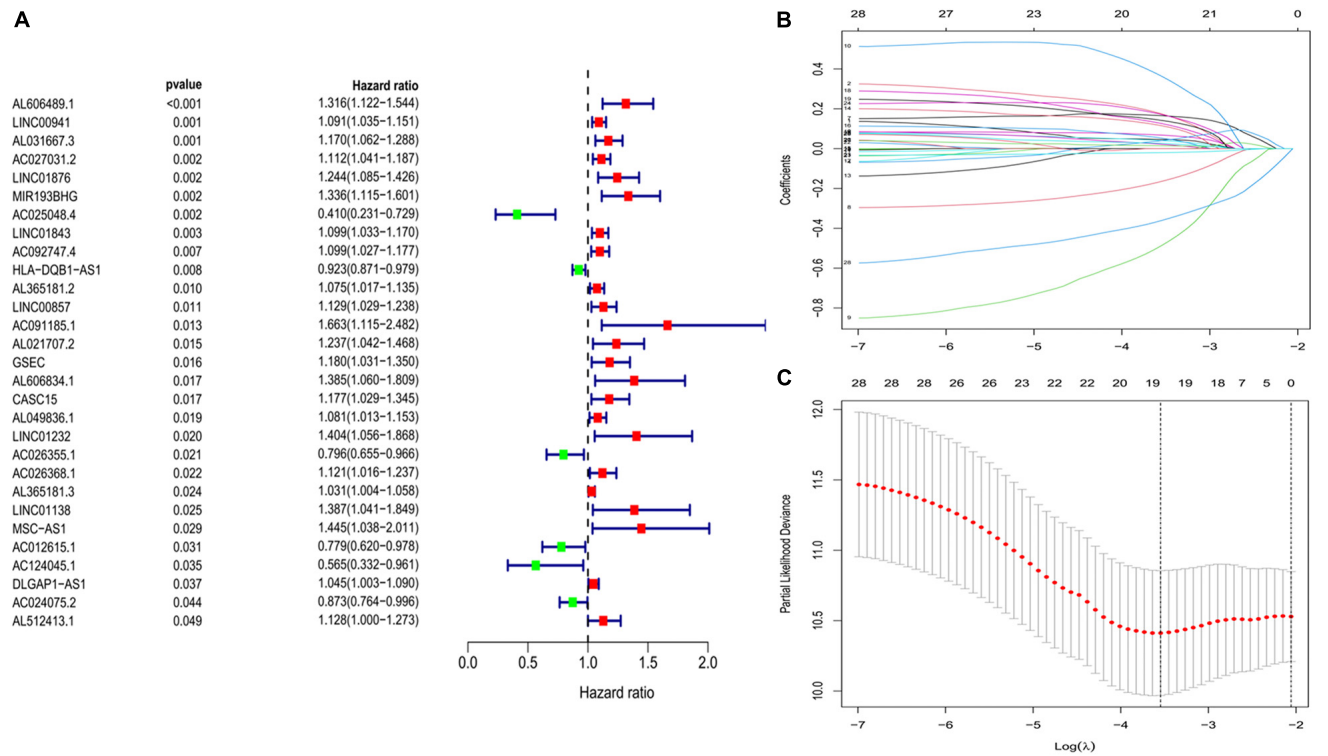


FIGURE 2 | Development of prognostic ferroptosis-associated lncRNA signature. **(A)** Forest plot of univariate Cox regression identified 29 ferroptosis-related lncRNAs associated with overall survival. **(B,C)** Lasso Cox regression to identify lncRNAs associated with ferroptosis that have implications for the prognosis of LUAD. lncRNAs, long non-coding RNAs; LASSO, least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma.

diagnostic efficacy of risk scores and other baseline approaches for patients with LUAD by drawing the ROC and calculating the area under the curve (AUC). The AUC values for risk scores in the training and validation groups were 0.741 and 0.771, respectively (**Figures 5E,F**), which were higher than the other indicators in both groups and further confirm that ferroptosis-associated lncRNA signatures have a higher diagnostic power than other indicators in predicting the prognosis of patients with LUAD.

Construction and Validation of the Predictive Nomogram in LUAD

We established a nomogram to predict the OS of patients with LUAD based on independent predictive factors derived from a multivariate Cox risk regression model (**Figure 6A**). The C-index of the nomogram was 0.763 [95% confidence interval (CI), 0.712–0.814)] and 0.736 (95% CI, 0.669–0.803) in the training and validation sets, respectively. The prediction model calibration curve also revealed consistent predicted and actual survival rates in the training and validation sets (**Figures 6B–G**).

Correlation Between Clinicopathological Characteristics and the Expression of 10 Ferroptosis-Associated lncRNAs

We further explored the relationship between clinicopathological characteristics and found that the expression of lncRNA AL031667.3 increased with age (**Figure 7A**), AC027031.2 was abundantly expressed in female patients (**Figure 7B**), the expression of AC091185.1 and AC124045.1 was associated with TNM stage (**Figure 7C**), that of AC091185.1, AC124045.1, AL021707.2, and LINC01843 was associated with pT stage (**Figure 7D**), and that of AC124045.1, AL021707.2, AL031667.3, and MIR193BHG was associated with pN stage (**Figure 7E**). Patients with decreased AC124045.1 expression were more likely to have distant metastases (**Figure 7F**).

TABLE 2 | Ten ferroptosis-associated lncRNAs determined *via* multivariate Cox regression analysis.

ID	Coef	HR	HR.95L	HR.95H	*P*-value
LINC01843	0.0717	1.0744	1.0024	1.1515	0.0426935115489
MIR193BHG	0.3744	1.4542	1.2138	1.7422	0.000048814974
AC026355.1	−0.2992	0.7414	0.6225	0.8831	0.000798318265
AC124045.1	−0.8269	0.4374	0.2313	0.8270	0.010950731667
AC091185.1	0.8407	2.3180	1.4739	3.6457	0.000273931536
AC027031.2	0.1046	1.1103	1.0416	1.1834	0.001327246427
AL021707.2	0.1700	1.1854	0.9948	1.4124	0.057252889138
AL031667.3	0.1748	1.1909	1.0791	1.3144	0.000512851317
AL606834.1	0.3496	1.4185	1.0584	1.9012	0.019277823214
AC025048.4	−0.6308	0.5321	0.2964	0.9554	0.034615991762

Coef, coefficient; HR, hazard ratio.

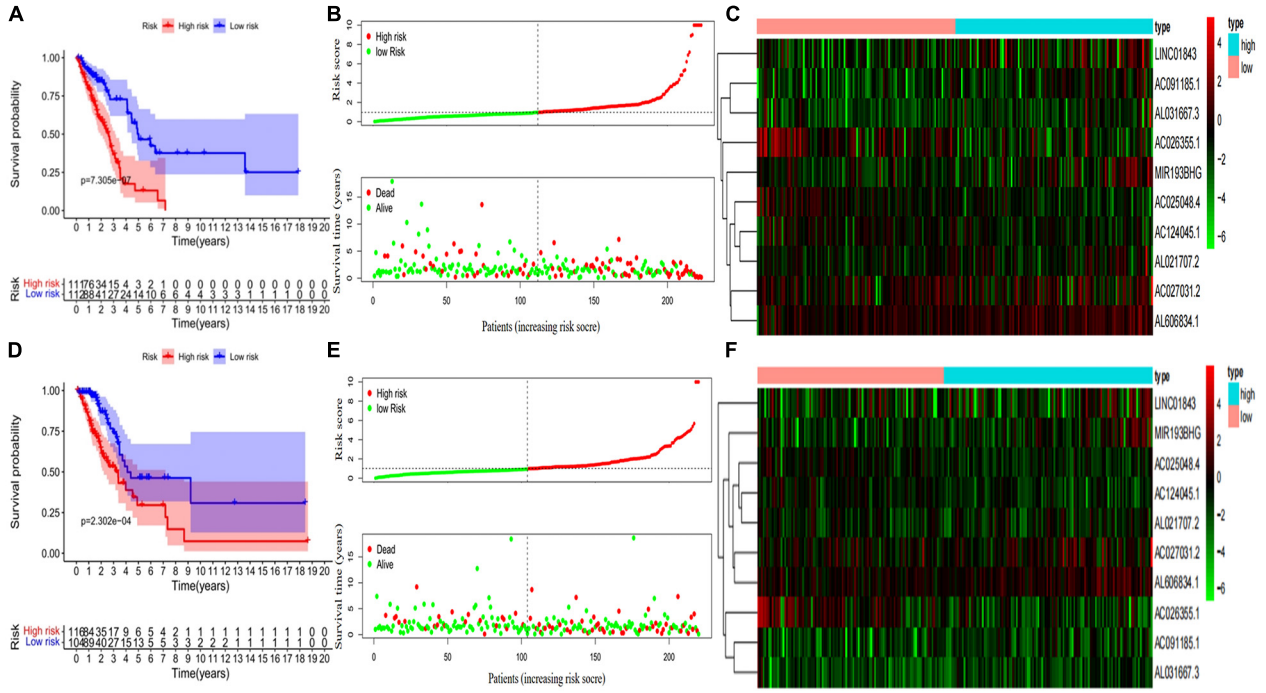


FIGURE 3 | Development and validation of prognostic ferroptosis-associated long non-coding RNA signature. **(A)** Kaplan–Meier curve, **(B)** risk score and survival status, and **(C)** heat map for the training group. **(D)** Kaplan–Meier curve, **(E)** risk score and survival status, and **(F)** heat map for the validation group.

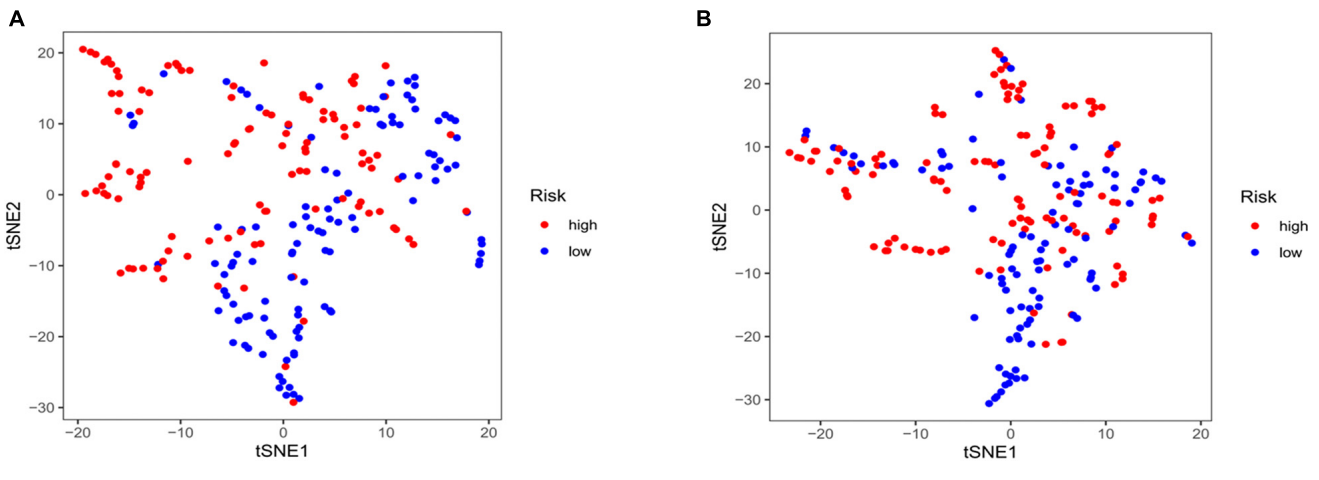


FIGURE 4 | T-distributed stochastic neighbor embedding analysis. **(A)** Training and **(B)** validation sets.

Construction of Co-expression Network and GSEA

We visualized the co-expression network between lncRNAs and mRNAs using Cytoscape (**Figure 8A**). We also differentiated protective and risk factors from Sankey diagrams prepared using the ggalluvial package in R (**Figure 8B**). The potential immune response functions of the 10 ferroptosis-associated lncRNAs were further explored by comparing immune-related GSEA enrichment between the high- and low-risk groups. The results showed that the immune response and the immune system process were significantly enriched in the low-risk group (all $P < 0.05$, FDR < 0.05; **Figure 9**).

DISCUSSION

Although screening, diagnosis, and treatment have recently progressed, LUAD remains one of the most aggressive and fatal malignancies due to its complex genetic and molecular mechanisms (Chang et al., 2021). As emerging biomarkers of genes and other molecules, lncRNAs play important roles in

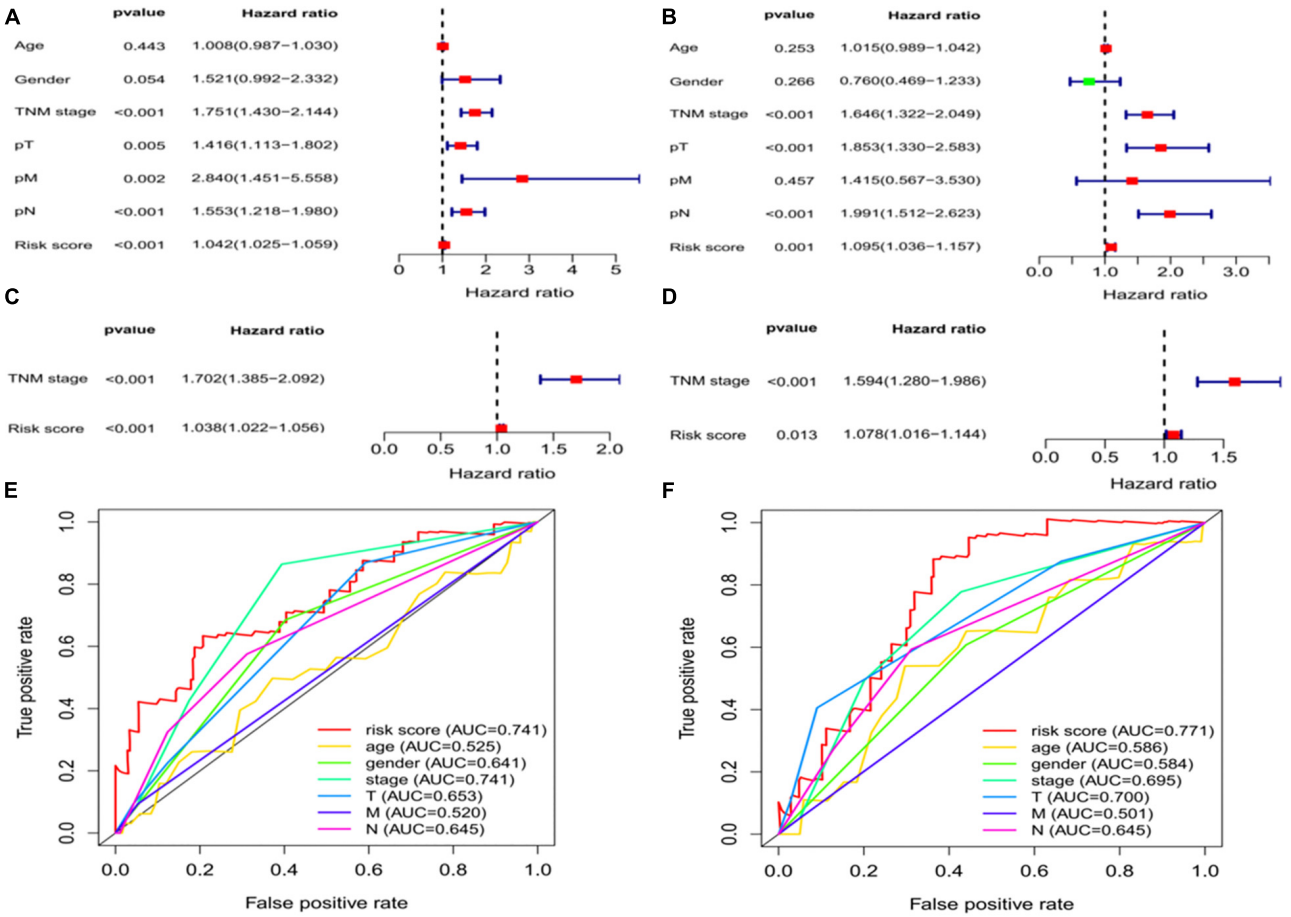


FIGURE 5 | Independent prognostic factors for LUAD overall survival. Univariate Cox regression analysis in the **(A)** training and **(B)** validation sets. Multivariate Cox regression analysis in the **(C)** training and **(D)** validation sets. Receiver operating characteristic curves of risk scores and other clinical characteristics based on overall survival in the **(E)** training and **(F)** validation sets. LUAD, lung adenocarcinoma.

the occurrence and development of various tumors, including LUAD (Peng et al., 2021; Zhang L. et al., 2021). Ferroptosis is a recently established type of regulatory cell death caused by the excessive accumulation of lipid peroxides and iron-dependent reactive oxygen species (Zhang J. et al., 2021). Ferroptosis is closely associated with the pathophysiological processes of many diseases, including LUAD (Wu Y. et al., 2020). However, lncRNAs associated with ferroptosis that can determine the prognosis of patients with LUAD have remained unknown. Therefore, we aimed to construct a prognostic model by exploring lncRNAs associated with ferroptosis that could detect LUAD and improve the survival rates of patients.

We obtained 259 ferroptosis-related genes and then identified 1,138 ferroptosis-related lncRNAs through a correlation analysis. Patients with LUAD were randomized into training and validation sets. We identified 29 prognostic ferroptosis-associated lncRNAs in the training set using univariate Cox regression. We avoided overfitting using lasso regression for dimensionality reduction and obtained 19 lncRNAs that were associated with ferroptosis. We then constructed a signature comprising 10 ferroptosis-related lncRNAs with the lowest AIC values using multivariate Cox regression analysis. All patients were then assigned to high- and low-risk groups based on risk scores. The results of univariate and multivariate Cox regression analysis of the training and validation sets indicated that the risk score is an independent risk factor affecting the prognosis of patients with LUAD. The AUC and t-SNE further confirmed the accuracy and distinguishing ability of the lncRNA signature.

Nomograms are intuitive and simple prognostic prediction models that have become more prevalent in medical research and clinical practice. They have the advantages of convenient operation and high prediction accuracy, and they have been widely applied to cancer research (Chen D. et al., 2021). Since few nomograms have been constructed based on ferroptosis-associated lncRNA, we established a model based on our multivariate findings, which directly reflected the degree of influence of risk scores on predictions of patient survival. The nomogram containing risk scores discriminated the training and validation sets with a C-index of 0.763 (95% CI, 0.712–0.814) and 0.736 (95% CI, 0.669–0.803), respectively (Van Oirbeek and Lesaffre, 2010). The established calibration curve also showed that the survival rate predicted by the nomogram in the training

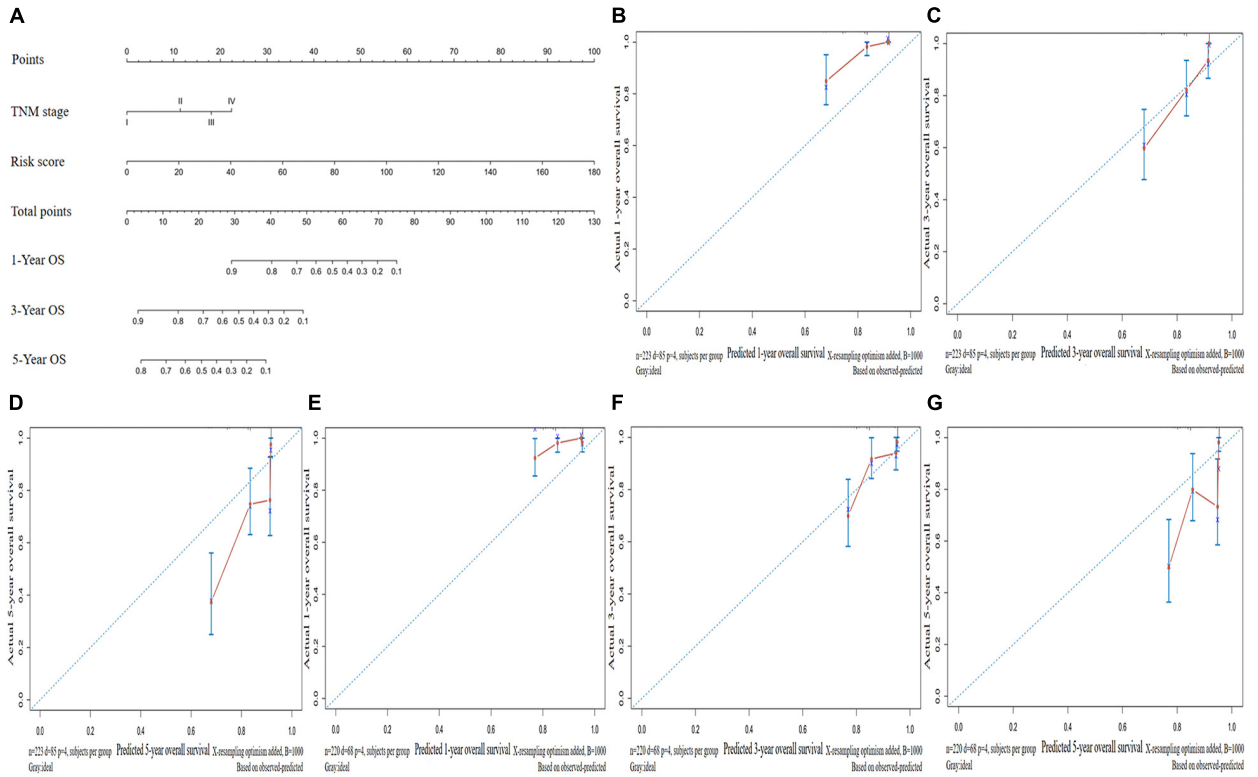


FIGURE 6 | Construction and validation of a predictive nomogram. **(A)** Nomogram for predicting the overall survival (OS) of patients with LUAD at 1, 3, and 5 years. Calibration curves of nomogram for OS prediction at 1, 3, and 5 years in the **(B–D)** training and **(E–G)** validation sets. LUAD, lung adenocarcinoma.

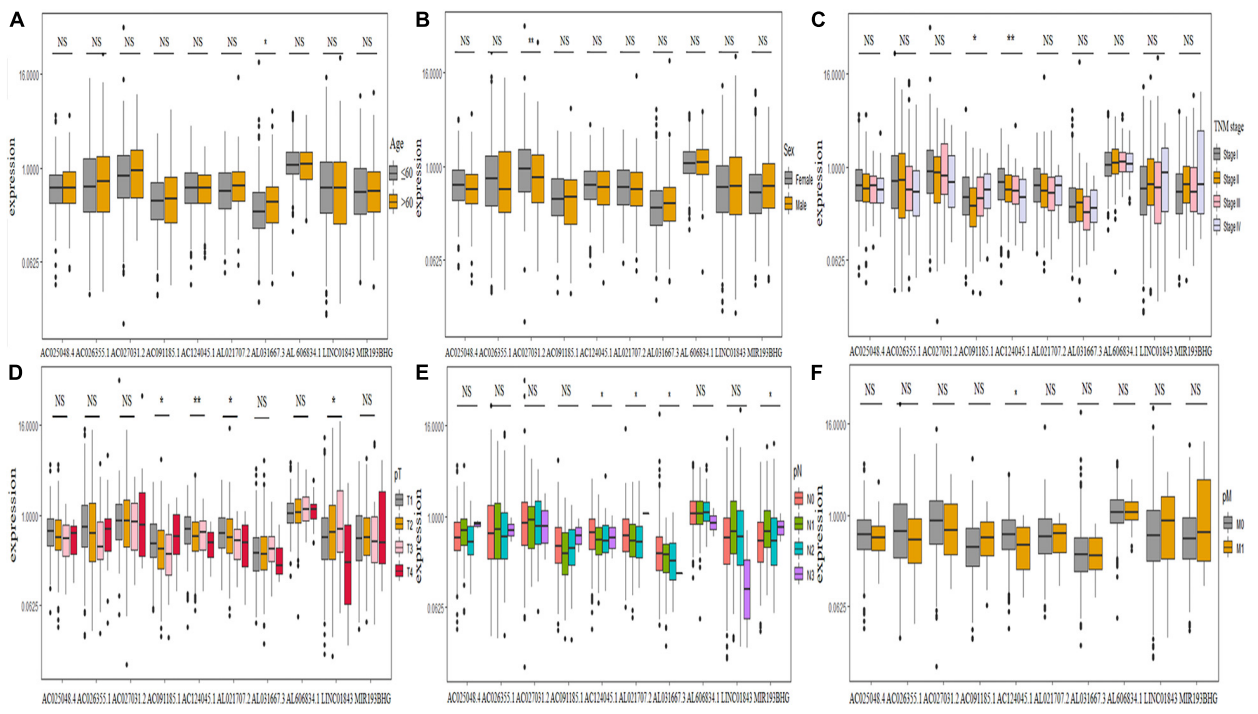


FIGURE 7 | Correlation between the expression level of 10 ferroptosis-associated long non-coding RNAs and clinicopathological characteristics. **(A–F)** Age, sex, TNM stage, pT stage, pN stage, and pM stage, respectively. TNM, tumor node metastasis; NS, not significant. $^{*}P < 0.05$; $^{**}P < 0.01$.

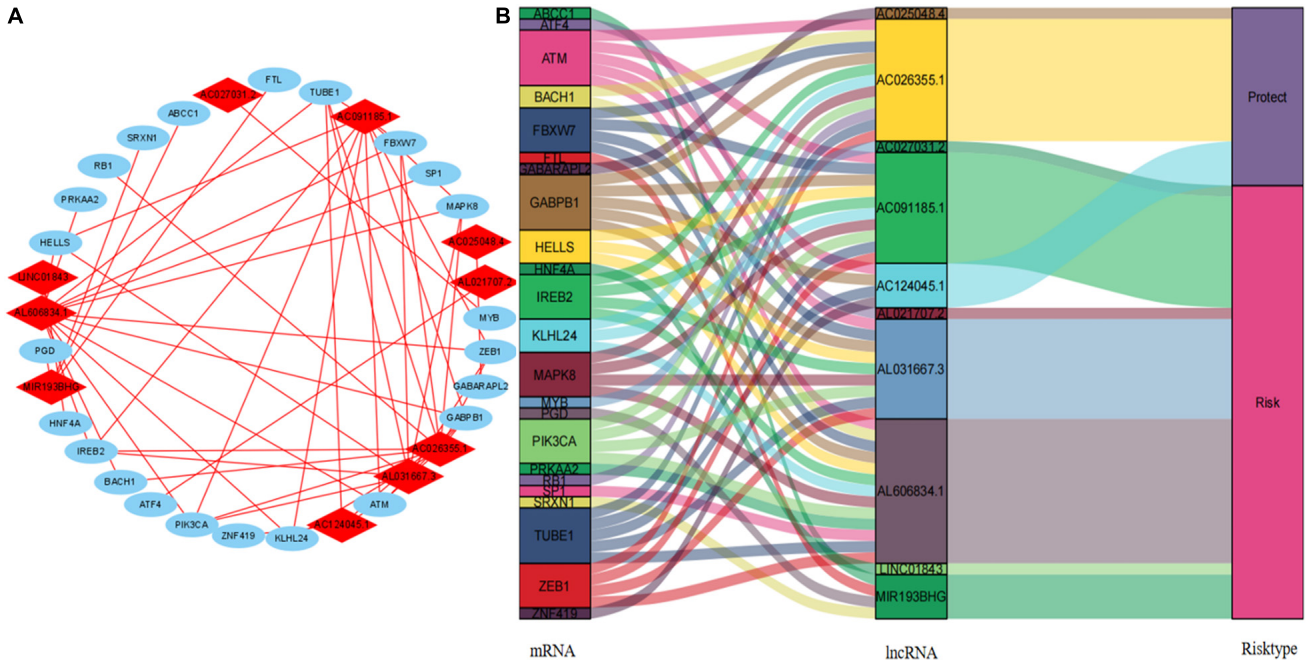


FIGURE 8 | Cytoscape and Sankey diagram of the lncRNA–mRNA co-expression network. **(A)** Cytoscape map: red and blue, lncRNA and mRNA, respectively. **(B)** Sankey diagram. lncRNAs, long non-coding RNAs.

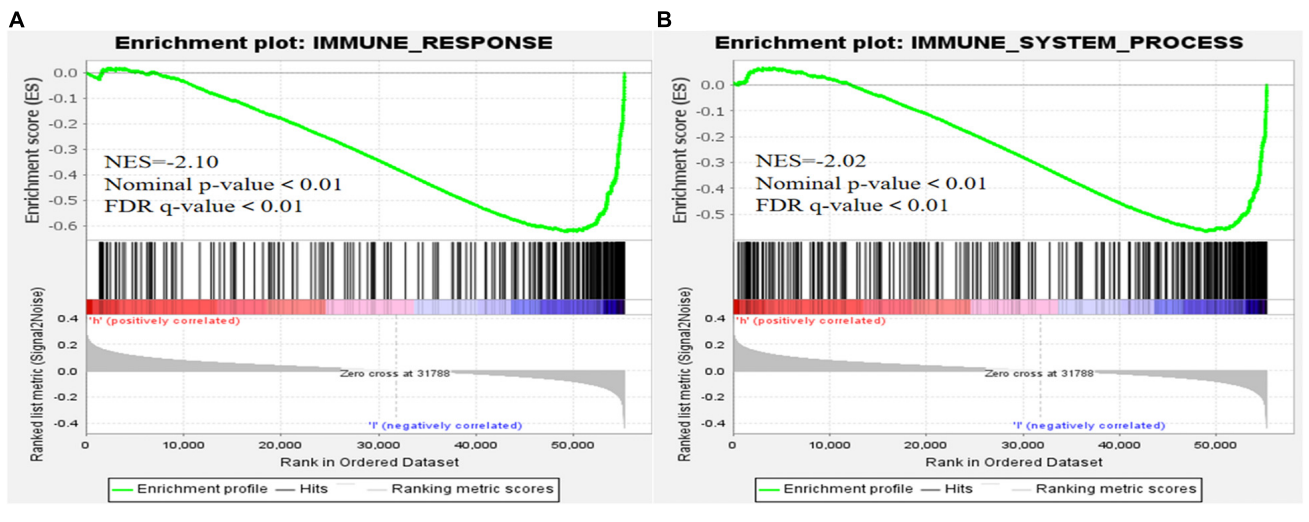


FIGURE 9 | Gene set enrichment analysis. **(A)** Immune response. **(B)** Immune system process (all $P < 0.05$; false discovery rate, < 0.05).

and validation sets was consistent with the actual survival rate (Zheng Z. et al., 2021). These results confirm the accuracy of our nomogram.

The Sankey diagram intuitively showed that seven lncRNAs were risk factors for prognosis, whereas three were prognostic protective factors. The abundant expression of LINC01843 might be associated with a poor prognosis in patients with LUAD (Li et al., 2020). MIR193BHG plays a key role in regulating physiological redox homeostasis, and lipid peroxidation is a key step in the process of ferroptosis (Wu X. et al., 2020). AC026355.1 might play a key role in the immune regulation of LUAD (You et al., 2021). We found that only AC124045.1 is associated with TNM stage, pT, pN, and pM. The expression of AC124045.1 decreased as TNM stage increased. Our results also found that AC124045.1 might be a key protective factor in terms of the prognosis of LUAD. However, little is known about this and other lncRNAs. Therefore, we plan to focus future studies on these lncRNAs to determine a new strategy for treating LUAD.

Tumor-related immune responses play important roles in cell infiltration and metastasis in the tumor microenvironment

(Marar et al., 2021; Zhu et al., 2021), whereas ferroptosis and lncRNAs play key regulatory roles in tumor-related immune responses (Wang W. et al., 2019; Sun et al., 2020; Tang et al., 2020). The results of our immune-related GSEA analyses of the high- and low-risk groups revealed that the immune response and immune system processes were significantly enriched in the low-risk group, suggesting that low-risk patients have ferroptosis-related, anti-tumor immune response processes that reduce the risk of death.

Continuous breakthroughs have recently occurred regarding ferroptosis and new treatments for diseases, and novel lncRNA functions are constantly being revealed. However, many gaps in knowledge about ferroptosis and lncRNAs remain to be filled. We identified 10 ferroptosis-associated lncRNAs using high-throughput sequencing technology and elaborated a possible immune response. As far as we can ascertain, this is the first study to identify a lncRNA prognostic signal associated with ferroptosis in LUAD. However, this study has some limitations. We only used test and validation sets from TCGA database to verify the effectiveness of the lncRNA prognostic model associated with ferroptosis. Relevant basic experiments to detect the expression levels of the identified ferroptosis-associated lncRNAs in cell lines and clinical samples are scarce. Therefore, further studies are needed to clarify the mechanisms of lncRNAs associated with ferroptosis in LUAD.

CONCLUSION

We identified and verified the signature of 10 lncRNAs associated with ferroptosis that have an independent prognostic value for LUAD patients and might be involved in the immune response. Therefore, we believe that our findings will serve as a potential prognostic indicator and inspire new treatment strategies involving ferroptosis to improve the prognosis of patients with 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/s.

AUTHOR CONTRIBUTIONS

DL and ZZ contributed to the conception and design. DL, ZZ, QZ, and WW analyzed and wrote the manuscript. ZZ, QZ, WW, YX, SL, and QC collected and processed the data. All the authors read and approved the final manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (approval number 82072355).

ACKNOWLEDGMENTS

We appreciate the free use of TCGA and FerrDb databases. We thank EditSprings (https://www.editsprings.com/yuyan) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.690509/full#supplementary-material

Supplementary Figure 1 | Correlation between the expression level of 10 ferroptosis-associated long non-coding RNAs and risk scores. **(A)** Training and **(B)** validation sets.

Supplementary Table 1 | List of the 259 ferroptosis-related genes.

REFERENCES

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18, 5–18. doi: 10.1038/nrc.2017.99

Chang, C., Sun, X., Wang, G., Yu, H., Zhao, W., Ge, Y., et al. (2021). A machine learning model based on PET/CT radiomics and clinical characteristics predicts ALK rearrangement status in lung adenocarcinoma. *Front. Oncol.* 11:603882. doi: 10.3389/fonc.2021.603882

Chen, D., Liu, Z., Liu, W., Fu, M., Jiang, W., Xu, S., et al. (2021). Predicting postoperative peritoneal metastasis in gastric cancer with serosal invasion using a collagen nomogram. *Nat. Commun.* 12:179. doi: 10.1038/s41467-020-20429-0

Chen, X., Kang, R., Kroemer, G., and Tang, D. (2021). Broadening horizons: the role of ferroptosis in cancer. *Nat. Rev. Clin. Oncol.* 18, 280–296. doi: 10.1038/s41571-020-00462-0

Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., et al. (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* 149, 1060–1072. doi: 10.1016/j.cell.2012.03.042

Garon, E. B., Aerts, J., Kim, J. S., Muehlenbein, C. E., Peterson, P., Rizzo, M. T., et al. (2021). Safety of pemetrexed plus platinum in combination with pembrolizumab for metastatic nonsquamous non-small cell lung cancer: a post hoc analysis of KEYNOTE-189. *Lung Cancer* 155, 53–60. doi: 10.1016/j.lungcan.2021.02.021

Ge, X., Chen, Y. E., Song, D., Mcdermott, M. L., and Li, J. J. (2020). Clipper: p-value-free FDR control on high-throughput data from two conditions. *bioRxiv* [Preprint] doi: 10.1101/2020.11.19.390773

Goulart, B., Chennupati, S., Fedorenko, C. R., and Ramsey, S. D. (2021). Access to tyrosine kinase inhibitors and survival in patients with advanced EGFR+ and ALK+ positive non-small-cell lung cancer treated in the real-world. *Clin. Lung Cancer* (in press). doi: 10.1016/j.cllc.2021.01.019

Hu, Y., Chen, H. Y., Yu, C. Y., Xu, J., Wang, J. L., Qian, J., et al. (2014). A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 5, 2230–2242. doi: 10.18632/oncotarget.1895

Jiang, X., Stockwell, B. R., and Conrad, M. (2021). Ferroptosis: mechanisms, biology and role in disease. *Nat. Rev. Mol. Cell Biol.* 22, 266–282. doi: 10.1038/s41580-020-00324-8

Kopp, F., and Mendell, J. T. (2018). Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172, 393–407. doi: 10.1016/j.cell.2018.01.011

Li, G., Yang, J., Zhao, G., Shen, Z., Yang, K., Tian, L., et al. (2021). Dysregulation of ferroptosis may involve in the development of non-small-cell lung cancer in Xuanwei area. *J. Cell. Mol. Med.* 25, 2872–2884. doi: 10.1111/jcmm.16318

Li, J. P., Li, R., Liu, X., Huo, C., Liu, T. T., Yao, J., et al. (2020). A seven immune-related lncRNAs model to increase the predicted value of lung adenocarcinoma. *Front. Oncol.* 10:560779. doi: 10.3389/fonc.2020.560779

Lortet-Tieulent, J., Soerjomataram, I., Ferlay, J., Rutherford, M., Weiderpass, E., and Bray, F. (2014). International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. *Lung Cancer* 84, 13–22. doi: 10.1016/j.lungcan.2014.01.009

Lou, J. S., Zhao, L. P., Huang, Z. H., Chen, X. Y., Xu, J. T., Tai, W. C., et al. (2021). Ginkgetin derived from *Ginkgo biloba* leaves enhances the therapeutic effect of cisplatin via ferroptosis-mediated disruption of the Nrf2/HO-1 axis in EGFR wild-type non-small-cell lung cancer. *Phytomedicine* 80:153370. doi: 10.1016/j.phymed.2020.153370

Lu, J., Xu, F., and Lu, H. (2020). LncRNA PVT1 regulates ferroptosis through miR-214-mediated TFR1 and p53. *Life Sci.* 260:118305. doi: 10.1016/j.lfs.2020.118305

Marar, C., Starich, B., and Wirtz, D. (2021). Extracellular vesicles in immunomodulation and tumor progression. *Nat. Immunol.* 22, 560–570. doi: 10.1038/s41590-021-00899-0

Meng, F., Zhou, Y., Dong, B., Dong, A., and Zhang, J. (2020). Long non-coding RNA LINC01194 promotes the proliferation, migration and invasion of lung adenocarcinoma cells by targeting miR-641/SETD7 axis. *Cancer Cell Int.* 20:588. doi: 10.1186/s12935-020-01680-3

Pang, B., Wang, Q., Ning, S., Wu, J., Zhang, X., Chen, Y., et al. (2019). Landscape of tumor suppressor long noncoding RNAs in breast cancer. *J. Exp. Clin. Cancer Res.* 38:79. doi: 10.1186/s13046-019-1096-0

Peng, J., Liu, F., Zheng, H., Wu, Q., and Liu, S. (2021). lncRNA ZFAS1 contributes to the radioresistance of nasopharyngeal carcinoma cells by sponging hsa-miR-7-5p to upregulate ENO2. *Cell Cycle* 20, 126–141. doi: 10.1080/15384101.2020.1864128

Qi, W., Li, Z., Xia, L., Dai, J., Zhang, Q., Wu, C., et al. (2019). LncRNA GABPB1-AS1 and GABPB1 regulate oxidative stress during erastin-induced ferroptosis in HepG2 hepatocellular carcinoma cells. *Sci. Rep.* 9:16185. doi: 10.1038/s41598-019-52837-8

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Sun, J., Zhang, Z., Bao, S., Yan, C., Hou, P., Wu, N., et al. (2020). Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J. Immunother. Cancer* 8:e000110. doi: 10.1136/jitc-2019-000110

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

Susko, E., and Roger, A. J. (2020). On the use of information criteria for model selection in phylogenetics. *Mol. Biol. Evol.* 37, 549–562. doi: 10.1093/molbev/msz228

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, necroptosis, and pyroptosis in anticancer immunity. *J. Hematol. Oncol.* 13:110. doi: 10.1186/s13045-020-00946-7

Van Oirbeek, R., and Lesaffre, E. (2010). An application of Harrell's C-index to PH frailty models. *Stat. Med.* 29, 3160–3171. doi: 10.1002/sim.4058

Wang, H., Peng, S., Cai, J., and Bao, S. (2021). Silencing of PTPN18 induced ferroptosis in endometrial cancer cells through p-P38-mediated GPX4/xCT down-regulation. *Cancer Manag. Res.* 13, 1757–1765. doi: 10.2147/CMAR.S278728

Wang, M., Mao, C., Ouyang, L., Liu, Y., Lai, W., Liu, N., et al. (2019). Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. *Cell Death Differ.* 26, 2329–2343. doi: 10.1038/s41418-019-0304-y

Wang, N., Wang, Y., Yang, Y., Shen, Y., and Li, A. (2013). "miRNA target prediction based on gene ontology," in *Proceedings of the 2013 Sixth International Symposium on Computational Intelligence and Design*, (Washington, DC: IEEE Computer Society), 430–433. doi: 10.1109/ISCID.2013.113

Wang, W., Green, M., Choi, J. E., Gijón, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8+ T cells regulate tumour ferroptosis during cancer immunotherapy. *Nature* 569, 270–274. doi: 10.1038/s41586-019-1170-y

Wang, Z., Chen, X., Liu, N., Shi, Y., Liu, Y., Ouyang, L., et al. (2021). A nuclear long non-coding RNA LINC00618 accelerates ferroptosis in a manner dependent upon apoptosis. *Mol. Ther.* 29, 263–274. doi: 10.1016/j.ymthe.2020.09.024

Wohlhieter, C. A., Richards, A. L., Uddin, F., Hulton, C. H., Quintanal-Villalonga, À, Martin, A., et al. (2020). Concurrent mutations in STK11 and KEAP1 promote ferroptosis protection and SCD1 dependence in lung cancer. *Cell Rep.* 33:108444. doi: 10.1016/j.celrep.2020.108444

Wu, X., Niculite, C. M., Preda, M. B., Rossi, A., Tebaldi, T., Butoi, E., et al. (2020). Publisher correction: regulation of cellular sterol homeostasis by the oxygen responsive noncoding RNA lincNORS. *Nat. Commun.* 11:5737. doi: 10.1038/s41467-020-19708-7

Wu, Y., Zhang, S., Gong, X., Tam, S., Xiao, D., Liu, S., et al. (2020). The epigenetic regulators and metabolic changes in ferroptosis-associated cancer progression. *Mol. Cancer* 19:39. doi: 10.1186/s12943-020-01157-x

You, J., Fang, W., Zhao, Q., Chen, L., Chen, L., and Chen, F. (2021). Identification of a RNA-seq based prognostic signature with seven immune-related lncRNAs for lung adenocarcinoma. *Clin. Lab.* 67. doi: 10.7754/Clin.Lab.2020.200663

Zhang, J., Xi, J., Huang, P., and Zeng, S. (2021). Comprehensive analysis identifies potential ferroptosis-associated mRNA therapeutic targets in ovarian cancer. *Front. Med.* 8:644053. doi: 10.3389/fmed.2021.644053

Zhang, L., Zhang, K., Liu, S., Zhang, R., Yang, Y., Wang, Q., et al. (2021). Identification of a ceRNA network in lung adenocarcinoma based on integration analysis of tumor-associated macrophage signature genes. *Front. Cell Dev. Biol.* 9:629941. doi: 10.3389/fcell.2021.629941

Zheng, Z., Lin, D., Chen, Q., Zheng, B., Liang, M., Chen, C., et al. (2021). Prognostic value of combined detection of preoperative albumin-to-fibrinogen ratio and neutrophil-to-lymphocyte ratio in operable esophageal squamous cell carcinoma patients without neoadjuvant therapy. *Cancer Manag. Res.* 13, 2359–2370. doi: 10.2147/CMAR.S296266

Zhou, N., and Bao, J. (2020). FerrDb: a manually curated resource for regulators and markers of ferroptosis and ferroptosis-disease associations. *Database* 2020:baaa021. doi: 10.1093/database/baaa021

Zhu, T., Ma, Z., Wang, H., Wei, D., Wang, B., Zhang, C., et al. (2021). Immune-related long non-coding RNA signature and clinical nomogram to evaluate survival of patients suffering esophageal squamous cell carcinoma. *Front. Cell Dev. Biol.* 9:641960. doi: 10.3389/fcell.2021.641960

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

frontiers in Genetics

REVIEW
published: 23 July 2021
doi: 10.3389/fgene.2021.689097

The CPNE Family and Their Role in Cancers

Haicheng Tang[1†], *Pei Pang*[2†], *Zhu Qin*[1†], *Zhangyan Zhao*[1], *Qingguo Wu*[1], *Shu Song*[3*] and *Feng Li*[1*]

[1] Department of Respiratory and Critical Care Medicine, Shanghai Public Health Clinical Center, Fudan University, Shanghai, China, [2] Department of Pathology, The First Affiliated Hospital of Soochow University, Suzhou, China, [3] Department of Pathology, Shanghai Public Health Clinical Center, Fudan University, Shanghai, China

Lung cancer is the leading cause of cancer-related deaths worldwide. Despite significant advances in cancer research and treatment, the overall prognosis of lung cancer patients remains poor. Therefore, the identification for novel therapeutic targets is critical for the diagnosis and treatment of lung cancer. CPNEs (copines) are a family of membrane-bound proteins that are highly conserved, soluble, ubiquitous, calcium dependent in a variety of eukaryotes. Emerging evidences have also indicated CPNE family members are involved in cancer development and progression as well. However, the expression patterns and clinical roles in cancer have not yet been well understood. In this review, we summarize recent advances concerning CPNE family members and provide insights into new potential mechanism involved in cancer development.

Keywords: copines, cancer, proliferation, metastasis, signaling pathway

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Gan Luan,
New Jersey Institute of Technology, United States
Flora Cimmino,
University of Naples Federico II, Italy
Peijie Wu,
Chengdu University of Traditional Chinese Medicine, China

***Correspondence:**
Shu Song
ycss1971@163.com
Feng Li
lifeng@shphc.org.cn

† *These authors contributed equally to this work*

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

Received: *31 March 2021*
Accepted: *21 June 2021*
Published: *23 July 2021*

Citation:
Tang H, Pang P, Qin Z, Zhao Z, Wu Q, Song S and Li F (2021) The CPNE Family and Their Role in Cancers. Front. Genet. 12:689097. doi: 10.3389/fgene.2021.689097

INTRODUCTION

Recently, based on its increased incidence and mortality, lung cancer has been listed as the leading cause of cancer-related deaths worldwide among males and females (Chen et al., 2016; Siegel et al., 2017). The application of traditional chemotherapeutic drugs can achieve response in some patients; however, these agents lack tissue specificity. Normal cells, especially bone marrow hematopoietic cells and epithelial cells in various organs, such as cells in the gastrointestinal tract, can also be killed by chemotherapeutic drugs (Perez-Herrero and Fernandez-Medarde, 2015; Somaiah et al., 2018). Therefore, traditional treatment methods are characterized by bottlenecks, limited means, and poor efficacy. And recurrence and metastases often develop in patients undergoing traditional treatments. In recent years, the concept of precision medicine has emerged, and corresponding treatments have been developed based on tumor tissue-specific gene changes. For example, genes, such as epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), and transmembrane receptor tyrosine kinase (ROS1), are detected in patient tissue specimens (Zhuang et al., 2019), and Corresponding first-, second- or even third-generation drugs have been developed to target these genes. In addition, RET rearrangement, NTRK fusion, BRAF mutation, MET14 mutation and HER2 mutation are also present in lung cancer (Staley et al., 1991; Farago and Azzoli, 2017; Liu et al., 2018a). At present, targeted therapy is mainly applied in patients with advanced non-small cell lung cancer. Despite remarkable breakthroughs in diagnosis and treatment, patient survival rates and treatment results have not changed significantly (Mayekar and Bivona, 2017; Ruiz-Cordero and Devine, 2020).

The 5-year survival rate remains only 15%, so it is important to develop new treatment strategies (Mulshine and Sullivan, 2005). Numerous potential mechanisms are involved in invasion and metastasis in non-small cell lung cancer, including tumor microenvironment (Wu et al., 2021). With the emergence of protein chips and gene chips, numerous lung cancer-related genes have been discovered, and it is important to identify specific and sensitive targets. Researchers are working to develop specific targeted drugs and to identify effective biomarkers for the prevention and diagnosis of non-small cell lung cancer. Several proteins have been identified as biomarkers and drug targets, but their exact role remains controversial. Therefore, there is a need to identify reliable biomarkers that can be used as new therapeutic targets for the effective treatment in non-small cell lung cancer patients.

CPNEs are a newly discovered class of phospholipid-binding proteins that are widely expressed, and their functions are Ca^{2+}-dependent and structurally evolutionarily conserved (Creutz et al., 1998). The CPNEs consisted of nine family members (Tomsig et al., 2003). Although their exact functions and biological roles remain unclear, an increasing number of studies have shown that CPNEs may mediate a variety of signaling pathways involved in tumorigenesis and development (**Figure 1**). This review highlights the biological properties of the copine family and their roles in membrane trafficking, tumor progression and metastasis.

THE CPNE FAMILY

The CPNE protein was originally discovered in nematodes and plants. Similar to other gene families, the CPNE family was also present throughout the evolution. CPNE1 was first reported in 1998, when Creutz isolated annexin in *Paramecium*. Nine CPNEs have been identified, and eight (CPNE1–8) are found in mammals (Tomsig and Creutz, 2000). CPNEs are highly homologous: CPNE2–5 exhibit 60%, 78%, 53% and 56% homology, respectively (Goel et al., 2019). The distribution of some CPNE genes is limited. Among them, CPNE1–3 are the most widely distributed, which expressed in almost all mammalian tissues, including brain, heart, lung, liver, intestine, spleen, testis and kidney. These genes induce the differentiation of granulocytes, and CPNE3 is expressed in the early stage of neutrophil differentiation (Cowland et al., 2003). CPNE4 is distributed in the heart and surrounding large blood vessels and cranial nerves (Goel et al., 2019). CPNE5 is mainly expressed in the lymphatic system, heart, stomach, spleen, lymph nodes and testis (Ding et al., 2008). CPNE6, which is also known as N-CPNE, is specifically expressed in brain neurons (Perestenko et al., 2010). CPNE6 has recently been found to be closely related to the learning and memory abilities of mice (Reinhard et al., 2016). In addition, CPNE6 expression is increased in epilepsy patients (Zhu et al., 2016). CPNE8 was initially identified as a major gene expressed in the prostate and testis. Further studies also found that CPNE8 is distributed in the prostate, heart and brain (Maitra et al., 2003).

MOLECULAR STRUCTURE AND BIOLOGICAL CHARACTERISTICS OF CPNEs

CPNEs are mainly composed of the N-terminus and the C-terminus, and the C-terminus contains an A-domain, which is structurally related to the extracellular structure of integrin, and its main function is protein binding (Nalefski and Falke, 1996). The N-terminus contains two C2 domains that play a key role in regulating the binding of calcium and phospholipids (Rizo and Sudhof, 1998). These domains are present in an increasing number of membrane-bound proteins and involved in a variety of cellular signaling pathways and cellular processes, including membrane trafficking, lipid messenger production, GTPase activation and protein phosphorylation (Mukhopadhyay et al., 2017; Yanez Arteta et al., 2018). The C2 domains are characterized by an eight-chain antiparallel β-sandwich structure and are classified into two different topologies with slightly different positions and connectivity in the structure of the β-chain (Viaud et al., 2021). The C-terminus of CPNE shares similarity with the protein binding domain of certain integrin known as the A domain (Singh et al., 2017). The domains of human CPNE1, 2 and 4 mediate the binding of CPNEs to target proteins. More than 20 target proteins have been identified, many of which involve intracellular signaling pathways, such as MAPK, ERK, CDC42, and C12 ubiquitin (Gupta et al., 2010; Guo et al., 2020). In addition, CPNEs recruit these target proteins to the phospholipid surface, suggesting that they can modulate the activity and localization of these proteins in cells in response to changes in intracellular Ca^{2+} (Creutz et al., 1998). Based on the current findings, we propose a hypothesis that one function of CPNEs is to facilitate Ca^{2+}-mediated regulation of intracellular signaling pathways (Ma et al., 2011). In addition, a variety of proteins have been identified as CPNE targets, many of which are involved in intracellular signaling pathways (Tomsig et al., 2004).

THE ROLE OF CPNEs IN MEMBRANE TRANSPORT

CPNEs are involved in the regulation of membrane fusion mainly because the C2 domain binds to annexin on the membrane under conditions of Ca^{2+} stimulation, thereby affecting lysosomal and endosomal fusion and regulating autophagy (Ghislat and Knecht, 2012). CPNEs are soluble membrane proteins that contain two tandem C2 domains at the N-terminus and an A domain at the C-terminus (Damer et al., 2005). The C2 domain acts as a calcium-dependent phospholipid binding motif and may be involved in cellular signaling and membrane trafficking pathways (Park et al., 2014). The A domain, which is named after von Willebrand factor, is a plasma and extracellular matrix protein that has been studied in integrins and several extracellular matrix proteins and appears to play a role in protein binding (Whittaker and Hynes, 2002). CPNEs have multiple targets in cells and their expression can be regulated in response to the activation of signal transduction pathways through crosstalk mechanisms,

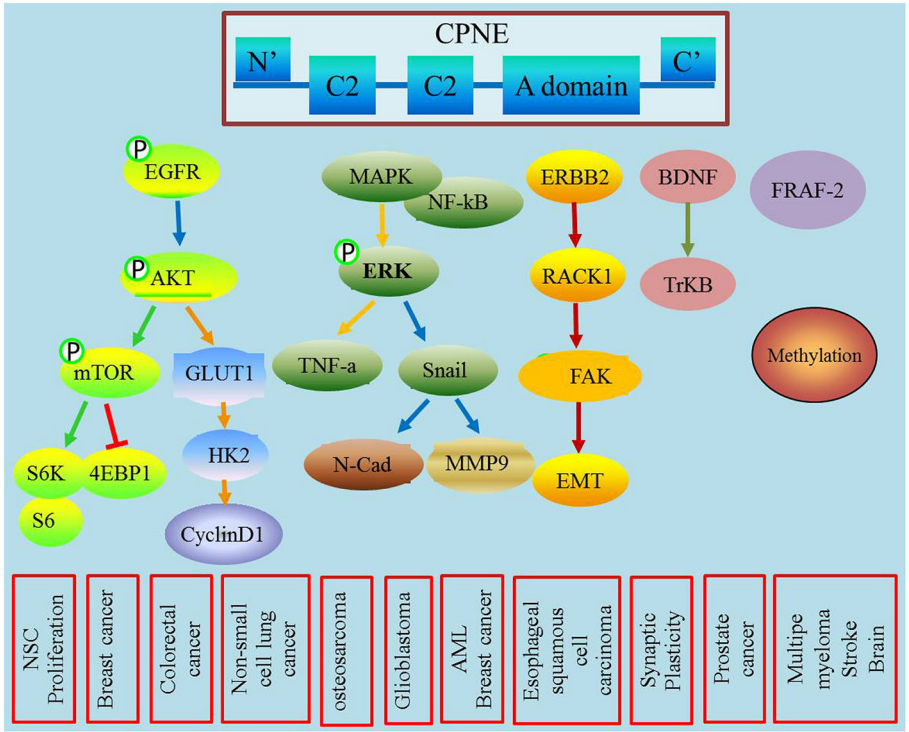


FIGURE 1 | This figure describes the role of CPNEs in mediating tumor biological process and also illustrated the detailed involved signal transducer molecules.

and CPNEs may serve as a skeleton protein during the membrane transport process (Tomsig et al., 2004).

THE ROLE OF CPNEs IN SIGNAL TRANSDUCTION

Different CPNEs could have different roles in different signaling pathways, which are dependent on their calcium sensitivities, lipid specificities, and target proteins (Ilacqua et al., 2018). Tomsig JL and other studies reported that CPNE1 is involved in cell signal transduction processes. CPNE1 plays an important role in regulating neuronal differentiation of HiB5 cells, mainly by activating the AKT signaling pathway via interacting with JAB-1 and 14-4-3 gamma (Kim et al., 2018). CPNE1 regulates the NF-κB-associated proteins MAPK, ERK and other signaling molecules through the A domain, which leads to the activation of the downstream TNF-α signaling pathway (Ramsey et al., 2008). Recent *in vitro* studies of positional candidates confirm that CPNE1 and STC2 are regulators of myogenesis (Hernandez Cordero et al., 2019). CPNE3 promotes migration and invasion in non-small cell lung cancer by interacting with RACK1 via FAK signaling activation (Lin et al., 2018).

THE ROLE OF CPNEs IN CANCER

CPNE1

In addition to its role in the central nervous system, accumulating studies have also highlighted its function in tumors. However, CPNE1 is rarely studied in tumors. Recent studies have shown higher CPNE1 expression in prostate cancer and that CPNE1expression is associated with the stage and prognosis of prostate cancer patients. Mechanistically, CPNE1 interacts with TRAF-2 to promote prostate cancer progression (Liang et al., 2017). In osteosarcoma, CPNE1 enhances cell proliferation and migration via the MAPK pathway and TGF-beta pathway (Jiang et al., 2018). Recent research revealed that CPNE1 promotes colorectal cancer progression by activating the AKT-GLUT1/HK2 cascade and enhances chemoresistance (Wang et al., 2021). Another study showed that CPNE1 promotes tumorigenesis and radioresistance in triple-negative breast cancer by regulating AKT activation, and targeted CPNE1 expression may be a strategy to sensitize triple-negative breast cancer cells to radiation therapy (Shao et al., 2020). For lung cancer, our previous studies demonstrated that CPNE1 is highly expressed in NSCLC tissues and is correlated with lymph node metastasis and poor survival in patients (Liu et al., 2018b). In addition, we also demonstrated that CPNE1 overexpression promotes cell proliferation and metastasis via the EGFR signaling pathway (Tang et al., 2018; Du et al., 2020). In hepatocellular carcinoma, overexpressed CPNE1 regulates the cell cycle process to mediate cell dedifferentiation (Skawran et al., 2008).

CPNE3

Studies have confirmed that CPNE3 is associated with schizophrenia, but the specific mechanism remains unclear. Other studies have confirmed that CPNE3 participates in

processes involved in acute myocardial infarction and coronary heart disease by potentially mediating the metabolism of nuclear fatty acids, serum total cholesterol and triglycerides (Tan et al., 2019). In addition, recent studies have also found that CPNE3 is highly expressed in breast cancer, prostate cancer and ovarian cancer and is involved in tumor cell proliferation and metastasis. Mechanistically, CPNE3 exhibits kinase activity, phosphorylates Hl histones and basic phospholipid proteins, activates downstream signaling pathways, and subsequently promotes tumor proliferation and metastasis (Thomas et al., 2008; Mo et al., 2013). Heinrich C found that CPNE3 induces EMT by activating the ErbB2 protein and induces tumor cell invasion and migration (Heinrich et al., 2010). Recent studies have shown that upregulated Copine3 with Jab1 activated downstream ErbB2 signaling and motility in breast cancer cell (Choi et al., 2016). Colorectal cancer patients with lower exosomal CPNE3 levels had substantially better disease-free survival and overall survival, implying that CPNE3 is a diagnostic and prognostic biomarker (Sun et al., 2019). In hepatocellular carcinoma, silencing the expression of CPNE3 enhances the sensitivity of cancer cells to the molecular targeted agent sorafenib (Chen et al., 2018). In NSCLC, CPNE3 expression level was positively correlated with clinical stage and TNM classification and quantitative proteomic analysis identifies CPNE3 as a novel metastasis-promoting gene in NSCLC (Lin et al., 2013). CPNE3 can promote migration and invasion in non-small cell lung cancer by interacting with RACK1 via FAK signaling activation (Lin et al., 2018). Upregulation of CPNE3 suppresses invasion, migration and proliferation of glioblastoma cells through FAK pathway inactivation (Shi et al., 2021). In addition, high expression of CPNE3 predicts adverse prognosis in acute myeloid leukemia (Fu et al., 2017).

CPNE5

Umeda S showed that CPNE5 expression is decreased in esophageal squamous cell carcinoma (ESCC), suggesting shorter overall survival in patients. Multivariate analysis showed that low CPNE5 expression was an independent prognostic factor for OS. Moreover, low CPNE5 expression potentially promotes the local growth of esophageal cancer and increases resistance to chemotherapy drugs. These findings suggest that CPNE5 can be used as a biomarker for predicting ESCC recurrence, especially in patients with local recurrence, and can help ensure that patients receive optimal treatment and follow-up (Umeda et al., 2018). It is also reported that multiple myeloma (MM) patients with higher CPNE5 expressions had longer event-free survival and overall survival, suggesting that CPNE5 might be used as a positive indicator for MM (Yang et al., 2018).

CPNE6

The report of CPNE6 in tumor progression in only limited in glioblastoma multiforme (GBM). GBM is a common type of brain tumor in adults; however effective candidate biomarkers for gene therapy in GBM remain unclear (Ni et al., 2018). It is reported that hub genes CPNE6, HAPLN2, CMTM3, NMI, CAPG, and PSMB8 were identified as potential liquid biopsy biomarkers for GBM diagnosis (Dent et al., 2008).

CPNE7

CPNE7 is a candidate tumor suppressor gene in breast cancer tissue. CPNE7 exhibits high homology with other members of the copine family, such as CPNE1, CPNE3 and CPNE6. CPNE7 is considered a potential tumor suppressor gene (Savino et al., 1999). Sequencing analysis of bladder transitional cell carcinoma revealed 565 candidate gene mutations, including CPNE7 and serine/arginine repetitive matrix 5, suggesting that CPNE7 mutations may be related to important mechanisms involved in bladder cancer. Identification of these genes may have therapeutic significance and may contribute to the development of future treatments for bladder cancer (Pan et al., 2016). To date, the role of CPNE7 in promoting or inhibiting cancer has not been clear, and further research is expected to confirm this hypothesis.

CPNE8

Ramsey found that in patients with acute myeloid leukemia (AML), the CPNE 8 gene can fuse with the AMLI gene to form an AML-CPNE8 chimera, thereby inhibiting AML gene transcription. It is hypothesized that CPNE8 negatively regulates the proliferation of AML cancer cells (Ramsey et al., 2003). A recent study showed that RP11-396F22.1 may represent an early diagnostic indicator of cervical cancer. After knocking down RP11-396F22.1, CPNE8 expression was significantly upregulated. It is hypothesized that CPNE8 is related to the occurrence of cervical cancer, but the specific mechanism requires further study (Zhao et al., 2018).

CPNE9

Recent research found that high CPNE5 and CPNE9 expression might serve as positive indicators of multiple myeloma, and the expression of both genes was a better predictor of survival in multiple myeloma patients (Zhu et al., 2021). A study by Liu found that CPNE9 is specifically expressed in pancreatic tumor tissues, indicating that CPNE9 is related to cancer progression (Liu and Liu, 2015). Moreover, CPNE9 was among the top five genes of the prognostic 14-gene signature that was used to construct the prognostic model that demonstrated a high predictive ability for glioblastoma (Hou et al., 2019). This finding suggests that CPNE9 has important clinical significance for the prognostic assessment of glioblastoma patients.

PROSPECTS

Early diagnosis and precise treatment of tumors are directly related to patient prognosis. Therefore, early diagnosis of tumors and marker screening are particularly important. At present, no clear biomarkers are available for the diagnosis of lung cancer, and most of the existing studies are limited. The CPNE family is an important

family that was discovered in recent years. CPNEs not only participate in the development and differentiation of the nervous system but also in the occurrence and development of numerous tumor types. CPNEs also affect the immune microenvironment of tumors. Studies on the roles of CPNEs in cancer are currently limited, and further research is needed. We hope that abnormal CPNEs expression will be validated as effective molecular markers for the early diagnosis, progression and prognosis of lung cancer.

AUTHOR CONTRIBUTIONS

HT: conception and design, manuscript writing, and review. PP and ZQ: data curation and analysis, study supervision. ZZ and QW: data curation and analysis. SS and FL: formal analysis, results discussion, and manuscript revision.

FUNDING

This work was supported by Shanghai Municipal Health Commission (No. 202040332), Shanghai Jinshan Science and Technology Commission (No. 2020-3-04), Shanghai Public Health Clinical Center (No. KY-GW-2021-16).

ACKNOWLEDGMENTS

We greatly apologize to the scientists, biologists, researchers, and colleagues whose works were not cited because of space limitations.

REFERENCES

Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer statistics in China. *CA. Cancer J. Clin*. 66, 115–132. doi: 10.3322/caac.21338

Chen, Z., Jiang, Z., Zhang, W., and He, B. (2018). Silencing the expression of copine-III enhances the sensitivity of hepatocellular carcinoma cells to the molecular targeted agent sorafenib. *Cancer Manage. Res*. 10, 3057–3067. doi: 10.2147/CMAR.S167781

Choi, H. Y., Park, N., Na, J. B., Ko, E. S., Park, J. Y., and Yoo, J. C. (2016). Direct binding of Copine3 with Jab1 activates downstream ErbB2 signaling and motility in SKBr3 breast cancer cells. *Oncol. Rep*. 35, 1147–1152. doi: 10.3892/or.2015.4472

Cowland, J. B., Carter, D., Bjerregaard, M. D., Johnsen, A. H., Borregaard, N., and Lollike, K. (2003). Tissue expression of copines and isolation of copines I and III from the cytosol of human neutrophils. *J. Leukoc. Biol*. 74, 379–388. doi: 10.1189/jlb.0203083

Creutz, C. E., Tomsig, J. L., Snyder, S. L., Gautier, M. C., Skouri, F., Beisson, J., et al. (1998). The copines, a novel class of C2 domain-containing, calcium-dependent, phospholipid-binding proteins conserved from *Paramecium* to humans. *J. Biol. Chem*. 273, 1393–1402. doi: 10.1074/jbc.273.3.1393

Damer, C. K., Bayeva, M., Hahn, E. S., Rivera, J., and Socec, C. I. (2005). Copine A, a calcium-dependent membrane-binding protein, transiently localizes to the plasma membrane and intracellular vacuoles in Dictyostelium. *BMC Cell Biol*. 6:46. doi: 10.1186/1471-2121-6-46

Dent, P., Yacoub, A., Park, M., Sarkar, D., Shah, K., Curiel, D. T., et al. (2008). Searching for a cure: gene therapy for glioblastoma. *Cancer Biol. Ther*. 7, 1335–1340. doi: 10.4161/cbt.7.9.6408

Ding, X., Jin, Y., Wu, Y., Wu, H., Xiong, L., Song, X., et al. (2008). Localization and cellular distribution of CPNE5 in embryonic mouse brain. *Brain Res*. 1224, 20–28. doi: 10.1016/j.brainres.2008.05.051

Du, W., Liu, T., Zhang, Y., Zeng, Y., Zhu, J., Tang, H., et al. (2020). MiR-195-5p is a potential factor responsible for CPNE1 differential expression between subtypes of non-small cell lung cancer. *J. Cancer* 11, 2610–2620. doi: 10.7150/jca.39884

Farago, A. F., and Azzoli, C. G. (2017). Beyond ALK and ROS1: RET, NTRK, EGFR and BRAF gene rearrangements in non-small cell lung cancer. *Trans. Lung Cancer R.* 6, 550–559. doi: 10.21037/tlcr.2017.08.02

Fu, L., Fu, H., Qiao, J., Pang, Y., Xu, K., Zhou, L., et al. (2017). High expression of CPNE3 predicts adverse prognosis in acute myeloid leukemia. *Cancer Sci*. 108, 1850–1857. doi: 10.1111/cas.13311

Ghislat, G., and Knecht, E. (2012). New Ca(2+)-dependent regulators of autophagosome maturation. *Commun. Integ. Biol*. 5, 308–311. doi: 10.4161/cib.20076

Goel, M., Li, T., and Badea, T. C. (2019). Differential expression and subcellular localization of Copines in mouse retina. *J. Comp. Neurol*. 527, 2245–2262. doi: 10.1002/cne.24684

Guo, Y. J., Pan, W. W., Liu, S. B., Shen, Z. F., Xu, Y., and Hu, L. L. (2020). ERK/MAPK signalling pathway and tumorigenesis. *Exp. Ther. Med*. 19, 1997–2007. doi: 10.3892/etm.2020.8454

Gupta, V. R., Wilson, B. A., and Blanke, S. R. (2010). Sphingomyelin is important for the cellular entry and intracellular localization of *Helicobacter pylori* VacA. *Cell. Microbiol*. 12, 1517–1533. doi: 10.1111/j.1462-5822.2010.01487.x

Heinrich, C., Keller, C., Boulay, A., Vecchi, M., Bianchi, M., Sack, R., et al. (2010). Copine-III interacts with ErbB2 and promotes tumor cell migration. *Oncogene* 29, 1598–1610. doi: 10.1038/onc.2009.456

Hernandez Cordero, A. I., Gonzales, N. M., Parker, C. C., Sokolof, G., Vandenbergh, D. J., Cheng, R., et al. (2019). Genome-wide associations reveal human-mouse genetic convergence and modifiers of myogenesis, CPNE1 and STC2. *Am. J. Hum. Genet*. 105, 1222–1236. doi: 10.1016/j.ajhg.2019.10.014

Hou, Z., Yang, J., Wang, H., Liu, D., and Zhang, H. (2019). A potential prognostic gene signature for predicting survival for glioblastoma patients. *Biomed. Res. Int*. 2019:9506461. doi: 10.1155/2019/9506461

Ilacqua, A. N., Price, J. E., Graham, B. N., Buccilli, M. J., McKellar, D. R., and Damer, C. K. (2018). Cyclic AMP signaling in Dictyostelium promotes the translocation of the copine family of calcium-binding proteins to the plasma membrane. *BMC Cell Biol*. 19:13. doi: 10.1186/s12860-018-0160-5

Jiang, Z., Jiang, J., Zhao, B., Yang, H., Wang, Y., Guo, S., et al. (2018). CPNE1 silencing inhibits the proliferation, invasion and migration of human osteosarcoma cells. *Oncol. Rep*. 39, 643–650. doi: 10.3892/or.2017.6128

Kim, T. H., Sung, S. E., Cheal Yoo, J., Park, J. Y., Yi, G. S., Heo, J. Y., et al. (2018). Copine1 regulates neural stem cell functions during brain development. *Biochem. Biophys. Res. Commun*. 495, 168–173. doi: 10.1016/j.bbrc.2017.10.167

Liang, J., Zhang, J., Ruan, J., Mi, Y., Hu, Q., Wang, Z., et al. (2017). CPNE1 is a useful prognostic marker and is associated with TNF Receptor-Associated Factor 2 (TRAF2) expression in prostate cancer. *Int. J. Clin. Exp. Med*. 23, 5504–5514. doi: 10.12659/MSM.904720

Lin, H., Zhang, X., Liao, L., Yu, T., Li, J., Pan, H., et al. (2018). CPNE3 promotes migration and invasion in non-small cell lung cancer by interacting with RACK1 via FAK signaling activation. *J. Cancer*. 9, 4215–4222. doi: 10.7150/jca.25872

Lin, H. C., Zhang, F. L., Geng, Q., Yu, T., Cui, Y. Q., Liu, X. H., et al. (2013). Quantitative proteomic analysis identifies CPNE3 as a novel metastasis-promoting gene in NSCLC. *J. Proteome Res*. 12, 3423–3433. doi: 10.1021/pr400273z

Liu, S., Li, S., Hai, J., Wang, X., Chen, T., Quinn, M. M., et al. (2018a). Targeting HER2 aberrations in non-small cell lung cancer with osimertinib. *Clin. Cancer Res*. 24, 2594–2604. doi: 10.1158/1078-0432.CCR-17-1875

Liu, S., Tang, H., Zhu, J., Ding, H., Zeng, Y., Du, W., et al. (2018b). High expression of Copine 1 promotes cell growth and metastasis in human lung adenocarcinoma. *Int. J. Oncol*. 53, 2369–2378. doi: 10.3892/ijo.2018.4558

Liu, X. X., and Liu, F. J. (2015). Novel bioinformatic identification of differentially expressed tissue-specific and cancer-related proteins from the human protein

atlas for biomarker discovery. *Genet. Mol. Res. GMR* 14, 4557–4565. doi: 10.4238/2015.May.4.14

Ma, C., Yao, Y., Yue, Q. X., Zhou, X. W., Yang, P. Y., Wu, W. Y., et al. (2011). Differential proteomic analysis of platelets suggested possible signal cascades network in platelets treated with salvianolic acid B. *PLoS ONE*. 6:e14692. doi: 10.1371/journal.pone.0014692

Maitra, R., Grigoryev, D. N., Bera, T. K., Pastan, I. H., and Lee, B. (2003). Cloning, molecular characterization, and expression analysis of Copine 8. *Biochem. Biophy. Res. Commun*. 303, 842–847. doi: 10.1016/S0006-291X(03)00445-5

Mayekar, M. K., and Bivona, T. G. (2017). Current landscape of targeted therapy in lung cancer. *Clin. Pharmacol. Ther*. 102, 757–764. doi: 10.1002/cpt.810

Mo, W., Zhang, J., Li, X., Meng, D., Gao, Y., Yang, S., et al. (2013). Identification of novel AR-targeted microRNAs mediating androgen signalling through critical pathways to regulate cell viability in prostate cancer. *PLoS ONE* 8:e56592. doi: 10.1371/journal.pone.0056592

Mukhopadhyay, S., Badgandi, H. B., Hwang, S. H., Somatilaka, B., Shimada, I. S., and Pal, K. (2017). Trafficking to the primary cilium membrane. *Mol. Biol. Cell*. 28, 233–239. doi: 10.1091/mbc.e16-07-0505

Mulshine, J. L., and Sullivan, D. C. (2005). Clinical practice. Lung cancer screening. *N. Engl. J. Med. NLM*. 352, 2714–2720. doi: 10.1056/NEJMcp042630

Nalefski, E. A., and Falke, J. J. (1996). The C2 domain calcium-binding motif: structural and functional diversity. *Protein Sci*. 5, 2375–2390. doi: 10.1002/pro.5560051201

Ni, Y., Zhang, F., An, M., Yin, W., and Gao, Y. (2018). Early candidate biomarkers found from urine of glioblastoma multiforme rat before changes in MRI. *Life Sci*. 61, 982–987. doi: 10.1007/s11427-017-9201-0

Pan, H., Xu, X., Wu, D., Qiu, Q., Zhou, S., He, X., et al. (2016). Novel somatic mutations identified by whole-exome sequencing in muscle-invasive transitional cell carcinoma of the bladder. *Oncol. Lett*. 11, 1486–1492. doi: 10.3892/ol.2016.4094

Park, N., Yoo, J. C., Lee, Y. S., Choi, H. Y., Hong, S. G., Hwang, E. M., et al. (2014). Copine1 C2 domains have a critical calcium-independent role in the neuronal differentiation of hippocampal progenitor HiB5 cells. *Biochem. Biophys. Res. Commun*.454, 228–233. doi: 10.1016/j.bbrc.2014.10.075

Perestenko, P. V., Pooler, A. M., Noorbakhshnia, M., Gray, A., Bauccio, C., and Jeffrey McIlhinney, R. A. (2010). Copines-1,−2,−3,−6 and−7 show different calcium-dependent intracellular membrane translocation and targeting. *FEBS J*. 277, 5174–5189. doi: 10.1111/j.1742-4658.2010.07935.x

Perez-Herrero, E., and Fernandez-Medarde, A. (2015). Advanced targeted therapies in cancer: drug nanocarriers, the future of chemotherapy. *Eur. J. Pharm. Biopharm*. 93, 52–79. doi: 10.1016/j.ejpb.2015.03.018

Ramsey, C. S., Yeung, F., Stoddard, P. B., Li, D., Creutz, C. E., and Mayo, M. W. (2008). Copine-I represses NF-kappaB transcription by endoproteolysis of p65. *Oncogene* 27, 3516–3526. doi: 10.1038/sj.onc.1211030

Ramsey, H., Zhang, D. E., Richkind, K., Burcoglu-O'Ral, A., and Hromas, R. (2003). Fusion of AML1/Runx1 to copine VIII, a novel member of the copine family, in an aggressive acute myelogenous leukemia with t(12;21) translocation. *Leukemia* 17, 1665–1666. doi: 10.1038/sj.leu.2403048

Reinhard, J. R., Kriz, A., Galic, M., Angliker, N., Rajalu, M., Vogt, K. E., et al. (2016). The calcium sensor Copine-6 regulates spine structural plasticity and learning and memory. *Nat. Commun*. 7:11613. doi: 10.1038/ncomms11613

Rizo, J., and Sudhof, T. C. (1998). C2-domains, structure and function of a universal Ca2+-binding domain. *J. Biol. Chem*. 273, 15879–15882. doi: 10.1074/jbc.273.26.15879

Ruiz-Cordero, R., and Devine, W. P. (2020). Targeted therapy and checkpoint immunotherapy in lung cancer. *Surgl. Pathol. Clin*. 13, 17–33. doi: 10.1016/j.path.2019.11.002

Savino, M., d'Apolito, M., Centra, M., van Beerendonk, H. M., Cleton-Jansen, A. M., Whitmore, S. A., et al. (1999). Characterization of copine VII, a new member of the copine family, and its exclusion as a candidate in sporadic breast cancers with loss of heterozygosity at 16q24.3. *Genomics*. 61, 219–226. doi: 10.1006/geno.1999.5958

Shao, Z., Ma, X., Zhang, Y., Sun, Y., Lv, W., He, K., et al. (2020). CPNE1 predicts poor prognosis and promotes tumorigenesis and radioresistance via the AKT singling pathway in triple-negative breast cancer. *Mol. Carcinogen* 59, 533–544. doi: 10.1002/mc.23177

Shi, D., Lin, B., Lai, J., Li, K., and Feng, Y. (2021). Upregulation of CPNE3 suppresses invasion, migration and proliferation of glioblastoma cells through FAK pathway inactivation. *J. Mol. Histol*. 52, 589–596. doi: 10.1007/s10735-021-09966-0

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer statistics. *CA. Cancer J. Clin*. 67, 7–30. doi: 10.3322/caac.21387

Singh, V., Ram, M., Kumar, R., Prasad, R., Roy, B. K., and Singh, K. K. (2017). Phosphorylation: implications in cancer. *Protein J*. 36, 1–6. doi: 10.1007/s10930-017-9696-z

Skawran, B., Steinemann, D., Becker, T., Buurman, R., Flik, J., Wiese, B., et al. (2008). Loss of 13q is associated with genes involved in cell cycle and proliferation in dedifferentiated hepatocellular carcinoma. *Modern pathol*. 21, 1479–1489. doi: 10.1038/modpathol.2008.147

Somaiah, C., Kumar, A., Sharma, R., Sharma, A., Anand, T., Bhattacharyya, J., et al. (2018). Mesenchymal stem cells show functional defect and decreased anti-cancer effect after exposure to chemotherapeutic drugs. *J. Biomed*. 25:5. doi: 10.1186/s12929-018-0407-7

Staley, J., Coy, D., Taylor, J. E., Kim, S., and Moody, T. W. (1991). [Des-Met14]bombesin analogues function as small cell lung cancer bombesin receptor antagonists. *Peptides* 12, 145–149. doi: 10.1016/0196-9781(91)90181-N

Sun, B., Li, Y., Zhou, Y., Ng, T. K., Zhao, C., Gan, Q., et al. (2019). Circulating exosomal CPNE3 as a diagnostic and prognostic biomarker for colorectal cancer. *J. Cell Physiol*. 234, 1416–1425. doi: 10.1002/jcp.26936

Tan, B., Liu, L., Yang, Y., Liu, Q., Yang, L., and Meng, F. (2019). Low CPNE3 expression is associated with risk of acute myocardial infarction: a feasible genetic marker of acute myocardial infarction in patients with stable coronary artery disease. *Cardiol. J*. 26, 186–193. doi: 10.5603/CJ.a2017.0155

Tang, H., Zhu, J., Du, W., Liu, S., Zeng, Y., Ding, Z., et al. (2018). CPNE1 is a target of miR-335-5p and plays an important role in the pathogenesis of non-small cell lung cancer. *J. Exp. Clin. Canc. Res*. 37:131. doi: 10.1186/s13046-018-0811-6

Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet*. 40, 310–315. doi: 10.1038/ng.91

Tomsig, J. L., and Creutz, C. E. (2000). Biochemical characterization of copine: a ubiquitous Ca2+-dependent, phospholipid-binding protein. *Biochemistry* 39, 16163–16175. doi: 10.1021/bi0019949

Tomsig, J. L., Snyder, S. L., and Creutz, C. E. (2003). Identification of targets for calcium signaling through the copine family of proteins. Characterization of a coiled-coil copine-binding motif. *J. Biol. Chem*. 278, 10048–10054. doi: 10.1074/jbc.M212632200

Tomsig, J. L., Sohma, H., and Creutz, C. E. (2004). Calcium-dependent regulation of tumour necrosis factor-alpha receptor signalling by copine. *Biochem. J*. 378, 1089–1094. doi: 10.1042/bj20031654

Umeda, S., Kanda, M., Koike, M., Tanaka, H., Miwa, T., Tanaka, C., et al. (2018). Copine 5 expression predicts prognosis following curative resection of esophageal squamous cell carcinoma. *Oncol. Rep*. 40, 3772–3780. doi: 10.3892/or.2018.6742

Viaud, J., Ceccato, L., Payrastre, B., and Gaits-Iacovoni, F. (2021). Liposome-based methods to study GTPase activation by phosphoinositides. *Methods Mol. Biol*. 2251, 185–194. doi: 10.1007/978-1-0716-1142-5_13

Wang, Y., Pan, S., He, X., Huang, H., Chen, J., Zhang, Y., et al. (2021). CPNE1 enhances colorectal cancer cell growth, glycolysis, and drug resistance through regulating the AKT-GLUT1/HK2 Pathway. *OncoTargets Ther*. 14, 699–710. doi: 10.2147/OTT.S284211

Whittaker, C. A., and Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol. Biol. Cell*. 13, 3369–3387. doi: 10.1091/mbc.e02-05-0259

Wu, P., Gao, W., Su, M., Nice, E. C., Zhang, W., Lin, J., et al. (2021). Adaptive mechanisms of tumor therapy resistance driven by tumor microenvironment. *Front Cell Dev. Biol*. 9:641469. doi: 10.3389/fcell.2021.641469

Yanez Arteta, M., Kjellman, T., Bartesaghi, S., Wallin, S., Wu, X., Kvist, A. J., et al. (2018). Successful reprogramming of cellular protein production through mRNA delivered by functionalized lipid nanoparticles. *PNAS* 115, E3351–E3360. doi: 10.1073/pnas.1720542115

Yang, Q., Wang, R., Wei, B., Peng, C., Wang, L., Hu, G., et al. (2018). Candidate biomarkers and molecular mechanism investigation for

glioblastoma multiforme utilizing WGCNA. *Biomed. Res. Int*. 2018:4246703. doi: 10.1155/2018/4246703

Zhao, Y., Huang, J., Liu, T., He, S., Shang, C., Guo, L., et al. (2018). Overexpression of long non-coding RNA RP11-396F22.1 correlates poor prognosis of patients with early-stage cervical cancer. *Am. J. Transl. Res.* 10, 684–695.

Zhu, B., Zha, J., Long, Y., Hu, X., Chen, G., and Wang, X. (2016). Increased expression of copine VI in patients with refractory epilepsy and a rat model. *J. Neurol. Sci.* 360, 30–36. doi: 10.1016/j.jns.2015.11.041

Zhu, P., Qian, T., Si, C., Liu, Y., Cui, L., Huang, W., et al. (2021). High expression of CPNE5 and CPNE9 predicts positive prognosis in multiple myeloma. *Cancer Biomark.* 31, 77–85. doi: 10.3233/CBM-203108

Zhuang, X., Zhao, C., Li, J., Su, C., Chen, X., Ren, S., et al. (2019). Clinical features and therapeutic options in non-small cell lung cancer patients with concomitant mutations of EGFR, ALK, ROS1, KRAS or BRAF. *Cancer Med.* 8, 2858–2866. doi: 10.1002/cam4.2183

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 August 2021
doi: 10.3389/fgene.2021.689801

A Prognostic Autophagy-Related Gene Pair Signature and Small-Molecule Drugs for Hepatocellular Carcinoma

*ZeBing Song, GuoPei Zhang, Yang Yu and ShaoQiang Li**

Department of Liver Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Dysregulation of autophagy-related genes (ARGs) is related to the prognosis of cancers. However, the aberrant expression of ARGs signature in the prognosis of hepatocellular carcinoma (HCC) remain unclear. Using The Cancer Genome Atlas and the International Cancer Genome Consortium database, 188 common autophagy-related gene pairs (ARGPs) were identified. Through univariate, least absolute shrinkage and selection operator analysis, and multivariate Cox regression analysis, a prognostic signature of the training set was constructed on the basis of 6 ARGPs. Further analysis revealed that the ARGP based signature performed more accurately in overall survival (OS) prediction compared to other published gene signatures. In addition, a high risk of HCC was closely related to CTLA4 upregulation, LC3 downregulation, low-response to axitinib, rapamycin, temsirolimus, docetaxel, metformin, and high-response to bleomycin. Univariate Cox and multivariate Cox analysis revealed that the risk score was an independent prognostic factor for HCC. These results were internally validated in the test and TCGA sets and externally validated in the ICGC set. A nomogram, consisting of the risk score and the TNM stage, performed well when compared to an ideal nomogram. In conclusion, a 6-ARGP-based prognostic signature was identified and validated as an effective predictor of OS of patients with HCC. Furthermore, we recognized six small-molecule drugs, which may be potentially effective in treating HCC.

Keywords: autophagy-related gene pair signature, small-molecule drugs, hepatocellular carcinoma, bioinformatic analysis, autophagy

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Ning Wang,
Arcus Biosciences, United States
Byoung Kuk Jang,
Keimyung University Dongsan Medical Center, South Korea

***Correspondence:**
ShaoQiang Li
lishaoq@mail.sysu.edu.cn

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

Received: *01 April 2021*
Accepted: *28 July 2021*
Published: *23 August 2021*

Citation:
Song ZB, Zhang GP, Yu Y and Li SQ (2021) A Prognostic Autophagy-Related Gene Pair Signature and Small-Molecule Drugs for Hepatocellular Carcinoma. Front. Genet. 12:689801. doi: 10.3389/fgene.2021.689801

INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common liver malignancies worldwide, with increasing rates of morbidity and mortality annually (Bray et al., 2018). The prognosis of patients with HCC is poor owing to its high recurrence rate. Hence, effective biomarkers are needed to help improve the prognosis of patients with HCC.

Autophagy, also called "programmed cell death type II," plays an important role in *tumorigenesis*, metastasis and drug resistance (Huang et al., 2018). Liu K. et al. (2018) reported that autophagy

is required for benign hepatic tumors to progress into malignant HCC. Another study revealed that the activation of autophagy can promote metastasis through the upregulation of MCT1 via activating Wnt/β-catenin signaling in HCC cells (Fan et al., 2018). Sorafenib is an effective molecular-targeted drug used to treat advanced-stage HCC by inducing autophagy, thus prolonging the survival of patients with HCC (Raza and Sood, 2014; Finn et al., 2018). While only approximately 30% of patients with advanced HCC respond well to sorafenib, resistance to sorafenib remains an open question, which may result from pro-survival pathways of autophagy induced by sorafenib (Sun T. et al., 2017; Jiang et al., 2018). These findings indicate that autophagy plays an important role in the prognosis of patients with HCC. Hence, autophagy-related genes (ARGs) can be effective biomarkers to diagnose, and guide treatments for HCC. However, some effective biomarkers to predict the prognosis of HCC have not been established thus far.

To construct an autophagy-related prognostic signature for HCC, The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) data sets were used as the data source in this study. A 6-autophagy-related gene pair (ARGP) prognostic signature was identified and validated for its predictability of overall survival (OS) among patients with HCC in comparison with 3 others previously reported prognostic gene signatures. Furthermore, the effectiveness of small-molecule drugs for HCC was assessed, and 6 types of drugs were identified and validated in this study.

TABLE 1 | Clinicopathological parameters of hepatocellular carcinoma patients in training set, test set and TCGA data set.

Covariates	Type	Total	Train	Test	*P*-value
Age	≤65	232 (62.7%)	113 (61.08%)	119 (64.32%)	0.5909
	>65	138 (37.3%)	72 (38.92%)	66 (35.68%)	
	Unknow	0 (0%)	0 (0%)	0 (0%)	
Gender	Female	121 (32.7%)	68 (36.76%)	53 (28.65%)	0.1208
	Male	249 (67.3%)	117 (63.24%)	132 (71.35%)	
Grade	G1-2	232 (62.7%)	122 (65.95%)	110 (59.46%)	0.3227
	G3-4	133 (35.95%)	62 (33.51%)	71 (38.38%)	
	Unknow	5 (1.35%)	1 (0.54%)	4 (2.16%)	
Stage	I-II	256 (69.19%)	127 (68.65%)	129 (69.73%)	0.7612
	III-IV	90 (24.32%)	47 (25.41%)	43 (23.24%)	
	Unknow	24 (6.49%)	11 (5.95%)	13 (7.03%)	
T	T1-2	274 (74.05%)	137 (74.05%)	137 (74.05%)	0.9534
	T3-4	94 (25.41%)	46 (24.86%)	48 (25.95%)	
	Unknow	2 (0.54%)	2 (1.08%)	0 (0%)	
M	M0	266 (71.89%)	134 (72.43%)	132 (71.35%)	1
	M1	4 (1.08%)	2 (1.08%)	2 (1.08%)	
	Unknow	100 (27.03%)	49 (26.49%)	51 (27.57%)	
N	N0	252 (68.11%)	126 (68.11%)	126 (68.11%)	0.1387
	N1	4 (1.08%)	4 (2.16%)	0 (0%)	
	Unknow	114 (30.81%)	55 (29.73%)	59 (31.89%)	

METERIALS AND METHODS

Data Source

Transcription profiling RNA data, along with the HCC clinical data were downloaded from TCGA[1] (Liu J. et al., 2018), and were used to identify differentially expressed ARGs. Finally, 370 patients with complete survival data were identified from TCGA and randomized into a training set ($n = 185$) and test set ($n = 185$); These two sets were used to develop and internally validate the HCC prognostic signature. The demographic characteristics of the training set, test set, and TCGA set are summarized in **Table 1**. To externally validate the prognostic value of the prognostic signature, the gene expression data, and the clinical data on patients with HCC from the ICGC database (ICGC-LIRI-JP)[2] were downloaded as well. The databases selection and data procession flow chart of this study were shown in **Figure 1**.

Identification of Differential Expression of ARGs Between HCC and Non-tumor Samples of TCGA

An ARG set was downloaded from the Human Autophagy Database[3]. We then extracted the expression profile of these ARGs on the basis of the gene set and TCGA transcription profiling data. Differentially expressed ARGs were identified using the limma package. Differential expression of ARGs with a log2 fold change ($|\log2FC| > 1$ and a false discovery rate < 0.05 were considered significant and were included in the subsequent analysis. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional analysis were performed to explore the potential molecular function of these differentially expressed ARGs (Bandyopadhyay and Mallick, 2014).

Construction of a Prognostic Signature Based on the Training Set

In this study, OS was considered the primary endpoint. Pairwise comparison was performed between the expression profiles of differentially expressed ARGs in each HCC sample to obtain a score for each ARGP by using the R software. According to the proposed algorithm (Heinäniemi et al., 2013), if the expression level of the first ARG was higher than that of the second ARG in an ARGP, the score of this ARGP was 1; otherwise, the score was 0. If the score of an ARGP was 0 or 1 in > 80% of the samples of the training or the test sets, the ARGP was discarded, and the rest of the ARGPs were involved in subsequent analysis. Univariate Cox regression analysis was performed for the training set to identify OS-related ARGPs (FDR < 0.05), and the Least absolute shrinkage and selection operator (LASSO) analysis was performed to avoid overfitting of the prognostic signature. The most stable ARGP prognostic signature was constructed through multivariate Cox regression analysis (FDR < 0.05). In this study, patients were categorized into high- and low-risk groups in accordance with the median risk score. The risk score of the prognostic signature was calculated by multiplying the expression level

[1]https://cancergenome.nih.gov/

[2]https://dcc.icgc.org/

[3]http://www.autophagy.lu/

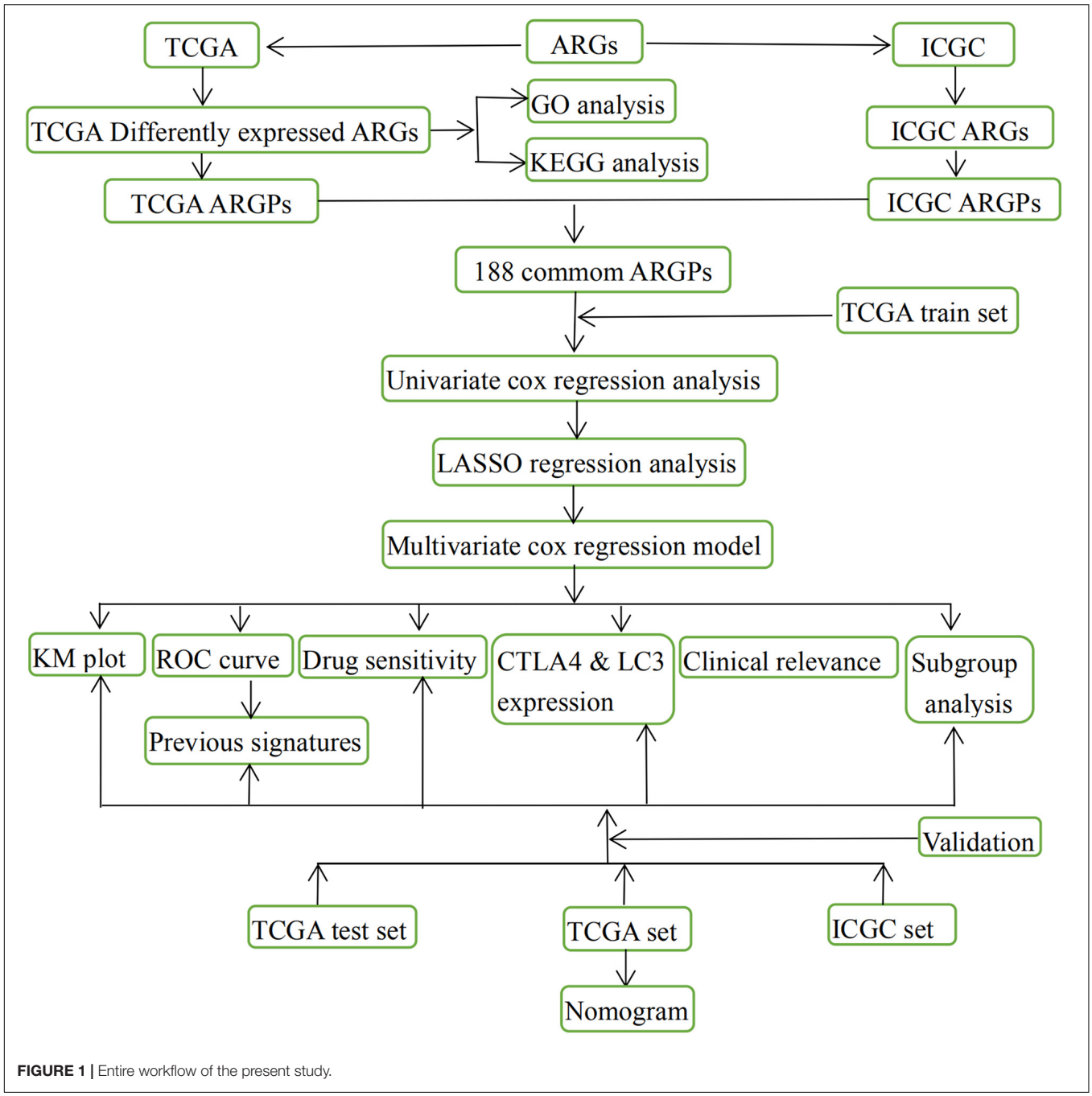


FIGURE 1 | Entire workflow of the present study.

with Cox regression coefficients of the ARGPs. The formula was as followed: risk score = Σ Cox regression coefficient of ARGPi * expression value of gene ARGPi.

Evaluation and Validation of the Prognostic Signature

The prognostic signature was evaluated by utilizing the training set and validated using the test set, TCGA set, and ICGC set. Following the median risk score of the train set, the patients were classified into high- and low-risk groups. The Kaplan-Meier (KM) method was applied to compare the OS between the high- and low-risk groups. The receiver operating characteristic (ROC) curve was plotted and the area under the curve (AUC) was calculated to ensure that the prognostic signature prediction efficacy can be estimated. Both univariate and multivariate Cox analysis were conducted with the clinicopathologic features and risk score to explore the HCC prognostic factors. Furthermore, we compared the AUC of the prognostic signature with that of 3 published gene prognostic signatures in the TCGA set. Subgroup analysis were performed to expand the application scope of the ARGP signature.

Prediction of Potential Small-Molecule Drugs

The drug response toward axitinib, rapamycin, temsirolimus, docetaxel, metformin, and bleomycin in each patient with HCC in the training set, test set, TCGA set, and ICGC set was calculated on the basis of the Genomics of Drug Sensitivity in Cancer (GDSC)[4] by using the prophetic R package[5]. The half maximal inhibitory concentration (IC50) value for patients with HCC was used to evaluate the effectiveness of these drugs, and $P < 0.05$ was set as the cutoff value.

Establishment and Evaluation of a Nomogram for Predicting the Survival of Patients With HCC

We included all independent clinicalpathological prognostic factors selected from multivariate Cox regression analysis to construct a nomogram that can assess an OS probability of 1, 3, and 5 years for patients with HCC. The prediction probability of the nomogram was compared with the observed actual probability form the calibration curve to verify its accuracy. Overlaps with the reference line indicate that the model is accurate.

Statistical Analysis

Statistical analysis was performed using the R software (version 3.6.3)[6] and Perl software (version 5.30)[7]. Cluster heatmaps and volcano maps were generated using gplots and heatmap packages. Univariate and multivariable Cox proportional hazards regression analysis were performed using the survival R software package. The KM analysis was performed using the survival R package and assessed using the log-rank test (Aalen, 1988). The survival ROC R package was used to calculate the AUC of the survival ROC curve.

RESULTS

Identification of Differential Expression of ARGs and Assessment of the Potential Molecular Function of ARGs

As shown in **Figures 2A,B** 59 ARGs were identified, including 4 down-regulated and 55 up-regulated genes. GO analysis of these ARGs revealed that "autophagy," "vacuolar membrane," and "protein kinase regulator activity" were the most frequent biological terms for biological processes, cellular components, and molecular functions, respectively (**Figures 2C,D**; Wang et al., 2021). KEGG analysis revealed that the primary pathways of these ARGs were "autophagy-animal," "IL-17 signaling pathway," "PI3K-Akt signaling pathway," and "mTOR signaling pathway," which were primarily correlated with autophagy, immune process, and carcinogenicity (**Figures 2E,F**).

[4]https://www.cancerrxgene.org

[5]https://github.com/paulgeeleher/pRRophetic

[6]https://cran.r-project.org/

[7]https://strawberryperl.com/

Establishment of an ARGP Signature in the Training Set

In total, 188 common ARGPs from among the TCGA and ICGC expression profile data were extracted. 17 ARGPs were found to be related to OS among patients with HCC, as revealed through univariate Cox regression analysis (**Figure 3B**). Thereafter, 9 ARGPs were found to be capable to construct a prognostic signature through LASSO analysis (**Figure 3A**). Finally, a 6-ARGP prognostic signature was constructed through multivariate Cox regression analysis, which included BAK1| PELP1, BIRC5| CDKN2A, BIRC5| RGS19, CAPN2| ULK3, DIRAS3| TMEM74, and PRKCD| RB1CC1 (**Figure 3C**). The risk score of our prognostic signature was as follows: risk score = (the expression level of BAK1| PELP1 * 0.6951) + (the expression level of BIRC5| CDKN2A * 0.5093) + (the expression level of BIRC5| RGS19 * 0.6322) + (the expression level of CAPN2| ULK3 * 0.5645) + (the expression level of DIRAS3| TMEM74 * -0.4505) + (the expression level of PRKCD| RB1CC1 * 0.5581).

The AUC of the 1-, 2-, and 3-year OS were 0.773, 0.761, and 0.761, respectively (**Figure 4E**). As shown in **Figure 4A**, the OS of the high-risk group was poorer than that of the low-risk group.

Validation of the Predictive Values of the ARGP Prognostic Signature

The predictive values of the ARGP prognostic signature were validated in the test set, TCGA set, and ICGC set, respectively. To improve the accuracy of validation, patients in the aforementioned 3 data sets were randomized into high- and low-risk groups using the same cutoff of the training set, i.e., the median risk score of the training set. As shown in **Figures 4B–D**, the OS of the high-risk group of the aforementioned 3 data sets had poorer OS than the low-risk group. The AUC of test set, TCGA set and ICGC set was 0.7, 0.735, and 0.726 after year 1; 0.712, 0.733, and 0.77 after year 2; and 0.648, 0.707, and 0.77 after year 3, respectively (**Figures 4F–H**). In addition, the AUC of the ARGP signature was higher than age, gender, TNM stage and tumor grade in the train, test, TCGA and ICGC set (**Figures 4I–L**). The survival status of patients, the rank of the risk score, and the heatmap of expression profiles of the 6 ARGPs in the low- and high-risk groups are indicated in **Supplementary Figure 1**.

As shown in **Figure 5**, the univariate and multivariate Cox regression analysis indicate that the risk score could potentially be an independent prognostic factor after adjustment by age, gender, tumor grade, and TNM stage in the training (HR: 2.066, 95% CI: 1.637–2.607, $P < 0.001$), test (HR: 1.801, 95% CI: 1.487–2.18, $P < 0.001$), TCGA (HR: 1.397, 95% CI: 1.203–1.622, $P < 0.001$) and ICGC sets (HR: 1.59, 95% CI: 1.248–2.025, $P < 0.001$).

Subgroup analysis based on age (>65 and ≤ 65 years), gender (male and female), tumor grade (G1-2 and G3-4), TNM stage (I-II and III-IV), tumor stage (T1-2 and T3-4), lymph node metastasis status (N0), and *distant metastasis* status (M0) in the training, test, TCGA, and ICGC sets were performed to further validate the predictive values of the ARGP signature. As shown in **Supplementary Figures 2, 3**, all subgroup analysis in the training, TCGA, and ICGC sets performed well in OS

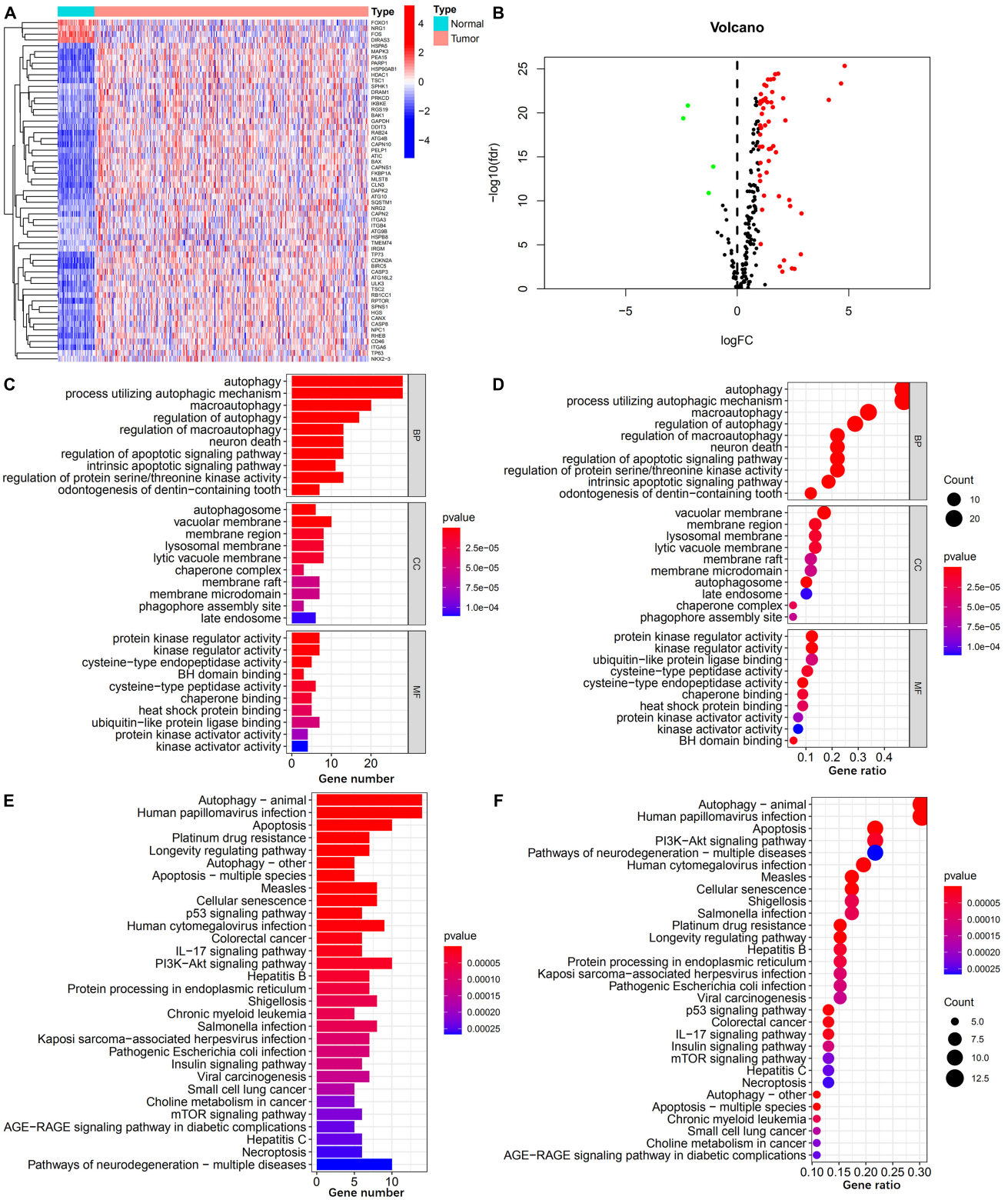


FIGURE 2 | Identification of differentially expressed ARGs. **(A,B)** Heatmap and volcano plot illustrates the expression of 59 differentially expressed ARGs between HCC tumor and non-tumor specimens. **(C,D)** Barplot and bubble plot of GO analysis shows the top 10 biological functions of the differentially expressed ARGs in the biological processes, cellular components, and molecular functions. **(E,F)** Barplot and bubble plot of KEGG analysis shows the top 30 signaling pathways of the differentially expressed ARGs participated in.

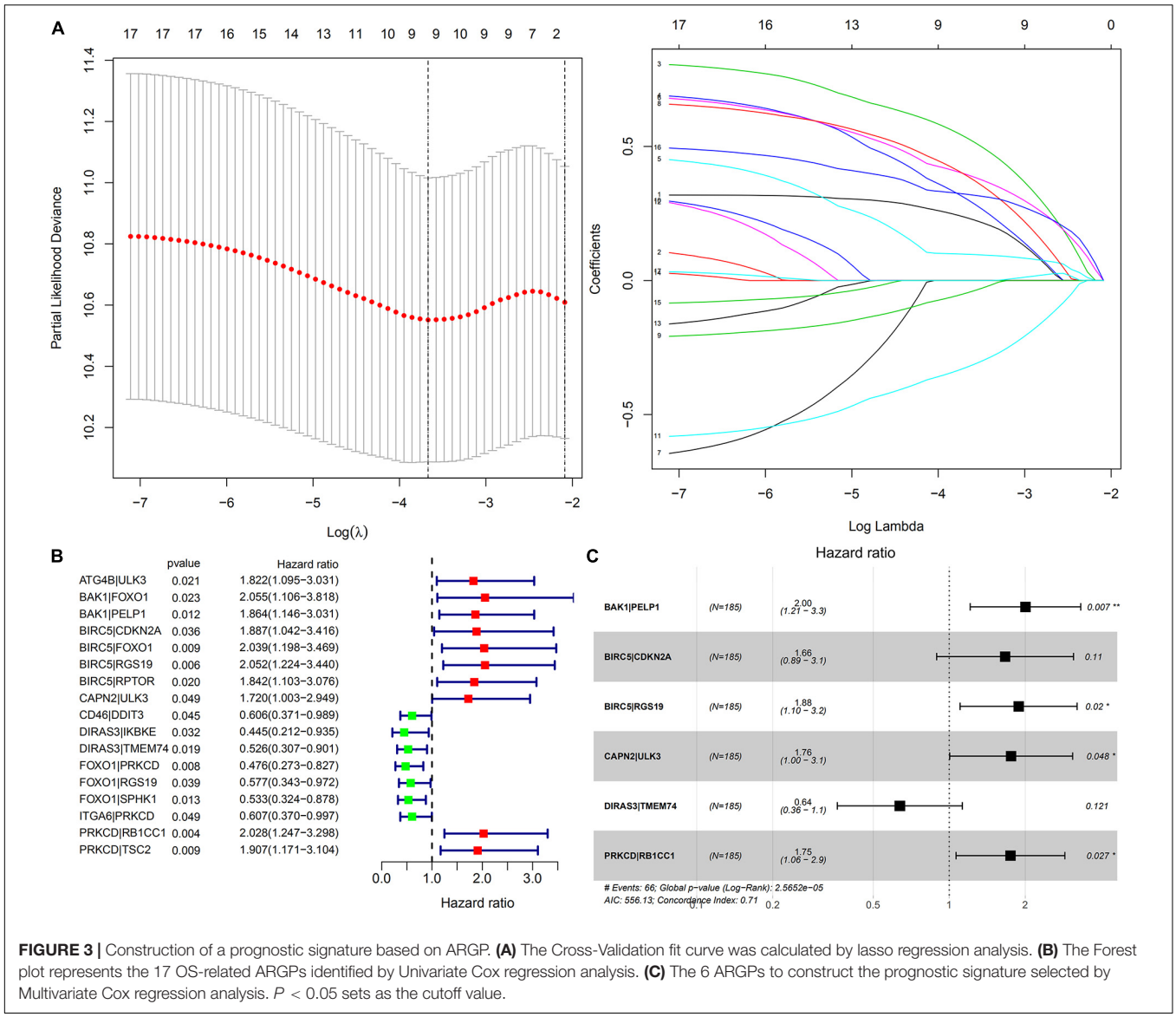


FIGURE 3 | Construction of a prognostic signature based on ARGP. **(A)** The Cross-Validation fit curve was calculated by lasso regression analysis. **(B)** The Forest plot represents the 17 OS-related ARGPs identified by Univariate Cox regression analysis. **(C)** The 6 ARGPs to construct the prognostic signature selected by Multivariate Cox regression analysis. $P < 0.05$ sets as the cutoff value.

prediction. In the test set, subgroup analysis performed well in OS prediction, except for patients in subgroups of female, G1-2, Stage III-IV, and T3-4.

As shown in **Supplementary Figure 4**, risk score of both the test and TCGA sets was associated with the pathological stage (I-II and III-IV), and tumor grade (G1-2 and G3-4), and the risk-score of training set was also correlated with the tumor grade (G1-2 and G3-4).

Analysis of Cytotoxic T-Lymphocyte-Associated Protein 4 (CTLA4) and LC3 Expression Levels Between the High-Risk and Low-Risk Groups

To further explore the role of autophagy and immune processes in the OS of patients with HCC, analysis of CTLA4 and LC3 expression levels between high-risk and a low-risk group of the ARGP prognostic signature was performed. As shown in **Figure 6**, the expression level of CTLA4 in the low-risk group of the training, test, TCGA, and ICGC sets was lower than that of the high-risk group, while the expression level of LC3 was higher in the low-risk group in the training, test, and TCGA sets.

Drug Sensitivity Analysis Between High and Low-Risk Groups

As shown in **Table 2**, the IC50 value of axitinib, rapamycin, temsirolimus, docetaxel, and metformin in the low-risk group was lower than that in the high-risk group, indicating that these 5 small molecule drugs were more effective for patients in the low-risk group. However, the IC50 value of bleomycin was greater in the low-risk group than in the high-risk group, indicating that bleomycin was more effective for patients in the high-risk group.

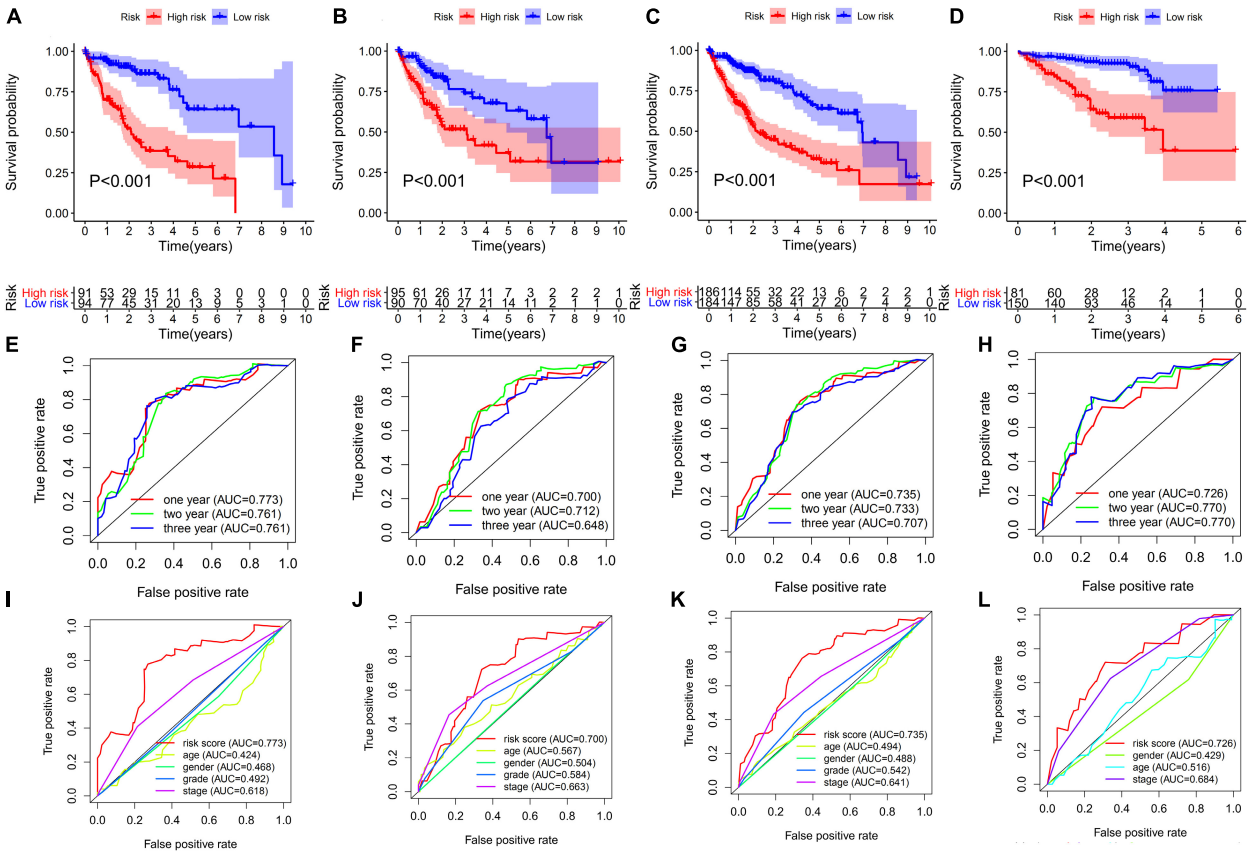


FIGURE 4 | Evaluation and validation of the risk score of the 6-ARGP signature. Survival analysis and ROC analysis based on risk score in the training set **(A,E,I)**, test set **(B,F,J)**, TCGA set **(C,G,K)**, and ICGC set **(D,H,L)**, respectively.

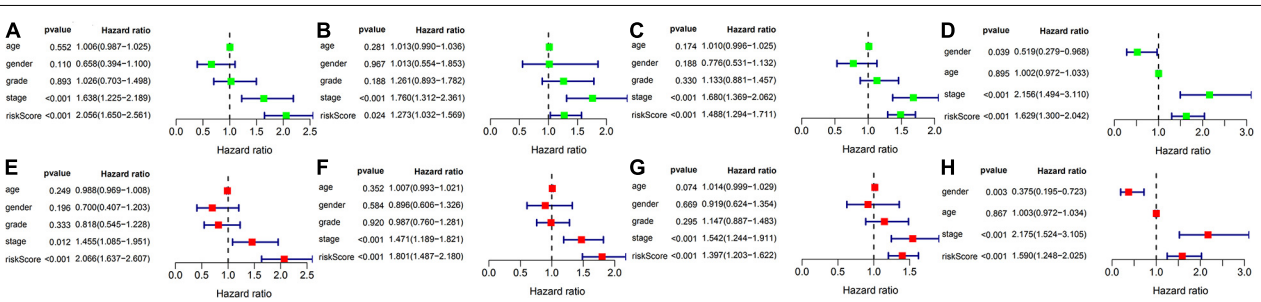


FIGURE 5 | Univariate Cox and Multivariate Cox regression analysis of risk score and clinicopathologic factors in different cohorts. **(A–D)** Forest plot represents the results of univariate Cox regression analysis in the training set, test set, TCGA set and ICGC set, respectively. **(E–H)** Forest plot represents the results of multivariate Cox regression analysis in the training set, test set, TCGA set and ICGC set, respectively.

Comparative Analysis of Predictive Values Between ARGP Signature and Published Gene Signatures

As shown in **Figure 7**, the AUC at 1-, 3-, 5-years OS compared between the ARGP signature and 3 published gene signatures in the same TCGA set, which included a 6-gene signature (FangGeneSig), a 7-gene signature (XieGeneSig), and an 8-gene signature (XuGeneSig) (Fang and Chen, 2020; Xie et al., 2020; Xu et al., 2020). Although the AUC of the ARGP signature at 1-year OS was 0.727, which was lower than that of XieGeneSig (0.732) and XuGeneSig (0.737), the AUC of the ARGP signature at 3- and 5-year OS was 0.717 and 0.672, respectively, which was higher than that of FangGeneSig (0.606 and 0.623), XieGeneSig (0.667 and 0.648) and XuGeneSig (0.679 and 0.643). These results suggest that the predictive value of our signature was more accurate than that of the aforementioned 3 published gene signatures in longer OS prediction.

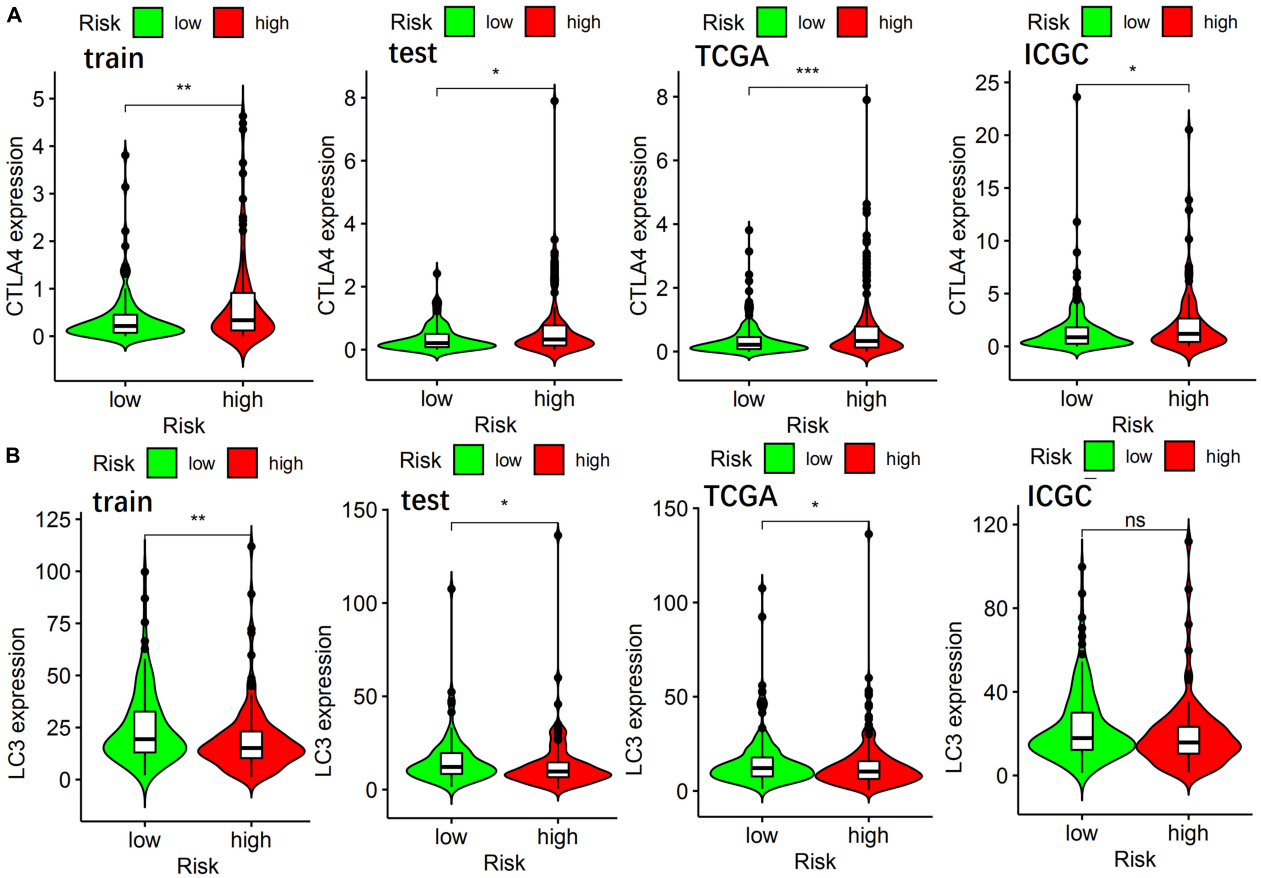


FIGURE 6 | The relationship of ARGP risk group with CTLA4, LC3 expression in different data sets. **(A)** The differential expression of CTLA4 in the training set, test set, TCGA set, and ICGC set, respectively. **(B)** The differential expression of LC3 in the training set, test set, TCGA set, and ICGC set, respectively.

TABLE 2 | The relationship of ARGP risk group with small molecule drug therapy response in different data sets.

Drugs (IC50)	Group	Train	Test	TCGA	ICGC	*P*-value
Axitinib	High risk	3.95 (– 4 to 7.5)	3.2 (– 2.8 to 6.3)	3.75 (– 4.1 to 7.7)	3.8 (– 2.8 to 6.6)	<0.05
	Low risk	3 (– 3.6 to 7.6)	2.7 (– 2.6 to 6)	3 (– 3.8 to 7.6)	2.9 (– 4.9 to 6.5)	
Rapamycin	High risk	–0.15 (– 10 to 5)	0.04 (– 8 to 3.5)	–0.03 (– 9 to 4.3)	0.1 (– 2.5 to 2.4)	<0.05
	Low risk	–0.8 (– 4.8 to 4)	–1 (– 6.8 to 3.6)	–0.8 (– 7.5 to 4)	–0.5 (– 4.9 to 2.7)	
Temsirolimus	High risk	–0.1 (– 8 to 9.6)	–0.1 (– 8.9 to 9.1)	–0.1 (– 9 to 9.6)	0.3 (– 7 to 9.5)	<0.05
	Low risk	–2.3 (– 9 to 8.7)	–2.4 (– 9 to 9.3)	–2.3 (– 9 to 9.3)	–2.2 (– 9.6 to 7.8)	
Docetaxel	High risk	–5 (– 8.3 to 0)	–4.9 (– 8.7 to 0)	–5 (– 9 to 0)	–4.9 (– 6.3 to 0)	<0.05
	Low risk	–5.7 (– 9.4 to 1)	–5.7 (– 8.7 to 0)	–5.7 (– 9.4 to 1)	–5.8 (– 9.8 to 0)	
Metformin	High risk	11.5 (7 to 14)	11 (6.5 to 15)	11.5 (6.5 to 15)	11.3 (9 to 13.5)	<0.05
	Low risk	10 (7.5 to 13.5)	10 (6.5 to 12.5)	10 (6.5 to 13.5)	10.2 (7.5 to 13)	
Bleomycin	High risk	0.3 (– 11 to 9)	0.5 (– 10 to 7.5)	0.3 (– 11 to 9)	0.1 (– 9.6 to 5.4)	<0.05
	Low risk	4.3 (– 7 to 10)	4.5 (– 10 to 12)	4.5 (– 10 to 12)	3.3 (– 9 to 10.3)	

IC50, the half maximal inhibitory concentration; ARGP, autophagy-related gene pair.

Establishment of a Nomogram to Predict the OS of HCC Patients

The TNM stage and risk score of the signature could potentially be independent prognostic factors, as revealed through multivariate Cox regression analysis. Hence, a nomogram that consists of the TNM stage and risk score was constructed, to predict the 1-, 3-, and 5-years OS among patients with HCC. As shown in **Figure 8**, calibration curves of the nomogram at 1-, 3-, and 5-years OS were proximal to the actual line, indicating that our nomogram performed well in predicting the OS of patients with HCC.

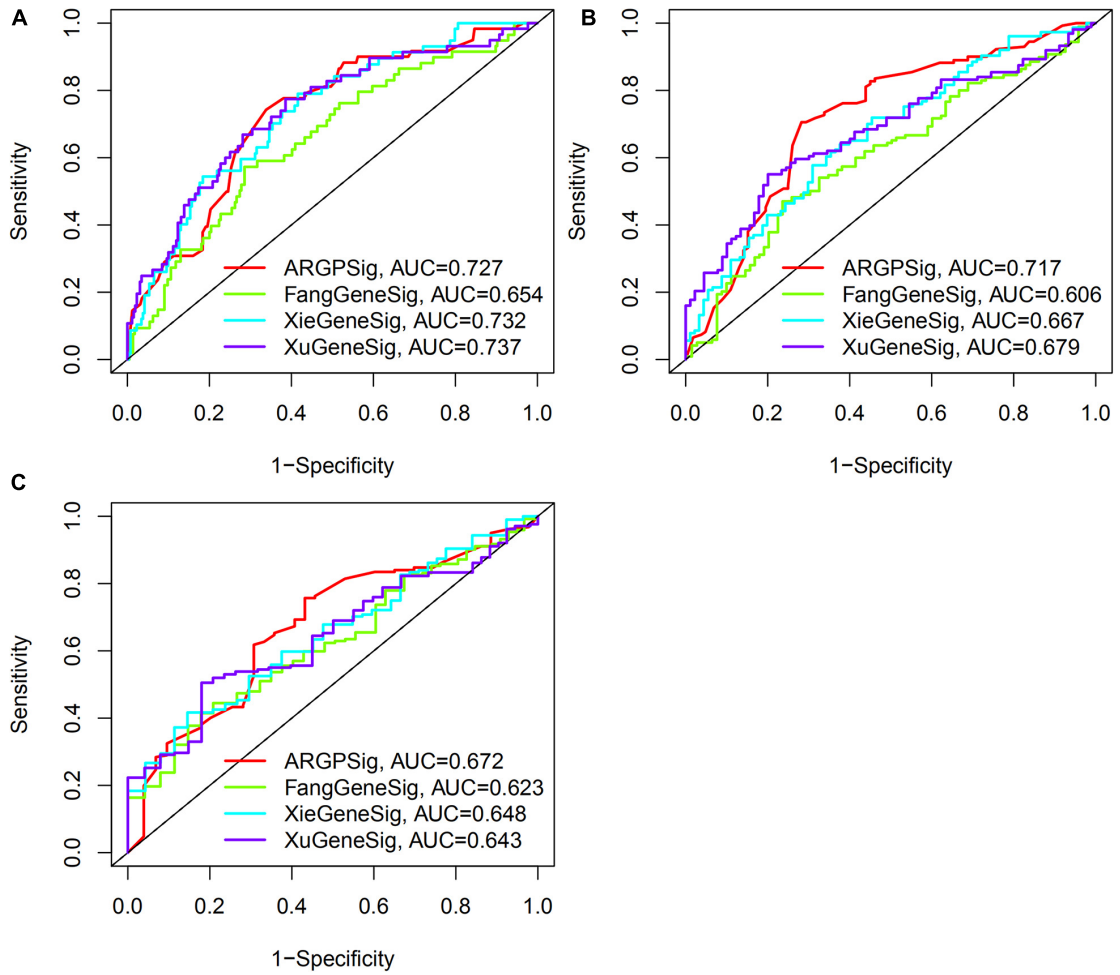


FIGURE 7 | Comparative analysis of the predictive value of ARGP signature with published gene signature. **(A–C)** The AUC at 1-, 3-, 5-years OS of ARGP signature, Fang gene signature, Xie gene signature, and Xu gene signature, respectively.

DISCUSSION

HCC is one of the most prevalent malignancies worldwide. Patients with HCC are suffering high risk for recurrence and metastasis. Over the past decades, *the therapeutic* effect of surgery for HCC and *adjuvant* therapy remained unsatisfactory. With studies on autophagy, attention has been shifting to studies on novel biomarkers for tumor autophagy for estimating treatment responses and survival outcomes.

Hence, based on comprehensive bioinformatics analysis, 6 OS-related ARGPs that could potentially serve as effective biomarkers for HCC were identified. The molecular function of these ARGs was primarily related to the immune function and autophagy, indicating that the potential molecular mechanisms underlying the effect of these ARGs on HCC prognosis are related to immune and autophagy. Furthermore, this 6-ARGP prognostic signature could help doctors to classify patients with HCC into 2 subgroups with significantly different OS. The ROC of the prognostic signature indicated moderate predictive accuracy in OS prediction for patients with HCC, and it revealed an adequate discrimination ability of OS in subgroup analysis, indicating that the prognostic signature was applicable for different subgroups of patients with HCC. Multivariate Cox regression analysis revealed that the risk score of the prognostic signature could serve as an independent prognostic factor. Moreover, a nomogram consisting of the TNM stage and risk score was constructed to better predict the OS of HCC patients more clearly, which performed well in OS prediction.

Although, some ARG based signatures have been published, the methods of our study are different. Most of them only choose ARGs to construct a prognostic signature and validated the signature in other data sets. While we performed pairwise comparison analysis to identify ARGPs and then construct a prognostic signature based on ARGPs. Compared with ARGs, ARGPs can reduce the batch effects when validating the prognostic signature in other data sets, such as ICGC, GEO data set, etc. So, the validated results are more reliable and accurate. In addition, we validate the signature both externally and internally, while the published articles only validate the signature externally. We also perform subgroup analysis to

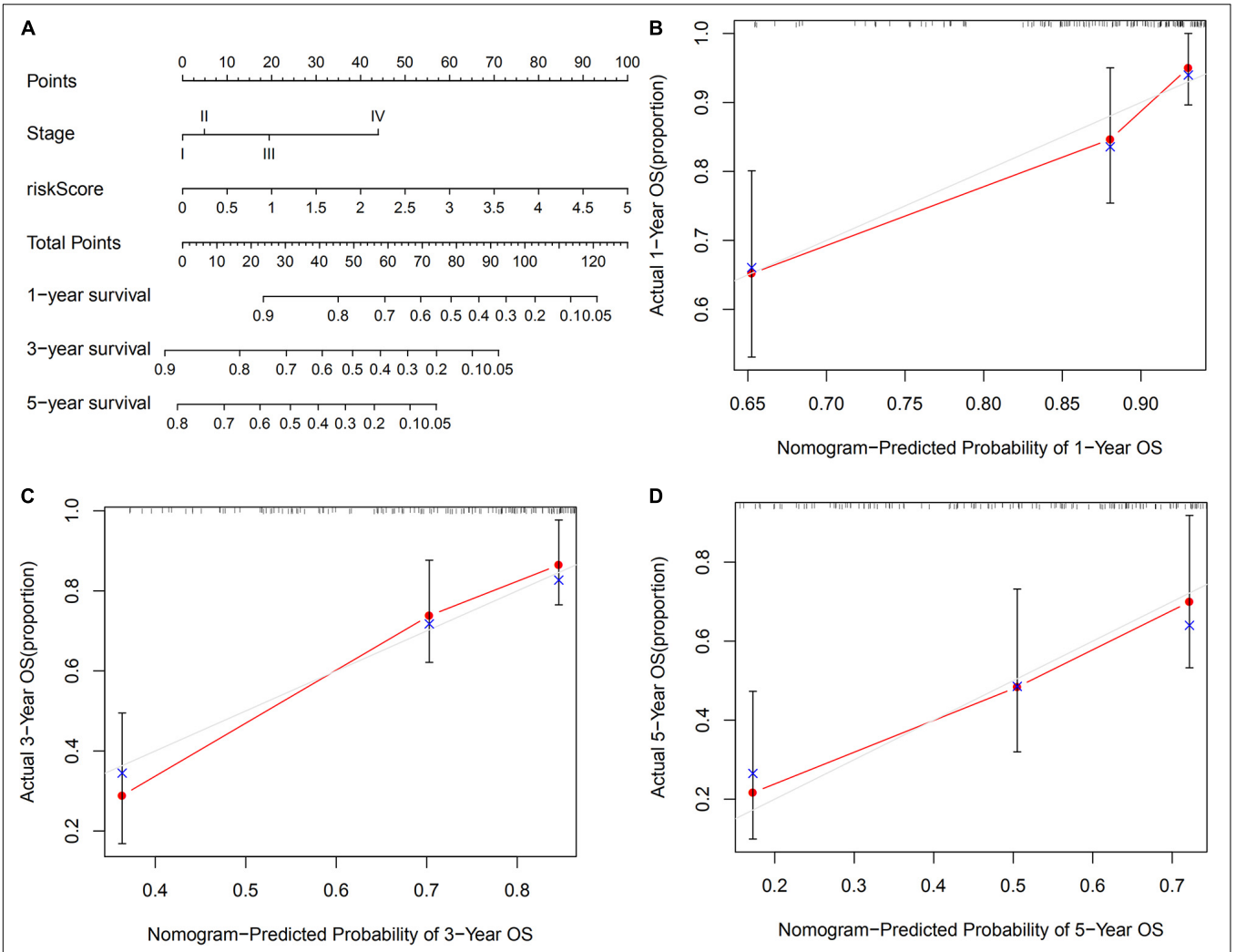


FIGURE 8 | Construction of a nomogram based on risk score and TNM stage in TCGA set. **(A)** A nomogram consists of risk score and stage to predict the OS of HCC patients at 1-, 3-, and 5-years. **(B–D)** Calibration curves to validate the prediction value of nomogram at 1-, 3-, and 5-year OS, respectively.

further validate the predictive value of the signature for HCC patients in different clinical features and the results show the signature performs well. To guide the therapy of HCC, a drug sensitivity analysis is performed to identify potential small molecular drugs and 6 drugs are identified. So, the results of our study may be more clinically meaningful compared with published articles, and further researches are needed.

The 6-ARGP signature highlighted 11 ARGs, including BAK1, PELP1, BIRC5, CDKN2A, RGS19, CAPN2, ULK3, DIRAS3, TMEM74, PRKCD, and RB1CC1. Most of these ARGs are correlated with the prognosis of HCC or other cancers. BAK1 is an important cell death regulator that can initiate mitochondria-mediated apoptosis and is reportedly correlated with the occurrence of several cancers (Slager et al., 2012; Wang Y. D. et al., 2013; Marcotte et al., 2017). PELP1 serves as a proto-oncogene in all hormone-responsive cancers, including breast, prostate cancers, ovarian, and other cancers (Dimple et al., 2008; Cortez et al., 2012; Yang et al., 2012; Daniel et al., 2015). BIRC5 overexpression was reported in breast cancer, lung adenocarcinoma, and neuroblastic malignance specimens (Hagenbuchner et al., 2016; Hamy et al., 2016; Cao et al., 2019). In a rat model of HCC, combination therapy with a CDKN2A inhibitor and transarterial chemoembilization promoted cancer cell necrosis (Gade et al., 2017). Wang Y. et al. (2013) reported that RGS19 suppressed the occurrence of non-small-cell carcinoma by downregulating Ras. ULK3 is reportedly involved in cancer-associated fibroblast conversion by activating 2 main signaling pathways (Goruppi et al., 2017). Recent studies have revealed that CAPN2 plays a vital role in tumorigenesis and tumor progression in breast cancer, and colon cancer (Storr et al., 2012; Mo et al., 2019). DIRAS3 is downregulated in 60% of ovarian cancers and negatively related to progression-free survival (Yu et al., 2003; Rosen et al., 2004). Sun Y. et al. (2017) reported that TMEM74 promotes tumor cell survival by inducing autophagy by interacting with ATG16L1 and ATG9A. PRKCD is downregulated in HCC cells

and PRKCD upregulation can suppress the viability of HCC cells (Nambotin et al., 2011). RB1CC1 also called FIP200, is crucial in autophagy and is associated with the prognosis of and drug resistance in multiple cancers, including HCC (Yeo et al., 2020). Our results show that the aforementioned ARGs are correlated with the prognosis of HCC; However, the underlying molecular mechanism of these ARGs in HCC prognosis requires further investigation.

To explore the mechanisms through which the ARGP signature effectively stratifies patients with HCC, the expression profiles of CTLA4 and LC3 between the high- and low-risk groups was performed. CTLA4 is a receptor on the surface of activated T cells and act as an effective immune therapy checkpoint, whose functions are to inhibit the production of IL-2, proliferation of T cells, and cell cycle (Fraser et al., 1999; Intlekofer and Thompson, 2013; Bhandaru and Rotte, 2019). LC3 is essential for the execution of autophagy. Therefore it is a widely accepted marker for autophagy, which can be a potential target for anticancer therapy (Schaaf et al., 2016). These results show that the low-risk group has a lower expression level of CTLA4 and a higher expression level of LC3. CTLA4 and LC3 dysregulation may be responsible for the difference in survival outcomes between the high- and low-risk groups.

Based on our results, we hypothesize that immunological and autophagy-related small-molecule drugs might be used to treat patients with HCC. Hence, a drug sensitivity analysis was performed to explore potentially effective small-molecule drugs for patients with HCC. 6 drugs were identified, including axitinib, rapamycin, temsirolimus, docetaxel, metformin, and bleomycin. Recent studies have reported that axitinib serves as a multi-receptor tyrosine kinase inhibitor to treat multiple cancers, and axitinib inhibits the VEGF receptor, platelet-derived growth factor receptor, and epidermal growth factor receptor (Ongkeko et al., 2005; Bran et al., 2009; Lawrence et al., 2015). Lin et al. (2020) reported that axitinib is an effective second-line therapy drug for advanced patients with HCC, who failed sorafenib therapy. Rapamycin and temsirolimus belong to Rapalogs, have been reported to suppress proliferation and promote autophagy in HCC cells by targeting the mTOR signaling pathway (Lu et al., 2020). In the work of Hui et al. (2010), rapamycin and temsirolimus can significantly inhibit the growth and metastasis of PLC/PRF/5 human HCC cells. Docetaxel belongs to the taxane family, and preclinical studies have reported the anticancer potential of docetaxel in suppressing HCC cell proliferation. For example, docetaxel treatment can reduce the tumor size in a nude mouse model of HCC, and suppress the proliferation capacity of the HepG2 cell line (Zhu et al., 2016). In addition, Zhang et al. (2019) revealed that docetaxel can induce HCC cell apoptosis by inhibiting the PI3K/AKT signaling pathway. Metformin is used to treat not only diabetes but also tumors. For instance, metformin can inhibit tumor cell proliferation by targeting mTOR complex 1 via an AMPK-independent mechanism (Kalender et al., 2010). DeWaal et al. (2018) reported that metformin can induce cell cycle arrest in HCC cells by targeting mTOR complex 1 through an AMPK-independent mechanism as well. These studies indicate that the AMPK-independent anticancer activities of metformin may be a novel finding Overall, this study and other preclinical studies have revealed that these small- molecule drugs can be potentially effective drugs in treating HCC, and further clinical trials are needed to validate these results.

In this study, we identified an ARGP based prognostic signature that performs well in predicting the OS of patients with HCC. For all we know, this is the first reported ARGP-based signature for HCC. However, our study has several limitations. First, the results were biased to an extent because we used fewer non-tumor specimens than HCC specimens. Second, the underlying molecular mechanisms of HCC in this study have not been determined on the basis of *in vitro* and *in vivo* studies. Further studies are needed to validate these results.

CONCLUSION

An ARGP prognostic signature was identified and validated in different data sets, this signature performed better in OS prediction of HCC in comparison with 3 previously published gene signatures. Furthermore, 6 small-molecule drugs were identified to be potentially effective drugs in treating HCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The data that support the findings of this study are openly available in TCGA data portal (https://portal.gdc.cancer.gov/), ICGC data base (https://dcc.icgc.org/), and HADb (http://www.autophagy.lu/).

AUTHOR CONTRIBUTIONS

SQL: conceptualization and manuscript revision. GPZ, ZBS, and YY: data curation. ZBS: data analysis and figure plot. ZBS and GPZ: manuscript writing. All authors final approval of manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (82072663).

ACKNOWLEDGMENTS

We would like to thank the TCGA, ICGC, and the Human Autophagy Database for the availability of the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.689801/full#supplementary-material

Supplementary Figure 1 | Risk score analysis of the prognostic signature in different cohorts. **(A)** Heatmap plot represents the expression of the 6 ARGPs between high- and low-risk groups in the training set, test set, TCGA set, and ICGC set, respectively. **(B)** Survival status of patients in the train set, test set, TCGA set, and ICGC set, respectively. **(C)** The rank of the risk score in the training set, test set, TCGA set, and ICGC set, respectively.

Supplementary Figure 2 | Subgroup analysis in the training set and ICGC set. **(A,B)** The KM curve represents the OS of high- and low-risk groups in different subgroups of the training set. **(C)** The KM curve represents the OS of high- and low-risk groups in different subgroups of the ICGC set.

Supplementary Figure 3 | Subgroup analysis in the test set and TCGA set. **(A,B)** The KM curve represents the OS of high- and low-risk groups in different subgroups of the test set. **(C,D)** The KM curve represents the OS of high- and low-risk groups in different subgroups of the TCGA set.

Supplementary Figure 4 | Clinical relevance analysis of risk score. **(A–D)** The relationship of stage group with risk score in the training set, test set, TCGA set, and ICGC set, respectively. **(E–G)** The relationship of grade group with risk score in the training set, test set, and TCGA set, respectively.

REFERENCES

Aalen, O. O. (1988). Heterogeneity in survival analysis. *Stat. Med.* 7, 1121–1137. doi: 10.1002/sim.4780071105

Bandyopadhyay, S., and Mallick, K. (2014). A new path based hybrid measure for gene ontology similarity. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11, 116–127. doi: 10.1109/tcbb.2013.149

Bhandaru, M., and Rotte, A. (2019). Monoclonal antibodies for the treatment of melanoma: present and future strategies. *Methods Mol. Biol.* 1904, 83–108. doi: 10.1007/978-1-4939-8958-4_4

Bran, B., Bran, G., Hörmann, K., and Riedel, F. (2009). The platelet-derived growth factor receptor as a target for vascular endothelial growth factor-mediated anti-angiogenetic therapy in head and neck cancer. *Int. J. Oncol.* 34, 255–261.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Cao, Y., Zhu, W., Chen, W., Wu, J., Hou, G., and Li, Y. (2019). Prognostic value of BIRC5 in lung adenocarcinoma lacking EGFR, KRAS, and ALK mutations by integrated bioinformatics analysis. *Dis. Markers* 2019:5451290. doi: 10.1155/2019/5451290

Cortez, V., Mann, M., Tekmal, S., Suzuki, T., Miyata, N., Rodriguez-Aguayo, C., et al. (2012). Targeting the PELP1-KDM1 axis as a potential therapeutic strategy for breast cancer. *Breast Cancer Res.* 14:R108. doi: 10.1186/bcr3229

Daniel, A. R., Gaviglio, A. L., Knutson, T. P., Ostrander, J. H., D'Assoro, A. B., Ravindranathan, P., et al. (2015). Progesterone receptor-B enhances estrogen responsiveness of breast cancer cells via scaffolding PELP1- and estrogen receptor-containing transcription complexes. *Oncogene* 34, 506–515. doi: 10.1038/onc.2013.579

DeWaal, D., Nogueira, V., Terry, A. R., Patra, K. C., Jeon, S. M., Guzman, G., et al. (2018). Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin. *Nat. Commun.* 9:446. doi: 10.1038/s41467-017-02733-4

Dimple, C., Nair, S. S., Rajhans, R., Pitcheswara, P. R., Liu, J., Balasenthil, S., et al. (2008). Role of PELP1/MNAR signaling in ovarian tumorigenesis. *Cancer Res.* 68, 4902–4909. doi: 10.1158/0008-5472.can-07-5698

Fan, Q., Yang, L., Zhang, X., Ma, Y., Li, Y., Dong, L., et al. (2018). Autophagy promotes metastasis and glycolysis by upregulating MCT1 expression and Wnt/β-catenin signaling pathway activation in hepatocellular carcinoma cells. *J. Exp. Clin. Cancer Res.* 37:9. doi: 10.1186/s13046-018-0673-y

Fang, Q., and Chen, H. (2020). Development of a Novel Autophagy-Related Prognostic Signature and Nomogram for Hepatocellular Carcinoma. *Front Oncol* 10:591356. doi: 10.3389/fonc.2020.591356

Finn, R. S., Zhu, A. X., Farah, W., Almasri, J., Zaiem, F., Prokop, L. J., et al. (2018). Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: a systematic review and meta-analysis. *Hepatology* 67, 422–435. doi: 10.1002/hep.29486

Fraser, J. H., Rincón, M., McCoy, K. D., and Le Gros, G. (1999). CTLA4 ligation attenuates AP-1, NFAT and NF-kappaB activity in activated T cells. *Eur. J. Immunol.* 29, 838–844.

Gade, T. P. F., Tucker, E., Nakazawa, M. S., Hunt, S. J., Wong, W., Krock, B., et al. (2017). Ischemia induces quiescence and autophagy dependence in hepatocellular carcinoma. *Radiology* 283, 702–710. doi: 10.1148/radiol.2017160728

Goruppi, S., Procopio, M. G., Jo, S., Clocchiatti, A., Neel, V., and Dotto, G. P. (2017). The ULK3 kinase is critical for convergent control of cancer-associated fibroblast activation by CSL and GLI. *Cell Rep.* 20, 2468–2479. doi: 10.1016/j.celrep.2017.08.048

Hagenbuchner, J., Kiechl-Kohlendorfer, U., Obexer, P., and Ausserlechner, M. J. (2016). BIRC5/Survivin as a target for glycolysis inhibition in high-stage neuroblastoma. *Oncogene* 35, 2052–2061. doi: 10.1038/onc.2015.264

Hamy, A. S., Bieche, I., Lehmann-Che, J., Scott, V., Bertheau, P., Guinebretière, J. M., et al. (2016). BIRC5 (survivin): a pejorative prognostic marker in stage II/III breast cancer with no response to neoadjuvant chemotherapy. *Breast Cancer Res. Treat.* 159, 499–511. doi: 10.1007/s10549-016-3961-2

Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J. X., et al. (2013). Gene-pair expression signatures reveal lineage control. *Nat. Methods* 10, 577–583. doi: 10.1038/nmeth.2445

Huang, F., Wang, B. R., and Wang, Y. G. (2018). Role of autophagy in tumorigenesis, metastasis, targeted therapy and drug resistance of hepatocellular carcinoma. *World J. Gastroenterol.* 24, 4643–4651. doi: 10.3748/wjg.v24.i41.4643

Hui, I. C., Tung, E. K., Sze, K. M., Ching, Y. P., and Ng, I. O. (2010). Rapamycin and CCI-779 inhibit the mammalian target of rapamycin signalling in hepatocellular carcinoma. *Liver Int.* 30, 65–75. doi: 10.1111/j.1478-3231.2009.02117.x

Intlekofer, A. M., and Thompson, C. B. (2013). At the bench: preclinical rationale for CTLA-4 and PD-1 blockade as cancer immunotherapy. *J. Leukoc. Biol.* 94, 25–39. doi: 10.1189/jlb.1212621

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558. doi: 10.1038/s41591-018-0136-1

Kalender, A., Selvaraj, A., Kim, S. Y., Gulati, P., Brûlé, S., Viollet, B., et al. (2010). Metformin, independent of AMPK, inhibits mTORC1 in a rag GTPase-dependent manner. *Cell Metab.* 11, 390–401. doi: 10.1016/j.cmet.2010.03.014

Lawrence, M. S., Sougnez, C., Lichtenstein, L., Cibulskis, K., Lander, E., Gabriel, S. B., et al. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517, 576–582. doi: 10.1038/nature14129

Lin, Z. Z., Chen, B. B., Hung, Y. P., Huang, P. H., Shen, Y. C., Shao, Y. Y., et al. (2020). A multicenter phase II study of second-line axitinib for patients with advanced hepatocellular carcinoma failing first-line sorafenib monotherapy. *Oncologist* 25, e1280–e1285. doi: 10.1634/theoncologist.2020-0143

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, 400–416.e411. doi: 10.1016/j.cell.2018.02.052

Liu, K., Lee, J., and Ou, J. J. (2018). Autophagy and mitophagy in hepatocarcinogenesis. *Mol. Cell Oncol.* 5:e1405142. doi: 10.1080/23723556.2017.1405142

Lu, X., Paliogiannis, P., Calvisi, D. F., and Chen, X. (2020). Role of the mammalian target of rapamycin pathway in liver cancer: from molecular genetics to targeted therapies. *Hepatology* 73 Suppl 1(Suppl 1), 49–61. doi: 10.1002/hep.31310

Marcotte, E. L., Pankratz, N., Amatruda, J. F., Frazier, A. L., Krailo, M., Davies, S., et al. (2017). Variants in BAK1, SPRY4, and GAB2 are associated with

pediatric germ cell tumors: a report from the children's oncology group. *Genes Chromosomes. Cancer* 56, 548–558. doi: 10.1002/gcc.22457

Mo, S., Dai, W., Xiang, W., Li, Y., Feng, Y., Zhang, L., et al. (2019). Prognostic and predictive value of an autophagy-related signature for early relapse in stages I-III colon cancer. *Carcinogenesis* 40, 861–870. doi: 10.1093/carcin/bgz031

Nambotin, S. B., Lefrancois, L., Sainsily, X., Berthillon, P., Kim, M., Wands, J. R., et al. (2011). Pharmacological inhibition of Frizzled-7 displays anti-tumor properties in hepatocellular carcinoma. *J. Hepatol.* 54, 288–299. doi: 10.1016/j.jhep.2010.06.033

Ongkeko, W. M., Altuna, X., Weisman, R. A., and Wang-Rodriguez, J. (2005). Expression of protein tyrosine kinases in head and neck squamous cell carcinomas. *Am. J. Clin. Pathol.* 124, 71–76. doi: 10.1309/btln5wtmj3pcnrrc

Raza, A., and Sood, G. K. (2014). Hepatocellular carcinoma review: current treatment, and evidence-based medicine. *World J. Gastroenterol.* 20, 4115–4127. doi: 10.3748/wjg.v20.i15.4115

Rosen, D. G., Wang, L., Jain, A. N., Lu, K. H., Luo, R. Z., Yu, Y., et al. (2004). Expression of the tumor suppressor gene ARHI in epithelial ovarian cancer is associated with increased expression of p21WAF1/CIP1 and prolonged progression-free survival. *Clin. Cancer Res.* 10, 6559–6566. doi: 10.1158/1078-0432.ccr-04-0698

Schaaf, M. B., Keulers, T. G., Vooijs, M. A., and Rouschop, K. M. (2016). LC3/GABARAP family proteins: autophagy-(un)related functions. *FASEB J.* 30, 3961–3978. doi: 10.1096/fj.201600698R

Slager, S. L., Skibola, C. F., Di Bernardo, M. C., Conde, L., Broderick, P., McDonnell, S. K., et al. (2012). Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood* 120, 843–846. doi: 10.1182/blood-2012-03-413591

Storr, S. J., Lee, K. W., Woolston, C. M., Safuan, S., Green, A. R., Macmillan, R. D., et al. (2012). Calpain system protein expression in basal-like and triple-negative invasive breast cancer. *Ann. Oncol.* 23, 2289–2296. doi: 10.1093/annonc/mds176

Sun, T., Liu, H., and Ming, L. (2017). Multiple roles of autophagy in the sorafenib resistance of hepatocellular carcinoma. *Cell Physiol. Biochem.* 44, 716–727. doi: 10.1159/000485285

Sun, Y., Chen, Y., Zhang, J., Cao, L., He, M., Liu, X., et al. (2017). TMEM74 promotes tumor cell survival by inducing autophagy via interactions with ATG16L1 and ATG9A. *Cell Death Dis.* 8:e3031. doi: 10.1038/cddis.2017.370

Wang, N., Lefaudeux, D., Mazumder, A., Li, J. J., and Hoffmann, A. (2021). Identifying the combinatorial control of signal-dependent transcription factors. *PLoS Comput. Biol.* 17:e1009095. doi: 10.1371/journal.pcbi.1009095

Wang, Y., Tong, Y., Tso, P. H., and Wong, Y. H. (2013). Regulator of G protein signaling 19 suppresses Ras-induced neoplastic transformation and tumorigenesis. *Cancer Lett.* 339, 33–41. doi: 10.1016/j.canlet.2013.07.025

Wang, Y. D., Cai, N., Wu, X. L., Cao, H. Z., Xie, L. L., and Zheng, P. S. (2013). OCT4 promotes tumorigenesis and inhibits apoptosis of cervical cancer cells by miR-125b/BAK1 pathway. *Cell Death Dis.* 4:e760. doi: 10.1038/cddis.2013.272

Xie, H., Liu, S., Zhang, Z., Chen, P., and Tao, Y. (2020). A novel seven-gene signature as prognostic biomarker in hepatocellular carcinoma. *J. Cancer* 11, 5768–5781. doi: 10.7150/jca.44573

Xu, D., Wang, Y., Zhou, K., Wu, J., Zhang, Z., Zhang, J., et al. (2020). Development and validation of a novel 8 immune gene prognostic signature based on the immune expression profile for hepatocellular carcinoma. *Oncol. Targets Ther.* 13, 8125–8140. doi: 10.2147/ott.s263047

Yang, L., Ravindranathan, P., Ramanan, M., Kapur, P., Hammes, S. R., Hsieh, J. T., et al. (2012). Central role for PELP1 in nonandrogenic activation of the androgen receptor in prostate cancer. *Mol. Endocrinol.* 26, 550–561. doi: 10.1210/me.2011-1101

Yeo, S. K., Wang, C., and Guan, J. L. (2020). Role of FIP200 in inflammatory processes beyond its canonical autophagy function. *Biochem. Soc. Trans.* 48, 1599–1607. doi: 10.1042/bst20191156

Yu, Y., Fujii, S., Yuan, J., Luo, R. Z., Wang, L., Bao, J., et al. (2003). Epigenetic regulation of ARHI in breast and ovarian cancer cells. *Ann. N. Y. Acad. Sci.* 983, 268–277. doi: 10.1111/j.1749-6632.2003.tb05981.x

Zhang, X., Shao, J., Li, X., Cui, L., and Tan, Z. (2019). Docetaxel promotes cell apoptosis and decreases SOX2 expression in CD133-expressing hepatocellular carcinoma stem cells by suppressing the PI3K/AKT signaling pathway. *Oncol. Rep.* 41, 1067–1074. doi: 10.3892/or.2018.6891

Zhu, D., Tao, W., Zhang, H., Liu, G., Wang, T., Zhang, L., et al. (2016). Docetaxel (DTX)-loaded polydopamine-modified TPGS-PLA nanoparticles as a targeted drug delivery system for the treatment of liver cancer. *Acta Biomater.* 30, 144–154. doi: 10.1016/j.actbio.2015.11.031

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 August 2021
doi: 10.3389/fgene.2021.696912

Tumor Expression Profile Analysis Developed and Validated a Prognostic Model Based on Immune-Related Genes in Bladder Cancer

Bingqi Dong[1†], Jiaming Liang[2,3†], Ding Li[4†], Wenping Song[4], Shiming Zhao[1], Yongkang Ma[1], Jinbo Song[1], Mingkai Zhu[1] and Tiejun Yang[1]*

[1]Department of Urology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2]The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [3]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, National Clinical Research Center for Respiratory Disease, Guangzhou, China, [4]Department of Pharmacy, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:
Xiangqian Guo,
Henan University, China

Reviewed by:
Shuangyu Lv,
Henan University, China
Jiateng Zhong,
Xinxiang Medical University, China

***Correspondence:**
Tiejun Yang
tiejunyang@126.com

†These authors have contributed equally to this work

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

Received: 18 April 2021
Accepted: 28 July 2021
Published: 27 August 2021

Citation:
Dong B, Liang J, Li D, Song W, Zhao S, Ma Y, Song J, Zhu M and Yang T (2021) Tumor Expression Profile Analysis Developed and Validated a Prognostic Model Based on Immune-Related Genes in Bladder Cancer. Front. Genet. 12:696912. doi: 10.3389/fgene.2021.696912

Background: Bladder cancer (BLCA) ranks 10th in incidence among malignant tumors and 6th in incidence among malignant tumors in males. With the application of immune therapy, the overall survival (OS) rate of BLCA patients has greatly improved, but the 5-year survival rate of BLCA patients is still low. Furthermore, not every BLCA patient benefits from immunotherapy, and there are a limited number of biomarkers for predicting the immunotherapy response. Therefore, novel biomarkers for predicting the immunotherapy response and prognosis of BLCA are urgently needed.

Methods: The RNA sequencing (RNA-seq) data, clinical data and gene annotation files for The Cancer Genome Atlas (TCGA) BLCA cohort were extracted from the University of California, Santa Cruz (UCSC) Xena Browser. The BLCA datasets GSE31684 and GSE32894 from the Gene Expression Omnibus (GEO) database were extracted for external validation. Immune-related genes were extracted from InnateDB. Significant differentially expressed genes (DEGs) were identified using the R package "limma," and Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for the DEGs were performed using R package "clusterProfiler." Least absolute shrinkage and selection operator (LASSO) regression analysis were used to construct the signature model. The infiltration level of each immune cell type was estimated using the single-sample gene set enrichment analysis (ssGSEA) algorithm. The performance of the model was evaluated with receiver operating characteristic (ROC) curves and calibration curves.

Results: In total, 1,040 immune-related DEGs were identified, and eight signature genes were selected to construct a model using LASSO regression analysis. The risk score of BLCA patients based on the signature model was negatively correlated with OS and the immunotherapy response. The ROC curve for OS revealed that the model had good accuracy. The calibration curve showed good agreement between the predictions and actual observations.

Conclusions: Herein, we constructed an immune-related eight-gene signature that could be a potential biomarker to predict the immunotherapy response and prognosis of BLCA patients.

Keywords: bladder cancer, immune-related signature, The Cancer Genome Atlas, Gene Expression Omnibus, immunotherapy

INTRODUCTION

Bladder cancer (BLCA) ranks 10th in incidence among malignant tumors and 6th in incidence among malignant tumors in males (Bray et al., 2018). The disease may present as muscularly invasive bladder cancer (MIBC), non-muscularly invasive bladder cancer (NMIBC) or as a metastatic form of other diseases (Tran et al., 2021). With the development of technologies such as electrosurgery, chemotherapy, and radical surgery, the overall survival (OS) rate of BLCA patients has greatly improved. However, primary BLCA are prone to recurrence after systematic treatment, the prognosis is not satisfactory (Rouprêt et al., 2021), and there are no clinically meaningful diagnostic markers. Based on the fact that the incidence of BLCA is gradually increasing, valuable biomarkers are particularly urgently needed.

Current studies have revealed that the tumor microenvironment is closely correlated with tumorigenesis, progression and prognosis, and immune cells of the tumor microenvironment exhibit complex interactions with tumor cells (Hanahan and Weinberg, 2011). As the landmark developments of immune checkpoint inhibitors (represented by anti-PD-1/PD-L1 antibodies), Immunotherapy is playing an increasingly important role in the treatment of BLCA (Powles et al., 2014; Pettenati and Ingersoll, 2018; Rouanne et al., 2018). Thus, immune molecules associated with the tumor microenvironment have a tremendous role in serving as prognostic markers for BLCA. Previous studies have proposed immune-related biomarkers in thyroid and ovarian cancers for risk stratification and clinical outcome prediction (Kim et al., 2018; Shen et al., 2019). A few studies have been conducted to assess the potential of immune-related genes to predict clinical outcomes and the immunotherapy response in BLCA (Qiu et al., 2020; Zhu et al., 2020; Lv et al., 2021), but the depth and results of these studies are not satisfactory.

Herein, the purpose of this study is to find suitable biomarkers of BLCA with high sensitivity and strong specificity and molecular targets that affect the clinicopathological process of BLCA, then provide an important reference for the diagnosis of BLCA. We identified immune-related DEGs for BLCA, and constructed an immune-related eight-gene signature model. The signature model showed good prognostic value for predicting OS and could be used to predict the immunotherapy response in BLCA patients.

MATERIALS AND METHODS

Data Collection and Preprocessing

The RNA sequencing (RNA-seq) data, probe annotation files and clinical data of the BLCA patients were extracted from The Cancer Genome Atlas (TCGA) and used to acquire the expression profiles of the BLCA patients. After screening, samples with no clinical data were excluded. A total of 406 tumor samples and 18 normal samples were included in the analysis. BLCA cohorts GSE31684 and GSE32894 were obtained from the Gene Expression Omnibus (GEO) database using the R package "GEOquery" (Davis and Meltzer, 2007). Immune-related genes were obtained from InnateDB.[1]

Identification of Differentially Expressed Genes and Functional Annotation

Significant DEGs between normal and BLCA samples were identified with screening criteria of adjusted value of $p < 0.05$ and $|\log \text{fold change (FC)}| > 1$ by the R package "limma" (Ritchie et al., 2015). Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for the DEGs were performed using R package "clusterProfiler" (Yu et al., 2012).

Estimation of the Infiltration Degree of Each Immune Cell in BLCA

A group of immune cell gene markers, consisting of 782 genes, which represent 28 immune cell types related to innate and adaptive immunity, were obtained from previous studies to estimate the infiltration level of different immune cell types in the tumor microenvironment (Charoentong et al., 2017). The assessed immune cell types included, B cells, natural killer (NK) cells, dendritic cells, myeloid-derived suppressor cells (MDSCs), neutrophils, and T cells. Subsequently, the expression

[1]https://www.innatedb.ca/

profiles of each sample were used to estimate the infiltration level of each immune cell type in BLCA using the single-sample gene set enrichment analysis (ssGSEA) algorithm with the R package "GSVA" (Hänzelmann et al., 2013).

Survival Analysis

Univariate Cox proportional hazards regression analysis was carried out to evaluate the association between the expression level of the immune-related DEGs and the OS of BLCA patients. Immune-related DEGs with a value of $p < 0.001$ based on the log-rank test were selected as candidate genes for construction of the prognostic model. The risk score of each patient was calculated based on the signature model and was used to evaluate the association between the gene signature model and the prognosis of BLCA patients. The samples were assigned to the high-risk or low-risk group based on the median risk score. Kaplan–Meier curves and log-rank tests were performed to compare the differences in OS and progression-free survival between the high-risk and low-risk groups. A statistically significant difference was defined as $p < 0.05$. Survival analysis and log-rank tests were performed using the R package "survival," and the R package "survminer" was used to plot the Kaplan–Meier curve.

Establishment and Evaluation of the Immune-Related Signature Model

The TCGA BLCA cohort was randomly divided into a training set ($n = 285$) and a testing set ($n = 121$) at a ratio of 7:3. The R package "glmnet" was used to perform LASSO regularization to reduce the coefficients from the training set. An immune related eight-gene model was constructed. The following formula was used:

$$\begin{aligned} Risk\ score = & -0.15744 \times DCHS1 + 0.05310 \\ & \times PTGIS - 0.32469 \times PTPN6 \\ & -0.07170 \times AIFM3 + 0.11454 \\ & \times FLRT2 + 0.04565 \times PCSK5 \\ & +0.15854 \times CLSTN2 - 0.11663 \times HSH2D \end{aligned}$$

The risk score was calculated for each sample using the signature model. The BLCA cohort was assigned to high- and low-risk groups based on the risk score. Receiver operating characteristic (ROC) curves for 1-, 3-, and 5-year OS were generated for the two groups using the R package "survivalROC." Calibration curves were derived from the R package "rms" to evaluate the precision of the 1-, 3-, and 5-year OS prediction. Decision curve analysis was performed using the R package "ggDCA" to quantify the net benefits at different threshold probabilities and evaluate the clinical usefulness of the immune-related signature model.

Prediction of the Immunotherapy Response

The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm[2] was used to estimate the response of each sample to anti-PD-1/PD-L1 and anti-CTLA4 immunotherapy based on the gene expression profiles of the BLCA cohort (Jiang et al., 2018).

Statistical Analysis

Univariate survival analysis was performed using the log-rank test. The Pearson correlation formula was used to calculate the correlations between the risk score and immune markers, the risk score and characteristic gene expression, characteristic gene expression and the immune cell infiltration score, and the risk score and immune cell infiltration score. Two-tailed Student's t tests were used for two-group comparisons. A statistically significant difference was defined as $p < 0.05$. All statistical analyses were performed in R version 4.0.2.

RESULTS

Identification of Immune-Related DEGs in BLCA

The overall design is shown in **Figure 1**. First, to identify DEGs in BLCA, a total of 406 BLCA samples and 18 normal samples were downloaded from the University of California, Santa Cruz (UCSC). A total of 3,677 up-regulated and 3,182 down-regulated genes were obtained (**Figure 2A**). Among these DEGs, there were 1,040 immune-related genes, of which 385 were up-regulated and 655 were down-regulated, respectively (**Figures 2B,C**).

GO and KEGG Annotation of the DEGs

The results of GO functional annotation analysis of the immune-related DEGs showed that the most significantly enriched biological processes (BPs) included homophilic cell adhesion *via* plasma membrane adhesion molecules, extracellular structure organization, and extracellular matrix organization. The most significantly enriched cellular components (CCs) included collagen trimers, the endoplasmic reticulum lumen, and cell–cell junctions, and the molecular functions (MFs) included integrin binding, cytokine receptor binding, and cytokine activity (**Figure 2D**). KEGG pathway enrichment analysis for the immune-related DEGs showed that the significant pathways included the Rap1 signaling pathway, the JAK–STAT signaling pathway, transcriptional misregulation in cancer and proteoglycans in cancer (**Figure 2E**). Most of the above results were related to immunity.

Construction of the Immune-Related Signature

Univariate Cox regression based on the survival and gene expression data of BLCA patients was used to evaluate the prognostic value of the immune-related DEGs. In total, 13 immune-related DEGs were selected as candidate genes to construct the prognostic model with the criterion of value of $p < 0.001$. A forest plot showing the value of p and hazard ratio of the candidate genes is shown in **Figure 3A**.

The BLCA cohort was randomly divided into training and testing sets at a ratio of 7:3. LASSO regression analysis was

[2]http://tide.dfci.harvard.edu

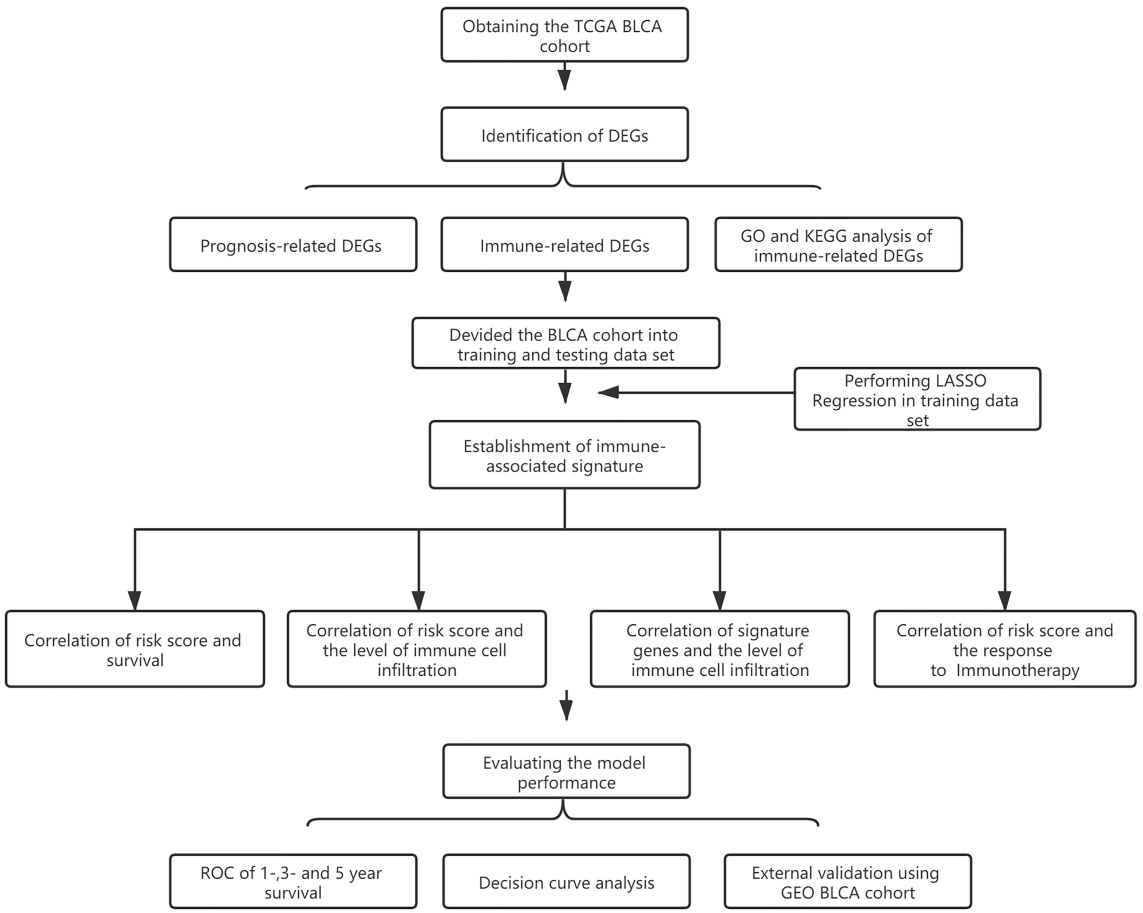


FIGURE 1 | Flow diagram showing the design of the study.

performed using the training data set to identify genes that were most significantly related to survival outcomes (**Figures 3B,C**). A total of eight signature genes were identified: DCHS1, PTGIS, PTPN6, AIFM3, FLRT2, PCSK5, CLSTN2, and HSH2D (**Figure 3D**); the coefficients of the signature genes are shown in **Supplementary Table 1**. Next, we investigated the expression profile of these eight genes in the BLCA cohort. The results indicated that AIFM3, HSH2D, and PTPN6 were significantly up-regulated, while DCHS1, PTGIS, FLRT2, PCSK5, and CLSTN2 were significantly down-regulated in tumor samples compared with normal samples (**Figure 3E**).

A High-Risk Score Is Correlated With Poor Clinical Outcome in BLCA

The risk score of each sample was calculated and ranked on the basis of the signature model in the training set (**Figure 4A**). The scatter plot represented the OS status of BLCA patients according to the risk score, and it suggested that the high-risk group had higher mortality than the low-risk group (**Figure 4B**). The expression profiles of the signature genes showed that tumors with higher risk scores tended to exhibit elevated PCSK5, DCHS1, CLSTN2, PTGIS, and FLRT2 levels, while those with lower risk scores tended to exhibit elevated AIFM3, PTPN6, and HSH2D levels (**Figure 4C**). The same analysis was performed using the testing set, and the results were consistent with those derived from the training data set (**Figures 4D–F**). Compared with the low-risk group, the high-risk group presented a significantly poorer clinical outcome in both the training ($p < 0.001$, **Figure 5A**) and testing sets ($p < 0.001$, **Figure 5B**). Then, the association between the risk score and OS was evaluated, and Kaplan–Meier analysis showed that the low-risk groups had a longer survival time than the high-risk group in both the training (**Figure 5C**, $p < 0.0001$) and testing sets (**Figure 5D**, $p = 0.0018$). These results demonstrated that the risk score was associated with OS and that a low-risk score predicts better survival outcomes for BLCA patients.

Estimating the Degree of Each Immune Cell Infiltration and Predicting the Response to Immunotherapy

Since the efficacy of immunotherapy depends on the level of immune cell infiltration, we analyzed the correlation between the level of immune cell infiltration and risk score. The results demonstrated that there were differences in the infiltration of most immune cells between the high-groups and low-risk

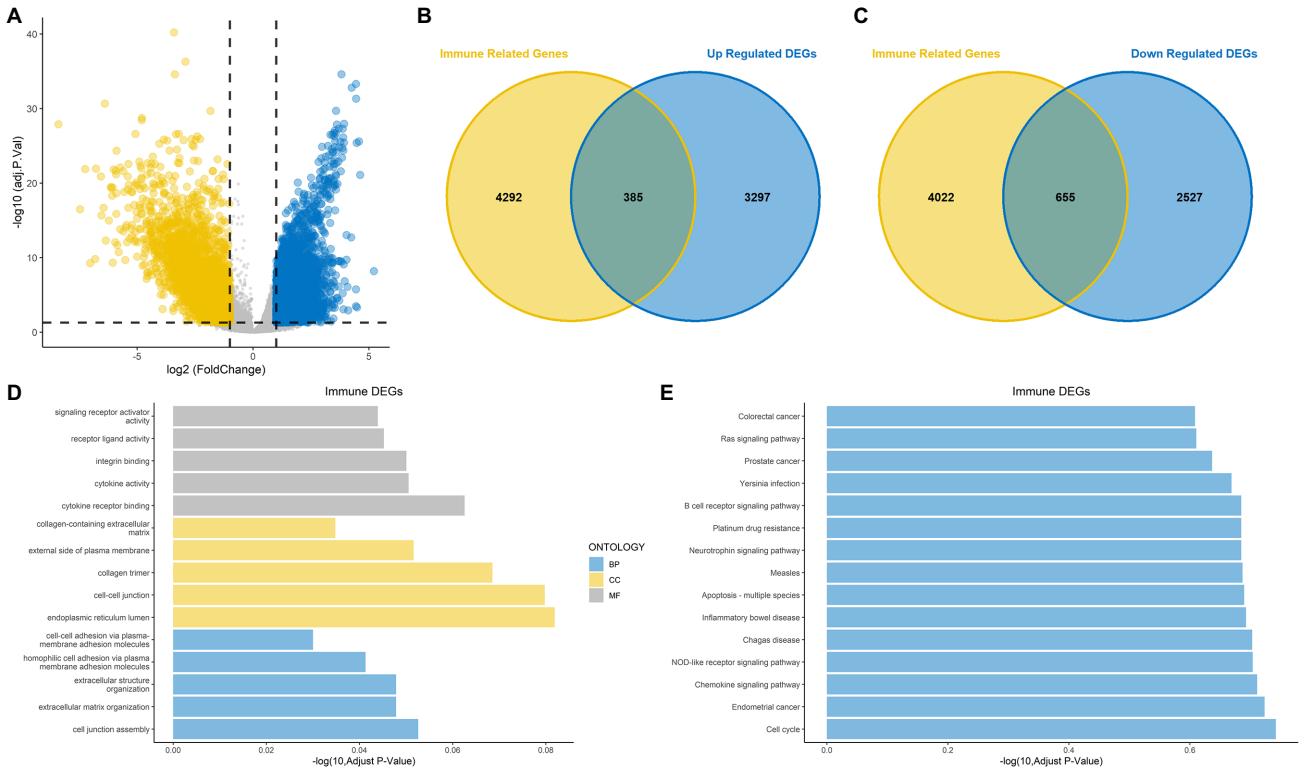


FIGURE 2 | Identification of differentially expressed genes (DEGs) and immune-related DEGs for bladder cancer (BLCA). **(A)** Volcano plot for DEGs for BLCA. **(B,C)** Venn diagram showing up-regulated and down-regulated immune-related DEGs. **(D)** Gene Ontology (GO) functional annotation for immune-related DEGs. **(E)** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for immune-related DEGs.

groups, except for CD56dim NK cells and Eosinophil cells, which demonstrated that signature was significantly correlated with immune infiltration (**Figure 6A**). In addition, we explored the correlation between each of these eight genes and immune cell infiltration. DCHS1, PTGIS, FLRT2, PCSK5, and CLSTN2 were positively related to the infiltration almost all immune cells, and AIFM3 was negatively correlated with the infiltration of most immune cells (**Figure 6B**). Since the eight signature genes were significantly correlated with the level of immune cell infiltration, the correlation of the risk score and the immunotherapy response was explored. Firstly, the relationship between the risk score derived from the risk signature and the expression of immune checkpoint molecules was analyzed. The risk score was positively correlated with the expression of the checkpoint markers, PD-1 (**Figure 6C**), PD-L1 (**Figure 6D**) and CTLA-4 (**Figure 6E**), implicating the potential roles of the signature model in the response to immunotherapy in BLCA patients. The immunotherapy response of the BLCA cohort was estimated by the TIDE algorithm, and the ability of the model to predict the immunotherapy response was evaluated. Furthermore, we then calculated the immunotherapy response rate of samples in the high-risk and low-risk groups. Overall, 49 and 66% patients in the high-risk and low-risk groups, respectively, were predicted to respond to immunotherapy (**Figure 6F**). Patients were divided into responsive and nonresponsive groups, then their risk scores were calculated. The results indicated that the responsive group had significantly lower scores than the nonresponsive group (**Figure 6G**, $p = 0.001$). An ROC curve was generated to determine the efficacy of the risk score in predicting the response to immunotherapy. The area under the ROC curve (AUC) was 0.595, suggesting that the immune-related signature model predicted the response to immunotherapy for BLCA with modest accuracy (**Figure 6H**). In summary, the eight-gene signature model was associated with immune cell infiltration and immunotherapy response in BLCA.

Evaluation of the Model Performance

Time-dependent ROC curves and AUCs were ploted to determine the prognostic values of the eight-gene risk score in the training and testing sets. The AUCs of the risk score for predicting 1-, 3-, and 5-year OS were 0.621, 0.7, and 0.737 in the training set (**Figure 7A**) and 0.655, 0.68, and 0.695 in testing set (**Figure 7B**). To compare the consistency of the model prediction with actual clinical outcomes, calibration curves for 1-, 3-, and 5-year OS in the training (**Supplementary Figures 1A–C**) and testing sets (**Supplementary Figures 1D–F**) were constructed. The results suggested that the calibration curves showed satisfactory agreement between the predicted and observed values for 1-, 3-, and 5-year OS. To determine the clinical usefulness of the risk signature, decision curve analysis in the training and testing sets was performed. Decision curve analysis

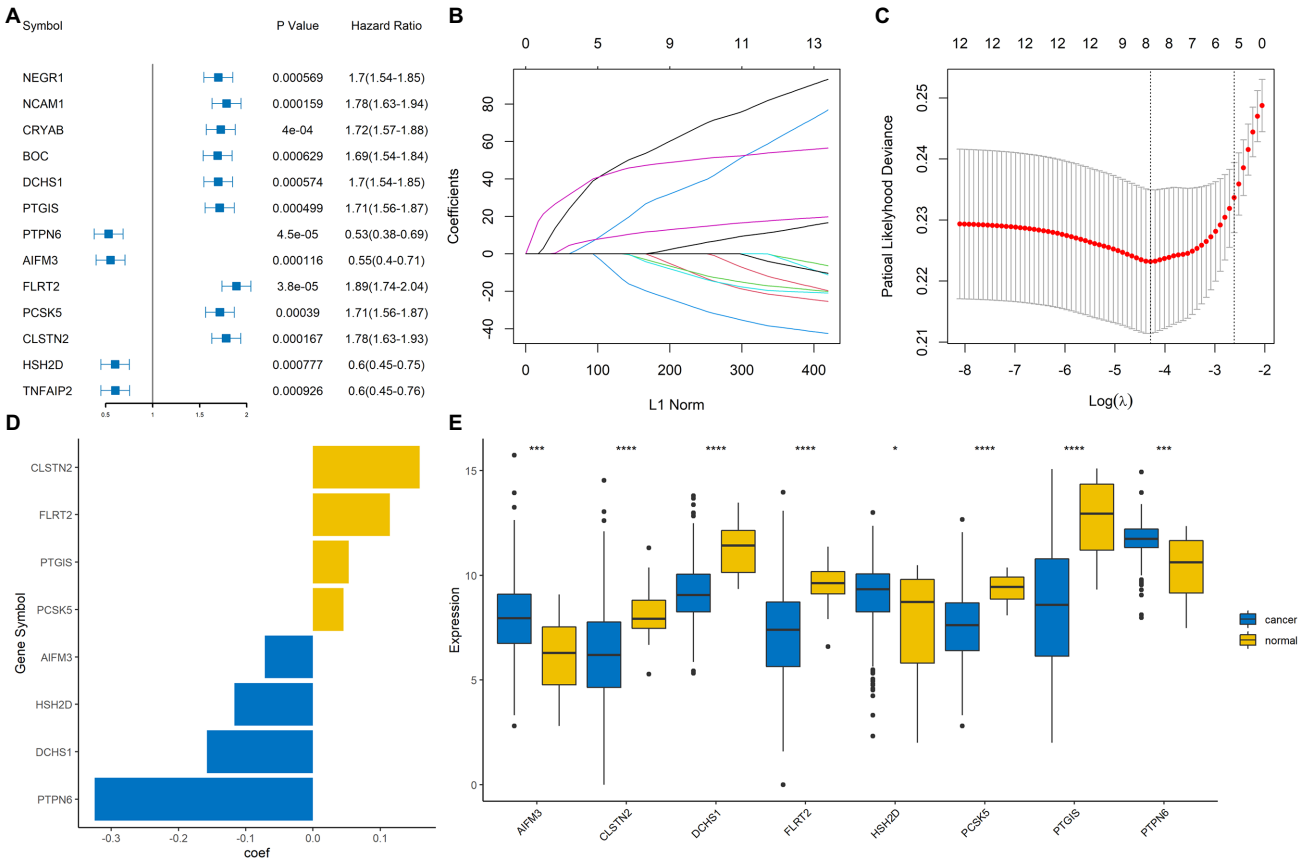


FIGURE 3 | Construction of an immune-related signature model. **(A)** Forrest plot showing the hazard ratio and value of *p* for the candidate genes. **(B,C)** Least absolute shrinkage and selection operator (LASSO) Cox regression analysis identified eight immune-related signature genes that were most correlated with overall survival (OS). **(D)** Bar plot showing the coefficient value for the eight immune-related signature genes. **(E)** Boxplot showing the expression profile of the eight immune-related signature genes in tumor and normal tissues. ****, ***, and * represented $p < 0.00001$, $p < 0.0001$, and $p < 0.05$, respectively.

showed that the risk scores offered a net benefit over the "treat-all" or "treat-none" strategy, which indicated that the model was clinically useful (**Figures 7C,D**). External validation was performed using the BLCA GEO database cohort (GSE31684 and GSE32894). All the samples were divided into high-risk and low-risk groups based on the optimal cutoff point of the risk score, then Kaplan–Meier analysis was performed. The results suggested that the prognosis of the low-risk group was better than that of the high-risk group in the BLCA cohorts GSE31684 (**Figure 7E**, $p = 0.047$) and GSE32894 (**Figure 7F**, $p = 0.012$). The results above indicate that this model had good predictive power in both the TCGA and other external cohorts.

DISCUSSION

Tumors arise as a result of the accumulation of genetic mutations (Williams and Stein, 2004), and a large number of point mutations and structural changes occur during the development of tumors (Stratton et al., 2009), which inevitably stimulates the production of corresponding tumor antigens and thus induces an immune response (Zhang and Zhang, 2020). In recent years, with in-depth research on tumor immunology and molecular biology, immunotherapy has provided a new direction for tumor treatment (Ma et al., 2021). A variety of immune checkpoint (PD-1, PD-L1, and CTLA-4) inhibitors (ICIs) are being used in the treatment of advanced BLCA, and clinical studies have shown that they are reliable in terms of safety and efficacy. However, not every BLCA patient benefits from immunotherapy (Fumet et al., 2020). Consequently, it is imperative to identify biomarkers that can predict patient response to immunotherapy. In the present study, we aimed to construct an immune-related DEGs model, to investigate the relationship between this model and patient prognosis as well as the immunotherapy response, and to assess the potential clinical applications of the model.

In the present study, we obtained 1,040 immune-related BLCA DEGs by analyzing the TCGA database; 385 were up-regulated and 655 were down-regulated. Thirteen immune-related DEGs with significant prognostic value were screened by a one-way Cox regression model ($p < 0.001$). Subsequently, we used LASSO regression analysis to screen eight signature genes from the 13 candidate genes in the univariate Cox regression model. The eight genes were CLSTN2, FLRT2, PTGIS,

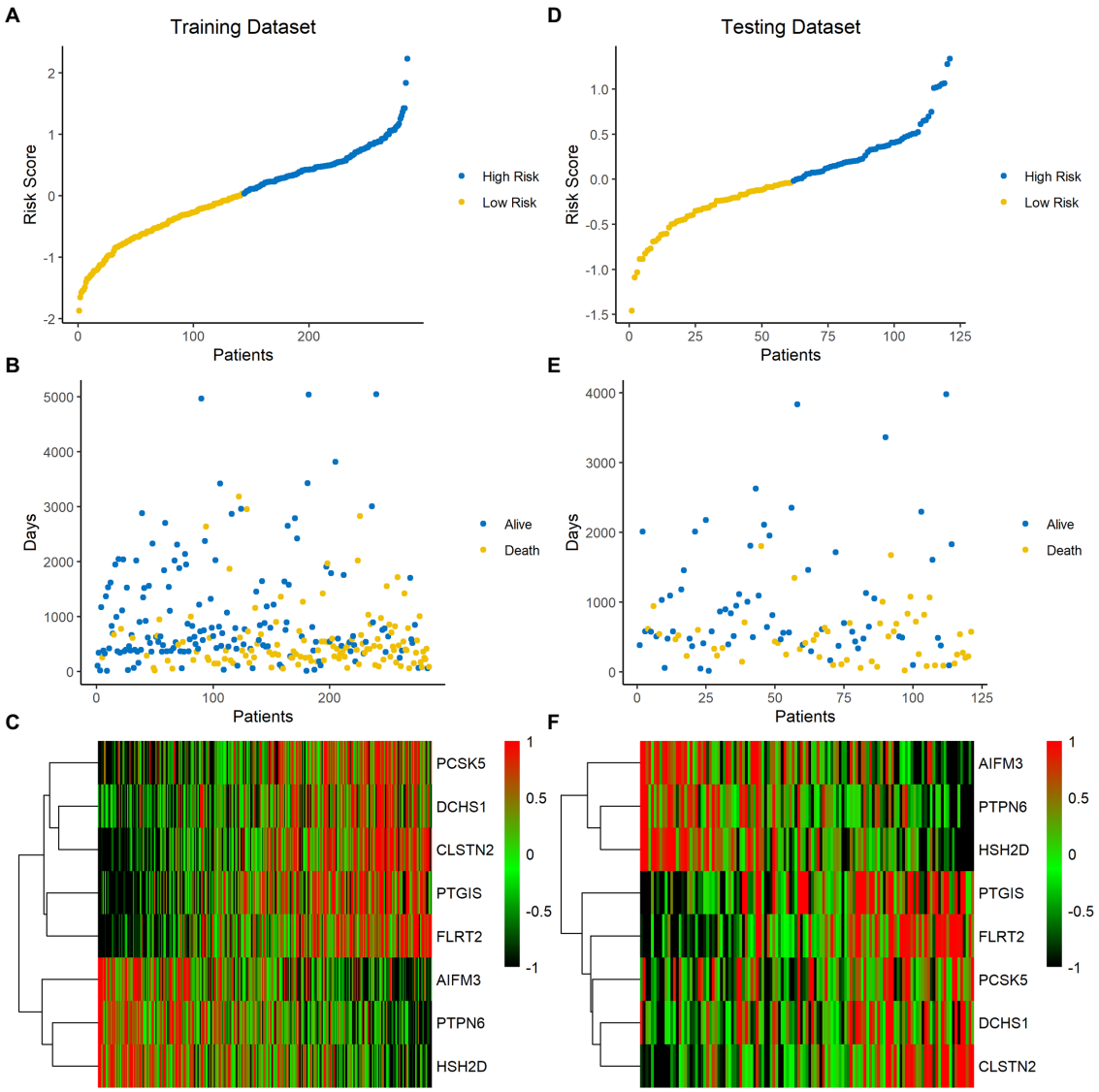


FIGURE 4 | Analysis of the signature score in the training and testing sets. **(A–C)** Training set. **(D–F)** Testing set. The ranked dot plot indicates the risk score distribution in the training set **(A)** and testing set **(D)**. Scatter plot presenting the patients' survival status in the training set **(B)** and testing set **(E)**. Heatmap showing the expression profile of the eight signature genes in BLCA patients from the training set **(C)** and testing set **(F)**.

PCSK5, AIFM3, HSH2D, DCHS1, and PTPN6. FLRT2 is involved in the development of ovarian and uterine carcinogenesis, and FLRT2 has been reported to be a tumor suppressor in breast and prostate cancer (Donninger et al., 2004; Santin et al., 2005; Wu et al., 2016; Bae et al., 2017). The MethHc database show that PTGIS has a high level of DNA methylation in BLCA. PTGIS is a HIF-1α target gene that plays a primary regulatory role in hypoxic tumor progression by activating the transcription of various oncogenes (Lu et al., 2018). PCSK5 is one member of the chymosin-like proprotein convertase family, which can regulate the cleavage and activation of the prestructural domain of TGFβ/bmp family members. PCSK5 plays a major role in the mouse skeleton and organogenesis (Szumska et al., 2017). AIFM3 express in a variety of tissues and aberrantly expressed in several cancers, widely. AIFM3 is a direct target of miR-210 and is associated with the proliferation of human liver cancer cells (Yang et al., 2012), and overexpression of AIFM3 predicts stronger proliferative and invasive behavior in breast cancer (Zheng et al., 2019). HSH2D is an important signaling molecule that can affect T cell activation (Oda et al., 2001; Lapinski et al., 2009). HSH2D inhibits the transcriptional activation of the IL-2 promoter, specifically at the RE/AP element of IL-2, which is regulated by CD28, and HSH2D expression contributes to methotrexate resistance in human T-cell acute lymphoblastic leukemia (Shapiro et al., 2004; Pegram et al., 2015). DCHS1 plays a significant role in innate immunity in the human kidney and bladder according to immunostaining studies, and DCHS1 also can participate in cell adhesion, growth, planar cell polarity

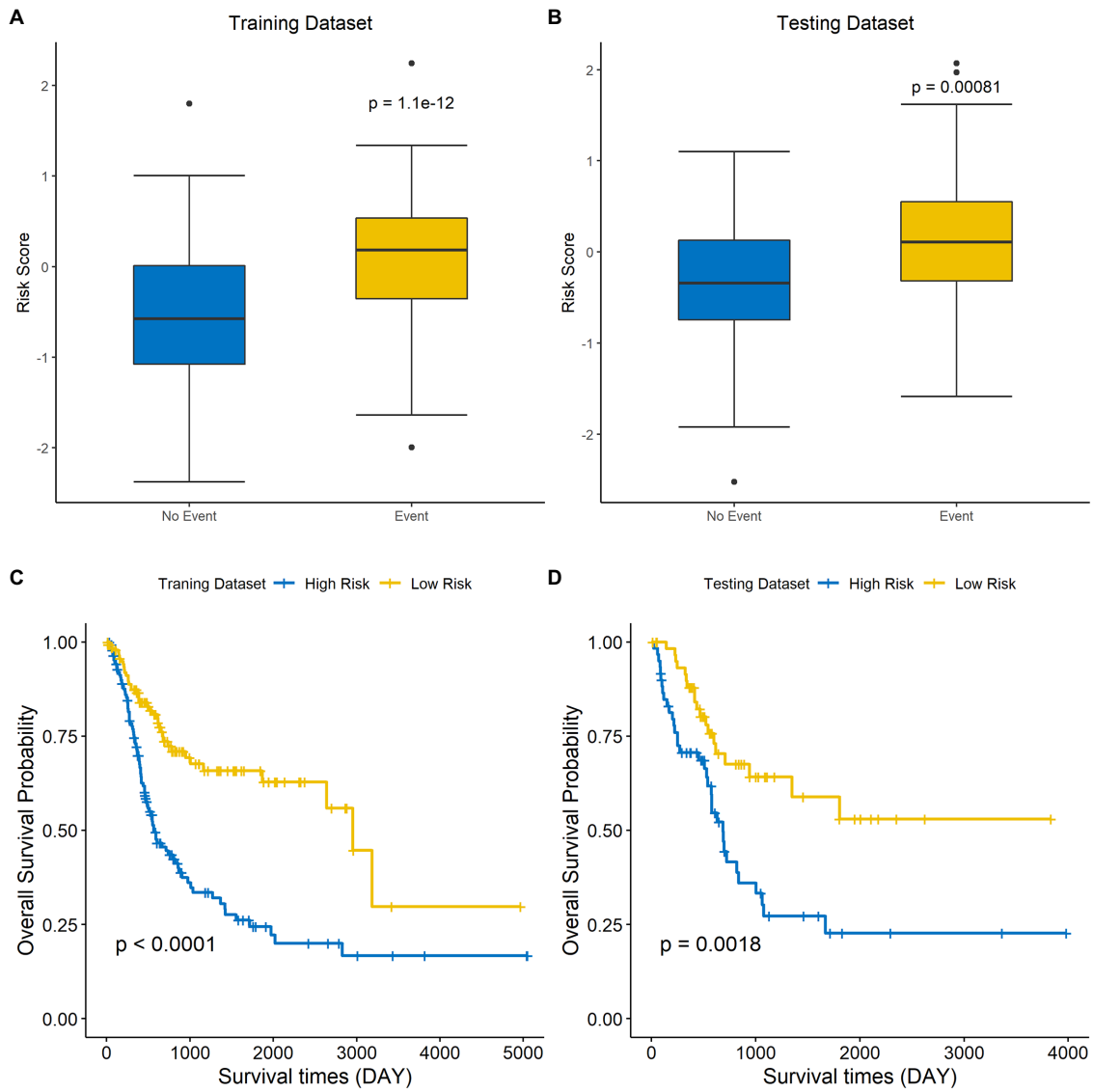


FIGURE 5 | A low-risk score is correlated with better OS in BLCA. The differences in risk scores for patients with endpoint events *vs* nonendpoint events in the training set **(A)** and testing set **(B)**. Survival analysis of the high-risk and low-risk groups in the training **(C)** and testing sets **(D)**.

and tissue pattern (Liang et al., 2019; Qureshi et al., 2020). PTPN6 is thought to be a signaling molecule that can regulate many of cellular processes, including cell growth, differentiation, oncogenic transformation and mitotic cycle, as well as can act as a tumor suppressor. PTPN6 may improve the chemotherapy efficacy and can be used in combination with blocking antibodies in immunotherapy; moreover, some studies suggest that PTPN6 may be an immune-related prognostic biomarker for BLCA (Shen et al., 2020).

There is growing evidence that the immune microenvironment, in which immune cells and molecules are important components, acts an important role in tumor development and the degree of immune cell infiltration is highly correlated with patient prognosis (Grivennikov et al., 2010; Seager et al., 2017). With the recent development of technologies such as RNA-seq, it is possible to systematically analyze the tumor microenvironment and the functional diversity of tumor-infiltrating immune cells, the sensitivity of patients to immunotherapy and the prognosis (Zhang and Zhang, 2019). In recent study, we constructed an eight-gene signature model. The risk scores were significantly associated with the infiltration level of immune cells. Furthermore, six of the eight signature genes showed a significant positive or negative correlation with the infiltration level of immune cells. We also assessed the correlation of the eight signature genes with the response to immunotherapy, and the results suggested that the risk score was significantly and positively correlated with the expression of the checkpoint markers, PD-1, PD-L1, and CTLA-4. The immune-related signature model predicted the response to immunotherapy for BLCA with good accuracy. We applied this signature model

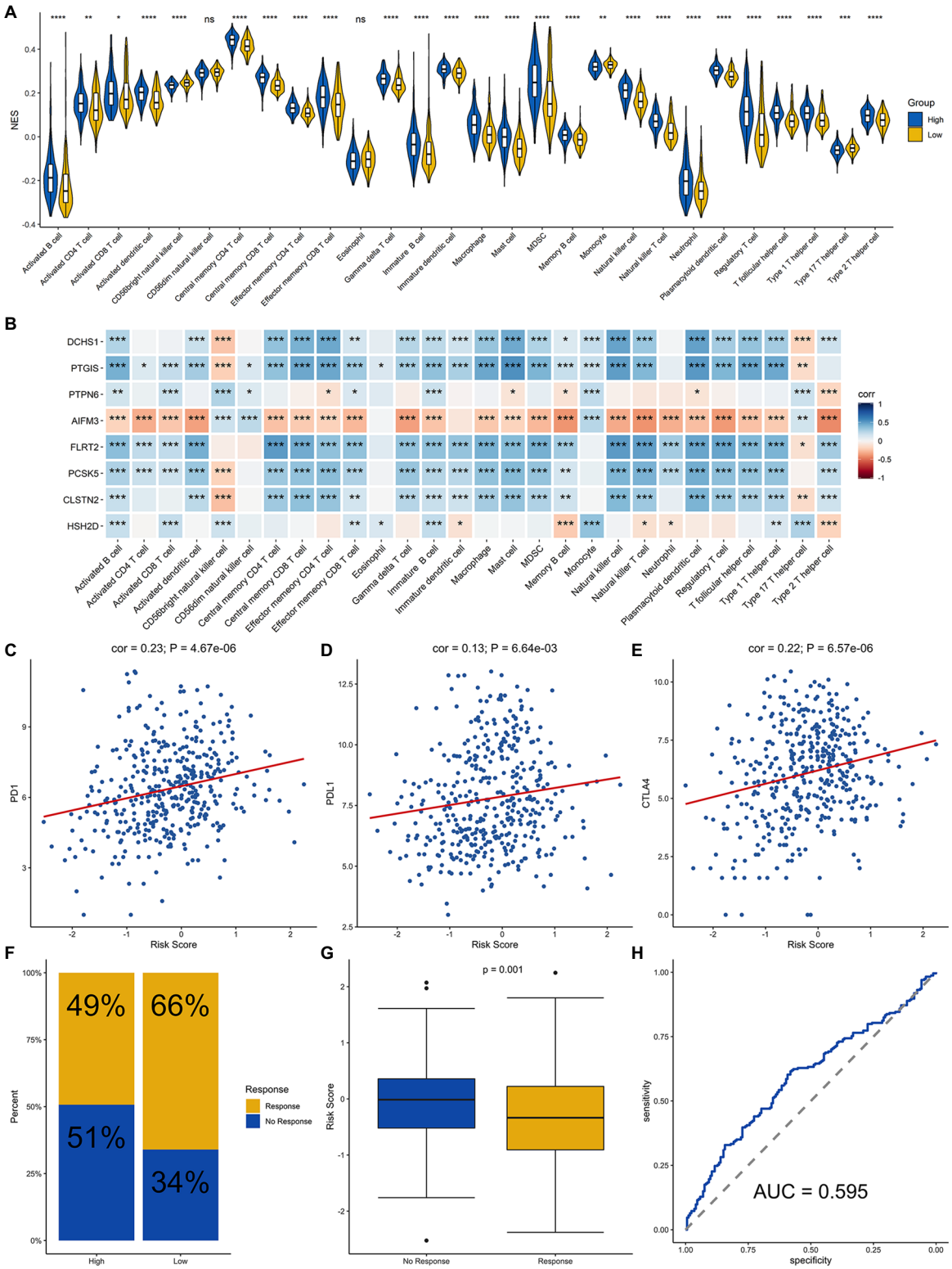


FIGURE 6 | Analysis of immune cell infiltration and prediction of the response to immunotherapy in BLCA patients. **(A)** Infiltration level of immune cells in the high-risk and low-risk groups. **(B)** Correlation analysis for the eight signature genes and the infiltration level of the immune cells. Correlation analysis of the risk score and PD-1 **(C)**, PD-L1 **(D)**, and CTLA4 **(E)**. **(F)** Proportion of patients who responded to immunotherapy in the high-risk group and the low-risk group estimated by the TIDE algorithm. **(G)** Comparison of the risk score of the response group with that of the nonresponse group. **(H)** Receiver operating characteristic (ROC) curve for predicting the response to immunotherapy based on the risk score for BLCA. ****, ***, **, * and ns represented $p < 0.00001$, $p < 0.0001$, $p < 0.001$, $p < 0.05$, and not significant, respectively.

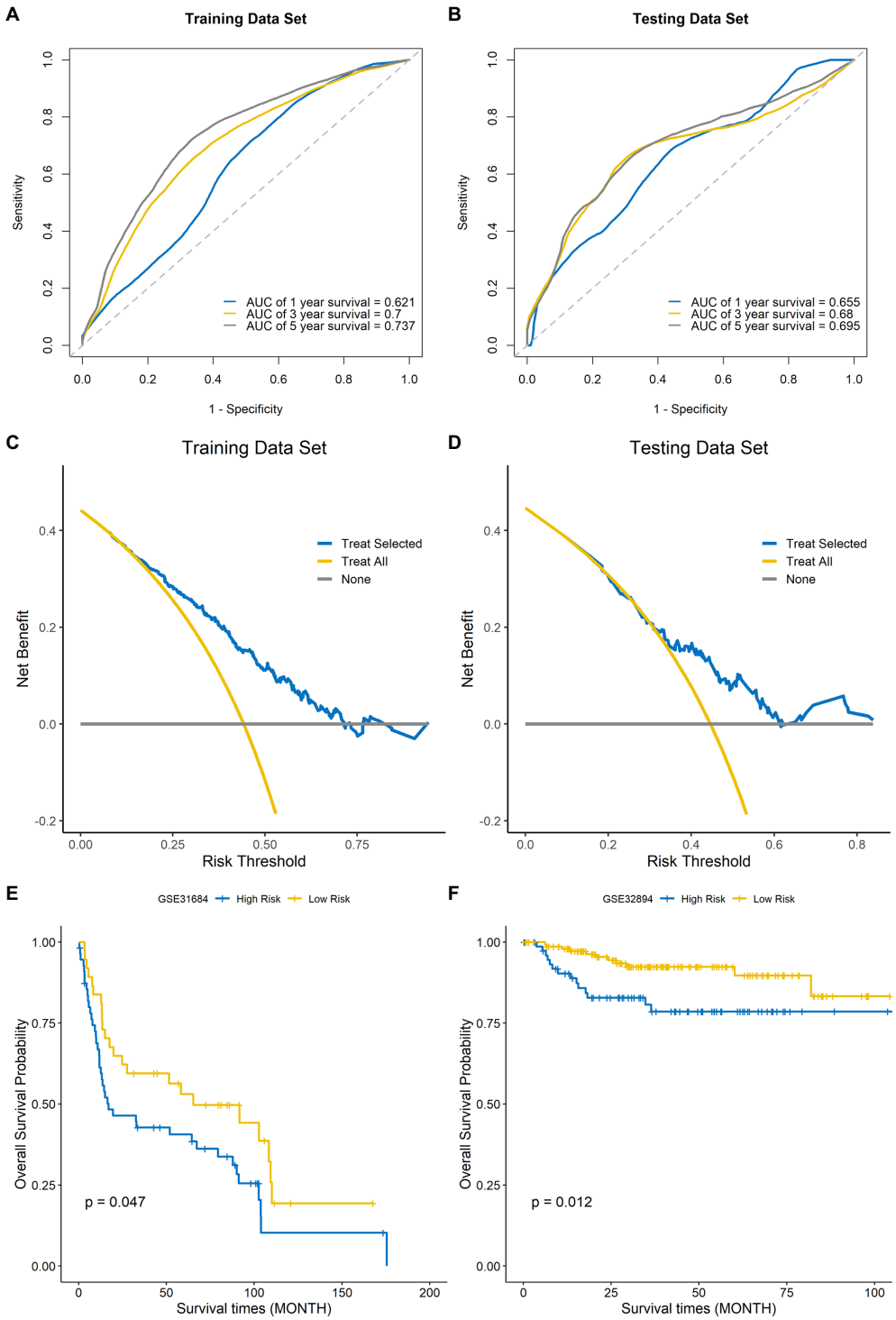


FIGURE 7 | Evaluation of the performance of the signature model. ROC curves used for predicting the 1-, 3-, and 5-year ROC curves in the training set **(A)** and testing set **(B)**. Decision curve analysis of the training set **(C)** and testing set **(D)**. External validation of the signature model using the Gene Expression Omnibus (GEO) BLCA cohorts GSE31684 **(E)** and GSE32894 **(F)**.

to evaluate the clinical data, and the results showed that patients in the low-risk group had better clinical outcomes than those in the high-risk group. We performed Kaplan–Meier analysis on the training set, and the results showed that the low-risk group had a longer survival time than the high-risk group ($p < 0.001$). The above data suggest that this model has good

clinical feasibility. We validated the accuracy and clinical usefulness of the signature model using several methods, including ROC curves for 1-, 3-, and 5-year OS and decision curve analysis. Finally, external validation of the model using the GEO BLCA cohorts further verified the prognostic ability of the model.

However, there were some limitations of this study. This study was retrospective, and further prospective studies are required to validate our findings, and in this study, only innate immunity genes were analyzed and adaptive immunity was not involved. In addition, some clinical characteristics, such as age and clinical stages, were not included in our model.

CONCLUSION

In this study, we established an immune-related eight-gene signature model for BLCA, that could be used to predict the immune response and prognosis of BLCA patients.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

TY made substantial contributions to the conception, design, interpretation, and preparation of the final manuscript. BD, JL, WS, SZ, YM, JS, and MZ participated in the coordination of data acquisition and data analysis and reviewed the manuscript. DL reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a project cosponsored by the Henan Province and Ministry of Health (201601025) and medical science and technology program of Henan Provincial Science and Technology Department (162102310040).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.696912/full#supplementary-material

Supplementary Figure 1 | Calibration curves for evaluating the accuracy of the model in predicting 1-, 3-, and 5-year OS in the training **(A–C)** and testing sets **(D–F)**.

Supplementary Figure 2 | Analysis of immune cell infiltration in GEO BLCA cohorts GSE31684 **(A)** and GSE32894 **(B)**.

Supplementary Table 1 | The coefficients of the signature genes.

Supplementary Table 2 | The clinical information of TCGA BLCA data.

Supplementary Table 3 | The clinical information of GSE32894.

Supplementary Table 4 | The clinical information of GSE31684.

REFERENCES

Bae, H., Kim, B., Lee, H., Lee, S., Kang, H. S., and Kim, S. J. (2017). Epigenetically regulated Fibronectin leucine rich transmembrane protein 2 (FLRT2) shows tumor suppressor activity in breast cancer cells. *Sci. Rep.* 7:272. doi: 10.1038/s41598-017-00424-0

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18, 248–262. doi: 10.1016/j.celrep.2016.12.019

Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* 23, 1846–1847. doi: 10.1093/bioinformatics/btm254

Donninger, H., Bonome, T., Radonovich, M., Pise-Masison, C. A., Brady, J., Shih, J. H., et al. (2004). Whole genome expression profiling of advance stage papillary serous ovarian cancer reveals activated pathways. *Oncogene* 23, 8065–8077. doi: 10.1038/sj.onc.1207959

Fumet, J. D., Truntzer, C., Yarchoan, M., and Ghiringhelli, F. (2020). Tumour mutational burden as a biomarker for immunotherapy: current data and emerging concepts. *Eur. J. Cancer* 131, 40–50. doi: 10.1016/j.ejca.2020.02.038

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

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. doi: 10.1016/j.cell.2011.02.013

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* 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, 1550–1558. doi: 10.1038/s41591-018-0136-1

Kim, K., Jeon, S., Kim, T. M., and Jung, C. K. (2018). Immune gene signature delineates a subclass of papillary thyroid cancer with unfavorable clinical outcomes. *Cancer* 10:494. doi: 10.3390/cancers10120494

Lapinski, P. E., Oliver, J. A., Bodie, J. N., Marti, F., and King, P. D. (2009). The T-cell-specific adapter protein family: TSAd, ALX, and SH2D4A/SH2D4B. *Immunol. Rev.* 232, 240–254. doi: 10.1111/j.1600-065X.2009.00829.x

Liang, D., McHugh, K. M., Brophy, P. D., Shaikh, N., Manak, J. R., Andrews, P., et al. (2019). DNA copy number variations in children with vesicoureteral reflux and urinary tract infections. *PLoS One* 14:e0220617. doi: 10.1371/journal.pone.0220617

Lu, M., Ge, Q., Wang, G., Luo, Y., Wang, X., Jiang, W., et al. (2018). CIRBP is a novel oncogene in human bladder cancer inducing expression of HIF-1α. *Cell Death Dis.* 9:1046. doi: 10.1038/s41419-018-1109-5

Lv, Z., Pang, C., Wang, J., Xia, H., Liu, J., Yan, Q., et al. (2021). Identification of a prognostic signature based on immune-related genes in bladder cancer. *Genomics* 113, 1203–1218. doi: 10.1016/j.ygeno.2021.03.012

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, 8688–8705. doi: 10.18632/aging.202681

Oda, T., Muramatsu, M. A., Isogai, T., Masuho, Y., Asano, S., and Yamashita, T. (2001). HSH2: a novel SH2 domain-containing adapter protein involved in tyrosine kinase signaling in hematopoietic cells. *Biochem. Biophys. Res. Commun.* 288, 1078–1086. doi: 10.1006/bbrc.2001.5890

Pegram, H. J., Purdon, T. J., van Leeuwen, D. G., Curran, K. J., Giralt, S. A., Barker, J. N., et al. (2015). IL-12-secreting CD19-targeted cord blood-derived T cells for the immunotherapy of B-cell acute lymphoblastic leukemia. *Leukemia* 29, 415–422. doi: 10.1038/leu.2014.215

Pettenati, C., and Ingersoll, M. A. (2018). Mechanisms of BCG immunotherapy and its outlook for bladder cancer. *Nat. Rev. Urol.* 15, 615–625. doi: 10.1038/s41585-018-0055-4

Powles, T., Eder, J. P., Fine, G. D., Braiteh, F. S., Loriot, Y., Cruz, C., et al. (2014). MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515, 558–562. doi: 10.1038/nature13904

Qiu, H., Hu, X., He, C., Yu, B., Li, Y., and Li, J. (2020). 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

Qureshi, A. H., Liang, D., Canas, J., Hooks, J., Arrregui, S. W., Saxena, V., et al. (2020). DCHS1 DNA copy number loss associated with pediatric urinary tract infection risk. *Innate Immun.* 26, 473–481. doi: 10.1177/1753425920917193

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:e47. doi: 10.1093/nar/gkv007

Rouanne, M., Roumiguié, M., Houédé, N., Masson-Lecomte, A., Colin, P., Pignot, G., et al. (2018). Development of immunotherapy in bladder cancer: present and future on targeting PD(L)1 and CTLA-4 pathways. *World J. Urol.* 36, 1727–1740. doi: 10.1007/s00345-018-2332-5

Rouprêt, M., Babjuk, M., Burger, M., Capoun, O., Cohen, D., Compérat, E. M., et al. (2021). European association of urology guidelines on upper urinary tract urothelial carcinoma: 2020 update. *Eur. Urol.* 79, 62–79. doi: 10.1016/j.eururo.2020.05.042

Santin, A. D., Zhan, F., Cane, S., Bellone, S., Palmieri, M., Thomas, M., et al. (2005). Gene expression fingerprint of uterine serous papillary carcinoma: identification of novel molecular markers for uterine serous cancer diagnosis and therapy. *Br. J. Cancer* 92, 1561–1573. doi: 10.1038/sj.bjc.6602480

Seager, R. J., Hajal, C., Spill, F., Kamm, R. D., and Zaman, M. H. (2017). Dynamic interplay between tumour, stroma and immune system can drive or prevent tumour progression. *Converg. Sci. Phys. Oncol.* 3:034002. doi: 10.1088/2057-1739/aa7e86

Shapiro, M. J., Powell, P., Ndubuizu, A., Nzerem, C., and Shapiro, V. S. (2004). The ALX Src homology 2 domain is both necessary and sufficient to inhibit T cell receptor/CD28-mediated up-regulation of RE/AP. *J. Biol. Chem.* 279, 40647–40652. doi: 10.1074/jbc.M404198200

Shen, C., Liu, J., Wang, J., Yang, X., Niu, H., and Wang, Y. (2020). The analysis of PTPN6 for bladder cancer: an exploratory study based on TCGA. *Dis. Markers* 2020:4312629. doi: 10.1155/2020/4312629

Shen, S., Wang, G., Zhang, R., Zhao, Y., Yu, H., Wei, Y., et al. (2019). Development and validation of an immune gene-set based prognostic signature in ovarian cancer. *EBioMedicine* 40, 318–326. doi: 10.1016/j.ebiom.2018.12.054

Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. *Nature* 458, 719–724. doi: 10.1038/nature07943

Szumska, D., Cioroch, M., Keeling, A., Prat, A., Seidah, N. G., and Bhattacharya, S. (2017). Pcsk5 is required in the early cranio-cardiac mesoderm for heart development. *BMC Dev. Biol.* 17:6. doi: 10.1186/s12861-017-0148-y

Tran, L., Xiao, J. F., Agarwal, N., Duex, J. E., and Theodorescu, D. (2021). Advances in bladder cancer biology and therapy. *Nat. Rev. Cancer* 21, 104–121. doi: 10.1038/s41568-020-00313-1

Williams, S. G., and Stein, J. P. (2004). Molecular pathways in bladder cancer. *Urol. Res.* 32, 373–385. doi: 10.1007/s00240-003-0345-y

Wu, Y., Davison, J., Qu, X., Morrissey, C., Storer, B., Brown, L., et al. (2016). Methylation profiling identified novel differentially methylated markers including OPCML and FLRT2 in prostate cancer. *Epigenetics* 11, 247–258. doi: 10.1080/15592294.2016.1148867

Yang, W., Sun, T., Cao, J., Liu, F., Tian, Y., and Zhu, W. (2012). Downregulation of miR-210 expression inhibits proliferation, induces apoptosis and enhances radiosensitivity in hypoxic human hepatoma cells in vitro. *Exp. Cell Res.* 318, 944–954. doi: 10.1016/j.yexcr.2012.02.010

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, 284–287. doi: 10.1089/omi.2011.0118

Zhang, L., and Zhang, Z. (2019). Recharacterizing tumor-infiltrating lymphocytes by single-cell RNA sequencing. *Cancer Immunol. Res.* 7, 1040–1046. doi: 10.1158/2326-6066.CIR-18-0658

Zhang, Y., and Zhang, Z. (2020). The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell. Mol. Immunol.* 17, 807–821. doi: 10.1038/s41423-020-0488-6

Zheng, A., Zhang, L., Song, X., Wang, Y., Wei, M., and Jin, F. (2019). Clinical implications of a novel prognostic factor AIFM3 in breast cancer patients. *BMC Cancer* 19:451. doi: 10.1186/s12885-019-5659-4

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.* 10:19695. 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.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 September 2021
doi: 10.3389/fgene.2021.684281

A Robust Seven-Gene Signature Associated With Tumor Microenvironment to Predict Survival Outcomes of Patients With Stage III–IV Lung Adenocarcinoma

*Hao Zhao[1], Xuening Zhang[2], Lan Guo[1], Songhe Shi[3] and Ciyong Lu[1]**

[1] Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China, [2] Department of Epidemiology and Biostatistics, College of Public Health, Shandong University, Jinan, China, [3] Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:
Wan Zhu,
Stanford University, United States

Reviewed by:
Daiwei Wan,
The First Affiliated Hospital of Soochow University, China
Ezhilarasi Chendamarai,
Washington University in St. Louis, United States

***Correspondence:**
Ciyong Lu
luciyong@mail.sysu.edu.cn

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

Received: *24 March 2021*
Accepted: *31 July 2021*
Published: *06 September 2021*

Citation:
Zhao H, Zhang X, Guo L, Shi S and Lu C (2021) A Robust Seven-Gene Signature Associated With Tumor Microenvironment to Predict Survival Outcomes of Patients With Stage III–IV Lung Adenocarcinoma. Front. Genet. 12:684281. doi: 10.3389/fgene.2021.684281

Background: Due to the relatively insidious early symptoms of lung adenocarcinoma (LUAD), most LUAD patients are at an advanced stage at the time of diagnosis and lose the best chance of surgical resection. Mounting evidence suggested that the tumor microenvironment (TME) was highly correlated with tumor occurrence, progress, and prognosis. However, TME in advanced LUAD remained to be studied and reliable prognostic signatures based on TME in advanced LUAD also had not been well-established. This study aimed to understand the cell composition and function of TME and construct a gene signature associated with TME in advanced LUAD.

Methods: The immune, stromal, and ESTIMATE scores of each sample from The Cancer Genome Atlas (TCGA) database were, respectively, calculated using an ESTIMATE algorithm. The LASSO and Cox regression model were applied to select prognostic genes and to construct a gene signature associated with TME. Two independent datasets from the Gene Expression Omnibus (GEO) were used for external validation. Twenty-two subsets of tumor-infiltrating immune cells (Tiics) were analyzed using the CIBERSORT algorithm.

Results: Favorable overall survival (OS) and progression-free survival (PFS) were found in patients with high immune score ($p = 0.048$ and $p = 0.028$; respectively) and stromal score ($p = 0.024$ and $p = 0.025$; respectively). Based on the immune and stromal scores, 453 differentially expressed genes (DEGs) were identified. Using the LASSO and Cox regression model, a seven-gene signature containing AFAP1L2, CAMK1D, LOXL2, PIK3CG, PLEKHG1, RARRES2, and SPP1 was identified to construct a risk stratification model. The OS and PFS of the high-risk group were significantly worse than that of the low-risk group ($p < 0.001$ and $p < 0.001$; respectively). The receiver operating characteristic (ROC) curve analysis confirmed the good potency of the seven-gene signature. Similar findings were validated in two independent cohorts. In addition,

the proportion of macrophages M2 and Tregs was higher in high-risk patients ($p = 0.041$ and $p = 0.022$, respectively).

Conclusion: Our study established and validated a seven-gene signature associated with TME, which might serve as a prognosis stratification tool to predict survival outcomes of advanced LUAD patients. In addition, macrophages M2 polarization may lead to worse prognosis in patients with advanced LUAD.

Keywords: tumor microenvironment, immune, lung adenocarcinoma, stromal, TCGA

INTRODUCTION

Lung cancer ranks first in the incidence and mortality of all malignant tumors worldwide (Bray et al., 2018). The 5-year survival rate of lung cancer patients is less than 20% (Herbst et al., 2018). Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small cell lung cancer (NSCLC), which accounts for about 40% of all lung malignancies and usually occurs in the outer area of the lung (Chen et al., 2014). Clinical studies have shown that nearly 70% of LUAD patients are discovered in stage III–IV, and 57% of LUAD patients have already developed distant metastasis at the time of initial diagnosis, and have lost the best opportunity for surgical resection (Jemal et al., 2017).

In recent years, significant progress has been made in the research of molecular genetics and immunotherapy of lung cancer, and molecular typing based on genetic characteristics has brought the treatment of advanced lung cancer into the era of personalized molecular targeted therapy (Subramanian and Govindan, 2008). EGFR-Inhibitor and BRAF(V600E)-mutant, rearrangements of ALK or ROS1 genes, as well as immune checkpoint inhibitor antibodies against PD-1 or PD-L1 have been approved for the treatment of advanced LUAD (Vargas and Harris, 2016; Hirsch et al., 2017). At present, the TNM staging system is still the most effective tool to judge the survival of patients and guide clinical treatment strategies, but the evaluation effect for advanced survival is not good (Goldstraw et al., 2016). Therefore, looking for a new survival predictor for advanced LUAD patients is particularly important for personalized treatment of clinical decision-making and prognostic health management.

Tumor microenvironment (TME) refers to the surrounding microenvironment of tumor cells, including immune cells, stromal cells, endothelial cells, inflammatory cells, and fibroblasts (Neal et al., 2018). Among them, immune cells and stromal cells are two major non-tumor cell components, which have been considered important for the diagnosis and prognostic evaluation of cancer patients (Gajewski et al., 2013). Therefore, understanding the cell composition and function of TME will bring a new dawn to patients with advanced LUAD in terms of immunity and targeted therapy and improvement of prognosis.

The continuous improvement and development of public databases, such as The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database, provide reliable data resources for TME research (Cancer Genome Atlas Research Network, Weinstein et al., 2013; NCBI Resource Coordinators, 2016). Yoshihara et al. (2013) first proposed the ESTIMATE algorithm in 2013. This algorithm uses the unique properties of the transcription profile of cancer samples to infer infiltrating stromal/immune cells. According to reports, researchers have explored the tumor characteristics and prognosis assessment of liver cancer (Li et al., 2019), breast cancer (Bai et al., 2019), and clear cell renal cell carcinoma (Luo et al., 2020) based on the ESTIMATE algorithm. However, the value of immune and stromal scores for advanced LUAD has not been verified.

In the present study, the immune and stromal scores were estimated using the ESTIMATE algorithm based on the transcription profile of LUAD patients with stage III–IV. A robust gene signature based on immune-stromal score was subsequently developed for prognosis stratification in advanced LUAD. Finally, we explored the relationship between high-/low-risk advanced LUAD patients and immune cell infiltration based on the CIBERSORT method, so as to provide some references for combined immunotherapy and targeted therapy for advanced LUAD patients.

MATERIALS AND METHODS

Data Collection and Processing

We obtained the fragments per kilobase per million (FPKM) data of RNA-Seq from the TCGA-LUAD cohort[1], including 535 LUAD patients and 59 normal samples. Next, the FPKM data were transferred to transcripts per million (TPM) expression data. The gene expression levels of duplicate samples were averaged, and normal samples were deleted for subsequent analysis.

We used the GDC tool and cBioPortal website[2] to download the corresponding clinical information, including age, gender, history of smoking, tumor laterality, metastasis, lymph node status, pathological T stage, stage, and prognostic information. In this study, only patients with stage III–IV were included and patients with incomplete key clinical information were excluded. Finally, a total of 103 advanced LUAD patients were included for follow-up analysis. We utilized the "limma" package for normalization processing, and then immune, stromal, and ESTIMATE scores were calculated using ESTIMATE algorithm. Two independent datasets from the GEO database[3], namely, 27

[1]https://portal.gdc.cancer.gov/

[2]http://www.cbioportal.org/

[3]https://www.ncbi.nlm.nih.gov/geo/

LUAD patients with stage III–IV from Series GSE81089 and 53 LUAD patients with stage III–IV from Series GSE41271, were used for external validation in this study. For all patients from the GEO database, a normalized expression matrix was used for subsequent analyses.

Correlations Between Prognoses and Immune/Stromal Scores

Overall survival (OS) was used as the primary prognosis endpoint, and progression-free survival (PFS) was used as the secondary prognosis endpoint. According to the stromal and immune scores of each advanced LUAD patient, the best cutoff value based on the R package "maxstat" (i.e., the maximum selective rank statistic method) (Ritchie et al., 2015) was used to divide the patients into high-score and low-score groups. Based on "survival" packages, the Kaplan–Meier (K–M) survival curve analysis and log-rank tests were used to compare the prognoses of the two groups.

Differentially Expressed Gene (DEG) Screening

The "limma" package in R software was used to screen for DEGs between high-score and low-score groups of immune score and stromal score. In this study, an adjusted p-value < 0.05 and fold change ≥ 1.5 were regarded as the critical value for screening DEGs. The immune-related DEGs and stromal-related DEGs showing the same expression trend were selected for further analysis using a Venn diagram. We used the "pheatmap" package to generate the immune-related heatmap and stromal-related heatmap.

DEG Functional Enrichment Analysis

The David online database[4] was used to explore the potential functions of DEGs. Gene ontology (GO) analysis included biological processes (BP), molecular functions (MF), and cellular components (CC), which are demonstrated by a bar plot. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed to conduct the pathway analysis, which was illustrated by a dot plot. With false discovery rate (FDR) < 0.05 as cutoff value, all enrichment results were visualized with "ggplot2" package.

Construction of Gene Signature and Survival Analysis

Firstly, the univariate Cox model was used to determine the relationship between TME-related DEG expression and patient's survival. Then, the least absolute shrinkage and selection operator (LASSO) regression analysis was used to further screen out key genes from significant DEGs in the univariate analysis. LASSO regression increases penalty function on the basis of the least

[4]http://david.ncifcrf.gov

TABLE 1 | Clinical characteristics of 103 advanced LUAD patients included in the study from the TCGA cohort.

Variables	Total (*N* = 103)	Death		Log-rank *P*	Progress or Death		Log-rank *P*
		No (*N* = 44)	Yes (*N* = 59)		No (*N* = 35)	Yes (*N* = 68)	
Age				0.766			0.378
<65	48 (46.60)	21 (43.75)	27 (56.25)		18 (37.50)	30 (62.50)	
≥65	55 (53.40)	23 (41.82)	32 (58.18)		17 (30.91)	38 (69.09)	
Gender				0.915			0.549
Female	55 (53.40)	24 (43.64)	31 (56.36)		21 (38.18)	34 (61.82)	
Male	48 (46.60)	20 (41.67)	28 (58.33)		14 (29.17)	34 (70.83)	
Smoke				0.148			0.130
No	16 (15.53)	12 (75.00)	4 (25.00)		10 (62.50)	6 (37.50)	
Yes	87 (84.47)	32 (36.78)	55 (63.22)		25 (28.74)	62 (71.26)	
Tumor laterality				0.227			0.493
Left	41 (39.81)	16 (39.02)	25 (60.98)		14 (34.15)	27 (65.85)	
Right	62 (60.19)	28 (45.16)	34 (54.84)		21 (33.87)	41 (66.13)	
Metastasis				0.928			0.820
M0	79 (76.70)	35 (44.30)	44 (55.70)		29 (36.71)	50 (63.29)	
M1	24 (23.30)	9 (37.50)	15 (62.50)		6 (25.00)	18 (75.00)	
Lymph node status				0.098			0.145
No	17 (16.50)	11 (64.71)	6 (35.29)		10 (58.82)	7 (41.18)	
Yes	86 (83.50)	33 (38.37)	53 (61.63)		25 (29.07)	61 (70.93)	
Pathological T stage				0.063			0.228
T1–T2	67 (65.05)	33 (49.25)	34 (50.75)		26 (38.81)	41 (61.19)	
T3–T4	36 (34.95)	11 (30.56)	25 (69.44)		9 (25.00)	27 (75.00)	
Stage				0.950			0.853
III	78 (75.73)	34 (43.59)	44 (56.41)		28 (35.90)	50 (64.10)	
IV	25 (24.27)	10 (40.00)	15 (60.00)		7 (28.00)	18 (72.00)	

squares method, which can reduce the number of variables and avoid overfitting effectively (Bovelstad et al., 2007). Finally, the key genes screened by LASSO were included in multivariate Cox analysis, and the gene signature (risk score) formula was constructed according to the analysis results.

The risk score was calculated as follows: risk score = $\sum$ (βi * Expi) ("i" = the number of prognostic hub genes, "βi" represents the coefficient of each gene, and "Expi" represents gene expression).

In addition, advanced LAUD patients were divided into high-risk and low-risk groups according to the median risk score. The receiver operating characteristic (ROC) curves and the consistency index (C-index) were then used to assess the predictive ability of the risk score. The K–M curves and log-rank tests were used to analyze the difference in survival between the high-risk group and the low-risk group. Furthermore, the independent prognostic value of the gene signature was explored by multivariate Cox analysis combined with other clinicopathologic characteristics.

Validation of Gene Signature in the Testing Dataset

The GSE81089 and GSE41271 independent datasets were used for verification. According to the gene signature calculation formula of the training dataset, the samples in the test dataset were divided into the high-risk group and the low-risk groups. The K–M survival analysis and ROC curves were used to evaluate the predictive ability of this model. Immunohistochemistry (IHC) images of the selected prognosis-related genes in tumor and normal tissue were retrieved from the Human Protein Atlas online database[5].

Estimating the Composition of Immune Cells

CIBERSORT is a deconvolution algorithm based on the principle of linear support vector regression to describe the infiltration of immune cells in the sample (Shen-Orr et al., 2010). LM22 is composed of 547 genes that accurately distinguish 22 human hematopoietic cell phenotypes, including seven T-cell types, naïve and memory B cells, plasma cells, NK cells, and myeloid subsets (Newman et al., 2015). We used CIBERSORT and LM22 to jointly estimate the scores of 22 human immune cell types in advanced LAUD patients from the TCGA cohort. For each sample, the sum of all estimated immune cell type scores was equal to 1. We compared differences in the composition of immune cell types between high-risk and low-risk groups.

[5]http://www.proteinatlas.org

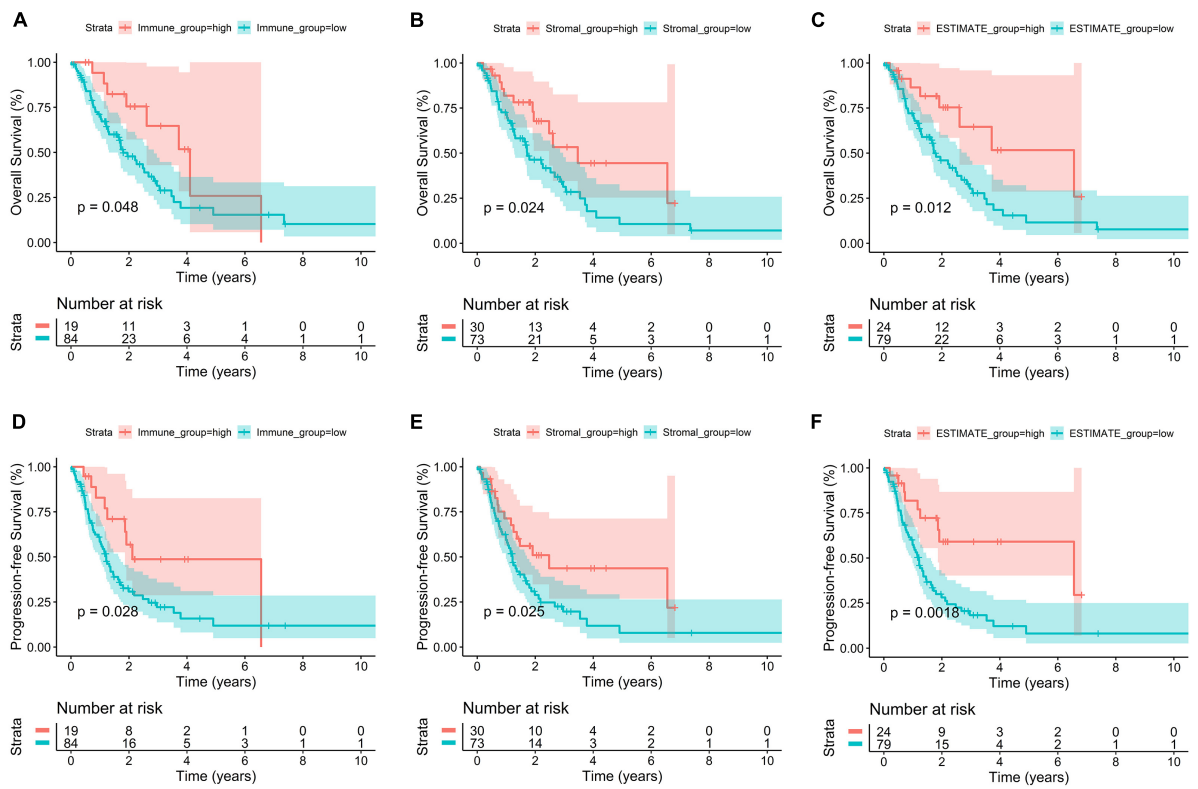


FIGURE 1 | Correlation of immune score and stromal score with advanced LUAD survival outcomes. **(A–C)** K–M survival curves of the relationships between different score levels and OS showed that patients with lower immune, stromal, and ESTIMATE scores had worse OS outcomes (p = 0.048, p = 0.024, and p = 0.012, respectively). **(D–F)** K–M survival curves of the relationships between different score levels and PFS showed that patients with lower immune, stromal, and ESTIMATE scores had worse PFS outcomes (p = 0.028, p = 0.025, and p = 0.002, respectively).

Statistical Analysis

Statistical analysis was performed using R software (version 3.6.1). All statistical tests were two sided and p-value < 0.05 indicated statistical significance.

RESULTS

Estimation of Immune Score and Stromal Score

We included 103 LUAD samples from the TCGA database, of which 78 (75.73%) were in stage III and 25 (24.27%) were in stage IV. The clinical and pathological characteristics of the included patients are listed in **Table 1**. Among them, elderly LUAD patients (≥65 years) accounted for 53.40%, and the proportion of LUAD patients with a history of smoking was as high as 84.47%.

The immune, stromal, and ESTIMATE scores of each sample were, respectively, calculated using an ESTIMATE algorithm. The immune score ranged from −941.95 to 2,940.32, the stromal score ranged from −1,755.55 to 1,923.43, and the ESTIMATE score ranged from −2,298.51 to 4,012.25.

Immune Score and Stromal Score Were Significantly Related to Advanced LUAD Survival Outcomes

Lung adenocarcinoma samples were divided into high-score and low-score groups, based on the best cutoff value of immune score, stromal score, and ESTIMATE score,

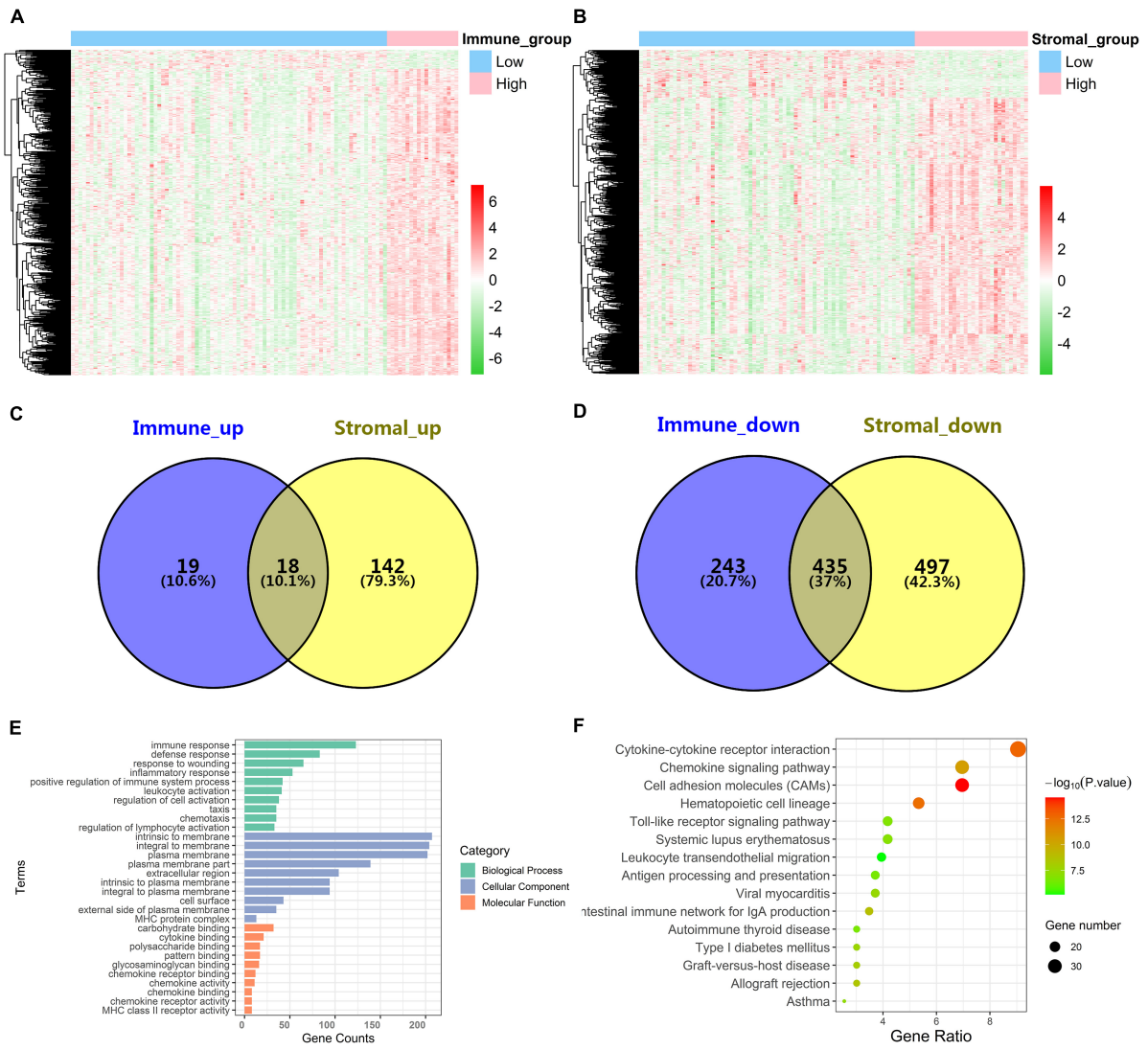


FIGURE 2 | Identification of DEGs and function and pathway enrichment analysis. **(A)** A heatmap of 715 DEGs between patients with high or low immune scores. **(B)** A heatmap of 1,092 DEGs between patients with high or low stromal scores. **(C,D)** Cross-upregulated and cross-downregulated DEGs between the immune and stromal groups. **(E,F)** Function and pathway enrichment analysis of DEGs by GO and KEGG.

respectively. The K–M survival curves were performed to evaluate the relationships between different score levels and survival outcome. K–M survival curves of the relationships between different score levels and OS showed that patients with lower immune, stromal, and ESTIMATE scores had worse OS outcomes (p = 0.048, p = 0.024, and p = 0.012, respectively; **Figures 1A–C**). Consistently, K–M survival curves of the relationships between different score levels and PFS showed that patients with lower immune, stromal, and ESTIMATE scores had worse PFS outcomes (p = 0.028, p = 0.025, and p = 0.002, respectively; **Figures 1D–F**). These observations consistently suggested that advanced LUAD patients with a higher immune score or stromal score had a more favorable outcome.

Identification of DEGs Based on Immune Score and Stromal Score in Advanced LUAD

In order to explore the DEGs closely related to the TME, the "limma" package was used to process the Affymetrix microarray data from 103 advanced LUAD patients. **Figure 2A** showed a heatmap of 715 DEGs between high and low immune scores, and **Figure 2B**

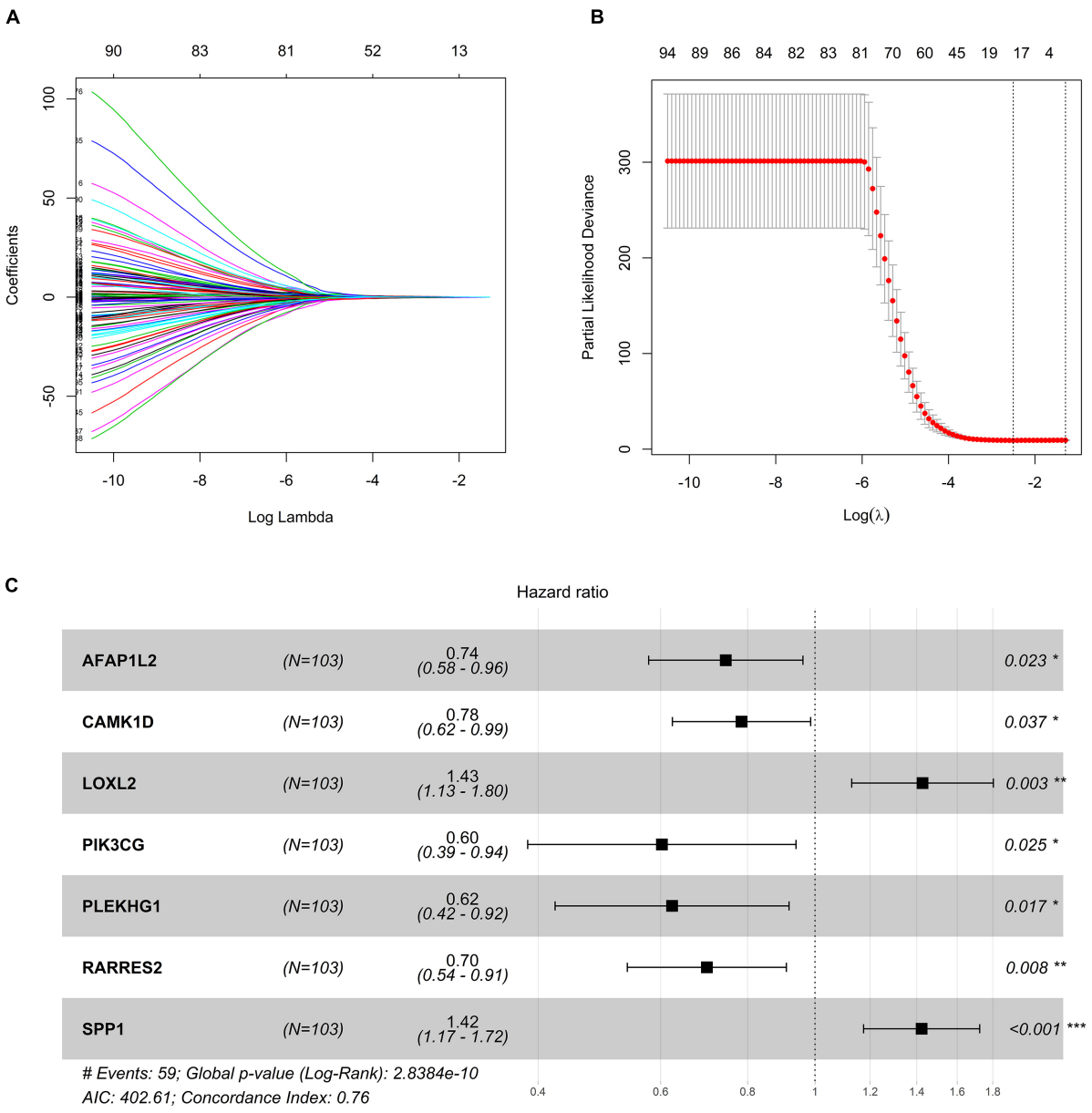


FIGURE 3 | Construction of seven-gene signature. **(A)** LASSO coefficient profiles. **(B)** Tenfold cross-validation result that identified optimal values of the penalty parameter λ. **(C)** Forest plot of seven hub genes based on stepwise regression method and multivariate Cox results. *p < 0.05, **p < 0.01, ***p < 0.001.

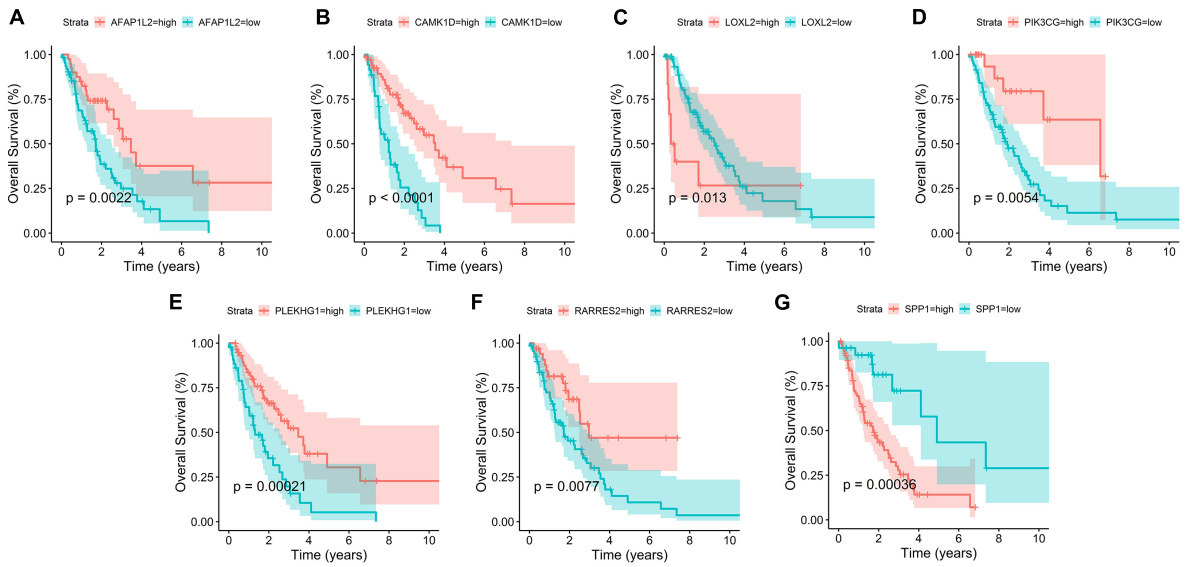


FIGURE 4 | Survival curves of seven prognostic hub genes. **(A–G)** Survival curves of seven DEGs were constructed to explore the prognostic value of each gene in the TCGA database.

showed a heatmap of 1,092 DEGs between high and low stromal scores.

For the immune score, there were 37 upregulated DEGs and 678 downregulated DEGs in the high group compared with the low group. For stromal score, compared with the low score group, there were 160 upregulated DEGs and 932 downregulated DEGs in the high-score group. A Venn diagram showed 18 cross-upregulated DEGs and 435 cross-downregulated DEGs between the immune and stromal groups (**Figures 2C,D**).

Function and Pathway Enrichment Analysis of DEGs

Functional enrichment analyses for DEGs, including BP, CC, MF, and KEGG pathways, were conducted using the David gene annotation tool. BP indicated that these genes may be associated with immune response, defense response, response to wounding, inflammatory response, and positive regulation of immune system process. CC indicated that these genes may be associated with intrinsic to membrane, integral to membrane, and plasma membrane. MF indicated that these genes may be associated with carbohydrate binding, cytokine binding, and polysaccharide binding (**Figure 2E**). The result of KEGG enrichment was related to immune response, including cytokine–cytokine receptor interaction, chemokine signaling pathway, cell adhesion molecules (CAMs), and hematopoietic cell lineage (**Figure 2F**). Overall, our results confirmed that TME-related DEGs were closely related to the anti-tumor immunity of advanced LUAD patients.

Construction of Seven-Gene Signature and Survival Analysis

In order to explore the potential role of DEGs in survival outcome, a univariate Cox proportional hazards regression model was first conducted, and the results showed that 96 DEGs were selected by univariate analysis. Next, according to the −2 log-likelihood test, the 10-fold cross-validation random sampling method was used, and LASSO regression analysis further screened out 18 DEGs (**Figures 3A,B**). Finally, a multivariate Cox proportional hazards model was performed, and a total of seven DEGs were selected to establish a seven-gene signature, and the seven-gene signature formula was as follows: risk score = (−0.29529*AFAP1L2) + (−0.24317*CAMK1D) + (0.35563*LOXL2) + (−0.50661*PIK3CG) + (−0.47294*PLEKHG1) + (−0.35771*RARRES2) + (0.35258*SPP1) (**Figure 3C**).

In addition, survival curves of seven DEGs were constructed to explore the prognostic value of each gene (**Figure 4**). Furthermore, a total of 51 patients with risk scores higher than the median were classified as "high-risk group," and the remaining 52 patients were classified as "low-risk group." K–M curves showed that the OS and PFS of high-risk patients were significantly worse ($p < 0.001$ and $p < 0.001$, respectively; **Figures 5A,B**).

In order to evaluate the predictive ability of the seven-gene signature, we drew the ROC curve based on the risk score and calculated the AUC of the area under the curve. The AUCs of the first, second, and third year of OS prognostic models were 0.783, 0.806, and 0.843, respectively (**Figure 5C**). Consistently, the AUCs of the first, second, and third year of PFS prognostic models were 0.733, 0.795, and 0.766, respectively (**Figure 5D**).

To explore the independent prognostic value of seven-gene signature, multivariate Cox analysis combined with other clinicopathologic characteristics showed that risk score was an independent predictor (For OS, HR: 6.42, 95% CI: 3.32–12.40; For PFS, HR: 4.74, 95% CI: 2.71–8.28) (**Table 2**).

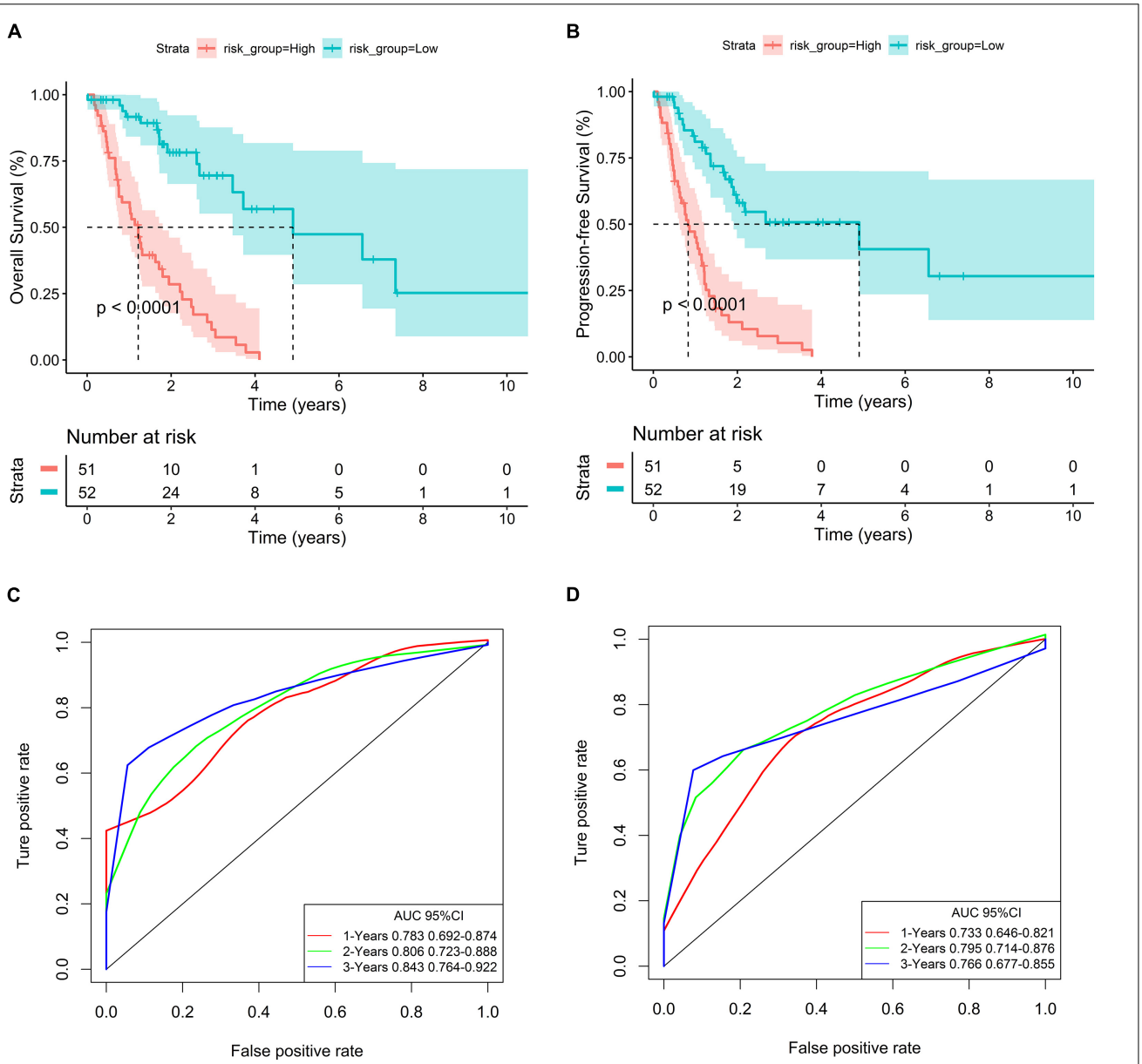


FIGURE 5 | Evaluation of the predictive ability of the seven-gene signature. **(A,B)** K–M curves for OS and PFS of high- and low-risk groups in the TCGA database. **(C,D)** ROC curves for OS and PFS based on the seven-gene signature in the TCGA database.

Validation of the Risk Stratification Model

In the GSE81089 and GSE41271 datasets (**Supplementary Figures 1, 2**, respectively), the correlation between seven genes and the risk score indicated that AFAP1L2, CAMK1D, PIK3CG, PLEKHG1, and RARRES2 were negatively correlated with the risk score, while LOXL2 and SPP1 were positively correlated with the risk score. The human protein atlas database was used to explore protein expression levels. Typical IHC of four favorable and two adverse prognostic genes (except RARRES2, which was not included in the database) in normal and tumor tissues is shown in **Supplementary Figure 3**.

In order to verify the generalization value of the seven-gene signature based on the TCGA cohort, we separately calculated the risk score of each sample for the 27 advanced LUAD patients in GSE81089 and the 53 advanced LUAD patients in GSE41271 using the above risk score formula. For the GSE81089 dataset, K–M survival curves indicated that the low-risk group had higher OS (p = 0.019) (**Figure 6A**). ROC curves based on the risk score model showed that the AUCs for the first, second, and third year of OS prognostic models were 0.746, 0.728, and 0.764, respectively (**Figure 6C**). Consistently, for the GSE41271 dataset, K–M survival curves indicated that the low-risk group also had higher OS (p = 0.04) (**Figure 6B**). ROC curves based on the risk

TABLE 2 | Multivariate Cox analysis of clinical information and risk group.

Variables	OS		PFS	
	HR (95% CI)	*p*	HR (95% CI)	*p*
Age		0.390		0.331
<65	1		1	
≥65	1.28 (0.73–2.26)		1.30 (0.77–2.19)	
Gender		0.946		0.355
Female	1		1	
Male	1.02 (0.58–1.80)		1.28 (0.76–2.17)	
History of smoking		0.092		0.122
No	1		1	
Yes	2.50 (0.86–7.28)		1.98 (0.83–4.70)	
Tumor laterality		0.719		0.546
Left	1		1	
Right	0.90 (0.49–1.64)		1.19 (0.68–2.09)	
Metastasis		0.205		0.506
M0	1		1	
M1	1.59 (0.78–3.27)		1.25 (0.65–2.39)	
Lymph node status		0.082		0.100
N0	1		1	
N1–N3	2.48 (0.89–6.87)		2.11 (0.87–5.14)	
Pathological T stage		0.360		0.895
T1–T2	1		1	
T3–T4	1.30 (0.74–2.28)		1.04 (0.61–1.76)	
Risk group		<0.001		<0.001
Low	1		1	
High	6.42 (3.32–12.40)		4.74 (2.71–8.28)	

score model showed that the AUCs for the first, second, and third year of OS prognostic models were 0.630, 0.653, and 0.623, respectively (**Figure 6D**).

Estimating the Composition of Immune Cells

We used CIBERSORT to estimate the immune cell composition of 103 samples and to quantify the relative levels of different cell types in the mixed cell population (**Figure 7A**). In patients with advanced LUAD, the expression level of macrophage M2 was significantly higher than that of macrophage M1 ($p < 0.001$). As shown in **Figure 7B**, we compared different cell types of patients in the low-risk group with those in the high-risk group. These results indicated that the expression levels of macrophages M2 and regulatory T cells (Tregs) in the high-risk group were significantly higher than those in the low-risk group ($p = 0.041$ and $p = 0.022$, respectively).

DISCUSSION

Early symptoms of LUAD are relatively insidious, without typical symptoms. As a result, most LUAD patients are at an advanced stage at the time of diagnosis, losing the best chance of surgical resection and affecting the treatment effect and quality of life of patients (Shapira, 2018). Fortunately, the treatment of LUAD continues to develop, from the original surgery, radiotherapy, chemotherapy, and targeted therapy to the current tumor immunotherapy, and the continuous innovation of treatment methods provides new treatment options for patients with advanced LUAD (Hanna et al., 2017). Previous studies have shown that TME plays a vital role in malignant progression, immune escape, and therapeutic resistance (Lambrechts et al., 2018). Therefore, it is important to study the TME of advanced LUAD in this study to determine biomarkers that can predict survival outcomes of patients.

In order to study the TME of advanced LUAD, we first calculated the immune score, stromal score, and estimate score of each advanced LUAD sample extracted from the TCGA database by applying an ESTIMATE algorithm. These patients were then divided into high/low immune score groups and high/low stromal score groups, and 453 cross-sectional DEGs were identified.

The GO and KEGG analyses of DEGs showed that DEGs mainly participated in TME, such as immune response, defense response, response to wounding, inflammatory response, and positive regulation of immune system process. These processes may inhibit tumor progression and metastasis, thereby improving the prognosis. We also found that these DEGs have a strong correlation with the immune response and tumor immune microenvironment.

In addition, we applied univariate Cox, LASSO, and multivariate Cox regression model to construct a gene signature based on seven DEGs that were screened from 453 cross-sectional DEGs. According to this gene signature, OS and PFS in the high-risk group were significantly worse than those in the low-risk group. Based on the LASSO model, Ma et al. (2020) established a prognostic model for patients with stage I–IV LUAD (AUC = 0.648). Based on the multivariate Cox model, our prognostic model for patients with advanced LUAD had more powerful predictive ability (The AUCs of the first, second, and third year of OS prognostic models were 0.783, 0.806, and 0.843, respectively). Therefore, survival outcomes in advanced LUAD patients could be well predicted by this seven-gene signature.

Among this seven-gene signature, we found that high expression levels of LOXL2 and SPP1 were associated with poor survival outcomes. In contrast, the higher the expression levels of AFAP1L2, CAMK1D, PIK3CG, PLEKHG1, and RARRES2, the better the survival outcomes. LOXL2 can promote the survival and drug resistance of tumor cells, regulate cell adhesion, movement and invasion, and reshape the TME (Barker et al., 2012). Upregulation of LOXL2 has been shown to promote lung cancer invasion and metastasis (Peng et al., 2017). Peinado et al. (2008) also showed that high LOXL2 expression was associated with reduced survival of patients with NSCLC. SPP1, also known as OPN, is a pleiotropic chemokine involved in the induction of tumor metastasis (Shi and Wang, 2017). In various types of cancer, elevated serum SPP1 levels are frequently detected in patients with metastatic cancer (Chiou et al., 2019). Advanced or metastatic LUAD patients with lower SPP1 levels had significantly

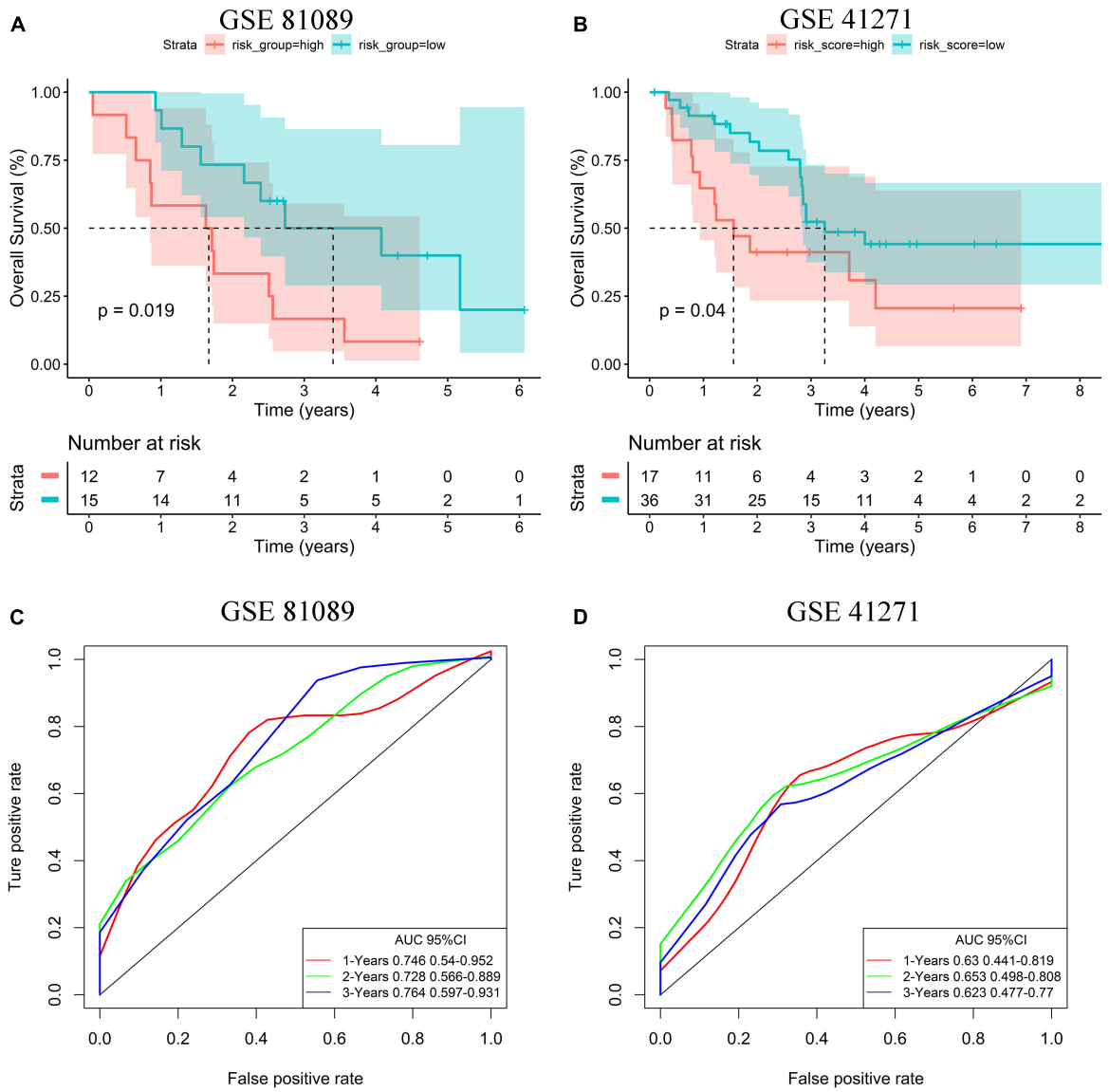


FIGURE 6 | Validation of the risk stratification model. **(A)** K–M curves for OS of high- and low-risk groups in the GSE81089 dataset. **(B)** K–M curves for OS of high- and low-risk groups in the GSE41271 dataset. **(C)** ROC curves for OS based on the seven-gene signature in the GSE81089 dataset. **(D)** ROC curves for OS based on the seven-gene signature in the GSE41271 dataset.

superior OS and PFS compared with patients with higher levels (Mack et al., 2008).

Actin filament-associated protein 1-Like 2 (AFAP1L2 also known as XB130) is a novel multifunctional adapter protein (Cho et al., 2019). AFAP1L2 mediates the innate immune response and inhibits tumor lung cancer cell proliferation and metastasis (Wang et al., 2020). CAMK1D, an inhibitory kinase, is a member of the calcium/calmodulin-dependent protein kinase 1 family. CAMK1D overexpression impairs tumor neoangiogenesis *in vivo*, thus achieving tumor inhibition (Dimitrova et al., 2016). PIK3CG is deemed to be a tumor suppressor gene (Kratz et al., 2002). Immunohistochemistry revealed a decreased PIK3CG expression in 85% of colorectal cancers, which was associated with tumor invasiveness and metastasis (Semba et al., 2002). RARRES2 is also known as chemerin (Shin et al., 2018). For LUAD, Yi et al. (Liu-Chittenden et al., 2017) found that the expression level of RARRES2 was positively correlated with NK cells in tumor invasion. Previous studies have also shown that higher RARRES2 expression was associated with positive prognosis in lung cancer patients (Zhao et al., 2011; Cai et al., 2016). PLEKHG1 belongs to a family of Rho-GEFs. Matthew et al. (Traylor et al., 2019) found that genetic variation in PLEKHG1 was associated with white matter hyperintensities and ischemic stroke. However, the relationship between PLEKHG1 and LUAD has not been reported, and PLEKHG1 may be a new therapeutic target for LUAD.

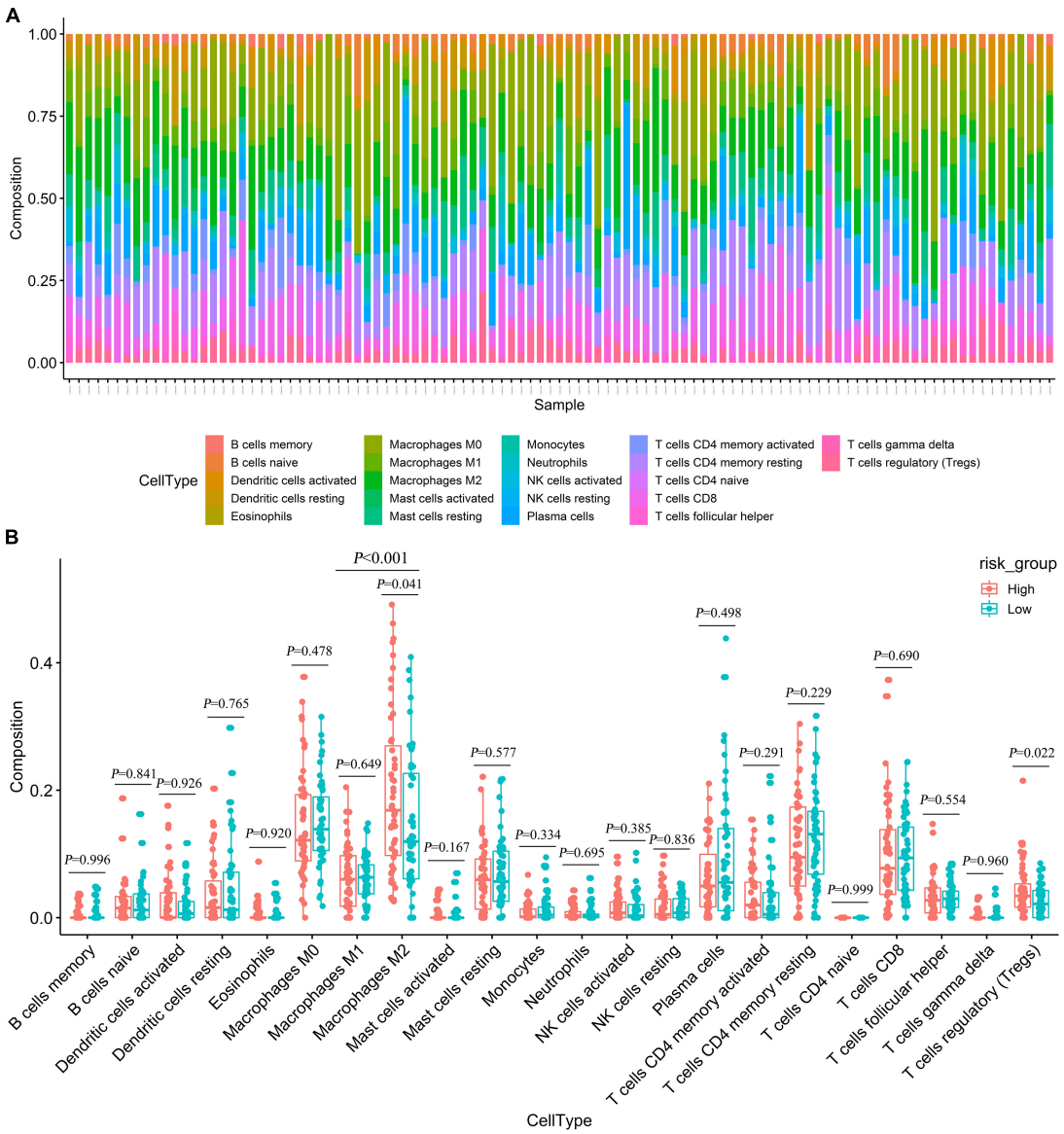


FIGURE 7 | Estimating the composition of immune cells **(A)** Relative proportion of 22 immune cell infiltration in high- and low-risk patients. **(B)** Differences of immune infiltration between high- and low-risk groups.

Currently, immunotherapy for advanced LUAD mainly uses checkpoint inhibitors, such as PD-1/PD-L1 inhibitors and CTLA-4 inhibitors, to activate the patient's own immune system to kill tumor cells. According to the different phenotypes and activation states of macrophages, they are classified into two polarized types: classically activated macrophages (macrophages M1) and alternatively activated macrophages (macrophages M2) (Martinez and Gordon, 2014). The macrophages M2 exhibit immunosuppression, which can promote tumorigenesis, angiogenesis, and metastasis (Noy and Pollard, 2014), and macrophages M1 play a key role in the anti-tumor immune effect (Mantovani et al., 2017). Our results showed that the proportion of macrophages M2 in advanced LUAD patients was significantly higher than that of macrophages M1. Although the ratio of macrophage M1/M2 in the high-risk group was lower than that in the low-risk group, it was not found to be statistically significant in our study, possibly due to the limitation of sample size. For the ratio of macrophage M1/M2, our study can be used as a hint, and further large sample data may be needed to verify this. These may indicate that the late stage of LUAD is related to the differentiation of macrophages M1 into macrophages M2. Interestingly, we also found that for advanced LUAD patients, the expression levels of Tregs and macrophages M2 in the high-risk group were significantly higher than those in the low-risk group. Tregs cells can attenuate the anti-tumor effects of CD4 T,

CD8 T, and NK cells (Frydrychowicz et al., 2017). Therefore, combination immunotherapy for inducing macrophages M2 to polarize macrophages M1 and regulating the function of Treg immunosuppressive cells may provide clues for the precise immune treatment of advanced LUAD patients and improving the effect of tumor immunotherapy.

However, this study also had certain limitations. First, this study only conducted bioinformatics research on public databases. Next, we should verify the results of this study through clinical patients in the prospective design. Second, our study provided evidence that seven TME-related genes were significantly related to the prognosis of advanced LUAD patients, but they were analyzed only through data mining merely. The biological function and mechanism of these genes depend on further experimental studies to elucidate.

CONCLUSION

In summary, our study established and validated a seven-gene signature associated with TME, which might serve as a prognosis stratification tool to provide a theoretical basis for predicting survival outcomes of advanced LUAD patients. In addition, macrophages M2 polarization may lead to worse prognosis in patients with advanced LUAD.

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, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethics committee approval for our study was not required because the data were obtained from publicly available databases.

AUTHOR CONTRIBUTIONS

HZ and CL conceived and designed the experiments. HZ and XZ analyzed the data. HZ and LG wrote the manuscript. SS and CL reviewed the draft. All authors have approved the final manuscript.

FUNDING

This research was funded by the National Key Research and Development Program of China (Grant Nos. 2017YFC1307705 and 2016YFC0106907). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are grateful for the joint efforts of the members of the research group. We also extend our thanks to all participants of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.684281/full#supplementary-material

Supplementary Figure 1 | Correlations between seven gene expression levels and risk score were found in GSE81089 dataset.

Supplementary Figure 2 | Correlations between seven gene expression levels and risk score were found in GSE41271 dataset.

Supplementary Figure 3 | Typical IHC of seven genes in normal and tumor tissues.

REFERENCES

Bai, F., Jin, Y., Zhang, P., Chen, H., Fu, Y., Zhang, M., et al. (2019). Bioinformatic profiling of prognosis-related genes in the breast cancer immune microenvironment. *Aging* 11, 9328–9347. doi: 10.18632/aging.102373

Barker, H. E., Cox, T. R., and Erler, J. T. (2012). The rationale for targeting the LOX family in cancer. *Nat. Rev. Cancer* 12, 540–552. doi: 10.1038/nrc3319

Bovelstad, H. M., Nygard, S., Storvold, H. L., Aldrin, M., Borgan, O., Frigessi, A., et al. (2007). Predicting survival from microarray data–a comparative study. *Bioinformatics* 23, 2080–2087. doi: 10.1093/bioinformatics/btm305

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Cai, Q., Huang, Z., Qi, L., Wang, T., Shen, Y., and Huang, J. (2016). Tazarotene-induced gene 2 is associated with poor survival in non-small cell lung cancer. *Oncol. Lett.* 12, 2680–2685. doi: 10.3892/ol.2016.5025

Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764

Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F., and Wong, K. K. (2014). Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat. Rev. Cancer* 14, 535–546. doi: 10.1038/nrc3775

Chiou, J., Chang, Y. C., Tsai, H. F., Lin, Y. F., Huang, M. S., Yang, C. J., et al. (2019). Follistatin-like Protein 1 inhibits lung cancer metastasis by preventing proteolytic activation of osteopontin. *Cancer Res.* 79, 6113–6125. doi: 10.1158/0008-5472.CAN-19-0842

Cho, H. R., Wang, Y., Bai, X., Xiang, Y. Y., Lu, C., Post, A., et al. (2019). XB130 deficiency enhances carcinogen-induced skin tumorigenesis. *Carcinogenesis* 40, 1363–1375. doi: 10.1093/carcin/bgz042

Dimitrova, N., Gocheva, V., Bhutkar, A., Resnick, R., Jong, R. M., Miller, K. M., et al. (2016). Stromal expression of miR-143/145 promotes neoangiogenesis in lung cancer development. *Cancer Discov.* 6, 188–201. doi: 10.1158/2159-8290.CD-15-0854

Frydrychowicz, M., Boruczkowski, M., Kolecka-Bednarczyk, A., and Dworacki, G. (2017). The dual role of treg in cancer. *Scand. J. Immunol.* 86, 436–443. doi: 10.1111/sji.12615

Gajewski, T. F., Schreiber, H., and Fu, Y. X. (2013). Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 14, 1014–1022. doi: 10.1038/ni.2703

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., et al. (2016). The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* 11, 39–51. doi: 10.1016/j.jtho.2015.09.009

Hanna, N., Johnson, D., Temin, S., Baker, S. Jr., Brahmer, J., Ellis, P. M., et al. (2017). Systemic therapy for stage IV non-small-cell lung cancer: American society of clinical oncology clinical practice guideline update. *J. Clin. Oncol.* 35, 3484–3515. doi: 10.1200/JCO.2017.74.6065

Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553, 446–454. doi: 10.1038/nature25183

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Wu, Y.-L., et al. (2017). Lung cancer: current therapies and new targeted treatments. *Lancet* 389, 299–311. doi: 10.1016/s0140-6736(16)30958-8

Jemal, A., Ward, E. M., Johnson, C. J., Cronin, K. A., Ma, J., Ryerson, B., et al. (2017). Annual report to the nation on the status of cancer, 1975-2014, featuring survival. *J. Natl. Cancer Inst.* 109:djx030 . doi: 10.1093/jnci/djx030

Kratz, C. P., Emerling, B. M., Bonifas, J., Wang, W., Green, E. D., Le Beau, M. M., et al. (2002). Genomic structure of the PIK3CG gene on chromosome band 7q22 and evaluation as a candidate myeloid tumor suppressor. *Blood* 99, 372–374. doi: 10.1182/blood.v99.1.372

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* 24, 1277–1289. doi: 10.1038/s41591-018-0096-5

Li, W., Xu, L., Han, J., Yuan, K., and Wu, H. (2019). Identification and validation of tumor stromal immunotype in patients with hepatocellular carcinoma. *Front. Oncol.* 9:664. doi: 10.3389/fonc.2019.00664

Liu-Chittenden, Y., Jain, M., Gaskins, K., Wang, S., Merino, M. J., Kotian, S., et al. (2017). RARRES2 functions as a tumor suppressor by promoting beta-catenin phosphorylation/degradation and inhibiting p38 phosphorylation in adrenocortical carcinoma. *Oncogene* 36, 3541–3552. doi: 10.1038/onc.2016.497

Luo, J., Xie, Y., Zheng, Y., Wang, C., Qi, F., Hu, J., et al. (2020). Comprehensive insights on pivotal prognostic signature involved in clear cell renal cell carcinoma microenvironment using the ESTIMATE algorithm. *Cancer Med.* 9, 4310–4323. doi: 10.1002/cam4.2983

Ma, C., Luo, H., Cao, J., Zheng, X., Zhang, J., Zhang, Y., et al. (2020). Identification of a novel tumor microenvironment-associated eight-gene signature for prognosis prediction in lung adenocarcinoma. *Front. Mol. Biosci.* 7:571641. doi: 10.3389/fmolb.2020.571641

Mack, P. C., Redman, M. W., Chansky, K., Williamson, S. K., Farneth, N. C., Lara, P. N. Jr., et al. (2008). Lower osteopontin plasma levels are associated with superior outcomes in advanced non-small-cell lung cancer patients receiving platinum-based chemotherapy: SWOG Study S0003. *J. Clin. Oncol.* 26, 4771–4776. doi: 10.1200/JCO.2008.17.0662

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, 399–416. doi: 10.1038/nrclinonc.2016.217

Martinez, F. O., and Gordon, S. (2014). The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* 6:13. doi: 10.12703/P6-13

NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 44, D7–D19. doi: 10.1093/nar/gkv1290

Neal, J. T., Li, X., Zhu, J., Giangarra, V., Grzeskowiak, C. L., Ju, J., et al. (2018). Organoid modeling of the tumor immune microenvironment. *Cell* 175, 1972–1988.e16. doi: 10.1016/j.cell.2018.11.021

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

Noy, R., and Pollard, J. W. (2014). Tumor-associated macrophages: from mechanisms to therapy. *Immunity* 41, 49–61. doi: 10.1016/j.immuni.2014.06.010

Peinado, H., Moreno-Bueno, G., Hardisson, D., Perez-Gomez, E., Santos, V., Mendiola, M., et al. (2008). Lysyl oxidase-like 2 as a new poor prognosis marker of squamous cell carcinomas. *Cancer Res.* 68, 4541–4550. doi: 10.1158/0008-5472.CAN-07-6345

Peng, D. H., Ungewiss, C., Tong, P., Byers, L. A., Wang, J., Canales, J. R., et al. (2017). ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis. *Oncogene* 36, 1925–1938. doi: 10.1038/onc.2016.358

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:e47. doi: 10.1093/nar/gkv007

Semba, S., Itoh, N., Ito, M., Youssef, E. M., Harada, M., Moriya, T., et al. (2002). Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-OH kinase, by CpG hypermethylation in human colorectal carcinoma. *Clin. Cancer Res.* 8, 3824–3831.

Shapira, O. M. (2018). Radial artery as the preferred second conduit for coronary bypass. *N. Engl. J. Med.* 378, 2134–2135. doi: 10.1056/NEJMe1804750

Shen-Orr, S. S., Tibshirani, R., Khatri, P., Bodian, D. L., Staedtler, F., Perry, N. M., et al. (2010). Cell type-specific gene expression differences in complex tissues. *Nat. Methods* 7, 287–289. doi: 10.1038/nmeth.1439

Shi, L., and Wang, X. (2017). Role of osteopontin in lung cancer evolution and heterogeneity. *Semin. Cell Dev. Biol.* 64, 40–47. doi: 10.1016/j.semcdb.2016.08.032

Shin, W. J., Zabel, B. A., and Pachynski, R. K. (2018). Mechanisms and functions of chemerin in cancer: potential roles in therapeutic intervention. *Front. Immunol.* 9:2772. doi: 10.3389/fimmu.2018.02772

Subramanian, J., and Govindan, R. (2008). Molecular genetics of lung cancer in people who have never smoked. *Lancet Oncol.* 9, 676–682. doi: 10.1016/s1470-2045(08)70174-8

Traylor, M., Tozer, D. J., Croall, I. D., Lisiecka-Ford, D. M., Olorunda, A. O., Boncoraglio, G., et al. (2019). Genetic variation in PLEKHG1 is associated with white matter hyperintensities (n = 11,226). *Neurology* 92, e749–e757. doi: 10.1212/WNL.0000000000006952

Vargas, A. J., and Harris, C. C. (2016). Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer* 16, 525–537. doi: 10.1038/nrc.2016.56

Wang, Q., Yang, G., Jiang, Y., Luo, M., Li, C., Zhao, Y., et al. (2020). XB130, regulated by miR-203, miR-219, and miR-4782-3p, mediates the proliferation and metastasis of non-small-cell lung cancer cells. *Mol. Carcinog.* 59, 557–568. doi: 10.1002/mc.23180

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4:2612. doi: 10.1038/ncomms3612

Zhao, S., Li, C., Ye, Y.-B., Peng, F., and Chen, Q. (2011). Expression of chemerin correlates with a favorable prognosis in patients with non-small cell lung cancer. *Lab. Med.* 42, 553–557. doi: 10.1309/lmww79nits6zadpt

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, Zhang, Guo, Shi 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 Genetics

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fgene.2021.694777

Identification of Signature Genes Associated With Invasiveness and the Construction of a Prognostic Model That Predicts the Overall Survival of Bladder Cancer

*Yang He[1], Yongxin Wu[1], Zhe Liu[1], Boping Li[1], Ning Jiang[1], Peng Xu[1] and Abai Xu[1,2]**

[1] Department of Urology, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [2] Department of Urology, The First People's Hospital of Kashgar Prefecture, Kashgar, China

Background: Bladder cancer has become the tenth most diagnosed cancer worldwide. The prognosis has been shown to differ between non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC). We aimed to identify signature genes that are associated with the invasiveness and survival of bladder cancer and to identify potential treatments.

Methods: We downloaded gene expression profiles of bladder cancer from the Gene Expression Omnibus database to identify differentially expressed genes and perform weighted gene co-expression network analysis. Functional enrichment was analyzed by GO and KEGG analyses. Hub genes were identified from the significant module. Another dataset was also acquired to verify the expression of hub genes. Univariate and multivariate Cox regression analyses were applied to the dataset downloaded from The Cancer Genome Atlas database. Risk scores were calculated and the effect was evaluated by Kaplan-Meier survival analysis. A nomogram was constructed and validated using training and testing samples, respectively. Analysis of the tumor immune microenvironment was conducted with the CIBERSORT algorithm.

Results: In total, 1,245 differentially expressed genes (DEGs) were identified. A distinct module was identified that was significantly correlated to invasiveness. The genes within this module were found to be significantly associated with extracellular exosomes, GTPase activity, metabolic pathways, etc. Three hub genes (VSIG2, PPFIBP2, and DENND2D) were identified as biomarkers of invasiveness; two of these (PPFIBP2 and DENND2D) were closely associated with prognosis. The risk score was regarded as an independent prognostic factor. The nomogram was associated with acceptable accuracy for predicting 1- and 5-year overall survival. The infiltrating levels of resting NK cells, activated natural killer (NK) cells, $CD8^+$ T cells, activated memory $CD4^+$ T cells, and T follicular helper cells, were significantly higher in the group with lower risk scores. The group with higher risk scores showed predominant infiltration by regulatory T cells (Tregs).

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Antonella Iuliano,
Telethon Institute of Genetics and Medicine (TIGEM), Italy
Chandra Shekar R. Ambati,
Baylor College of Medicine, United States

***Correspondence:**
Abai Xu
lc96xab@163.com

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

Received: *13 April 2021*
Accepted: *25 August 2021*
Published: *13 September 2021*

Citation:
He Y, Wu Y, Liu Z, Li B, Jiang N, Xu P and Xu A (2021) Identification of Signature Genes Associated With Invasiveness and the Construction of a Prognostic Model That Predicts the Overall Survival of Bladder Cancer. Front. Genet. 12:694777. doi: 10.3389/fgene.2021.694777

Conclusion: We successfully identified three signature genes related to invasiveness and constructed a nomogram of bladder cancer with acceptable performance. Differences suggested by risk scores between groups of patients showing diverse patterns of immune cell infiltration may be beneficial for selecting therapeutic approaches and predicting prognosis.

Keywords: bladder cancer (BLCA), invasiveness, weighted gene co-expression network analysis (WGCNA), nomogram, tumor immune microenvironment

INTRODUCTION

Bladder cancer (BLCA) has become the tenth most commonly diagnosed type of cancer worldwide, with approximately 573,000 new cases per year, a morbidity of 3.0%, and 213,000 deaths; the mortality rate associated with BLCA is 2.1% (Sung et al., 2021). In clinical diagnosis, approximately 75% of patients with bladder cancer are diagnosed with non-muscle invasive bladder cancer (NMIBC); the others are diagnosed with muscle invasive bladder cancer (MIBC); diagnosis is made according to whether the tumor invades the muscular layer of the bladder (Avgeris et al., 2018). The postoperative recurrence rate of NMIBC exceeds 70%; approximately 15% of these patients progress to MIBC. The postoperative recurrence rate of patients with MIBC can exceed 50% after radical cystectomy (RC), and many patients can die from this disease (Kotolloshi et al., 2021; Zhang et al., 2021). Because of the different prognoses and biological pathways underlying NMIBC and MIBC, a range of different treatment strategies may be required (Nouhaud et al., 2021). Therefore, it is crucial and beneficial to identify genes or features that are of prognostic value and to establish a prognostic model that could identify potential treatments and predict prognoses.

In this study, we analyzed mRNA expression data relating to BLCA from the Gene Expression Omnibus (GEO) database using differential gene expression analysis and weighted gene co-expression network analysis (WGCNA). Then, we identified the genes in the significant module that was most relevant to invasiveness and analyzed these genes using pathway and functional enrichment analyses. We also identified survival-associated hub genes and clinical signatures to predict the prognoses of patients with BLCA and developed a robust prognostic model to help direct treatment strategies and decision-making in the clinical treatment of BLCA patients. Finally, we investigated the diversity of tumor immune cell infiltration between different groups of patients with different risk scores. **Figure 1** shows a flowchart of the entire study process.

MATERIALS AND METHODS

Data Download and Processing

The GSE13507 microarray dataset features the gene expression profile of patients with BLCA (Lee et al., 2010) and contains 102 NMIBC samples and 63 MIBC samples. This dataset was downloaded from the GEO database.[1] We also downloaded the GSE120736 dataset (Song et al., 2019) which contains 78 NMIBC samples and 61 MIBC samples; this was used to verify the expression profiles of the signature genes. We also downloaded a standardized RNA-seq fragments per kilobase per million mapped reads (FPKM) dataset for BLCA from The Cancer Genome Atlas (TCGA) database[2] in order to construct a reliable prognostic model and investigate the tumor immune microenvironment (TIME).

The Identification of DEGs Between NMIBC and MIBC Samples

We utilized the "limma" package downloaded from Bioconductor[3] in R to analyze differentially expressed genes (DEGs) between NMIBC and MIBC samples. In order to select significant DEGs, we first normalized gene expression levels and then set the adjusted P-value to < 0.05 and the $|\log2 \text{ fold change}|$ to > 0.5 as thresholds.

Weighted Gene Co-expression Network Analysis

We used all of the genes in the GSE13507 dataset for WGCNA; this was performed using the "WGCNA" R package (Langfelder and Horvath, 2008). Our aim was to explore the relationships between significant expression modules and invasiveness. We created a sample clustering tree to detect and eliminate an outlier and set the soft thresholding power value to 9 in order to obtain a scale-free network. The resulting adjacency matrix derived from the gene expression set was then converted to a topological overlap matrix (TOM) for module clustering. We set 30 as the minimum number of genes in each module, and similar modules were merged with a threshold cut-off of 0.25. Next, we generated a hierarchical clustering dendrogram; distinct colors were assigned to diverse branches to reveal different modules. The vital clinical trait was then integrated into the eigengene network as an auxiliary node to explore the connection between the trait and the modules. Module-trait associations were then evaluated by analyzing the correlation between module membership (MM) and gene significance (GS). Modules that were highly correlated with invasiveness were selected and extracted to perform subsequent analysis.

Enrichment Annotation Analysis

Genes identified in the crucial module were then analyzed by Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

[1] https://www.ncbi.nlm.nih.gov/geo/

[2] https://portal.gdc.cancer.gov/

[3] http://www.bioconductor.org/

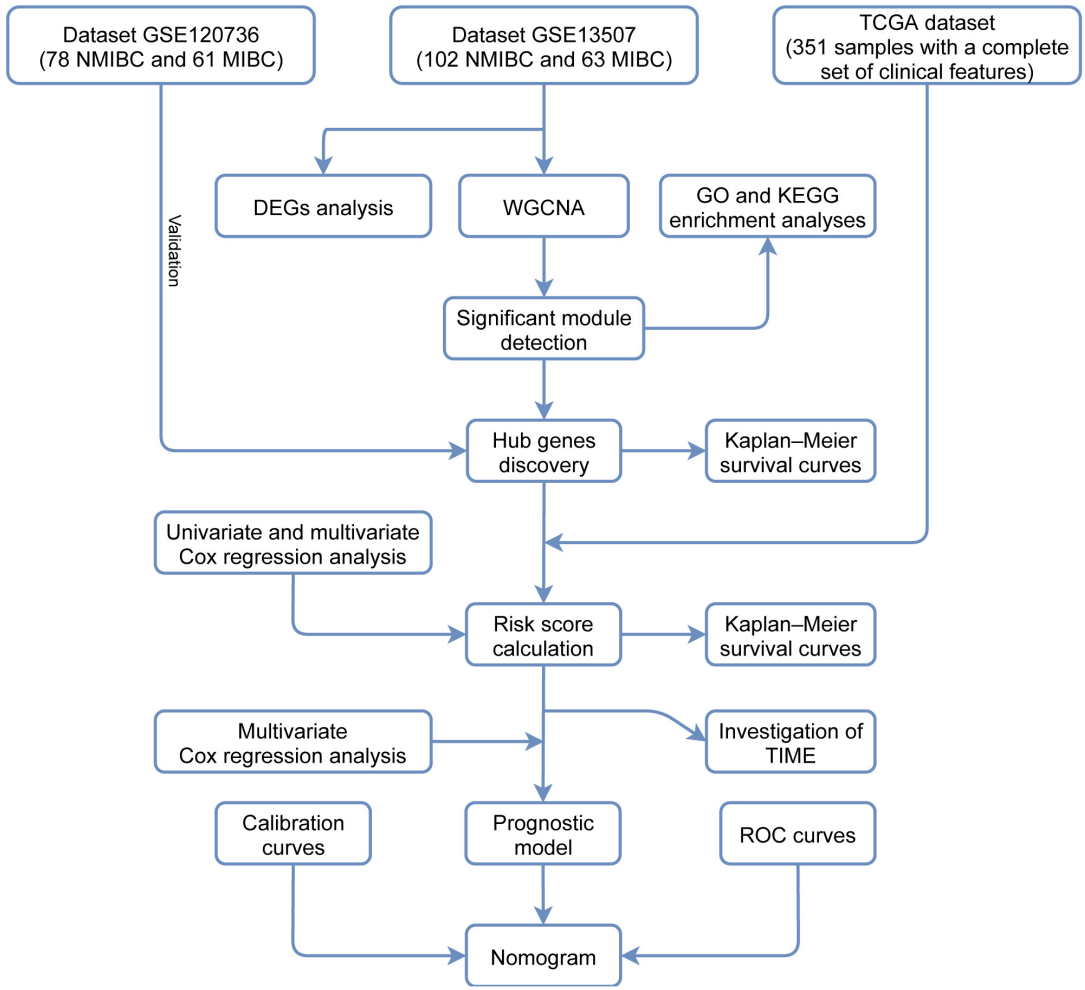


FIGURE 1 | A flowchart depicting the entire study process.

For this, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID)[4] an adjusted $P < 0.05$ was considered significant.

Hub Genes

To ensure quantity and accuracy, genes in the most significant module (with $|MM| > 0.86$ and $|GS| > 0.42$), as determined by WGCNA, were recognized as hub genes. Kaplan-Meier survival analyses were also adopted to determine the differences in overall survival (OS) between groups of hub genes when expressed at high and low levels.

Identification of Prognostic Signature Genes

We used a TCGA dataset with a complete set of clinical features ($n = 351$) to identify genes that were relevant to prognosis. Univariate Cox regression analysis was applied to explore the prognostic value of hub genes. After filtration, genes with $P < 0.1$ were selected for multivariate Cox regression analysis to evaluate the interactions between prognosis-related genes; this was carried out with the "survival" package in the R environment.

Establishment of a Prognostic Model

Next, a prognostic risk score model was developed based on prognosis-associated genes, expression levels, and coefficients. The risk score was calculated using an established formula (Sullivan et al., 2004), as follows:

$$\text{Risk Score} = \sum_{i=1}^{n} \text{Coefficient}_i \times \text{Expression}_i$$

Based on the median risk score, we separated all samples from TCGA dataset into two different groups. We carried out multivariate Cox regression analysis again and calculated hazard ratios (HRs) to identify the independence of the risk score for predicting overall survival.

[4]https://david.ncifcrf.gov

Construction and Validation of a Nomogram

Samples from TCGA dataset (with a complete set of clinical features) were divided into training (n = 252) and testing (n = 99) cohorts randomly. A nomogram derived from the training dataset was constructed by the "rms" package in R with the following clinical features: age, gender, risk score, T stage, and N stage. We created calibration plots to examine the predictive performance of the nomogram. A receiver operating characteristic (ROC) curve was derived from the testing cohort and used to check the accuracy of the nomogram based on a prognostic model; this was performed with the "survivalROC" package in R.

Investigation of the Tumor Immune Microenvironment

TIME analysis was applied to samples from TCGA dataset; this was carried out with the "CIBERSORT" analytic tool. We determined the proportions of 22 different tumor-infiltrating immune cells; $P < 0.05$ was considered to be the level of statistical significance (Newman et al., 2015). We also used the Wilcoxon rank-sum test to detect significant differences in the proportions of immune cell infiltration between low- and high-risk groups of patients.

Statistical Analysis

R software (version 4.0.3), and a range of tools within the R environment, were used for statistical analysis. We employed univariate and multivariate Cox regression analyses to determine prognostic factors. Kaplan-Meier curves were utilized to compare the OS of different groups, and statistical significances were verified with the log-rank test. Two groups of independent non-parametric samples were evaluated by the Wilcoxon rank-sum test.

RESULTS

A Comparison of DEGs Between NMIBC and MIBC Samples

Samples from the GSE13507 dataset were normalized and separated according to invasiveness. We identified 1245 DEGs (780 upregulated genes and 465 downregulated genes) using specific cut-off criteria (adjusted $P < 0.05$ and $|\log2 \text{ fold change}| > 0.5$) (**Figures 2A,B**).

The Construction of Weighted Gene Co-expression Network

WGCNA was adopted on samples from the GSE13507 dataset in order to identify genes related to invasiveness. A sample clustering tree was obtained, and an outlier (GSM340606) was detected and eliminated (**Figure 3A**). Next, we needed to set an appropriate soft threshold; power values ranged from 1 to 20 and a power (β) of 9 was selected to obtain a scale-free network (**Figure 3B**). The scale-free topology fitting index R^2 reached 0.85 (**Figure 3C**), thus fulfilling the requirements of scale-free topology. TOM was transformed from an adjacency matrix for clustering modules. Using the dynamic tree cutting and merging method and taking 0.25 as the cut-off point and 30 as the minimum number of genes, we merged similar modules. This strategy ultimately revealed 24 modules from 18,575 genes (**Figure 3D**). Correlation factors were calculated and then displayed as a heatmap (**Figure 4A**). The midnight-blue module, containing 240 genes (**Supplementary Table 1**), were highly correlated with invasiveness (correlation coefficient =–0.51; $P = 2 \times 10^{-12}$) and grade (correlation coefficient = −0.44; $P = 3 \times 10^{-9}$). The trait was then rescaled, using MIBC as the reference; the connection between the invasiveness and each module is shown in **Figure 4B** and **Supplementary Figure 1**. There was a strong correlation between the midnight-blue module and phenotype (correlation coefficient = 0.69; $P = 2.9 \times 10^{-35}$), as shown by **Figure 4C**.

Function and Pathway Enrichment Annotation Analysis of the Significant Module

Next, we employed GO and KEGG analyses to explore the function and pathway enrichment of the genes involved. The genes in the midnight-blue module were significantly related to protein homodimerization activity, integral component of Golgi membrane, extracellular exosome, lipid metabolic process, epithelial cell differentiation, positive regulation of GTPase activity, negative regulation of transforming growth factor beta receptor signaling pathway, and thymic T cell selection, etc. (**Figure 5A**, $P < 0.05$). KEGG pathway analysis identified significant enrichment in metabolic pathways; valine, leucine and isoleucine degradation; and ovarian steroidogenesis (**Figure 5B**, $P < 0.05$).

The Identification of Hub Genes From the Significant Module

Hub genes were selected using specific criteria ($|\text{MM}| > 0.86$, and $|\text{GS}| > 0.42$). Three genes were identified: *VSIG2*, *PPFIBP2*, and *DENND2D* (**Supplementary Table 2**). The expression levels of hub genes in the NMIBC and MIBC groups were visualized as violin plots (**Figures 6A–C**) using the "ggstatsplot" R package and validated with the GSE120736 dataset (**Figures 6D–F**); this analysis demonstrated good levels of consistency. The clinicopathological characteristics of patients in the two GEO cohorts are shown in **Table 1**. Kaplan-Meier survival curves were used to demonstrate the differences in OS between the low- and high- expression groups of hub genes (**Figures 6G–I**).

The Identification of Prognostic Signature Genes

Hub genes extracted from the midnight-blue module were then subjected to univariate Cox regression (**Table 2**). The exclusion criterion was set to $P > 0.1$. Multivariate Cox regression was then applied to the hub genes that passed the exclusion criterion (**Figure 7A**). The clinical characteristics of patients in TCGA cohort used for analysis are shown in **Supplementary Table 3**.

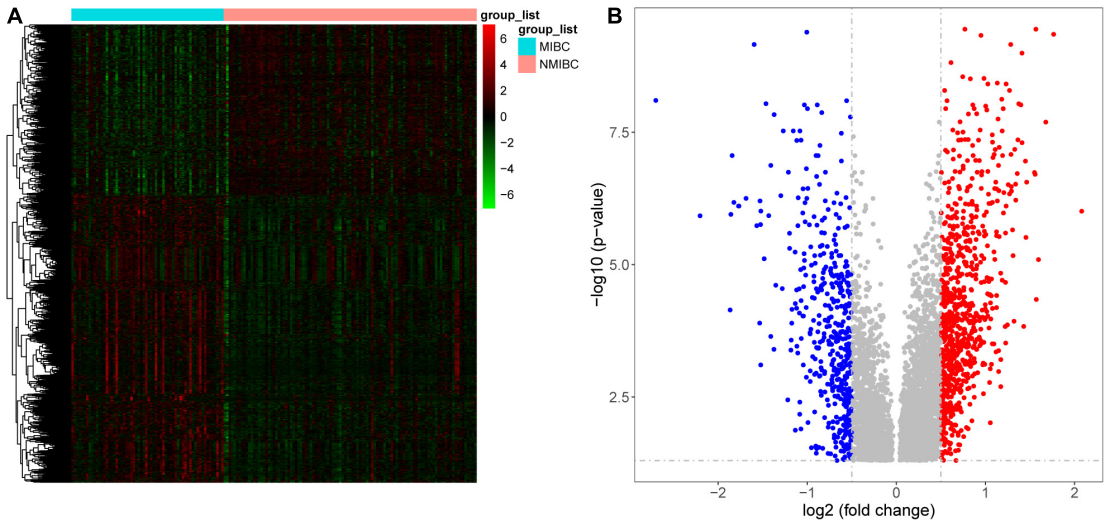


FIGURE 2 | Differentially expressed genes between patients with non-muscle invasive bladder cancer (NMIBC) and muscle invasive bladder cancer (MIBC) displayed as a heatmap **(A)** and a volcano map **(B)**.

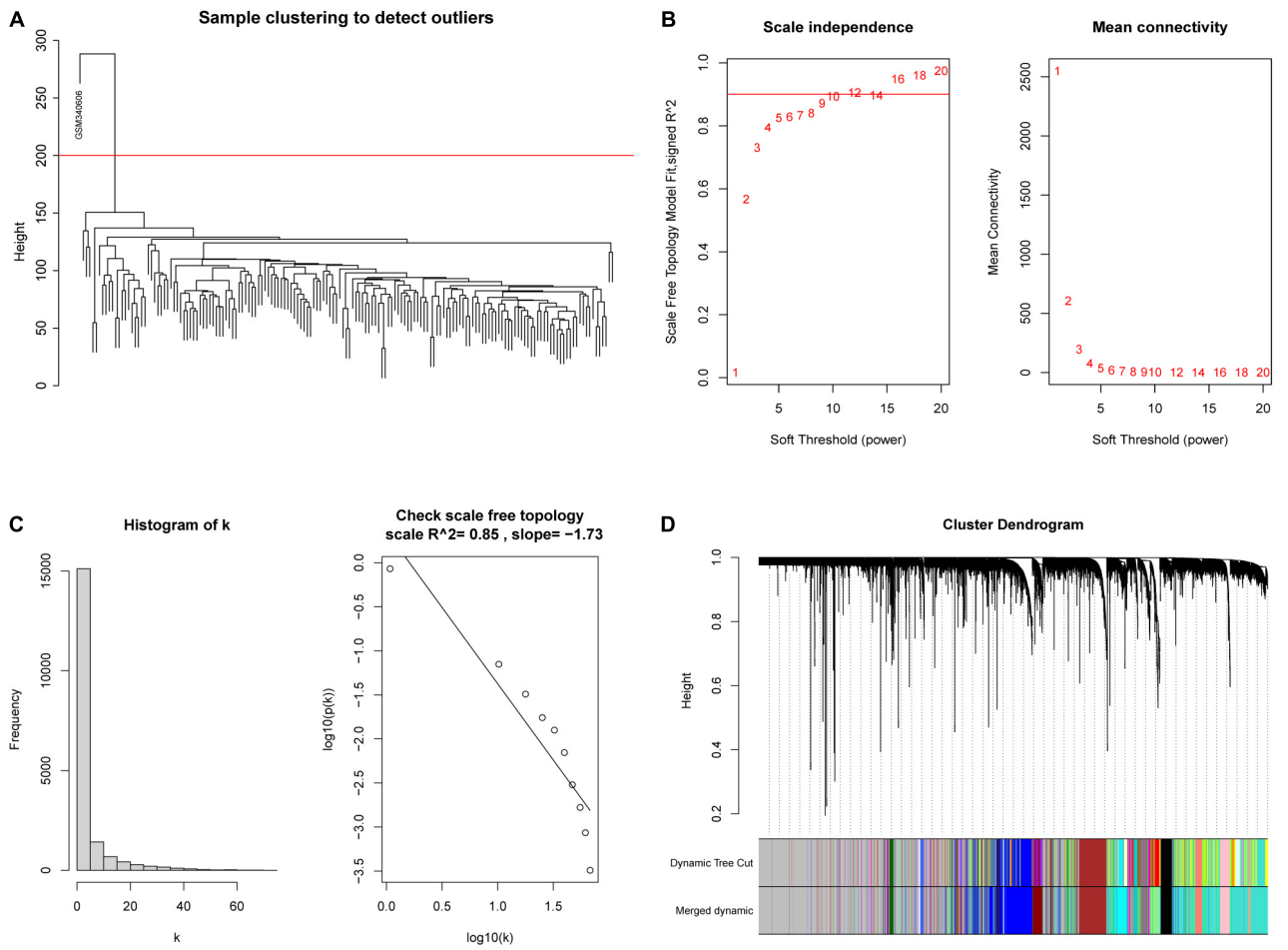


FIGURE 3 | Construction of a weighted co-expression network. **(A)** An outlier (GSM340606) was detected and eliminated by sample clustering. **(B)** β = 9 was selected as the soft threshold. **(C)** Validation of the scale-free topology network. **(D)** The dynamic tree cutting and merging method resulted in the identification of 24 modules.

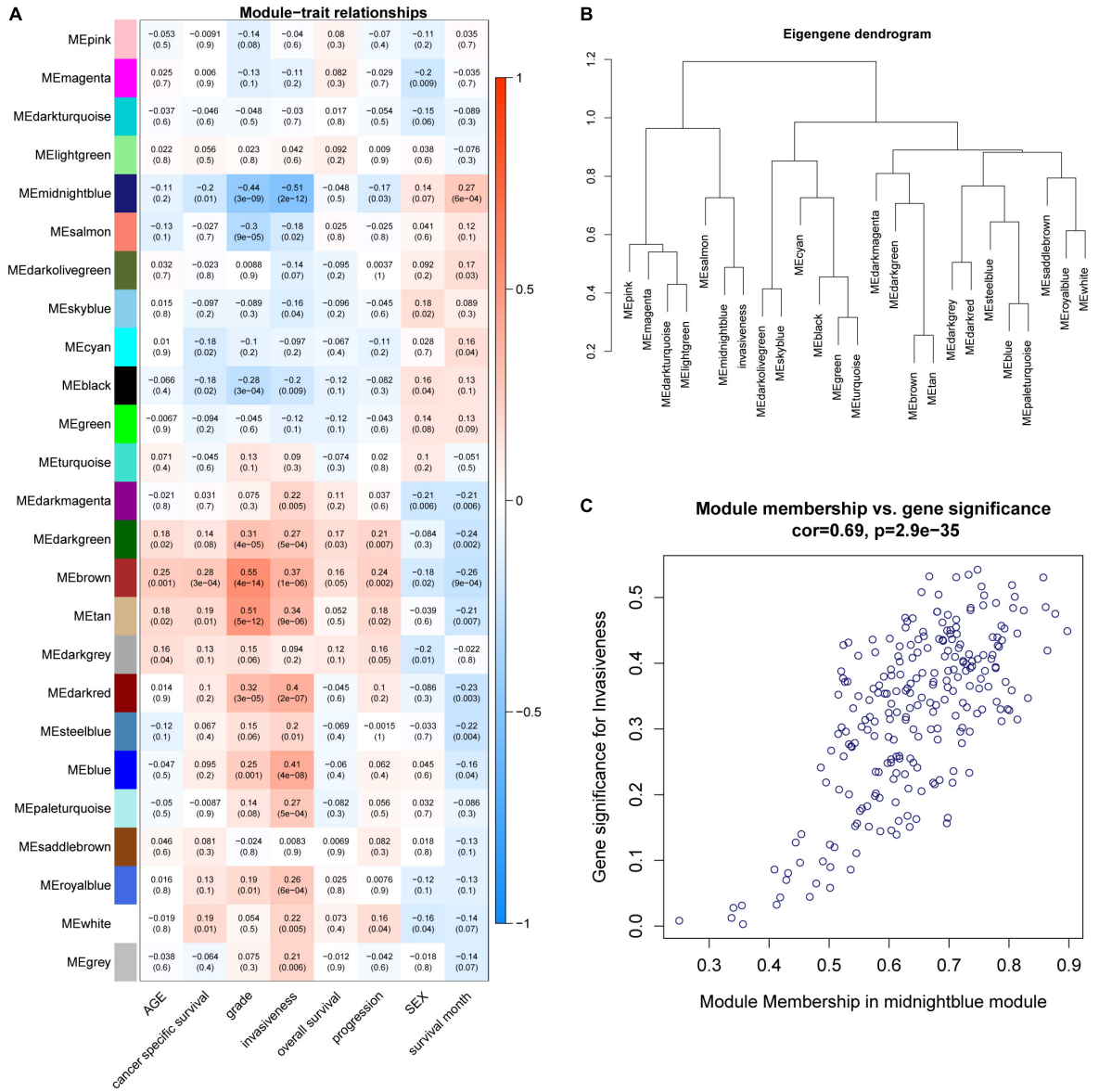


FIGURE 4 | (A) A heatmap was created to display the relationships between modules and traits; correlation coefficients and *P*-values are also shown for each module. **(B)** Invasiveness was integrated into the eigengene network to explore the connection between the trait and modules. **(C)** Correlation between module membership (MM) and gene significance (GS).

PPFIBP2 and *DENND2D* were identified as the hub genes that had the most influence on prognosis.

Establishment of a Prognostic Model Based on Hub Genes

Next, we calculated the risk score for each sample used expression levels of hub genes and coefficients, as follows: Risk Score = (0.06219 × *VSIG2*) + (−0.20005 × *PPFIBP2*) + (−0.29484 × *DENND2D*). The risk score, based on invasiveness, was regarded as an independent prognostic factor for survival, as demonstrated (**Figure 7B**). According to the median risk score in TCGA cohort, we separated patients into different groups (**Figure 7C**). The overall survival of low-risk patients was significantly longer than that of high-risk patients, as indicated by survival analysis (**Figure 7D**).

Construction and Validation of a Nomogram for Predicting Prognosis of Patients With Bladder Cancer

To predict the prognosis of patients with BLCA, we developed a nomogram using the training cohort from TCGA dataset (**Figure 7E**). Consequently, 1- and 5-year OS can be estimated according to the total number of points; the risk score had the greatest weighing in this calculation. A patient with a lower

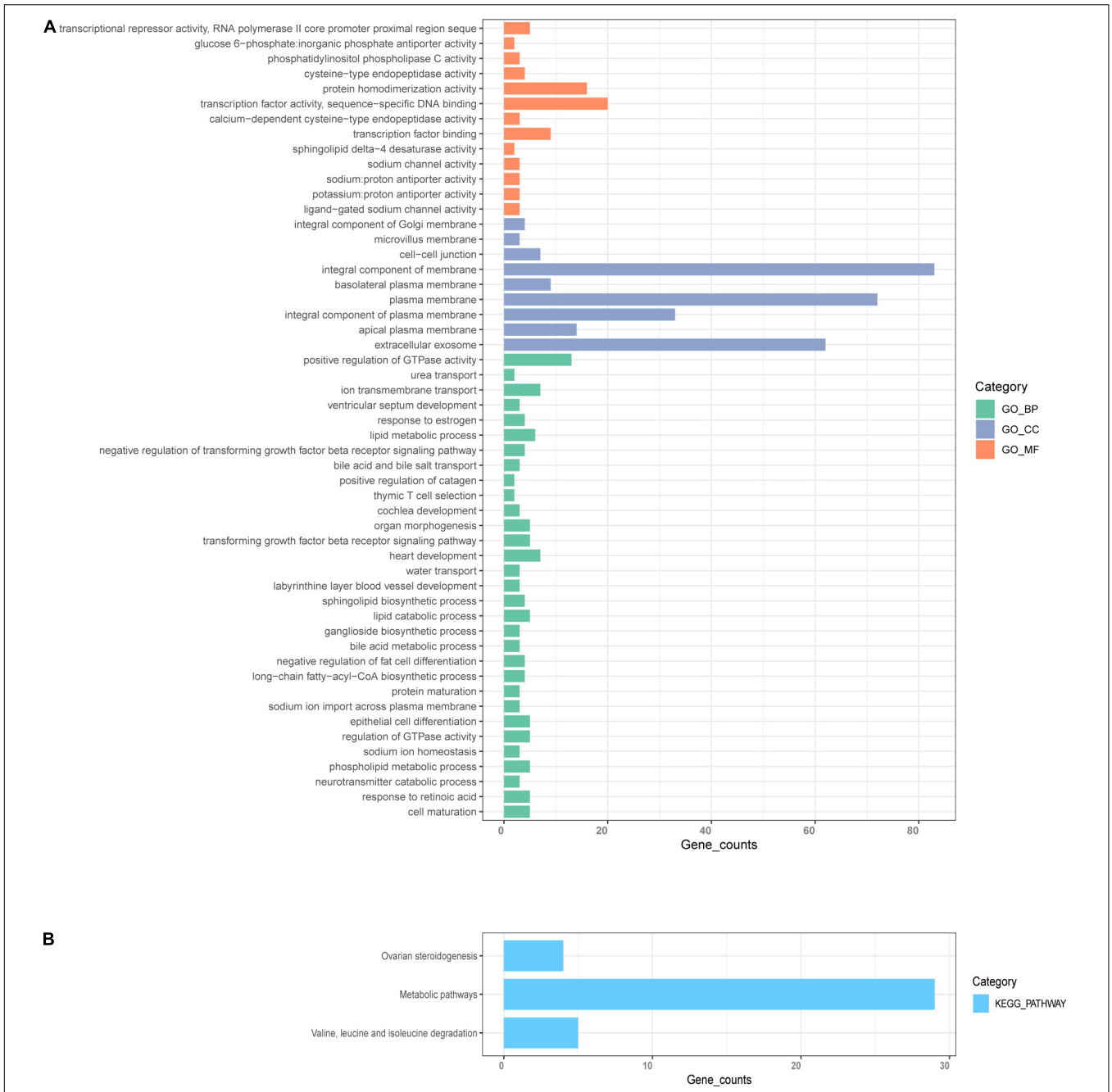


FIGURE 5 | GO and KEGG enrichment analysis. **(A)** GO analysis for genes in significant modules, including biological process (BP), cellular component (CC), and molecular function (MF). **(B)** KEGG enrichment analysis for genes in the significant modules.

risk score, T stage, N stage, or age, has a higher likelihood of a better prognosis. 0.67375977 was found as the C-index for the nomogram. Acceptable consistency between nomogram predictions and actual 1- and 5-year OS was determined by calibration curves (**Figures 7F,G**). In addition, we also developed time-dependent ROC curves for the established nomogram using the testing cohort (**Figure 7H**); the area under the curve (AUC) values for 1- and 5-year OS were 0.756 and 0.739, respectively.

Association of Risk Score With Immune Cell Infiltration in the Tumor Microenvironment

The proportions of 22 types of immune cells were determined for different groups of BLCA patients (**Figure 8A**). We then compared the abundances of tumor-infiltrating immune cells in groups with different levels of risk scores (**Figure 8B**). The infiltrating levels of resting NK cells, activated natural killer (NK)

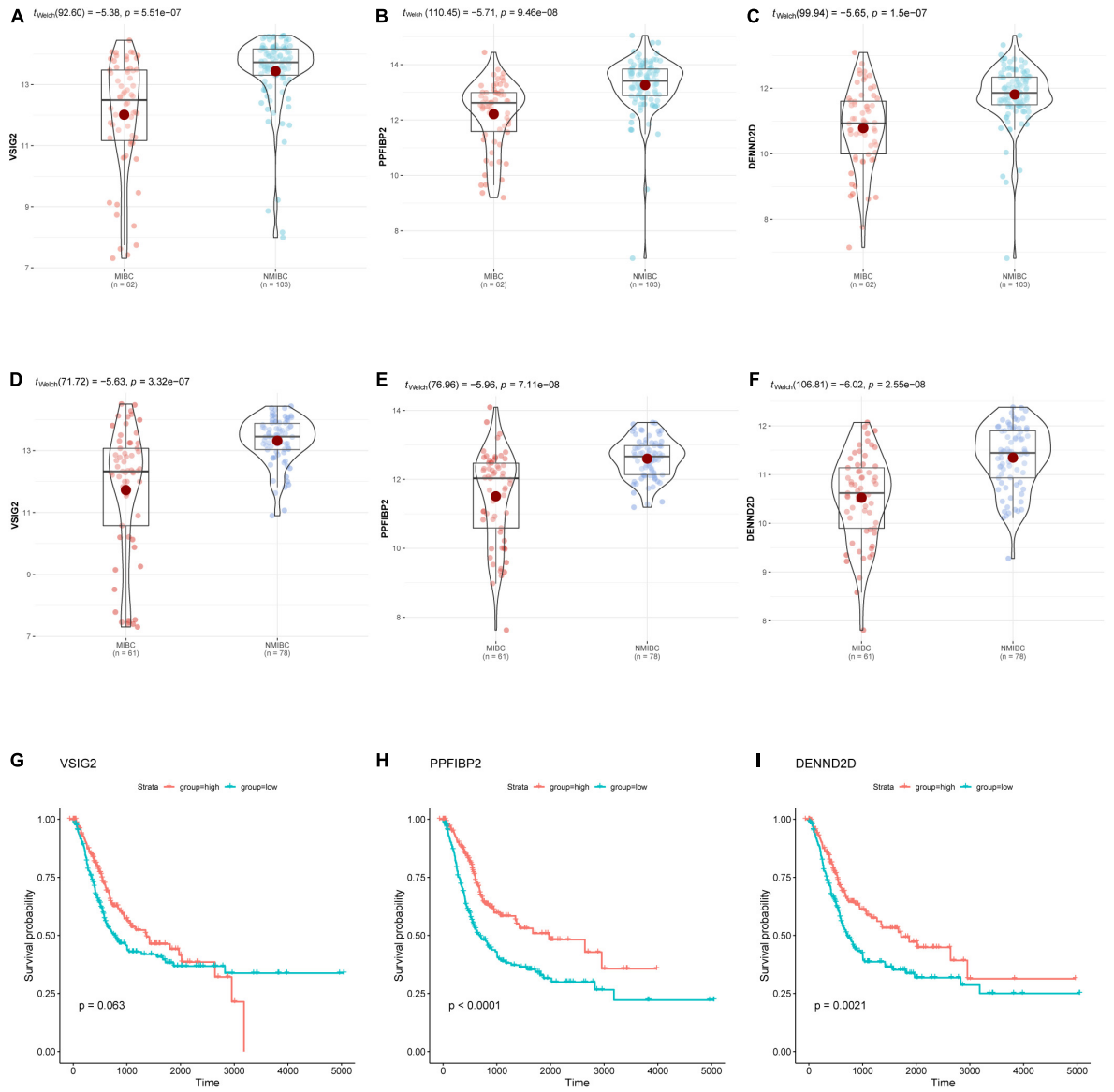


FIGURE 6 | The expression of hub genes (*VSIG2*, *PPFIBP2*, and *DENND2D*) in the GSE13507 dataset **(A–C)** and the GSE120736 dataset **(D–F)**. Kaplan-Meier survival curves for patients with different expression levels of hub genes **(G–I)**. The expression levels of groups were classified according to the median gene expression value.

cells, $CD8^+$ T cells, activated memory $CD4^+$ T cells, and T follicular helper cells, were significantly higher in the low-risk group. The high-risk group showed predominant infiltration by regulatory T cells (Tregs).

DISCUSSION

BLCA is associated with morbidity and mortality rates of 3.0 and 2.1% worldwide, respectively (Sung et al., 2021). Because of the diverse range of biological mechanisms that exist between NMIBC and MIBC, it is important that we consider whether to provide different treatment strategies. Previous studies have established multiple prognostic models (Chen et al., 2021; Xu et al., 2021). Furthermore, Yan et al. (2020) built an eight-gene signature to predict OS in patients with BLCA using WGCNA. Qiu et al. (2020) and Shen et al. (2020) also constructed prognostic models based on immune-associated genes; however, none of these established prognostic models were based on invasiveness. Identifying biomarkers relating to the invasiveness of BLCA and using these biomarkers to establish a prognostic model may provide new options for treatment selection and the prediction of prognosis.

In the present study, we aimed to identify signature genes associated with invasiveness and construct a prognostic model based on these signature genes. In total, we identified 1245

TABLE 1 | Basic clinical characteristics of patients in the GEO datasets.

Variables	GSE13507	GSE120736	*P*-value
Patients	165	139	
Type			
NMIBC	102 (61.8)	78 (56.1)	0.313
MIBC	63 (38.2)	61 (43.9)	
Gender			
Female	30 (18.2)	24 (17.3)	0.835
Male	135 (81.8)	115 (82.7)	
Size			
<3 cm	69 (41.8)	58 (41.7)	0.776
≥3 cm	89 (53.9)	80 (57.6)	
NA	7 (4.2)	1 (0.7)	
Tumor number			
Single	98 (59.4)	88 (63.3)	0.198
2–7	45 (27.3)	48 (34.5)	
>8	21 (12.7)	3 (2.2)	
NA	1 (0.6)		
Stage			
Ta	24 (14.5)	37 (26.6)	0.676
T1	78 (47.3)	41 (29.5)	
T2–T4	63 (38.2)	61 (43.9)	

TABLE 2 | Univariate Cox regression analysis of hub genes.

Hub genes	Beta	HR (95% CI for HR)	Wald	*P*-value
VSIG2	–0.055	0.95 (0.89–1)	3.5	0.062
PPFIBP2	–0.21	0.81 (0.71–0.92)	10	0.0016
DENND2D	–0.33	0.72 (0.6–0.86)	13	0.00 035

DEGs, thus illustrating clear differences in gene expression between NMIBC and MIBC. A specific module, containing 240 genes, was identified by WGCNA, and found to be significantly associated with invasiveness. Several vital pathways were revealed by functional enrichment analysis. Three hub genes (*VSIG2*, *PPFIBP2*, and *DENND2D*) were extracted from this module; two of these genes (*PPFIBP2*, *DENND2D*) played an important role in prognosis as protective factors. We established a prognostic model and presented this model as a nomogram based on the prognosis-associated signatures. Further validation of the nomogram, with calibration curves and time-dependent ROC curves, suggested an acceptable level of accuracy. As well as identifying new biomarkers related to invasiveness and developing a prognostic model, we also used TIME to investigate potential treatment options.

The midnight-blue module was mainly associated with metabolic-related activities, showed by GO functional analysis; this was consistent with the results derived from KEGG pathway enrichment analysis. Analyses showed that several pathways were enriched, including epithelial cell differentiation, and negative regulation of transforming growth factor beta receptor signaling pathway that inhibits the progression of tumor metastasis, as described previously (Tan et al., 2017; Lyu et al., 2019). The current analyses also identified enrichment in positive regulation of GTPase activity, thymic T cell selection, and extracellular exosome, etc.

Three hub genes (*VSIG2*, *PPFIBP2*, and *DENND2D*) were identified as biomarkers that showed upregulation in patients with NMIBC and downregulated in patients with MIBC. *VSIG2* is expressed in the thymus and may be related to antigen presentation (Kariuki et al., 2010). The function of *VSIG2* in cancer has not been described previously, although this gene may serve as a potential biomarker for BCLA. The *PPFIBP2* product is known to be associated with axon guidance and the development of neuronal synapses (Iyama et al., 2017). The protein encoded by *PPFIBP2* is liprin-β2; previous studies have shown that low levels of liprin-β2 is associated with a poor prognosis for urothelial, renal, prostate, lung, head, and neck cancers (Tan et al., 2008; Wu et al., 2018). Tumor cell migration, invasion, and progression, are controlled by the ERK pathway (Ueoka et al., 2000; Reddy et al., 2003). ERK2 is a major facilitator of cell migration and invasion within the tumor microenvironment; liprin-β2 represents a specific target for ERK2. Liprin-β2 can inhibit cell migration and invasion and acts downstream of ERK2. Consequently, liprin-β2 may act by facilitating the transporting of anti-migratory molecules or by halting the recycling of pro-invasive molecules. ERK2 is also known to drive invasiveness by inhibiting liprin-β2 (von Thun et al., 2012). *DENND2D* is considered to act as a tumor suppressor gene and has been implicated in several types of cancer, including hepatocellular, lung, esophageal, and gastric cancer (Ling et al., 2013; Hibino et al., 2014; Kanda et al., 2014, 2015). *DENND2D* is a regulator of Rab GTPases and is highly associated with carcinogenesis and the progression of cancer. In addition, several Rab GTPase family members are known to influence the secretion of exosomes *via* the trans-Golgi network or inducible vesicular transporting (Ponnambalam and Baldwin, 2003; Ostrowski et al., 2010). Adjacent cells can take up exosomes derived from cancer cells, which are able to induce pathways that are involved in the initiation and progression of cancer (Rink et al., 2005; Henderson and Azorsa, 2012). MiR-1246 in tumor exosomes can directly target and downregulate *DENND2D*, as reported previously (Sakha et al., 2016). *DENND2D* is also involved in the miR-522-induced migration and invasion of non-small cell lung cancer cells by targeting *DENND2D* (Zhang et al., 2016). Higher expression levels of *PPFIBP2* and *DENND2D* are known to be associated with lower levels of tumor invasiveness and a better prognosis; our present findings were consistent with these earlier observations.

TIME plays a key role in tumor initiation and progression; furthermore, immunotherapy is often performed as a component of neoadjuvant therapy (Lyu et al., 2020). We assessed tumor immune cell infiltration to explore potential therapies and other prognostic factors, and we then investigated the differences in TIME between different groups. The proportions of resting NK cells, activated NK cells, $CD8^+$ T cells, activated memory $CD4^+$ T cells, and Tfh cells, were significantly higher in the low-risk group, thus indicating a better prognosis. In contrast, the proportion of Tregs was lower in the low-risk group than in the high-risk group. NK cells are known to kill adjacent cells that express surface markers that are associated with oncogenic transformation (Shimasaki et al., 2020). By secreting cytokines and chemokines, NK cells may induce T cell infiltration and inflammation; they may also prevent metastasis by eliminating

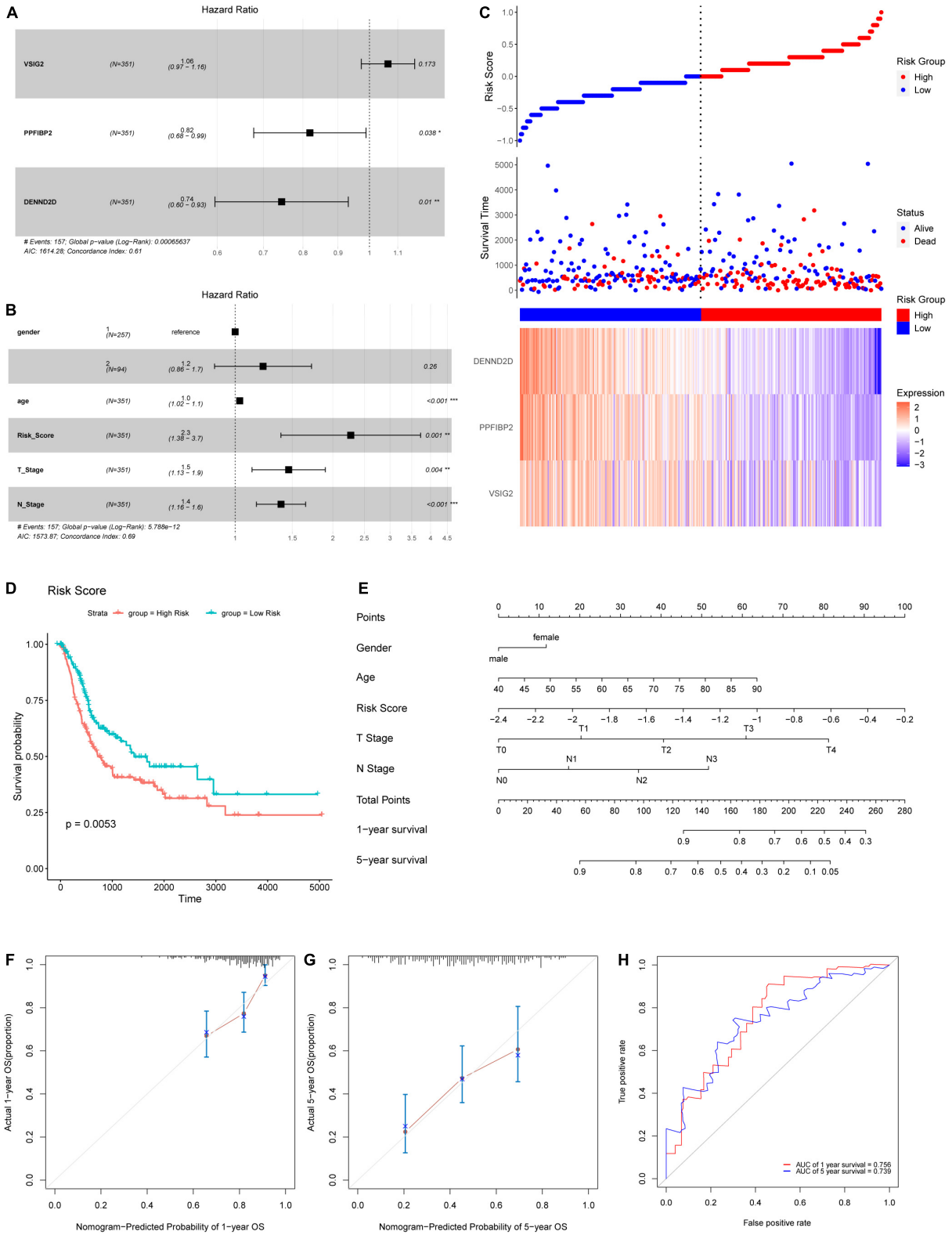


FIGURE 7 | (A) A Forest map of hub genes, as determined by multivariate Cox regression analysis. **(B)** A Forest map of risk score and other clinical features, as determined by multivariate Cox regression analysis. **(C)** Risk score distribution, survival status, and an expression heatmap for the three hub genes. **(D)** Kaplan-Meier survival curves for patients in the low- and high-risk groups. **(E)** Nomogram for predicting the 1- and 5-year OS of patients with BLCA. **(F,G)** A calibration curve of the constructed nomogram for predicting 1- and 5-year OS. **(H)** The time-dependent ROC curves for the constructed nomogram using the test data.

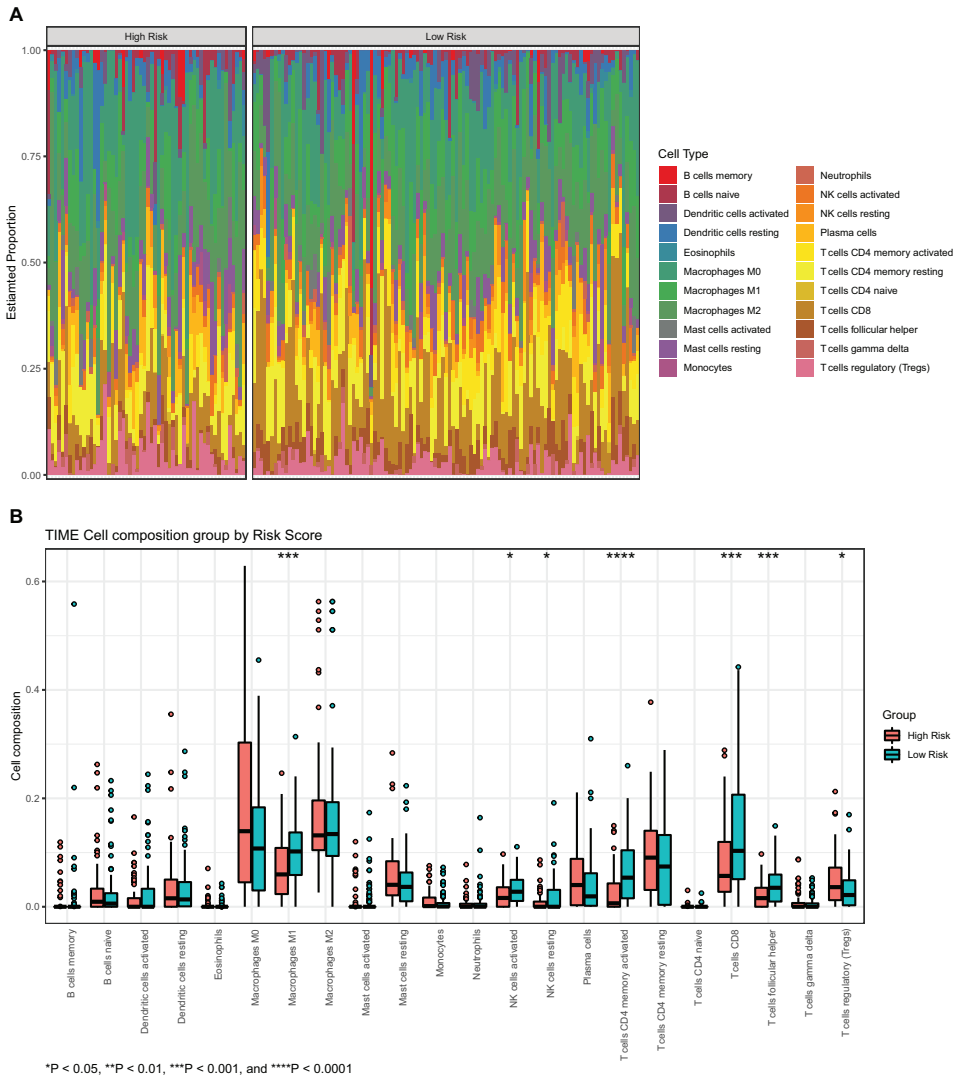


FIGURE 8 | (A) The proportion of infiltrating immune cells, as estimated by the CIBERSORT algorithm. **(B)** A box plot of immune cells in the low- and high-risk groups.

circulating tumor cells (Fauriat et al., 2010; Gooden et al., 2011; Lopez-Soto et al., 2017; Malmberg et al., 2017). In pancreatic cancer, memory CD4$^+$ T cells are closely related to gemcitabine resistance (Gu et al., 2020). A similar association may exist for BLCA, although further validation is still needed. Patients with a high proportion of CD8$^+$ T cells are more likely to show a favorable response to neoadjuvant chemotherapy (Green et al., 2017). In triple-negative breast cancer, tumors with high levels of infiltrating CD8$^+$ T cells and memory CD4$^+$ T cells might result in a better prognosis (Matsumoto et al., 2016; Oshi et al., 2020); this is consistent with BLCA. A high abundance of CD8$^+$ T cells is closely related to high expression levels of multiple immune checkpoint molecules, thus implying that treatment involving immune checkpoint inhibitors may be effective (Oshi et al., 2020). Tfh cells are an independent subset of CD4$^+$ T cells derived from naïve T cells that localize to lymphoid follicles and mediate the selection, proliferation, and survival of B cells to generate antibody signals (Eivazi et al., 2016). Tregs constitutively express CTLA-4 and are able to suppress the activation of leukocytes and maintain immune homeostasis (Lu et al., 2017). Patients with high levels of Tregs infiltration may respond effectively to therapy involving ipilimumab and tremelimumab (Simeone et al., 2014; He et al., 2017).

In summary, we successfully identified signature genes associated with invasiveness and used these genes to establish a reliable prognostic model for BLCA. These gene signatures represent potential biomarkers and targets for prognosis and treatment. Risk score acted as an independent prognostic factor and could guide the selection of therapy involving immune checkpoint inhibitors. However, there were some limitations to this study that need to be considered. First, further laboratory experiments are still required to validate the potential mechanisms underlying the action of the signature genes and TIME. Second, our sample size was not sufficient; therefore, we were unable to detect additional risk signatures associated with invasiveness in patients with BLCA.

CONCLUSION

In conclusion, we identified three signature genes associated with the invasiveness of BLCA; two of these showed strong associations with prognosis. We also constructed a prognostic risk model that featured the three signature genes and other clinical features; this model showed acceptable levels of performance. Differences in TIME between the patient groups showing different risk scores were also analyzed to guide the selection of therapeutic approaches and to help predict prognosis.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/ and https://portal.gdc.cancer.gov/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YH and AX designed this study. YH obtained and analyzed the data. YH, AX, YW, ZL, BL, NJ, and PX helped to review relevant literature and discuss the results. YH wrote the manuscript. All authors approved the final submitted manuscript.

FUNDING

This work was supported by grants from the Mobility Programme of National Natural Science Foundation of China (Grant No. M-0299), the National Natural Science Foundation of China (Grant No. 81902559), and Clinical Research Startup Program of Southern Medical University by High-level University Construction Funding of Guangdong Provincial Department of Education (Grant No. LC2016PY030).

ACKNOWLEDGMENTS

We would like to thank the researchers who submit data to the GEO and TCGA databases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.694777/full#supplementary-material

Supplementary Figure 1 | The relationships between modules and the invasiveness of BLCA.

Supplementary Table 1 | All genes from the midnight-blue module and their expression levels in each sample.

Supplementary Table 2 | The identified hub genes and their MM in each module.

Supplementary Table 3 | Clinical information of BLCA samples in TCGA.

REFERENCES

Avgeris, M., Tsilimantou, A., Levis, P. K., Tokas, T., Sideris, D. C., Stravodimos, K., et al. (2018). Loss of GAS5 tumour suppressor lncRNA: an independent molecular cancer biomarker for short-term relapse and progression in bladder cancer patients. *Br. J. Cancer* 119, 1477–1486. doi: 10.1038/s41416-018-0320-6

Chen, F., Wang, Q., and Zhou, Y. (2021). The construction and validation of an RNA binding protein-related prognostic model for bladder cancer. *BMC Cancer* 21:244. doi: 10.1186/s12885-021-07930-5

Eivazi, S., Bagheri, S., Hashemzadeh, M. S., Ghalavand, M., Qamsari, E. S., Dorostkar, R., et al. (2016). Development of T follicular helper cells and their role in disease and immune system. *Biomed. Pharmacother.* 84, 1668–1678. doi: 10.1016/j.biopha.2016.10.083

Fauriat, C., Long, E. O., Ljunggren, H. G., and Bryceson, Y. T. (2010). Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood* 115, 2167–2176. doi: 10.1182/blood-2009-08-238469

Gooden, M., Lampen, M., Jordanova, E. S., Leffers, N., Trimbos, J. B., Van Der Burg, S. H., et al. (2011). HLA-E expression by gynecological cancers restrains tumor-infiltrating CD8(+) T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10656–10661. doi: 10.1073/pnas.1100354108

Green, A. R., Aleskandarany, M. A., Ali, R., Hodgson, E. G., Atabani, S., De Souza, K., et al. (2017). Clinical Impact of Tumor DNA Repair Expression and T-cell Infiltration in Breast Cancers. *Cancer Immunol. Res.* 5, 292–299. doi: 10.1158/2326-6066.CIR-16-0195

Gu, J., Zhang, J., Huang, W., Tao, T., Huang, Y., Yang, L., et al. (2020). Activating miRNA-mRNA network in gemcitabine-resistant pancreatic cancer cell associates with alteration of memory CD4(+) T cells. *Ann. Transl. Med.* 8:279. doi: 10.21037/atm.2020.03.53

He, M., Chai, Y., Qi, J., Zhang, C. W. H., Tong, Z., Shi, Y., et al. (2017). Remarkably similar CTLA-4 binding properties of therapeutic ipilimumab and tremelimumab antibodies. *Oncotarget* 8, 67129–67139. doi: 10.18632/oncotarget.18004

Henderson, M. C., and Azorsa, D. O. (2012). The genomic and proteomic content of cancer cell-derived exosomes. *Front. Oncol.* 2:38. doi: 10.3389/fonc.2012.00038

Hibino, S., Kanda, M., Oya, H., Takami, H., Shimizu, D., Nomoto, S., et al. (2014). Reduced expression of DENND2D through promoter hypermethylation is an adverse prognostic factor in squamous cell carcinoma of the esophagus. *Oncol. Rep.* 31, 693–700. doi: 10.3892/or.2013.2901

Iyama, K., Matsuse, M., Mitsutake, N., Rogounovitch, T., Saenko, V., Suzuki, K., et al. (2017). Identification of Three Novel Fusion Oncogenes, SQSTM1/NTRK3, AFAP1L2/RET, and PPFIBP2/RET, in Thyroid Cancers of Young Patients in Fukushima. *Thyroid* 27, 811–818. doi: 10.1089/thy.2016.0673

Kanda, M., Nomoto, S., Oya, H., Takami, H., Hibino, S., Hishida, M., et al. (2014). Downregulation of DENND2D by promoter hypermethylation is associated with early recurrence of hepatocellular carcinoma. *Int. J. Oncol.* 44, 44–52. doi: 10.3892/ijo.2013.2165

Kanda, M., Shimizu, D., Nomoto, S., Takami, H., Hibino, S., Oya, H., et al. (2015). Prognostic impact of expression and methylation status of DENN/MADD domain-containing protein 2D in gastric cancer. *Gastric Cancer* 18, 288–296. doi: 10.1007/s10120-014-0372-0

Kariuki, S. N., Franek, B. S., Kumar, A. A., Arrington, J., Mikolaitis, R. A., Utset, T. O., et al. (2010). Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. *Arthritis Res. Ther.* 12:R151. doi: 10.1186/ar3101

Kotolloshi, R., Hölzer, M., Gajda, M., Grimm, M.-O., and Steinbach, D. (2021). SLC35F2, a Transporter Sporadically Mutated in the Untranslated Region, Promotes Growth, Migration, and Invasion of Bladder Cancer Cells. *Cells* 10:80. doi: 10.3390/cells10010080

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, J. S., Leem, S. H., Lee, S. Y., Kim, S. C., Park, E. S., Kim, S. B., et al. (2010). Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors. *J. Clin. Oncol.* 28, 2660–2667. doi: 10.1200/JCO.2009.25.0977

Ling, B., Zheng, H., Fu, G., Yuan, J., Shi, T., Chen, S., et al. (2013). Suppression of non-small cell lung cancer proliferation and tumorigenicity by DENND2D. *Lung Cancer* 79, 104–110. doi: 10.1016/j.lungcan.2012.10.012

Lopez-Soto, A., Gonzalez, S., Smyth, M. J., and Galluzzi, L. (2017). Control of Metastasis by NK Cells. *Cancer Cell* 32, 135–154. doi: 10.1016/j.ccell.2017.06.009

Lu, L., Barbi, J., and Pan, F. (2017). The regulation of immune tolerance by FOXP3. *Nat. Rev. Immunol.* 17, 703–717. doi: 10.1038/nri.2017.75

Lyu, L., Xiang, W., Zhu, J. Y., Huang, T., Yuan, J. D., and Zhang, C. H. (2019). Integrative analysis of the lncRNA-associated ceRNA network reveals lncRNAs as potential prognostic biomarkers in human muscle-invasive bladder cancer. *Cancer Manag. Res.* 11, 6061–6077. doi: 10.2147/CMAR.S207336

Lyu, Q., Lin, A., Cao, M., Xu, A., Luo, P., and Zhang, J. (2020). Alterations in TP53 Are a Potential Biomarker of Bladder Cancer Patients Who Benefit From Immune Checkpoint Inhibition. *Cancer Control* 27:1073274820976665. doi: 10.1177/1073274820976665

Malmberg, K. J., Carlsten, M., Bjorklund, A., Sohlberg, E., Bryceson, Y. T., and Ljunggren, H. G. (2017). Natural killer cell-mediated immunosurveillance of human cancer. *Semin. Immunol.* 31, 20–29. doi: 10.1016/j.smim.2017.08.002

Matsumoto, H., Thike, A. A., Li, H., Yeong, J., Koo, S. L., Dent, R. A., et al. (2016). Increased CD4 and CD8-positive T cell infiltrate signifies good prognosis in a subset of triple-negative breast cancer. *Breast Cancer Res. Treat.* 156, 237–247. doi: 10.1007/s10549-016-3743-x

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

Nouhaud, F. X., Chakroun, M., Lenormand, C., Ouzaid, I., Peyronnet, B., Gryn, A., et al. (2021). Comparison of the prognosis of primary vs. progressive muscle invasive bladder cancer after radical cystectomy: results from a large multicenter study. *Urol. Oncol.* 39, e191–e195. doi: 10.1016/j.urolonc.2020.09.006

Oshi, M., Asaoka, M., Tokumaru, Y., Yan, L., Matsuyama, R., Ishikawa, T., et al. (2020). CD8 T Cell Score as a Prognostic Biomarker for Triple Negative Breast Cancer. *Int. J. Mol. Sci.* 21:6968. doi: 10.3390/ijms21186968

Ostrowski, M., Carmo, N. B., Krumeich, S., Fanget, I., Raposo, G., Savina, A., et al. (2010). Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat. Cell Biol.* 12, 19–30. doi: 10.1038/ncb2000

Ponnambalam, S., and Baldwin, S. A. (2003). Constitutive protein secretion from the trans-Golgi network to the plasma membrane. *Mol. Membr. Biol.* 20, 129–139. doi: 10.1080/0968768031000084172

Qiu, H., Hu, X., He, C., Yu, B., Li, Y., and Li, J. (2020). 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

Reddy, K. B., Nabha, S. M., and Atanaskova, N. (2003). Role of MAP kinase in tumor progression and invasion. *Cancer Metastasis Rev.* 22, 395–403. doi: 10.1023/A:1023781114568

Rink, J., Ghigo, E., Kalaidzidis, Y., and Zerial, M. (2005). Rab conversion as a mechanism of progression from early to late endosomes. *Cell* 122, 735–749. doi: 10.1016/j.cell.2005.06.043

Sakha, S., Muramatsu, T., Ueda, K., and Inazawa, J. (2016). Exosomal microRNA miR-1246 induces cell motility and invasion through the regulation of DENND2D in oral squamous cell carcinoma. *Sci. Rep.* 6:38750. doi: 10.1038/srep38750

Shen, C., Xu, T., Sun, Y., Wang, L., Liang, Z., Niu, H., et al. (2020). Construction of an Immune-Associated Gene-Based Signature in Muscle-Invasive Bladder Cancer. *Dis. Markers* 2020, 1–13. doi: 10.1155/2020/8866730

Shimasaki, N., Jain, A., and Campana, D. (2020). NK cells for cancer immunotherapy. *Nat. Rev. Drug Discov.* 19, 200–218. doi: 10.1038/s41573-019-0052-1

Simeone, E., Gentilcore, G., Giannarelli, D., Grimaldi, A. M., Caraco, C., Curvietto, M., et al. (2014). Immunological and biological changes during ipilimumab treatment and their potential correlation with clinical response and survival in patients with advanced melanoma. *Cancer Immunol. Immunother.* 63, 675–683. doi: 10.1007/s00262-014-1545-8

Song, B. N., Kim, S. K., Mun, J. Y., Choi, Y. D., Leem, S. H., and Chu, I. S. (2019). Identification of an immunotherapy-responsive molecular subtype of bladder cancer. *EBioMedicine* 50, 238–245. doi: 10.1016/j.ebiom.2019.10.058

Sullivan, L. M., Massaro, J. M., and D'agostino, R. B., Sr. (2004). Presentation of multivariate data for clinical use: the framingham study risk score functions. *Stat. Med.* 23, 1631–1660. doi: 10.1002/sim.1742

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

Tan, K. D., Zhu, Y., Tan, H. K., Rajasegaran, V., Aggarwal, A., Wu, J., et al. (2008). Amplification and overexpression of PPFIA1, a putative 11q13 invasion suppressor gene, in head and neck squamous cell carcinoma. *Genes Chromosomes Cancer* 47, 353–362. doi: 10.1002/gcc.20539

Tan, M., Zhang, D., Zhang, E., Xu, D., Liu, Z., Qiu, J., et al. (2017). SENP2 suppresses epithelial-mesenchymal transition of bladder cancer cells through deSUMOylation of TGF-betaRI. *Mol. Carcinog.* 56, 2332–2341. doi: 10.1002/mc.22687

Ueoka, Y., Kato, K., Kuriaki, Y., Horiuchi, S., Terao, Y., Nishida, J., et al. (2000). Hepatocyte growth factor modulates motility and invasiveness of ovarian carcinomas via Ras-mediated pathway. *Br. J. Cancer* 82, 891–899. doi: 10.1054/bjoc.1999.1016

von Thun, A., Birtwistle, M., Kalna, G., Grindlay, J., Strachan, D., Kolch, W., et al. (2012). ERK2 drives tumour cell migration in three-dimensional microenvironments by suppressing expression of Rab17 and liprin-beta2. *J. Cell Sci.* 125, 1465–1477. doi: 10.1242/jcs.092916

Wu, Y., Yu, H., Zheng, S. L., Feng, B., Kapron, A. L., Na, R., et al. (2018). Germline mutations in PPFIBP2 are associated with lethal prostate cancer. *Prostate* 78, 1222–1228. doi: 10.1002/pros.23697

Xu, Z., Gujar, H., Fu, G., Ahmadi, H., Bhanvadia, S., Weisenberger, D. J., et al. (2021). A Novel DNA Methylation Signature as an Independent Prognostic Factor in Muscle-Invasive Bladder Cancer. *Front. Oncol.* 11:614927. doi: 10.3389/fonc.2021.614927

Yan, X., Fu, X., Guo, Z. X., Liu, X. P., Liu, T. Z., and Li, S. (2020). Construction and validation of an eight-gene signature with great prognostic value in bladder cancer. *J. Cancer* 11, 1768–1779. doi: 10.7150/jca.38741

Zhang, Q., Lai, Q., Wang, S., Meng, Q., and Mo, Z. (2021). Clinical Value of Postoperative Neutrophil-to-Lymphocyte Ratio Change as a Detection Marker of Bladder Cancer Recurrence. *Cancer Manag. Res.* 13, 849–860. doi: 10.2147/CMAR.S289986

Zhang, T., Hu, Y., Ju, J., Hou, L., Li, Z., Xiao, D., et al. (2016). Downregulation of miR-522 suppresses proliferation and metastasis of non-small cell lung cancer cells by directly targeting DENN/MADD domain containing 2D. *Sci. Rep.* 6:19346. doi: 10.1038/srep19346

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fgene.2021.759953

TUBA1C is a Prognostic Marker in Low-grade Glioma and Correlates with Immune Cell Infiltration in the Tumor Microenvironment

Hua Zhu[1,2†], Xinyao Hu[3†], Lijuan Gu[4], Zhihong Jian[2], Liqin Li[1], Siping Hu[5], Sheng Qiu[1] and Xiaoxing Xiong[1,2*]*

[1]Department of Neurosurgery, The Affiliated Huzhou Hospital, Zhejiang University School of Medicine (Huzhou Central Hospital), Huzhou, China, [2]Department of Neurosurgery, Renmin Hospital of Wuhan University, Wuhan, China, [3]Cancer Center, Renmin Hospital of Wuhan University, Wuhan, China, [4]Central Laboratory, Renmin Hospital of Wuhan University, Wuhan, China, [5]Department of Anesthesiology, The Affiliated Huzhou Hospital, Zhejiang University School of Medicine (Huzhou Central Hospital), Huzhou, China

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Yu Xin,
Capital Medical University, China
Liang Wang,
Tangdu Hospital, China
Xuejun Li,
Central South University, China

***Correspondence:**
Sheng Qiu
qius2001@126.com
Xiaoxing Xiong
xiaoxingxiong@whu.edu.cn

[†]These authors have contributed equally to this work

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

Received: 17 August 2021
Accepted: 22 September 2021
Published: 14 October 2021

Citation:
Zhu H, Hu X, Gu L, Jian Z, Li L, Hu S, Qiu S and Xiong X (2021) TUBA1C is a Prognostic Marker in Low-grade Glioma and Correlates with Immune Cell Infiltration in the Tumor Microenvironment. Front. Genet. 12:759953. doi: 10.3389/fgene.2021.759953

TUBA1C, a microtubule component, contributes to the development of several cancers. Our purpose was to study the expression of TUBA1C, its potential prognostic value, and its effects on the infiltration of immune cells of low-grade glioma (LGG). Through applying multiple bioinformatics analyses, we extracted and analyzed datasets from TCGA, TIMER, GTEx, GEPIA, and HPA to investigate the potential oncogenic mechanisms of TUBA1C, including the correlation between TUBA1C and prognosis, immune-checkpoints, tumor microenvironment (TME), and infiltration of immune cells in LGG. GO functional annotations and KEGG pathway analyses were further applied to investigate the potential action of TUBA1C in LGG. We revealed that the mRNA levels of TUBA1C were increased in LGG tumor tissues than in normal tissues. Additionally, TUBA1C was up-regulated in the grade III of LGG than in grade II. Moreover, we found that TUBA1C may be an independent prognostic factor of LGG, and high TUBA1C expression correlated to a poor prognosis of LGG. TUBA1C expression was positively associated with the infiltration of B cells, CD8 T+ cells, $CD4^+$ T cells, macrophages, dendritic cells, and neutrophils. TUBA1C was also verified to be co-expressed with immune-related genes and immune-checkpoints. GO and KEGG pathway analyses indicated that TUBA1C may potentially regulate the pathogenesis of LGG through immune-related pathways, including chemokine pathway; JAK-STAT pathway; natural killer cell mediated cytotoxicity; T cell receptor pathway; leukocyte migration; negative regulation of immune system process; regulation of lymphocyte activation; T cell activation and other pathways. In conclusion, TUBA1C expression is increased in LGG and high TUAB1C expression is related to a poor prognosis. TUBA1C may influence tumor development by regulating the tumor-infiltrating cells in the TME. TUBA1C may be a potential target for immunotherapy.

Keywords: LGG, TUBA1C, prognosis, tumor microenvironment, tumor immune cell infiltration

INTRODUCTION

Glioma is a common and fatal tumor type in the central nervous system. Lower-grade glioma (LGG, WHO II, III) grow slowly with less malignancy and have superior survival outcomes compared with glioblastoma (GBM, WHO IV). Bioinformatics studies of gliomas have an important role in improving the accuracy of diagnosis and treatment, for example, thanks to bioinformatics, WHO have added molecular biomarkers such as isocitrate dehydrogenase mutation status to the diagnostic guidelines of glioma (Louis et al., 2016). Currently, although surgical resection of tumors combined with chemotherapy, radiotherapy, and neurorestorative therapy somewhat improves the prognosis of the patient, over 50% of LGG patients eventually develop into highly aggressive gliomas (Deng et al., 2019; Wang et al., 2020). Hence, new prognostic factors need to be identified for LGG.

Microtubules, which are assembled from highly conserved α/β-tubulin heterodimers (Bodakuntla et al., 2019), are one of the components of the backbone of eukaryotic cells, and exert a vital action in depolymerization and dynamic aggregation *via* cell division and replication (Kim et al., 2010). Previous studies have shown that α-tubulin participates in the development of several cancers, such as lung, prostate, and breast cancers (Boggs et al., 2015; Li et al., 2019; Zhang et al., 2016). Additionally, α-tubulin is involved in the occurrence of astrocytoma and chemoresistance in hepatocellular carcinoma (HCC) which greatly affects the prognosis of patients with liver cancer (Tsai et al., 2014; Hu et al., 2021). Notably, overexpression of TUBA1C, which is one of the α-tubulin subtypes is involved in the poor prognosis of HCC (Li et al., 2017), pancreatic ductal adenocarcinoma (Albahde et al., 2020), and lung cancer (Bian et al., 2021). However, whether TUBA1C can affect the prognosis of LGG has not been explored.

In recent years, immunotherapies have been applied for the treatment of glioma patients and have changed the treatment paradigm for glioma (Lim et al., 2018). Tumor-infiltrating immune cells (TIICs) are known to affect the immune system, processing abnormal biological behavior in a complex manner and exerting an essential action in response to immunotherapies (Xiong et al., 2018). In addition, genes associated with immune components of the TME are of great value as prognostic biomarkers (Le Rhun et al., 2019). Recently, several studies have proposed that TUBA1C is related to poor prognosis in HCC (Wang et al., 2017), pancreatic ductal adenocarcinoma (Albahde et al., 2020), and lung adenocarcinoma (Bian et al., 2021). However, no studies have yet explored whether TUBA1C overexpression affects the tumor immune microenvironment of LGG.

Through applying multiple bioinformatics analyses, we extracted and analyzed datasets from The Cancer Genome Atlas (TCGA), genotype-tissue expression (GTEx), Tumor Immune Estimation Resource (TIMER), Gene Expression Profiling Interactive Analysis (GEPIA), and the Human Protein Atlas (HPA) to explore the potential oncogenic mechanisms of TUBA1C, including correlations between TUBA1C and the prognosis, immune-checkpoints, TME and TIICs in LGG. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), pathway enrichment analyses were applied to investigate the potential functions of TUBA1C.

METHODS

Data Source and Analysis of Differential Expressions

We downloaded cancer-related RNA sequences, clinicopathological, and survival data of LGG and GBM on the UCSC Xena web (https://xena.ucsc.edu/, originally from the TCGA). GTEx (http://commonfund.nih.gov/GTEx/) contains publicly available gene expression data from the RNA sequencing of 54 normal tissue sites from about 1,000 individuals (GTEx Consortium, 2013). Normal samples from the GTEx database and tumor samples from TCGA were applied to compare the differential expressions of TUBA1C between cancer and normal tissue. The TUBA1C expression data in TCGA (https://tcga.xenahubs.net) were extracted by using Perl software for subsequent analysis. The "wilcox.test" method was used to evaluate the differential mRNA expressions of TUBA1C in LGG. The cut-off was established as a False Discovery Rate (FDR) value <0.05. The "ggpubr" R package was used for the boxplot.

Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/index .html) is a publicly available database developed by Peking University, China (Tang et al., 2017). To verify the differential TUBA1C expressions in LGG and normal tissues, we further used GEPIA to verify our results and constructed boxplot, and survival curves, such as overall survival (OS) curve and disease-free survival (DFS) curve.

Immunohistochemistry (IHC) Staining

Immunohistochemical images of TUBA1C protein expression analyses, assessment of the differences in TUBA1C expression at the protein level, were performed in normal, LGG, and GBM tissues from the HPA (http://www.proteinatlas.org/). The antibody HPA043684 was used for IHC.

Identification of the Prognostic Factors for OS in LGG

Univariate and multivariate Cox regression analyses were used to assess TUBA1C and five major clinical and prognostic factors, including age, race, sex, grade, radiotherapy, to identify the proper terms to build the nomogram. The forest plots showed the HR, 95% CI, and p-value of each variable by using the "forestplot" R package. Using R software (version 4.0.2) with the packages "rms," a nomogram was formulated according to the results of multivariate Cox proportional hazards analysis to predict the 1-yr, 3-yr, and 5-yr overall recurrence.

Correlations Between TUBA1C and Clinicopathology or Survival in LGG

We extracted the survival information for each sample in the TCGA. We then selected several indicators: OS, disease-specific survival (DSS), DFS, and progression-free interval (PFS), to clarify the association of TUBA1C expression with the prognosis of LGG patients. The Kaplan-Meier (KM) and log-

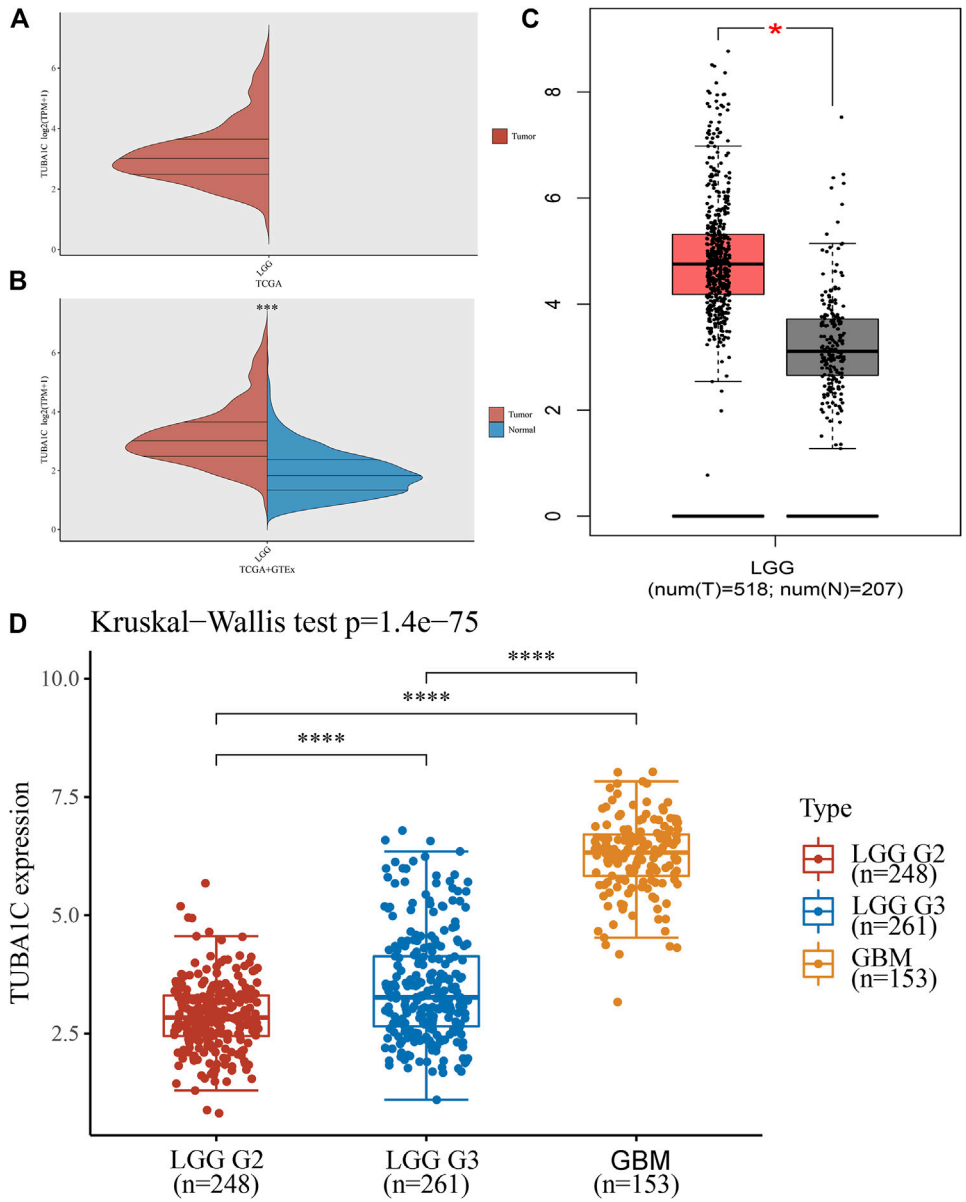


FIGURE 1 | The different mRNA expressions of TUBA1C in LGG, GBM, and normal tissues. **(A,B)** mRNA expressions of TUBA1C in LGG and normal tissues. **(C)** mRNA expression of TUBA1C in LGG and normal tissues in the GEPIA. **(D)** Differential TUBA1C expressions in tumor stages in LGG and GBM. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$.

rank test were applied for survival analysis of LGG ($p < 0.05$) and survival curves were performed through R packages "survminer" and "survival." Then, R packages "survival" was used for Cox analysis to identify the correlation of TUBA1C with survival. The R packages "ggpubr" and "limma" were applied for clinicopathological correlation analyses.

Relationship Between TUBA1C and Immune Cells Infiltration in the TME of LGG

The TIMER (https://cistrome.shinyapps.io/timer/) database was applied to explore the correlations between LGG and TIICs. The TIICs in TIMER contained dendritic cells, B cells, neutrophils, $CD4^+$ T cells, macrophages, and $CD8^+$ T cells. Subsequently, we applied the ESTIMATE algorithm in the R package "estimate" and "limma" to calculate immune and stromal scores. We analyzed tumor purity and the infiltration of stromal/immune cells in the tissue of LGG based on TUBA1C expression data using CIBERSORT, which was developed to estimate the abundance of particular cells in hybrid cell populations applying gene expression datasets. We next analyzed the correlation of TUBA1C with TME or infiltration of immune cells by using R packages "ggplot2," "ggpubr," and "ggExtra" (cut-off: value of $p < 0.01$).

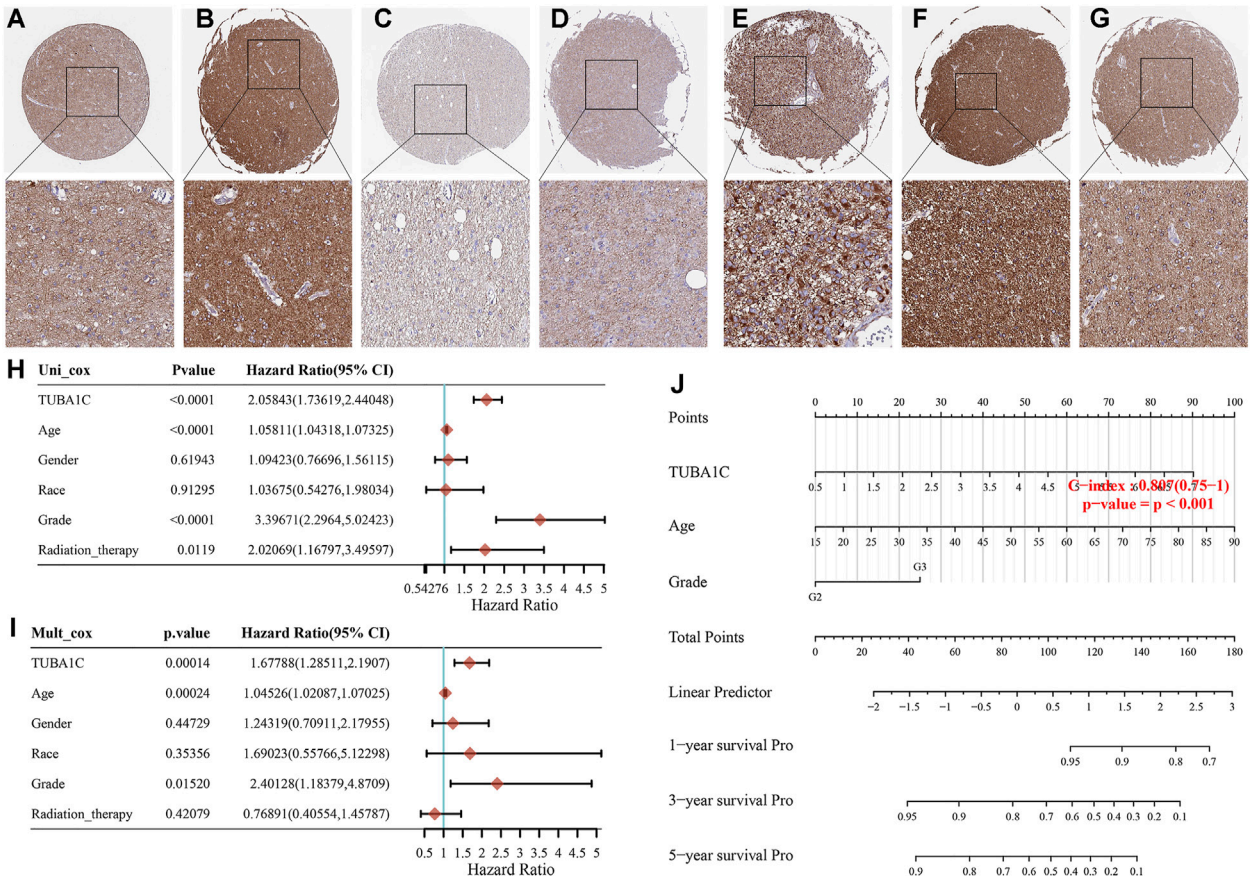


FIGURE 2 | The TUBA1C expression detected by immunohistochemistry and the construction of prognostic signature in LGG. Representative immumohistochemical stainings in various normal **(A)**, LGG **(B–D)** and GBM **(E–G)** tissues. Hazard ratio and *p*-value of constituents involved in the univariable **(H)** and multivariate **(I)** Cox regression. **(J)** Nomogram consisting of risk score and other clinical indicators to predict the 1-yr, 3-yr and 5-yr OS of the patients with LGG.

Immune-checkpoint Analysis

SIGLEC15, TIGIT, CTLA4, CD274, HAVCR2, LAG3, PDCD1, and PDCD1LG2 were selected to be immune-checkpoint–relevant transcripts, and the expression data of these eight genes were extracted. R package "ggplot2," "pheatmap" and "immuneeconv" were used to assess the expression of the immune-checkpoints and co-expression of TUBA1C with these immune-checkpoints. Potential immune checkpoint blockade response was predicted with the TIDE algorithm (Jiang et al., 2018).

Co-Expression Analyses and Pathways Enrichment Analyses of TUBA1C in LGG

R packages "limma," "reshape2" and "RColorBrewer" were applied to perform the co-expression analyses. GO and KEGG gene sets were from the Gene Set Enrichment Analysis website (https://www.gsea-msigdb.org/gsea/downloads.jsp). The GO functional annotations and the KEGG enrichment pathways analyses of TUBA1C were conducted using the R package "enrichplot2," "clusterProfiler," "limma," and "org.Hs.eg.db."

Statistical Analysis

Gene expression data were normalized by log2-transformation. Normal and tumor tissues were compared using a two-group *t*-test. Kruskal-Wallis one-way ANOVA was used to make comparisons between groups greater than or equal to three. The Cox proportional hazards model, KM analyses, and log-rank test were conducted for all survival analyses. Spearman's test or Pearson's test was applied to analyze the correlation between two variables; p value <0.05 were considered significant. R software (version 4.0.2) was used for statistical analysis.

RESULTS

The Different Expressions of TUBA1C in LGG, GBM, and Normal Tissues

To evaluate the TUBA1C expression in LGG, GBM, and normal tissues, RNA sequencing data obtained from the TCGA were analyzed using R software. The TCGA profiles (including 509 LGG and 153 GBM samples) of mRNA expression were gained. Since the TCGA database lacked paracancerous tissue samples of

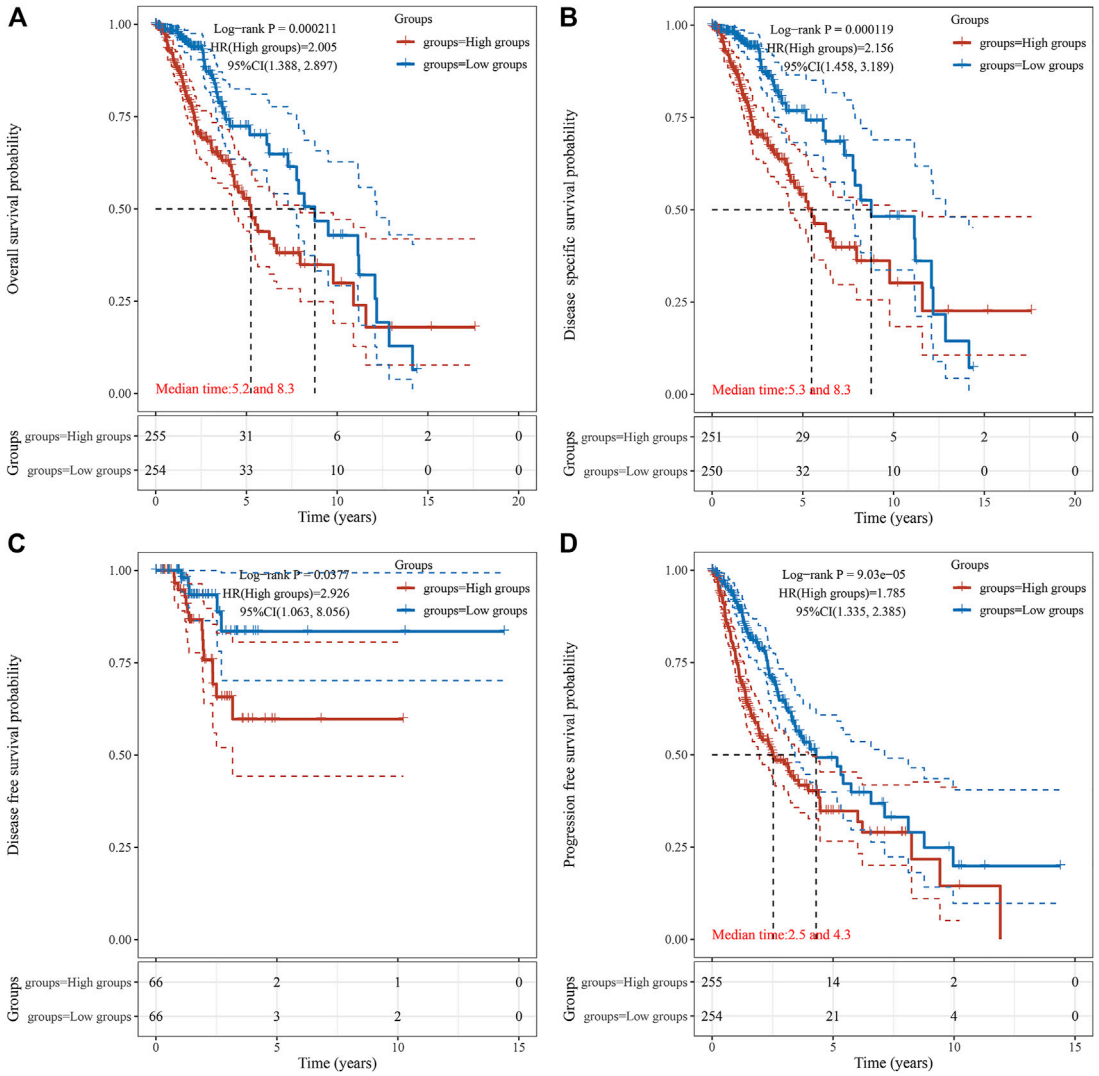


FIGURE 3 | The correlation between TUBA1C expression and OS **(A)**, DSS **(B)**, DFS **(C)**, and PFS **(D)**. The correlation of TUBA1C expression with TME and TIICs.

LGG (**Figure 1A**), we downloaded 2642 normal samples from the GTEx database. We found that TUBA1C was overexpressed in LGG tumor tissues compared to the normal tissues ($p < 0.001$) (**Figure 1B**). We further verified the high TUBA1C expression in LGG tumor tissues in the GEPIA, the result indeed showed that TUBA1C was over-expressed in LGG than normal tissues (**Figure 1C**), which was consistent with our findings. Additionally, we evaluated the correlation between TUBA1C expression and tumor stage and found that the expression of TUBA1C was increased in stage III than stage II of LGG ($p < 0.0001$). Moreover, the expression of TUBA1C was elevated in GBM (WHO, stage IV) than stage III of LGG ($p < 0.0001$) (**Figure 1D**).

To assess the expression of TUBA1C at the protein level, the IHC results were obtained and analyzed from the HPA. The results illustrated that the TUBA1C IHC staining was weak in the normal cerebral cortex (**Figure 2A**), whereas LGG tumor tissues had not detected the TUBA1C IHC staining or had high TUBA1C IHC staining (**Figures 2B–D**). However, the GBM tumor tissues had strong TUBA1C IHC staining (**Figures 2E–G**).

Prognostic Value of TUBA1C in LGG

We further applied the univariable and multivariable Cox regression model to investigate the independent prognostic force of the signature. Univariable analysis results indicated that TUBA1C (HR = 2.05843, $p < 0.0001$), grade (HR = 3.39671, $p < 0.0001$), age (HR = 1.05811, $p < 0.0001$), and radiation therapy (HR = 2.02069, $p = 0.0119$) had prognostic value for the OS of LGG (**Figure 2H**). Similarly, in the subsequent multivariable stepwise cox regression analysis, TUBA1C (HR = 1.67788, $p = 0.00014$), age (HR = 1.04526, $p = 0.00024$), grade (HR = 2.40128, $p = 0.0152$) still maintained their prognostic values (**Figure 2I**). These results demonstrated that TUBA1C expression was an independent prognostic factor in LGG.

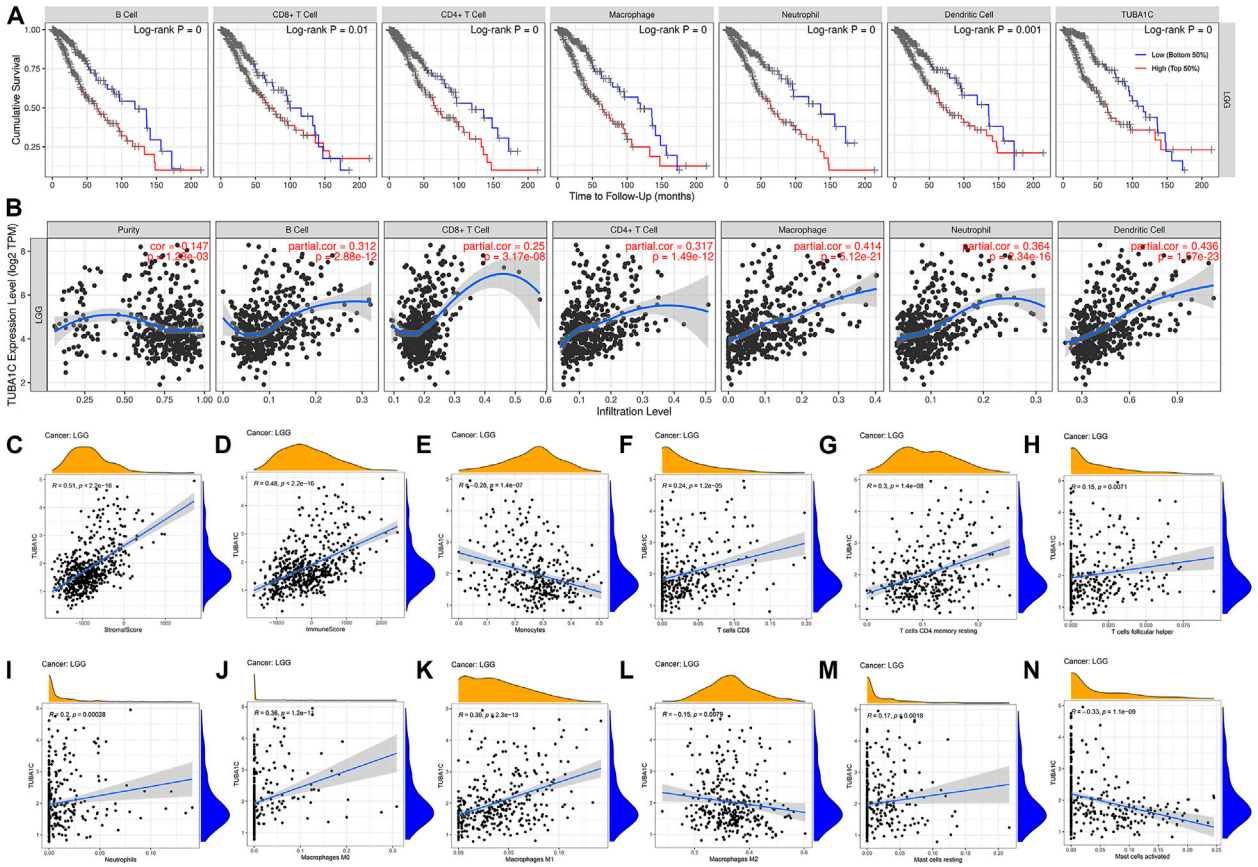


FIGURE 4 | The correlations between TUBA1C expression and immune cell infiltration in the tumor microenvironment. **(A)** Correlation of infiltrating levels of immune cells with the cumulative survival of LGG. **(B,E–N)** Relationships between TUBA1C expression levels and immune cell infiltration in LGG. **(C)** Correlation of TUBA1C expression and stromal score. **(D)** Correlation of TUBA1C expression and immune score. TUBA1C co-expressed with immune-related genes and immune-checkpoints.

Thereafter, TUBA1C, age, and grade were visualized in the nomogram. Nomograms of 1-yr, 3-yr, and 5-yr OS in the cohort are shown in **Figure 2J**.

Survival association analyses, including OS, DSS, DFS, and PFS were applied to further verify the correlation of TUBA1C expression with prognosis in LGG. KM plotter results showed that among the individuals with LGG (p = 0.0002) (**Figure 3A**), those with high TUBA1C expression had less time of survival. Furthermore, DSS analyses indicated a correlation between high TUBA1C expression and adverse outcomes in the LGG patients (p = 0.0001) (**Figure 3B**). Notably, LGG patients with high TUBA1C expression had poor DFS (p = 0.0377) (**Figure 3C**) and (p < 0.0001) PFS (**Figure 3D**). Furthermore, the OS and DFS curves were performed with the help of GEPIA. LGG patients with high TUBA1C expression had a poor OS (p = 0.00027) (**Supplementary Figure S1A**) and DFS (p = 0.00068) (**Supplementary Figure S1B**), which is consistent with our results in **Figure 3**.

To investigate the mechanism by which high expression of TUBA1C leads to a worse prognosis, we applied the TIMER to reveal the correlation of TUBA1C with infiltrating levels of six immune cell subtypes, including dendritic cells, neutrophils, B cells, CD8 T+ cells, macrophages, and $CD4^+$ T cells. As presented in **Figure 4A**, LGG patients with high infiltrating levels of these immune cells had a poor cumulative survival compared with low infiltration with these immune cells. Moreover, the TUBA1C expression was positively correlated with the infiltration of dendritic cells (Cor = 0.436, p = 1.57e−23), neutrophils (Cor = 0.364, p = 2.34e−16), B cells (Cor = 0.312, p = 2.88e−12), CD8 T+ cells (Cor = 0.25, p = 3.17e−08), macrophages (Cor = 0.414, p = 5.12e−21), and $CD4^+$ T cells (Cor = 0.317, p = 1.49e−12). These results illustrate that LGG patients with high TUBA1C expression had a poor cumulative survival, which is consistent with the results in **Figure 2**. Additionally, ESTIMATE algorithm was applied to investigate stromal and immune scores for LGG and analyze the correlations of TUBA1C expression levels with these two scores. We revealed that in LGG, TUBA1C expression was significantly positively related to the stromal (R = 0.51, p < 2.2e−16) (**Figure 4C**) and immune scores (R = 0.48, p < 2.2e−16) (**Figure 4D**). To further evaluate the TIICs in the TME of LGG, we investigated the correlation of TUBA1C with the infiltration of 22 immune cell subtypes, including regulatory T cells, gamma delta T cells, follicular helper T cells, $CD8^+$ T cells, $CD4^+$ naïve T cells, $CD4^+$ memory resting T cells, $CD4^+$ memory activated T cells, neutrophils, resting NK cells,

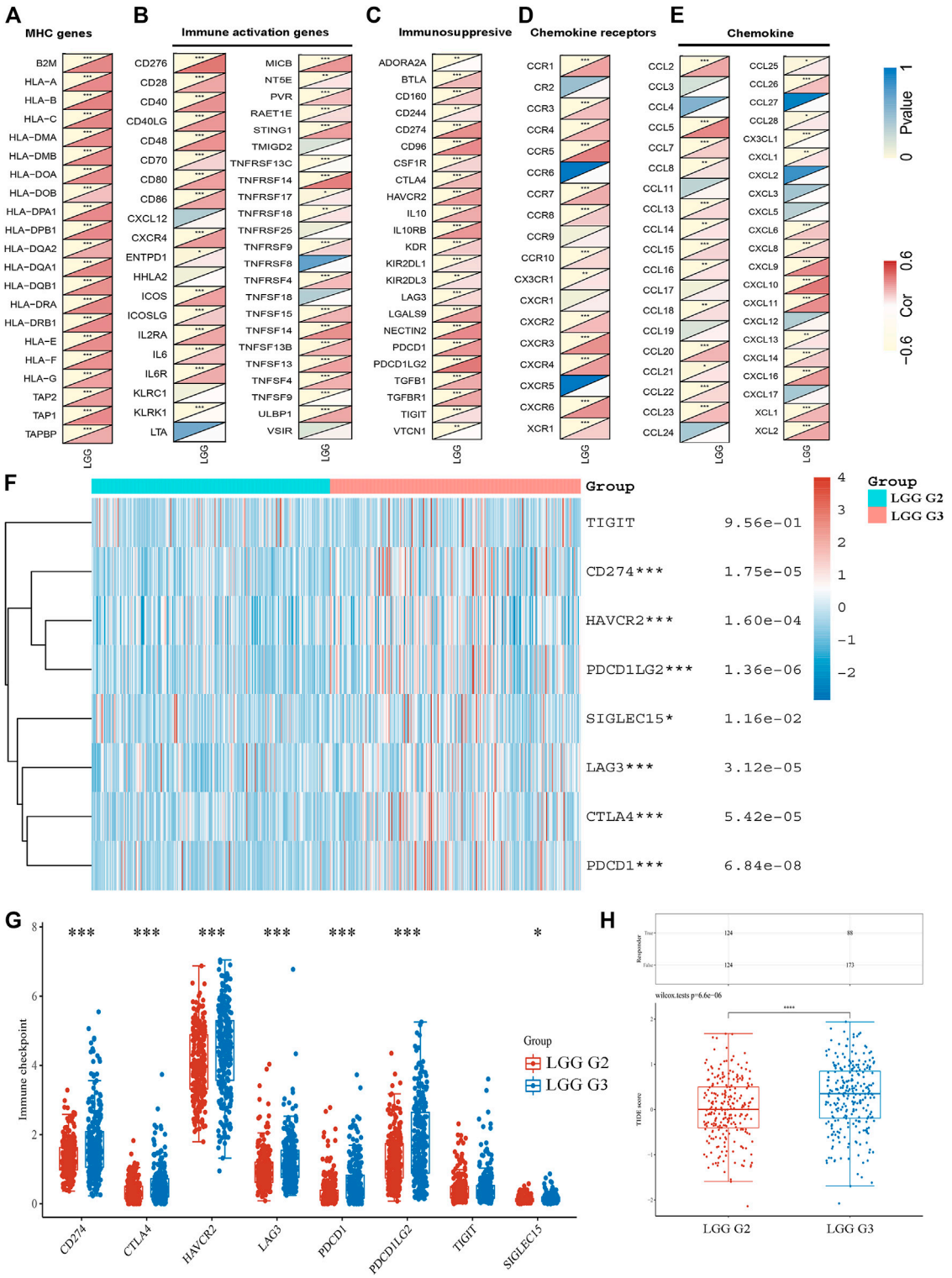


FIGURE 5 | The co-expression of TUBA1C with immune-related genes and the expression of immune-checkpoints in LGG. Co-expression of TUBA1C with MHC genes **(A)**, immune activation genes **(B)**, immunosuppresive genes **(C)**, chemokine receptors related genes **(D)**, and chemokine related genes **(E)**. **(F,G)** Different expressions of immune-checkpoints in grade II and III of LGG. **(H)** Different responses to immune checkpoint blockade in grade II and III of LGG.

TABLE 1 | Correlation of TUBA1C with immune-checkpoint in LGG.

Genes	Cor	*p*-value
CD274	0.395	[a]/2.20e−116
CTLA4	0.242	[a]/3.21e−08
HAVCR2	0.461	[a]/2.20e−116
LAG3	0.216	[a]/9.25e−07
PDCD1	0.530	[a]/2.74e−38
PDCD1LG2	0.498	[a]/2.20e−116
TIGIT	0.290	[a]/2.86e−11
SIGLEC15	0.229	[a]/1.71e−07

[a]p < *0.001*.

activated NK cells, resting myeloid dendritic cells, activated myeloid dendritic cells, monocyte, resting mast cells, activated mast cells, macrophage M0, M1, M2, eosinophil, plasma B cells, naïve B cells, memory B cells. The results further demonstrated that TUBA1C expression was positively related to the filtrating levels of $CD8^+$ T cells (R = 0.24, p = 1.2e−05) (**Figure 4F**) and neutrophils (R = 0.2, p = 0.00028) (**Figure 4I**), which was in line with the results from the TIMER. Moreover, we revealed that the TUBA1C expression was positively associated with the infiltration of resting CD4 memory T cells (R = 0.3, p = 1.4e−08) (**Figure 4G**), follicular helper T cells (R = 0.15, p = 0.0071) (**Figure 4H**), macrophages M0 (R = 0.36, p = 1.2e−11) (**Figure 4J**), macrophages M1 (R = 0.39, p = 2.3e−13) (**Figure 4K**), resting mast cells (R = 0.17, p = 0.0018) (**Figure 4M**), while they were negatively related to the infiltration of macrophages M2 (R = −0.15, p = 0.0079) (**Figure 4L**) and activated mast cells (R = −0.33, p = 1.1e−09) (**Figure 4N**).

We further performed gene co-expression analysis to evaluate the mechanisms that TUBA1C was correlated to the infiltration of immune cells in LGG. MHC genes, immune activation genes, immunosuppresive, and chemokine (receptors) related genes were assessed. The results showed that TUBA1C positively co-expressed with all listed MHC genes (**Figure 5A**). TUBA1C had a positive relationship with most of the immune activation genes, such as genes that encoding CD276, CD28, IL6, CXCR4 (**Figure 5B**). Notably, almost all immunosuppresive genes positively co-expressed with TUBA1C (**Figure 5C**). Similarly, TUBA1C had a positive correlation with some chemokine (receptors) (**Figures 5D,E**).

The expression of immune-checkpoints, including SIGLEC15, TIGIT, CTLA4, CD274, HAVCR2, LAG3, PDCD1, and PDCD1LG2, were further investigated in the WHO grade II and III of LGG. The results illustrated that CD274 (p = 1.75e−05), HAVCR2 (p = 1.60e−05), PDCD1 (p = 6.84e−08), PDCD1LG2 (p = 1.36e−02), LAG3 (p = 3.12e−05), CTLA4 (p = 5.42e−05) and SIGLEC15 (p = 1.16e−02) were upregulated in WHO grade III compared with WHO grade II of LGG (**Figures 5F,G**). In addition, we revealed that LGG patients in WHO grade III had a better response to immune checkpoint blockade compared with LGG patients in WHO grade II (**Figure 5H**). Moreover, the results demonstrated that TUBA1C positively co-expressed with these immune-checkpoints (**Table 1**), indicating that TUBA1C may be a potential immunotherapy target.

PATHWAY ENRICHED ANALYSES OF TUBA1C IN LGG

GO functional annotations and KEGG pathway analyses were further applied to investigate the potential actions of TUBA1C in LGG. KEGG pathway analyses revealed six pathways that correlated with the high TUBA1C expression: cell adhesion molecules cams; chemokine pathway; JAK-STAT pathway; T cell receptor signaling pathway; cytokine cytokine receptor interaction; natural killer cell mediated cytotoxicity (**Figure 6A**). GO functional annotations showed the following six positively correlated pathways that correlated with the high TUBA1C expression: leukocyte migration; negative regulation of immune system process; positive regulation of cytokine production; regulation of hemopoiesis; regulation of lymphocyte activation; and T cell activation (**Figure 6B**). These findings demonstrated that TUBA1C was involved in immune-related pathways in the TME of LGG.

DISCUSSION

LGG is a prevalent primary malignant tumor of the central nervous system (Ostrom et al., 2013). Despite the fact that surgical resection combined with chemotherapy and radiotherapy can improve clinical outcomes, over 50% of LGG patients evolve to therapy-resistant high-grade aggressive glioma over time (Zhang et al., 2020). Therefore, we urgently need to identify new prognostic factors for LGG. The immune microenviroment plays a vital role in the development of tumors (Gajewski et al., 2013). Preclinical and clinical immunotherapies, including immune checkpoint inhibitors (ICI), active or passive immunotherapies, and gene therapies, have been used in gliomas (Barzon et al., 2006; Tan et al., 2017; Lim et al., 2018), further confirming that immunotherapies are essential for LGG treatment. Therefore, it is necessary to determine more potential immune-related prognostic biomarkers for immunotherapies.

It has been reported that TUBA1C, a multifunctional cytoskeleton protein, is a kind of α-tubulin subtype and that it is correlated with microtubules, participating in the progress of cellular mitosis and division (Jordan and Wilson, 2004; Albahde et al., 2020). It has been revealed that high the expression of TUBA1C predicts a poor prognosis of hepatocellular carcinoma (Wang et al., 2017), lung adenocarcinoma (Bian et al., 2021), breast cancer (Wang et al., 2019), and osteosarcomas (Li et al., 2010). However, the role of TUBA1C in LGG and whether TUBA1C has immune-related biological functions in LGG have not been reported. The present study is the first to discover that TUBA1C was elevated in LGG tumors compared to normal tissues (**Figures 1B,C**). In addition, the mRNA expression of TUBA1C in grade III was more increased than in grade II of LGG (**Figure 1D**). Moreover, our findings also indicate that the mRNA expression of TUBA1C in GBM was higher than in LGG (**Figure 1D**). High expression of TUBA1C was verified to be an independent prognostic factor of LGG (**Figures 2H,I**). We further found that high TUBA1C expression

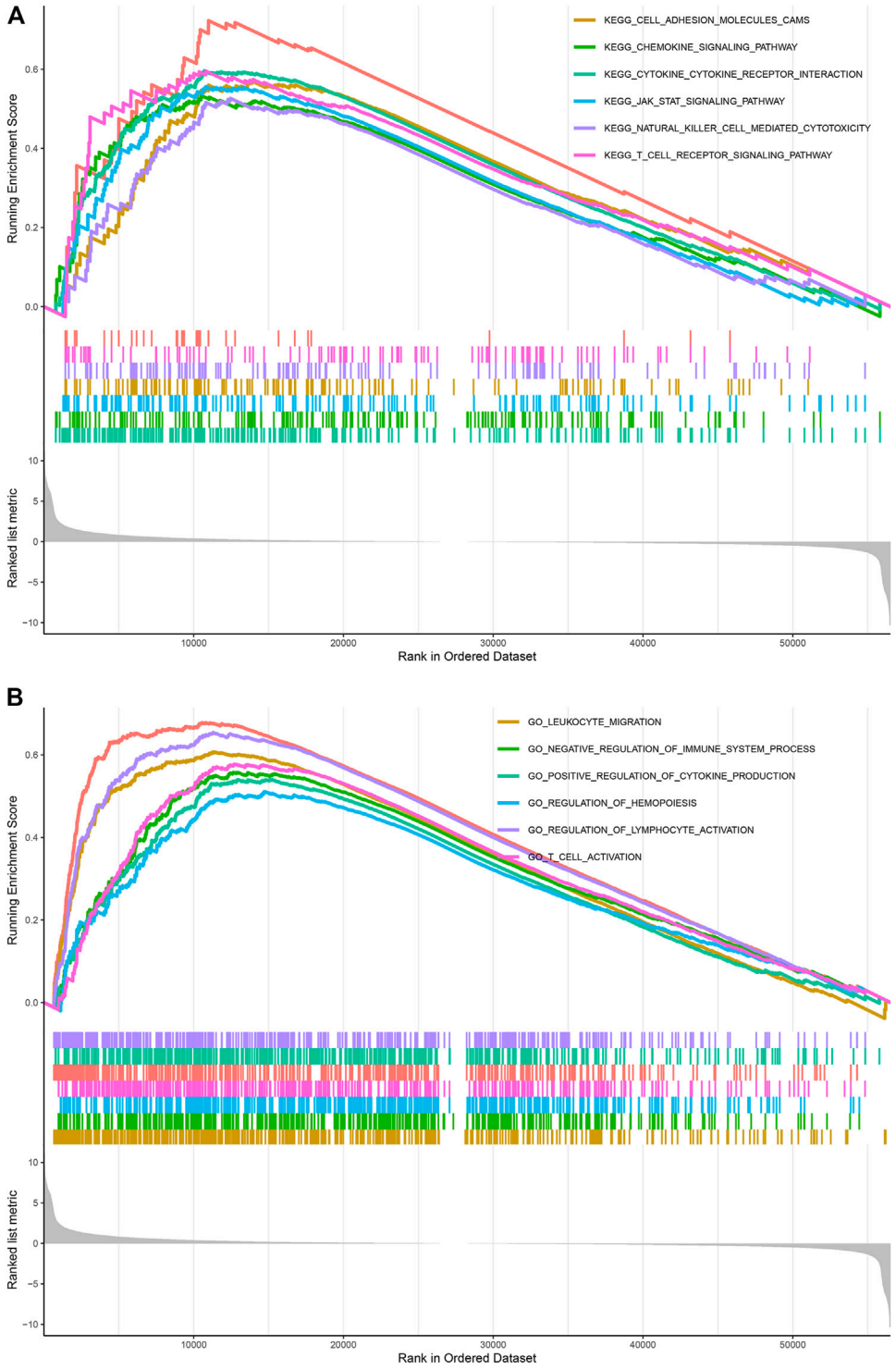


FIGURE 6 | KEGG and GO pathways enriched analyses results. **(A)** Enrichment of pathways in KEGG with high TUBA1C expression. **(B)** Enrichment of pathways in GO with high TUBA1C expression.

was related to poor survival in patients with LGG (**Figure 3**). These results demonstrated that TUBA1C may be served as a prognostic biomarker of LGG.

TME features have been verified to serve as biomarkers to evaluate tumor cell responses to immunotherapy and affect prognostic outcomes (Wu and Dai, 2017). Our results

demonstrated the TUBA1C played an important role in cancer immunity. TUBA1C expression positively correlated with stromal and immune cell content in TME of LGG based on ESTIMATE score (**Figures 4C,D**). We further used the TIMER database to investigate whether the expression of TUBA1C correlated with the TIICs. The results demonstrated that TUBA1C was positively correlated with the infiltration of B cells, CD8 T+ cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells (**Figure 4B**). Additionally, LGG patients with high infiltration of these immune cells had a poor cumulative survival compared with low infiltration with these immune cells. These findings indicate that high TUBA1C expression may be involved in poor survival through upregulating the infiltrating levels of B cells, CD8 T+ cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells. Moreover, TUBA1C had different correlations with different immune cell subtypes, such as macrophages M0, M1, M2 (**Figures 4J–L**). Similarly, a previous study has illustrated that TUBA1C relates to the tumor-infiltrating cells of lung adenocarcinoma (Bian et al., 2021). We also demonstrated that TUBA1C was co-expressed with multiple immune-related genes, such as particular MHC genes, immune activation genes, immunosuppresive genes, chemokine receptor related genes, and chemokine related genes (**Figures 5A–E**). In addition, the results demonstrated that TUBA1C positively co-expressed with immune-checkpoints (**Table 1**), including CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG2, TIGIT, and SIGLEC15. Moreover, LGG patients in grade III had higher expression of these immune-checkpoints and better response to immune checkpoint blockade than LGG patients in grade II (**Figures 5F–H**). These results indicate that LGG patients with high expression of TUAB1C may have a better prognosis after immune checkpoint blockade treatment, further illustrating that TUBA1C may be a potential target for immunotherapy.

Moreover, the enrichment analyses showed that TUBA1C influences the tumor development process through multiple immune-related pathways, including cell adhesion molecules cams; chemokine pathway; JAK-STAT pathway; cytokine cytokine receptor interaction; T cell receptor pathway; natural killer cell mediated cytotoxicity; (**Figure 6A**); leukocyte migration; positive regulation of cytokine production; negative regulation of immune system process; regulation of lymphocyte activation; and T cell activation (**Figure 6B**). These results indicated that TUBA1C was involved in immune-related pathways in the TME of LGG.

In conclusion, our work demonstrated that TUBA1C expression is increased in LGG tumor tissues and high expression of TUAB1C is associated with a poor prognosis. TUBA1C may influence the tumor development process by regulating the tumor-infiltrating cells in the TME. TUBA1C may be a potential target for immunotherapy.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All the listed authors directly, substantially, and intellectually contributed to the preparation of this manuscript and approved the publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (no. 81870939 to XX, no. 81771283 and 82071339 to LG) and the Natural Science Foundation of Hubei Province, China (no. 2019CFB498 to LG).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.759953/full#supplementary-material

REFERENCES

Albahde, M. A. H., Zhang, P., Zhang, Q., Li, G., and Wang, W. (2020). Upregulated Expression of TUBA1C Predicts Poor Prognosis and Promotes Oncogenesis in Pancreatic Ductal Adenocarcinoma via Regulating the Cell Cycle. *Front. Oncol.* 10, 49. doi:10.3389/fonc.2020.00049

Barzon, L., Zanusso, M., Colombo, F., and Palù, G. (2006). Clinical Trials of Gene Therapy, Virotherapy, and Immunotherapy for Malignant Gliomas. *Cancer Gene Ther.* 13 (6), 539–554. doi:10.1038/sj.cgt.7700930

Bian, T., Zheng, M., Jiang, D., Liu, J., Sun, H., Li, X., et al. (2021). Prognostic Biomarker TUBA1C Is Correlated to Immune Cell Infiltration in the Tumor Microenvironment of Lung Adenocarcinoma. *Cancer Cel Int* 21 (1), 144. doi:10.1186/s12935-021-01849-4

Bodakuntla, S., Jijumon, A. S., Villablanca, C., Gonzalez-Billault, C., and Janke, C. (2019). Microtubule-Associated Proteins: Structuring the Cytoskeleton. *Trends Cel Biol.* 29 (10), 804–819. doi:10.1016/j.tcb.2019.07.004

Boggs, A. E., Vitolo, M. I., Whipple, R. A., Charpentier, M. S., Goloubeva, O. G., Ioffe, O. B., et al. (2015). α-Tubulin Acetylation Elevated in Metastatic and Basal-like Breast Cancer Cells Promotes Microtentacle Formation, Adhesion, and Invasive Migration. *Cancer Res.* 75 (1), 203–215. doi:10.1158/0008-5472.Can-13-3563

Deng, Z., Yu, H., Wang, N., Alafate, W., Wang, J., Wang, T., et al. (2019). Impact of Preoperative Karnofsky Performance Scale (KPS) and American Society of Anesthesiologists (ASA) Scores on Perioperative Complications in Patients with Recurrent Glioma Undergoing Repeated Operation. *J. Neurorestoratology* 7 (3), 143–152. doi:10.26599/jnr.2019.9040015

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

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653

Hu, X., Zhu, H., Shen, Y., Zhang, X., He, X., and Xu, X. (2021). The Role of Non-coding RNAs in the Sorafenib Resistance of Hepatocellular Carcinoma. *Front. Oncol.* 11, 696705. doi:10.3389/fonc.2021.696705

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

Jordan, M. A., and Wilson, L. (2004). Microtubules as a Target for Anticancer Drugs. *Nat. Rev. Cancer* 4 (4), 253–265. doi:10.1038/nrc1317

Kim, N. D., Park, E.-S., Kim, Y. H., Moon, S. K., Lee, S. S., Ahn, S. K., et al. (2010). Structure-based Virtual Screening of Novel Tubulin Inhibitors and Their Characterization as Anti-mitotic Agents. *Bioorg. Med. Chem.* 18 (19), 7092–7100. doi:10.1016/j.bmc.2010.07.072

Le Rhun, E., Preusser, M., Roth, P., Reardon, D. A., van den Bent, M., Wen, P., et al. (2019). Molecular Targeted Therapy of Glioblastoma. *Cancer Treat. Rev.* 80, 101896. doi:10.1016/j.ctrv.2019.101896

Li, H., Jiang, X., Zhu, S., and Sui, L. (2017). Identification of Personalized Dysregulated Pathways in Hepatocellular Carcinoma. *Pathol. - Res. Pract.* 213 (4), 327–332. doi:10.1016/j.prp.2017.01.015

Li, W., Zhang, J., Zou, L., Cui, J., Su, F., Jin, J., et al. (2019). Palmitoylome Profiling Indicates that Androgens Regulate the Palmitoylation of α-tubulin in P-rostate C-ancer-derived LNCaP C-ells and S-upernatants. *Oncol. Rep.* 42 (6), 2788–2796. doi:10.3892/or.2019.7333

Li, Y., Liang, Q., Wen, Y.-q., Chen, L.-l., Wang, L.-t., Liu, Y.-l., et al. (2010). Comparative Proteomics Analysis of Human Osteosarcomas and Benign Tumor of Bone. *Cancer Genet. Cytogenet.* 198 (2), 97–106. doi:10.1016/j.cancergencyto.2010.01.003

Lim, M., Xia, Y., Bettegowda, C., and Weller, M. (2018). Current State of Immunotherapy for Glioblastoma. *Nat. Rev. Clin. Oncol.* 15 (7), 422–442. doi:10.1038/s41571-018-0003-5

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

Ostrom, Q. T., Gittleman, H., Farah, P., Ondracek, A., Chen, Y., Wolinsky, Y., et al. (2013). CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2006-2010. *Neuro-Oncology* 15 (Suppl. 2), ii1–ii56. doi:10.1093/neuonc/not151

Tan, A. C., Heimberger, A. B., and Khasraw, M. (2017). Immune Checkpoint Inhibitors in Gliomas. *Curr. Oncol. Rep.* 19 (4), 23. doi:10.1007/s11912-017-0586-5

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

Tsai, H.-P., Tsai, C.-Y., Lieu, A.-S., Chai, C.-Y., Kwan, A.-L., Howng, S.-L., et al. (2014). Association of Aurora A and Gamma-Tubulin Expression in Astrocytomas and Patient Survival. *Neurol. Res.* 36 (8), 746–751. doi:10.1179/1743132813y.0000000310

Wang, C. C. N., Li, C. Y., Cai, J. H., Sheu, P. C., Tsai, J. J. P., Wu, M. Y., et al. (2019). Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *Jcm* 8 (8), 1160. doi:10.3390/jcm8081160

Wang, J., Chen, W., Wei, W., and Lou, J. (2017). Oncogene TUBA1C Promotes Migration and Proliferation in Hepatocellular Carcinoma and Predicts a Poor Prognosis. *Oncotarget* 8 (56), 96215–96224. doi:10.18632/oncotarget.21894

Wang, X., Sun, N., Meng, X., Chen, M., Jiang, C., and Cai, J. (2020). Review of Clinical Nerve Repair Strategies for Neurorestoration of Central Nervous System Tumor Damage. *J. Neurorestoratology* 8 (3), 172–181. doi:10.26599/jnr.2020.9040018

Wu, T., and Dai, Y. (2017). Tumor Microenvironment and Therapeutic Response. *Cancer Lett.* 387, 61–68. doi:10.1016/j.canlet.2016.01.043

Xiong, Y., Wang, K., Zhou, H., Peng, L., You, W., and Fu, Z. (2018). Profiles of Immune Infiltration in Colorectal Cancer and Their Clinical Significant: A Gene Expression-based Study. *Cancer Med.* 7 (9), 4496–4508. doi:10.1002/cam4.1745

Zhang, M., Wang, X., Chen, X., Zhang, Q., and Hong, J. (2020). Novel Immune-Related Gene Signature for Risk Stratification and Prognosis of Survival in Lower-Grade Glioma. *Front. Genet.* 11, 363. doi:10.3389/fgene.2020.00363

Zhang, Y. Z., Chen, X., Fan, X. X., He, J. X., Huang, J., Xiao, D. K., et al. (2016). Compound Library Screening Identified Cardiac Glycoside Digitoxin as an Effective Growth Inhibitor of Gefitinib-Resistant Non-small Cell Lung Cancer via Downregulation of α-Tubulin and Inhibition of Microtubule Formation. *Molecules*, 21(3), 374. doi:10.3390/molecules21030374

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 October 2021
doi: 10.3389/fgene.2021.767259

Analysis of Genome-Wide Alternative Splicing Profiling and Development of Potential Drugs in Lung Adenocarcinoma

*Jing Song[1], Jia Liu[2], Dekang Lv[3], Xuan Meng[4] and Xiaodong Li[3]**

[1]Department of Respiratory Medicine, Qinzhou First People's Hospital, The Tenth Affiliated Hospital of Guangxi Medical University, Qinzhou, China, [2]Department of Gynecology, Cancer Hospital of China Medical University, Dalian, China, [3]Institute of Cancer Stem Cell, Dalian Medical University, Dalian, China, [4]School of Medicine, Southern University of Science and Technology, Shenzhen, China

Alternative splicing (AS) is significantly related to tumor development as well as a patient's clinical characteristics. This study was designed to systematically analyze the survival-associated AS signatures in Lung adenocarcinoma (LUAD). Among 30,735 AS events in 9,635 genes, we found that there were 1,429 AS in 1,125 genes which were conspicuously related to the overall survival of LUAD patients. Then, according to the seven types of AS events, we established AS signatures and constructed a new combined prognostic model. The Kaplan-Meier curve results showed that seven types of AS signatures and the combined prognostic model could divide patients into distinct prognoses. The ROC curve shows that all eight AS signatures had powerful predictive properties with different AUCs ranging from 0.708 to 0.849. Additionally, the elevated risk scores were positively related to higher TNM stage and metastasis. Interestingly, AS events and splicing factors (SFs) network shed light on a meaningful connection between prognostic AS genes and corresponding SFs. Moreover, we found that the combined prognostic model signature has a higher predictive ability than the mRNA signature. Furthermore, tumors at high risk might evade immune recognition by decreasing the expression of antigen presentation genes. Finally, we predicted the three most significant small molecule drugs to inhibit LUAD. Among them, NVP-AUY922 had the lowest IC50 value and might become a potential drug to prolong a patient's survival. In conclusion, our study established a potential prognostic signature for LUAD patients, revealed a splicing network between AS and SFs and possible immune escape mechanism, and provided several small-molecule drugs to inhibit tumorigenesis.

Keywords: lung adenocarcinoma, alternative splicing, splicing factor, prognosis, immune infiltration, small molecule drugs

OPEN ACCESS

Edited by:
Xiangqian Guo,
Henan University, China

Reviewed by:
Shuangyu Lv,
Henan University, China
Huimin Li,
Henan University, China

****Correspondence:***
Xiaodong Li
xd_li201212@dmu.edu.cn

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

Received: *30 August 2021*
Accepted: *30 September 2021*
Published: *19 October 2021*

Citation:
Song J, Liu J, Lv D, Meng X and Li X (2021) Analysis of Genome-Wide Alternative Splicing Profiling and Development of Potential Drugs in Lung Adenocarcinoma. Front. Genet. 12:767259. doi: 10.3389/fgene.2021.767259

INTRODUCTION

Lung cancer is a leading cause of cancer-related mortality worldwide (Chen et al., 2016; Siegel et al., 2021). Non-small cell lung cancer (NSCLC) is the most prevalent and heterogeneous subtype of lung cancer, including lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) (Chang et al., 2015). Generally, when LUAD and LUSC are in the same stage, the growth of LUAD is relatively slow with smaller masses. However, LUAD are more likely to start metastasis at the early

stages. Although genomic data such as gene expression, DNA methylation, and copy number variation have been extensively studied in cancer, comprehensive and systematic analysis of alternative splicing is insufficient despite the fact that it has played a significant role in cancer (Ludwig and Weinstein, 2005; Koch et al., 2018; Han et al., 2019). Thus, it is necessary to further investigate the function of alternative splicing events on cancer recurrence and metastasis, especially in LUAD.

Alternative splicing (AS) is an extensive and sophisticated mechanism to increase the diversity of the proteome structurally and functionally (Climente-González et al., 2017). AS could exert a far-reaching influence on a protein's biological characteristics by changing its stability, adding or deleting functional domains, controlling gene location, or regulating protein-protein interactions (Lee and Abdel-Wahab, 2016). Furthermore, it is a ubiquitous process and over 95% of genes undergo variable AS to produce a variety of transcripts (Feng et al., 2013). Normal alternative splicing could generate a multi-functional proteome to exert healthy cellular functions, while unusual alternative splicing can result in the occurrence and deterioration of cancers (Baralle and Giudice, 2017; Park et al., 2018). For instance, aberrant AS events could regulate the development and progression of a tumor by participating in several biological processes such as cell cycle progression, cell proliferation, and RNA processing. Accumulated studies have also highlighted that AS events have gradually become one of the hallmarks of carcinogenesis due to their unbalanced or incorrect expression (Ladomery, 2013). Thus, unusual AS events can be potential targets for cancer treatment.

The functions and roles of AS events in various cancers have been explored in a number of studies. For example, Bechara et al. (2013) demonstrated that NUMB alternative splicing regulated by RBM5, 6, and 10 could control lung cancer cell proliferation. Wan et al. (2019) verified that SRSF6 could mediate colorectal cancer progression by regulating alternative splicing. The alternative splicing of CCDC50 regulated by HnRNP A1 can result in clear cell renal cell carcinoma (ccRCC) tumorigenesis and development (Sun et al., 2020). However, few articles have systematically reported LUAD-specific AS events correlated with clinical characteristics.

Extensive dysregulated AS events in many types of cancers are easily programmed by many SFs, particularly the serine/arginine-rich (SR) and the heterogeneous nuclear ribonucleoproteins (hnRNPs) family. Different variate mature mRNAs were produced by SFs to assist spliceosome recognition and binding specific sequence of precursor mRNA (de Almeida and Carmo-Fonseca, 2012). The expression level of hnRNPs is different in many types of cancers, suggesting their extraordinary roles in tumorigenesis. Therefore, it is essential to depict an exhaustive regulatory network of SFs (Kędzierska and Piekiełko-Witkowska, 2017; Ratnadiwakara et al., 2018). Because the intimate correlation between AS and SFs was only superficially understood in view of their complexity, it is significant to investigate their potential prognostic performance, as well as their regulatory mechanism in LUAD.

The present study analyzed genome-wide LUAD-specific AS events using RNA-seq data in The Cancer Genome Atlas (TCGA) program, providing a systematically new understanding of the potential prognostic effects of AS events in LUAD. The purpose of this study was to clarify the roles of splice variants that could be considered as prognostic biomarkers in LUAD. Finally, the research uncovered interesting splicing networks in LUAD which could contribute to a better understanding of the fundamental mechanisms of LUAD.

MATERIAL AND METHODS

Alternative Splicing Events Data Collection

SpliceSeq data of TCGA-LUAD were downloaded from the TCGA database (https://tcga-data.nci.nih.gov/tcga/) (Tomczak et al., 2015). SpliceSeq tool, a java-based application, was usually used to unambiguously quantify the mRNA splicing levels of samples in TCGA. A novel value could be calculated by SpliceSeq based on seven types of AS events about each protein-coding gene provided from the Ensemble gene database (Ryan et al., 2012). For the following seven kinds of AS events, the Percent Spliced In (PSI) value was calculated, quantifying splicing event levels range from 0 to 1: Mutually Exclusive Exons (ME), Exon Skip (ES), Alternate Promoter (AP), Retained Intron (RI), Alternate Acceptor site (AA), Alternate Donor site (AD), and Alternate Terminator (AT). The schematic graph explaining these seven types of AS is shown in **Supplementary Figure S1A**.

Establishment of the Prognostic Model

Patients' clinical parameters of LUAD were downloaded and extracted from the TCGA database. In total, 444 LUAD patients were included in this analysis. Clinical information for these patients and the pathological details obtained from TCGA are provided in **Supplementary Table S1**. The PSI value of AS events in samples were collected and subjected to univariate Cox analysis. All the AS events screen for a p value <0.05, and these events were considered as candidate prognosis-related events. The "glmnet R″ software package was used to perform the least absolute shrinkage and selection operator (LASSO) analysis to filter out the most valuable and concise AS events in all AS events filtered in univariate Cox analysis ($p < 0$ .05). Afterward, the prognostic independence of the AS signature was constructed by multivariate Cox analysis. Then, based on the coefficient of each above AS event, each patients' risk score could be calculated by the signature, respectively. Meanwhile, all patients were divided into distinct subgroups based on the median value of risk scores.

Survival Analysis

The Kaplan-Meier curve was implemented to evaluate the differential survival status in both groups. The receiver operating characteristic curves (ROC) were conducted to detect both the sensitivity and specificity of prognostic signatures using the "survivalROC" R package (https://www.r-project.org/, v3.5.3).

UpSet Plot and Splicing Factor Regulatory Network Establishment

We developed the Upset intersective plot, a more scalable visualizing diagram than Venn, which was used to explore the

interactive sets of AS events. The "UpSet" R package was used to visualize their potential interrelationship. The expression data of the Splicing factors (SFs) was extracted from TCGA-LUAD mRNA-seq data. All SF genes were subjected to univariate Cox analysis when their $p < 0.05$. These SFs were considered the survival-associated splicing factors. The relationship closeness between SFs expression value and AS's PSI value were calculated by the Spearman test. At the same time, the interaction network diagram of these SFs and prognosis-related AS events was illustrated using Cytoscape 3.7.0 (https://cytoscape.org/).

Evaluation of Tumor Immune Cells Infiltrating

22 types of tumor-infiltrating immune cells were estimated using the CIBERSOFT algorithm, which characterized the cellular composition of complex tissues based on normalized gene expression profiles. The comparison of immune cell distribution between high and low-risk groups was made using the Mann-Whitney *U*-test.

Identification of Potential Small Molecule Drugs

Potential drugs for the treatment of LUAD were selected using the Connectivity Map (CMap) database (https://clue.io/). We uploaded differentially expressed genes between high and low-risk groups into the CMap database for genomic enrichment analysis. We screened the small drug molecules with an enrichment score of <90, and obtained 3D structure through PubChem database (http://www.pubchem.ncbi.nlm.gov), a public repository of small molecules in properties. The IC50 value of these small molecule drugs was provided by the GDSC database (https://www.cancerrxgene.org/).

RESULTS

Overview of AS Events in TCGA-LUAD

Comprehensive AS events were examined in a cohort of 444 TCGA-LUAD patients (**Supplementary Figure S1A**).A total of 30,735 AS events from 9,635 genes were detected, including 11,768 ES events in 5,467 genes, 6,129 AP events in 3,424 genes, 5,782 AT events in 3,372 genes, 2,605 AA events in 1,961 genes, 2,199 AD events in 1,659 genes, 2,103 RI events in 1,461 genes, 149 ME events in 146 genes (**Supplementary Figure S1B**). In TCGA-LUAD, ES events were the most common AS events, accounting for approximately just over one-third of all events, followed by the number of AP and AT events, while the number of ME events was the least. Remarkably, the number of AS events went far beyond those of their corresponding mRNAs. Furthermore, a subset of overlapping AS events in the seven types of AS in LUAD was displayed by the UpSet plot diagram (**Supplementary Figure S1C**).

Identification of Prognosis-Related AS Events in LUAD

First, we conducted a univariate Cox analysis based on the 30,735 AS events related to 444 patients to appraise the relationship between AS events and overall survival (OS) status in LUAD. Consequently, 1,429 AS events and corresponding 1,125 genes were conspicuously related to the overall survival of LUAD patients (**Figure 1A**). **Figures 1B–H** showed the top 20 most important AS events related to OS among these seven types of AS events. Interestingly, some special survival-associated AS genes contain multiple types of AS events. For example, AD, AT, AA, RI, and ES of C1orf159 and AP, AA, RI, AD, and ES of MRPL55 were all related to OS of LUAD patients.

Establishment of Prognostic AS Signatures

We selected the significant prognostic associated AS events as candidates by univariate Cox analysis, aiming to further screen out the most significant AS events related to patient prognosis by LASSO Cox regression analysis (**Supplementary Figure S2**). Furthermore, several prediction signatures based on these prognostic associated AS events were constructed by multivariate Cox analyses. Eventually, a combined prognostic model was built, integrated from different types of AS events (**Supplementary Table S2**). The median value of risk scores was considered as the cutoff criteria for dividing patients into a high-risk group and a low-risk group. The Kaplan–Meier curves shown in **Figures 2A–H**, showed that LUAD patients in the high-risk group had appreciably shorter OS than patients in the low-risk group, demonstrated that these AS signatures could be powerful biomarkers to distinguish patient prognosis. The combined prognostic signature showed better predictive properties than the single type of AS events (**Figure 2H**). Then, the ROC curve was performed to appraise the prognostic efficiency of prognostic AS models. The results show that all signatures had a robust predictive property with AUC values from 0.708 to 0.849, except the ME signature (AUC = 0.582, **Figure 2I**). Conceivably, the combined model contains different types of AS events that had the highest efficiency (AUC = 0.849). The distribution diagram of patients' risk score, survival status, and expression profiles of related AS model events are shown in **Supplementary Figures S3A–H**.

Validation of Combined AS Signature

To confirm the prognostic value of our combined AS signature, we randomly selected 50% of patients in all LUAD patients as a testing cohort and performed Kaplan-Meier cure and ROC analysis (**Supplementary Figure S4**). The results were consistent with the above results and demonstrated the combined AS signature had a robust predictive ability (AUC = 0.856).

AS Signatures Are Independent Factors for Other Clinical Characteristics

To explore the predicted performance of AS signatures and other clinical characteristics of survival, univariate and multivariate

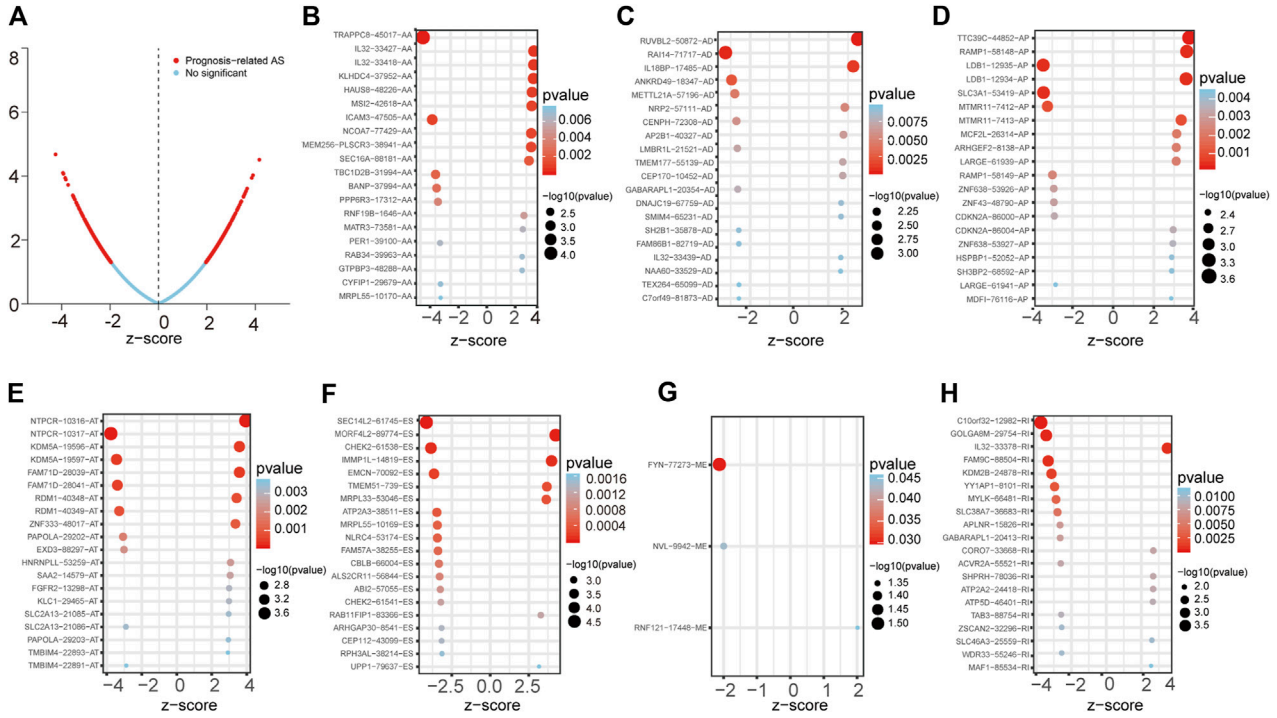


FIGURE 1 | Forest plots analyses of survival-associated AS events. **(A)** Volcano plot depicting the *p* values from univariate Cox analysis of the 30,735 AS events. **(B–H)** Forest plots of z-score of the top 20 significantly survival-related AS events for seven splicing types (ME only three events).

Cox analyses were performed to find out if these AS signatures were independent prognostic factors for LUAD patients. The univariate Cox analysis showed that almost all risk score signatures (except ME-risk score signature), TNM stage, T, N, and M stage were remarkably related to LUAD patients' overall survival (**Supplementary Figure S5**). Furthermore, multivariate Cox analysis results show that most risk score signature, T and N stage still have a predictive ability when all univariate significant factors are considered together, suggesting that risk score, T and N stage were unassisted risk elements (**Figures 3A–G**). Taken together, all these results demonstrated that AS signatures exhibit powerful predictive performance in LUAD patients. In addition, circos plots were depicted to display the details of AS events and their interacting genes in the chromosome (**Figure 3H**).

Network of Prognosis-Related AS Events and SFs

Interestingly, extensive dysregulated AS events in many types of cancers are easily programmed by some specific SFs. Thus, an interesting issue is whether several key SFs could regulate these prognosis-associated AS events in LUAD. To determine those specific SFs which had a close connection with prognosis-associated AS events in LUAD, univariate Cox analysis of all SFs were performed based on the gene expression value of LUAD patients. The results showed that there were 30 SFs related to the OS of LUAD patients (**Supplementary Table S3**). Furthermore, correlations between SFs and prognostic AS events were tested by Spearman's test (**Figure 4A**). In the correlation networks, 27 SFs (purple dots) were related to 248 prognosis-associated AS events, involving 141 favorable AS events (green dots) and 107 adverse AS events (red dots). Intriguingly, there was a positive relationship (red lines) between most of the poor survival-related AS events (red dots) and SFs (purple dots), while there was a negative correlation (green lines) between most of the favorable survival-related AS events (green dots) and SFs. For example, SFs RNF34 and HNRNPK were related to worse survival of LUAD patients (**Figures 4B,C**). ES events of YPEL5 were an adverse factor, while the AP events of PCNA and AP events of PDHX were related to a favorable prognosis. The relationship between RNF34 and the AP of PCNA or ES of YPEL5 were shown in dot plots, implicating the high expression of RNF34 was negatively related to adverse prognosis (**Figures 4D,E**). Similarly, the relationship between HNRNPK and the AP of PDHX or ES of YPEL5 were shown in dot plots (**Figures 4F,G**), illustrating that the high expression of HNRNPK was positively related to poor prognosis.

The AS Signature has Better Predictive Ability Than the mRNA Signature

We finally constructed an mRNA signature for LUAD patients by subjecting differentially expressed mRNAs in LUAD to univariate and multivariate Cox analysis: mRNA risk score = (0.1674*XPR1) + (0.0529*MMP1) + (-0.2108*MAP3K8) + (0.0882*RHOV) + (0.1099*CDH17) + (0.1488*UCA1). Then, Kaplan-Meier

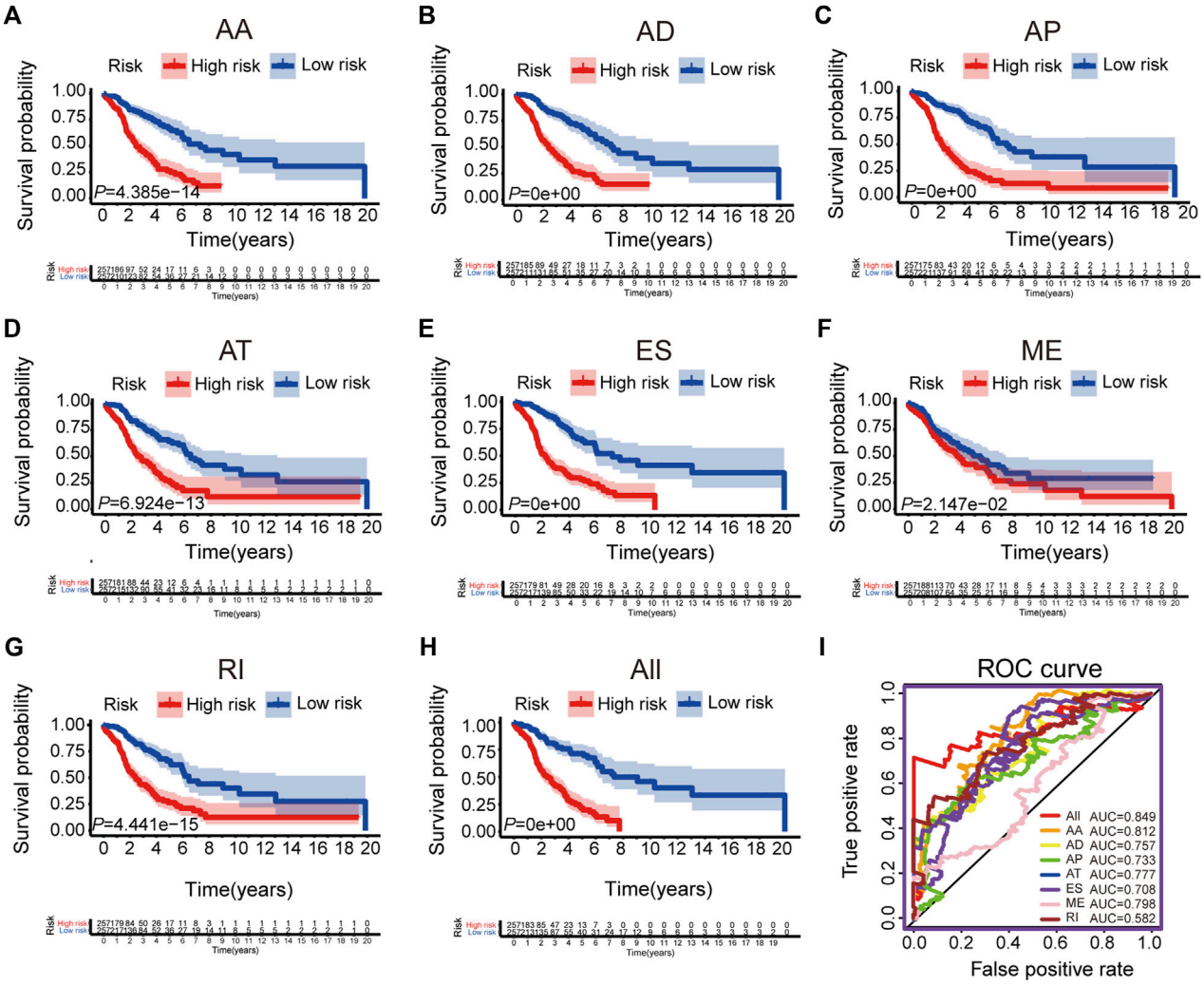


FIGURE 2 | The Kaplan-Meier curves and ROC curves of prognostic AS models. **(A–G)** The Kaplan-Meier plots of seven types of AS events, respectively. **(H)** The Kaplan-Meier plots of combined prognostic model. **(I)** The ROC curves for overall survival of seven types of AS events and combined prognostic model.

and ROC curves were implemented to contrast prognostic ability between AS signature and mRNA signature. Both results from Kaplan-Meier and ROC analyses illustrated that AS signature had significantly better survival and higher ROC than mRNA signature (**Figures 5A,B**). These data demonstrated that the predictive power of AS signature is greater than the mRNA signature. In general, AS signature could be used as a superior indicator to predict the prognosis of LUAD patients.

Tumor Microenvironment Cell Infiltration Characteristics Related to Drug Resistance

We also analyzed tumor microenvironment cell infiltration of signature. We calculated the median absolute score that CIBERSORT gave for 22 cell types in two groups. The results showed that the fraction of T cells follicular helper, NK cells resting, Monocytes, and Macrophages M1 was significantly higher in high-risk than that in the low-risk group. However, B cells naive, Macrophages M0 and Mast cells resting were remarkably higher in the low-risk than in high-risk group (**Figure 6A**). Furthermore, the results showed that the expression of some immunomodulator agonists was correlated with a risk score. We found that some immune ligands and receptors were significantly higher in the high-risk group, while antigen presentation was mainly higher in the low-risk group (**Figure 6B**).

Related Small Molecule Drugs Screening

To predict small molecule drugs that can inhibit the resistance of LUAD patients, DEGs of high-risk and low-risk groups were assigned into up-regulated and down-regulated groups. Then we matched it to a small molecule drug in the CMap database. Finally, we selected the three most important small molecular compounds, including AZ-628 (score = −95.17), NVP-AUY922 (score = −91.77) and Nomifensine (score = −90.86). The 3D structure of these three small molecular compounds was

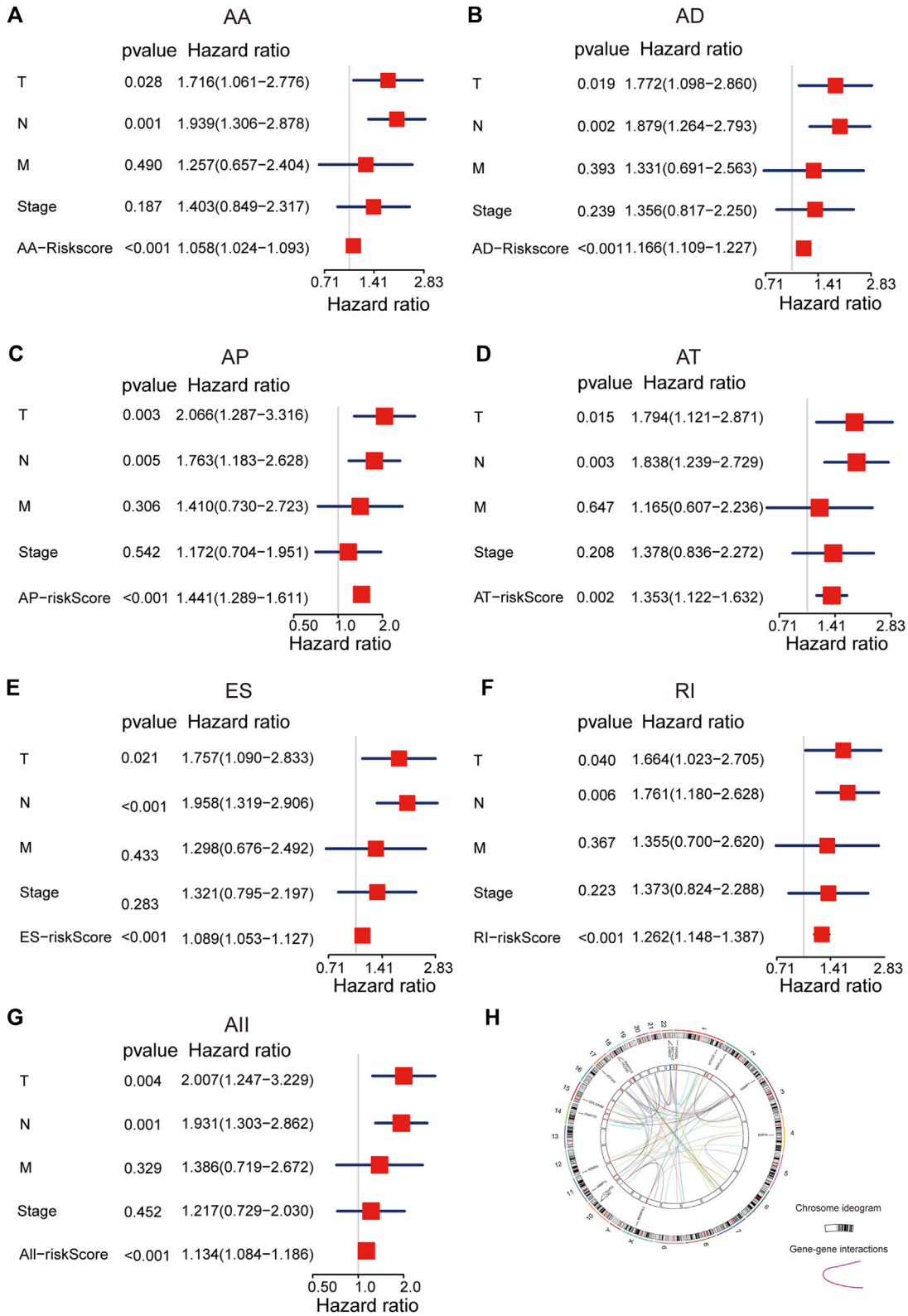


FIGURE 3 | The multivariate Cox analysis of AS signatures and clinical characteristics. **(A–G)** Forest plots of hazard ratios of risk scores and clinical characteristics from multivariate Cox analyses. **(H)** Circos plots of the detail of AS events and its interacting genes in chromosome.

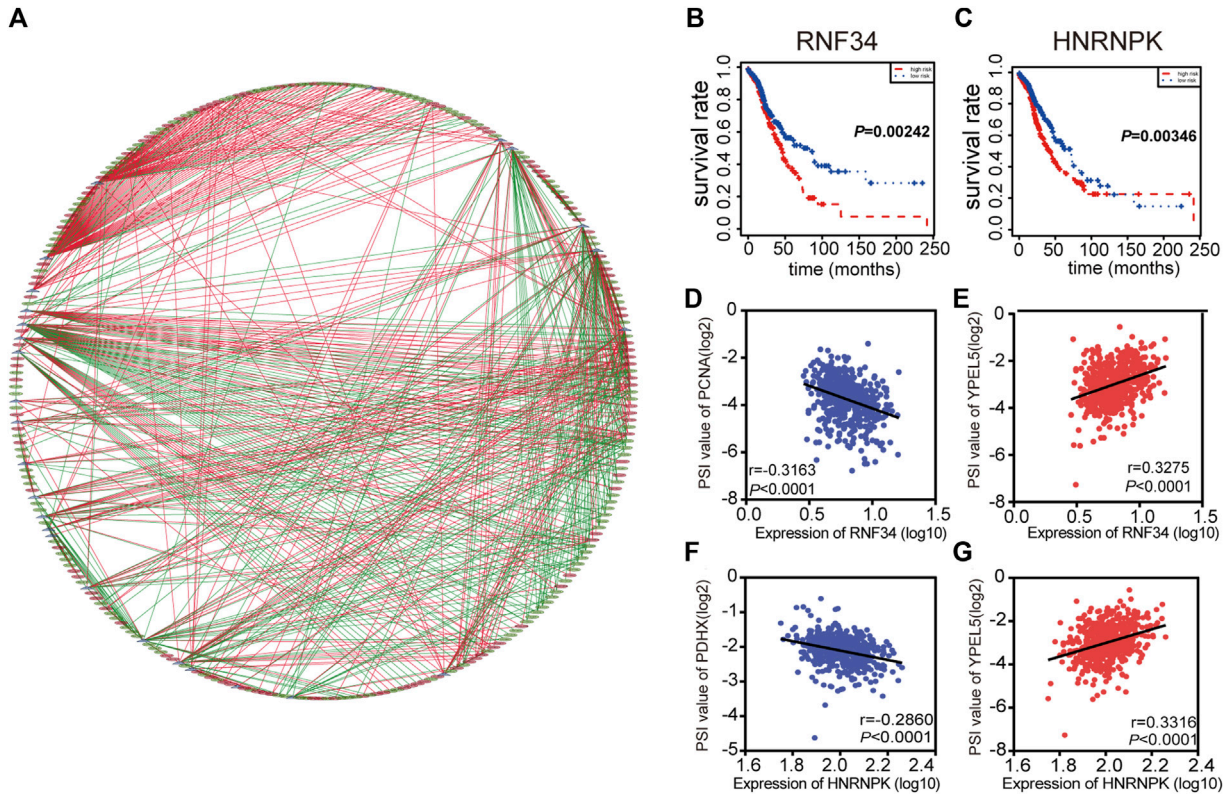


FIGURE 4 | Splicing correlation network in LUAD. **(A)** Correlation network between expression of survival Splicing factors and PSI values of AS genes generated using Cytoscape. Purple dots were survival associated splicing factors. Green/Red dots were favorable/adverse AS events. Red/Green lines represent positive/negative correlations between substances. **(B–C)** Kaplan-Meier curve of splicing factors RNF34 and HNRNPK. **(D–G)** Representative dot plots of correlations between expression of splicing factors and PSI values of AS events.

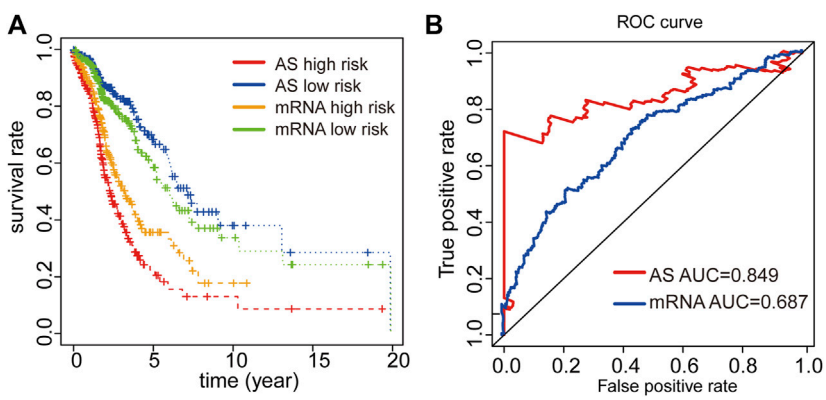


FIGURE 5 | Comparison of Kaplan-Meier and time-dependent ROC analysis of our AS signature with mRNA signature. **(A)** Comparison of Kaplan-Meier analysis of our AS signature and mRNA signature. **(B)** Comparison of ROC analysis of the sensitivity and specificity of our AS signature and mRNA signature.

downloaded from the PubChem database (**Figure 7**). These small molecules might potentially improve the outcomes of LUAD patients and provide recommendations for the selection of LUAD-targeted drugs, while the specific mechanism and effectiveness need to be further studied. We also explored the IC50 values of these small molecule drugs in LUAD cells through the GDSC database. The results showed that NVP-AUY622 (IC50 = 0.004–10.222 μM) had lower IC50 values than AZ-628 (IC50 = 0.050–46.073 μM), indicating that NVP-AUY622 has a stronger effect in LUAD cells (nomifensine not provided). These results demonstrated that NVP-AUY922 might become a novel drug to improve the survival of LUAD patients.

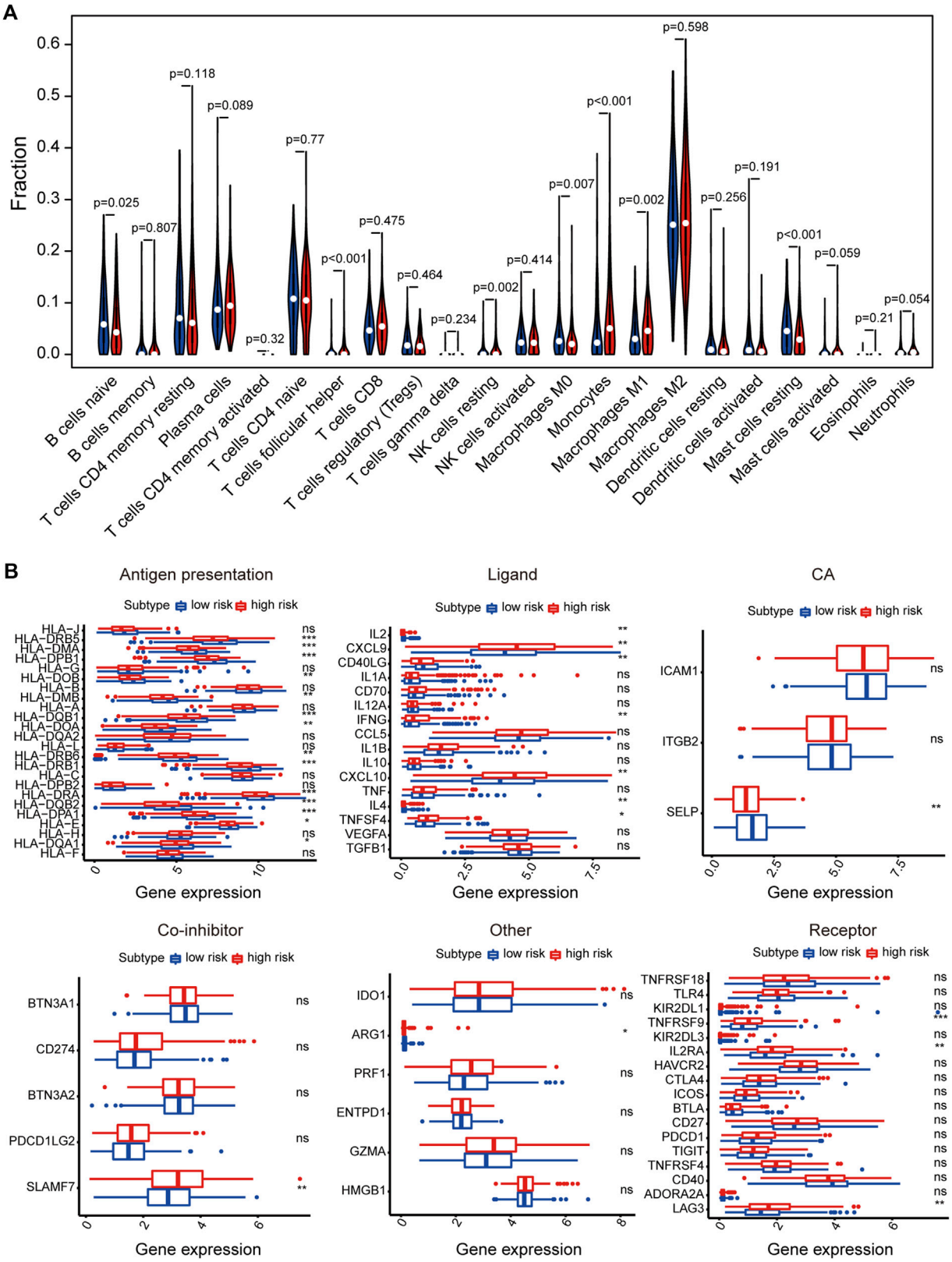


FIGURE 6 | The relationship between signature and immune-related features. **(A)** The violin plot of the 22 immune cell proportions between high and low-risk groups. **(B)** Effect of signature on the expression of different immunomodulators. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

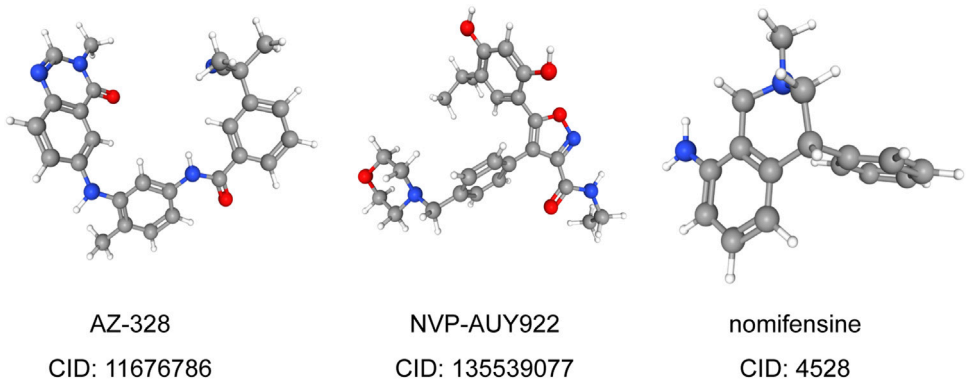


FIGURE 7 | 3D conformer of most significant small molecule drugs.

DISCUSSION

Alternative splicing of pre-mRNA is a ubiquitous and flexible process, which could provide the opportunity for cells to generate different or even the opposite functions of protein isoforms from a single gene. Furthermore, this flexibility is usually used by tumor cells to generate proteins that facilitate growth and progression (David and Manley, 2010). Changes in AS might affect all aspects of tumor biology, including metastasis, invasion, metabolism, and apoptosis, as well as angiogenesis (Blencowe, 2006; Moore et al., 2010; Coomer et al., 2019). For example, *David et al* suggested that regulating AS events can contribute to tumor cell proliferation (David et al., 2010). KLF6-SV1 (splice variant 1) exerts an essential role in the development and progression of cancer (Narla et al., 2005; DiFeo et al., 2009). All in all, it is necessary to further investigate not only mRNA expression levels but also some splice variants.

In recent decades, accumulating studies have shown that AS events affect proliferation, migration, radioresistance, and so on (Yae et al., 2012; Sheng et al., 2018; Wang et al., 2020a). For example, Sheng et al. (2018) found that SRSF1 is involved in radioresistance in lung cancer cells through modulating the aberrant splicing of PTPMT1. In addition, QKI-5 could inhibit cell proliferation and the migration of lung cancer via regulating the splicing of ADD3 exon 14 (Wang et al., 2020a). However, prognosis-related AS events in LUAD remain mostly unstudied, especially in patients with metastasis.

SpliceSeq, a novel convenient exploited analysis pipeline (integration tool), was used to detect AS events, which could help analyze complex or low-frequency AS events (Ryan et al., 2012). Here, we systematically and comprehensively analyzed a total of 30,735 AS events in a TCGA-LUAD cohort and identified 1,429 prognosis-associated AS events in 1,125 genes. Our results indicated that a combined prognostic signature containing several kinds of AS events had excellent performance in survival prediction. More importantly, we revealed that some AS events were associated with the metastasis of LUAD patients, which might be a meaningful discovery in further exploring the mechanisms of LUAD metastasis. In a previous study, Tyler et al. detected some splicing variants and found that CHEK2 is a prominent suppressor gene in cancer (Landrith et al., 2020). KDM5A could promote SCLC proliferation and metastasis *in vivo* by repressing the NOTCH signaling pathway (Oser et al., 2019). These results have been verified in our analysis.

As an increasing number of studies have demonstrated that AS events are modulated by some pivotal SFs (Cieply and Carstens, 2015; Song et al., 2018), the analysis of SFs was also underlined in this study. Intriguingly, the splicing correlation network suggested that there was a positive relationship between most poor survival-related AS events and SFs, while there was a negative correlation between most favorable survival-related AS events and SFs. For example, HNRNPK have adverse factors in LUAD and were found to be positively associated with ES of YPEL5, which was also considered an adverse factor. Moreover, HNRNPK has been reported and could promote metastasis in lung cancer (Li et al., 2018; Li et al., 2019). It is fair to say that poor-survival SFs might facilitate the occurrence of adverse prognostic AS events. Nevertheless, it is necessary to further explore the more specific regulatory mechanisms of AS-SF networks.

In the analysis of the correlation between risk score and immune cells, we found that T cells follicular helper, NK cells resting, Monocytes and Macrophages M1 were mainly enriched high-risk group, while B cells naive, Macrophages M0 and Mast cells resting were enriched in the low-risk group. Jhunjhunwala et al. (2021) that found tumors can down-regulate the expression of HLA-1 by interfering with antigen processing and presentation mechanisms to achieve tumor evasion immune recognition. In our study, most of the antigen presentation genes were a low expression in a high-risk group, which might promote tumor immune escape.

More importantly, we identified the three most significant small molecule drugs, including AZ-628, NVP-AUY922, and Nomifensine, that might improve the survival of LUAD patients, In particular, NVP-AUY622 has the lowest IC50 value in LUAD cell lines. Among these three drugs, AZ-628, an RAF Kinase inhibitor, can reverse cancer multidrug resistance (MDR) by mediating ATP-Binding Cassette Transporter G2 (ABCG2) (Wang et al., 2020b). NVP-AUY922 has been found to have potent anti-tumor activity and can inhibit tumor growth, including NSCLC, breast cancer, colorectal cancer, and so on (Jensen et al., 2008; Garon et al., 2013; Lee et al., 2015). In colorectal cancer cells, HSP90 inhibitor NVP-AUY922 can

suppress the JAK2-STAT3-Mcl-1 signaling pathway to enhance TRAIL-induced apoptosis (Lee et al., 2015). At present, as an antidepressant, Nomifensine is mainly used to diagnose and test hyperprolactinemia (Müller et al., 1978). However, the anti-tumor effect of Nomifensine needs further confirmation.

In summary, this study indicated the prognostic value of some AS events in LUAD and patients with metastasis, which could modulate some key SFs. These results further understanding of the interaction between AS and SFs in LUAD, indicating that the systematic analysis of AS signatures in LUAD might contribute multiple potential biomarkers and the underlying mechanisms of LUAD metastasis. In addition, we provided three small molecule drugs for treatment selection of 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

XL and JS designed the study and wrote the paper. JS and JL performed research. JS, JL, DL, and XM analyzed data. JS and XL wrote the first draft of the paper and edited the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported with grants from a Liaoning Provincial Natural Science Foundation of China (2015020670 to XL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.767259/full#supplementary-material

REFERENCES

Baralle, F. E., and Giudice, J. (2017). Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cel Biol* 18 (7), 437–451. doi:10.1038/nrm.2017.27

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., and Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. *Mol. Cel.* 52 (5), 720–733. doi:10.1016/j.molcel.2013.11.010

Blencowe, B. J. (2006). Alternative Splicing: New Insights from Global Analyses. *Cell* 126 (1), 37–47. doi:10.1016/j.cell.2006.06.023

Chang, J. T.-H., Lee, Y. M., and Huang, R. S. (2015). The Impact of the Cancer Genome Atlas on Lung Cancer. *Translational Res.* 166 (6), 568–585. doi:10.1016/j.trsl.2015.08.001

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

Cieply, B., and Carstens, R. P. (2015). Functional Roles of Alternative Splicing Factors in Human Disease. *WIREs RNA* 6 (3), 311–326. doi:10.1002/wrna.1276

Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cel Rep.* 20 (9), 2215–2226. doi:10.1016/j.celrep.2017.08.012

Coomer, A. O., Black, F., Greystoke, A., Munkley, J., and Elliott, D. J. (2019). Alternative Splicing in Lung Cancer. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1862 (11-12), 194388. doi:10.1016/j.bbagrm.2019.05.006

David, C. J., Chen, M., Assanah, M., Canoll, P., and Manley, J. L. (2010). HnRNP Proteins Controlled by C-Myc Deregulate Pyruvate Kinase mRNA Splicing in Cancer. *Nature* 463 (7279), 364–368. doi:10.1038/nature08697

David, C. J., and Manley, J. L. (2010). Alternative Pre-mRNA Splicing Regulation in Cancer: Pathways and Programs Unhinged. *Genes Dev.* 24 (21), 2343–2364. doi:10.1101/gad.1973010

de Almeida, S. F., and Carmo-Fonseca, M. (2012). Design Principles of Interconnections between Chromatin and Pre-mRNA Splicing. *Trends Biochemical Sciences* 37 (6), 248–253. doi:10.1016/j.tibs.2012.02.002

DiFeo, A., Martignetti, J. A., and Narla, G. (2009). The Role of KLF6 and its Splice Variants in Cancer Therapy. *Drug Resist. Updates* 12 (1-2), 1–7. doi:10.1016/j.drup.2008.11.001

Feng, H., Qin, Z., and Zhang, X. (2013). Opportunities and Methods for Studying Alternative Splicing in Cancer with RNA-Seq. *Cancer Lett.* 340 (2), 179–191. doi:10.1016/j.canlet.2012.11.010

Garon, E. B., Finn, R. S., Hamidi, H., Dering, J., Pitts, S., Kamranpour, N., et al. (2013). The HSP90 Inhibitor NVP-Auy922 Potently Inhibits Non-small Cell Lung Cancer Growth. *Mol. Cancer Ther.* 12 (6), 890–900. doi:10.1158/1535-7163.Mct-12-0998

Han, X., Tan, Q., Yang, S., Li, J., Xu, J., Hao, X., et al. (2019). Comprehensive Profiling of Gene Copy Number Alterations Predicts Patient Prognosis in Resected Stages I-III Lung Adenocarcinoma. *Front. Oncol.* 9, 556. doi:10.3389/fonc.2019.00556

Jensen, M. R., Schoepfer, J., Radimerski, T., Massey, A., Guy, C. T., Brueggen, J., et al. (2008). NVP-AUY922: a Small Molecule HSP90 Inhibitor with Potent Antitumor Activity in Preclinical Breast Cancer Models. *Breast Cancer Res.* 10 (2), R33. doi:10.1186/bcr1996

Jhunjhunwala, S., Hammer, C., and Delamarre, L. (2021). Antigen Presentation in Cancer: Insights into Tumour Immunogenicity and Immune Evasion. *Nat. Rev. Cancer* 21 (5), 298–312. doi:10.1038/s41568-021-00339-z

Kędzierska, H., and Piekiełko-Witkowska, A. (2017). Splicing Factors of SR and hnRNP Families as Regulators of Apoptosis in Cancer. *Cancer Lett.* 396, 53–65. doi:10.1016/j.canlet.2017.03.013

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 (7), 459–466. doi:10.1038/s41571-018-0004-4

Ladomery, M. (2013). Aberrant Alternative Splicing Is Another Hallmark of Cancer. *Int. J. Cel. Biol.* 2013, 1–6. doi:10.1155/2013/463786

Landrith, T., Li, B., Cass, A. A., Conner, B. R., LaDuca, H., McKenna, D. B., et al. (2020). Splicing Profile by Capture RNA-Seq Identifies Pathogenic Germline Variants in Tumor Suppressor Genes. *Npj Precis. Onc.* 4, 4. doi:10.1038/s41698-020-0109-y

Lee, D.-H., Sung, K. S., Bartlett, D. L., Kwon, Y. T., and Lee, Y. J. (2015). HSP90 Inhibitor NVP-Auy922 Enhances TRAIL-Induced Apoptosis by Suppressing the JAK2-STAT3-Mcl-1 Signal Transduction Pathway in Colorectal Cancer Cells. *Cell Signal.* 27 (2), 293–305. doi:10.1016/j.cellsig.2014.11.013

Lee, S. C.-W., and Abdel-Wahab, O. (2016). Therapeutic Targeting of Splicing in Cancer. *Nat. Med.* 22 (9), 976–986. doi:10.1038/nm.4165

Li, K., Sun, D., Gou, Q., Ke, X., Gong, Y., Zuo, Y., et al. (2018). Long Non-coding RNA Linc00460 Promotes Epithelial-Mesenchymal Transition and Cell Migration in Lung Cancer Cells. *Cancer Lett.* 420, 80–90. doi:10.1016/j.canlet.2018.01.060

Li, M., Zhang, W., Yang, X., Liu, H., Cao, L., Li, W., et al. (2019). Downregulation of HNRNPK in Human Cancer Cells Inhibits Lung Metastasis. *Anim. Models Exp. Med.* 2 (4), 291–296. doi:10.1002/ame2.12090

Ludwig, J. A., and Weinstein, J. N. (2005). Biomarkers in Cancer Staging, Prognosis and Treatment Selection. *Nat. Rev. Cancer* 5 (11), 845–856. doi:10.1038/nrc1739

Moore, M. J., Wang, Q., Kennedy, C. J., and Silver, P. A. (2010). An Alternative Splicing Network Links Cell-Cycle Control to Apoptosis. *Cell* 142 (4), 625–636. doi:10.1016/j.cell.2010.07.019

Müller, E. E., Genazzani, A. R., and Murru, S. (1978). Nomifensine: Diagnostic Test in Hyperprolactinemic States*. *J. Clin. Endocrinol. Metab.* 47 (6), 1352–1357. doi:10.1210/jcem-47-6-1352

Narla, G., DiFeo, A., Yao, S., Banno, A., Hod, E., Reeves, H. L., et al. (2005). Targeted Inhibition of the KLF6 Splice Variant, KLF6 SV1, Suppresses Prostate Cancer Cell Growth and Spread. *Cancer Res.* 65 (13), 5761–5768. doi:10.1158/0008-5472.can-05-0217

Oser, M. G., Sabet, A. H., Gao, W., Chakraborty, A. A., Schinzel, A. C., Jennings, R. B., et al. (2019). The KDM5A/RBP2 Histone Demethylase Represses NOTCH Signaling to Sustain Neuroendocrine Differentiation and Promote Small Cell Lung Cancer Tumorigenesis. *Genes Dev.* 33 (23-24), 1718–1738. doi:10.1101/gad.328336.119

Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am. J. Hum. Genet.* 102 (1), 11–26. doi:10.1016/j.ajhg.2017.11.002

Ratnadiwakara, M., Mohenska, M., and Änkö, M.-L. (2018). Splicing Factors as Regulators of miRNA Biogenesis - Links to Human Disease. *Semin. Cel Dev. Biol.* 79, 113–122. doi:10.1016/j.semcdb.2017.10.008

Ryan, M. C., Cleland, J., Kim, R., Wong, W. C., and Weinstein, J. N. (2012). SpliceSeq: a Resource for Analysis and Visualization of RNA-Seq Data on Alternative Splicing and its Functional Impacts. *Bioinformatics (Oxford, England)* 28 (18), 2385–2387. doi:10.1093/bioinformatics/bts452

Sheng, J., Zhao, Q., Zhao, J., Zhang, W., Sun, Y., Qin, P., et al. (2018). SRSF1 Modulates PTPMT1 Alternative Splicing to Regulate Lung Cancer Cell Radioresistance. *EBioMedicine* 38, 113–126. doi:10.1016/j.ebiom.2018.11.007

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654

Song, X., Zeng, Z., Wei, H., and Wang, Z. (2018). Alternative Splicing in Cancers: From Aberrant Regulation to New Therapeutics. *Semin. Cel Dev. Biol.* 75, 13–22. doi:10.1016/j.semcdb.2017.09.018

Sun, G., Zhou, H., Chen, K., Zeng, J., Zhang, Y., Yan, L., et al. (2020). HnRNP A1 - Mediated Alternative Splicing of CCDC50 Contributes to Cancer Progression of clear Cell Renal Cell Carcinoma via ZNF395. *J. Exp. Clin. Cancer Res.* 39 (1), 116. doi:10.1186/s13046-020-01606-x

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *wo* 1A (1a), 68–77. doi:10.5114/wo.2014.47136

Wan, L., Yu, W., Shen, E., Sun, W., Liu, Y., Kong, J., et al. (2019). SRSF6-regulated Alternative Splicing that Promotes Tumour Progression Offers a Therapy Target for Colorectal Cancer. *Gut* 68 (1), 118–129. doi:10.1136/gutjnl-2017-314983

Wang, J.-Q., Teng, Q.-X., Lei, Z.-N., Ji, N., Cui, Q., Fu, H., et al. (2020). Reversal of Cancer Multidrug Resistance (MDR) Mediated by ATP-Binding Cassette Transporter G2 (ABCG2) by AZ-628, a RAF Kinase Inhibitor. *Front. Cel Dev. Biol.* 8, 601400. doi:10.3389/fcell.2020.601400

Wang, J.-Z., Fu, X., Fang, Z., Liu, H., Zong, F.-Y., Zhu, H., et al. (2020). QKI-5 Regulates the Alternative Splicing of Cytoskeletal Gene ADD3 in Lung Cancer. *J. Mol. Cel. Biol.* 13, 347–360. doi:10.1093/jmcb/mjaa063

Yae, T., Tsuchihashi, K., Ishimoto, T., Motohara, T., Yoshikawa, M., Yoshida, G. J., et al. (2012). Alternative Splicing of CD44 mRNA by ESRP1 Enhances Lung Colonization of Metastatic Cancer Cell. *Nat. Commun.* 3, 883. doi:10.1038/ncomms1892

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fgene.2021.689715

Identification of Pathway-Based Biomarkers with Crosstalk Analysis for Overall Survival Risk Prediction in Breast Cancer

Xiaohua Liu[1†], *Lili Su*[2†], *Jingcong Li*[1] *and Guoping Ou*[1*]

[1]State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China, [2]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

Recently, many studies have investigated the role of gene-signature on the prognostic assessment of breast cancer (BC), however, the tumor heterogeneity and sequencing noise have limited the clinical usage of these models. Pathway-based approaches are more stable to the perturbation of certain gene expression. In this study, we constructed a prognostic classifier based on survival-related pathway crosstalk analysis. We estimated pathway's deregulation scores (PDSs) for samples collected from public databases to select survival-related pathways. After pathway crosstalk analysis, we conducted K-means clustering analysis to cluster the patients into G1 and G2 subgroups. The survival outcome of the G2 subgroup was significantly worse than the G1 subgroup. Internal and external dataset exhibits high consistency with the training dataset. Significant differences were found between G2 and G1 subgroups on pathway activity, gene mutation, immune cell infiltration levels, and in particular immune cells/pathway's activities were significantly negatively associated with BC patient's outcomes. In conclusion, we established a novel classifier reflecting the overall survival risk of BC and successfully validated its clinical usage on multiple BC datasets, which could offer clinicians inspiration in formulating the clinical treatment plan.

Keywords: breast cancer, deep-learning, pathway's deregulation scores, prognosis, classifier

OPEN ACCESS

Edited by:
Wan Zhu,
Stanford University, United States

Reviewed by:
Ning Wang,
Arcus Biosciences, United States
Salima Akter,
Kyung Hee University, South Korea

***Correspondence:**
Guoping Ou
ouguop@sysucc.org.cn

†These authors share co-first authorship

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

Received: *01 April 2021*
Accepted: *28 September 2021*
Published: *21 October 2021*

Citation:
Liu X, Su L, Li J and Ou G (2021) Identification of Pathway-Based Biomarkers with Crosstalk Analysis for Overall Survival Risk Prediction in Breast Cancer. Front. Genet. 12:689715. doi: 10.3389/fgene.2021.689715

INTRODUCTION

As a highly metastatic and invasive malignant tumor with high incidence, breast cancer (BC) seriously threatens women's health and quality of life (Veronesi et al., 2005; Siegel et al., 2019; Rüschoff et al., 2020). BC occupied a quarter of all malignant tumors, which has received numerous clinical attention worldwide (Ferlay et al., 2015). At present, the primary treatment options for BC are chemotherapy, surgery, and radiotherapy (Shi et al., 2019). However, BCs tend to exhibit drug resistance and high recurrence rates on account of heterogeneity, making the therapeutic effects and prognosis of the disease unsatisfactory (Natarajan et al., 2012). Screening biomarkers for BC has a significant effect on reducing mortality, early diagnosis, and the improvement of prognosis in BC.

With the development of RNA-Seq high-throughput sequencing technology, various gene expression profiles of BC have been accumulated. Plenty of excellent models have been constructed to decode BC, the majority of them were built based on a single gene list. For example, van de Vijver et al. (2002) established a 70-gene prognosis profile to classify 295 BC patients, which is a powerful predictor for monitoring the prognosis of young BC patients. Tekpli

et al. (2019) identified clinically relevant immune clusters by integrating 15 BC cohorts, and discovered that patients with pro-tumorigenic immune infiltration were associated with poor prognosis. PAM50 (Parker et al., 2009) gene signature is a well-known molecular subtyping signature for BC, which could classify the BC into five molecular intrinsic subtypes: Normal-like, Basal-like, HER2+, Luminal A, and Luminal B. These efforts have helped us gain a deeper understanding of BC. Nevertheless, studies have found that due to the tumor heterogeneity and sequencing noise, gene-based signatures were highly unstable and the identified biomarkers were dramatically affected by the selection of training datasets (Michiels et al., 2005; Domany, 2014). In recent years many researchers indicated pathways could be helpful to extract more stable and interpretable features for risk prediction. Efforts have been made to decode cancer at levels of predefined pathways available in biological databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016), Reactome (Fabregat et al., 2018), and Gene-Set Enrichment Analysis (GSEA) (He et al., 2018). However, most existing pathways are general rather than disease-specific, and disease progression can only be partially affected by them. For pathway pairs with many common genes, we call it crosstalk. Taking the impact of overlapping genes on the pathway activity score (PAS) quantification of the two pathways into consideration can help identify disease-related features. Although it is intuitively believed that pathways will influence each other, especially when genes are shared, the existence of this phenomenon has not been studied in PAS estimation. And few studies have explored the PAS in cancer with crosstalk accommodated among well-established pathways to identify cancer-specific sub-pathways that could be used to predict the prognosis of cancer patients. Therefore, subtyping patients based on PAS and pathway crosstalk analysis is essential to promote personalized medicine.

In this study, we constructed a novel classifier reflecting the overall survival risk of BC based on survival-related pathway crosstalk analysis. We calculated the PAS for each pathway obtained from KEGG and GO resources based on the expression matrix. And then investigated the influence of crosstalk between these selected pathways on different cohorts to select the most critical 100 sub-pathways among all cohorts. we further conducted a K-means clustering analysis to cluster the patients into G1 (moderate) and G2 (aggressive) subgroups. Internal and external dataset exhibits high consistency with the training dataset.

MATERIALS AND METHODS

Data Source

We collected BC gene expression profiles from TCGA and GEO datasets, and the dataset with less than 20 samples or without overall survival information was excluded from our selection. TCGA mRNA expression data (level 3) and clinical features were downloaded from the UCSC Xena webserver (https://xenabrowser.net/datapages), while GSE16446, GSE42568, GSE7390, GSE20711, GSE1456A, GSE1456B and GSE20685 microarray data and relevant clinical information were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/gds/). After removing normal and non-survival information samples, we finally obtained 1,090 (TCGA), 107 (GSE16446), 104 (GSE42568), 198 (GSE7390), 88 (GSE20711), 159 (GSE1456A), 159 (GSE1456B) and 327 (GSE20685) BC samples for each dataset.

Pathway Activity Score

The pathway activity score (PAS) for each dataset was calculated based on the method proposed by Bhandari et al. (Bhandari et al., 2019). We downloaded all pathways from the gene ontology (GO) database (http://geneontology.org/) and generated a new mRNA expression matrix that contains only genes that exist in it for each pathway. After that, for each gene, based on its expression level, we classified the tumors into two subgroups, the samples in the higher group were scored +1, while the others were scored −1. Finally, we averaged all gene scores in this pathway as the pathway activity score for each tumor sample. A higher PAS indicates a higher pathway activity in the sample, and otherwise, a lower score means lower activity in the sample.

Selection of Survival-Related Sub-Pathways

Based on PASs and survival information, we calculated the log-rank *p*-value for each pathway by regression analysis. The pathways were then ranked based on the log-rank *p*-value. To minimize the false positive rate, we used the common significant pathways of these three large breast cancer cohorts, instead of using any single data set. The combined rank of each pathway was determined by the sure independence screening (SIS) method. We further selected the top n pathways for pathway crosstalk analysis. The threshold n was set to 100, which is much bigger than N/log(N), where N is the sample size of each cohort.

Different pathways often share some of the same genes, which can lead to crosstalk in the prognostic associations of different pathways. Considering the influence of overlapping genes on the PAS quantification of the two pathways can help identify cancer-related features. We further identified the crosstalk among the 100 selected survival-related pathways to define sub-pathways related to survival. The crosstalk between two pathways with at least three genes in common could be classified into three types. The overlapped genes between pathway A and pathway B could be defined as $P_A \cap P_B$, while the unique genes that specifically exist in pathway A or pathway B were defined as $P_A - (P_A \cap P_B)$ and $P_B - (P_A \cap P_B)$. Based on this classifier, each pathway pair could generate three sub-pathways.

To make sure each sub-pathway contains enough genes for further analysis, we obtained sub-pathways that consist of at least three genes. The Cox-PH model was used to calculate the survival risk *p*-value based on the recalculated PAS for each sub-pathway. After Bonferroni correction, we identified critical survival-related pathways and sub-pathways (FDR *p*-value < 0.01) for each dataset, and the overlapped pathways were finally adopted for further modeling.

Model Construction and Evaluation

With the pathways generated above, we constructed a pathway activity matrix with the row names are sub-pathways and the column names are sample IDs for each dataset. We performed consensus clustering with the pathway features acquired above to classify the TCGA samples and obtained the best cluster number as 2 based on three metrics, including C-index, Brier score, and log-rank *p*-value to redefine the samples as G1 and G2 subgroups. To predict these two subgroups for other datasets, we used several machine learning methods, including SVM, Adaboost, and Gaussian, to build a prediction model and obtained the best performance based on the pathway activity matrix. The robustness of the model was evaluated in the TCGA testing dataset and several external individual GEO datasets. We further built a classification model using several machine learning methods, including SVM, Adaboost, and Gaussian, based on these labels. We used the grid search to slightly turn the hyperparameters of the classifier. In the cross-validation procedure, TCGA samples were divided into training and testing datasets in a 4:1 ratio, and the training dataset was used to perform 10-fold cross-validation. To predict the GEO dataset, we used all TCGA samples to build the classification model.

We then compared three metrics, including C-index, Brier score, and log-rank *p*-value, to evaluate the model's performance. These metrics can quantify the proportion of patient pairs in a cohort whose risk prediction is highly consistent with survival outcomes. Usually, a higher C-index indicates more precise prediction performance, and 1 means perfect prediction, while 0.5 means the prediction performance is similar to random prediction. To calculate the C-index, we built a Cox-PH model based on the clustering labels and the patient's survival information in the TCGA training dataset and predicted the survival rate in the testing dataset, which was calculated by the R "survcomp" package. Brier score reflects the mean difference between observed and predicted survival after a certain period in survival analysis, and a lower score means good performance. Log-rank *p*-value was calculated by the R "survival" package to show the survival difference between the two groups, and a lower score means a more significant survival difference.

Survival Analysis

The log-rank test compares the survival difference of two groups at each observed event time was performed by R "survival" package. Kaplan-Meier analysis was applied to obtain a survival-curve plot of BRCA subtypes. Multivariate Cox regression analysis determined the independent role of this newly established predictor. Besides, we adopted Fisher's exact test to compare the census gene mutation differences between G2 and G1 subgroups in the TCGA cohort. We also compared the distributions of G2 and G1 in different clinicopathological features, such as tumor stage, new tumor event, and sex, by using Fisher exact tests.

We used the "DESeq2" R package (Love et al., 2014) to real the differential expressed genes between G2 and G1 subgroup; the significant DEGs were identified as $|\text{LogFoldChange}| > 1$ and false discovery rate (FDR) < 0.05.

Gene Set Enrichment Analysis

GSEA analysis was used to compare the pathway activity difference between G2 and G1, in which the R "ClusterProfier" package was performed. We adopted Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways to perform GSEA analysis and the top 20 significant pathways were displayed.

Mutation Analysis

The R "maftools" package was utilized to analyze and visualize the mutation data. The mutation data were compared between one group and the other groups using the chi-square test. A *p*-value of less than 0.05 was considered significant.

RESULTS

Identification of Overall Survival Risk Subtypes in BC

We obtained seven BC datasets (TCGA, GSE1456A, GSE1456B, GSE7390, GSE16446, GSE20685, GSE20711, and GSE42568) gene expression profiles and available clinical survival information from the TCGA and GEO databases. After calculating the PAS for each pathway obtained from KEGG and GO resources and selecting the survival-related pathways, we investigated the influence of crosstalk between these selected pathways on different cohorts, and then the most critical 100 sub-pathways among all cohorts were identified (**Supplementary Table S1**). K-means clustering analysis was used to divide the TCGA patients into two subgroups, defined as group 1 (G1, moderate) and group 2 (G2, aggressive) (**Supplementary Figure S1A**, and **Supplementary Table S2**). Notably, patients from the G2 subgroup show significantly worse clinical outcomes (overall survival and relapse-free survival) compared to the G1 subgroup (**Supplementary Figures S1B,C**; $p = 0.0053$ and $p = 0.0031$, respectively; log-rank test). We further built a classifier based on the TCGA training dataset with the label defined by k-means clustering analysis (*Materials and Methods*).

Evaluation of the Performance

To evaluate the robustness of OS risk prediction of the model, we tested the model performance on the TCGA testing dataset and several external GEO datasets, including GSE1456A, GSE7390, GSE16446, GSE20685, GSE20711, and GSE42568. As shown in **Figure 1** and **Table 1**, the model was stable and exhibited excellent classification capability, indicated by C-index and log-rank *p*-values between G2 and G1. The TCGA test cohort generated a high C-index (0.661), low Brier score (0.179), and significant average log-rank *p*-value ($p = 0.00123$) on survival difference. In different datasets, our classifier can significantly divide the samples into a good prognostic group and a poor prognostic group, which suggested that the newly developed classifier is able to universally predict the overall survival outcomes for patients with BC.

In order to compare the risk prediction capabilities of our predictor with some other clinical information, we performed univariate cox regression analysis for each clinical information

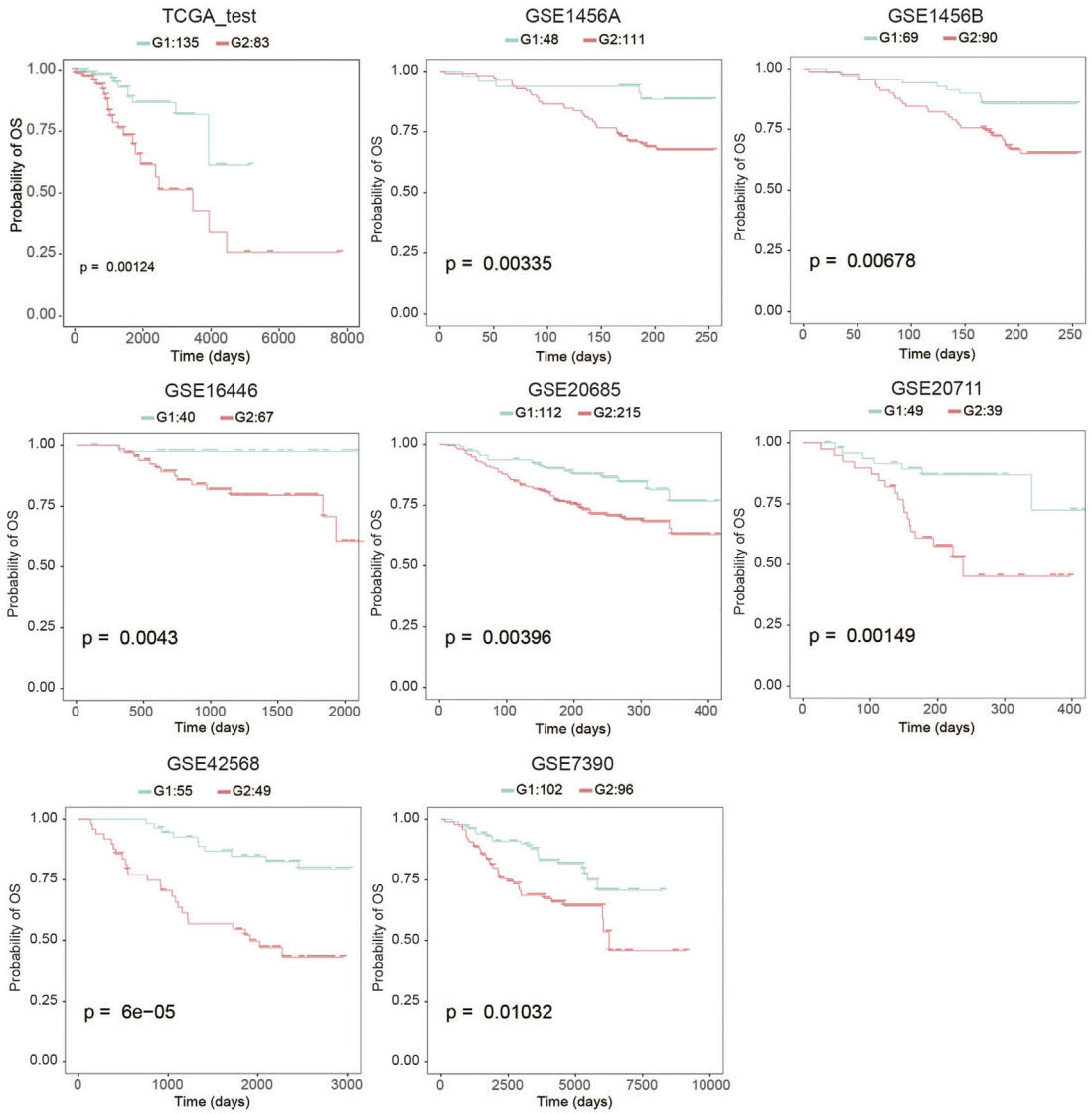


FIGURE 1 | Kaplan-Meier analysis to determine the survival differences between group 2 (G2) and group 1 (G1).

TABLE 1 | Cross-validation based performance robustness of classifier on TCGA training and testing cohorts.

Cohorts	Omics type	Samples	C-index	Brier score	Log-rank *p*
TCGA_test	RNA-Seq	218	0.661	0.179	1.23E-03
GSE1456A	Microarray	159	0.600	0.122	3.35E-03
GSE1456B	Microarray	159	0.599	0.121	6.78E-03
GSE16446	Microarray	107	0.646	0.112	4.30E-03
GSE20685	Microarray	327	0.579	0.153	3.96E-03
GSE20711	Microarray	88	0.656	0.189	1.49E-03
GSE42568	Microarray	104	0.679	0.195	6.27E-05
GSE7390	Microarray	198	0.597	0.180	1.03E-02

TCGA, The Cancer Genome Atlas.

(including age, tumor stage, and PAM50 subtyping) in the TCGA dataset as well as the external validation datasets. As shown in **Figure 2**, our classifier has a more general prognostic ability than other clinical information ($p < 0.05$ in all datasets). We further introduced several published transcriptomic-based predictors as previous study (Lee et al., 2021), including the proliferation index

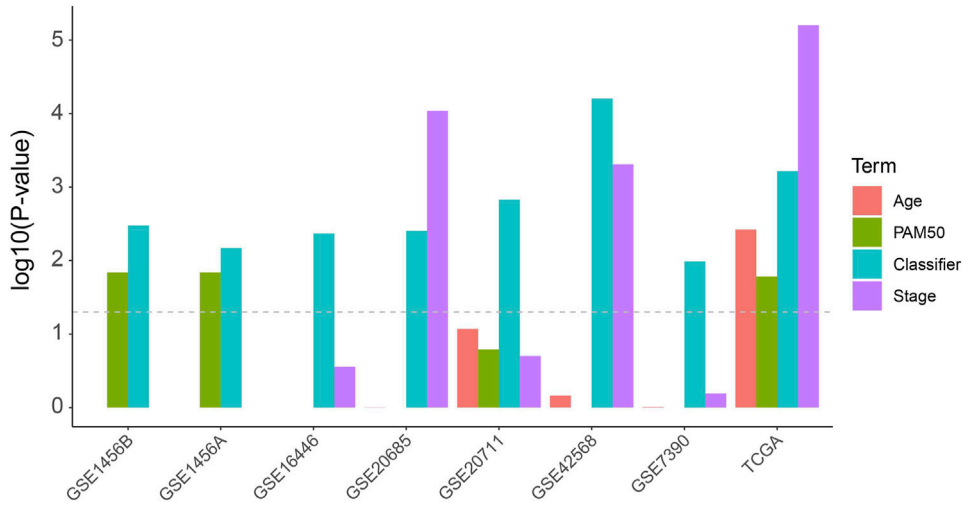


FIGURE 2 | Univariate Cox analysis of the classifier as well as regular clinical classification (Age, PAM50 and tumor stage) in TCGA and other extermal validation cohorts.

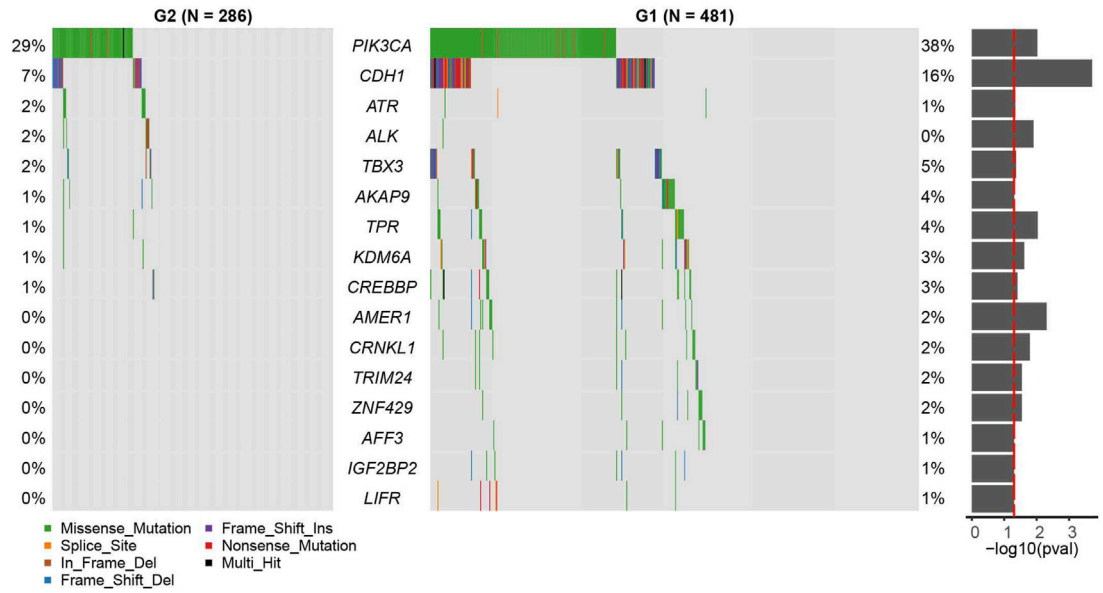


FIGURE 3 | Census mutation landscape between group 2 (G2) and group 1 (G1). Only cancer census genes in the COSMIC database are shown in the plot. The significance of the difference in gene mutation frequency between the two groups is shown in the barplot on the right (fisher's exact test).

(Whitfield et al., 2006), interferon-γ (IFNγ) signature score (Ayers et al., 2017) as well as cytolytic activity score (Rooney et al., 2015), and performed a multivariate Cox regression analysis with age, tumor stage, and our classifier (**Supplementary Figure S2**). In this analysis, the proliferation index and the IFNγ signature score were estimated as ssGSEA score (Yi et al., 2020a) of each gene signature, respectively, and the cytolytic activity score was calculated as the mean expression level of *GZMA* and *PRF1* (Rooney et al., 2015). As shown in **Supplementary Figure S2**, the proliferation index and IFNy signature score show higher predictive power (hizard ratios were 2.13 and 0.27, respectively), but also have larger confidence intervals and *p*-values, which suggesting that they cannot be used as independent prognostic factors of BRCA. Reassuringly, our classifier had a more stable hizard ratioa near statistically significant *p*-value. In addition, we also test the risk prediction performance in different subgroups of age and tumor stage (**Supplementary Figure S3**). This result suggests

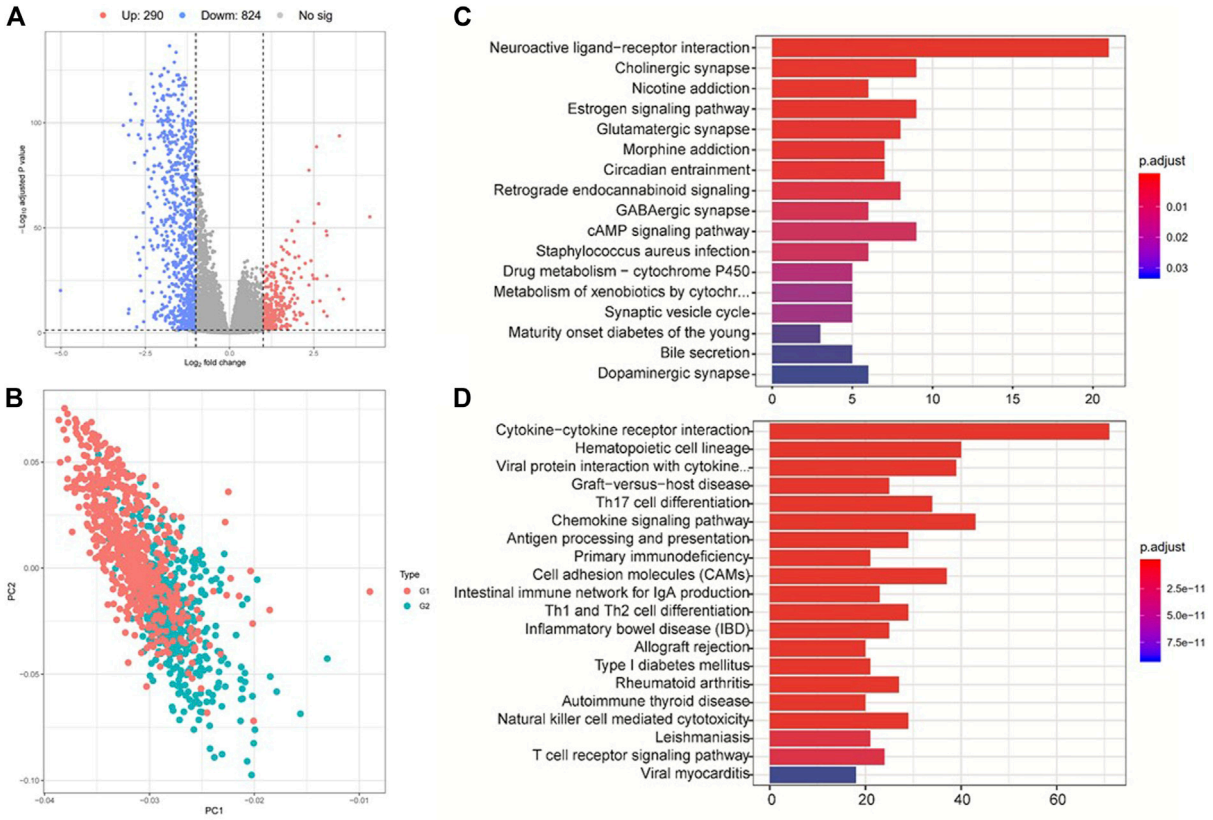


FIGURE 4 | Analysis of differentially expressed genes and their corresponding pathways. **(A)** Gene expression differences between group 2 (G2) and group 1 (G1). **(B)** PCA analysis shows the clustering results of group 2 (G2) and group 1 (G1). **(C)** and **(D)** KEGG pathway enrichment analysis for the up- (up panel) and down-regulated (down panel) genes between G2 and G1 subgroups.

that our classifier can be used essentially for the typing of different clinical subgroups, although in the low-age group and low-level group of TCGA the *p*-values did not reach significance.

Association Between Different Survival Subtypes and Genomic Feature

We found that the mutation rates of *PI3KCA* and *CDH1* were significantly higher in the G1 group than G2 group (*PI3KCA*: OR = 0.655, 95%CI: 0.471–0.907, $p = 0.00951$; *CDH1*: OR = 0.389, 95%CI: 0.220–0.660, $p = 0.000195$, **Figure 3**, and **Supplementary Table S3**, Fisher's exact test). Other significant differentially mutated genes between the two groups, including *ATR*, *ALK*, *TBX3*, *AKAP9*, *TPR*, *KDM6A*, *CREBBP*, *AMER1*, *CRNKL1*, *TRIM24*, *ZNF429*, *AFF3*, *IGF2BP2*, and *LIFR* (**Supplementary Table S3**, all $p < 0.05$). No significant tumor mutation burden (TMB) level difference was found between G1 and G2 subgroups (**Supplementary Figure S4**). *PI3KCA* and *CDH1* are two frequently mutated genes in many cancers, including breast cancer, gastric cancer, colorectal carcinoma, and head and neck squamous cell carcinoma (Hansford et al., 2015; Millis et al., 2016; Zhang et al., 2017; An et al., 2018). However, the association of PIK3CA mutation and prognosis has not been clarified clearly, PI3KCA mutation can be associated with a better prognosis (Barbareschi et al., 2007; Maruyama et al., 2007; Pérez-Tenorio et al., 2007; Kalinsky et al., 2009) or a worse prognosis (Li et al., 2006; Lerma et al., 2008). In some studies, PIK3CA mutation even has no obvious relationship with the prognosis (Saal et al., 2005; Lai et al., 2008; Stemke-Hale et al., 2008; Michelucci et al., 2009; Loi et al., 2010; Boyault et al., 2012). A similar phenomenon was found for CDH1 mutation as well (Corso et al., 2018).

We then performed differential expression analysis between G2 and G1 subgroups, and identified 290 upregulated and 824 downregulated genes ($|\log_2 \text{fold change}| > 1$ and FDR *p*-value > 0.05) (**Figures 4A,B**, and **Supplementary Table S4**) based on the TCGA cohort. KEGG pathway analysis indicated that these upregulated genes were mostly enriched in neuroactive ligand–receptor interaction, cholinergic synapse and estrogen signaling pathways (**Figure 4C**). The downregulated genes were mostly enriched in immune-related pathways, such as cytokine–cytokine receptor interaction, hematopoietic cell lineage, graft–versus–host disease, and Th17 cell differentiation (**Figure 4D**). These results prompted that the G1 subgroup might be immune activated subtype, which could be associated with its better overall survival. We also tested the correlations between the two survival subtypes (G2 and G1) and

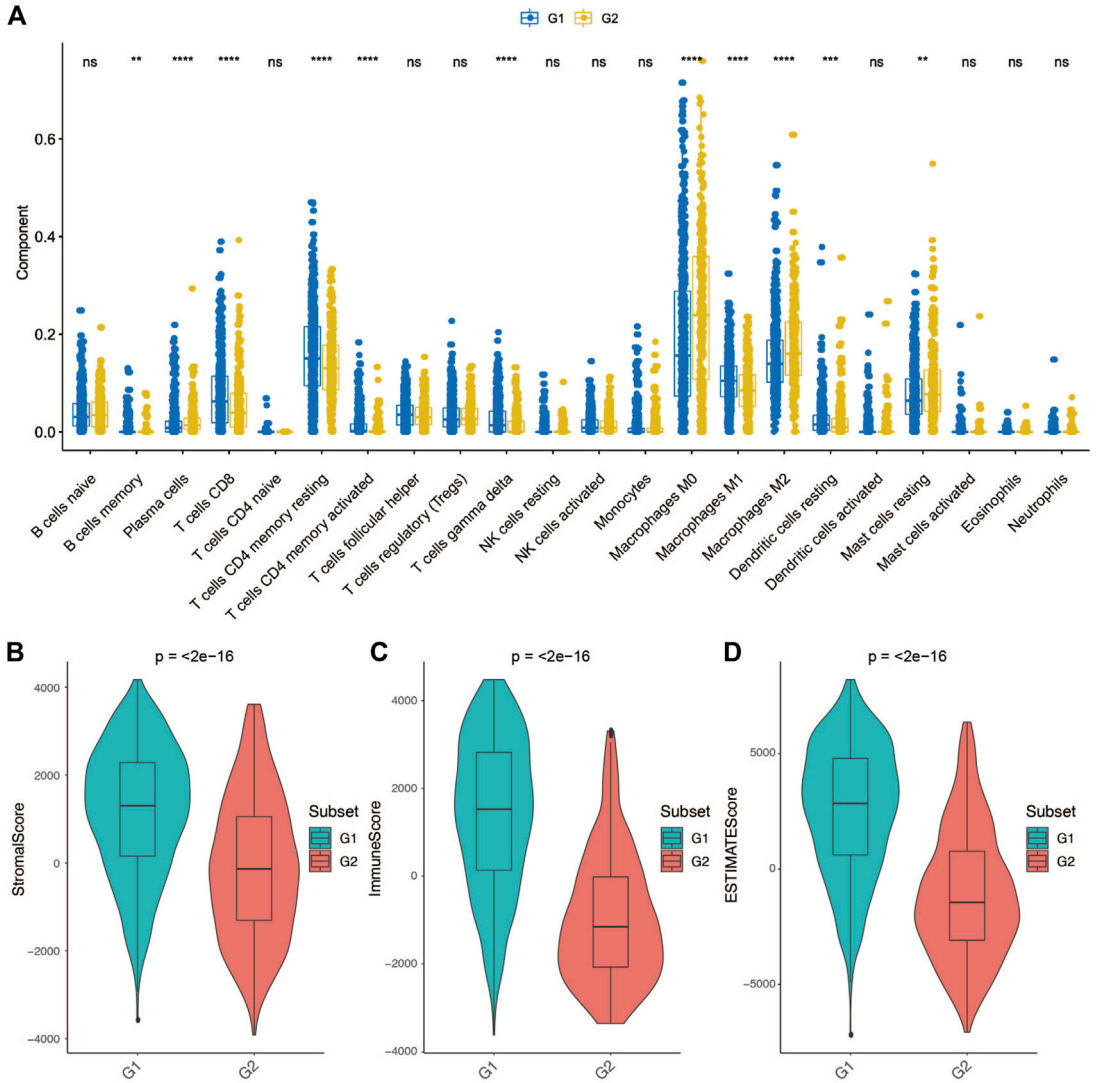


FIGURE 5 | Tumor microenvironment differences between group 2 (G2) and group 1 (G1) subgroups. **(A)** Comparison of each immune cell component between G1 and G2. **(B)** Comparison of stroma score between G1 and G2. **(C)** Comparison of immune score between G1 and G2. **(D)** Comparison of ESTIMATE score between G1 and G2.

clinicopathological characteristics from the TCGA cohort and found that no significant differences were revealed in age, sex, tumor stage, metastasis coded, estrogen receptor status, progesterone receptor status, and histological type subgroups, instead of PAM50 subtype (**Supplementary Figure S5**, **Supplementary Table S5**).

We then performed GSEA analysis to compare the G2 and G1 subgroups, aiming to identify critical pathways that displayed different activities between the G1 and G2 subgroups (**Supplementary Figure S6**, **Supplementary Tables S6–S8**). Hallmark pathway enrichment analysis showed that immune-related pathways including the inflammatory response, allograft rejection, interferon-gamma response and TNFA-signaling *via* NFκB were enriched in the G1 subgroup, while metabolic-related pathways such as oxidative phosphorylation signaling were activated in the G2 subgroup (**Supplementary Table S6**). Pathway enrichment analysis indicated that the differences between these two groups were concentrated in the KEGG pathways of "Graft *vs.* host disease", "primary immunodeficiency", and "allograft rejection" (**Supplementary Table S7**) and Reactome pathways related to co-stimulation by the CD28 family, generation of second messenger molecules, and cytokine signaling in the immune system (**Supplementary Table S8**). Previous studies have proved that metabolic pathway activities like oxidative phosphorylation signaling were negatively correlated with immune infiltration

and contributed to a worse prognosis in TNBC (Gong et al., 2021), which is consistent with our results.

Comparison of Tumor Microenvironment Between G2 and G1

We further employed the CIBERSORT algorithm to investigate the distributions of infiltrated immune cells between the G2 and G1 subgroups (**Supplementary Figure S7**). The result revealed that significant differences were obtained between two groups in CD8+T cells, CD4+T memory cells (resting), CD4+T memory (activated), γδ T cells, Macrophages M0, Macrophages M1, Macrophages M2, Dendritic cells (resting), and Mast cells (resting) (**Figure 5A**). Among macrophages, Macrophages M1 accounts for a higher proportion of the G1 subgroup, while the G2 subgroup consists of a higher proportion of Macrophages M2. Macrophages M2 was found to be dominant in BC and associated with poor clinical outcomes of BC (Bao et al., 2021), which could be the reason that the G2 subgroup patients have a worse overall survival.

Considering that the tumor tissue has tumor cells, stromal cells and immune cells, we measured stromal score and immune score based on specific gene expression signature to represent the level of immune infiltration and stroma infiltration of each tumor following the previous reported method-ESTIMATE (Yoshihara et al., 2013). Also, an ESTIMATE score also calculated which reflects the overall level of both immune infiltration and stromal infiltration. As shown in **Figures 5B–D**, G1 presented a higher stromal score, immune score and ESTIMATE score compared with G2. These results consistent with the previous definition that the G1 subgroup might be immune activated subtype, there was abundant crosstalk in the tumor microenvironment of this type of tumor, which could benefit from immunotherapy.

DISCUSSION

In the era of personalized medicine, there is an urgent need for a molecular marker-based approach to predict the prognostic outcomes of cancer patients accurately. Previous studies have reported many gene-based signatures to subtype BC (van de Vijver et al., 2002; Pu et al., 2020) (Tekpli et al., 2019) (Parker et al., 2009). Here, we constructed an overall survival risk model to classify samples into two subgroups. Internal and external datasets validation exhibits high consistency with the training dataset. Significant differences were found between the G2 and G1 subgroups including pathway activity, gene mutation, immune cell infiltration levels. In particular, immune cells/pathway's activities were significantly negatively associated with BC patient's outcomes.

In order to test whether our classifier is applicable to all ages and tumor grades, we performed prognostic association analysis for different clinical subgroups. For a data set with sufficient samples (more than 20 samples for each subgroup), our classifier can basically distinguish patients with different overall survival periods, although the high-age group and the low-stage group of TCGA have not reached statistical significance. Although the *p*-value of the high-age group of TCGA does not reach statistical significance (0.076), a clear trend can still be seen. However, it is challenging to explain why our classifier is unable to distinguish OS in the low-stage samples of TCGA with prognostic significance, though it performed well in the other two verification sets (**Supplementary Figures S3M,O**).

We found a significant mutation rate difference of PI3KCA and CDH1 gene between the G1 and G2 subgroups. It is not yet clear whether PIK3CA mutation is associated with clinical outcome, PI3KCA mutation can be associated with a better prognosis (Barbareschi et al., 2007; Maruyama et al., 2007; Pérez-Tenorio et al., 2007; Kalinsky et al., 2009) or a worse prognosis (Li et al., 2006; Lerma et al., 2008). In some studies, PIK3CA mutation even has no obvious relationship with the prognosis (Saal et al., 2005; Lai et al., 2008; Stemke-Hale et al., 2008; Michelucci et al., 2009; Loi et al., 2010; Boyault et al., 2012). A similar phenomenon was found for CDH1 mutation as well (Corso et al., 2018). Our results suggested that mutations of PI3KCA are positively associated with a favorable prognosis, but future studies are needed to investigate the potential mechanisms.

We also found that the G1 subgroup displayed significant higher level of immune infiltration, stromal infiltration level than the G2 subgroup. As reported, Th1 and cytotoxic types of memory T cells and CD8+ T cells can predict better prognosis in diverse cancers (Wei et al., 2018; Yi et al., 2020b; St. Paul and Ohashi, 2020). Several studies showed that the existence of mature antigen-presenting dendritic cells (DCs) could infiltrate colon cancer and theoretically increase immune response, which are correlated with improved survival as well (Schwaab et al., 2001). Other immune cells such as macrophages always produce plenty of factors influencing tumor cell's survival and growth, chemotaxis, cell invasion, angiogenesis, or repress T cell responses (Pagès et al., 2010). Therefore, a high rate of tumor-associated macrophages typically serves as a poor prognostic factor. It is valuable to predict the efficacy of specific therapies, especially immunotherapy. For example, Peng et al. recently developed a computational method named TIDE to accurately predict immunotherapy outcomes of melanoma (Jiang et al., 2018). The level of immune infiltration was significantly associated with the efficacy of immunotherapy (Galon and Bruni, 2019). The significant immunological differences between G1 and G2 suggest that our classifier may be predictive of immunotherapy efficacy. We will collect relevant data resources for more in-depth study in our future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA and Gene Expression Omnibus (GSE1456A, GSE16446, GSE20685, GSE20711, GSE42568, and GSE7390).

AUTHOR CONTRIBUTIONS

XL and GO designed this experiment. XL and LS performed the research and analyzed the data. XL and JL wrote the manuscript, and XL, LS, JL, and GO revised the manuscript. All these authors 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/fgene.2021.689715/full#supplementary-material

Supplementary Figure S1 | The establishment and preliminary validation of the classifier. **(A)** Heatmap shows the clustering of two subclassification groups. **(B)** Kaplan-Meier plot shows the OS difference between group 2 (G2) and group 1(G1). **(C)** Kaplan-Meier plot shows the RFS difference between group 2 (G2) and group 1(G1).

Supplementary Figure S2 | Multivariate Cox regression analysis of prognostic factors for overall survival in TCGA patients. Age, tumor stage, proliferation score, IFNg score, cytolytic activity as well as our classifier were taken into consideration in this analysis.

Supplementary Figure S3 | KM curves show the prognostic performance of the classifier in different age level and tumor stage level of all included cohorts. Only cohorts with available clinical information are shown in the plot.

Supplementary Figure S4 | Comparison of tumor mutation burden between the G1 and G2 subgroups.

Supplementary Figure S5 | Sankey plot shows the distribution of group 2 (G2) and group 1 (G1) in different clinicopathological and genetic subgroups.

Supplementary Figure S6 | Gene Set Enrichment Analysis (GSEA) analysis shows the differences between group 2 (G2) and group 1 (G1) at Hallmark, (Kyoto Encyclopedia of Genes and Genomes) KEGG, and Reactome pathway aspects.

Supplementary Figure S7 | CIBERSORT analysis shows the distributions of 22 tumor infiltrated immune cells in each breast cancer sample.

REFERENCES

An, Y., Adams, J. R., Hollern, D. P., Zhao, A., Chang, S. G., Gams, M. S., et al. (2018). Cdh1 and Pik3ca Mutations Cooperate to Induce Immune-Related Invasive Lobular Carcinoma of the Breast. *Cel Rep.* 25 (3), 702–714. doi:10.1016/j.celrep.2018.09.056

Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D. R., et al. (2017). IFN-γ-related mRNA Profile Predicts Clinical Response to PD-1 Blockade. *J. Clin. Invest.* 127 (8), 2930–2940. doi:10.1172/jci91190

Bao, X., Shi, R., Zhao, T., Wang, Y., Anastasov, N., Rosemann, M., et al. (2021). Integrated Analysis of Single-Cell RNA-Seq and Bulk RNA-Seq Unravels Tumour Heterogeneity Plus M2-like Tumour-Associated Macrophage Infiltration and Aggressiveness in TNBC. *Cancer Immunol. Immunother.* 70 (1), 189–202. doi:10.1007/s00262-020-02669-7

Barbareschi, M., Buttitta, F., Felicioni, L., Cotrupi, S., Barassi, F., Del Grammastro, M., et al. (2007). Different Prognostic Roles of Mutations in the Helical and Kinase Domains of the PIK3CA Gene in Breast Carcinomas. *Clin. Cancer Res.* 13 (20), 6064–6069. doi:10.1158/1078-0432.Ccr-07-0266

Bhandari, V., Hoey, C., Liu, L. Y., Lalonde, E., Ray, J., Livingstone, J., et al. (2019). Molecular Landmarks of Tumor Hypoxia across Cancer Types. *Nat. Genet.* 51 (2), 308–318. doi:10.1038/s41588-018-0318-2

Boyault, S., Drouet, Y., Navarro, C., Bachelot, T., Lasset, C., Treilleux, I., et al. (2012). Mutational Characterization of Individual Breast Tumors: TP53 and PI3K Pathway Genes Are Frequently and Distinctively Mutated in Different Subtypes. *Breast Cancer Res. Treat.* 132 (1), 29–39. doi:10.1007/s10549-011-1518-y

Corso, G., Veronesi, P., Sacchini, V., and Galimberti, V. (2018). Prognosis and Outcome in CDH1-Mutant Lobular Breast Cancer. *Eur. J. Cancer Prev.* 27 (3), 237–238. doi:10.1097/cej.0000000000000405

Domany, E. (2014). Using High-Throughput Transcriptomic Data for Prognosis: a Critical Overview and Perspectives. *Cancer Res.* 74 (17), 4612–4621. doi:10.1158/0008-5472.Can-13-3338

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46 (D1), D649–d655. doi:10.1093/nar/gkx1132

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *Int. J. Cancer* 136 (5), E359–E386. doi:10.1002/ijc.29210

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

Gong, Y., Ji, P., Yang, Y.-S., Xie, S., Yu, T.-J., Xiao, Y., et al. (2021). Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. *Cel Metab.* 33 (1), 51–64. doi:10.1016/j.cmet.2020.10.012

Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., et al. (2015). Hereditary Diffuse Gastric Cancer Syndrome. *JAMA Oncol.* 1 (1), 23–32. doi:10.1001/jamaoncol.2014.168

He, W., Chen, L., Yuan, K., Zhou, Q., Peng, L., and Han, Y. (2018). Gene Set Enrichment Analysis and Meta-Analysis to Identify Six Key Genes Regulating and Controlling the Prognosis of Esophageal Squamous Cell Carcinoma. *J. Thorac. Dis.* 10 (10), 5714–5726. doi:10.21037/jtd.2018.09.55

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

Kalinsky, K., Jacks, L. M., Heguy, A., Patil, S., Drobnjak, M., Bhanot, U. K., et al. (2009). PIK3CA Mutation Associates with Improved Outcome in Breast Cancer. *Clin. Cancer Res.* 15 (16), 5049–5059. doi:10.1158/1078-0432.Ccr-09-0632

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* 44 (D1), D457–D462. doi:10.1093/nar/gkv1070

Lai, Y.-L., Mau, B.-L., Cheng, W.-H., Chen, H.-M., Chiu, H.-H., and Tzen, C.-Y. (2008). PIK3CA Exon 20 Mutation Is Independently Associated with a Poor Prognosis in Breast Cancer Patients. *Ann. Surg. Oncol.* 15 (4), 1064–1069. doi:10.1245/s10434-007-9751-7

Lee, J. S., Nair, N. U., Dinstag, G., Chapman, L., Chung, Y., Wang, K., et al. (2021). Synthetic Lethality-Mediated Precision Oncology via the Tumor Transcriptome. *Cell* 184 (9), 2487–2502. doi:10.1016/j.cell.2021.03.030

Lerma, E., Catasus, L., Gallardo, A., Peiro, G., Alonso, C., Aranda, I., et al. (2008). Exon 20 PIK3CA Mutations Decreases Survival in Aggressive (HER-2 Positive) Breast Carcinomas. *Virchows Arch.* 453 (2), 133–139. doi:10.1007/s00428-008-0643-4

Li, S. Y., Rong, M., Grieu, F., and Iacopetta, B. (2006). PIK3CA Mutations in Breast Cancer Are Associated with Poor Outcome. *Breast Cancer Res. Treat.* 96 (1), 91–95. doi:10.1007/s10549-005-9048-0

Loi, S., Haibe-Kains, B., Majjaj, S., Lallemand, F., Durbecq, V., Larsimont, D., et al. (2010). PIK3CA Mutations Associated with Gene Signature of Low mTORC1 Signaling and Better Outcomes in Estrogen Receptor-Positive Breast Cancer. *Proc. Natl. Acad. Sci.* 107 (22), 10208–10213. doi:10.1073/pnas.0907011107

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Maruyama, N., Miyoshi, Y., Taguchi, T., Tamaki, Y., Monden, M., and Noguchi, S. (2007). Clinicopathologic Analysis of Breast Cancers with PIK3CA Mutations in Japanese Women. *Clin. Cancer Res.* 13 (2 Pt 1), 408–414. doi:10.1158/1078-0432.Ccr-06-0267

Michelucci, A., Di Cristofano, C., Lami, A., Collecchi, P., Caligo, A., Decarli, N., et al. (2009). PIK3CA in Breast Carcinoma. *Diagn. Mol. Pathol.* 18 (4), 200–205. doi:10.1097/PDM.0b013e31818e5fa4

Michiels, S., Koscielny, S., and Hill, C. (2005). Prediction of Cancer Outcome with Microarrays: a Multiple Random Validation Strategy. *The Lancet* 365 (9458), 488–492. doi:10.1016/s0140-6736(05)17866-0

Millis, S. Z., Ikeda, S., Reddy, S., Gatalica, Z., and Kurzrock, R. (2016). Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations across 19 784 Diverse Solid Tumors. *JAMA Oncol.* 2 (12), 1565–1573. doi:10.1001/jamaoncol.2016.0891

Natarajan, K., Xie, Y., Baer, M. R., and Ross, D. D. (2012). Role of Breast Cancer Resistance Protein (BCRP/ABCG2) in Cancer Drug Resistance. *Biochem. Pharmacol.* 83 (8), 1084–1103. doi:10.1016/j.bcp.2012.01.002

Pagès, F., Galon, J., Dieu-Nosjean, M.-C., Tartour, E., Sautès-Fridman, C., and Fridman, W.-H. (2010). Immune Infiltration in Human Tumors: a Prognostic

Factor that Should Not Be Ignored. *Oncogene* 29 (8), 1093–1102. doi:10.1038/onc.2009.416

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Jco* 27 (8), 1160–1167. doi:10.1200/jco.2008.18.1370

Pérez-Tenorio, G., Alkhori, L., Olsson, B., Waltersson, M. A., Nordenskjöld, B., Rutqvist, L. E., et al. (2007). PIK3CA Mutations and PTEN Loss Correlate with Similar Prognostic Factors and Are Not Mutually Exclusive in Breast Cancer. *Clin. Cancer Res.* 13 (12), 3577–3584. doi:10.1158/1078-0432.Ccr-06-1609

Pu, M., Messer, K., Davies, S. R., Vickery, T. L., Pittman, E., Parker, B. A., et al. (2020). Research-based PAM50 Signature and Long-Term Breast Cancer Survival. *Breast Cancer Res. Treat.* 179 (1), 197–206. doi:10.1007/s10549-019-05446-y

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

Rüschoff, J., Lebeau, A., Sinn, P., Schildhaus, H.-U., Decker, T., Ammann, J., et al. (2020). Statistical Modelling of HER2-Positivity in Breast Cancer: Final Analyses from Two Large, Multicentre, Non-interventional Studies in Germany. *The Breast* 49, 246–253. doi:10.1016/j.breast.2019.12.005

Saal, L. H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., et al. (2005). PIK3CA Mutations Correlate with Hormone Receptors, Node Metastasis, and ERBB2, and Are Mutually Exclusive with PTEN Loss in Human Breast Carcinoma. *Cancer Res.* 65 (7), 2554–2559. doi:10.1158/0008-5472-can-04-3913

Schwaab, T., Weiss, J. E., Schned, A. R., and Barth, R. J., Jr. (2001). Dendritic Cell Infiltration in colon Cancer. *J. Immunother.* 24 (2), 130–137. doi:10.1097/00002371-200103000-00007

Shi, W., Luo, Y., Zhao, D., Huang, H., and Pang, W. (2019). Evaluation of the Benefit of Post-mastectomy R-adiotherapy in P-atients with E-arly-stage B-reast C-ancer: A P-ropensity S-core M-atching S-tudy. *Oncol. Lett.* 17 (6), 4851–4858. doi:10.3892/ol.2019.10197

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

Stemke-Hale, K., Gonzalez-Angulo, A. M., Lluch, A., Neve, R. M., Kuo, W.-L., Davies, M., et al. (2008). An Integrative Genomic and Proteomic Analysis of PIK3CA, PTEN, and AKT Mutations in Breast Cancer. *Cancer Res.* 68 (15), 6084–6091. doi:10.1158/0008-5472.Can-07-6854

St. Paul, M., and Ohashi, P. S. (2020). The Roles of CD8+ T Cell Subsets in Antitumor Immunity. *Trends Cel Biol.* 30 (9), 695–704. doi:10.1016/j.tcb.2020.06.003

Tekpli, X., Lien, T., Lien, T., Røssevold, A. H., Nebdal, D., Borgen, E., et al. (2019). An Independent Poor-Prognosis Subtype of Breast Cancer Defined by a Distinct Tumor Immune Microenvironment. *Nat. Commun.* 10 (1), 5499. doi:10.1038/s41467-019-13329-5

van de Vijver, M. J., He, Y. D., van 't Veer, L. J., Dai, H., Hart, A. A. M., Voskuil, D. W., et al. (2002). A Gene-Expression Signature as a Predictor of Survival in Breast Cancer. *N. Engl. J. Med.* 347 (25), 1999–2009. doi:10.1056/NEJMoa021967

Veronesi, U., Boyle, P., Goldhirsch, A., Orecchia, R., and Viale, G. (2005). Breast Cancer. *The Lancet* 365 (9472), 1727–1741. doi:10.1016/S0140-6736(05)66546-4

Wei, M., Shen, D., Mulmi Shrestha, S., Liu, J., Zhang, J., and Yin, Y. (2018). The Progress of T Cell Immunity Related to Prognosis in Gastric Cancer. *Biomed. Res. Int.* 2018, 1–6. doi:10.1155/2018/3201940

Whitfield, M. L., George, L. K., Grant, G. D., and Perou, C. M. (2006). Common Markers of Proliferation. *Nat. Rev. Cancer* 6 (2), 99–106. doi:10.1038/nrc1802

Yi, M., Nissley, D. V., McCormick, F., and Stephens, R. M. (2020). ssGSEA Score-Based Ras Dependency Indexes Derived from Gene Expression Data Reveal Potential Ras Addiction Mechanisms with Possible Clinical Implications. *Sci. Rep.* 10 (1), 10258. doi:10.1038/s41598-020-66986-8

Yi, S., Zhang, Y., Xiong, W., Chen, W., Hou, Z., Yang, Y., et al. (2020). Prominent Immune Signatures of T Cells Are Specifically Associated with Indolent B-cell Lymphoproliferative Disorders and Predict Prognosis. *Clin. Transl Immunol.* 9 (1), e01105. doi:10.1002/cti2.1105

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, Y., Kwok-Shing Ng, P., Kucherlapati, M., Chen, F., Liu, Y., Tsang, Y. H., et al. (2017). A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell* 31 (6), 820–832. doi:10.1016/j.ccell.2017.04.013

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 November 2021
doi: 10.3389/fgene.2021.689676

Molecular Subclassification Based on Crosstalk Analysis Improves Prediction of Prognosis in Colorectal Cancer

Xiaohua Liu[1†], *Lili Su*[2†], *Jingcong Li*[1] *and Guoping Ou*[1*]

[1]State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China, [2]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

The poor performance of single-gene lists for prognostic predictions in independent cohorts has limited their clinical use. Here, we employed a pathway-based approach using embedded biological features to identify reproducible prognostic markers as an alternative. We used pathway activity score, sure independence screening, and K-means clustering analyses to identify and cluster colorectal cancer patients into two distinct subgroups, G2 (aggressive) and G1 (moderate). The differences between these two groups with respect to survival, somatic mutation, pathway activity, and tumor-infiltration by immunocytes were compared. These comparisons revealed that the survival rates in the G2 subgroup were significantly reduced compared to that in the G1 subgroup; further, the mutational burden rates in several oncogenes, including *KRAS*, *DCLK1*, and *EPHA5*, were significantly higher in the G2 subgroup than in the G1 subgroup. The enhanced activity of the critical pathways such as MYC and epithelial-mesenchymal transition may also lead to the progression of colorectal cancer. Taken together, we established a novel prognostic classification system that offers meritorious insights into the hallmarks of colorectal cancer.

Keywords: colorectal cancer, pathways deregulation score, overall survival, signature, personalized medicine

OPEN ACCESS

Edited by:
Liuyang Wang,
Duke University, United States

Reviewed by:
Deli Liu,
Weill Cornell Medicine, United States
Anqi Zhu,
23andMe, United States

***Correspondence:**
Guoping Ou
ouguop@sysucc.org.cn

†These authors share first authorship

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

Received: *01 April 2021*
Accepted: *23 September 2021*
Published: *04 November 2021*

Citation:
Liu X, Su L, Li J and Ou G (2021) Molecular Subclassification Based on Crosstalk Analysis Improves Prediction of Prognosis in Colorectal Cancer. Front. Genet. 12:689676. doi: 10.3389/fgene.2021.689676

INTRODUCTION

Colorectal cancer is the third most common form of cancer and the second leading cause of cancer-related deaths worldwide (Bray et al., 2018). Currently, the TNM staging system is widely used to predict the prognosis of colorectal cancer patients; however, even patients within the same TNM stage often present with distinct prognosis outcomes in clinical practice. For instance, patients with stage I and II colorectal cancer generally exhibit a favorable prognosis and are treated with surgical resection alone. Unfortunately, approximately 10–30% of stage I and II colorectal cancer patients experience tumor recurrence within 5 years of curative surgery and require more intense treatment, such as adjuvant chemotherapy (Lin et al., 2017; Guraya, 2019). Therefore, more precise prognostic tools for colorectal cancer will enable individualized therapy and improve patient prognosis.

Bioinformatic subtyping methods are generally based on gene expression data. There are two examples of these systems for colorectal cancer, where subtypes are established based on their molecular features, with these subtypes often demonstrating significant differences in clinical outcomes (Luo et al., 2020; Yang et al., 2020). Recently, several research groups have indicated that pathway analysis may be helpful in extracting more stable and interpretable features for risk prediction (Alcaraz et al., 2017; Sheng et al., 2019; Su et al., 2021). Several efforts have been made to

decode cancer at the gene, protein, and metabolite levels, and with the help of predefined pathways available from various biological databases, including Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016), Reactome (Fabregat et al., 2018), and Gene-Set Enrichment Analysis (GSEA) (He et al., 2018a), more stable and interpretable characteristics could be obtained. Usually we calculate the pathway activity difference based the differentially expressed genes calculated based on the comparison of two groups, this method was used to compare the pathway difference between two groups and reveal novel mechanisms, but not suitable for model construction. (Alcaraz et al., 2017). Notably, PARADIGM and Pathifier are exceptions (Vaske et al., 2010; Drier et al., 2013). The Pathifier algorithm only needs the gene expression data from each pathway to produce a coarse-grained score, which represents the degree of dysregulation within related pathways. This means that this algorithm can produce a useful score for evaluating disease subtypes and has been used for subtyping tumors and predicting prognosis in cancer patients (Livshits et al., 2015; Huang et al., 2016; Fa et al., 2019). PARADIGM has been used to infer patient-specific pathway activities from multidimensional cancer genomics data, and could be helpful in integrating multi-omics data and facilitating biomarker discovery in specific diseases (Han et al., 2021; Park et al., 2021). However, since most of the pathways share genes between them, which we call "crosstalk," the specificity of the pathway activity score (PAS) is compromised. Taking intersection genes among pathways into account on PAS quantification will help identify disease-specific features. Thus, developing new prognostic classifier of colorectal cancer based on crosstalk eliminated PAS would be valuable.

Here we developed a new PAS estimation method based on crosstalk factorization, and established a novel pathway-level-feature-based signature for colorectal cancer that can be used to predict overall survival (OS) outcomes, and serve as a complement to the currently available staging system.

MATERIAL AND METHODS

Data Sources

Both mRNA normalized level 3 expression and colorectal cancer clinical data were downloaded from the TCGA portal (https://portal.gdc.cancer.gov/). The microarray data and clinical information from GSE17537, GSE29623, and GSE87211 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). After removing samples without survival data, 613, 55, 65, and 196 samples were retained in these four datasets, respectively.

Pathway Activity Score

The pathway activity score (PAS) for each dataset was calculated based on the method proposed by Bhandari et al. (2019). We downloaded all pathways from the gene oncology (GO) database (http://geneontology.org/) and generated a new mRNA expression matrix that contains only genes exist in it for each pathway. After that, for each gene, based on its expression level, we classified the tumors into two subgroups, the samples in the higher group were scored +1, while the others were scored −1. Finally, we averaged all gene scores in this pathway as the pathway activity score for each tumor sample. A higher PAS indicates a higher pathway activity in the sample, and otherwise, a lower score means lower activity in the sample.

Overall Design and Construction of the Prediction Pipeline

The overall methodology used to define the cancer survival risk subtypes identified in this study is shown in **Figure 1**. Given a series of data sets, based on PASs and survival information, we calculated the log-rank pvalue for each pathway by regression analysis. The pathways were then ranked based on the log-rank pvalue. Then we applied sure independence screening (SIS) to identify the main pathways associated with overall survival in each cohort at a critical threshold of 100. This value was much larger than the default n/log(n) for each cohort, where n is the sample size. We then used these 100 survival-related pathways (**Supplementary Table S1**) to evaluate the impact of crosstalk between these pathways on different datasets. The crosstalk between the two pathways can be divided into three categories: $P_{i \cap j}$, representing the overlapping genes between pathways i and j; $P_i - (P_{i \cap j})$, representing the genes specific to pathway i; and $P_j - (P_{i \cap j})$, representing the genes specific to pathway j. We only retained the crosstalk results where each pathway included at least three genes in the $P_{i \cap j}$ category. We then recalculated the PAS for each sub-pathway and calculated their survival risk p-value using the Cox-PH model. After correction, important (FDR p-value < 0.01) pathways/sub-pathways were determined for each of the three cohorts, and the common pathways/sub-pathways in each data set were used as sample features for further analysis.

We then divided the TCGA data into a training set and a test set using a 4:1 ratio. We then used the activity matrix for each sample to complete a K-means clustering evaluation on the training set and defined the number of clusters as n ($2 \leq n \leq 6$) and then used this evaluation to divide the samples into several subgroups. The optimal number of clusters was determined using three parameters: The C index for the prognostic differences, the Silhouette index, and the Calinski–Harabasz criterion (**Supplementary Table S4**). This analysis allowed us to classify these samples into two distinct subgroups which were redefined as G1 and G2. We then went on to evaluate these two subgroups in other datasets, using various machine-learning frameworks, such as SVM, Adaboost, knearest neighbor (KNN), and Gaussian, and used the pathway activity matrix to build a novel classification model. These algorithms were applied to our frameworks via the Python package "sklearn," which integrated many classification and regression algorithms, using default parameters. We were able to use a

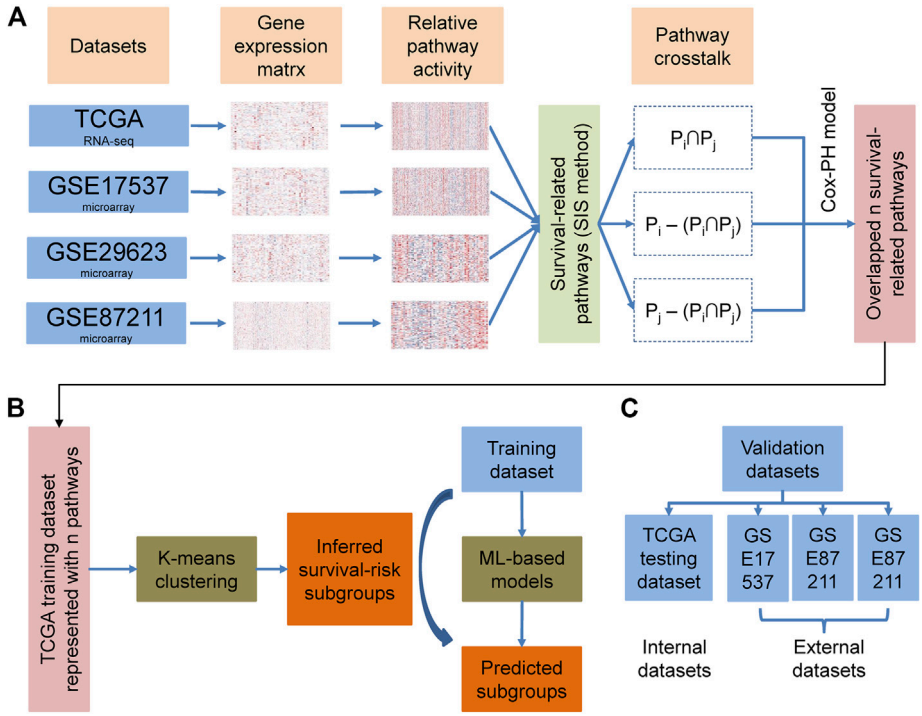


FIGURE 1 | Overall workflow describing the design and validation processes used in this study.

comparison of these predictions to produce an optimal algorithm which was then applied to our final model. We finally settled on the KNN classification algorithm and this was used for further implementation (**Supplementary Table S2**). The number of neighbors was set to 5, the leaf size passed to the classification tree was set to 30, and the power parameter for the Minkowski metric was set to 2.

Evaluation Metrics for Model

We trained our novel classification model using 10-fold cross-validation of the TCGA training dataset and used this to determine the best machine learning framework, as described above. We then went on to evaluate the performance of this classification method using the TCGA test dataset as well as the three GEO datasets. We combined the C-index, Brier score, and log-rank p-value with a deep learning-based study to reflect the prediction accuracy of these methods. The application of these metrics allowed us to quantify the proportion of patient pairs whose prediction was consistent with their OS outcome (Harrell et al., 1996). The C-index was determined using the R "survcomp" package (Schröder et al., 2011), and the Brier score was used to measure the accuracy of the probabilistic predictions. Kaplan–Meier and log-rank analyses were used to compare the survival differences between the groups (R survival package available from http://CRAN.R-project.org/package=survival) and the mean differences between the predicted and observed survival rates at specific time points were determined using the survival analyses metrics, and this score was negatively associated with accuracy and determined using the R "survcomp" package.

Differentially Expressed Genes and GSEA Analyses

Differentially expressed genes between the G2 and G1 subgroups were identified using the DESeq2 package (Love et al., 2014) and GSEA analyses were used to compare the differences between these two subgroups at the hallmark pathway level (Mootha et al., 2003; Subramanian et al., 2005).

Clinical Covariate and Somatic Mutation Analyses

We compared the somatic mutation rates between the G2 and G1 subgroups in the TCGA training cohort using Fisher's exact test and compared the distributions of various clinicopathological features, such as tumor stage, new tumor event, and sex in each of these subgroups using Fisher exact tests.

RESULTS

Identification of Prognostic Subtypes in Colorectal Cancer

The overall methodology used to identify the different cancer survival risk subtypes defined in this study is shown in **Figure 1**. We produced four curated colorectal cancer datasets (TCGA, GSE17537, GSE29623, and GSE87211) using the survival information from the TCGA and GEO databases (**Table 1**). We then used the gene expression matrix from these datasets

TABLE 1 | Clinicopathological features for each of the enrolled cohorts.

Parameters		TCGA training (*n* = 510)	TCGA testing (*n* = 128)	GSE17537 (*n* = 55)	GSE29623 (*n* = 65)	GSE87211 (*n* = 196)
Age	<60	152 (29.8%)	31 (24.22%)	24 (43.64%)	0 (0%)	73 (37.24%)
	≥60	358 (70.2%)	97 (75.78%)	31 (56.36%)	0 (0%)	123 (62.76%)
	Not reported	0 (0%)	0 (0%)	0 (0%)	65 (100%)	0 (0%)
Gender	Male	272 (53.33%)	65 (50.78%)	26 (47.27%)	40 (61.54%)	136 (69.39%)
	Female	237 (46.47%)	61 (47.66%)	29 (52.73%)	25 (38.46%)	60 (30.61%)
	Not reported	1 (0.2%)	2 (1.56%)	0 (0%)	0 (0%)	0 (0%)
OS	Alive	386 (75.69%)	101 (78.91%)	35 (63.64%)	40 (61.54%)	168 (85.71%)
	Dead	104 (20.39%)	22 (17.19%)	20 (36.36%)	25 (38.46%)	28 (14.29%)
	Not reported	20 (3.92%)	5 (3.91%)	0 (0%)	0 (0%)	0 (0%)

OS, overall survival; TCGA, the Cancer Genome Atlas; n*, number.*

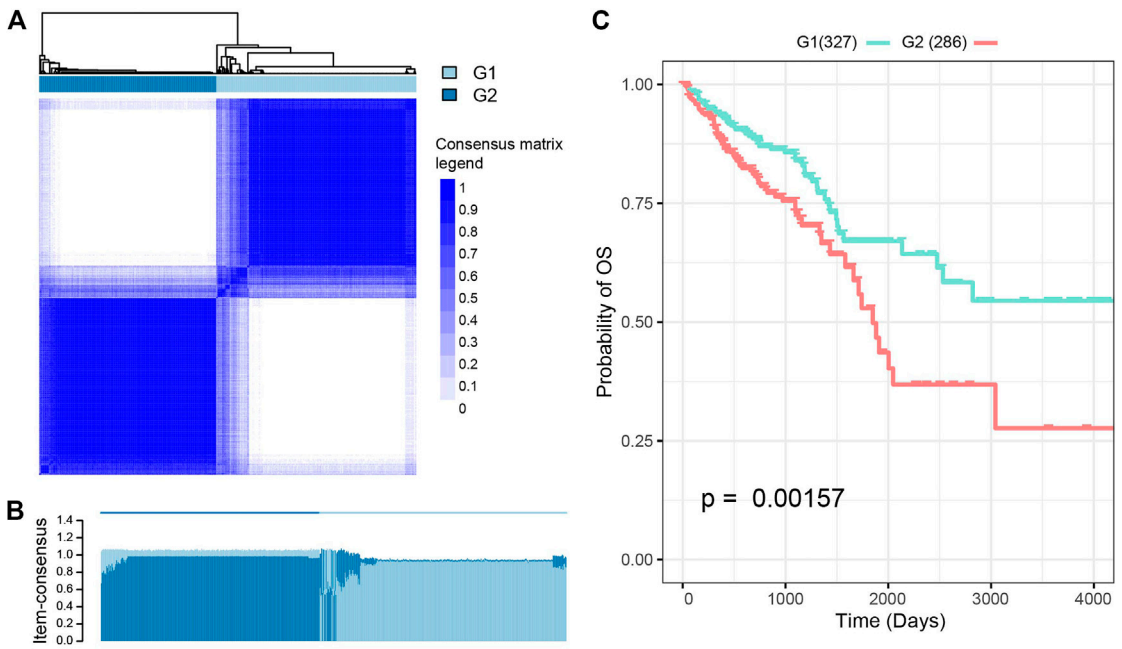


FIGURE 2 | Subtyping analysis of the Cancer Genome Atlas-colorectal cancer (TCGA-CRC) patients. **(A)** K-means clustering analysis split these patients into two subgroups. The optimal number of clusters was determined using three parameters: The C index for prognostic differences, the Silhouette index, and the Calinski–Harabasz criterion. **(B)** Performance of K-means clustering when k was set at 2. **(C)** There were significant differences in survival rates between these two CRC subtypes.

to calculated the PAS for each pathway identified using the KEGG and Gene Ontology Resource databases and then extracted the pathways most closely associated with overall survival. These were then evaluated using the SIS method to identify the central pathways associated with overall survival when using a critical threshold of 100 (**Supplementary Table S1**). We then explored the impact of crosstalk between these pathways on each of the datasets and selected the most important sub-pathways (FDR p-value < 0.01) for each.

These analyses identified 11 central features (**Supplementary Table S3**) which were then used to perform a K-means clustering analysis which produced two subgroups: G1 and G2 (**Figures 2A,B**; **Supplementary Table S4**). Patients from the G2 subgroup displayed significantly worse overall survival than those in the G1 subgroup (**Figure 2C**; $p = 0.0015$, log-rank test).

Evaluating the Performance of Pathway-Based Prognosis

The AUC values for the training and test datasets were 0.9305 and 0.8909, respectively, and the ROC curves are shown in **Supplementary Figure S1**. In addition, we observed a considerable performance in prediction accuracy when using the PAS features as evidenced by the C-index value and significant differences in OS between G1 and G2, as evidenced by the log-rank p-value test (**Figure 3**; **Table 2**). We also noted that this model produced outcomes with low Brier error rates.

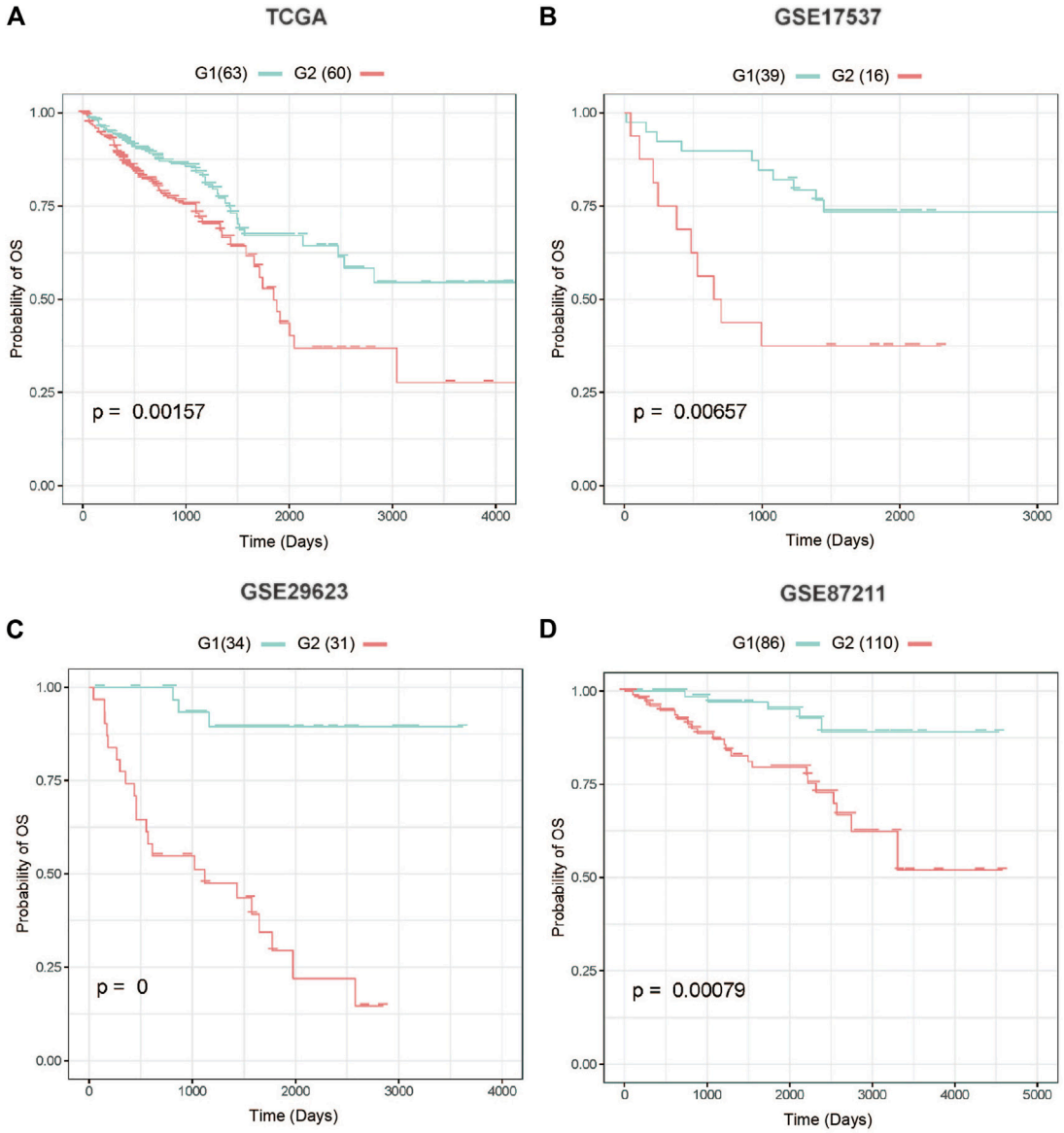


FIGURE 3 | Four test cohorts demonstrating significant differences in survival rates. **(A)** The Cancer Genome Atlas (TCGA) test cohort, **(B)** GSE17537 cohort, **(C)** GSE29623 cohort, **(D)** GSE87211 cohort. G2: aggressive (higher-risk survival) subtype; G1: moderate (lower-risk survival) subtype.

TABLE 2 | Performance of the gene signature in the test TCGA cohort and three external validation cohorts using all 11 features.

Cohort	Omics type	Sample number	C-index	Brier score	Log-rank *P*-value
TCGA	RNA-Seq	123	0.626	0.229	2.90E-03
GSE17537	Microarray	55	0.651	0.227	6.57E-03
GSE29623	Microarray	65	0.747	0.301	3.56E-07
GSE87211	Microarray	196	0.658	0.15	7.91E-04

TCGA, the Cancer Genome Atlas.

The test data from the TCGA-COADREAD samples produced a high C-index (0.626), low Brier score (0.229), and significant average log-rank p-value (2.90e-3) when evaluated for differences in survival. We then validated this model using three external cohorts (**Supplementary Table S5**). The log-rank and Kaplan-Meier analyses of the GSE17537, GSE29623, and GSE87211

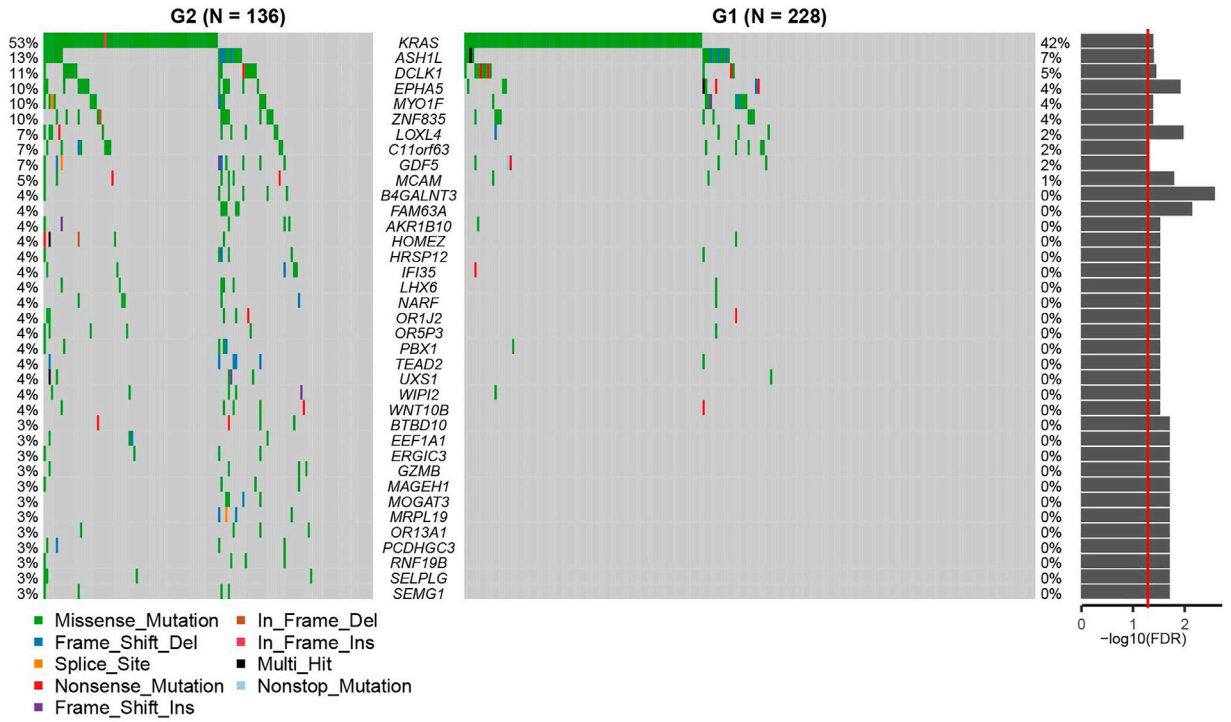


FIGURE 4 | The Oncoprint demonstrating the differences between G2 and G1 subgroups at the genetic level. G2: aggressive (higher-risk survival) subtype; G1: moderate (lower-risk survival) subtype. The p values from the Fisher-exact test are displayed on the right as a bar plot. The red line indicates $p = 0.05$.

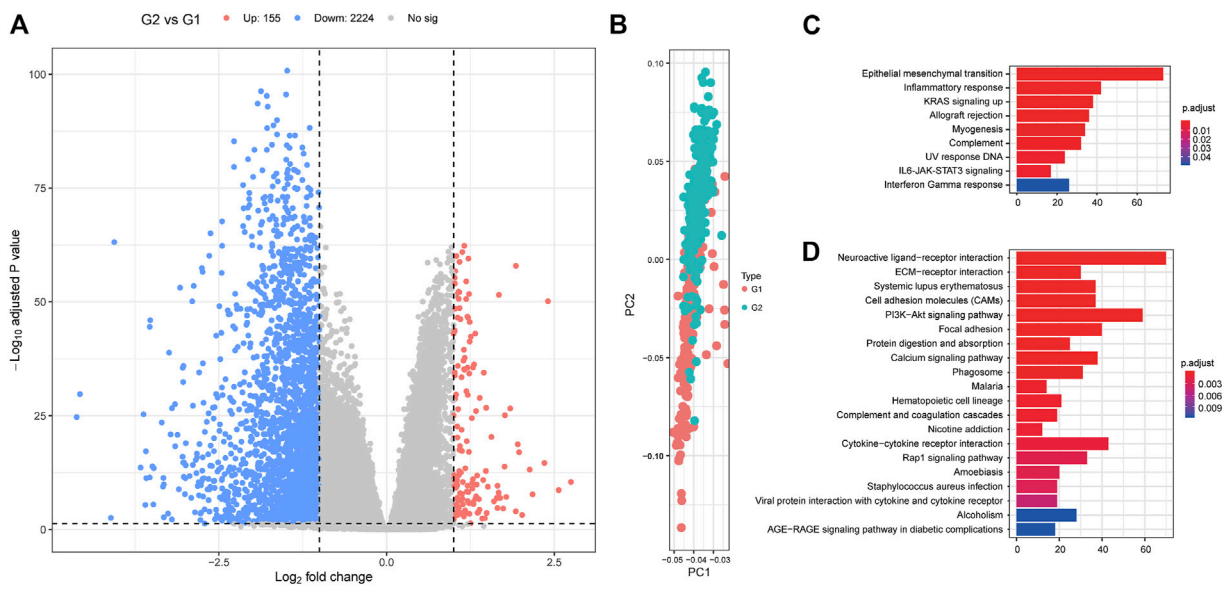


FIGURE 5 | The differentially expressed genes between G1 and G2 subgroups. **(A)** Volcano plot displaying the differentially expressed genes between G2 and G1 subgroups. **(B)** Principal component analysis (PCA) analysis describing the differences in clustering between the G1 and G2 subgroups. Hallmark **(C)** and Kyoto Encyclopedia of Genes and Genomes (KEGG) **(D)** analyses for the differentially expressed genes. G2: aggressive (higher-risk survival) subtype; G1: moderate (lower-risk survival) subtype.

cohorts produced similar results to the TCGA cohort (log-rank p-value < 0.05, **Figure 3**; **Table 2**) indicating that the model was stable. Taken together, these results indicate that this newly developed method can accurately predict OS outcomes in colorectal cancer patients.

Associations Between Different Survival Subtypes and Genomic Features

We then used a Fisher's exact test to evaluate differences between the two survival subtypes identified in the TCGA cohort. These evaluations revealed that KRAS mutations were significantly more frequent in the aggressive subgroup G2 (OR = 1.573, 95%CI: 1.004–2.470, FDR = 0.039, **Figure 4**; **Supplementary Table S6**). This was also shown to be true in several other oncogenes including ASH1L, DCLK1, EPHA5, MYO1F, ZNF835, LOXL4, C11orf63, GDF5, MCAM, B4GALNT3, FAM63A, AKR1B10, HOMEZ, HRSP12, IFI35, LHX6, NARF, OR1J2, OR5P3, PBX1, TEAD2, UXS1, WIPI2, WNT10B, BTBD10, EEF1A1, ERGIC3, GZMB, MAGEH1, MOGAT3, MRPL19, OR13A1, PCDHGC3, RNF19B, SELPLG, and SEMG1 (all $p < 0.05$). Differentially expressed genes from each subtype were identified using the DESeq2 package in R and evaluated using a cutoff value of $|\log_2$ fold change$| > 1$ and FDR P-value < 0.05 (Love et al., 2014). These analyses identified 155 upregulated and 2,224 downregulated genes in the aggressive G2 subgroup (**Figure 5**; **Supplementary Table S7**). Of the upregulated genes, several were found to be associated with survival and pathogenesis in other studies including one study which found that ectopic expression of HBE1 decreased the production of radiation-induced intracellular reactive oxygen species (ROS) and cell mortality (Park et al., 2019); similarly, ORM1 serves as a prognostic factor and can be used to predict therapeutic response in advanced extranodal NK/T cell lymphoma patients treated with pegaspargase/gemcitabine (Zhou et al., 2016). We also tested the correlations between the two survival subtypes (G1 and G2) and the clinicopathological characteristics of the patients in the TCGA cohort and found no significant differences in age, sex, clinical stage or new tumor event subgroups (**Supplementary Figure S2**).

The Distribution of Tumor-Infiltrating Immune Cells in G1 and G2 Patients

Differences in the distribution of 22 subpopulations of TIICs in the G1 and G2 subgroups were determined using the CIBERSORT algorithm and the information from the TCGA dataset and the results of this evaluation are summarized in **Supplementary Figure S3**. We evaluated the average proportion of each immune cell type in both the G1 and G2 subgroups and demonstrated that there were no significant differences in TIIC distribution between these two subgroups. These results suggest that the differences in G1 and G2 survival are not reflected in the TIIC population in these patients.

Hallmark Analyses Compared the Differences Between G2 and G1 Subgroups

We used GSEA analyses to compare the G2 and G1 subgroups to identify the critical pathways involved in the progression of colorectal cancer. This hallmark pathway enrichment analysis revealed that the differences between these two groups were concentrated in the coagulation, MYC targets v2, epithelial-mesenchymal transition, bile acid metabolism, and peroxisome pathways (FDR p-value < 0.05, **Supplementary Figure S4**; **Supplementary Table S8**). This is supported by the fact that most basic research suggests that there is a close connection between hemostatic components and cancer biology as they interact in multiple ways. The coagulation system can be activated by cancer cells, and hemostatic factors can promote tumor progression. In the case of both the MYC (Sikora et al., 1987; Arango et al., 2003; Castell and Larsson, 2015; Boudjadi and Beaulieu, 2016; He et al., 2018b) and EMT (Bates et al., 2007; Vu and Datta, 2017a; Vu and Datta, 2017b; Lamprecht et al., 2018; Ieda et al., 2019; Huang et al., 2020) pathways, many studies have investigated their role in colorectal cancer. We also analyzed the KEGG and Reactome GSEA analyses, and the detailed results are presented in **Supplementary Tables S9, S10**. These evaluations identified several pathways that are significantly associated with tumor progression and resistance to drug treatment. We also used a GSEA-Reactome analysis to compare the G2 and G1 subgroups, and noted the appearance of several significantly enriched pathways including 3'UTR Mediated Translational Regulation, Bile Acid, and Bile Salt Metabolism, Translation, Cytochrome P450 Arranged by Substrate Type, CDK-Mediated Phosphorylation and Removal of CDC6, P53 dependent G1 DNA damage response, P53 independent G1 S DNA damage checkpoint, S phase, and base excision repair. These results suggest that the oncogenic role of these critical pathways may promote colorectal cancer progression.

DISCUSSION

Here, we classified colorectal cancer patients into different risk categories based on the results of our pathway activity score with crosstalk evaluations. Our pathway-level features produced satisfactory outcomes in the TCGA training and three external validation cohorts derived from the GEO database. These results were then used to separate the colorectal cancer patients into aggressive (G2) and moderate (G1) subgroups in the TCGA cohort and the other three colorectal cancer datasets (GSE17537, GSE29623, and GSE87211) with reasonable accuracy. We then went on to evaluate these subtype distinctions in terms of other clinicopathological criteria and revealed that while vastly different in terms of survival there were no significant differences in their clinical presentation. However, when we compared the somatic mutation landscape between these two subgroups we found that the mutation rate in oncogene *KRAS* was significantly higher in the G2 group, which may explain why these patients were seen as experiencing a more aggressive progression than the G1 group. Strategies to improve

outcomes for patients with *KRAS* mutations should be developed. Taken together our data suggests that evaluating the specific genes governing these important features may provide valuable insights into the hallmarks of colorectal cancer and that these may be combined to produce specific prognostic signatures.

RAS is one of the most investigated proto-oncogenes in the world with gain-of-function mutations in this gene being identified in approximately 30% of all human cancers (Artale et al., 2008; Liu et al., 2019). Furthermore, KRAS mutations are associated with poor prognosis in several cancers, and there is still a lack of effective targeted therapeutics designed to counteract the effects of this mutation. Phipps et al. (2013) enrolled 1,989 colorectal patients in their study which was designed to investigate the association between KRAS mutation and survival, and their results suggested that 31% of these patients had KRAS mutations and that these mutations were closely associated with unfavorable outcomes when compared to the wild type (HR = 1.37, 95%CI: 1.13–1.66). Kim et al. (2016) evaluated the impact of KRAS mutations on time to recurrence (TTR) and overall survival (OS) in patients with metastatic colorectal cancer who underwent curative surgery with perioperative chemotherapy. They found that 37.8% of these patients has a KRAS mutation but that these mutations were not associated with TTR or OS (log-rank $p = 0.425$ for TTR; log-rank $p = 0.137$ for OS). In addition, several KRAS positive patients from a set of clinical trials did not respond to treatment with epidermal growth factor receptor inhibitors, cetuximab, or panitumumab (Artale et al., 2008; Van Cutsem et al., 2008) suggesting that these mutations may also be implicated in therapeutic response. Notably, the National Comprehensive Cancer Network guidelines suggest that every colorectal cancer patient, once confirmed to have developed liver metastases, should be screened for KRAS mutations. KRAS mutations serve as a predictor of unfavorable prognosis for colorectal cancer patients in both stage II and III tissues and indicate that these patients could benefit from postoperative FOLFOX chemotherapy (Deng et al., 2015). Given this more studies should investigate the underlying mechanisms of *KRAS* mutation-mediated effects on chemo- and immunotherapy.

We used GSEA analyses to compare the differences between G2 and G1 at the hallmark pathway level. Of the significantly enriched pathways, we found that coagulation, Myc-Target-V2, and epithelial-mesenchymal transition (EMT) were the most significantly correlated with the G2 group, which included patients with poorer outcomes than the G1 group. Most cancer patients demonstrate some biochemical evidence of the systematic activation of coagulation at diagnosis, and hemostatic changes may disappear after curative treatment. Iversen and Thorlacius-Ussing (2003) reported coagulation reactivation in response to cancer recurrence which was demonstrated by significantly increased plasma thrombin antithrombin III complex and Serum ferritin expression. Mandoj et al. (2018) established a risk predicting signature for OS which was shown to be closely associated with age ($p = 0.043$), tumor size ($p = 0.001$), levels of D-dimer ($p = 0.029$), and factor VIII ($p = 0.087$) when evaluated using a multivariate model. Coagulation abnormalities in cancer patients increase the tendency of these patients to develop both hemorrhages and thrombosis (Falanga et al., 2013) and evidence from basic research suggests that hemostatic components and cancer biology interact in multiple ways.

In summary, the PAS-based features and crosstalk evaluation provide an accurate and robust stratification of colorectal cancer patients. This stratification can be clearly linked to prognosis and the signature holds the promise of facilitating precision medicine for colorectal cancer patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA testing cohort, (B) GSE17537 cohort, (C) GSE29623 cohort, (D) GSE87211.

AUTHOR CONTRIBUTIONS

XL and GO designed the experiments. XL and LS performed the experiments and analyzed the data. XL, LS, and JL wrote the manuscript, and XL, JL, and GO revised the manuscript. All the authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.com) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.689676/full#supplementary-material

REFERENCES

Alcaraz, N., List, M., Batra, R., Vandin, F., Ditzel, H. J., and Baumbach, J. (2017). De Novo pathway-based Biomarker Identification. *Nucleic Acids Res.* 45 (16), e151. doi:10.1093/nar/gkx642

Arango, D., Mariadason, J. M., Wilson, A. J., Yang, W., Corner, G. A., Nicholas, C., et al. (2003). c-Myc Overexpression Sensitises colon Cancer Cells to Camptothecin-Induced Apoptosis. *Br. J. Cancer* 89 (9), 1757–1765. doi:10.1038/sj.bjc.6601338

Artale, S., Sartore-Bianchi, A., Veronese, S. M., Gambi, V., Sarnataro, C. S., Gambacorta, M., et al. (2008). Mutations of KRAS and BRAF in Primary and Matched Metastatic Sites of Colorectal Cancer. *Jco* 26 (25), 4217–4219. doi:10.1200/jco.2008.18.7286

Bates, R. C., Pursell, B. M., and Mercurio, A. M. (2007). Epithelial-mesenchymal Transition and Colorectal Cancer: Gaining Insights into Tumor Progression

Using LIM 1863 Cells. *Cells Tissues Organs* 185 (1-3), 29–39. doi:10.1159/000101300

Bhandari, V., Hoey, C., Liu, L. Y., Lalonde, E., Ray, J., Livingstone, J., et al. (2019). Molecular Landmarks of Tumor Hypoxia across Cancer Types. *Nat. Genet.* 51 (2), 308–318. doi:10.1038/s41588-018-0318-2

Boudjadi, S., and Beaulieu, J.-F. (2016). MYC and Integrins Interplay in Colorectal Cancer. *Oncoscience* 3 (2), 50–51. doi:10.18632/oncoscience.293

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

Castell, A., and Larsson, L.-G. (2015). Targeting MYC Translation in Colorectal Cancer. *Cancer Discov.* 5 (7), 701–703. doi:10.1158/2159-8290.cd-15-0660

Deng, Y., Wang, L., Tan, S., Kim, G. P., Dou, R., Chen, D., et al. (2015). KRAS as a Predictor of Poor Prognosis and Benefit from Postoperative FOLFOX Chemotherapy in Patients with Stage II and III Colorectal Cancer. *Mol. Oncol.* 9 (7), 1341–1347. doi:10.1016/j.molonc.2015.03.006

Drier, Y., Sheffer, M., and Domany, E. (2013). Pathway-based Personalized Analysis of Cancer. *Proc. Natl. Acad. Sci.* 110 (16), 6388–6393. doi:10.1073/pnas.1219651110

Fa, B., Luo, C., Tang, Z., Yan, Y., Zhang, Y., and Yu, Z. (2019). Pathway-based Biomarker Identification with Crosstalk Analysis for Robust Prognosis Prediction in Hepatocellular Carcinoma. *EBioMedicine* 44, 250–260. doi:10.1016/j.ebiom.2019.05.010

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46 (D1), D649–d655. doi:10.1093/nar/gkx1132

Falanga, A., Marchetti, M., and Vignoli, A. (2013). Coagulation and Cancer: Biological and Clinical Aspects. *J. Thromb. Haemost.* 11 (2), 223–233. doi:10.1111/jth.12075

Guraya, S. Y. (2019). Pattern, Stage, and Time of Recurrent Colorectal Cancer after Curative Surgery. *Clin. Colorectal Cancer* 18 (2), e223–e228. doi:10.1016/j.clcc.2019.01.003

Han, S., Wang, P.-F., Cai, H.-Q., Wan, J.-H., Li, S.-W., Lin, Z.-H., et al. (2021). Alterations in the RTK/Ras/PI3K/AKT Pathway Serve as Potential Biomarkers for Immunotherapy Outcome of Diffuse Gliomas. *Aging* 13 (11), 15444–15458. doi:10.18632/aging.203102

Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist. Med.* 15 (4), 361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4

He, W., Chen, L., Yuan, K., Zhou, Q., Peng, L., and Han, Y. (2018). Gene Set Enrichment Analysis and Meta-Analysis to Identify Six Key Genes Regulating and Controlling the Prognosis of Esophageal Squamous Cell Carcinoma. *J. Thorac. Dis.* 10 (10), 5714–5726. doi:10.21037/jtd.2018.09.55

He, W. L., Weng, X. T., Wang, J. L., Lin, Y. K., Liu, T. W., Zhou, Q. Y., et al. (2018). Association between C-Myc and Colorectal Cancer Prognosis: A Meta-Analysis. *Front. Physiol.* 9 (1549), 1549. doi:10.3389/fphys.2018.01549

Huang, K., Gao, N., Bian, D., Zhai, Q., Yang, P., Li, M., et al. (2020). Correlation between FAK and EGF-Induced EMT in Colorectal Cancer Cells. *J. Oncol.* 2020, 5428920. doi:10.1155/2020/5428920

Huang, S., Chong, N., Lewis, N. E., Jia, W., Xie, G., and Garmire, L. X. (2016). Novel Personalized Pathway-Based Metabolomics Models Reveal Key Metabolic Pathways for Breast Cancer Diagnosis. *Genome Med.* 8 (1), 34. doi:10.1186/s13073-016-0289-9

Ieda, T., Tazawa, H., Okabayashi, H., Yano, S., Shigeyasu, K., Kuroda, S., et al. (2019). Visualization of Epithelial-Mesenchymal Transition in an Inflammatory Microenvironment-Colorectal Cancer Network. *Sci. Rep.* 9 (1), 16378. doi:10.1038/s41598-019-52816-z

Iversen, L. H., and Thorlacius-Ussing, O. (2003). Systemic Coagulation Reactivation in Recurrence of Colorectal Cancer. *Thromb. Haemost.* 89 (4), 726–734. doi:10.1055/s-0037-1613579

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* 44 (D1), D457–D462. doi:10.1093/nar/gkv1070

Kim, H. S., Heo, J. S., Lee, J., Lee, J. Y., Lee, M.-Y., Lim, S. H., et al. (2016). The Impact of KRAS Mutations on Prognosis in Surgically Resected Colorectal Cancer Patients with Liver and Lung Metastases: a Retrospective Analysis. *BMC Cancer* 16 (1), 120. doi:10.1186/s12885-016-2141-4

Lamprecht, S., Kaller, M., Schmidt, E. M., Blaj, C., Schiergens, T. S., Engel, J., et al. (2018). PBX3 Is Part of an EMT Regulatory Network and Indicates Poor Outcome in Colorectal Cancer. *Clin. Cancer Res.* 24 (8), 1974–1986. doi:10.1158/1078-0432.CCR-17-2572

Lin, H.-H., Wei, N.-C., Chou, T.-Y., Lin, C.-C., Lan, Y.-T., Chang, S.-C., et al. (2017). Building Personalized Treatment Plans for Early-Stage Colorectal Cancer Patients. *Oncotarget* 8 (8), 13805–13817. doi:10.18632/oncotarget.14638

Liu, P., Wang, Y., and Li, X. (2019). Targeting the Untargetable KRAS in Cancer Therapy. *Acta Pharmaceutica Sinica B* 9 (5), 871–879. doi:10.1016/j.apsb.2019.03.002

Livshits, A., Git, A., Fuks, G., Caldas, C., and Domany, E. (2015). Pathway-based Personalized Analysis of Breast Cancer Expression Data. *Mol. Oncol.* 9 (7), 1471–1483. doi:10.1016/j.molonc.2015.04.006

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Luo, D., Liu, Q., Shan, Z., Cai, S., Li, Q., and Li, X. (2020). Development and Validation of a Novel Epigenetic Signature for Predicting Prognosis in colon Cancer. *J. Cel Physiol.* doi:10.1002/jcp.29715

Mandoj, C., Pizzuti, L., Sergi, D., Sperduti, I., Mazzotta, M., Di Lauro, L., et al. (2018). Observational Study of Coagulation Activation in Early Breast Cancer: Development of a Prognostic Model Based on Data from the Real World Setting. *J. Transl Med.* 16 (1), 129. doi:10.1186/s12967-018-1511-x

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1α-responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes. *Nat. Genet.* 34 (3), 267–273. doi:10.1038/ng1180

Park, S. Y., Lee, S.-J., Cho, H. J., Kim, J.-T., Yoon, H. R., Lee, K. H., et al. (2019). Epsilon-Globin HBE1 Enhances Radiotherapy Resistance by Down-Regulating BCL11A in Colorectal Cancer Cells. *Cancers* 11 (4), 498. doi:10.3390/cancers11040498

Park, Y., Heider, D., and Hauschild, A. C. (2021). Integrative Analysis of Next-Generation Sequencing for Next-Generation Cancer Research toward Artificial Intelligence. *Cancers (Basel)* 13 (13), 3148. doi:10.3390/cancers13133148

Phipps, A. I., Buchanan, D. D., Makar, K. W., Win, A. K., Baron, J. A., Lindor, N. M., et al. (2013). KRAS-mutation Status in Relation to Colorectal Cancer Survival: the Joint Impact of Correlated Tumour Markers. *Br. J. Cancer* 108 (8), 1757–1764. doi:10.1038/bjc.2013.118

Schröder, M. S., Culhane, A. C., Quackenbush, J., and Haibe-Kains, B. (2011). Survcomp: an R/Bioconductor Package for Performance Assessment and Comparison of Survival Models. *Bioinformatics* 27 (22), 3206–3208. doi:10.1093/bioinformatics/btr511

Sheng, M., Xie, X., Wang, J., and Gu, W. (2019). A Pathway-Based Strategy to Identify Biomarkers for Lung Cancer Diagnosis and Prognosis. *Evol. Bioinform Online* 15, 1176934319838494. doi:10.1177/1176934319838494

Sikora, K., Chan, S., Evan, G., Gabra, H., Markham, N., Stewart, J., et al. (1987). c-Myc Oncogene Expression in Colorectal Cancer. *Cancer* 59 (7), 1289–1295. doi:10.1002/1097-0142(19870401)59::7<1289:aid-cncr2820590710>3.0.co;2-o

Su, K., Yu, Q., Shen, R., Sun, S.-Y., Moreno, C. S., Li, X., et al. (2021). Pan-cancer Analysis of Pathway-Based Gene Expression Pattern at the Individual Level Reveals Biomarkers of Clinical Prognosis. *Cel Rep. Methods* 1 (4), 100050. doi:10.1016/j.crmeth.2021.100050

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

Van Cutsem, E., Lang, I., D'haens, G., Moiseyenko, V., Zaluski, J., Folprecht, G., et al. (2008). KRAS Status and Efficacy in the First-Line Treatment of Patients with Metastatic Colorectal Cancer (mCRC) Treated with FOLFIRI with or without Cetuximab: The CRYSTAL Experience. *Jco* 26 (15_Suppl. l), 2. doi:10.1200/jco.2008.26.15_suppl.2

Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., et al. (2010). Inference of Patient-specific Pathway Activities from Multi-Dimensional

Cancer Genomics Data Using PARADIGM. *Bioinformatics* 26 (12), i237–i245. doi:10.1093/bioinformatics/btq182

Vu, T., and Datta, P. K. (2017). Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. *Cancers (Basel)* 9 (12). doi:10.3390/cancers9120171

Vu, T., and Datta, P. (2017). Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. *Cancers* 9 (12), 171. doi:10.3390/cancers9120171

Yang, J., Song, K., Guo, W., Zheng, H., Fu, Y., You, T., et al. (2020). A Qualitative Transcriptional Signature for Predicting Prognosis and Response to Bevacizumab in Metastatic Colorectal Cancer. *Mol. Cancer Ther.* doi:10.1158/1535-7163.mct-19-0864

Zhou, Z., Li, Z., Sun, Z., Zhang, X., Lu, L., Wang, Y., et al. (2016). S100A9 and ORM1 Serve as Predictors of Therapeutic Response and Prognostic Factors in Advanced Extranodal NK/T Cell Lymphoma Patients Treated with Pegaspargase/gemcitabine. *Sci. Rep.* 6 (1), 23695. doi:10.1038/srep23695

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fgene.2021.710944

High Expression of C1ORF112 Predicts a Poor Outcome: A Potential Target for the Treatment of Low-Grade Gliomas

*Zhe Zhang[1], Zilong Tan[1], Qiaoli Lv[2], Lichong Wang[1], Kai Yu[1], Huan Yang[1], Huaizhen Liang[3], Tianzhu Lu[4], Yulong Ji[2], Junjun Chen[2], Wei He[1], Zhen Chen[1], Shuhui Chen[4] and Xiaoli Shen[1]**

[1]Department of Neurosurgery, The Second Affifiliated Hospital of Nanchang University, Nanchang, China, [2]Jiangxi Key Laboratory of Translational Cancer Research, Jiangxi Cancer Hospital, Nanchang, China, [3]Department of Orthopaedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4]Department of Radiation Oncology, Jiangxi Cancer Hospital of Nanchang University, Nanchang, China

Background: Glioma is the most common primary tumor of the central nervous system and is associated with poor overall survival, creating an urgent need to identify survival-associated biomarkers. C1ORF112, an alpha-helical protein, is overexpressed in some cancers; however, its prognostic role has not yet been explored in gliomas. Thus, in this study, we attempted to address this by determining the prognostic value and potential function of C1ORF112 in low-grade gliomas (LGGs).

Methods: The expression of C1ORF112 in normal and tumor tissues was analyzed using data from The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA), Oncomine, and Rembrandt databases. The genetic changes of C1ORF112 in LGG were analyzed using cBioPortal. Survival analysis was used to evaluate the relationship between C1ORF112 expression and survival in patients with LGG. Correlation between immune infiltration and C1ORF112 expression was determined using Timer software. Additionally, data from three online platforms were integrated to identify the co-expressed genes of C1ORF112. The potential biological functions of C1ORF112 were investigated by enrichment analysis.

Results: C1ORF112 mRNA was highly expressed in LGGs ($p < 0.01$). Area under the ROC curve (AUC) showed that the expression of C1ORF112 in LGG was 0.673 (95% confidence interval [CI] = 0.618–0.728). Kaplan-Meier survival analysis showed that patients with high C1ORF112 expression had lower OS than patients with low C1ORF112 expression ($p < 0.05$). Multivariate analysis showed that high expression of C1ORF112 was an independent prognostic factor for the overall survival in patients from TCGA and CGGA databases. C1ORF112 expression was positively correlated with six

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Felipe J Núñez,
IIBBA-CONICET Leloir Institute Foundation, Argentina
Yang Xu,
Renmin Hospital of Wuhan University, China

***Correspondence:**
Xiaoli Shen
shenxldoc@126.com

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

Received: *17 May 2021*
Accepted: *01 November 2021*
Published: *22 November 2021*

Citation:
Zhang Z, Tan Z, Lv Q, Wang L, Yu K, Yang H, Liang H, Lu T, Ji Y, Chen J, He W, Chen Z, Chen S and Shen X (2021) High Expression of C1ORF112 Predicts a Poor Outcome: A Potential Target for the Treatment of Low-Grade Gliomas. Front. Genet. 12:710944. doi: 10.3389/fgene.2021.710944

Abbreviations: CGGA, Chinese Glioma Genome Atlas; CNS, central nervous system; DAVID, Database for Annotation, Visualization, and Integrated Discovery; GBM, Glioblastoma; GEPIA, Gene Expression Profiling Interactive Analysis; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LGG, low-grade glioma; OS, overall survival; PPI, protein-protein interaction; ROC, receiver operating characteristic curve; TCGA, The Cancer Genome Atlas; WHO, World Health Organization.

immunoinfiltrating cells (all $p < 0.001$). The enrichment analysis suggested the enrichment of C1ORF112 and co-expressed genes in cell cycle and DNA replication.

Conclusion: This study suggested that C1ORF112 may be a prognostic biomarker and a potential immunotherapeutic target for LGG.

Keywords: C1ORF112, biomarker, immunoinfiltration, low-grade glioma, prognosis

INTRODUCTION

Glioma, the most commonly diagnosed and fatal type of primary tumor of the central nervous system (CNS) (Jiang et al., 2016), is often associated with poor prognosis (Demuth and Berens, 2004). As per the new classification of tumors of the CNS by the World Health Organization (WHO) in 2016, gliomas of the brain can be classified into four grades (I–IV) (Diamandis and Aldape, 2018; Aiman and Rayi, 2021) accordingly, gliomas are considered as "high-grade" and "low-grade", wherein, high-grade gliomas show a high proliferative activity and strong invasion ability; whereas, low-grade gliomas (LGGs) show slow proliferation and relatively long survival time. Therefore, it is imperative to further investigate the key drivers of survival in LGGs and identify potential therapeutic targets to improve the overall prognosis.

The malignant development of a tumor is closely associated with gene expression. In 2012, Van Dam et al. determined that the mouse BC055324 gene (human homologous gene is C1ORF112) showed strong co-expression with cancer-related genes, such as, RAD51 and CCDC6 (van Dam et al., 2012). C1ORF112, an α-helical protein, is co-expressed with many genes in the BRCA-Fanconi anemia-associated DNA damage response pathway, including BRCA1, BRCA2, FANCD2, and FANCI (Nalepa and Clapp, 2018), and is also modified in some tumors with TP53 mutation (Edogbanya et al., 2021). Although, at present, only a few studies have reported the possible dysregulation of C1ORF112 in gastric cancer (Chen et al., 2020), this does suggest its biological and clinical significance in cancer. However, the underlying molecular functions of C1ORF112 and its expression and prognostic value in glioma remain undetermined.

Thus, in this study, we used gene expression and clinical data from Oncomine, The Cancer Genome Atlas (TCGA), and Chinese Glioma Genome Atlas (CGGA) to investigate the relationship between C1ORF112 and LGG, and determine its potential prognostic value in patients with LGG. Additionally, the genes co-expressed with C1ORF112 were collected, and their expression levels were verified in LGG. The results showed that C1ORF112 was significantly overexpressed in LGG samples and was an independent prognostic factor of the overall survival (OS) of patients with LGG. Further, C1ORF112 expression was closely related to the immune response of LGG, and played a crucial role in the malignant progression as well. Thus, C1ORF112, as a new prognostic factor, may be a new therapeutic target for the diagnosis and treatment of LGG.

MATERIALS AND METHODS

Collection of Patient Data

We used TCGA database (TCGA-GBM; TCGA-LGG; https://portal.gdc.cancer.gov/) to download RNA-sequencing transcriptomic data and corresponding clinical information of 511 patients with LGG and 95 normal participants, 163 GBM patients and 207 normal participants. The inclusion criteria were defined as WHO II or III classified patients with complete prognostic information. From 161 patients with LGG and 28 normal participants, GlioVis (http://gliovis.bioinfo.cnio.es/) was used to download RNA-sequencing and corresponding clinical data, and used to verify C1ORF112 expression and prognostic potential in LGGs. Accordingly, patients with LGGs were then categorized into high and low expression groups according to the median expression value of C1ORF112. Additionally, C1ORF112 expression and clinical data of 381 LGG patients were downloaded from the CGGA (http://www.cgga.org.cn/) (Liu et al., 2018) database to analyze the relationship between C1ORF112 and patient prognosis. The patients we studied included both children and adults.

Oncomine Database

Oncomine (https://www.oncomine.org/resource/login.html) database presents integrated RNA and DNA sequence data from the Gene Expression Omnibus, TCGA, and published literature. Using this database, we determined C1ORF112 expression in different types of cancers by setting the following criteria: $p < 0.01$, $|\log2 \text{ fold change}| > 1.5$, gene level 10%, and data type "mRNA".

cBioPortal Database

cBioPortal for Cancer Genomics (http://cBioportal.org) integrates data from more than 100 tumor genomic studies, and records the mutation site and possibility of a copy number variation at the mutation site. Here, we used high-throughput cBioPortal data to analyze the genetic changes associated with C1ORF112 in LGG samples.

Immunoinfiltration Analysis

The relationship between C1ORF112 expression and immune cell infiltration in LGG samples from TCGA database was investigated using the Timer online website tool (Li et al., 2016).

Enrichment and Protein-Protein Interaction Analyses

Multi Experiment Matrix (https://biit.cs.ut.ee/mem/index.cgi) (Kolde et al., 2012) and COXPRESdb (https://coxpresdb.jp/)

(Obayashi et al., 2019) platforms were used to obtain C1ORF112 co-expression genes. Inclusion criterion was $p < 0.05$. According to TCGA data, genes with similar expression patterns as that of C1ORF112 in LGG were analyzed, and the inclusion criterion was $p < 0.05$. Using Database for Annotation, Visualization, and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) (Huang et al., 2009), we performed gene ontology (GO) and Kyoto Gene and Genome Encyclopedia (KEGG) pathway analyses. A protein-protein interaction (PPI) network of C1ORF112 and co-expressed genes was constructed using STRING database (https://string-db.org/) (Szklarczyk et al., 2019), and co-expressed hub genes of C1ORF112 were obtained.

Gene Expression Profiling Interactive Analysis Database

The Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/) (Tang et al., 2019) integrates TCGA data with GTEx normal tissue data to provide key interactive analysis and customization capabilities. Here, we used GEPIA to evaluate the expression and prognostic value of C1ORF112 in GBM, and to evaluate the expression and prognostic value of key co-expressed genes in LGG. The relationship between C1ORF112 and key co-expressed genes with OS was further analyzed.

Statistical Analysis

Unpaired *t*-test was used to compare the expression levels of C1ORF112 between different groups, and $p < 0.05$ was considered significant. Receiver operating characteristic (ROC) curves were generated to evaluate the diagnostic performance of C1ORF112 expression using the SPSS. The median expression level of C1ORF112 was used to distinguish between the OS of patients with LGG. Kaplan-Meier method was used to plot the survival curves, and the OS differences between the groups were evaluated using log-rank test; here as well, $p < 0.05$ was considered significant. Univariate and multivariate analyses were performed to determine whether C1ORF112 expression was an independent prognostic marker in patients with LGG using the Cox proportional risk model. All statistical analyses were performed using R (version 4.0.2) and SPSS (version 26.0).

RESULTS

C1ORF112 was Highly Expressed in Low-Grade Gliomas

Using data from Oncomine, we analyzed the transcriptional levels of C1ORF112 in different cancer types. Compared with the normal tissues ($p < 0.01$, $|\log2 \text{ fold change}| > 1.5$), we found that C1ORF112 was upregulated in almost all cancer types (**Figure 1A**), including colorectal cancer, breast cancer, lung cancer, sarcomas, and tumors of the CNS. Further, multiple data sets showed that C1ORF112 expression was significantly elevated in the CNS (**Table 1**). For instance, in the Sun Brain dataset, C1ORF112 expression in diffuse astrocytomas was 2.313 times higher than that in normal tissues ($p = 2.23E\text{-}4$). Similarly, in the French Brain dataset, C1ORF112 expression was 1.9 times higher ($p = 0.001$) in anaplastic oligodendrocytomas and 1.544 times higher ($p = 1.13E\text{-}5$) in anaplastic oligodendrocytomas than in normal tissues. We have analyzed the relationship between C1orf112 and high-grade glioma using TCGA database. The results indicated that C1orf112 was highly expressed in high-grade gliomas, but its prognostic value was not statistically significant ($p = 0.59$) (**Supplementary Figures S1A,B**). We then obtained C1ORF112 expression profiling data of 511 patients with LGG using TCGA, and observed that C1ORF112 was significantly upregulated in the tumor tissues than in the non-tumor tissues (**Figure 1B**; $p < 0.01$). In addition, we performed a validation using C1ORF112 profiling data from the Rembrandt database (**Figure 1C**; $p < 0.01$), and observed that C1ORF112 was highly expressed in LGGs. Further, area under the ROC curve (AUC) showed that the expression of C1ORF112 in LGG was 0.673 (95% confidence interval [CI] = 0.618–0.728; **Figure 1D**).

Correlation of C1ORF112 With Clinical Features in Low-Grade Gliomas

Using cBioPortal, we found that in LGGs, C1ORF112 had a mutation with a relatively low rate of genetic change. Therefore, the role of highly expressed C1ORF112 in LGG development may not be mediated by mutations or amplification (**Figures 3A,B**). We analyzed the relationship between C1ORF112 and WHO grades using TCGA database, and found that the C1ORF112 mRNA expression was positively correlated with the WHO grades (**Figure 2A**). In addition, We analyzed the relationship between C1ORF112 expression and mutational status of IDH1, ATRX and 1P19Q co-deletion status (**Figures 2B–E**). The relationship between LGG subtype (includes Astrocytoma, Oligoastrocytoma and Oligodendrogliom) and C1ORF112 expression was also analyzed (**Figure 2F**).

Correlation Between C1ORF112 Overexpression and Overall Survival of Low-Grade Gliomas

To investigate the relationship between C1ORF112 expression and OS, we classified 255 patients into the high expression group and 256 patients into the low expression group according to the median expression value of C1ORF112 in PANCAN-LGG in TCGA. Kaplan-Meier survival analysis showed that patients with high C1ORF112 expression had lower OS than patients with low C1ORF112 expression ($p < 0.001$; **Figure 3C**). To further validate the prognostic value of C1ORF112 expression in LGG, data from 161 patients in the Rembrandt database were analyzed (**Figure 3D**; $p < 0.05$). The results showed that patients with high expression of C1ORF112 in LGG had significantly lower OS than those with low expression of C1ORF112. Furthermore, univariate analysis using TCGA data showed that C1ORF112, age, and grade were high-risk factors (**Table 2**). Multivariate

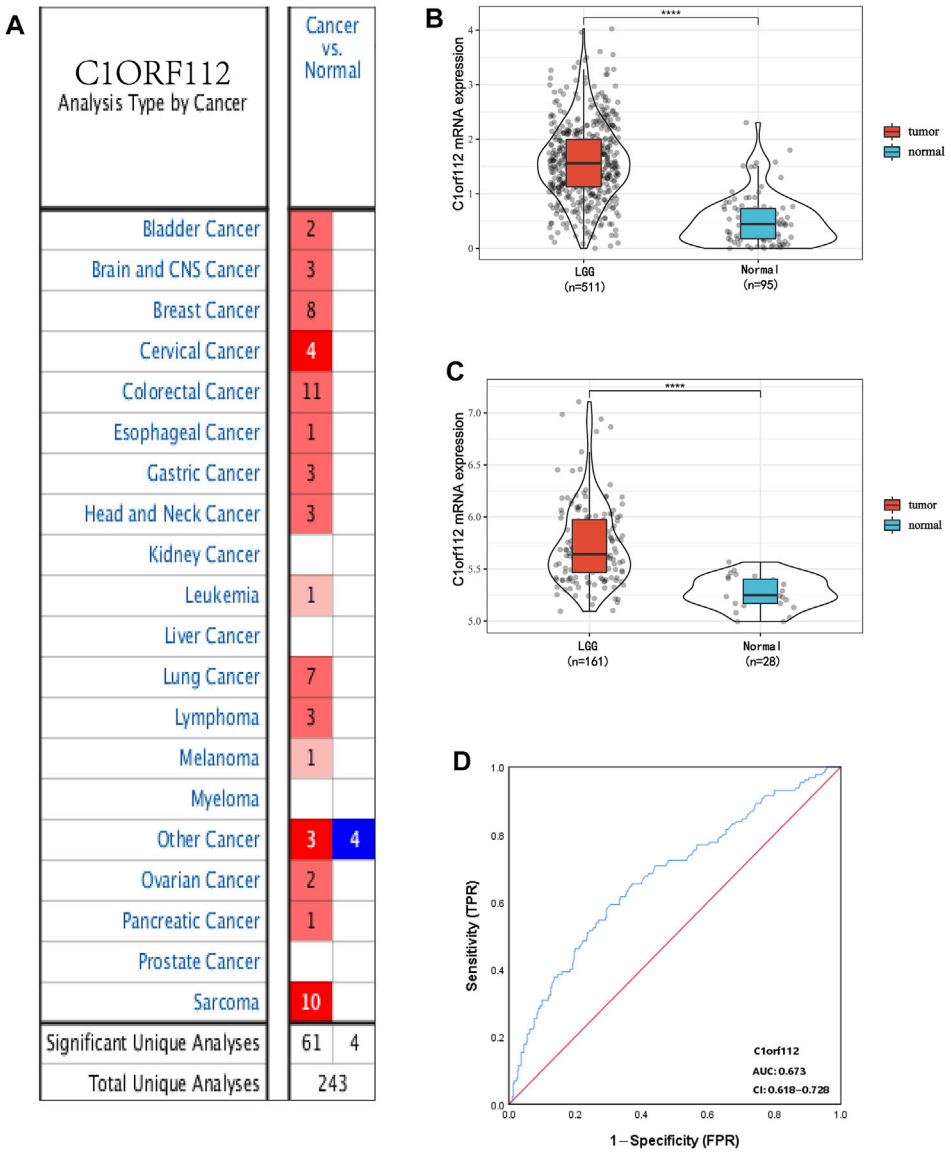


FIGURE 1 | C1ORF112 expression between cancer and normal tissues in LGG patients. **(A)**. Transcriptional expression of C1ORF112 in different types of cancer diseases. C1ORF112 mRNA is highly expressed in low-grade glioma tissues in TCGA dataset **(B)** and Rembrandt dataset **(C)**. **(D)** Receiver operating characteristic analysis (ROC) of C1ORF112 in LGG. ****$p < 0.0001$.

TABLE 1 | Significant changes of C1ORF112 expression in transcription level between Brain glioma and Normal brain tissues (ONCOMINE).

	Types of brain glioma VS normal	Fold change	*p* Value	*t*-Test	Ref
C1ORF112	Diffuse Astrocytoma	2.313	2.23E-4	4.004	Sun Brain
	Anaplastic Oligoastrocytoma	1.900	0.001	5.828	French Brain
	Anaplastic Oligodendroglioma	1.544	1.13E-5	5.545	French Brain

analysis confirmed that C1ORF112 was an independent prognostic factor for the OS of LGG [hazard ratio (HR) = 1.554, 95% CI = 1.040–2.321; $p = 0.031$; **Table 2**]. Similarly, validation using the CGGA database confirmed that C1ORF112 was indeed an independent prognostic factor for the OS of LGG (HR = 1.500, 95% CI = 1.109–2.209, $p = 0.009$; **Table 2**).

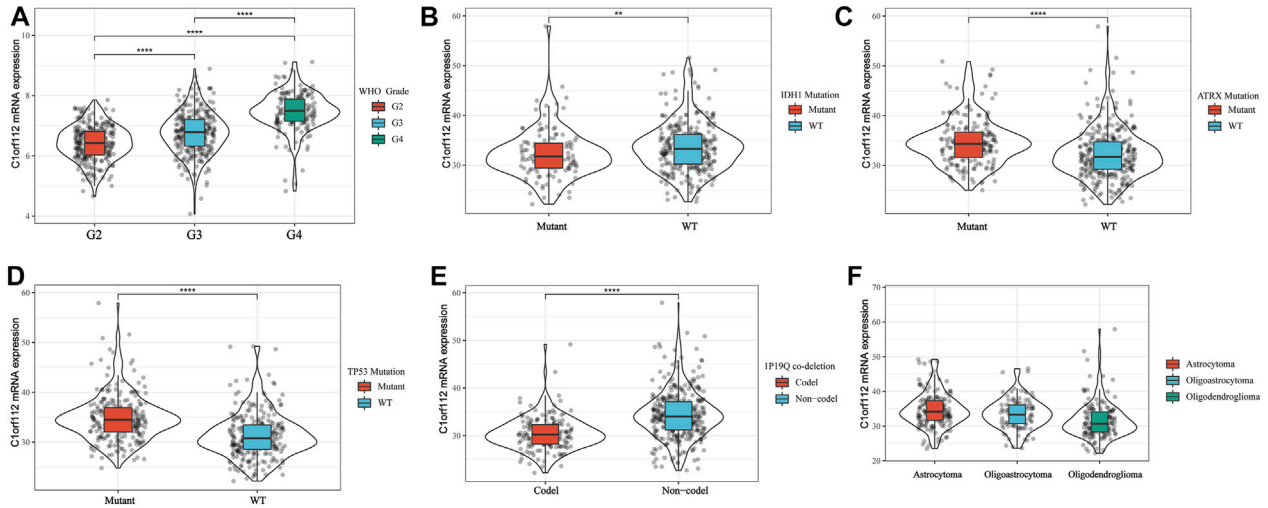


FIGURE 2 | C1ORF112 mRNA was related to WHO Grade **(A)**, IDH1 **(B)**, ATRX **(C)**, TP53 **(D)**, status of 1P19Q co-deletion **(E)**, and LGG subtypes **(F)**. $^{**}p < 0.01$, $^{****}p < 0.0001$.

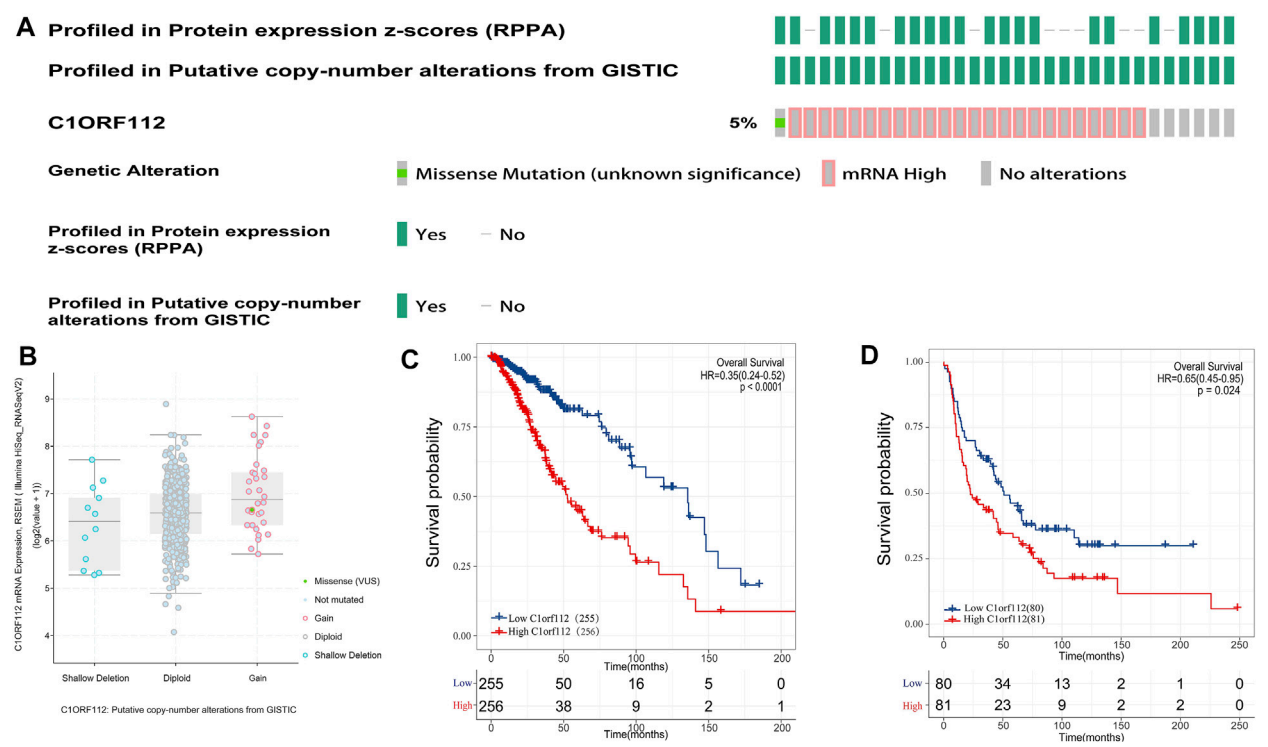


FIGURE 3 | Genetic alterations and prognostic value of C1ORF112 expression in low-grade gliomas. **(A)** Mutation rate of C1ORF112 in LGGs. **(B)** Putative copy number alterations of C1ORF112 in LGGs. **(C)** Survival curves of OS from TCGA dataset (n = 511). **(D)** Survival curves of OS from Rembrandt dataset (n = 161).

Relationship Between C1ORF112 and Immune Infiltration

Immune cell infiltration may be an important pathophysiological factor in the development of glioma. We analyzed the relationship between C1ORF112 expression and infiltration of six common immune cells: B cells, $CD8^+$ T cells, $CD4^+$ T cells, macrophages, neutrophils, and dendritic cells. C1ORF112 expression was positively correlated with all six immune cells ($p < 0.001$; **Figure 4**), indicating that patients with high C1ORF112 expression in LGGs had higher immune cell infiltration than patients with low C1ORF112 expression.

TABLE 2 | Univariate and multivariate analysis of C1ORF112 expression profile in TCGA database and CGGA database.

Datasets	Univariate			Multivariate	
	Variable	HR (95% CI)	P	HR (95% CI)	P
TCGA	Age	4.764 (3.042–7.462)	<0.001	4.998 (3.111–8.030)	<0.001
	Gender	1.159 (0.796–1.686)	0.441	1.053 (0.715–1.550)	0.795
	Histological_type	0.752 (0.607–0.932)	0.009	0.800 (0.634–1.010)	0.061
	Histologic_grade	3.630 (2.397–5.496)	<0.001	3.067 (2.015–4.668)	<0.001
	Idh1_mutation	0.895 (0.544–1.472)	0.663	0.957 (0.572–1.601)	0.867
	C1ORF112	1.876 (1.284–2.740)	0.001	1.554 (1.040–2.321)	0.031
CGGA	Age	1.486 (0.760–2.905)	0.246	1.012 (0.507–2.023)	0.972
	Gender	1.085 (0.811–1.452)	0.583	1.146 (0.846–1.553)	0.378
	Histological_type	1.224 (1.114–1.343)	<0.001	0.558 (0.395–0.786)	0.001
	Histologic_grade	2.932 (2.096–4.100)	<0.001	17.005 (5.715–50.604)	<0.001
	Idh1_mutation	0.463 (0.339–0.630)	<0.001	0.517 (0.367–0.728)	<0.001
	C1ORF112	1.772 (1.324–2.372)	<0.001	1.500 (1.109–2.209)	0.009

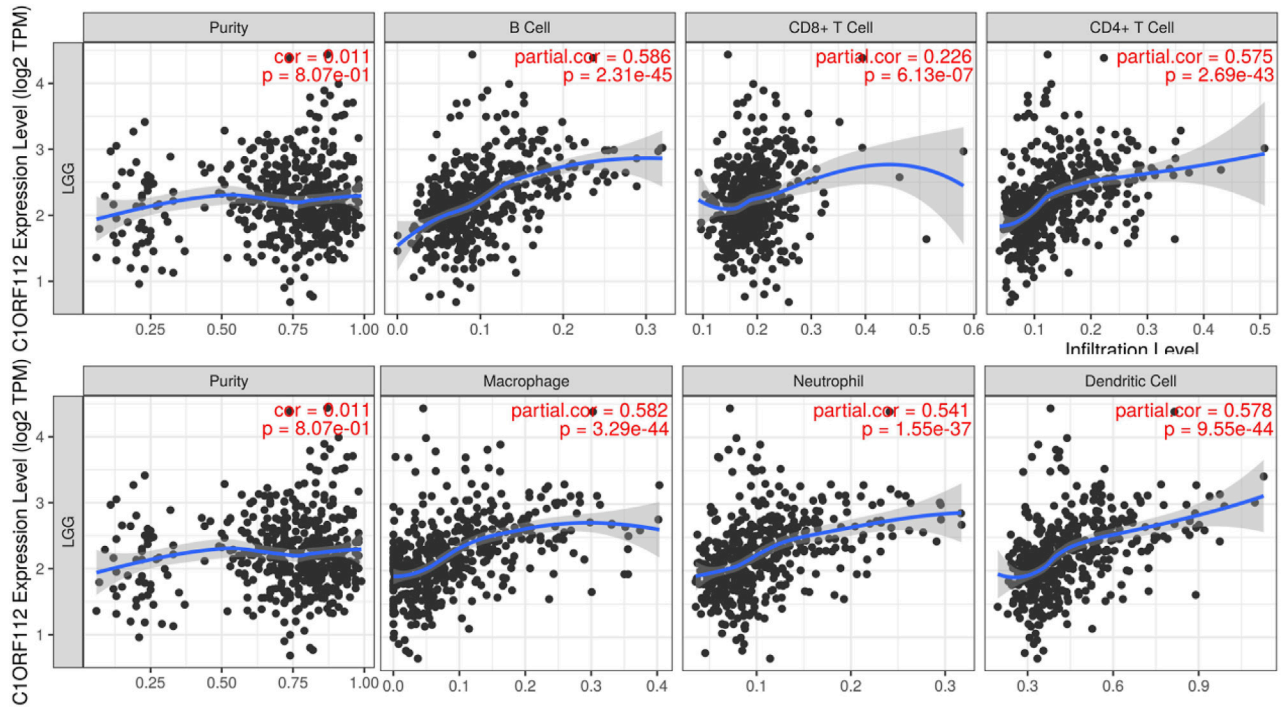


FIGURE 4 | The relationship between C1ORF112 and the level of immune infiltration.

Enrichment and Protein-Protein Interaction Analyses of C1ORF112 Co-expressed Genes

We obtained 2000 C1ORF112 co-expressed genes from Multi Experiment Matrix and Coxpresdb platforms. Then, 1,000 genes similar to C1ORF112 expression patterns in LGGs were obtained by calculating the TCGA database. Finally, 319 overlapping genes (that were overlapping in all three databases) were considered as the co-expressed genes of C1ORF112 in LGGs (**Figure 5A**). GO enrichment analysis showed that C1ORF112 and the co-expressed genes were mainly enriched in cell division, DNA repair, and ATP binding (**Figures 5B–D**). KEGG analysis revealed that C1ORF112 and the co-expressed genes were mainly enriched in cell cycle, DNA replication, pyrimidine metabolism, and RNA transport (**Figure 5E**). PPI analysis showed that CDK1, CCNB1, CCNB2, CDC20 were the key co-expressed genes of C1ORF112 in LGGs (**Figures 6A,B**).

Correlation Between the Key C1ORF112 Co-expressed Genes and Low-Grade Gliomas

We used GEPIA to analyze the RNA-sequencing data of 518 LGG tissues from TCGA and 207 normal samples from the

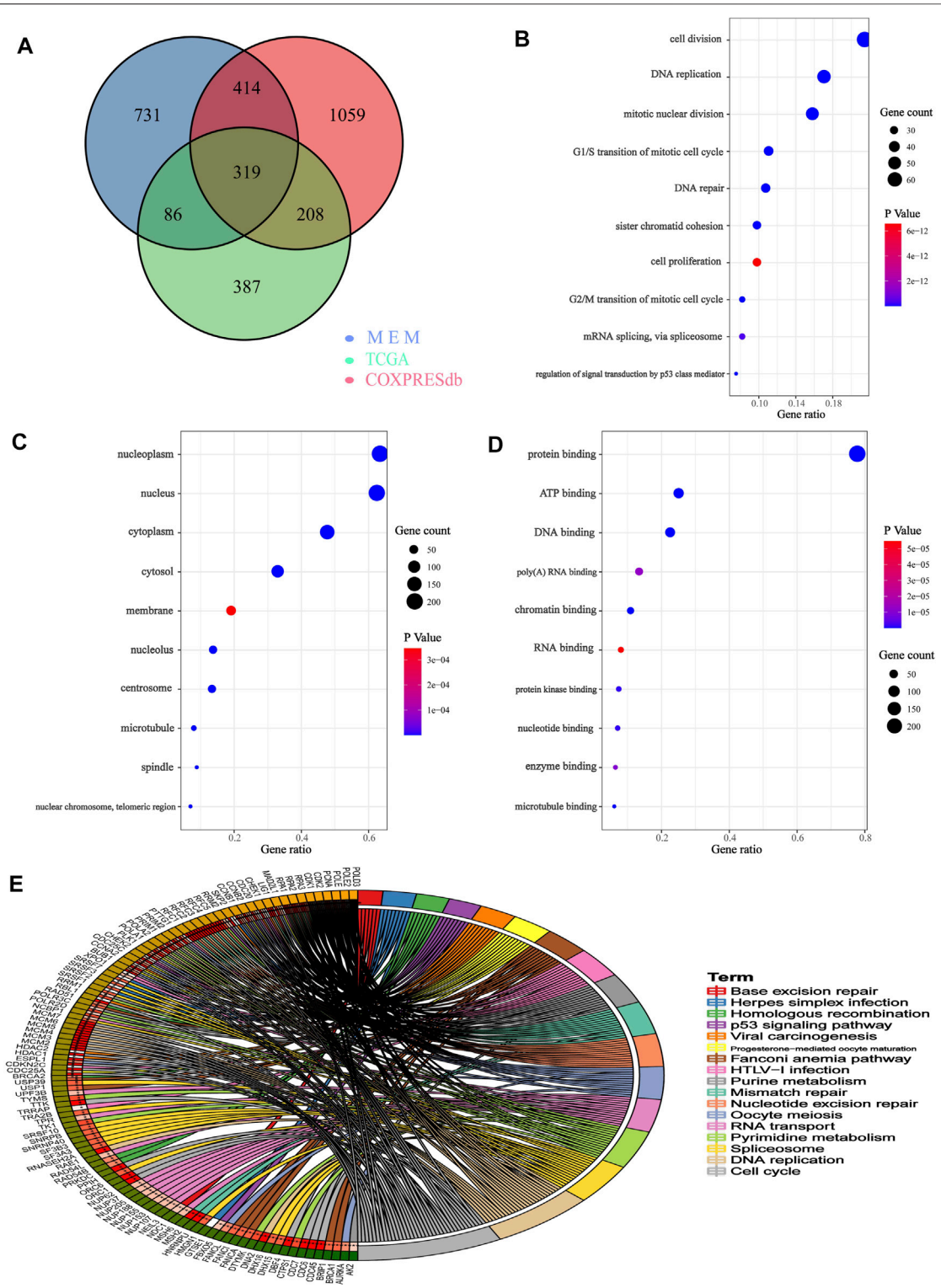


FIGURE 5 | Functional enrichment of C1ORF112 and its co-expressed genes in low-grade gliomas. **(A)** Venn diagram of C1ORF112 co-expressed genes in LGG. **(B)** Enriched GO terms in the "biological process" category. **(C)** Enriched GO terms in the "cellular component" category. **(D)** Enriched GO terms in the "molecular function" category. **(E)** Kyoto Encyclopedia of Genes and Genomes Pathway.

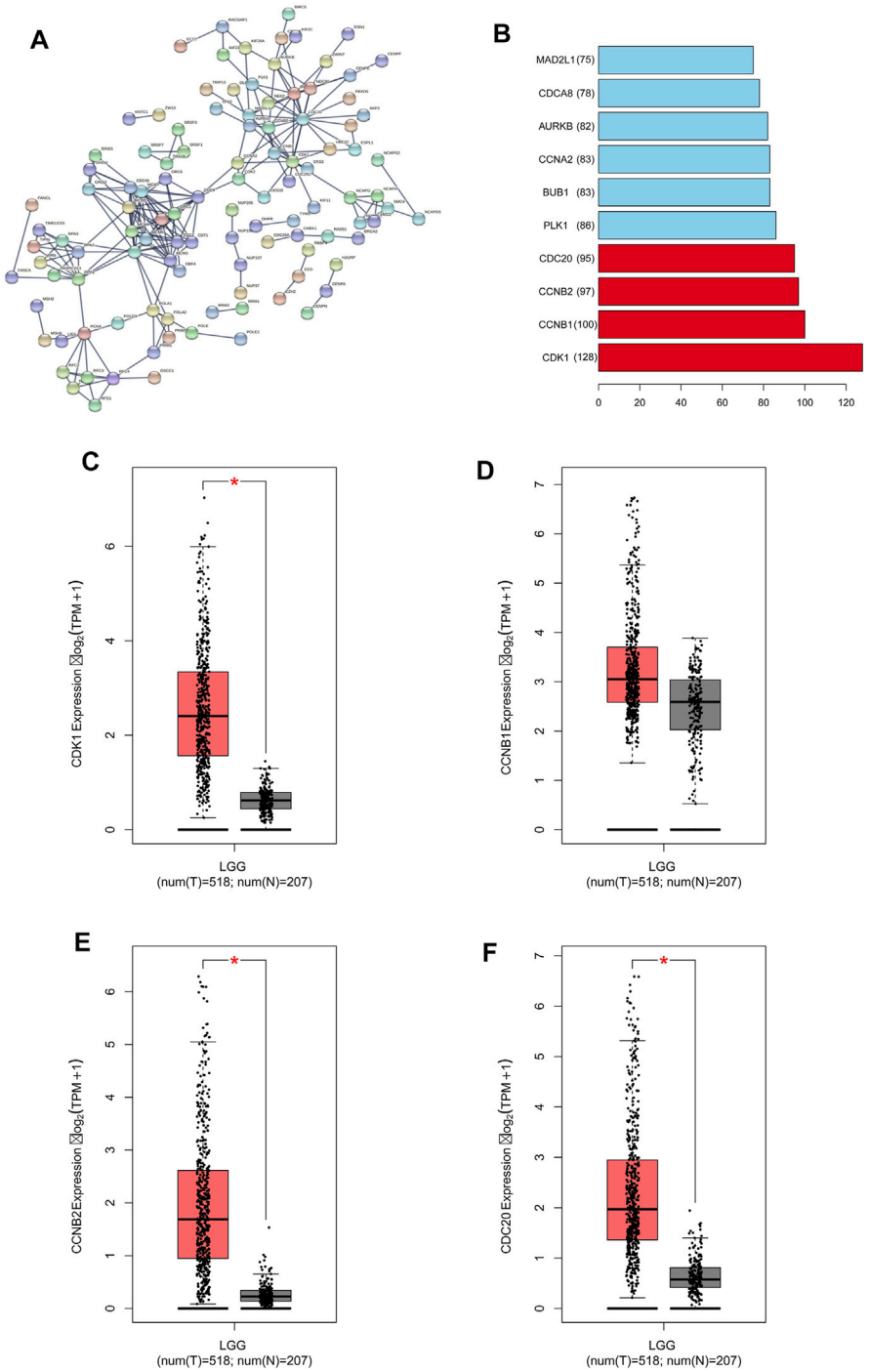


FIGURE 6 | C1ORF112 and its co-expressed genes in a protein-protein interaction network in low-grade gliomas. **(A)** PPI network. **(B)** Top 10 of key co-expressed genes of C1ORF112. The expression level of key co-expressed genes in LGGs was : **(C)** CDK1; **(D)** CCNB1; **(E)** CCNB2; **(F)** CDC20. *p < 0 .05.

GTEx project, and found that CDK1, CCNB1, CCNB2, and CCDC20 were highly expressed in the LGG tissues and poorly expressed in the normal tissues (**Figures 6C–F**). Kaplan-Meier survival analysis showed that LGG patients with high CDK1, CCNB1, CCNB2, and CDC20 expression had a significantly lower OS than patients with low CDK1 expression (p < 0.001; **Figures 7A–D**). LGG analysis in TCGA showed that C1ORF112 was significantly positively correlated with the key co-expressed genes (**Figures 7E–H**).

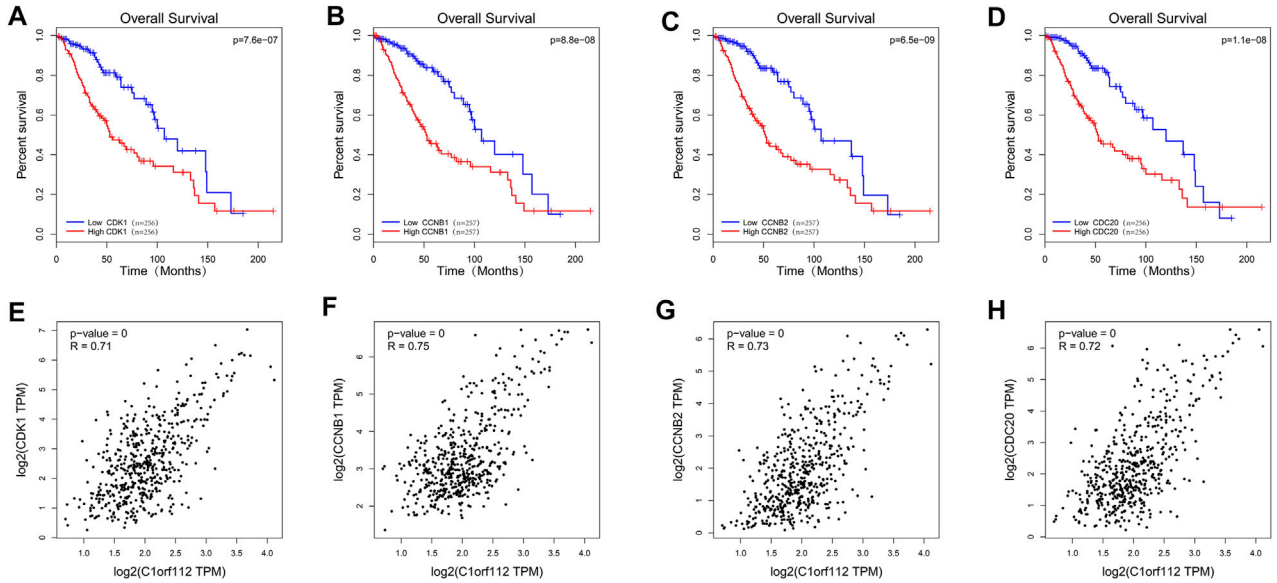


FIGURE 7 | The TCGA database was used to analyze the prognostic value and correlation of co-expressed genes. Survival curves of OS: **(A)** CDK1; **(B)** CCNB1; **(C)** CCNB2; **(D)** CDC20. Correlation between C1ORF112 and key co-expressed genes: **(E)** CDK1; **(F)** CCNB1; **(G)** CCNB2; **(H)** CDC20.

DISCUSSION

Glioma, a highly heterogenous tumor, is the most commonly diagnosed tumor in the CNS and is associated with poor OS (Ostrom et al., 2018; Waker and Lober, 2019). While there have been a few studies on C1ORF112 mRNA, to our knowledge, no study has investigated the correlation between C1ORF112 and LGG. In this study, 163 GBM tissues and 207 normal tissues were studied in TCGA database. Our results showed that C1ORF112 was significantly overexpressed in GBM, but the prognostic analysis was not statistically significant. This might be explained by the small number of glioblastoma samples and different molecular mechanisms between LGG and GBM. Therefore, we integrated and screened clinical RNA-sequencing data from TCGA, CGGA, and Rembrandt databases, and obtained a total of 1053 LGG tissues and 123 normal tissues. Our results revealed that C1ORF112 was significantly overexpressed in LGGs with ATRX mutation, TP53 mutation, 1p19q non-codel, or Astroglioma. In addition, univariate and multivariate analyses showed that C1ORF112 expression was an independent prognostic factor for LGG. The enrichment analysis showed that C1ORF112 and its co-expressed genes were associated with cell cycle, DNA replication, pyrimidine metabolism, nucleotide excision, and repair, RNA transport, purine metabolism, and the Fanconi anemia pathway. Therefore, C1ORF112 may play an important role in glioma pathogenesis, and may be a potential LGG biomarker.

High C1ORF112 mRNA expression has been previously reported in breast cancer (Leo et al., 2005), gastric cancer (Chen et al., 2020), desmoid tumors (Bowden et al., 2007), bladder cancer (Sanchez-Carbayo et al., 2007), head and neck squamous cell carcinoma (Renkonen et al., 2017), cervical cancer, and others. Increased copy number of C1ORF112 has been reported in breast cancer studies (Gonzalez-Perez et al., 2013; Rubio-Perez et al., 2015). However, to date, the expression of C1ORF112 in LGG has not been studied, and the expression of C1ORF112 in other cancers has only been verified *via* the co-expression analysis of related genes. In this study, we confirmed that C1ORF112 was significantly overexpressed in most tumors in TCGA database. Further analysis showed that C1ORF112 was highly expressed in LGG than in normal tissues ($p < 0.001$). Additionally, the AUC was 0.673. Together, these results suggest that C1ORF112 has the potential to be a diagnostic marker for many cancers, including LGG. The expression of C1ORF112 is closely related to the survival of patients with endometrial cancer, wherein, higher the expression, worse the prognosis. Notably, our study is the first to show that C1ORF112 expression may influence the prognosis in LGG. By analyzing TCGA-LGG data, we found that patients with high C1ORF112 mRNA expression had poor OS, and this was an independent prognostic factor for OS and progression-free survival. This result was supported by clinical LGG data from the Rembrandt database as well. In addition, multivariate analysis showed that C1ORF112 was an independent prognostic factor for LGG. Therefore, it is necessary to further investigate the role of C1ORF112 in LGG.

To explore the possible mechanism of C1ORF112 in LGG, we performed enrichment analysis of C1ORF112 and its co-expressed genes, and found that they were enriched in cell cycle, DNA replication, Fanconi anemia, Mismatch repair, Nucleotide excision repair etc. Previous studies have found

that C1ORF112 may influence the Fanconi anemia pathway or its regulation (Liu et al., 1998; Auerbach, 2009). Zhaojing et al. has reported that LINC00152 promotes the proliferation, migration, and invasion of gastric cancer cells *in vitro* through the cell cycle pathway (Zhao et al., 2015). Qiuni et al. reported that cullin-7 is a predictor of poor prognosis in patients with breast cancer, and is involved in the regulation of breast cancer cells by regulating the cell cycle (Qiu et al., 2018). Yun et al. reported the involvement of the cell cycle pathway in cerebellar meningeal metastasis of non-small cell lung cancer (Fan et al., 2018). Hung-Wei et al. reported that overexpression of cell cycle regulating nuclear cell protein L2DTL was associated with the progression and poor prognosis of hepatocellular carcinoma (Pan et al., 2006). Thus, the cell cycle not only plays an important role in tumor regulation, but also affects the prognostic evaluation (Williams and Stoeber, 2012). C1ORF112 and its co-expressed genes were positively correlated with the cell cycle. Four key co-expression genes (CDK1, CCNB1, CCNB2, and CDC20) of C1ORF112 were further analyzed. CDK1-mediated perturbations in chromosome stability and G2/M control that promotes cell cycle progression are key tumorigenic events (Asghar et al., 2015). Overexpression of FOXM1 and upregulation of CCNB1 leads to a malignant phenotype (Katoh et al., 2013). CCNB2 is overexpressed in non-small cell lung cancer and is closely associated with poor prognosis (Qian et al., 2015). High expression of CDC20 is significantly associated with reduced survival of most human tumors (Wang et al., 2018). Therefore, we speculate that C1ORF112 may be involved in the progression of LGG *via* the cell cycle. DNA damage repair is a phenomenon that DNA molecules of normal cells are damaged followed by a series of activation of various enzymes to restore their structures (Ciccia and Elledge, 2010). The mechanism plays an important role in maintaining gene stability (Yan et al., 2016). DNA damage repair includes four types: nucleotide excision repair, base excision repair, recombination repair, and mismatch repair. DNA damage repair is quite important for regulating the therapeutic response of cancer (Squatrito and Holland, 2011). Numerous chemotherapeutic drugs exert anti-tumor effects through DNA damage (Pang et al., 2020). Temozolomide, for example, is the first-line drugs to kill glioma cells by damaging their DNA. Our analysis showed that C1ORF112 and its co-expressed genes were involved in nucleotide excision repair and mismatch repair. Therefore, we speculate that C1ORF112 may be involved in the progression of LGG through facilitating DNA damage repair. Finally, we concluded that four key co-expression genes of C1ORF112 (CDK1, CCNB1, CCNB2, and CDC20) were highly expressed in LGG, and the high expression of these genes was closely associated with poor prognosis in patients with LGG.

Infiltration of immune cells plays a crucial role in tumor growth, metastasis, and treatment response (Gieryng et al., 2017). Therefore, we analyzed the correlation between C1ORF112 and infiltrating immune cells. The results showed that the expression of C1ORF112 was negatively correlated with tumor purity and positively correlated with the infiltration of $CD4^+$ T cells, $CD8^+$ T cells, B cells, neutrophils, dendritic cells, and other immune cells. This is the first report of the potential involvement of C1ORF112 in immunity. According to the immune response, low-grade gliomas can form an immunosuppressive tumor microenvironment similar to other tumors that impair T cell antitumor responses *via* immune checkpoint inhibition pathways. Immune checkpoint inhibitors (i.e., monoclonal antibody inhibitors) were develped to block these inhibitory signaling pathways, activate systemic immunity, and thus enhance T cell activity. Typically, Programmed cell death 1 (PD-1) and its ligand Programmed cell death ligand 1 (PD-L1) mediates tumor immunosuppression by promoting T cell apoptosis and Treg induction (Nduom et al., 2016). Therefore, PD-1/PD-L1 is an important immunosuppressive interaction for tumor cells to escape from the immune killing of the matrix. Dung et al. reported the mismatch repair deficiency predicts response of tumors to PD-1 blockade (Le et al., 2017). Our current study found that C1ORF112 and its co-expressed genes are functionally enriched in mismatch repair. However, further studies are needed to evaluate the role C1ORF112 and mismatch repair on PD-L1 immune checkpoint therapy. Immunotherapy aims to strengthen the immune system of a patient to recognize and attack tumor cells. Therefore, C1ORF112 may be a target for future immunotherapy.

Through this study, we have improved the understanding of the relationship between C1ORF112 and LGG; however, there are still some limitations of our study. The knock-down and knock-out experiments of C1ORF112 is supposed to be performed to explore and verify its function in LGG in our future study. For example, proliferation, migration, invasion, and immune response of LGG cells, and *in vivo* study using LGG mouse model can be performed to verify the mechanism of C1ORF112. Translating these cell cycle-associated biomarkers into practical clinical applications also requires further investigation. In conclusion, the overexpression of C1ORF112 mRNA in LGG was closely related to the poor prognosis of patients with LGG. Enrichment analysis showed that C1ORF112 may regulate the progression of LGG *via* the cell cycle, affect the prognosis of patients with LGG, and thus play a potential role as a carcinogenic factor. Finally, this study suggests that C1ORF112 may be a potential biomarker for the diagnosis and prognosis of LGG, and a potential immunotherapeutic target.

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

ZZ and XS designed the manuscript. LW, QL, ZT, KY, HY, HL, and TL conducted the study and analyzed the data. YJ, JC, WH, SC, and ZC contributed to the literature retrieval of the manuscript. ZZ wrote the manuscript and the writing was

directed by XS. All authors read and approved the final manuscript.

FUNDING

This study by the national natural science fund project (81960458, 81860664, 82060680), The Province Natural Science Foundation of Jiangxi Province (Nos 20192BAB215063, 20202BABL216080), The research program of science and technology in jiangxi province department of education (180077), Youth Project of Jiangxi Provincial Department of Education (GJJ200244), Science and Technology Program of Jiangxi Provincial Health Commission (202130347), Project of the Second Affiliated Hospital of Nanchang University (2014YNLC12009).

ACKNOWLEDGMENTS

We thank editage (http://www.fabiao@editage.cn) for editing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.710944/full#supplementary-material

Supplementary Figure S1 | C1ORF112 expression and prognostic value between cancer and normal tissues in GBM patients. **(A)** C1ORF112 mRNA is highly expressed in GBM tissues in TCGA dataset. **(B)** Survival curves of OS from TCGA dataset (*n*=162).

REFERENCES

Aiman, W., and Rayi, A. (2021). *Low Grade Gliomas*. StatPearls. Treasure Island FL.

Asghar, U., Witkiewicz, A. K., Turner, N. C., and Knudsen, E. S. (2015). The History and Future of Targeting Cyclin-dependent Kinases in Cancer Therapy. *Nat. Rev. Drug Discov.* 14, 130–146. doi:10.1038/nrd4504

Auerbach, A. D. (2009). Fanconi Anemia and its Diagnosis. *Mutat. Research/Fundamental Mol. Mech. Mutagenesis* 668, 4–10. doi:10.1016/j.mrfmmm.2009.01.013

Bowden, N. A., Croft, A., and Scott, R. J. (2007). Gene Expression Profiling in Familial Adenomatous Polyposis Adenomas and Desmoid Disease. *Hered. Cancer Clin. Pract.* 5, 79–96. doi:10.1186/1897-4287-5-2-79

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

Ciccia, A., and Elledge, S. J. (2010). The DNA Damage Response: Making it Safe to Play with Knives. *Mol. Cel* 40, 179–204. doi:10.1016/j.molcel.2010.09.019

Demuth, T., and Berens, M. E. (2004). Molecular Mechanisms of Glioma Cell Migration and Invasion. *J. Neurooncol.* 70, 217–228. doi:10.1007/s11060-004-2751-6

Diamandis, P., and Aldape, K. (2018). World Health Organization 2016 Classification of Central Nervous System Tumors. *Neurol. Clin.* 36, 439–447. doi:10.1016/j.ncl.2018.04.003

Edogbanya, J., Tejada-Martinez, D., Jones, N. J., Jaiswal, A., Bell, S., Cordeiro, R., et al. (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

Fan, Y., Zhu, X., Xu, Y., Lu, X., Xu, Y., Wang, M., et al. (2018). Cell-Cycle and DNA-Damage Response Pathway Is Involved in Leptomeningeal Metastasis of Non-small Cell Lung Cancer. *Clin. Cancer Res.* 24, 209–216. doi:10.1158/1078-0432.ccr-17-1582

Gieryng, A., Pszczolkowska, D., Walentynowicz, K. A., Rajan, W. D., and Kaminska, B. (2017). Immune Microenvironment of Gliomas. *Lab. Invest.* 97, 498–518. doi:10.1038/labinvest.2017.19

Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M. P., Jene-Sanz, A., et al. (2013). IntOGen-mutations Identifies Cancer Drivers across Tumor Types. *Nat. Methods* 10, 1081–1082. doi:10.1038/nmeth.2642

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

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

Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H., and Katoh, M. (2013). Cancer Genetics and Genomics of Human FOX Family Genes. *Cancer Lett.* 328, 198–206. doi:10.1016/j.canlet.2012.09.017

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust Rank Aggregation for Gene List Integration and Meta-Analysis. *Bioinformatics (Oxford, England)* 28, 573–580. doi:10.1093/bioinformatics/btr709

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., et al. (2017). Mismatch Repair Deficiency Predicts Response of Solid Tumors to PD-1 Blockade. *Science* 357, 409–413. doi:10.1126/science.aan6733

Leo, J. C. L., Wang, S. M., Guo, C. H., Aw, S. E., Zhao, Y., Li, J. M., et al. (2005). Gene Regulation Profile Reveals Consistent Anticancer Properties of Progesterone in Hormone-independent Breast Cancer Cells Transfected with Progesterone Receptor. *Int. J. Cancer* 117, 561–568. doi:10.1002/ijc.21186

Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol.* 17, 174. doi:10.1186/s13059-016-1028-7

Liu, N., Lamerdin, J. E., Tebbs, R. S., Schild, D., Tucker, J. D., Shen, M. R., et al. (1998). XRCC2 and XRCC3, New Human Rad51-Family Members, Promote Chromosome Stability and Protect against DNA Cross-Links and Other Damages. *Mol. Cel* 1, 783–793. doi:10.1016/s1097-2765(00)80078-7

Liu, X., Li, Y., Qian, Z., Sun, Z., Xu, K., Wang, K., et al. (2018). A Radiomic Signature as a Non-invasive Predictor of Progression-free Survival in Patients with Lower-Grade Gliomas. *NeuroImage: Clin.* 20, 1070–1077. doi:10.1016/j.nicl.2018.10.014

Nalepa, G., and Clapp, D. W. (2018). Fanconi Anaemia and Cancer: an Intricate Relationship. *Nat. Rev. Cancer* 18, 168–185. doi:10.1038/nrc.2017.116

Nduom, E. K., Wei, J., Yaghi, N. K., Huang, N., Kong, L.-Y., Gabrusiewicz, K., et al. (2016). PD-L1 Expression and Prognostic Impact in Glioblastoma. *Neuro Oncol.* 18, 195–205. doi:10.1093/neuonc/nov172

Obayashi, T., Kagaya, Y., Aoki, Y., Tadaka, S., and Kinoshita, K. (2019). COXPRESdb V7: a Gene Coexpression Database for 11 Animal Species Supported by 23 Coexpression Platforms for Technical Evaluation and Evolutionary Inference. *Nucleic Acids Res.* 47, D55–D62. doi:10.1093/nar/gky1155

Ostrom, Q. T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J. S. (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. *Neuro Oncol.* 20, iv1–iv86. doi:10.1093/neuonc/noy131

Pan, H.-W., Chou, H.-Y. E., Liu, S.-H., Peng, S.-Y., Liu, C.-L., and Hsu, H.-C. (2006). Role of L2DTL, Cell Cycle-Regulated Nuclear and Centrosome Protein, in Aggressive HepatocellularCarcinoma. *Cell Cycle* 5, 2676–2687. doi:10.4161/cc.5.22.3500

Pang, F.-M., Yan, H., Mo, J.-L., Li, D., Chen, Y., Zhang, L., et al. (2020). Integrative Analyses Identify a DNA Damage Repair Gene Signature for Prognosis Prediction in Lower Grade Gliomas. *Future Oncol.* 16, 367–382. doi:10.2217/fon-2019-0764

Qian, X., Song, X., He, Y., Yang, Z., Sun, T., Wang, J., et al. (2015). CCNB2 Overexpression Is a Poor Prognostic Biomarker in Chinese NSCLC Patients. *Biomed. Pharmacother.* 74, 222–227. doi:10.1016/j.biopha.2015.08.004

Qiu, N., He, Y., Zhang, S., Hu, X., Chen, M., and Li, H. (2018). Cullin 7 Is a Predictor of Poor Prognosis in Breast Cancer Patients and Is Involved in the

Proliferation and Invasion of Breast Cancer Cells by Regulating the Cell Cycle and Microtubule Stability. *Oncol. Rep.* 39, 603–610. doi:10.3892/or.2017.6106

Renkonen, S., Lee, M., Mäkitie, A., Lindström, L. S., and Czene, K. (2017). Site-specific Familial Risk and Survival of Familial and Sporadic Head and Neck Cancer. *Int. J. Cancer* 141, 497–502. doi:10.1002/ijc.30751

Rubio-Perez, C., Tamborero, D., Schroeder, M. P., Antolín, A. A., Deu-Pons, J., Perez-Llamas, C., et al. (2015). In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell* 27, 382–396. doi:10.1016/j.ccell.2015.02.007

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, 1650–1658. doi:10.2353/ajpath.2007.070338

Squatrito, M., and Holland, E. C. (2011). DNA Damage Response and Growth Factor Signaling Pathways in Gliomagenesis and Therapeutic Resistance: Figure 1. *Cancer Res.* 71, 5945–5949. doi:10.1158/0008-5472.can-11-1245

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

van Dam, S., Cordeiro, R., Craig, T., van Dam, J., Wood, S. H., and de Magalhães, J. P. (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

Waker, C. A., and Lober, R. M. (2019). Brain Tumors of Glial Origin. *Adv. Exp. Med. Biol.* 1190, 281–297. doi:10.1007/978-981-32-9636-7_18

Wang, S., Chen, B., Zhu, Z., Zhang, L., Zeng, J., Xu, G., et al. (2018). CDC20 Overexpression Leads to Poor Prognosis in Solid Tumors. *Medicine (Baltimore)* 97, e13832. doi:10.1097/md.0000000000013832

Williams, G. H., and Stoeber, K. (2012). The Cell Cycle and Cancer. *J. Pathol.* 226, 352–364. doi:10.1002/path.3022

Yan, M., Tang, C., Ma, Z., Huang, S., and Dong, Z. (2016). DNA Damage Response in Nephrotoxic and Ischemic Kidney Injury. *Toxicol. Appl. Pharmacol.* 313, 104–108. doi:10.1016/j.taap.2016.10.022

Zhao, J., Liu, Y., Zhang, W., Zhou, Z., Wu, J., Cui, P., et al. (2015). Long Non-coding RNA Linc00152 Is Involved in Cell Cycle Arrest, Apoptosis, Epithelial to Mesenchymal Transition, Cell Migration and Invasion in Gastric Cancer. *Cell Cycle* 14, 3112–3123. doi:10.1080/15384101.2015.1078034

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, Tan, Lv, Wang, Yu, Yang, Liang, Lu, Ji, Chen, He, Chen, Chen and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fgene.2021.764270

Comprehensive Analysis of Inhibitor of Apoptosis Protein Expression and Prognostic Significance in Non–Small Cell Lung Cancer

Jun Liu[1], *Yi Lu*[1], *Wenan Huang*[1] *and Zhibo He*[2]*

[1]*Medical College, Jiujiang University, Jiujiang, China,* [2]*School of Literature and Communication, Jiujiang University, Jiujiang, China*

Inhibitors of apoptosis proteins (IAPs) have been associated with tumor development and progression by affecting apoptosis through cell death signaling pathways. To date, eight IAPs (BIRC1–8) have been identified in mammalian cells. However, the role of IAPs in non–small cell lung cancer (NSCLC) development and progression has not been explored in depth. In this study, we used public datasets and bioinformatics tools to compare the expression, prognostic significance, and function of IAPs in NSCLC and its subtypes. Expression of IAPs in cancer and normal tissues and at different stages of NSCLC was compared with gene expression profiling interactive analysis, and their prognostic significance was analyzed with the Kaplan–Meier Plotter database. The correlations among IAPs were analyzed with the STRING database and SPSS19.0. Functional annotation of IAPs was analyzed by Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment on the basis of the DAVID tool. Among patients with lung adenocarcinoma (LUAD), the expression level of *BIRC5* was higher than that in normal samples, and the expression of *BIRC1* and *BIRC5* significantly varied in different stages. Moreover, the *BIRC1–3* and *BIRC5* mRNA levels were associated with overall survival (OS), and the *BIRC1–2* and *BIRC5–6* mRNA levels were associated with progression-free survival (PFS). Among patients with lung squamous cell carcinoma (LUSC), the expression level of *BIRC1* was lower and that of *BIRC5* was higher than those in normal tissues, and *BIRC5* expression significantly varied in different stages. *BIRC1* expression was associated with OS, whereas *BIRC2* and *BIRC6* expression was associated with PFS. Enrichment analysis showed that most IAPs are associated with ubiquitin- and apoptosis-related pathways. Collectively, this study suggests *BIRC5* as a potential diagnostic and staging marker, *BIRC1* as a potential marker of OS, and *BIRC2* and *BIRC6* as potential PFS markers for patients with NSCLC. These highlight new targets for the early detection, treatment, and management of NSCLC.

Keywords: IAPS, LUAD, LUSC, diagnose biomarker, clinical stages, prognostic values, correlationship

OPEN ACCESS

Edited by:
Jing Zhao,
Chongqing Medical University, China

Reviewed by:
Laura Mondragón Martínez,
Josep Carreras Leukaemia Research Institute (IJC), Spain
Nicola Bougen-Zhukov,
University of Otago, New Zealand

***Correspondence:**
Zhibo He
6040109@jju.edu.cn

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

Received: *25 August 2021*
Accepted: *01 November 2021*
Published: *02 December 2021*

Citation:
Liu J, Lu Y, Huang W and He Z (2021) Comprehensive Analysis of Inhibitor of Apoptosis Protein Expression and Prognostic Significance in Non–Small Cell Lung Cancer. Front. Genet. 12:764270. doi: 10.3389/fgene.2021.764270

INTRODUCTION

Non–small cell lung cancer (NSCLC) has one of the highest mortality rates among malignant tumors globally, which accounts for approximately 80% of all lung cancers (Siegel et al., 2020). The two predominant histological phenotypes of NSCLC are lung adenocarcinoma (LUAD,~50% of cases) and lung squamous cell carcinoma (LUSC,~40% of cases) (Davidson et al., 2013; Langer et al., 2015). Unfortunately, currently, available biomarkers mainly reflect sex and age variations (Tsao et al., 2012) but cannot accurately identify the stage or prognosis of a tumor. Consequently, NSCLC remains difficult to detect at an early stage, and most patients are commonly diagnosed when the cancer has already progressed to an advanced stage (40% of NSCLC cases are diagnosed at stage IV) and thus not eligible for curative treatments. Therefore, the prognosis of NSCLC remains poor with a 5-year survival rate of only 2%–13% (Liu et al., 2020). Currently, the primary treatment of NSCLC is surgery, radiotherapy, and chemotherapy (Upadhya et al., 2021). Because of tumor heterogeneity, the current biomarkers used to predict NSCLC prognosis have some limitations; thus, it is necessary to explore new biomarkers as diagnostic and prognostic indicators to effectively improve survival and individualized treatment.

Inhibitors of apoptosis proteins (IAPs) are among the most extensively studied molecular and therapeutic targets in treating cancers, and their dysregulated expression has been reported in NSCLC (De-Xuan et al., 2017; Mazur et al., 2018). IAPs play essential roles in preventing apoptosis or programmed cell death. To date, eight IAPs have been identified in mammalian cells (BIRC1–8; see **Table 1**). The common feature of IAP family members is the presence of one or more baculoviral IAP repeats (Kumar et al., 2020). In addition to inhibiting apoptosis, IAPs play various biological roles, including regulation of innate immunity and inflammation, cell proliferation, cell migration, and apoptosis (Ji et al., 2018; Khan et al., 2021). Accordingly, IAPs act as pivotal regulators in oncogenesis by directly or indirectly affecting apoptosis through intrinsic and extrinsic cell death signaling pathways (Ji et al., 2018; Khan et al., 2021). Therefore, dysregulation of IAPs may lead cells toward cancerization (Yang and Wang, 2016; Yang et al., 2020; Zhang et al., 2021).

Downregulating *BIRC2* expression indirectly induces NSCLC cell apoptosis by preventing the formation of the caspase-8–activating platform (Yang and Wang, 2016; Jian et al., 2019). Moreover, the positive rates of *BIRC4* mRNA expression in pathological tissues of patients with NSCLC were reported to be significantly higher than those in the para-cancerous tissues (De-Xuan et al., 2017). *BIRC5* is strongly expressed in different types of tumors but is not expressed or is only expressed at low levels in most normal differentiated tissues (Xiao and Li, 2015; Mazur et al., 2018). These findings suggest a role of IAPs in NSCLC. However, the underlying mechanism and functions of IAPs in different subtypes of NSCLC or at different stages of cancer progression have yet to be fully elucidated.

RNA and DNA research, an essential component of biological and biomedical studies, has been revolutionized with the development of microarray technology, providing vast molecular data for comparative analysis. However, to the best of our knowledge, bioinformatics analysis of IAPs has yet to be applied for NSCLC. In this study, we comprehensively analyzed the expression of IAPs in patients with NSCLC using public datasets to determine their expression patterns, potential functions, and distinct prognostic value. This study can provide new insight into understanding the molecular mechanisms of IAPs in NSCLC toward development of drugs to inhibit aberrantly expressed IAPs that can help to induce apoptosis in cancerous cells. Moreover, exploring biomarker to diagnose lung cancer and distinguish stages is very necessary.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Academic Committee of Jiujiang University. All datasets were retrieved from the published literature, in which written informed consent from patients was confirmed for the individual studies.

IAPs Expression Analysis

The differential mRNA expression of IAPs between NSCLC and normal samples was evaluated separately for LUAD and LUSC with gene expression profiling interactive analysis (GEPIA2; http://gepia2.cancer-pku.cn/#index/). The "expression analysis" mode was selected, with each IAP (*BIRC1 BIRC2*, *BIRC3*, *BIRC4*, *BIRC5*, *BIRC6*, *BIRC7*, and *BIRC8*) added as input. LUAD and LUSC were selected as cancer types. Differentially expressed genes were selected according to a log2 fold change cutoff of 2

TABLE 1 | Inhibitor of apoptosis proteins information.

No	Gene symble	Gene ID	Also known as
1	NAIP	4,671	**BIRC1**; NLRB1; psiNAIP
2	BIRC2	329	API1; MIHB; HIAP2; RNF48; cIAP1; Hiap-2; c-IAP1
3	BIRC3	330	API2; MIHC; CIAP2; HAIP1; HIAP1; IAP-1; MALT2; RNF49; c-IAP2
4	XIAP	331	API3; ILP1; MIHA; XLP2; **BIRC4**; IAP-3; hIAP3; hIAP-3
5	BIRC5	332	API4; EPR-1
6	BIRC6	57,448	APOLLON; BRUCE
7	BIRC7	79,444	KIAP, LIVIN; ML-IAP; MLIAP; RNF50
8	BIRC8	112,401	ILP-2; ILP2; RNF136; hILP2

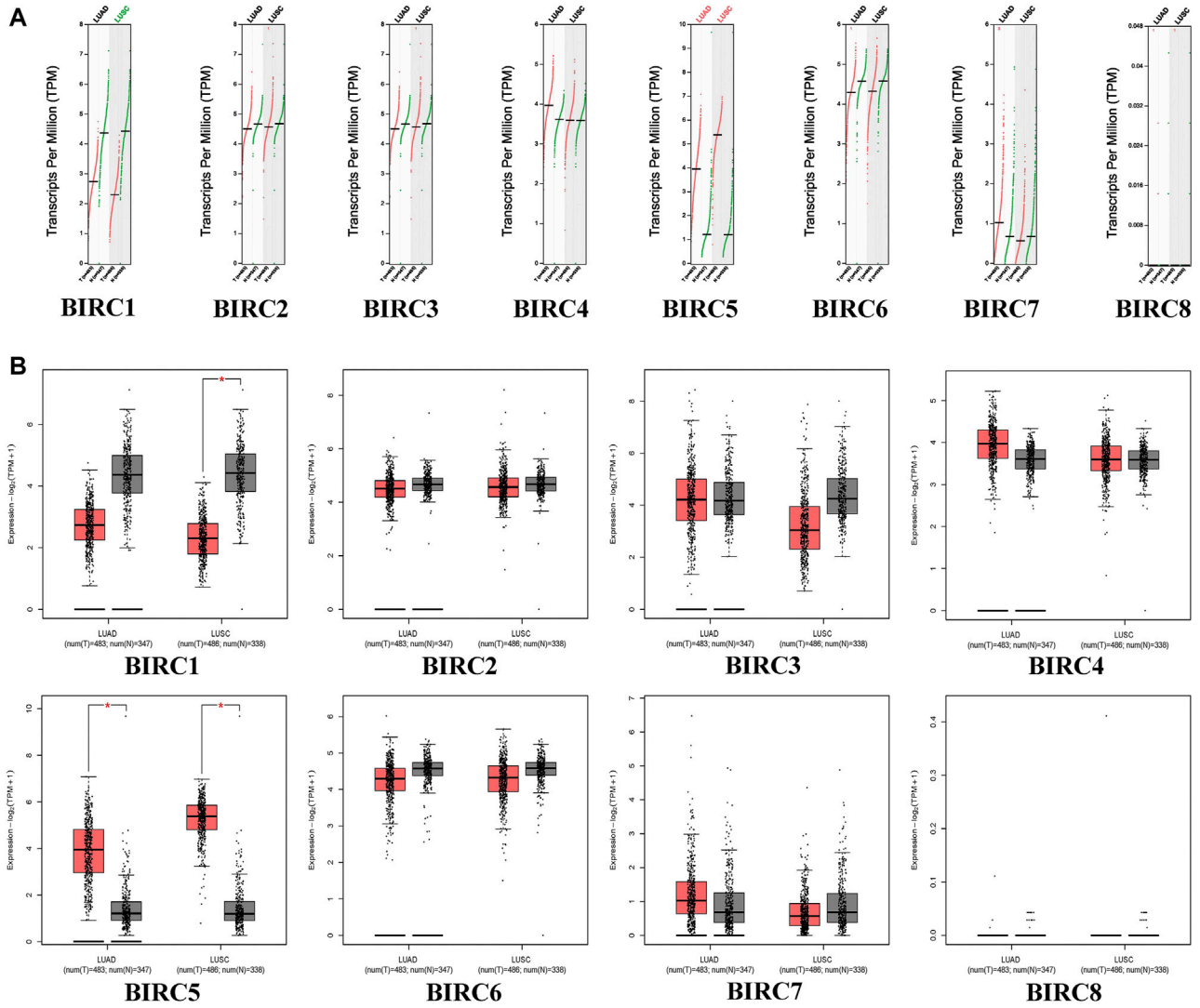


FIGURE 1 | mRNA expression of IAPs between NSCLC and normal lung tissues. **(A)** Scatter diagram of IAPs expression in NSCLC; red dots indicate tumor tissues and green dots indicate normal tissue. Green text indicates that the gene expression level in the tumor tissues was lower than that in normal tissues, and red text indicates that the gene expression level in tumor tissues was higher than that in normal tissues. **(B)** Box plot of IAPs expression in NSCLC; red boxes indicate tumor samples and the gray boxes indicate normal samples. *$p < 0.05$. NSCLC: non–small cell lung cancer; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma.

and q-value < 0.05. All other options were set to the default values.

Prognostic Significance of IAPs Expression in NSCLC

To further explore whether IAPs can be potential prognostic biomarkers in NSCLC, we evaluated the prognostic value of *BIRC1–7* mRNA expression in the survival of patients with LUAD and LUSC separately using the Kaplan–Meier Plotter database (https://kmplot.com/analysis/); data on *BIRC8* mRNA expression and survival of patients with LUAD and LUSC are lacking from the database. Patient samples were split into two groups according to the median expression level (high versus low expression). The Kaplan–Meier curve, hazard ratio with 95% confidence interval, and log-rank *p*-value were used to evaluate the relationship between the expression of each IAP and the overall survival (OS) or progression-free survival (PFS) of patients with NSCLC (LUAD and LUSC).

Construction of the IAPs Protein-Protein Interaction Network

The PPI network was constructed from the STRING database (https://string-db.org/), which includes data compiled from several sources. "BIRC1, BIRC2, BIRC3, BIRC4, BIRC5, BIRC6, BIRC7, and BIRC8" were input to the "multiple proteins" box with "*Homo sapiens*" selected as the organism. Other options were left as default options. Cytoscape 3.7.1

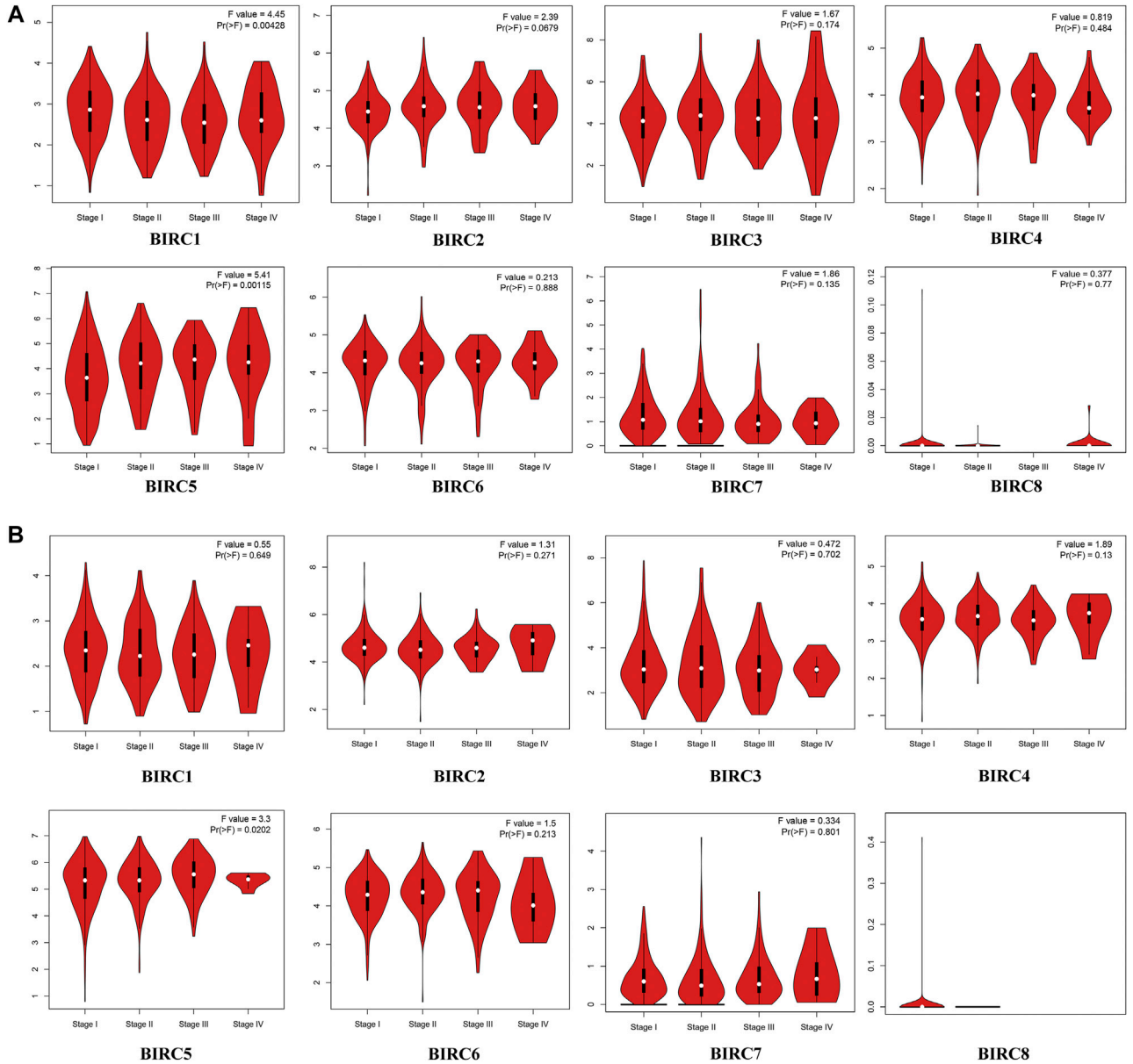


FIGURE 2 | mRNA Expression of IAPs in Different Stages of NSCLC. **(A)** Correlation between mRNA expression of IAPs and tumor stage in LUAD patients. **(B)** Correlation between mRNA expression of IAPs and tumor stage in LUSC Patients. *Pr* < 0.05 indicates that the gene expression differs across stages. NSCLC, non–small cell lung cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

software was used for construction of the PPI network and further visualization for analysis.

Correlations of IAP mRNA Levels in Patients With NSCLC

Gene Expression Omnibus profiles (https://www.ncbi.nlm.nih.gov/geoprofiles/?term=) were used to determine the correlations among expression levels of IAPs in NSCLC, using the keywords "BIRCX NSCLC" (where X refers to 1–8 for the eight IAPs), and each profile was obtained (https://www.ncbi.nlm.nih.gov/geoprofiles/62790008). Scatter plots were constructed and pairwise correlations between all IAPs were analyzed according to the Pearson correlation coefficient using SPSS 19.0 software; $p < 0.05$ was considered statistically significant.

Functional Enrichment Analysis of IAPs

The biological functions of IAPs were analyzed using Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways on the basis of the DAVID tool (https://david.ncifcrf.gov/). GO annotation enrichment analysis was conducted to identify the unique biological properties of

TABLE 2 | Correlation of IAPs mRNA expression and prognosis in NSCLC by Kaplan–Meier plotter.

IAPs name (LUAD*)	Overall survival		Progression-free survival		IAPs name (LUSC*)	Overall survival		Progression-free survival	
	HR	*p* value	HR	*p* value		HR	*p* value	HR	*p* value
BIRC1	0.61(0.48–0.78)	**5.8E-05**	0.56(0.41–0.77)	**0.00300**	BIRC1	0.78(0.61–0.99)	**0.0375**	0.81(0.48–1.35)	0.4167
BIRC2	0.61(0.48–0.77)	**0.00003**	0.71(0.52–0.97)	**0.03220**	BIRC2	1.08(0.85–1.37)	0.5178	1.74(1.03–2.94)	**0.0353**
BIRC3	0.66(0.52–0.83)	**0.00039**	0.73(0.53–1.00)	0.04870	BIRC3	1.08(0.85–1.37)	0.5111	0.94(0.56–1.56)	0.8036
BIRC4	1.00(0.79–1.26)	0.99730	1.33(0.97–1.82)	0.07180	BIRC4	0.91(0.72–1.15)	0.4217	1.26(0.75–2.10)	0.3801
BIRC5	2.42(1.90–3.09)	**2.2E-13**	3.13(2.23–4.40)	**4.0E-12**	BIRC5	0.99(0.78–1.25)	0.9072	0.98(0.59–1.64)	0.9436
BIRC6	0.87(0.68–1.10)	0.24910	0.55(0.40–0.77)	**0.00030**	BIRC6	1.36(1.00–1.86)	0.0504	1.88(1.11–3.20)	**0.0172**
BIRC7	1.23(0.97–1.55)	0.08360	0.89(0.65–1.22)	0.48775	BIRC7	1.01(0.79–1.28)	0.9608	1.09(0.65–1.82)	0.7470

**LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.*
Bold values indicate p < *0.05.*

IAPs, including biological processes, cellular components, and molecular functions. The top five terms were selected according to the *p* value. KEGG pathway enrichment analysis was performed to explore the key pathways of IAPs; $p < 0.05$ was considered statistically significant.

RESULTS

Transcriptional Levels of IAPs Are Altered in NSCLC

In the GEPIA dataset, *BIRC5* expression levels in LUAD and LUSC tissues were significantly higher than those in normal tissues, whereas the expression level of *BIRC1* was significantly lower in LUSC tissues than in the normal tissues (**Figure 1**). The expression of *BIRC1* and *BIRC5* significantly varied across LUAD stages, and the expression of *BIRC5* significantly varied across LUSC stages (**Figure 2**), suggesting that these IAPs may serve as potential biomarkers for diagnosis and cancer staging in NSCLC patients.

IAPs are Associated with the Prognosis of Patients with NSCLC

The Kaplan–Meier curve and associated statistical analyses revealed that decreased *BIRC1–3* mRNA levels and increased *BIRC5* mRNA levels were significantly associated with the OS, whereas the decreased *BIRC1-2* and *6* mRNA levels and the increased *BIRC5* mRNA levels were significantly associated with the PFS of patients with LUAD (**Table 2**; **Figure 3A**). A decreased *BIRC1* mRNA level was significantly associated with OS, whereas increased *BIRC2* and *BIRC6* mRNA levels were significantly associated with the PFS of the patients with LUSC (**Table 2**; **Figure 3B**).

PPI Network

The PPI network indicated that BIRC1 is co-expressed with BIRC6; BIRC2 is co-expressed with BIRC3, BIRC4, and BIRC6; BIRC3 is co-expressed with BIRC4; BIRC4 is co-expressed with BIRC6 and 7; BIRC5 is co-expressed with BIRC6; and BIRC6 is co-expressed with BIRC7 and 8 (**Figure 4A**). The interactions among these IAPs (except for BIRC5 with BIRC7 and BIRC8) have been experimentally validated. Overall, eight nodes formed a network of interactions with 19 edges (**Figure 4B**). The degree was greater than 4.75 for five nodes (average score): BIRC6, BIRC4, BIRC5, BIRC7, and BIRC2, from the highest to lowest (**Supplementary Table S1**). The combined score for 10 edges was greater than 0.686 (average score). The highest combined score was 0.971 based on the interaction of BIRC2 with BIRC4, followed by 0.943 (BIRC3 with BIRC4), 0.937 (BIRC2 with BIRC3), 0.827 (BIRC7 with BIRC8), 0.819 (BIRC6 with BIRC7), 0.779 (BIRC7 with BIRC4), 0.777 (BIRC2 with BIRC5), 0.771 (BIRC8 with BIRC4), 0.751 (BIRC5 with BIRC4), and 0.718 (BIRC3 with BIRC5) (**Supplementary Table S2**).

Correlations Among IAPs in NSCLC

In LUAD, *BIRC1* was positively correlated with *BIRC7*, *BIRC2* was positively correlated with *BIRC3* and *BIRC5*, and *BIRC3* was positively correlated with *BIRC5*. Significant and negative correlations were identified between the following IAPs in LUAD: *BIRC2* with *BIRC7*, *BIRC3* with *BIRC7*, and *BIRC5* with *BIRC7* (**Figure 5A** and **Table 3**). In LUSC, *BIRC1* was positively correlated with *BIRC3* and *BIRC3* was also positively correlated with *BIRC7*, whereas *BIRC5* was negatively correlated with *BIRC7* (**Figure 5B** and **Table 4**).

IAPs Play Roles in Apoptosis and Ubiquitination in NSCLC

GO enrichment analysis (**Figure 6A**, **Supplementary Table S3**) showed that the IAPs in NSCLC were significantly enriched in the biological process terms inhibition of cysteine-type endopeptidase activity involved in apoptotic process, mitotic spindle assembly, protein ubiquitination, negative regulation of apoptotic process, and apoptotic process terms; in the cellular component terms spindle microtubule, cytoplasm, nucleus, midbody, and membrane raft; and in the molecular function terms ubiquitin-protein transferase activity, cysteine-type endopeptidase inhibitor activity involved in apoptotic process, cysteine-type endopeptidase inhibitor activity, ligase activity, and zinc ion binding. More than half of the IAP members mainly participate in ubiquitin-protein transferase activity, protein ubiquitination, negative regulation of apoptotic process,

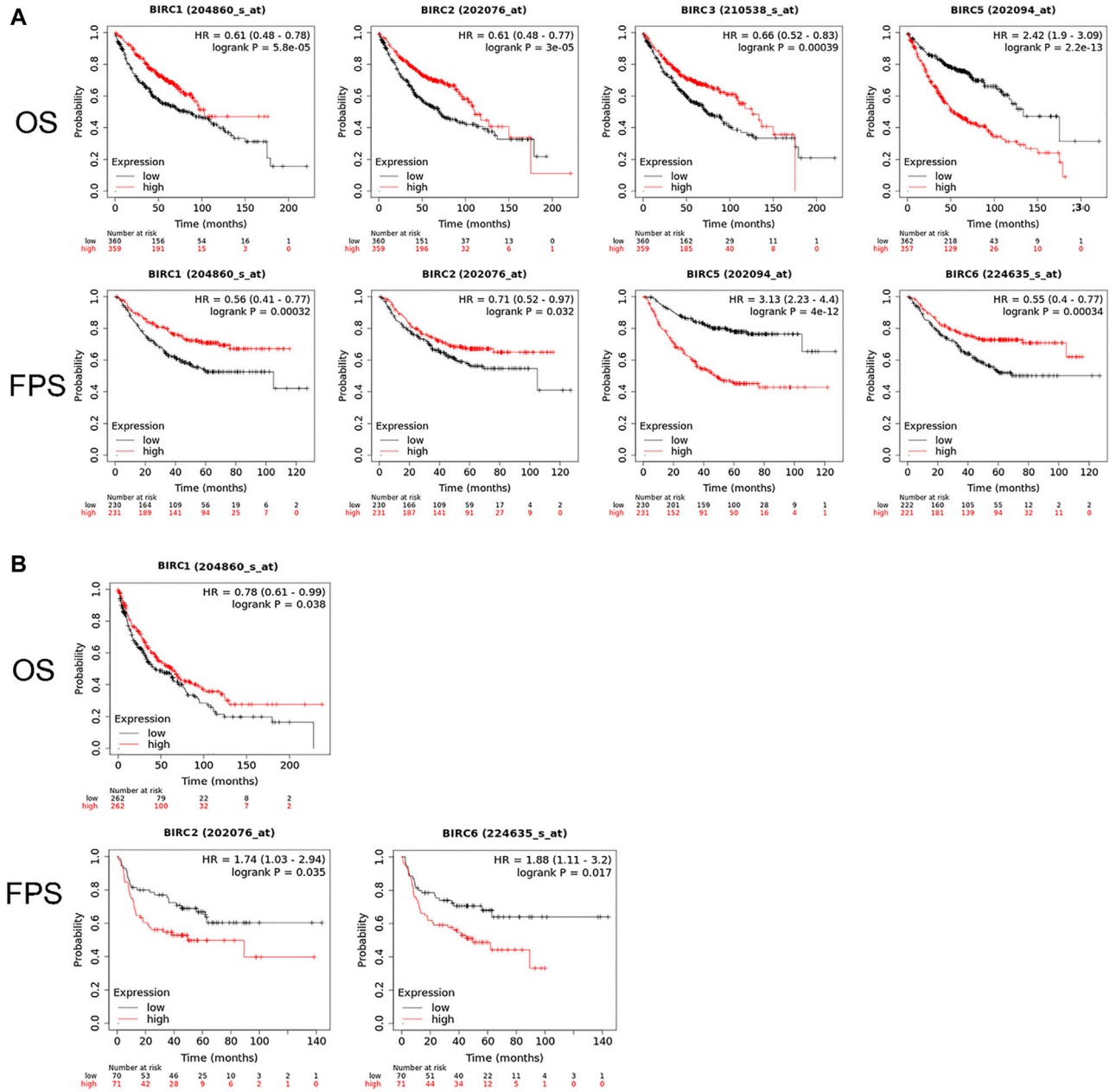


FIGURE 3 | Prognostic value of IAP mRNA expression in patients with NSCLC. **(A)** IAPs significantly associated with the prognosis of LUAD patients. **(B)** IAPs significantly associated with the prognosis of LUSC patients. NSCLC, non–small cell lung cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OS, overall survival; PFS, progression-free survival.

apoptotic process, cytoplasm, and inhibition of cysteine-type endopeptidase activity involved in apoptotic process, mitotic spindle assembly, spindle microtubule, nucleus, and zinc ion binding (**Figure 6B**).

KEGG pathway analysis (**Figure 7** and **Supplementary Table S4**) showed that the IAPs were most significantly enriched in ubiquitin-mediated proteolysis (p = 5.93E-08), followed by small cell lung cancer, toxoplasmosis, pathways in cancer, NOD-like receptor signaling pathway, NF-κB signaling pathway, focal adhesion, and apoptosis. More than half of the IAP members are mainly involved in the top four pathways (**Figure 7B**).

DISCUSSION

Although the role of IAPs in tumor development and progression has been partially confirmed in NSCLC, further bioinformatics analysis has yet to be performed (Rashed et al., 2019; Frazzi, 2021). This is the first study to comprehensively explore

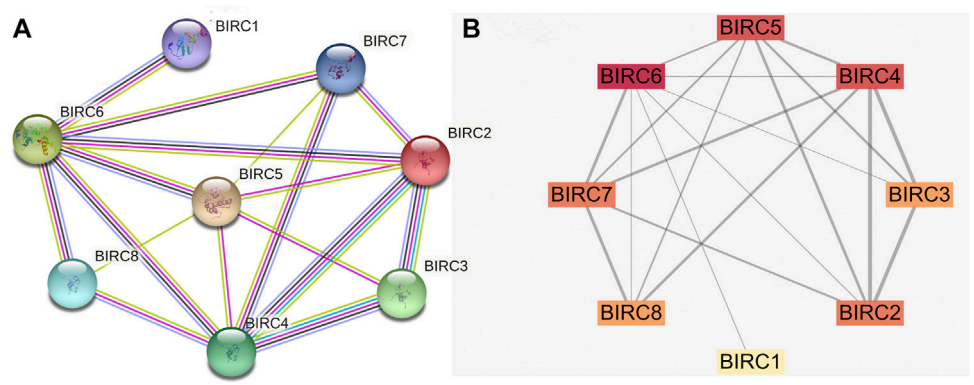


FIGURE 4 | Protein-protein interaction network of IAPs. **(A)** STRING analysis. Different colors of lines indicate a different source of evidence: light blue, curated databases; rose, experimentally determined; green, gene neighborhood; red, gene fusions; dark blue, gene co-occurrence; light green, text mining; black, co-expression; purple, protein homology. **(B)** Cytoscape analysis. The darker the color, the greater the degree; the wider the line, the stronger the interaction.

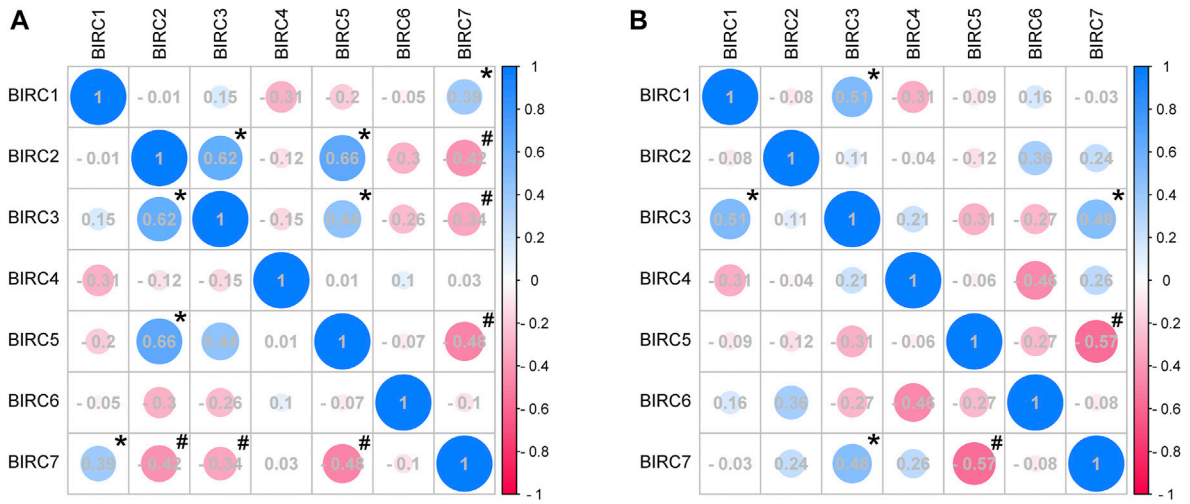


FIGURE 5 | Correlation analysis of mRNA expression of IAPs in NSCLC. **(A)** Correlation analysis of IAPs expression in LUAD. **(B)** Correlation analysis of IAPs expression in LUSC. NSCLC, non–small cell lung cancer; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma. Pink indicates a negative correlation and blue indicates a positive correlation; the darker the color, the large the Pearson correlation coefficient. * indicates a significantly positive relationship ($p < 0.05$); # indicates a significantly negative relationship ($p < 0.05$).

the transcriptional profiles, prognostic values, interactions, and functional enrichment of IAPs in different subtypes of NSCLC and across different tumor stages. Our findings can provide guidance for the development of IAPs as markers in the prevention, treatment, and prognosis for patients with NSCLC.

There has been no evidence of a role of *BIRC1* in NSCLC until now, and little has been reported of any association of *BIRC1* in cancer. A PTV-loaded nanocarrier was developed to trigger the apoptosis of glioblastoma multiforme cells by reducing the mRNA levels of *NFKB*, *IL6*, *BIRC1*, and *BIRC5* (Psc et al., 2021). Mig-6 exerts a tumor-suppressor function in murine endometrial cancer through downregulation of *BIRC1* expression (Kim et al., 2019). In our study, the GEPIA dataset revealed that the expression of *BIRC1* was lower in LUSC than that in normal tissues. Moreover, *BIRC1* mRNA expression was significantly different at least between two stages of LUAD and was associated with OS or PFS in patients with NSCLC. These data suggest that downregulation of *BIRC1* possibly plays a tumor-suppressor function in NSCLC development.

A previous study showed that *BIRC2* expression regulates the apoptosis and survival of NSCLC cells: downregulating *BIRC2* expression indirectly induces NSCLC cell apoptosis by preventing formation of the caspase-8–activating platform (Yang and Wang, 2016; Jian et al., 2019). The overexpression of *BIRC2*, regulated by Pellino-1, contributes to the oncogenesis of A549 and H1299 cells, which are both LUAD cell lines, and promotes cancer cell survival (Jeon et al., 2016; Xv et al., 2021). Consistently, we found that

TABLE 3 | IAPs correlations in lung adenocarcinoma.

Gene name	Statistical indicator	BIRC1	BIRC2	BIRC3	BIRC4	BIRC5	BIRC6	BIRC7
BIRC1	Pearson	1	–0.009	0.149	–0.306	–0.195	–0.05	**0.388**
	Sig		0.954	0.36	0.055	0.228	0.761	**0.013**
	N	40	40	40	40	40	40	**40**
BIRC2	Pearson	–0.009	1	**0.617**	–0.118	**0.659**	–0.3	**–0.416**
	Sig	0.954		**0**	0.47	**0**	0.06	**0.008**
	N	40	40	**40**	40	**40**	40	**40**
BIRC3	Pearson	0.149	**0.617**	1	–0.15	**0.44**	–0.257	**–0.343**
	Sig	0.36	**0**		0.356	**0.005**	0.109	**0.03**
	N	40	**40**	40	40	**40**	40	**40**
BIRC4	Pearson	–0.306	–0.118	–0.15	1	0.008	0.095	0.034
	Sig	0.055	0.47	0.356		0.96	0.559	0.835
	N	40	40	40	40	40	40	40
BIRC5	Pearson	–0.195	**0.659**	**0.44**	0.008	1	–0.072	**–0.475**
	Sig	0.228	**0**	**0.005**	0.96		0.659	**0.002**
	N	40	**40**	**40**	40	40	40	**40**
BIRC6	Pearson	–0.05	–0.3	–0.257	0.095	–0.072	1	–0.105
	Sig	0.761	0.06	0.109	0.559	0.659		0.52
	N	40	40	40	40	40	40	40
BIRC7	Pearson	**0.388**	**–0.416**	**–0.343**	0.034	**–0.475**	–0.105	1
	Sig	**0.013**	**0.008**	**0.03**	0.835	**0.002**	0.52	
	N	**40**	**40**	**40**	40	**40**	40	40

Bold values indicate $p < 0.05$.

TABLE 4 | IAPs correlations in lung squamous cell carcinoma.

Gene name	Statistical indicator	BIRC1	BIRC2	BIRC3	BIRC4	BIRC5	BIRC6	BIRC7
BIRC1	Pearson	1	–0.083	**0.51**	–0.314	–0.088	0.16	–0.027
	sig		0.742	**0.031**	0.204	0.729	0.525	0.915
	N	18	18	**18**	18	18	18	18
BIRC2	Pearson	–0.083	1	0.109	–0.04	–0.12	0.357	0.238
	sig	0.742		0.667	0.875	0.637	0.146	0.342
	N	18	18	18	18	18	18	18
BIRC3	Pearson	**0.51**	0.109	1	0.211	–0.307	–0.274	**0.48**
	sig	**0.031**	0.667		0.402	0.216	0.271	**0.044**
	N	**18**	18	18	18	18	18	**18**
BIRC4	Pearson	–0.314	–0.04	0.211	1	–0.055	–0.46	0.26
	sig	0.204	0.875	0.402		0.829	0.055	0.297
	N	18	18	18	18	18	18	18
BIRC5	Pearson	–0.088	–0.12	–0.307	–0.055	1	–0.267	**–0.568**
	sig	0.729	0.637	0.216	0.829		0.284	**0.014**
	N	18	18	18	18	18	18	**18**
BIRC6	Pearson	0.16	0.357	–0.274	–0.46	–0.267	1	–0.079
	sig	0.525	0.146	0.271	0.055	0.284		0.754
	N	18	18	18	18	18	18	18
BIRC7	Pearson	–0.027	0.238	**0.48**	0.26	**–0.568**	–0.079	1
	sig	0.915	0.342	**0.044**	0.297	**0.014**	0.754	
	N	18	18	**18**	18	**18**	18	18

Bold values indicate $p < 0.05$.

downregulated *BIRC2* expression was associated with the prolonged survival time of patients with LUAD.

In our study, only *BIRC3* expression was positively correlated with the OS of patients with LUAD. There is substantial evidence pointing to the pro-survival and anti-apoptotic roles of BIRC3 in cancer cells; however, not all data are consistent (Frazzi, 2021). An *in vitro* study showed that RNA-binding motif 10 overexpression inhibited the malignant behaviors of A549 and H1299 cells by inducing the expression of AKT2, BIRC3, and JUN (Guan et al., 2017). However, Dubois et al. (2019) reported that overexpression of *BIRC3* regulated by *RASSF1A* depletion decreased the rate of cancer cell apoptosis. Similarly, upregulation of *BIRC3* expression *via* Pellino-1 overexpression in A549 and H1299 cells promoted lung oncogenesis and survival, and *BIRC3* also demonstrated a strong positive correlation with Pellino-1 in human LUAD

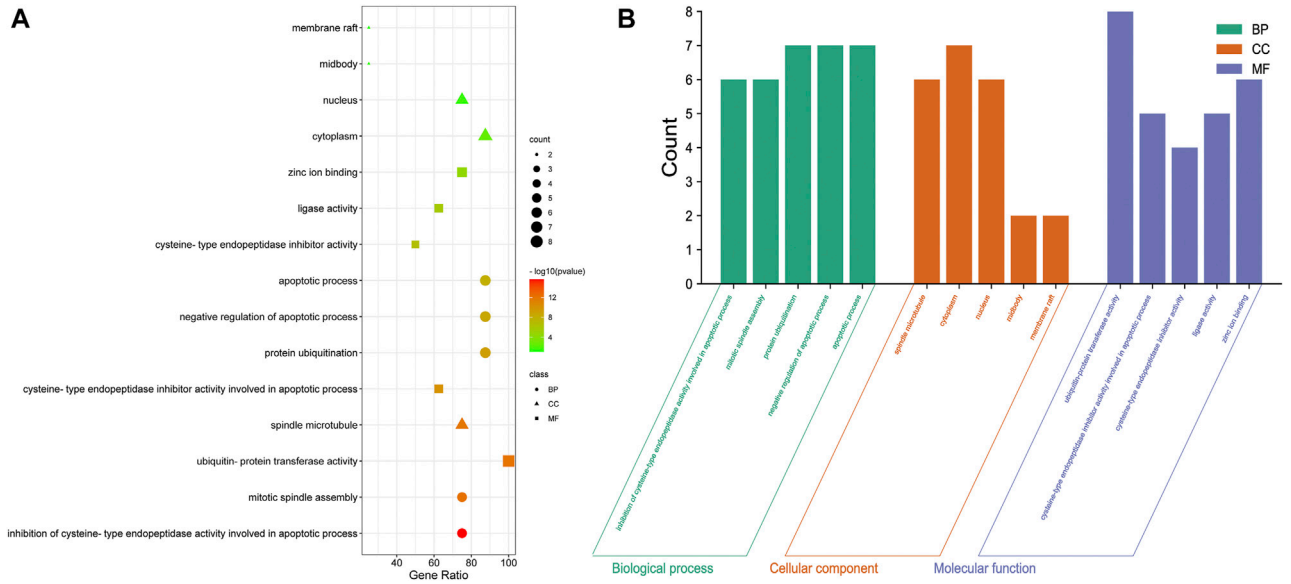


FIGURE 6 | GO analysis of IAPs. **(A)** Enrichment dot bubble diagram. **(B)** Enrichment histogram. GO, Gene Ontology; BP, biological processes; CC, cellular components; MF, molecular functions.

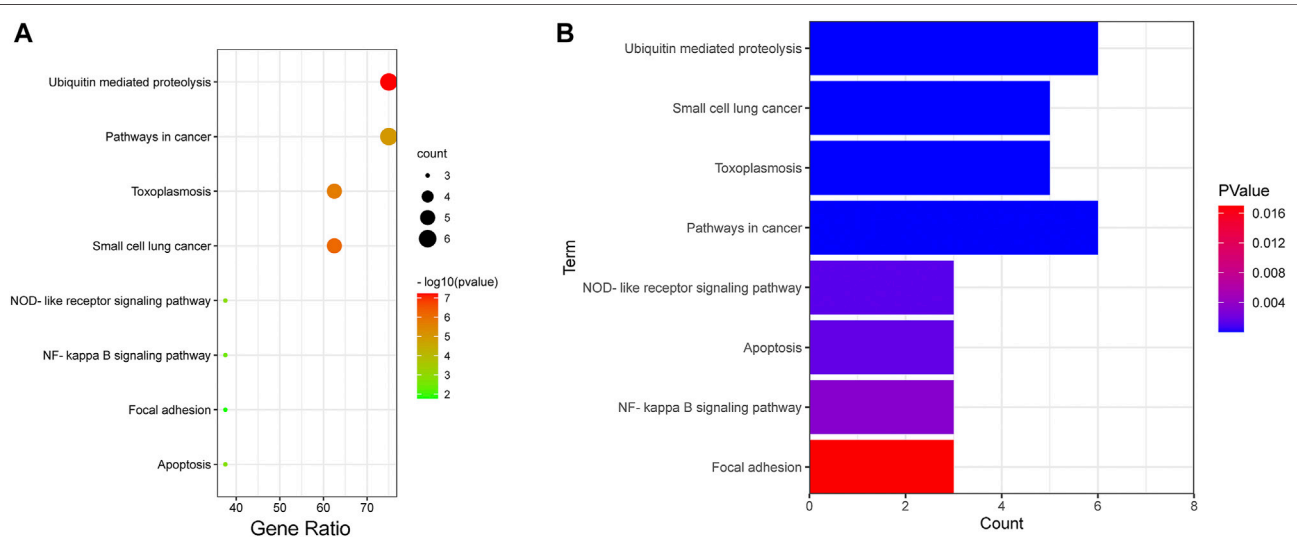


FIGURE 7 | KEGG analysis of IAPs. **(A)** Enrichment dot bubble diagram. **(B)** Enrichment histogram. KEGG: Kyoto Encyclopedia of Genes and Genomes.

tissues (Yang and Wang, 2016). Therefore, the mechanism of BIRC3 in cancer needs further study.

Surprisingly, we did not identify a specific role of *BIRC4* in the patients with LUAD or LUSC on the basis of the databases analyzed in this study. However, several *in vitro* studies have suggested an anti-NSCLC role of *BIRC4*. Hydrogen gas was suggested to promote the apoptosis of A549 cells by reducing the expression of *BIRC4* (Zhang et al., 2020a). Combined with other drugs in treating NSCLC *in vitro*, TRAIL induced cell apoptosis by inhibiting *BIRC4* expression and increasing cytotoxicity (Deok et al., 2018; Kim et al., 2019). Moreover, the positive rate of *BIRC4* mRNA expression in the pathological tissues of NSCLC patients was significantly higher than that in paracancerous tissues (De-Xuan et al., 2017). Although the expression of BIRC4 varies *in vitro* and *in vivo*, further *in vitro* experiments can represent an important starting point to better understand its regulation mechanisms and functions *in vivo*.

Unlike other IAPs, *BIRC5* is strongly expressed in most tumors but is not expressed or is expressed at only low levels in most normal differentiated tissues (Xiao and Li, 2015; Mazur et al., 2018). Consistently, we found that *BIRC5* expression levels were significantly higher in tumor tissues than in normal tissues. Previous studies have suggested *BIRC5* as a predictive biomarker in NSCLC, especially for LUAD (Zhang et al., 2020b; Haakensen et al., 2020). In addition, in the present study, BIRC5 emerged as

the most significant IAP that could be developed as a marker for preventing and treating NSCLC patients. Low expression of *BIRC5* mRNA was also previously positively correlated with NSCLC patient survival (Cao et al., 2019; Nitschkowski et al., 2019; Rashed et al., 2019; Zhang et al., 2020b).

BIRC6 has been suggested as a progression marker in NSCLC (Dong et al., 2013; Gharabaghi and Asadi, 2016), which was also associated with the PFS of the patients with NSCLC in our study. However, previous studies did not distinguish among different subtypes of NSCLC. Here, we show that *BIRC6* expression actually shows an opposite association with prognosis in patients with LUAD and LUSC: Increased *BIRC6* expression was significantly associated with the PFS of patients with LUAD, whereas decreased *BIRC6* was significantly associated with the PFS of patients with LUSC. This suggests that *BIRC6* is a potential biomarker for differentiating different types of NSCLC. No specific roles of *BIRC7* and *BIRC8* in NSCLC were identified in this study or in the literature to date.

From GO and KEGG enrichment analysis, we found that all eight members of the IAP family are enriched in ubiquitin-protein transferase activity, and most of them (six of eight) are enriched in ubiquitin-mediated proteolysis. The ubiquitin–proteasome system has become a key system of pathogenesis in several cancers (Senft et al., 2018). Thevebioside (an active ingredient from Traditional Chinese Medicine) was reported to inhibit the tumor growth of NSCLC through inhibiting SRC-3–mediated IGF-1R–PI3K-AKT signaling *via* ubiquitination to induce cellular apoptosis (Yao et al., 2020). In addition, deregulation of APC/C (a representative E3 ligase) together with its co-activators cell division cycle 20 (CDC20) or CDC20-like protein 1 (CDH1) has been associated with cancers (Jeon et al., 2016). Overexpression of Pellino-1 (an E3 ubiquitin ligase) is dependent on the expression of BIRC3 in human lung cancer cells, resulting in increased cell survival and colony forming ability (Jeon et al., 2016). SKP2 promotes programmed cell death protein 4 degradation through phosphorylation and ubiquitination, resulting in increased proliferation and radiation tolerance of breast cancer cells (Li et al., 2019). In addition, IAPs were also found to play a role in the NOD-like receptor signaling pathway and NF-κB signaling pathway (Mann and Oakley, 2005; Liu et al., 2019; Kumar et al., 2021). Thus, we speculate that the dysregulation of IAPs has more effective role in the inflammatory response.

CONCLUSION

In this study, we systematically analyzed the expression and prognostic value of IAPs in different subtypes of NSCLC, which can help to provide a more thorough understanding of the molecular biological properties of this cancer. Our results indicate that *BIRC1* and *BIRC5* are potential diagnostic markers for both LUAD and LUAC. *BIRC1*, *BIRC2*, and *BIRC5* are potential prognostic markers for LUAD, whereas *BIRC2* and *BIRC6* are prognostic markers for patients with NSCLC. From GO and KEGG enrichment analysis, we found that most IAP members are associated with ubiquitin and apoptosis. These highlight new targets for the early detection, treatment, and management of NSCLC.

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 Medical Ethics Review Committee of Jiujiang University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JL and ZH conceived and designed the research. JL performed gene expression profiling analysis. YL performed Kaplan–Meier survival curve analysis and PPI Network Construction. WH performed GO and KEGG enrichment analysis. ZH performed SPSS analysis. JL and ZH wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Science Foundation of Department of Education of Jiangxi Province (grant number GJJ201804), Science Foundation of Health Commisson of Jiangxi Province (grant number 202131076), Jiangxi Students' Platform for innovation and entrepreneurship training program (grant numbers s202111843037 and s202111843067).

ACKNOWLEDGMENTS

We would like to thank Bioinformatics (http://www.bioinformatics.com.cn/) for bioinformatics analysis. We would like to thank Editage (www.editage.cn) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.764270/full#supplementary-material

Supplementary Table S1 | The degree of IAPs from PPI network.

Supplementary Table S2 | The combined score of IAPs from PPI network.

Supplementary Table S3 | Go Function enrichment analysis of IAPs.

Supplementary Table S4 | KEGG pathway enrichment.

REFERENCES

Cao, Y., Zhu, W., Chen, W., Wu, J., Hou, G., and Li, Y. (2019). Prognostic Value of BIRC5 in Lung Adenocarcinoma Lacking EGFR, KRAS, and ALK Mutations by Integrated Bioinformatics Analysis. *Dis. markers* 2019 (3), 1–12. doi:10.1155/2019/5451290

Chen, J., Huang, X., Tao, C., Xiao, T., Li, X., Zeng, Q., et al. (2019). Artemether Attenuates the Progression of Non-small Cell Lung Cancer by Inducing Apoptosis, Cell Cycle Arrest and Promoting Cellular Senescence. *Biol. Pharm. Bull.* 42 (10), 1720–1725. doi:10.1248/bpb.b19-00391

De-Xuan, L. I., Ya-Shan, L. I., Wang, J., Zhang, B. J., Hong, Y., Zhang, Q. Y., et al. (2017). Expression and Clinical Significance of Survivin, Skp2 and XIAP mRNA in Non-small Cell Lung Cancer. *Prog. Mod. Biomed.* 17 (31), 6159–6162. doi:10.13241/j.cnki.pmb.2017.31.037

Deok, A., Hyo, L., Jisung, H., Han, H., Bonglee, K., Bumsang, S., et al. (2018). Lambertianic Acid Sensitizes Non-small Cell Lung Cancers to TRAIL-Induced Apoptosis via Inhibition of XIAP/NF-κB and Activation of Caspases and Death Receptor 4. *Int. J. Mol. ences* 19 (5), 1476. doi:10.3390/ijms19051476

Dong, X., Lin, D., Low, C., Vucic, E. A., English, J. C., Yee, J., et al. (2013). Elevated Expression of BIRC6 Protein in Non-small-cell Lung Cancers Is Associated with Cancer Recurrence and Chemoresistance. *J. Thorac. Oncol.* 8 (2), 161–170. doi:10.1097/jto.0b013e31827d5237

Dubois, F., Keller, M., Hoflack, J., Maille, E., Antoine, M., Westeel, V., et al. (2019). Role of the YAP-1 Transcriptional Target cIAP2 in the Differential Susceptibility to Chemotherapy of Non-small-cell Lung Cancer (NSCLC) Patients with Tumor RASSF1A Gene Methylation from the Phase 3 IFCT-0002 Trial. *Cancers* 11 (12), 1835. doi:10.3390/cancers11121835

Frazzi, R. (2021). BIRC3 and BIRC5: Multi-faceted Inhibitors in Cancer. *Cell Biosci.* 11 (1), 8. doi:10.1186/s13578-020-00521-0

Gharabaghi, M. A., and Asadi, M. (2016). Diagnostic Investigation of BIRC6 and SIRT1 Protein Expression Level as Potential Prognostic Biomarkers in Patients with Non-small Cell Lung Cancer. *Clin. Respir. J.* 12 (2), 633–638. doi:10.1111/crj.12572

Guan, G., Li, R., Tang, W., Liu, T., Su, Z., Wang, Y., et al. (2017). Expression of RNA-Binding Motif 10 Is Associated with Advanced Tumor Stage and Malignant Behaviors of Lung Adenocarcinoma Cancer Cells. *Tumour Biol.* 39 (3), 1010428317691740. doi:10.1177/1010428317691740

Haakensen, V. D., Khadse, A., Sandhu, V., Halvorsen, A. R., Solberg, S. K., Jørgensen, L. H., et al. (2020). Molecular Characterisation of TP53 Mutated Squamous Cell Carcinomas of the Lung to Identify Putative Targets for Therapy. *Int. J. Cancer* 147 (10), 2957–2966. doi:10.1002/ijc.33121

Jeon, Y. K., Kim, C. K., Koh, J., Chung, D. H., and Ha, G.-H. (2016). Pellino-1 Confers Chemoresistance in Lung Cancer Cells by Upregulating cIAP2 through Lys63-Mediated Polyubiquitination. *Oncotarget* 7 (27), 41811–41824. doi:10.18632/oncotarget.9619

Ji, J., Yu, Y., Li, Z.-L., Chen, M.-Y., Deng, R., Huang, X., et al. (2018). XIAP Limits Autophagic Degradation of Sox2 and Is A Therapeutic Target in Nasopharyngeal Carcinoma Stem Cells. *Theranostics* 8 (6), 1494–1510. doi:10.7150/thno.21717

Khan, S. A., Burke, M., Zhu, F., Yang, D.-H., Dubyk, C., Mehra, R., et al. (2021). Survivin Expression and Impact on Head and Neck Cancer Outcomes. *Oral Oncol.* 112, 105049. doi:10.1016/j.oraloncology.2020.105049

Kim, Y. H., Shin, E. A., Jung, J. H., Park, J. E., Koo, J., Koo, J. I., et al. (2019). Galbanic Acid Potentiates TRAIL Induced Apoptosis in Resistant Non-small Cell Lung Cancer Cells via Inhibition of MDR1 and Activation of Caspases and DR5. *Eur. J. Pharmacol.* 847, 91–96. doi:10.1016/j.ejphar.2019.01.028

Kumar, S., Nandi, A., Singh, S., Regulapati, R., Li, N., Tobias, J. W., et al. (2021). Dll1+ Quiescent Tumor Stem Cells Drive Chemoresistance in Breast Cancer through NF-Kb Survival Pathway. *Nat. Commun.* 12 (1), 432. doi:10.1038/s41467-020-20664-5

Li, C., Du, L., Ren, Y., Liu, X., Jiao, Q., Cui, D., et al. (2019). SKP2 Promotes Breast Cancer Tumorigenesis and Radiation Tolerance through PDCD4 Ubiquitination. *J. Exp. Clin. Cancer Res.* 38 (1), 76. doi:10.1186/s13046-019-1069-3

Liu, B., Ricarte Filho, J., Mallisetty, A., Villani, C., Kottorou, A., Rodgers, K., et al. (2020). Detection of Promoter DNA Methylation in Urine and Plasma Aids the Detection of Non-small Cell Lung Cancer. *Clin. Cancer Res.* 26 (16), 4339–4348. doi:10.1158/1078-0432.ccr-19-2896

Liu, P., Lu, Z., Liu, L., Li, R., Liang, Z., Shen, M., et al. (2019). NOD-like Receptor Signaling in Inflammation-Associated Cancers: From Functions to Targeted Therapies. *Phytomedicine* 64, 152925. doi:10.1016/j.phymed.2019.152925

Mann, D. A., and Oakley, F. (2005). NF-κB: a Signal for Cancer. *J. Hepatol.* 42 (4), 610–611. doi:10.1016/j.jhep.2005.01.007

Mazur, J., Roy, K., and Kanwar, J. R. (2018). Recent Advances in Nanomedicine and Survivin Targeting in Brain Cancers. *Nanomedicine* 13 (1), 105–137. doi:10.2217/nnm-2017-0286

Nitschkowski, D., Marwitz, S., Kotanidou, S. A., Reck, M., Kugler, C., Rabe, K. F., et al. (2019). Live and let die: epigenetic modifications of Survivin and Regucalcin in non-small cell lung cancer tissues contribute to malignancy. *Clin. Epigenet* 11 (1), 157. doi:10.1186/s13148-019-0770-6

Psc, A., Mk, B., Amc, C., As, B., and Dda, D. (2021). Multifunctional Silica-Coated Mixed Polymeric Micelles for Integrin-Targeted Therapy of Pediatric Patient-Derived Glioblastoma. *Mater. Sci. Eng. C* 128, 112261.

Rashed, R. A., Rahouma, M., Abo Elfetouh, R., Aziz, H., and Kamel, M. K. (2019). Effect of Serum Survivin on Survival Among Non-small Cell Lung Cancer Patients: NCI Experience. *Ann. Oncol.* 30, v51. doi:10.1093/annonc/mdz239.072

Senft, D., Qi, J., and Ronai, Z. e. A. (2018). Ubiquitin Ligases in Oncogenic Transformation and Cancer Therapy. *Nat. Rev. Cancer* 18 (2), 69–88. doi:10.1038/nrc.2017.105

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

Upadhya, A., Yadav, K. S., and Misra, A. (2021). Targeted Drug Therapy in Non-small Cell Lung Cancer: Clinical Significance and Possible Solutions-Part I. *Expert Opin. Drug Deliv.* 18 (1), 73–102. doi:10.1080/17425247.2021.1825377

Xiao, M., and Li, W. (2015). Recent Advances on Small-Molecule Survivin Inhibitors. *Curr. Med. Chem.* doi:10.2174/0929867322666150114102146

Xv, F., Liu, W., Chen, Y., Fang, H., and Zuo, J. (2021). miR-623 Targets DNM2 to Regulate the Apoptosis of Non-small Cell Lung Cancer Cell A549. *Int. J. Respiration* 41 (10), 771–776. Available at https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDZHYX&filename=GWHX202110009&uniplatform=NZKPT&v=HnAahxUC1PVMULeG3ei6y8TMPsoZqOr2vmEqExwGNL1kGi_5pQWuDYbgyXowxzuk

Yang, C., and Wang, H. (2016). LCL161 Increases Paclitaxel-Induced Apoptosis by Degrading cIAP1 and cIAP2 in NSCLC. *J. Exp. Clin. Cancer Res. CR* 35 (1), 158. doi:10.1186/s13046-016-0435-7

Yao, C., Su, L., Zhang, F., Zhu, X., and Xu, Z. (2020). Thevebioside, the Active Ingredient of Traditional Chinese Medicine, Promotes Ubiquitin-Mediated SRC-3 Degradation to Induce NSCLC Cells Apoptosis. *Cancer Lett.* 493, 167–177. doi:10.1016/j.canlet.2020.08.011

Zhang, Y., Chen, G., Yan, F. Z., Wang, F. L., and Wang, C. D. (2020a). Mechanism of Hydrogen Gas Promoted Apoptosis of Lung Adenocarcinoma A549 Cells through XIAP and BIRC3. Preprint. doi:10.21203/rs.3.rs-111317/v1

Zhang, Y., Sun, Y., Jia, Y., Zhang, Q., Zhu, P., and Ma, X. (2020b). α5-nAChR and Survivin: Two Potential Biological Targets in Lung Adenocarcinoma. *J. Cell Physiol.* 236 (3), 1787–1797. doi:10.1002/jcp.29956

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, Lu, Huang 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 Genetics

ORIGINAL RESEARCH
published: 13 December 2021
doi: 10.3389/fgene.2021.768140

High Expression of CEMIP Correlates Poor Prognosis and the Tumur Microenvironment in Breast Cancer as a Promisingly Prognostic Biomarker

Xingxing Dong†, *Yalong Yang*†, *Qianqian Yuan*, *Jinxuan Hou** and *Gaosong Wu**

Department of Thyroid and Breast Surgery, Zhongnan Hospital of Wuhan University, Wuhan, China

Cell migration-inducing hyaluronidase 1 (CEMIP), a Wnt-related protein and also known as KIAA1199, is implicated in the process of metastatic colonization in a variety of malignant tumors, including breast cancer (BC), which is one of the most frequently diagnosed tumors in women worldwide. In this study, multiple public databases, online analytical tools, and bioinformatics approaches were applied to explore the expression levels, regulatory mechanisms, and biological functions of CEMIP in BC. We illustrated that *CEMIP* was highly expressed in various kinds of carcinomas, including BC, especially advanced subtypes, and predicted less favorable prognosis (negatively associated with overall survival) in BC patients, which might be an independent prognostic factor. Then, we revealed that the mutation and high expression of *CEMIP* might lead to it as an oncogene. We also demonstrated that *TP53* mutation, DNA hypo-methylation, and the expression changes of three potential upstream transcription factors (*EZH2, EGR1*, and *JUN*) of *CEMIP* were likely to cause the hyperexpression of *CEMIP* in BC. Moreover, our findings suggested that CEMIP might exert its carcinogenic roles in the tumor microenvironment via participation in the extracellular matrix formation, increasing cancer-associated fibroblast (CAF), M2 macrophage, and neutrophil infiltration and decreasing $CD8^+$ T cell infiltration. In summary, our study provided more solid evidence for CEMIP as a prognostic and metastatic biomarker and a potential therapeutic target in BC. Of course, these findings also need more confirmations of basic experiments and further clinical trials in the future.

Keywords: breast cancer, CEMIP, prognosis, tumor microenvironment, biomarker

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Guo Zhao,
Henan University, China
Peijie Wu,
Chengdu University of Traditional Chinese Medicine, China
Lu Zhang,
Henan University, China

****Correspondence:***
Jinxuan Hou
jhou@whu.edu.cn
Gaosong Wu
wugaosongtj@163.com

†*These authors have contributed equally to this work*

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

Received: *31 August 2021*
Accepted: *18 November 2021*
Published: *13 December 2021*

Citation:
Dong X, Yang Y, Yuan Q, Hou J and Wu G (2021) High Expression of CEMIP Correlates Poor Prognosis and the Tumur Microenvironment in Breast Cancer as a Promisingly Prognostic Biomarker. Front. Genet. 12:768140. doi: 10.3389/fgene.2021.768140

INTRODUCTION

Cancer of the female breast is the leading malignancy worldwide with the highest incidence of 11.7% and the fourth in mortality (Sung et al., 2021). Some known risk factors contribute to the development of breast cancer (BC), for instance, gene mutations and lactation deficiency (Harbeck et al., 2019). The genetic predisposition of *BRCA1* and *BRCA2* mutations inherited from the family has been widely accepted. For sporadic BC, the advanced maternal age for a first pregnancy, early menarche, lack of breastfeeding, and late-onset menopause are recognized as risk factors. Besides, some modifiable risk factors, like obesity, physical inactivity, and alcohol use should also be noted (Harbeck et al., 2019). However, the exact mechanism of BC is still ambiguous. BC is highly heterogeneous (Aleskandarany et al., 2018) up to now, which can be divided into four major

subtypes according to molecular biomarkers (ER/PR, HER2, and Ki$^-$67) in clinical. In detail, they are luminal A (ER$^+$ and/or PR$^+$, HER2$^-$, Ki-67$^-$), luminal B (ER$^+$ and/or PR$^+$, HER2$^+$, Ki-67$^+$), HER2$^+$ (ER$^-$ and PR$^-$, HER2$^+$), and basal or triple-negative (ER$^-$, PR$^-$, and HER2$^-$) (Dai et al., 2015). The clinical treatment and prognosis of BC patients vary depending on these subtypes. Based on the worldwide clinical retrospective analysis, patients with luminal BC have the best outcomes with surgery, endocrine therapy, and chemotherapy; the HER2$^+$ second and the patients with triple-negative BC (TNBC) usually have the worst prognosis due to limited therapeutic options (Waks and Winer, 2019). Of course, other prognostic factors like age, stage, tumor grade, and lymphovascular status should be taken into consideration as well (Harbeck et al., 2019). Fortunately, the PD-L1 inhibition as a single-molecule target for therapy has been proved to ameliorate the progression-free survival (PFS) in TNBC patients, and CDK4/6 inhibitors substantially improve PFS of BC patients with endocrine resistance in recent years, although not for all BC patients (Schmid et al., 2018; Turner et al., 2018; Harbeck et al., 2019). Thereby, more effective therapeutic targets are to be explored urgently.

In recent decades, with the rise of high-throughput technologies and next-generation sequencing (NGS), followed by bioinformatics approaches and a variety of online analysis tools applied, the search for the key targeted genes in tumor genesis and progress has become a trend in tumor research in an attempt to reveal the mechanism of tumor genesis and development and thus to provide more precise treatment for patients and improve their prognosis. For instance, by applying these biological analysis tools, Lou et al. (2021) elucidated that *SEMA3F* was associated with poor prognosis and tumor immune infiltration of hepatocellular carcinoma, mediated by the TMPO-AS1/SNHG16-let-7c-5p axis. Cui et al. (2020) carried out a comprehensive pan-cancer analysis of the oncogenic role of *SND1* in human tumors, and Zeng et al. (2019) evaluated the potential of CXC chemokines as therapeutic targets and prognostic biomarkers in renal cell carcinoma.

Cell migration-inducing protein (*CEMIP*), previously known as *KIAA1199* or hyaluronan binding protein (*HYBID*), included in the Human Unidentified Gene-Encoded (HUGE) large protein database, is located on human chromosome 15q25.1 that encodes a 153 kDa protein which contains two GG domains and a special G8 domain (Li et al., 2017; Liu et al., 2021a). CEMIP is a kind of secreted protein, identified as an inner ear-specific protein at first, and mutations in it are related to non-syndromic hearing loss. In the past decades, an increasing number of studies have revealed that high expression of *CEMIP* promotes numerous malignancy progresses and metastasis and predicts poor prognosis of cancer patients, including breast (Jami et al., 2014), colorectal (Fink et al., 2015), liver, gastric, pancreatic, lung (Li et al., 2020a), prostate, cholangitis (Zhai et al., 2020), ovarian (Shen et al., 2019), and papillary thyroid cancers (Liu et al., 2021a). On the other hand, accumulated cell- and animal-based evidence shows that the over-expression of *CEMIP* could enhance proliferation, survival (Michishita et al., 2006), adhesion, motility, invasiveness, and epithelial-to-mesenchymal transition (EMT) (Liu et al., 2021b) of various cancer cells. Some researchers have elucidated that CEMIP is involved in Wnt/β-catenin, MEK/ERK, and PI3K/Akt signal pathways that contribute to the tumorigenesis (Liu et al., 2021a). A recent study reported that CEMIP can accelerate BC cell proliferation and migration by activating the STAT3 pathway (Chen et al., 2021). It is widely acknowledged that the expression of *CEMIP* is regulated by both genetic and epigenetic mechanisms. The two key transcription factors are nuclear factor-κB (NF-κB) (Shostak et al., 2014) and activator protein-1 (AP-1), although the basic promoter activity of this gene mainly lied in the DNA methylation of the CpG island (Kuscu et al., 2012). For example, *CEMIP* over-expression was associated with hypomethylation of the CpG island in BC (Kuscu et al., 2012). In addition, the increased presence of lysine 4 of histone H3 trimethylation (H3K4me3) was reported to be an activation marker for *CEMIP* transcription, and reduced H3K27me3 was demonstrated to promote its expression in the development of BC (Hsieh et al., 2020; Liu et al., 2021a). Recently, several microRNAs and pro-inflammatory cytokines have been described to participate in the expression regulation of *CEMIP* as concluded in a review (Liu et al., 2021a). All works above indicate that *CEMIP* plays an important role in the oncogenesis and progress of carcinomas, and the exact mechanism remains to be explored. Moreover, the correlations of *CEMIP* with tumor immune infiltration in BC are still not determined.

In the present study, we performed expression analysis for *CEMIP* in different kinds of human cancers, particularly in BC, and assessed the association of *CEMIP* expression with the prognosis of patients with BC. Next, the potentially genetic regulatory mechanisms of *CEMIP*, including DNA methylation, genetic alteration, and upstream transcription factors, were explored in BC and other cancers. Then, we also manipulated functional annotation of *CEMIP*-related genes and *CEMIP*-interacted kinases and determined the correlation of *CEMIP* mRNA expression with the infiltration level of immune cells, biomarkers of immune cells, and immune checkpoints in BC. Finally, gene set enrichment analysis (GSEA) of a single gene was performed to identify the underlying pathways and hallmark perturbations caused by *CEMIP* in BC. All analyses were principally based on TCGA and GEO databases. In conclusion, our findings uncover that the up-regulation of *CEMIP* mediated by *TP53* mutation, DNA hypomethylation, and transcription factors correlates with worse outcomes and higher immune cell infiltration levels of patients with BC.

MATERIALS AND METHODS

mRNA and Protein Expression of *CEMIP* in Normal Tissues and Cellular Localization Analysis

First, GeneCards®: The Human Gene Database (https://www.genecards.org/) was used to explore mRNA and protein expression of *CEMIP* in normal tissues and cells in the "expression module" and visualize its subcellular locations in the "Localization module". The Human Protein Atlas (https://www.proteinatlas.org/) was as well applied to visualize the location of CEMIP.

CEMIP Expression Analysis Among Tumors

Subsequently, the "Gene_DE Module" in the TIMER2.0 (Li et al., 2020b) (tumor immune estimation resource, version 2) webserver (http://timer.cistrome.org/) was applied to estimate the *CEMIP* mRNA expression level in all The Cancer Genome Atlas (TCGA) (Tomczak et al., 2015) cancer types compared to their corresponding adjacent normal tissues, which were displayed by box plots. Next, we employed the UALCAN (Chandrashekar et al., 2017) (http://ualcan.path.uab.edu/analysis-prot.html) web resource to investigate *CEMIP* expression among 1097 BC samples and 114 normal samples of TCGA data based on sample types (that is., BC versus normal tissues), individual cancer stages, patients' ages, major BC subclasses, nodal metastasis statuses, and *TP53* mutation statuses with the "Expression Link" of the "TCGA analysis module". Then, Breast Cancer Gene-Expression Miner v4.7 (Jézéquel et al., 2021) (bcGenExMiner v4.7, http://bcgenex.ico.unicancer.fr/BC-GEM/GEM-Accueil.php?js=1) was applied to validate *CEMIP* expression with DNA microarray data (n = 11,359) in the "EXPRESSION" of the "ANALYSIS Module" based on intrinsic molecular subtypes (PAM50 subtypes), patients' ages, nodal metastasis statuses, and *TP53* mutation statuses. Furthermore, we also downloaded the GSE42568 dataset (raw CEL file and GPL file) from The Gene Expression Omnibus (Edgar et al., 2002) (GEO, https://www.ncbi.nlm.nih.gov/geo/), which was based on GPL570 Platforms ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) and contained 104 primary BC and 17 normal breast biopsies gene chips. After quality control, a total of 94 BC and 14 normal breast samples were obtained, based on which we identified the differentially expressed genes (DEGs) in BC compared with normal samples using R package limma (Smyth, 2005) with $\log_2$Fold Change > 1, adjust p-value < 0.05 for expression validation of *CEMIP* and the following analysis.

Survival Analysis

We first performed survival analysis through the Kaplan-Meier plotter (Lanczky et al., 2010) (https://kmplot.com/analysis/index.php?p=service) based on the GEO database and divided the samples into high-expression and low-expression cohorts according to the median expression value of *KIAA1199* (one of alias of *CEMIP*) for exploring the associations of its expression level with patients overall, relapse-free, and distant metastasis-free survival (OS, RFS, and DMFS). Next, Breast Cancer Gene-Expression Miner v4.7 (Campone et al., 2012) as well was applied to carry out survival analysis based on DNA microarrays (n = 11,359) in "PROGNOSIS" of the "ANALYSIS Module", with which we explored the association of *CEMIP* expression level with patients' OS, disease-free survival (DFS), and DMFS among all BC patients, ER/PR-positive patients, and ER/PR-negative patients. Moreover, we downloaded the TCGA dataset of BC, including gene expression RNA-seq data (n = 1,218), clinical phenotype data (n = 1,247) data, curated survival data (n = 1,236), and somatic mutation data (MC3 gene-level non-silent mutation, n = 791) from the UCSC Xena web (https://xenabrowser.net/DATAPAGES/) for the univariate and multivariate OS analysis using the Cox Proportional Hazards model carried out by R package survival. We first applied separate univariate Cox regressions to assess the statistical significance for each of the variables with OS, comprising BC patient stage, age, ER/PR/HER2 status, molecular subtype, tumor/node/metastasis (TNM) status, *TP53* mutation status, and the expression level of *CEMIP*, and then performed multivariate Cox regression analysis with these variables.

Promoter Methylation and Genetic Alteration Analysis

Promoter methylation analysis was as well implemented using the UALCAN web based on TCGA data containing 793 BC samples and 97 normal samples, of which the BC samples were divided into different groups according to individual cancer stages, patient age, and major BC subclasses. The genetic alteration analysis of *CEMIP* was carried out in the cBioPortal (Cerami et al., 2012) web (https://www.cbioportal.org/) following these steps: (1) select "Breast"; (2) choose 11 studies (consisting of 4717 samples) except for "TCGA, Cell 2015", "TCGA, Nature 2012", and "TCGA, PanCancer Atlas" in the "Invasive Breast Carcinoma" section; (3) click "Query By Gene"; (4) input "CEMIP"; (5) submit Query; and (6) choose "Cancer Type Detailed" in "Cancer Types Summary" to obtain the genomic alteration of CEMIP and choose the "Mutations" option to acquire diagrams of mutation sites.

Upstream Transcription Factor and Kinase Interaction Analysis

ARCHS^4 (Lachmann et al., 2018) (https://maayanlab.cloud/archs4/help.html) was applied to predict the upstream transcription factor (TF) targets and kinase interactions of *CEMIP* in humans, which predicted upstream TFs based on ChIP-seq data from the ChEA and ENCODE gene set libraries and predicted protein kinases based on known kinase substrates from KEA. Subsequently, we employed the cBioPortal web to estimate genetic alterations of predicted TFs; used the Draw Venn Diagram online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) to obtain the differentially expressed upstream TF and kinase targets of *CEMIP* in BC by taking the intersection of all predicted upstream TFs, kinase targets of *CEMIP*, and DEGs identified from the GSE42568 dataset; and further validated their expression levels by the UALCAN web. For validated TFs, Breast Cancer Gene-Expression Miner v4.7 (Jezequel et al., 20132013) was applied to explore the correlations of their expression levels with *CEMIP* and investigate the correlations of significantly correlated TFs with the survival of BC patients, while the JASPAR^{2022} database (https://jaspar.genereg.net/) was used to predict binding sites of transcription factors in the *CEMIP* promoter region. Meanwhile, the functional annotation of validated kinase targets and *CEMIP* was investigated by using the Metascape (Zhou et al., 2019) resource (http://metascape.org).

CEMIP Interacted and Correlated Gene Analysis

We first acquired *CEMIP*-interacted genes through the STRING (Szklarczyk et al., 2019) website (https://string-db.org/) with default settings, and then, we put "CEMIP" into the UALCAN

web in "TCGA" gene analysis with breast invasive cancer of the TCGA dataset to explore *CEMIP*-correlated genes in the "Correlation" module. Next, we again employed the STRING tool to construct the protein–protein interaction (PPI) network of *CEMIP*-interacted genes and *CEMIP*-correlated genes with Pearson correlation coefficient ≥ 0.4 and reproduced the network using Cytoscape (Smoot et al., 2011). Moreover, we also performed the Gene Ontology (GO) (Ashburner et al., 2000) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2019) pathway enrichment analysis of these genes with R package clusterProfiler (version 3.16.1).

Immune Infiltration Analysis

We also used TIMER2.0 to evaluate the correlations between the expression level of *CEMIP* and the infiltration levels of immune cells, including $CD8^+$ T cells, $CD4^+$ T cells, B cells, macrophages, neutrophils, dendritic cells (DCs), natural killer (NK) cells, and CAF in the "Gene module" of the "Immune Association" section with all algorithms provided, like EPIC, TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, XCELL, and MCPCOUNTER algorithms. Then, for obviously correlated immune cells, we further employed the Breast Cancer Gene-Expression Miner v4.7 tool to estimate the correlations between the expression of *CEMIP* and the biomarkers of immune cells as well as three well-known immune checkpoints (PD1, PD-L1, and CTLA-4).

Gene Set Enrichment Analysis of *CEMIP*

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) is a powerful method to annotate gene expression data based on defined gene sets consisting of genes that have been proved to have common biological functions, chromosomal location or regulation, gene expression data, and information of phenotypes of samples. Herein, we used GSEA software (version 4.1.0) to assess the pathway variations correlated with the expression level of *CEMIP* based on the GSE42568 dataset by dividing BC samples into high/low-expression groups of *CEMIP* according to its median expression value with the cutoff: nominal p-value < 0.05.

Statistical Analysis

In this study, the statistical analysis carried out by online tools was automatically calculated and the p-value, log rank p-value, or nominal p-value < 0.05 was considered as statistically significant. With the analysis performed by R, adjust p-value and p-value < 0.05 were considered as statistically significant.

RESULTS

The mRNA and Protein Expression of *CEMIP* in Normal Tissues, Immune Cells, and Cellular Localization

We explored the expression profiling of *CEMIP* across different normal tissues, immune cells, and its cellular localization through two public protein databases. As displayed in **Supplementary Figure S1**, the mRNA of *CEMIP* is detected in most tissues using RNA-seq and microarray approaches, such as whole blood, the brain, lungs, the pancreas, the skin, etc., showing low tissue specificity, while in the breast, it can only be probed with RNA-seq (**Supplementary Figure S1A**). Moreover, *CEMIP* is expressed in immune cells, especially in T-reg cells, plasmacytoid DCs, naive B cells, memory B cells, and naive $CD8^+$ T cells (**Supplementary Figure S1B**). At the protein level, *CEMIP* was distinctly detected in plasma, pancreas, and bone marrow stromal cells and stem cells, while in the breast, it was rarely expressed (**Supplementary Figure S1C**). On the other hand, we also dug into cellular localization of *CEMIP*, which was discovered in the plasma membrane, extracellular regions, nucleus, and endoplasmic reticulum with confidence = 5 according to GeneCards®: The Human Gene Database, while in The Human Protein Atlas, CEMIP was predicted to be secreted (**Supplementary Figure S1D**).

The Expression of *CEMIP* in Cancers

TIMER2.0 was employed to investigate the mRNA expression level of *CEMIP* across a variety of cancers based on the TCGA dataset, which was expressed higher in 13 kinds of cancers other than normal controls ($p < 0.05$), including BC, but lower in other three types of cancer (**Figure 1A**). With UALCAN web tools, we again observed increased expression of *CEMIP* in BC relative to normal samples with statistical significance ($p < 0.001$) (**Figure 1B**) and further analyzed the expression level of *CEMIP* in BC based on subclasses (luminal, $HER2^+$, and TNBC), individual cancer stages, patients' age, nodal metastasis status, and *TP53* mutation status between 114 normal and 1,097 primary BC of TCGA. However, no significant difference existed among these clinical features (**Supplementary Figures S2A–D**) except that *CEMIP* was significantly expressed higher in the *TP53*-mutant group compared to the *TP53*-non-mutant group (**Figure 1C**). Applying the Breast Cancer Gene-Expression Miner v4.7 resource, which integrated almost all BC data comprising DNA microarrays (n = 11,359, most of which were obtained from the GEO dataset) and RNA-seq (n = 4712, TCGA data), we validated the over-expression of *CEMIP* in BC and the *TP53*-mutant group (**Figures 1D,F**) and found higher expression of *CEMIP* in basal-like/TNBC and $HER2^+$ BC compared with luminal BC (**Figure 1D**). Meanwhile, the expression of *CEMIP* showed discrepancy based on BC patients' age as well, higher in patients whose age was over 51 years, but without significant difference based on nodal metastasis status (**Figure 1E**; **Supplementary Figure S2E**). In addition, using bioinformatics approaches, we identified a total of 3,869 DEGs based on the GSE42568 dataset and also found that *CEMIP* expression was higher in BC than normal breast samples ($\log_2$fold Change = 1.20, adjust p-value < 0.0001).

The Prognostic Value of *CEMIP* in Patients With BC

Subsequently, we evaluated the correlation between *CEMIP* expression level and BC patients' outcomes. The Kaplan-Meier plotter was first applied to carry out survival analysis, according

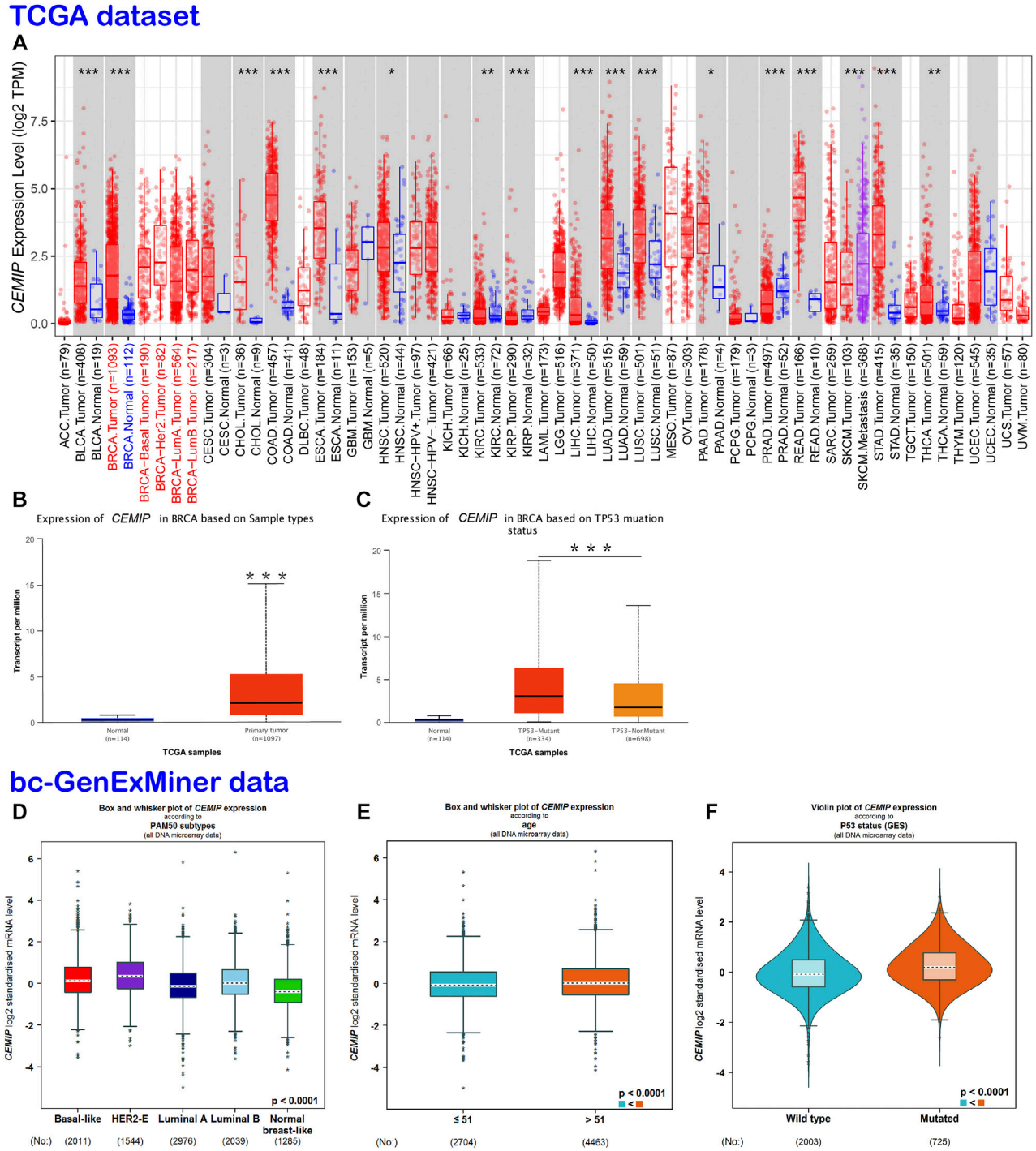


FIGURE 1 | Expression analysis of *CEMIP* across a variety of cancers and clinical features of BC. **(A)** Expression profile of *CEMIP* in various cancers analyzed by TIMER2.0 based on the TCGA database. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. **(B,C)** Expression analysis of *CEMIP* performed by the UALCAN web based on sample types and *TP53* mutation status with the TCGA database. ***$p < 0.001$. **(D–F)** Expression analysis of *CEMIP* carried out through the Breast Cancer Gene-Expression Miner v4.7 resource based on BC subtype, patient age, and *TP53* mutation status with 11,359 DNA microarrays of bc-GenExMiner data. bc-GenExMiner, the Breast Cancer Gene-Expression Miner; TCGA, The Cancer Genome Atlas.

to which no significant correlation was displayed between patients' OS and the expression level of *CEMIP* (**Figure 2A**, $p > 0.05$), while significantly negative associations were observed between patients' RFS and DMFS with *CEMIP* expression (**Figures 2B,C**, $p < 0.01$). Then, the Breast Cancer Gene-Expression Miner v4.7 resource was employed to confirm the

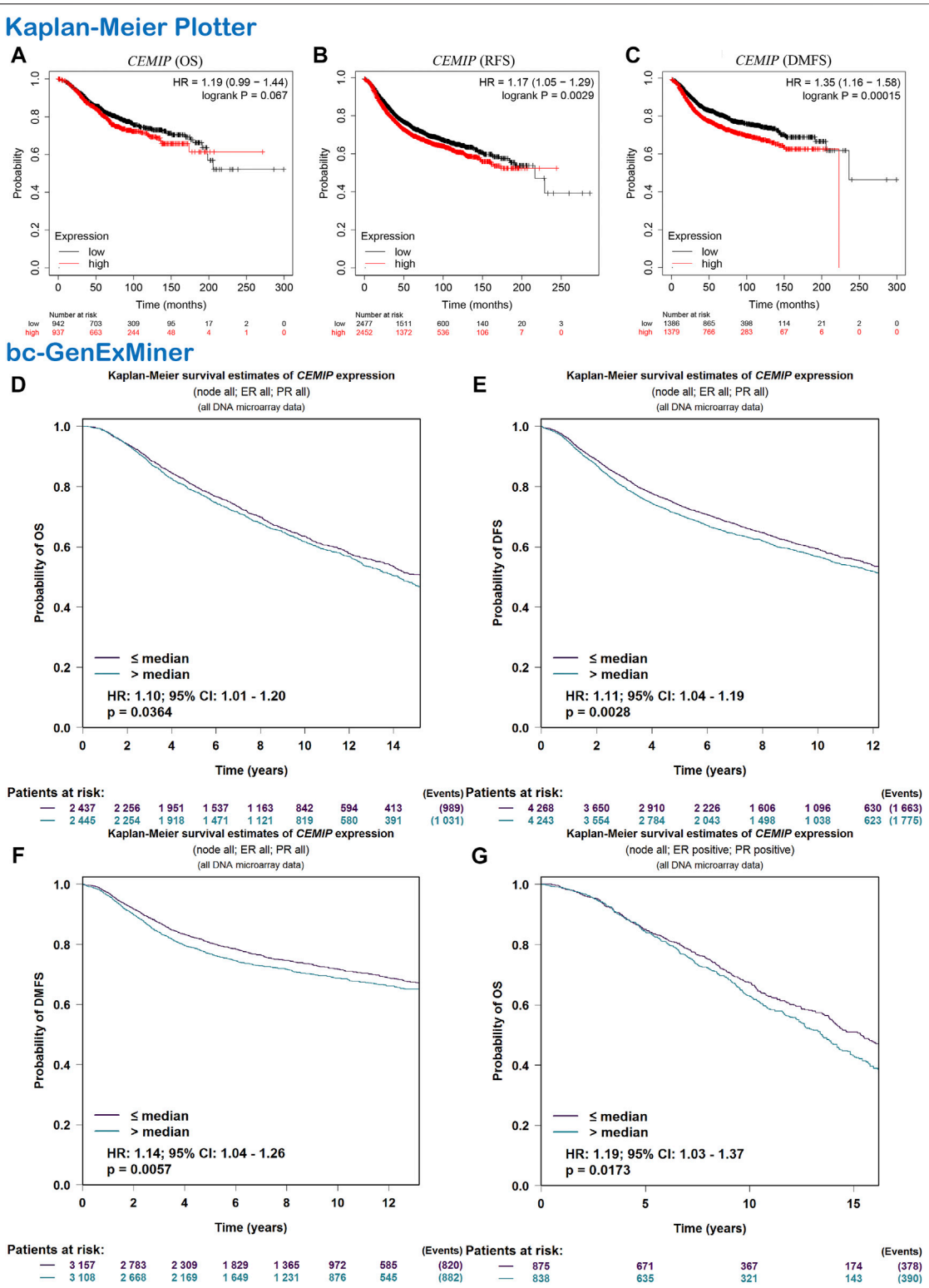


FIGURE 2 | Survival analysis. **(A–C)** OS, RFS, and DMFS plots of *CEMIP* in BC with the Kaplan-Meier plotter. **(D–F)** OS, DFS, and DMFS plots of *CEMIP* in all BC subtypes with the bc-GenExMiner. **(G)** OS plot of *CEMIP* in ER/PR^{+} BC with bc-GenExMiner. Log rank $p < 0.05$ and $p < 0.05$ were considered to be statistically significant. OS, overall survival; RFS, elapse-free survival; DMFS, distant metastasis-free survival; DFS, disease-free survival.

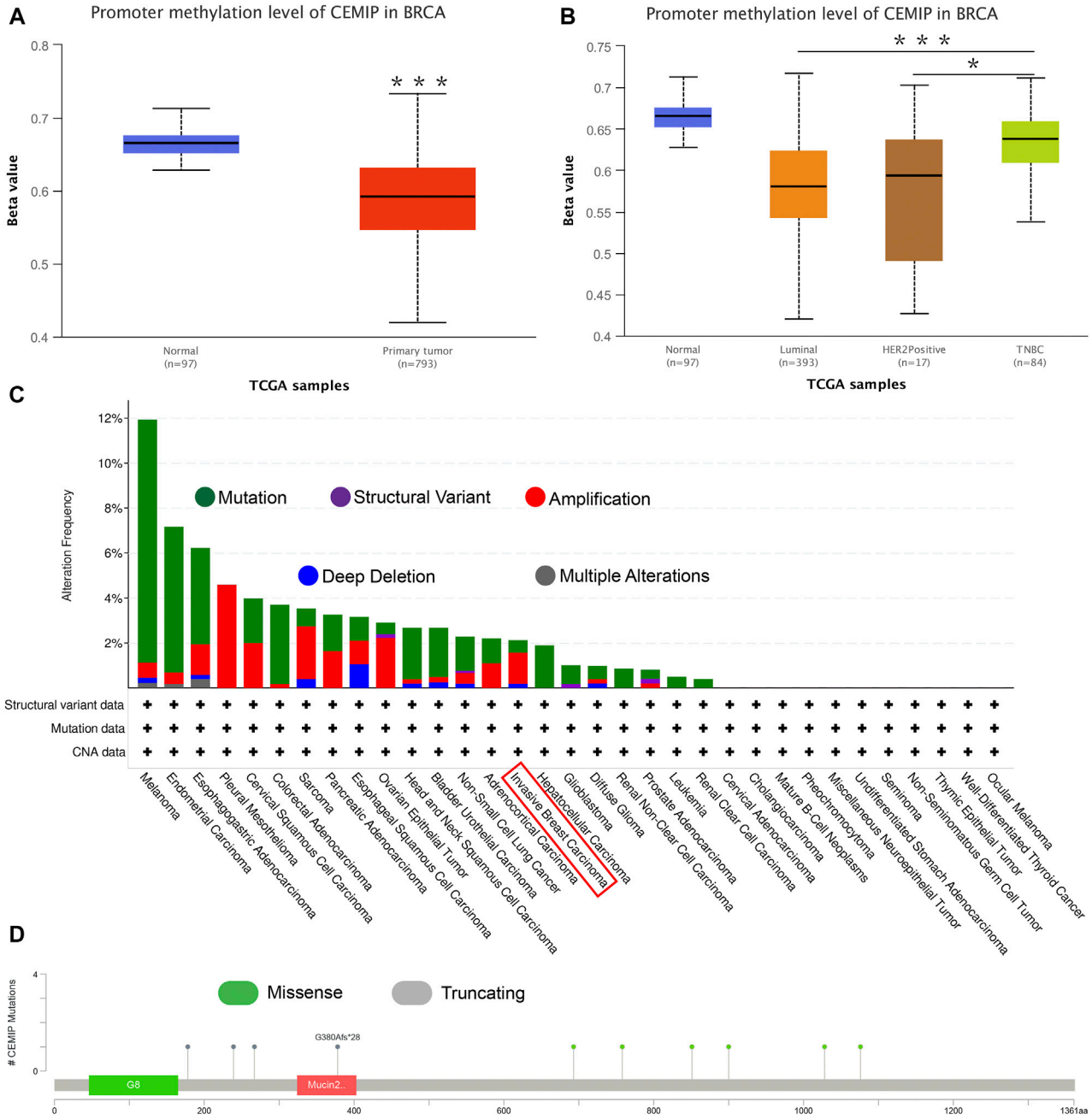


FIGURE 3 | Promoter methylation analysis using the UALCAN web and genetic alteration analysis performed by the cBioPortal web based on TCGA. **(A)** Promoter methylation level of *CEMIP* in BC compared with normal breast samples. ***$p < 0.001$. **(B)** Promoter methylation level of *CEMIP* in BC subtypes. *$p < 0.05$; ***$p < 0.001$. **(C)** Genetic alteration frequency of *CEMIP* in multiple malignancies. **(D)** Mutation site of *CEMIP* in BC.

prognostic value of *CEMIP*, with which we further performed survival analysis in ER/PR-positive BC patients and ER/PR-negative BC patients. Based on the DNA microarray data gathered in the Breast Cancer Gene-Expression Miner v4.7 resource, OS, DFS, and DMFS were negatively associated with *CEMIP* expression level (**Figures 2D–F**, $p < 0.05$) by enrolling all BC patients and OS was also negatively correlated with *CEMIP* expression level in ER/PR-positive BC patients (**Figure 2G**, $p < 0.05$), whereas no significant correlation existed between patients' DFS, DMFS, and the expression level of *CEMIP* (**Supplementary Figures S3A,B**, $p > 0.05$) as well as the OS, DFS, and DMFS in all ER/PR-negative BC patients (**Supplementary Figures S3C–E**, $p > 0.05$). Finally, we further predicted the prognostic value of *CEMIP* in BC patients by taking the stage, age, molecular subtype, ER/PR/HER2 status, TNM status, and *TP53* mutation status into consideration in addition to

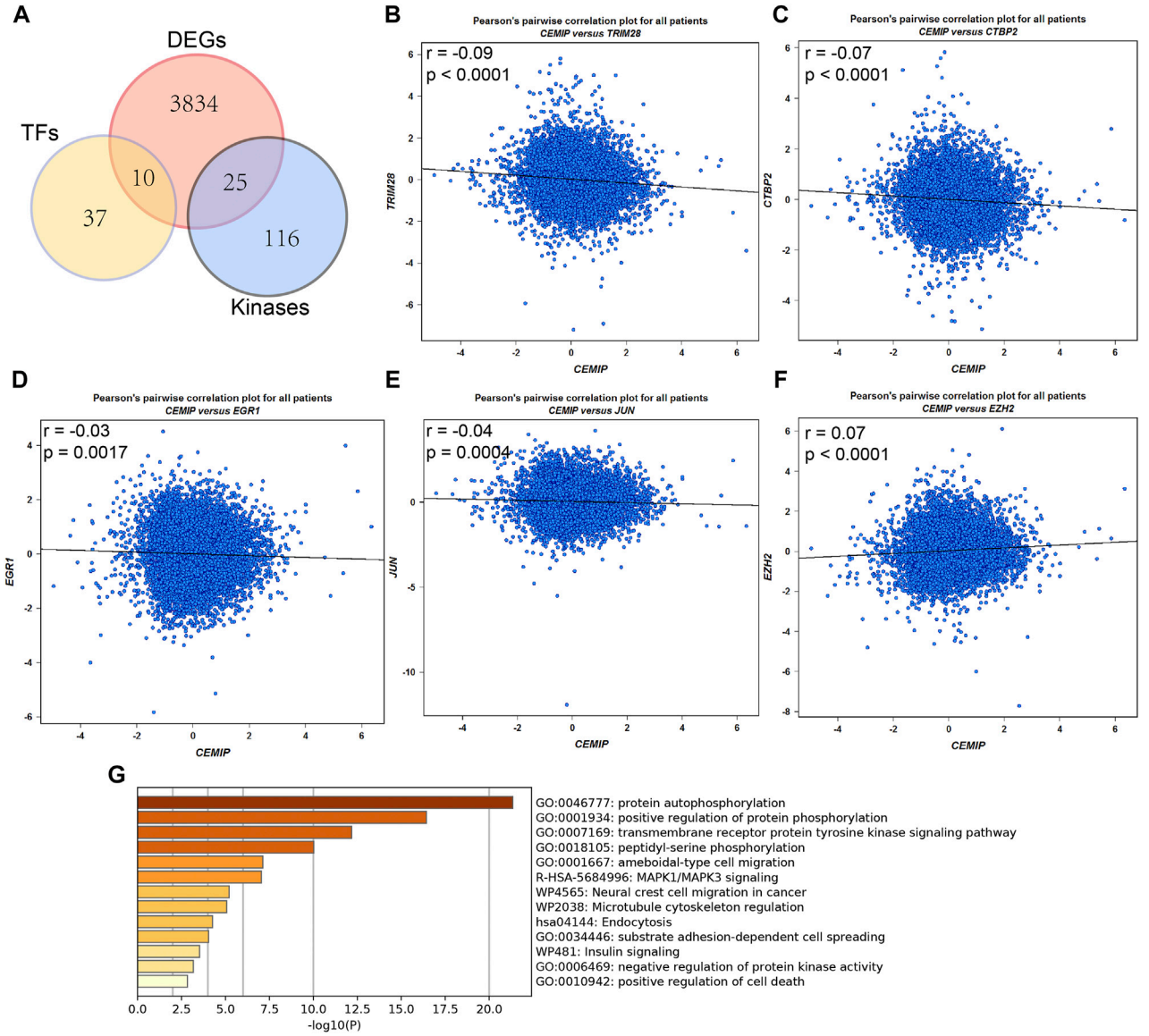


FIGURE 4 | Upstream transcription factor and kinase interaction analysis. **(A)** Differentially expressed predicted upstream TFs of *CEMIP* and kinases interacted with *CEMIP* identified in BC based on DEGs detected from the GSE42568 dataset. **(B–F)** Expression correlations of *CEMIP* and validated TFs in BC. **(G)** Functions and pathways enriched by kinases interacted with *CEMIP* and differentially expressed in BC. DEGs, differentially expressed genes.

the expression level of *CEMIP* using the Cox Proportional Hazards model based on TCGA cohorts. As a result, a total of 789 BC patients were enrolled, and the clinical characteristics and *TP53* state of them are listed in **Supplementary Table S1**. According to the univariate Cox regression analysis, we found that stage (HR = 2.2, p = 2.3e-07), age (HR = 1, p = 3.3e-06), node state (HR = 1.5, p = 0.00015), tumor states (HR = 1.5, p = 8e-04), metastasis status (HR = 4, p = 0.00045), and the expression level of *CEMIP* (HR = 1.1, p = 0.042) were significantly negative correlated to OS of BC patients as risk factors (**Supplementary Table S2**). Next, we included all variables into multivariate analysis and further revealed that only age (HR = 1.05, p = 4.18e-07) and the expression level of *CEMIP* (HR = 1.17, p = 0.028) were still significantly negatively associated with OS of BC patients in the presence of various factors (**Supplementary Table S2**).

DNA Methylation and Genetic Alteration of *CEMIP*

Next, the UALCAN web was applied to investigate the methylation level of promoter extension of *CEMIP*. As shown in **Figure 3**, the promoter methylation level of *CEMIP* is significantly lower in BC than in normal tissues (**Figure 3A**, p < 0.001). For subtypes of BC, the promoter methylation level of *CEMIP* was higher in TNBC than in luminal and HER2$^+$ BC

(**Figure 3B**, $p < 0.001$ and $p < 0.05$, respectively). Additionally, the cBioPortal resource was employed to assess the genetic alteration of *CEMIP* based on TCGA data. We found that the main genetic alterations of *CEMIP* in various cancers were mutation, amplification, deep deletion, structural variant, and multiple alterations. As for invasive breast carcinoma, the mutation, amplification with the highest alteration frequency, and deep deletion were involved (**Figure 3C**). The types and sites of *CEMIP* mutation in BC were further explored. As presented in **Figure 3D**, the missense and truncating are the dominating types and the latter alteration occurs to G380Afs*28 of the mucin2 domain with the highest frequency (**Figure 3D**).

Predicted Upstream Transcription Factor Targets and Kinase Interactions of *CEMIP*

Meanwhile, the ARCHS[4] resource was used to predict upstream TF targets and kinase interactions of *CEMIP*. As a result, a total of 47 unique predicted upstream TFs in humans and 141 kinases were obtained (**Figure 4A**). Then, we found that 43 TFs existed with genetic alterations in BC with the cBioPortal resource based on TCGA data, among which *ARNT*, *ATF3*, *ESR1*, *TFAP2C*, *TP53*, and *ZNF217* were observed to appear with higher rates of genetic alteration (>5%), and the main alteration type of them was amplification except for *TP53* (**Supplementary Figure S4**). Furthermore, combined with DEGs identified from the GSE42568 dataset, the expression levels of 10 TFs and 25 kinases were significantly different in BC compared with normal tissues (**Figure 4A**). Subsequently, we performed validation of their expression through the UALCAN web based on TCGA data as well, and 8 TFs and 22 kinases showed the same results (**Supplementary Table S3**). For validated TFs, we further evaluated the expression correlations between them and *CEMIP* using the Breast Cancer Gene-Expression Miner based on 11,359 DNA microarrays and found that the expression levels of two up-regulated genes, *TRIM28* and *CTBP2*, and two down-regulated genes, *EGR1* and *JUN*, in BC were negatively associated with *CEMIP*, and one up-regulated gene *EZH2* was positively correlated with *CEMIP* (**Figures 4B–F**; **Supplementary Table S3**). Taking expression analysis and correlation analysis into account, we speculated that *EZH2*, *EGR1*, and *JUN* might be the most potential upstream TFs of *CEMIP*. Thereby, expression and survival analyses were further carried out using the Breast Cancer Gene-Expression Miner. Accordingly, *EZH2* was expressed higher in basal-like/TNBC, $HER2^+$ and luminal B BC compared to the luminal A BC and normal breast tissues (**Figure 5A**) and showed negative associations with OS, DFS, and DMFS of BC patients (**Figures 5D–F**), while the expressions of *EGR1* and *JUN* were lower in these three types of BC (**Figures 5B,C**) and displayed positive associations with OS, DFS, and DMFS of BC patients (**Figures 5G–K**) except that the *JUN* expression had no statistically significant correlation with DMFS (**Figure 5L**). Moreover, the $JASPAR^{2022}$ database was used to predict binding sites of *EGR1*, *JUN*, and *EZH2* in the *CEMIP* promoter region. As shown in **Supplementary Table S4**, *EGR1* has predicted six binding sites and *JUN* has predicted five binding sites in the *CEMIP* promoter region, whereas *EZH2* could not be retrieved. Subsequently, through literature reviews about *EGR1*, *JUN*, and *EZH2*, we recognized that *EGR1* and *JUN* could be tumor suppressors, whereas *EZH2* can inhibit the expression of tumor suppressors (Baron et al., 2006; Shaulian, 2010; Eich et al., 2020). Therefore, we further assess the correlations among these three TFs. As a result, *EZH2* expression level was negatively correlated with *EGR1* (**Figure 5M**, $r = -0.28$, $p < 0.0001$) and *JUN* (**Figure 5N**, $r = -0.17$, $p < 0.0001$), while the expression of *EGR1* was positively correlated with *JUN* (**Figure 5O**, $r = 0.60$, $p < 0.0001$).

For validated kinases, we carried out enrichment analysis among them and *CEMIP* with the Metascape resource and found that they were mainly enriched in protein autophosphorylation, positive regulation of protein phosphorylation, and the transmembrane receptor protein kinase signaling pathway of GO terms (**Figure 4G**). Moreover, we integrated all biological processes that *CEMIP* participated in and noticed that two down-regulated genes *DDR2* and *TGFBR2* in BC compared with normal breast tissues and four up-regulated genes, *PTK2*, *RET*, *PRKD2*, and *CEMIP*, were involved in the positive regulation of cell migration, positive regulation of cell motility, positive regulation of cellular component movement, and positive regulation of locomotion (**Supplementary Tables S5, S6**).

CEMIP-Related Genes and Annotation

To further inquire about the molecular mechanism of *CEMIP* in BC, we integrated genes that interacted with *CEMIP* obtained from the STRING tool and the genes whose expressions were positively correlated with *CEMIP* expression. **Figure 6A** shows interacting genes, from which we could observe that the expression levels of most of these genes were distinct in BC compared with normal samples based on the DEG analysis of the GSE42568 dataset. Furthermore, gene annotation of GO functions, including biological process (BP), cellular component (CC), and molecular function (MF), and the KEGG pathway was executed by R package clusterProfiler. The top 10 terms of them are exhibited in **Figures 6B–E**. Most of these genes participated in the extracellular matrix organization, extracellular structure organization, and cell-substrate adhesion of BP (**Figure 6B**); located in the collage-containing extracellular matrix, focal adhesion, cell-substrate junction, and endoplasmic reticulum lumen of CC (**Figure 6C**); possessed extracellular matrix structural constitutes, cell adhesion molecule binding, glycosaminoglycan binding, integrin binding and collagen binding of MF (**Figure 6D**); and were involved in ECM–receptor interaction, focal adhesion, the PI3K-Akt signaling pathway, and the TGF-beta signaling pathway (**Figure 6E**).

Correlations of *CEMIP* Expression With the Immune Cell Infiltration Level in Cancers

Given that the tumor immune microenvironment (TIM) plays an important role in the progress and metastasis of cancers, we evaluated the correlations of *CEMIP* expression with several immune cell infiltration levels in various types of cancers, consisting of $CD8^+$ T cells, $CD4^+$ T cells, B cells, macrophages, neutrophils, DC, NK cells, and CAF using the TIMER2.0 web

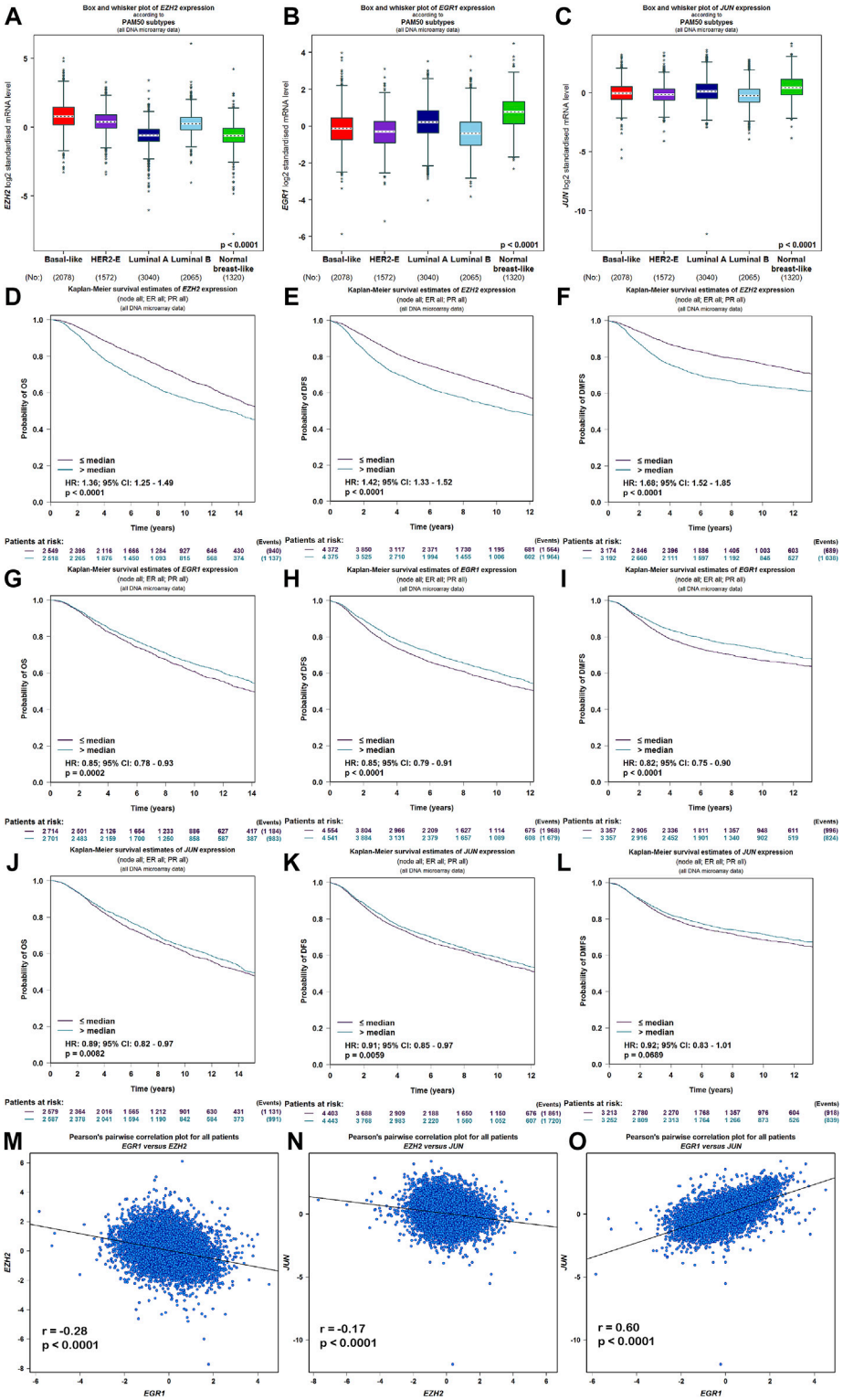


FIGURE 5 | Expression, survival, and correlation analysis of upstream TFs of *CEMIP* in BC. **(A–C)** Expression levels of *EZH2*, *EGR1*, and *JUN* among subtypes of BC. **(D–F)** OS, DFS, and DMFS analysis of *EZH2* in BC patients. **(G–I)** OS, DFS, and DMFS analysis of *EGR1* in BC patients. **(J–L)** OS, DFS, and DMFS analysis of *JUN* in BC patients. **(M–O)** Correlations between *EZH2* and *EGR1*, *EZH2* and *JUN*, and *EGR1* and *JUN* in BC.

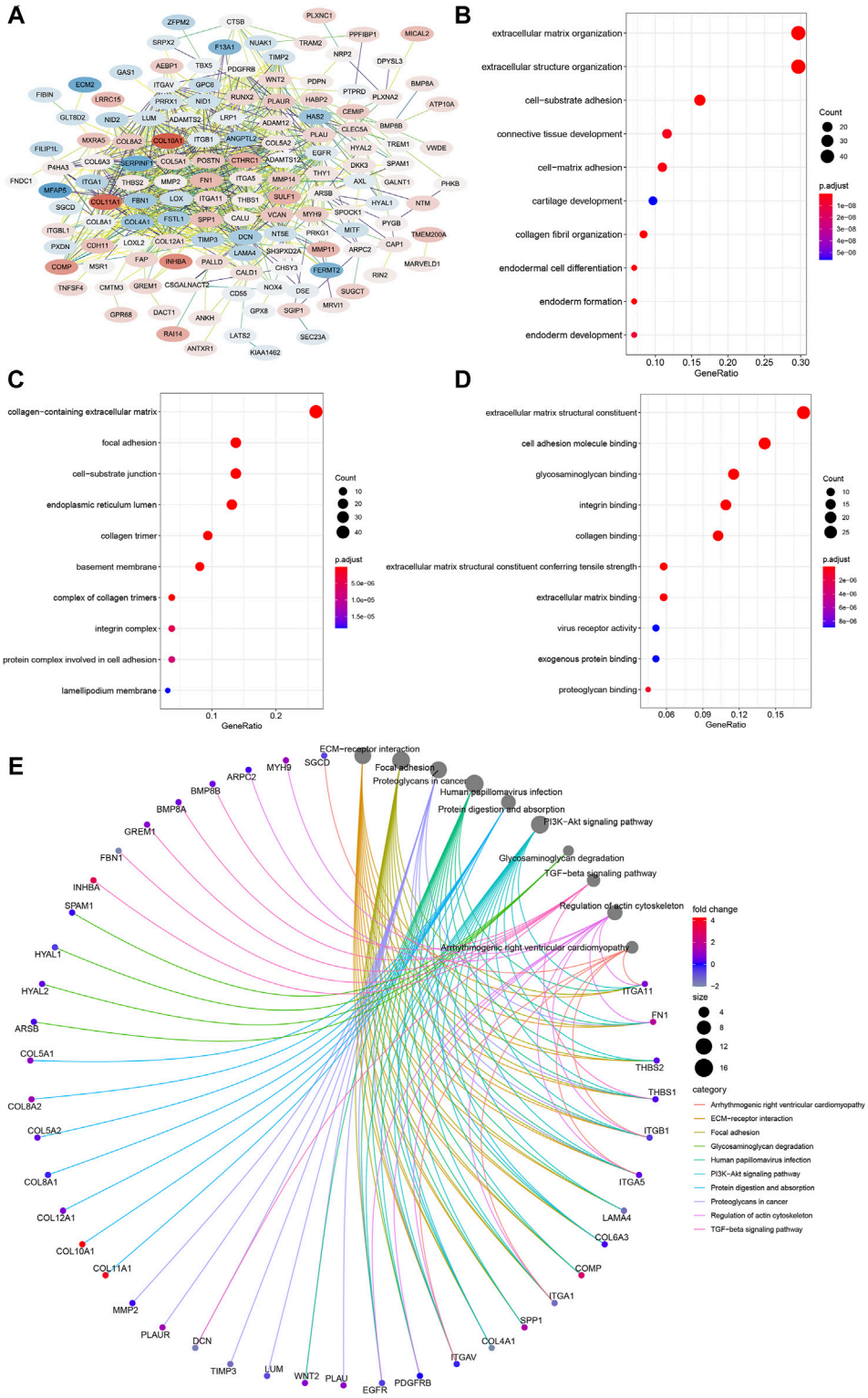


FIGURE 6 | PPI network construction and enrichment analysis of *CEMIP*-related genes. **(A)** PPI network of *CEMIP*-related genes generated by the STRING tool. The red nodes represent up-regulated genes in BC compared with normal samples, while the blue nodes represent down-regulated genes. **(B)** Top 10 BP terms of GO enriched by *CEMIP*-related genes. **(C)** Top 10 CC terms of GO enriched by *CEMIP*-related genes. **(D)** Top 10 MF terms of GO enriched by *CEMIP*-related genes. **(E)** Top 10 KEGG pathways enriched by *CEMIP*-related genes. PPI, protein–protein interaction; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological processes; CC, cellular components; MF, molecular functions.

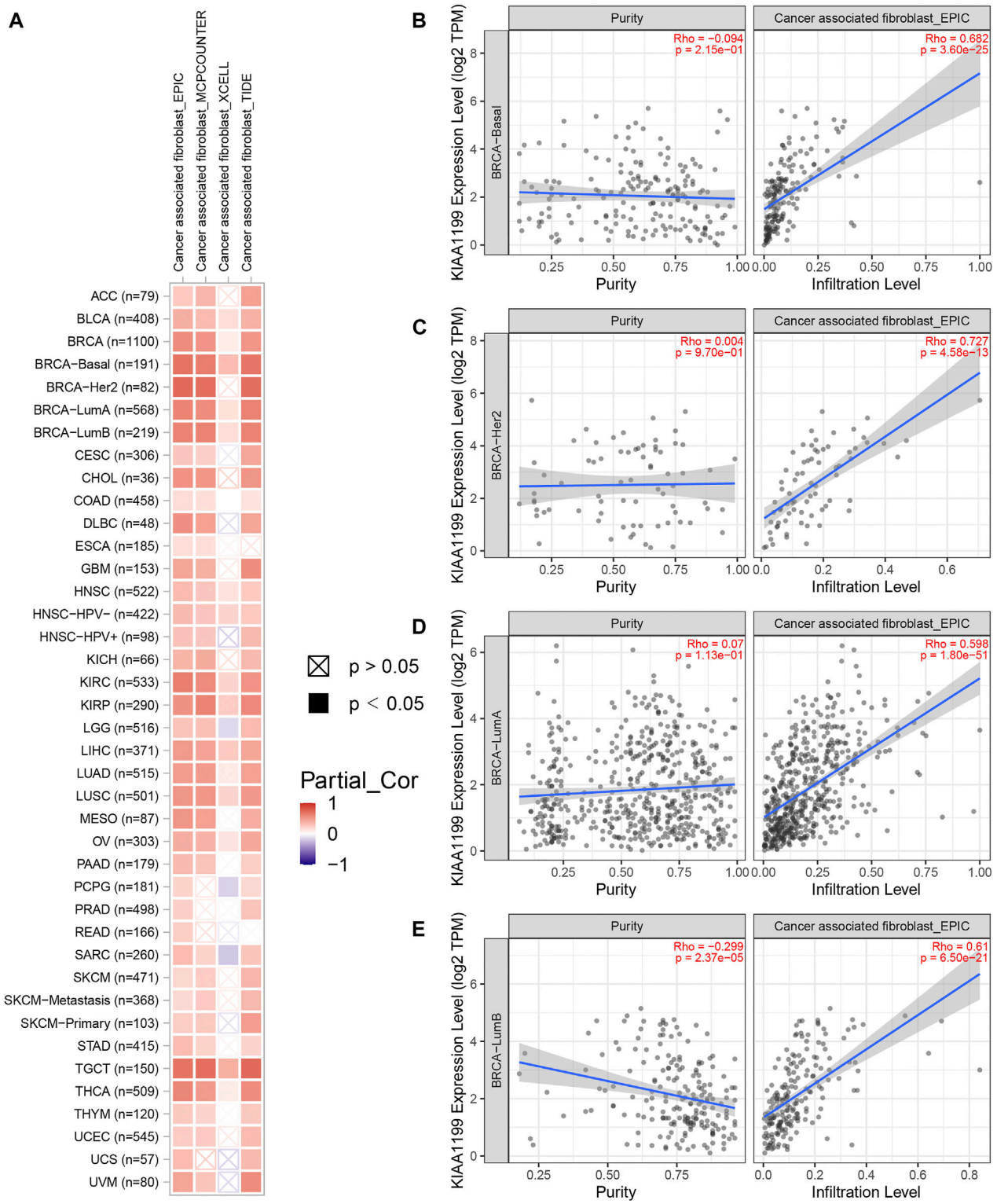


FIGURE 7 | Correlation of *CEMIP* expression with CAF infiltration level explored by TIMER2.0 web. **(A)** Correlations between *CEMIP* expression and the infiltration level of CAF across various types of cancers analyzed by four algorithms provided (EPIC, MCPCOUNTER, XCELL, and TIDE). The red squares represent positive correlations, while the blue squares represent negative correlations with statistical significance ($p < 0.05$). **(B–E)** Correlations between *CEMIP* expression and the infiltration level of CAF among four subtypes of BC (basal-like/TNBC, HER2$^+$, luminal A, and luminal B) using the EPIC algorithm. CAF, cancer-associated fibroblast.

TABLE 1 | Expression correlations of *CEMIP* with biomarkers of immune cells and immune checkpoints.

Immune cell	Biomarker	r value	*p*-value
B cell	CD19	−0.02	0.0558
	CD79A	−0.00	0.8227
$CD8^+$ T cell	CD8A	−0.02[a]	0.0331*
	CD8B	−0.02	0.0611
M1 macrophage	NOS2	0.07[a]	<0.0001****
	IRF5	0.12[a]	<0.0001****
	PTGS2	−0.01	0.5330
M2 macrophage	CD163	0.23[a]	<0.0001****
	VSIG4	0.24[a]	<0.0001****
	MS4A4A	0.17[a]	<0.0001****
Neutrophil	CEACAM8	0.00	0.6818
	ITGAM	0.22[a]	<0.0001****
	CCR7	−0.06[a]	<0.0001****
Checkpoints	PDCD1 (PD1)	−0.01	0.3018
	CD274(PD-L1)	0.08[a]	<0.0001****
	CTLA4	0.06[a]	<0.0001****

[a]*These results are statistically significant.* *p-*value* < 0.05; **p-*value* < 0.01; ***p-*value* < 0.001; ****p-*value* < 0.0001.

with all algorithms provided. Generally, the infiltration of macrophages, neutrophils, resting NK cells, and CAF was positively related to *CEMIP* expression with all or most algorithms in most carcinomas, including BC (**Supplementary Figures S5, S6A** and **Figure 7A**), while activated NK cells, $CD8^+$ T cells, and $CD4^+$ Th1 cells were opposite (**Supplementary Figures S6, S7A**) and other immune cell (that is., $CD4^+$ T cells, B cells, and DC) infiltrations showed no clear unifying trends with these available algorithms (**Supplementary Figures S7, S8**). We noticed that the infiltration of B cells was negatively associated with the expression level of *CEMIP* in BC (**Supplementary Figure S7B**). We also noted that the CAF infiltration had the highest correlations with *CEMIP* expression in BC, the scatterplots of which among subtypes of BC generated using the EPIC algorithm are presented in **Figures 7B–E**.

Expression Correlations of *CEMIP* With Biomarkers of Immune Cells and Checkpoints in BC

To further estimate the role of *CEMIP* in BC immune regulation, we explored the expression correlations of *CEMIP* with biomarkers of immune cells whose infiltration levels were significantly associated with *CEMIP* expression and immune checkpoints in BC. As listed in **Table 1**, *CEMIP* had significant positive correlations with M1 macrophages' biomarkers *NOS2* ($r = 0.07$, $p < 0.0001$) and *IRF5* ($r = 0.12$, $p < 0.0001$); M2 macrophages' biomarkers *CD163* ($r = 0.23$, $p < 0.0001$), *VSIG4* ($r = 0.24$, $p < 0.0001$), and *MS4A4A* ($r = 0.17$, $p < 0.0001$); neutrophils' biomarker *ITGAM* ($r = 0.22$, $p < 0.0001$); and immune checkpoints PD-L1 ($r = 0.08$, $p < 0.0001$) and CTLA4 ($r = 0.06$, $p < 0.0001$) and negative correlations with $CD8^+$ T cells' biomarker *CD8A* ($r = -0.02$, $p = 0.0331$) and neutrophils' biomarker *CCR7* ($r = -0.06$, $p < 0.0001$).

KEGG Pathway and HALLMARK Aberrations Correlated With CEMIP in BC

To further investigate the effect of *CEMIP* on BC, we manipulated Gene Set Enrichment Analysis according to the expression of *CEMIP* based on the gene expression matrix of the GSE42568 dataset and gene sets of the KEGG pathway and HALLMARK. As listed in **Supplementary Table S7**, a total of 10 gene sets were up-regulated in the high-expression group of *CEMIP* and were significantly enriched at nominal p-value < 0.05 based on the KEGG pathway, while two gene sets were based on HALLMARK. The top two KEGG pathways and HALLMARK processes enriched were "ANTIGEN_PROCESSING_AND_PRESENTATION, GLYCOSAMINOGLYCAN_ BIOSYNTHESIS_CHONDROITIN_ SULFATE" of KEGG" (**Figures 8A,B**) and "ALLOGRAFT_REJECTION, EPITHELIAL_MESENCHYMAL_TRANSITION" (**Figures 8C,D**), respectively.

DISCUSSION

To date, BC has been the most frequently diagnosed malignancy worldwide and leads to 684,996 deaths (Rodrigues et al., 2019; Zhai et al., 2020). Although the prognosis of cancer patients has improved dramatically with the advanced medical technology (Rodrigues et al., 2019; Zhai et al., 2020), a complete cure for BC patients is not yet possible with the currently available therapies due to the ambiguous mechanism of tumorigenesis and tumor progression in BC. Accumulated evidence shows that the over-expression of *CEMIP* could enhance proliferation, adhesion, motility, invasiveness, and EMT of various carcinomas, including BC, as well as its transcriptional regulation mechanism in many cancers (Liu et al., 2021a), indicating that *CEMIP* might play an important role in BC. Therefore, we performed an integrated analysis of *CEMIP* in BC and other cancers using bioinformatics approaches and a variety of online analysis tools mainly based on TCGA and GEO databases.

CEMIP is a type of secreted protein in exosomes and also exists in normal human tissues, like the brain, lung, pancreas, and testis (Zhai et al., 2020; Rodrigues et al., 2019). It has been described that CEMIP could mediate depolymerization of hyaluronic acid (HA) (Yoshida et al., 2013), bind to HA, hydrolyze high-molecular-weight HA, and be involved in hyaluronan catabolism as summarized in https://www.uniprot.org/uniprot/Q8WUJ3. In the present study, we also explored CEMIP expression in normal tissues, immune cells, and its subcellular locations. According to previous findings, CEMIP can be detected in normal tissues, including plasma, and even in immune cells, which mainly are located in the plasma membrane, extracellular regions, nucleus, and endoplasmic reticulum. It was reported that over-expression of CEMIP in exosomes might facilitate the metastasis of BC (Rodrigues et al., 2019; Zhai et al., 2020). Other research studies revealed that CEMIP

residing in the endoplasmic reticulum may enhance BC cell survival in hypoxia and cancer cell migration by upregulating and interacting with binding immunoglobulin protein (Evensen et al., 2013; Banach et al., 2019). Thereby, the over-expression of CEMIP detected in blood might be a convenient way to early diagnose BC metastasis. However, targeting CEMIP therapy for BC patients might be more complicated in the consideration of its normal biological functions, which acquires a precise delivery route, like targeting CEMIP located in the endoplasmic reticulum.

Subsequently, we conducted a pan-cancer analysis of *CEMIP* expression based on the TCGA database and discovered that it was highly expressed in most cancers, consistent with early studies (Evensen et al., 2013; Jami et al., 2014; Zhang et al., 2014; Fink et al., 2015; Shen et al., 2019; Li et al., 2020a; Zhai et al., 2020; Liu et al., 2021a; Chen et al., 2021), after which we focused on *CEMIP* expression in BC. However, we did not observe a statistically significant difference in *CEMIP* expression based on BC subtypes (luminal, $HER2^+$, and TNBC), individual cancer stages, patient age, and nodal metastasis status except for the *TP53* mutation status at first. Then, the Breast Cancer Gene-Expression Miner v4.7 resource, which integrated almost all public BC data comprising DNA microarrays and RNA-seq, was employed to evaluate *CEMIP* expression based on BC subtypes, patient age, nodal metastasis status, and *TP53* mutation status again with 11,359 DNA microarrays. We noticed that *CEMIP* was highly expressed in $HER2^+$ and TNBC compared with luminal type and was expressed higher in the group of patients with an age over 51 years than those with an age less than 51, suggesting that *CEMIP* could be an indicator of aggressive types of BC, like $HER2^+$ and TNBC, and might be responsible for some characteristics of invasive BC. In addition, the inconsistency of results from the two web tools in our findings might be primarily caused by the sample size. Several studies described that *CEMIP* was highly expressed in invasive breast cancer specimens and in invasive MDA-MB-231 TNBC cell lines, whereas some researchers observed much lower expression of *CEMIP* in non-invasive BC cells with a low-invasive potential, like MCF-7, T-47D, and ZR-75-1 cell lines (Evensen et al., 2013; Zhang et al., 2014). Of note, *CEMIP* was expressed lower in samples without *TP53* mutation than *TP53* mutation samples with one accord using these two webs and databases. *TP53* is a well-known cancer suppressor, whose mutation has been reported in plenty of malignancies (Miller et al., 2005). Our finding shows that the high expression of *CEMIP* in invasive BC might be partly caused by *TP53* mutation since it has been reported that TNBC exhibits more *TP53* mutation (Verigos and Magklara, 2015). To our knowledge, no study has yet linked them.

After identifying the discrepancy of *CEMIP* expression in BC, we further investigated its prognostic value among BC patients. We found that the expression of *CEMIP* was negatively associated with OS, RFS, DFS, and DMFS of BC patients when all types of BC were included. However, when we divided BC patients into ER^+/PR^+ and ER^-/PR^- groups, only OS of BC patients with ER^+/PR^+ was significantly negatively correlated with *CEMIP* expression, indicating that its prognostic prediction in BC patients with ER^+/PR^+ might be more significant since the patients with TNBC have been widely proved to have poor prognoses (Waks and Winer, 2019). More convincingly, we discovered that high expression of *CEMIP* had a strong relationship with increased risk of death in BC patients according to both univariate and multivariate Cox regression analysis with hazard ratio (HR) = 1.17 and p-value = 0.028 of the latter, that is, holding the other covariates, that is., BC patient stage, age, ER/PR/HER2 status, TNM status, and *TP53* mutation status, constant, a higher expression value of *CEMIP* predicted a poor survival of BC patients. Given these results, *CEMIP* might be a promisingly prognostic and therapeutic target for BC patients.

It has been reported that *CEMIP* mutation in the GG domain leads to non-syndromic hearing loss, while its over-expression prevailingly contributes to its oncogenic roles (Liu et al., 2021a). However, in this study, we found that the genetic alterations of *CEMIP* appeared in multiple cancers, including invasive BC, suggesting that mutations (that is, missense and truncating) of *CEMIP* may be responsible for the formation and progression of aggressive BC, especially truncating in the G380Afs*28 of mucin2 domain of *CEMIP*. Although studies found that *CEMIP* mutation happens in invasive BC cell lines, like MDA-MB-435 and MDA-MB-231 (Zhang et al., 2014), the mechanism of this needs more experiments to explore.

Kuscu et al. (2012) elucidated that the regulatory mechanisms which control *CEMIP* expression were genetic and epigenetic. In agreement with the previous study, we discovered a link between DNA hypomethylation and high expression of *CEMIP*. Furthermore, we also identified some predicted upstream TFs mutated or aberrantly expressed in BC. The result again suggested that *CEMIP* expression might be influenced by *TP53* mutation with the highest mutation frequency. Except for *TP53* mutation, *ARNT*, *ATF3*, *ESR1*, *TFAP2C*, and *ZNF217* alterations might also affect the expression of *CEMIP*. Meanwhile, we found three potential up-regulated genes (*TRIM28*, *CTBP2*, and *EZH2*) and two down-regulated genes (*EGR1* and *JUN*) in BC compared with the normal breast tissues which might regulate *CEMIP* expression, especially *EZH2*, *EGR1*, and *JUN*, due to the fact that the correlations between their expression levels and *CEMIP* expression were consistent with the directions of change. That is, the up-regulated gene *EZH2* showed a positive correlation with *CEMIP*, and the down-regulated genes *EGR1* and *JUN* had negative correlations with *CEMIP*, indicating that the abnormal expression levels of these three TFs were likely to be responsible for *CEMIP* over-expression in BC. Moreover, *EGR1* and *JUN* were included in the $JASPAR^{2022}$ database, where they had six and five possible binding sites in the *CEMIP* promoter region, respectively. Additionally, the high expression of *EZH2* predicted low OS, DFS, and DMFS of BC patients, while *EGR1* and *JUN* were opposite. *JUN*, a member of AP-1, is a controversial gene in cancer, which could be an oncogene or a tumor suppressor (Shaulian, 2010). It was reported to regulate *CEMIP* expression but as an activator of the *CEMIP* promoter (Shaulian, 2010). Hence, the specific circumstance needs more experiments to find out. *EGR1* is a cancer suppressor and has been confirmed to be down-regulated in BC compared with normal tissues (Baron et al., 2006), but no studies revealed the regulatory relationship between it and *CEMIP*. *EZH2* is classified as an oncogene, shows high expression in numerous cancers

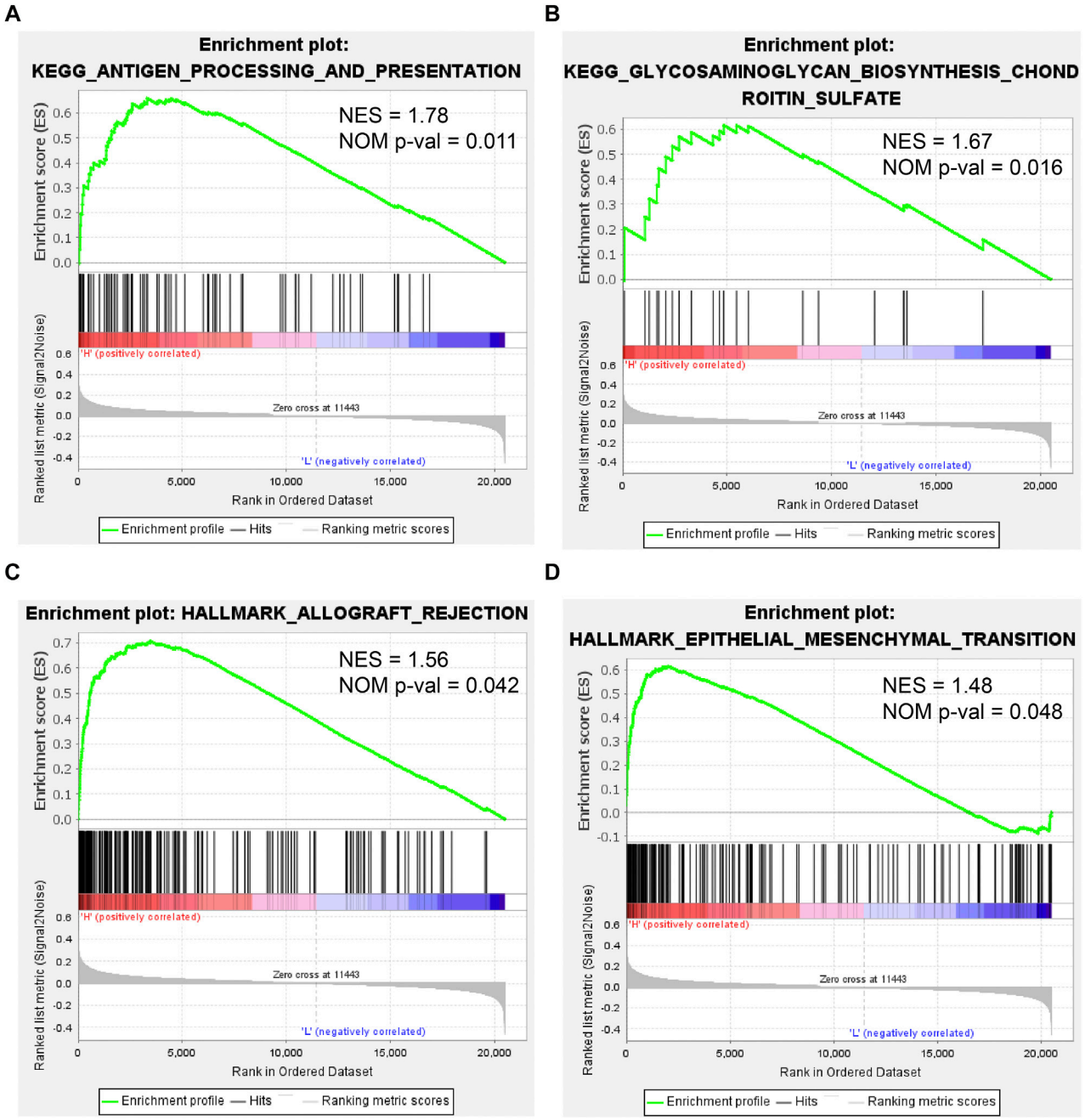


FIGURE 8 | Gene Set Enrichment Analysis based on the GSE42568 dataset and the gene sets of KEGG pathways and HALLMARK. **(A,B)** Top two KEGG pathways enriched by the *CEMIP* high expression group with the highest NES and NOM *p*-val < 0.05. **(C,D)** Two HALLMARK terms enriched by the *CEMIP* high-expression group with NOM *p*-value < 0.05. NES, normalized enrichment score; NOM *p*-value, nominal *p*-value.

including breast cancer, and is discovered to contribute to global transcriptional repression, mainly targeting tumor suppressor genes (Eich et al., 2020). However, no studies revealed the regulatory relationship between it and *CEMIP* either. Then, given that *EZH2* can inhibit the expression of tumor suppressors, we further assess the correlations among these three TFs. As expected, the *EZH2* expression level was negatively correlated with *EGR1* and *JUN*, while the expression of *EGR1* was positively correlated with *JUN*, suggesting that *EZH2*/*EGR1* and *JUN*/*CEMIP* might be the potential regulatory pathways in BC, especially in invasive BC, like TNBC and HER2^{+} BC, or *EZH2*, *EGR1*, and *JUN* mediate *CEMIP* expression directly in BC.

Cancer invasion and metastasis are basically dependent on cell migration (Hanahan and Weinberg, 2011; Evensen et al., 2013). It has been demonstrated that *CEMIP* can promote prostate, breast,

and colon cancer cell motility (Evensen et al., 2013). In our study, we first identified differentially expressed kinases that interacted with *CEMIP* in BC and performed enrichment analysis on these genes, including *CEMIP*. We noted that *DDR2*, *PTK2*, *RET*, *TGFBR2*, *PRKD2*, and *CEMIP* were involved in positive regulation of cell migration, indicating that *CEMIP* might participate in cancer cell migration by interacting with these genes, of which *DDR2* and *TGFBR2* were down-regulated in BC and proved to inhibit cancer metastasis (Lo Sardo et al., 2021; Mehta et al., 2021), while *PTK2*, *RET*, and *PRKD2* were up-regulated in BC and were shown to promote cancer development (Azoitei et al., 2014; Fan et al., 2019; Subbiah and Cote, 2020). However, no studies have revealed the relationships between them yet. Moreover, we also integrated *CEMIP*-related genes and found them mainly enriched in the ECM–receptor interaction, focal adhesion, the PI3K-Akt signaling pathway, and the TGF-beta signaling pathway, suggesting that *CEMIP* might promote cancer cell invasion and metastasis through these pathways. Some studies reported that *CEMIP* is involved in EMT, Wnt/β-catenin, MEK/ERK, and PI3K/Akt signal pathways to promote cancer progression, while the exact mechanism is still ambiguous.

In the present study, we first evaluated the correlations of the expression of *CEMIP* with the infiltration levels of immune cells and found that biomarkers of these infiltrated immune cells were significantly related to *CEMIP* expression as well as three well-known immune checkpoints. Through rigorous evaluation of a variety of algorithms, we observed that the infiltration of macrophages, neutrophils, resting NK cells, and CAF was positively related to *CEMIP* expression in most carcinomas, including BC, while activated NK cells and $CD8^+$ T cells were opposite, and the infiltration of B cells was negatively associated with the expression level of *CEMIP* in BC. Moreover, we discovered that *CEMIP* was correlated with some biomarkers of these immune cells and immune checkpoints, especially, *CD163* and *VSIG4* of M2 macrophages, and *ITGAM* of neutrophils. It has been widely accepted that the body's immune system plays a dual role in tumor initiation and progression, which suppress tumor growth in the early phase of oncogenesis, but promoting tumor progression once a tumor becomes invasive (Schreiber et al., 2011). The TME, consisting of the extracellular matrix (ECM), stromal cells (such as fibroblasts), and immune cells (comprising T and B lymphocytes, NK cells, and tumor-associated macrophages), provides mechanical support for the tumor (Roma-Rodrigues et al., 2019). Additionally, accumulated evidence has documented that the immune cell infiltration in TME is associated with BC patient outcomes. For instance, the high level of lymphocytic infiltration may predict a better prognosis in patients with early-stage TNBC and $HER2^+$ BC (Savas et al., 2016; Gao et al., 2020), whereas the infiltration of CAF and M2 macrophages may contribute to cancer progression (Ishii et al., 2016; Xia et al., 2020). Combining these existing research studies and our findings, *CEMIP* might promote the occurrence and development of tumors via participating in the formation of TME including both ECM and the immune microenvironment (immune cells), indicating that *CEMIP* may be a promisingly therapeutic target for advanced BC.

Finally, we performed GSEA according to the expression of *CEMIP* based on KEGG pathways and HALLMARK. In the group with high expression of *CEMIP,* the pathways and hallmarks significantly enriched were related to antigen presentation and EMT, which again revealed the entanglement of *CEMIP* with the EMT pathway and tumor immune infiltration.

In conclusion, we illustrated that *CEMIP* was highly expressed in various kinds of carcinomas, including BC, especially advanced subtypes, and predicted less favorable prognosis (negatively associated with OS, RFS, DFS, and DMFS) in BC patients, and the higher the expression of it, the worse the outcomes BC patients have. We revealed that the mutation and high expression of *CEMIP* might lead it to an oncogene. We also demonstrated that *TP53* mutation, DNA hypomethylation, and the expression changes of upstream TFs of *CEMIP* were likely to cause hyper-expression of *CEMIP*, and we further identified three potential upstream TFs in BC, namely, *EZH2*, *EGR1*, and *JUN*. Moreover, our findings suggested that *CEMIP* was closely related to TME and might exert its oncogenic roles by participating in the extracellular matrix formation, mainly increasing CAF, M2 macrophage, and neutrophil infiltration and decreasing $CD8^+$ T cell infiltration. Of course, these findings need more solid confirmations of further experiments and clinical trials in the future.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

XD, YY, JH, and GW conceived the study; XD, YY, and QY dealt with the data and performed the visualization; XD and YY wrote the manuscript; JH and GW revised the article. All authors contributed to the article and approved the submitted version.

FUNDING

The project was financially supported by the research fund from the Medical Sci-Tech Innovation Platform of Zhongnan Hospital, Wuhan University (PTXM2021017).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.768140/full#supplementary-material

REFERENCES

Aleskandarany, M. A., Vandenberghe, M. E., Marchiò, C., Ellis, I. O., Sapino, A., and Rakha, E. A. (2018). Tumour Heterogeneity of Breast Cancer: From Morphology to Personalised Medicine. *Pathobiology* 85, 23–34. doi:10.1159/000477851

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., and Cherry, J. M. (2000). Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556

Azoitei, N., Diepold, K., Brunner, C., Rouhi, A., Genze, F., Becher, A., et al. (2014). HSP90 Supports Tumor Growth and Angiogenesis through PRKD2 Protein Stabilization. *Cancer Res.* 74, 7125–7136. doi:10.1158/0008-5472.can-14-1017

Banach, A., Jiang, Y.-P., Roth, E., Kuscu, C., Cao, J., and Lin, R. Z. (2019). CEMIP Upregulates BiP to Promote Breast Cancer Cell Survival in Hypoxia. *Oncotarget* 10, 4307–4320. doi:10.18632/oncotarget.27036

Baron, V., Adamson, E. D., Calogero, A., Ragona, G., and Mercola, D. (2006). The Transcription Factor Egr1 Is a Direct Regulator of Multiple Tumor Suppressors Including TGFβ1, PTEN, P53, and Fibronectin. *Cancer Gene Ther.* 13, 115–124. doi:10.1038/sj.cgt.7700896

Campone, P. M., Gouraud, W., Guérin-Charbonnel, C., Leux, C., Ricolleau, G., and Campion, L. (2012). Bc-GenExMiner: an Easy-To-Use Online Platform for Gene Prognostic Analyses in Breast Cancer. *Breast Cancer Res. Treat.* 131, 765–775. doi:10.1007/s10549-011-1457-7

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

Chen, Y., Li, L., and Zhang, J. (2021). Cell Migration Inducing Hyaluronidase 1 (CEMIP) Activates STAT3 Pathway to Facilitate Cell Proliferation and Migration in Breast Cancer. *J. Receptors Signal Transduction* 41, 145–152. doi:10.1080/10799893.2020.1800732

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, 3958–3967. doi:10.1016/j.ygeno.2020.06.044

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., et al. (2015). Breast Cancer Intrinsic Subtype Classification, Clinical Use and Future Trends. *Am. J. Cancer Res.* 5, 2929–2943. doi:10.1534/g3.114.014894

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.* 30, 207–210. doi:10.1093/nar/30.1.207

Eich, M.-L., Athar, M., Ferguson, J. E., 3rd, and Varambally, S. (2020). EZH2-Targeted Therapies in Cancer: Hype or a Reality. *Cancer Res.* 80, 5449–5458. doi:10.1158/0008-5472.can-20-2147

Evensen, N. A., Kuscu, C., Nguyen, H.-L., Zarrabi, K., Dufour, A., Kadam, P., et al. (2013). Unraveling the Role of KIAA1199, a Novel Endoplasmic Reticulum Protein, in Cancer Cell Migration. *J. Natl. Cancer Inst.* 105, 1402–1416. doi:10.1093/jnci/djt224

Fan, Z., Duan, J., Wang, L., Xiao, S., Li, L., Yan, X., et al. (2019). PTK2 Promotes Cancer Stem Cell Traits in Hepatocellular Carcinoma by Activating Wnt/beta-Catenin Signaling. *Cancer Lett.* 450, 132–143. doi:10.1016/j.canlet.2019.02.040

Fink, S. P., Myeroff, L. L., Kariv, R., Platzer, P., Xin, B., Mikkola, D., et al. (2015). Induction of KIAA1199/CEMIP Is Associated with colon Cancer Phenotype and Poor Patient Survival. *Oncotarget* 6, 30500–30515. doi:10.18632/oncotarget.5921

Gao, G., Wang, Z., Qu, X., and Zhang, Z. (2020). Prognostic Value of Tumor-Infiltrating Lymphocytes in Patients with Triple-Negative Breast Cancer: a Systematic Review and Meta-Analysis. *BMC Cancer* 20, 179. doi:10.1186/s12885-020-6668-z

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144, 646–674. doi:10.1016/j.cell.2011.02.013

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast Cancer. *Nat. Rev. Dis. Primers* 5, 66. doi:10.1038/s41572-019-0111-2

Hsieh, I.-y., He, J., Wang, L., Lin, B., Liang, Z., Lu, B., et al. (2020). H3K27me3 Loss Plays a Vital Role in CEMIP Mediated Carcinogenesis and Progression of Breast Cancer with Poor Prognosis. *Biomed. Pharmacother.* 123, 109728. doi:10.1016/j.biopha.2019.109728

Ishii, G., Ochiai, A., and Neri, S. (2016). Phenotypic and Functional Heterogeneity of Cancer-Associated Fibroblast within the Tumor Microenvironment. *Adv. Drug Deliv. Rev.* 99, 186–196. doi:10.1016/j.addr.2015.07.007

Jami, M.-S., Hou, J., Liu, M., Varney, M. L., Hassan, H., Dong, J., et al. (2014). Functional Proteomic Analysis Reveals the Involvement of KIAA1199 in Breast Cancer Growth, Motility and Invasiveness. *BMC Cancer* 14, 194–2407. doi:10.1186/1471-2407-14-194

Jézéquel, P., Gouraud, W., Ben Azzouz, F., Guérin-Charbonnel, C., Juin, P. P., and Lasla, H. (2021). Bc-GenExMiner 4.5: New Mining Module Computes Breast Cancer Differential Gene Expression Analyses. *Database (Oxford)* 2021. doi:10.1093/database/baab007,

Jezequel, P., Frenel, J.-S., Campion, L., Guerin-Charbonnel, C., Gouraud, W., Ricolleau, G., et al. (20132013). Bc-GenExMiner 3.0: New Mining Module Computes Breast Cancer Gene Expression Correlation Analyses. *Database* 2013, bas060. doi:10.1093/database/bas060

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New Approach for Understanding Genome Variations in KEGG. *Nucleic Acids Res.* 47, D590–D595. doi:10.1093/nar/gky962

Kuscu, C., Evensen, N., Kim, D., Hu, Y.-J., Zucker, S., and Cao, J. (2012). Transcriptional and Epigenetic Regulation of KIAA1199 Gene Expression in Human Breast Cancer. *PLoS One* 7, e44661. doi:10.1371/journal.pone.0044661

Lachmann, A., Torre, D., Keenan, A. B., Jagodnik, K. M., Lee, H. J., Wang, L., et al. (2018). Massive Mining of Publicly Available RNA-Seq Data from Human and Mouse. *Nat. Commun.* 9, 1366. doi:10.1038/s41467-018-03751-6

Lanczky, B. A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An Online Survival Analysis Tool to Rapidly Assess the Effect of 22,277 Genes on Breast Cancer Prognosis Using Microarray Data of 1,809 Patients. *Breast Cancer Res. Treat.* 123, 725–731. doi:10.1007/s10549-009-0674-9

Li, L., Pan, Y., Mo, X., Wei, T., Song, J., Luo, M., et al. (2020). A Novel Metastatic Promoter CEMIP and its Downstream Molecular Targets and Signaling Pathway of Cellular Migration and Invasion in SCLC Cells Based on Proteome Analysis. *J. Cancer Res. Clin. Oncol.* 146, 2519–2534. doi:10.1007/s00432-020-03308-5

Li, L., Yan, L.-H., Manoj, S., Li, Y., and Lu, L. (2017). Central Role of CEMIP in Tumorigenesis and its Potential as Therapeutic Target. *J. Cancer* 8, 2238–2246. doi:10.7150/jca.19295

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

Liu, J., Yan, W., Han, P., and Tian, D. (2021). The Emerging Role of KIAA1199 in Cancer Development and Therapy. *Biomed. Pharmacother.* 138, 111507. doi:10.1016/j.biopha.2021.111507

Liu, S., Song, A., Wu, Y., Yao, S., Wang, M., Niu, T., et al. (2021). Analysis of Genomics and Immune Infiltration Patterns of Epithelial-Mesenchymal Transition Related to Metastatic Breast Cancer to Bone. *Translational Oncol.* 14, 100993. doi:10.1016/j.tranon.2020.100993

Lo Sardo, F., Pulito, C., Sacconi, A., Korita, E., Sudol, M., Strano, S., et al. (2021). YAP/TAZ and EZH2 Synergize to Impair Tumor Suppressor Activity of TGFBR2 in Non-small Cell Lung Cancer. *Cancer Lett.* 500, 51–63. doi:10.1016/j.canlet.2020.11.037

Lou, W., Wang, W., Chen, J., Wang, S., and Huang, Y. (2021). ncRNAs-Mediated High Expression of SEMA3F Correlates with Poor Prognosis and Tumor Immune Infiltration of Hepatocellular Carcinoma. *Mol. Ther. - Nucleic Acids* 24, 845–855. doi:10.1016/j.omtn.2021.03.014

Mehta, V., Chander, H., and Munshi, A. (2021). Complex Roles of Discoidin Domain Receptor Tyrosine Kinases in Cancer. *Clin. Transl Oncol.* 23, 1497–1510. doi:10.1007/s12094-021-02552-6

Michishita, E., Garces, G., Barrett, J., and Horikawa, I. (2006). Upregulation of the KIAA1199 Gene Is Associated with Cellular Mortality. *Cancer Lett.* 239, 71–77. doi:10.1016/j.canlet.2005.07.028

Miller, L. D., Smeds, J., George, J., Vega, V. B., Vergara, L., and Ploner, A. (2005). An Expression Signature for P53 Status in Human Breast Cancer Predicts Mutation Status, Transcriptional Effects, and Patient Survival. *Proc. Natl. Acad. Sci. U S A.* 102, 13550–13555. doi:10.1073/pnas.0506230102

Rodrigues, G., Hoshino, A., Kenific, C. M., Matei, I. R., Steiner, L., Freitas, D., et al. (2019). Tumour Exosomal CEMIP Protein Promotes Cancer Cell Colonization in Brain Metastasis. *Nat. Cel Biol.* 21, 1403–1412. doi:10.1038/s41556-019-0404-4

Roma-Rodrigues, C., Mendes, R., Baptista, P. V., and Fernandes, A. R. (2019). Targeting Tumor Microenvironment for Cancer Therapy. *Int. J. Mol. Sci.* 20. doi:10.3390/ijms20040840

Savas, P., Salgado, R., Denkert, C., Sotiriou, C., Darcy, P. K., Smyth, M. J., et al. (2016). Clinical Relevance of Host Immunity in Breast Cancer: from TILs to the Clinic. *Nat. Rev. Clin. Oncol.* 13, 228–241. doi:10.1038/nrclinonc.2015.215

Schmid, P., Adams, S., Rugo, H. S., Schneeweiss, A., Barrios, C. H., Iwata, H., et al. (2018). Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* 379, 2108–2121. doi:10.1056/nejmoa1809615

Schreiber, R. D., Old, L. J., and Smyth, M. J. (2011). Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion. *Science* 331, 1565–1570. doi:10.1126/science.1203486

Shaulian, E. (2010). AP-1 - the Jun Proteins: Oncogenes or Tumor Suppressors in Disguise? *Cell Signal.* 22, 894–899. doi:10.1016/j.cellsig.2009.12.008

Shen, F., Zong, Z.-h., Liu, Y., Chen, S., Sheng, X.-j., and Zhao, Y. (2019). CEMIP Promotes Ovarian Cancer Development and Progression via the PI3K/AKT Signaling Pathway. *Biomed. Pharmacother.* 114, 108787. doi:10.1016/j.biopha.2019.108787

Shostak, K., Zhang, X., Hubert, P., Göktuna, S. I., Jiang, Z., Klevernic, I., et al. (2014). NF-κB-induced KIAA1199 Promotes Survival through EGFR Signalling. *Nat. Commun.* 5, 5232. doi:10.1038/ncomms6232

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* 27, 431–432. doi:10.1093/bioinformatics/btq675

Smyth, G. K. (2005). "Limma: Linear Models for Microarray Data." in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, (Springer: New York, NY), 397–420. doi:10.1007/0-387-29362-0_23

Subbiah, V., and Cote, G. J. (2020). Advances in Targeting RET-dependent Cancers. *Cancer Discov.* 10, 498–505. doi:10.1158/2159-8290.cd-19-1116

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). Review the Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *wo* 1A, 68–77. doi:10.5114/wo.2014.47136

Turner, N. C., Slamon, D. J., Ro, J., Bondarenko, I., Im, S.-A., Masuda, N., et al. (2018). Overall Survival with Palbociclib and Fulvestrant in Advanced Breast Cancer. *N. Engl. J. Med.* 379, 1926–1936. doi:10.1056/nejmoa1810527

Verigos, J., and Magklara, A. (2015). Revealing the Complexity of Breast Cancer by Next Generation Sequencing. *Cancers (Basel)* 7, 2183–2200. doi:10.3390/cancers7040885

Waks, A. G., and Winer, E. P. (2019). Breast Cancer Treatment. *JAMA* 321, 288–300. doi:10.1001/jama.2018.19323

Xia, Y., Rao, L., Yao, H., Wang, Z., Ning, P., and Chen, X. (2020). Engineering Macrophages for Cancer Immunotherapy and Drug Delivery. *Adv. Mater.* 32, e2002054. doi:10.1002/adma.202002054

Yoshida, H., Nagaoka, A., Kusaka-Kikushima, A., Tobiishi, M., Kawabata, K., Sayo, T., et al. (2013). KIAA1199, a Deafness Gene of Unknown Function, Is a New Hyaluronan Binding Protein Involved in Hyaluronan Depolymerization. *Proc. Natl. Acad. Sci.* 110, 5612–5617. doi:10.1073/pnas.1215432110

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

Zhai, X., Wang, W., Ma, Y., Zeng, Y., Dou, D., Fan, H., et al. (2020). Serum KIAA1199 Is an Advanced-Stage Prognostic Biomarker and Metastatic Oncogene in Cholangiocarcinoma. *AGING* 12, 23761–23777. doi:10.18632/aging.103964

Zhang, Y., Jia, S., and Jiang, W. G. (2014). KIAA1199 and its Biological Role in Human Cancer and Cancer Cells (Review). *Oncol. Rep.* 31, 1503–1508. doi:10.3892/or.2014.3038

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fgene.2021.681809

Identification and Validation of DEPDC1B as an Independent Early Diagnostic and Prognostic Biomarker in Liver Hepatocellular Carcinoma

*Xiaoyan Fan[1], Junye Wen[2], Lei Bao[2], Fei Gao[1], You Li[3] and Dongwei He[3]**

[1]*Department of Oncology, Hebei General Hospital, Shijiazhuang, China,* [2]*Department of Hepatobiliary Surgery, Hebei General Hospital, Shijiazhuang, China,* [3]*Laboratory of Pathology, Hebei Cancer Institute, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China*

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Myvizhi Esai Selvan,
Icahn School of Medicine at Mount Sinai, United States
Gan Luan,
New Jersey Institute of Technology, United States

***Correspondence:**
Dongwei He
dongweihe@hebmu.edu.cn

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

Received: *17 March 2021*
Accepted: *10 December 2021*
Published: *13 January 2022*

Citation:
Fan X, Wen J, Bao L, Gao F, Li Y and He D (2022) Identification and Validation of DEPDC1B as an Independent Early Diagnostic and Prognostic Biomarker in Liver Hepatocellular Carcinoma. Front. Genet. 12:681809. doi: 10.3389/fgene.2021.681809

Liver hepatocellular carcinoma (LIHC) is one of the most lethal tumors worldwide, and while its detailed mechanism of occurrence remains unclear, an early diagnosis of LIHC could significantly improve the 5-years survival of LIHC patients. It is therefore imperative to explore novel molecular markers for the early diagnosis and to develop efficient therapies for LIHC patients. Currently, DEPDC1B has been reported to participate in the regulation of cell mitosis, transcription, and tumorigenesis. To explore the valuable diagnostic and prognostic markers for LIHC and further elucidate the mechanisms underlying DEPDC1B-related LIHC, numerous databases, such as Oncomine, Gene Expression Profiling Interactive Analysis (GEPIA), UALCAN, Kaplan-Meier plotter, and The Cancer Genome Atlas (TCGA) were employed to determine the association between the expression of DEPDC1B and prognosis in LIHC patients. Generally, the DEPDC1B mRNA level was highly expressed in LIHC tissues, compared with that in normal tissues ($p < 0.01$). High DEPDC1B expression was associated with poor overall survival (OS) in LIHC patients, especially in stage II, IV, and grade I, II, III patients (all $p < 0.05$). The univariate and multivariate Cox regression analysis showed that DEPDC1B was an independent risk factor for OS among LIHC patients (HR = 1.3, 95% CI: 1.08–1.6, $p = 0.007$). In addition, the protein expression of DEPDC1B was validated using Human Protein Atlas database. Furthermore, the expression of DEPDC1B was confirmed by quantitative real-time polymerase chain reaction (qRT-PCR) assay using five pairs of matched LIHC tissues and their adjacent noncancerous tissues. The KEGG pathway analysis indicated that high expression of DEPDC1B may be associated with several signaling pathways, such as MAPK signaling, the regulation of actin cytoskeleton, p53 signaling, and the Wnt signaling pathways. Furthermore, high DEPDC1B expression may be significantly associated with various cancers. Conclusively, DEPDC1B may be an independent risk factor for OS among LIHC cancer patients and may be used as an early diagnostic marker in patients with LIHC.

Keywords: DEPDC1B, prognosis, diagnosis, biomarker, ROC curve, liver hepatocellular carcinoma

INTRODUCTION

Liver hepatocellular carcinoma (LIHC) is the most common type of lethal cancer and is the fourth leading cause of death among cancer patients worldwide (Bray et al., 2018; Villanueva, 2019). The World Health Organization estimates that more than 1 million patients will die from LIHC in 2030 (Villanueva, 2019). Several risk agents, such as chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), and exposure to alcohol and aflatoxins are significantly involved in the intrinsic mechanisms (Wang et al., 2002; El-Serag and Rudolph 2007). Furthermore, recent studies have established that accumulated genetic alterations, such as somatic mutations, and chromosomal aberrations may be involved in this process (Villanueva, 2019). Somatic mutations in the TERT promoter, which is a recurrent insertion site for the genome of HBV, are the most frequent genetic alterations (approximately 60% of cases) (Schulze et al., 2016). Other mutated genes could affect the cell cycle (TP53, approximately 30% of cases) and WNT signaling (CTNNB1 and AXIN1 occurring in approximately 30 and 10% of cases, respectively) (Villanueva, 2019). In addition, chromatin remodeling (ARID1A and ARID2) may also account for approximately 10 and 5% of LIHC patients, respectively (Villanueva, 2019). Though the molecular mechanisms of LIHC remain far from being fully understood, the survival rate of LIHC patients could be improved by more than 50% with early detection of hepatocellular carcinoma (Kim et al., 2016; Lu et al., 2020). Conversely, the early diagnosis of LIHC is far from satisfactory, hence the exploration of novel molecular markers for early diagnosis and therapies is of great value for LIHC patients.

DEP domain-containing protein 1B (DEPDC1B), which is located at chromosome 5 (5q12.1), was initially discovered by mRNA expression profiling in MDA-MB231 human breast cancer cells (Boudreau et al., 2007). DEPDC1B contains two conserved domains: the DEP domain and the RhoGAP domain. The DEP domain is a globular domain containing about 90 amino acids, and was named from three proteins: *Drosophila* disheveled, *Caenorhabditis elegans* EGL-10, and *mammalian* Pleckstrin (Wong et al., 2000; Wharton, 2003). Being more than just a membrane anchor, the DEP domain could negatively interact with charged phospholipids located in membranes to activate Wnt signaling (Sokol, 2000). In addition, the DEP domain could interact directly with the G protein-coupled receptors to regulate GPCR signaling pathways (Ballon et al., 2006; Chen and Hamm, 2006). Moreover, the RhoGAP domain is a key participator in Rho GTPase signaling (Martemyanov et al., 2003). As a protein accumulating during G2 phase, Marchesi et al. have reported that the role of DEPDC1B in coordinating de-adhesion and cell-cycle progression at mitotic entry (Marchesi et al., 2014). Furthermore, increasing evidence in recent years suggests that the overexpression of DEPDC1B is associated with tumor aggressiveness and poor prognosis in cancers, such as oral cancer (Su et al., 2014), malignant melanoma (Xu et al., 2019), glioblastoma (Chen et al., 2020), non-small cell lung (Yang et al., 2014), and pancreatic cancers (Mishra et al., 2019; Liu et al., 2020). These findings strongly suggest that DEPDC1B could potentially contribute to tumorigenesis. However, the role of DEPDC1B in LIHC remains unclear.

In the present study, public databases were used to analyze the correlation between the expression of DEPDC1B and patient diagnostic and prognosis for LIHC. In addition, the results were confirmed by quantitative real-time polymerase chain reaction (qPCR) assay, and the findings from the study suggest that DEPDC1B may have a utility as a potential biomarker for the diagnosis and prognosis in LIHC patients.

MATERIALS AND METHODS

Oncomine Database Analysis

The expression level of DEPDC1B in LIHC was determined by Oncomine database analysis (https://www.oncomine.org/resource/login.html) (Rhodes et al., 2007). The thresholds ($p \leq 0.0001$, fold change: 2, and gene rank: Top 10%) were considered statistically significant.

UALCAN Database Analysis

UALCAN is a comprehensive, use-friendly, and interactive web resource for analyzing cancer OMICS data (http://ualcan.path.uab.edu/index.html). UALCAN is designed to provide easy access to publicly available cancer OMICS data (TCGA, MET500, and CPTAC databases), allowing users to identify biomarkers of interest (Chandrashekar et al., 2017). In this study, DEPDC1B expression was analyzed from the TCGA database and $p < 0.01$ was considered statistically significant.

TCGA Database Analysis

Gene expression data and patient data for LIHC were downloaded from the Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/) using the GDC data transfer tool. Gene expression data were analyzed using R (version: 3.6.1) with related R packages. Clinical parameters, such as age, gender, survival, and tumor grade and stage were extracted from the patient data and then matched to each patient using a PERL script.

GEPIA Database Analysis

The Gene Expression Profiling Interactive Analysis (GEPIA) platform (http://gepia.cancer-pku.cn/) is a newly developed interactive web server for analyzing RNA sequencing expression data for 9,736 tumors and 8,587 normal samples from TCGA and the Genotype-Tissue Expression database projects, using a standard processing pipeline (Tang et al., 2017). The database was used to evaluate DEPDC1B expression in LIHC patients. In the survival analysis, the threshold was determined according to the following values: group cutoff: median; cutoff-high (%): 50; cutoff-low (%): 50.

Kaplan-Meier Plotter Database Analysis

Based on the Kaplan Meier plotter (http://kmplot.com/analysis/) (Nagy et al., 2018), the correlation between DEPDC1B mRNA

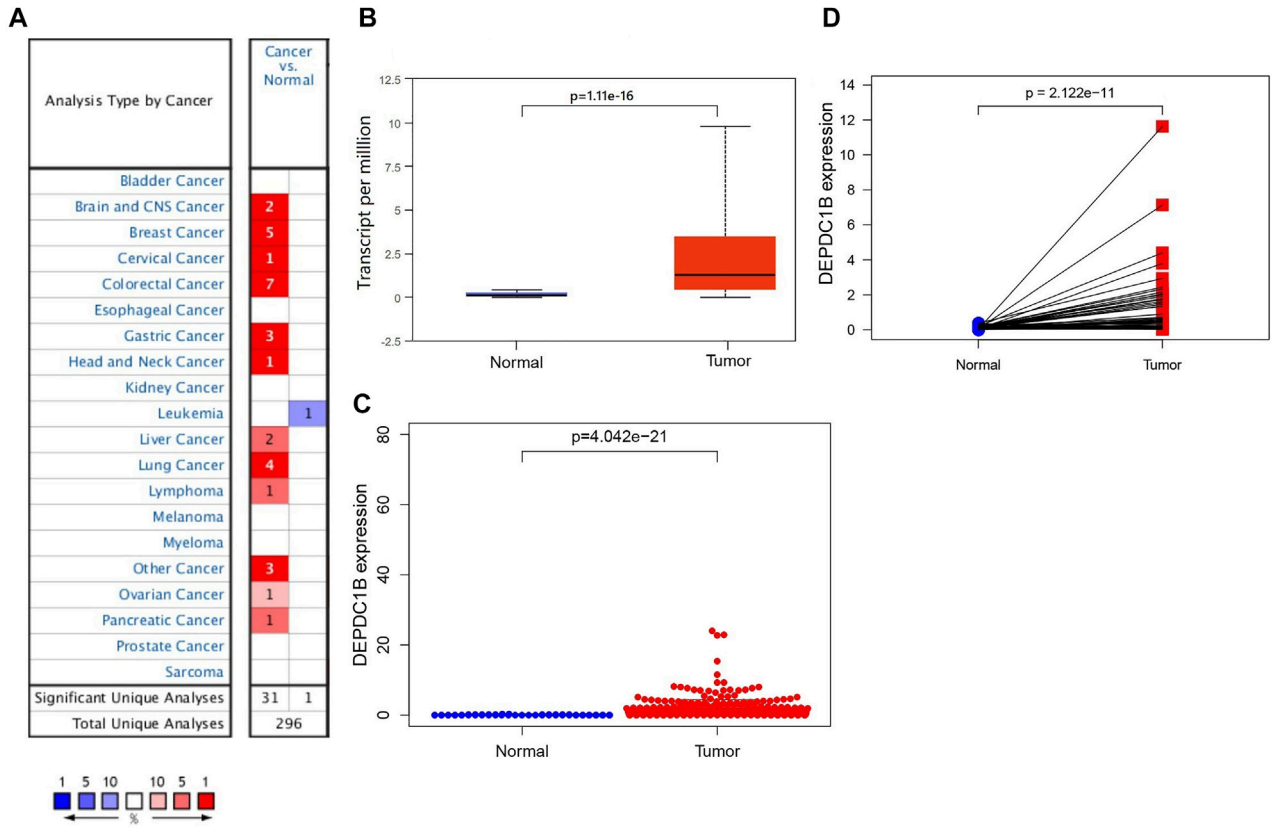


FIGURE 1 | DEPDC1B expression in LIHC tissues using online databases. **(A)** Expression of DEPDC1B (normal compared with cancer tissue) were analyzed using Oncomine database. The table shows the number of significant analyses across the whole Oncomine database (715 datasets and 86.733 samples). DEPDC1B shows high/low expression ratio of 31/1 from 32 analyses. The cell color is determined by the best gene rank percentile for the analyses within the cell. The number in each cell represents the number of analyses that meet the threshold within those analysis and cancer types. The gene rank was analyzed by percentile of target gene in the top of all genes measured in each research. **(B)** UALCAN database. The number of patients in normal and primary tumor groups was 50 and 371, respectively. p = 1.11e-16. **(C)** TCGA database using R script. Normal (n = 50), Tumor (n = 374). p = 4.042e-21. **(D)** DEPDC1B expression in a paired comparison of LIHC and their adjacent tissues. Data were extracted from the TCGA database. n = 50. p = 2.122e-11.

expression and survival in LIHC was analyzed using RNA-seq data. The patients were divided into low and high expression groups according to median expression, and the cutoff value was set to "auto select".

ROC Curve Generation

Receiver operating characteristic (ROC) curves were generated to evaluate the diagnostic value of DEPDC1B using IBM SPSS Statistics 26. And the area under the curve (AUC) was also determined and showed in the panel.

KEGG Analysis

To identify the potential mechanisms of DEPDC1B expression in LIHC, KEGG analysis was performed to detect whether a priori defined set of genes showed statistically significant differential expression between the high and low DEPDC1B expression groups using GSEA (Subramanian et al., 2005). Gene sets with a normal p-value < 0.05 and false discovery rate (FDR) < 0.05 were significantly enriched.

Immunohistochemistry Validation of DEPDC1B Expression Using Human Protein Atlas Database

To further confirm the expression level of DEPDC1B in LIHC tissues, DEPDC1B protein expression was analyzed in clinical specimens using The Human Protein Atlas (https://www.proteinatlas.org/).

RNA Extraction and qPCR Assay

qPCR was performed to determine the expression of DEPDC1B mRNA in LIHC and their adjacent tissues. Briefly, total RNA from the surgically obtained paired tissues (n = 5) was isolated using TRI Reagent RNA Isolation Reagent (Sigma-Aldrich) according to the manufacturer's instructions. A reverse transcription system was used to obtain the first-strand template Complementary DNA (cDNA). The primer sequences were used as follows: DEPDC1B: 5′- GAG CTACCAGGCTGTGGAAT-3′ (forward) and 5′- GCCGAAGTT TTGACTGCACC -3′ (reverse); GAPDH: 5′-CCATGTTCGTCATGGGTGTGAACCA-3′(forward) and 5′-GCCAGTAGAGGCAGGGATGATGTTC-3′(reverse) (Li et al.,

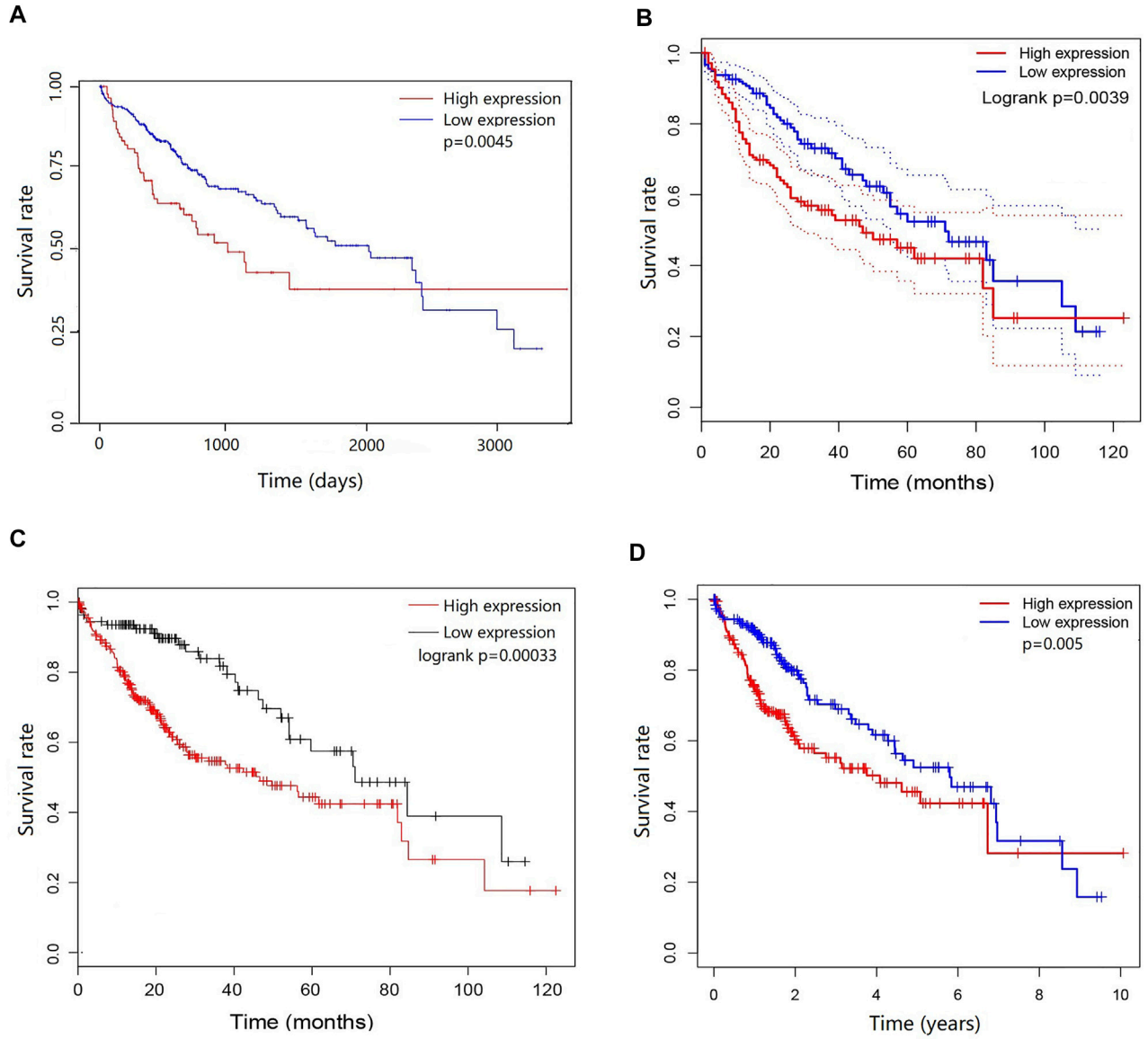


FIGURE 2 | Prognostic potential of DEPDC1B in LIHC. Four online tools were used to evaluate the relationships between DEPDC1B expression and patient survival. **(A)** OS in UALCAN database, the samples were divided into high DEPDC1B expression (n = 88) and low/medium DEPDC1B expression (n = 277) groups according to the median of DEPDC1B expression. p = 0.0045 **(B)** OS in GEPIA database. The samples were divided into high DEPDC1B expression (n = 181) and low DEPDC1B expression (n = 181) groups according to the median of DEPDC1B expression. p = 0.0039 **(C)** OS in Kaplan-Meier Plotter database. The number of patients in high and low groups was 252, 112, respectively. p = 0.00033 **(D)** OS in TCGA database, p = 0.005. The samples were divided into high DEPDC1B expression (n = 260) and low DEPDC1B expression (n = 110) groups according to the median of DEPDC1B expression. p = 0.005. OS: overall survival.

2020; Zhang et al., 2020). The expression of GAPDH was considered as an internal control. Each reaction was performed in triplicate. The study was approved by the Institute Research Ethics Committee at the Fourth Hospital of Hebei Medical University.

Statistical Analysis

Two-tailed p values less than 0.05 were considered statistically significant. TCGA-associated expression and prognosis analyses were conducted using R software (version 3.6.1). The univariate Cox analysis was used to select potential prognostic factors, and multivariate Cox analysis was performed to verify the correlations between DEPDC1B expression and survival, along with other clinical features.

RESULTS

High DEPDC1B mRNA Expression in LIHC

Oncomine and UALCAN online databases were used to determine the expression levels of DEPDC1B mRNA in LIHC and their normal tissues. Both databases showed higher levels of DEPDC1B expression in LIHC tissues when compared with normal tissues (all $p < 0.01$;

TABLE 1 | Clinical characteristics of the 349 LIHC patients downloaded from the TCGA database.

Characteristic	N (%)
Age (years)	
≤60	174 (49.86)
>60	175 (50.14)
Gender	
Female	110 (31.52)
Male	239 (68.48)
Stage	
I	173 (49.57)
II	85 (24.36)
III	86 (24.64)
IV	5 (1.43)
T classification	
T1	175 (50.14)
T2	87 (24.93)
T3	77 (22.06)
T4	10 (2.87)
Survival status	
Death	113 (32.39)
Survival	236 (67.61)

Figures 1A,B). For validation, we downloaded RNA-seq data for LIHC from TCGA database and analyzed the expression of DEPDC1B using R. **Figure 1C** indicates that when compared with that in normal tissues, the DEPDC1B expression was significantly upregulated in TCGA LIHC tissues (p = 4.042e-21; **Figure 1C**). Additionally, the pair-wise comparison of TCGA LIHC tissues and their adjacently matched tissues revealed a significantly higher level of DEPDC1B expression in the former (p = 2.122e-11; **Figure 1D**). These results suggest that the expression of DEPDC1B is highly elevated in LIHC when compared with normal tissues.

Prognostic Potential of DEPDC1B in LIHC

To determine the potential prognostic significance of the expression of DEPDC1B in LIHC, UALCAN, GEPIA, Kaplan-Meier Plotter, and R script were used to evaluate the relationships between DEPDC1B expression and the survival rate of patients. **Figure 2** shows that high DEPDC1B mRNA expression was significantly associated with poor overall survival (OS) in LIHC patients using UALCAN (p = 0.0045; **Figure 2A**), GEPIA (logrank p = 0.0039, HR = 1.7; **Figure 2B**), Kaplan-Meier Plotter [logrank p = 0.00033, HR = 2.14 (1.4–3.28); **Figure 2C**] analyses. In addition, clinical data for LIHC cases were downloaded from TCGA database, and the OS was subsequently analyzed using R (**Table 1**; **Figure 2D**). Consistent with previous findings reported above, high DEPDC1B expression was significantly and negatively associated with the survival of patients with TCGA LIHC (p = 0.005; **Figure 2D**). These results suggest that high DEPDC1B expression could lead to a poor prognosis in patients with LIHC.

Association Between DEPDC1B Expression and Clinical Characteristics of LIHC Patients

Since the expression of DEPDC1B was negatively associated with the prognosis of LIHC patients, to better understand the relevance, and mechanisms of DEPDC1B in LIHC, we determined the relationship

TABLE 2 | Correlation of DEPDC1B expression and clinical prognosis in LIHC with different clinical characteristics by Kaplan-Meier plotter.

Clinical characteristics	OS		RFS	
	HR (95%CI)	Logrank P	HR (95%CI)	Logrank P
Gender				
Female	1.96 (1.11–3.46)	0.019*	2.82 (1.54–5.16)	0.00046***
Male	2.2 (1.4–3.44)	4.42e-04***	1.69 (1.08–2.65)	0.02*
Race				
White	1.86 (1.13–3.08)	0.014*	2.13 (1.27–3.58)	0.0035**
Asian	4.52 (2.47–8.27)	8.9e-8***	2.51 (1.3–4.83)	0.0044**
Stage				
1	1.78 (0.97–3.27)	0.06	1.63 (0.94–2.81)	0.078
2	3.08 (1.15–8.23)	0.018*	–1.9 (0.79–4.58)	–0.14
3	2.23 (1.23–4.05)	0.0067**	2.23 (1.23–4.05)	0.0067**
Grade				
1	5.02 (1.66–15.13)	–0.0021**	–3.08 (0.7–13.66)	–0.12
2	2.53 (1.31–4.88)	0.0043**	–2.13 (1.21–3.74)	–0.0075**
3	1.98 (1.09–3.61)	0.022*	1.78 (1.02–3.1)	0.04*
Vascular invasion				
None	1.87 (1.12–3.14)	0.016*	1.44 (0.88–2.361)	0.15
Micro	1.88 (0.79–4.46)	0.15	1.88 (0.79–4.46)	0.15
Risk factors				
Alcohol consumption				
Yes	1.84 (0.97–3.5)	0.059	2.6 (1.41–4.79)	0.0015**
None	2 (1.26–3.18)	0.0027**	1.63 (1.05–2.55)	0.029*
Hepatitis virus				
Yes	1.39 (0.73–2.67)	0.31	1.27 (0.74–2.15)	0.39
None	3.68 (1.97–6.87)	1.4e-05***	3.92 (2.01–7.62)	1.7e-05***

- values indicate the sample number too low for meaningful analysis.

values indicate* p < *0.05, ** values indicate* p < *0.01, * values indicate* p < *0.001.*

TABLE 3 | Univariate and multivariate analysis of the correlation of DEPDC1B expression with OS among LIHC patients.

Parameter	Univariate analysis			Multivariate analysis		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Age	1.01	0.995–1.025	0.177	1.015	1.000–1.031	0.057
Gender	0.82	0.557–1.209	0.317	1.081	0.716–1.632	0.710
Grade	1.12	0.868–1.446	0.382	1.104	0.838–1.453	0.483
Stage	1.67	1.400–2.056	1.12e-06***	1.053	0.468–2.367	0.901
T classification	1.65	1.357–2.011	5.82e-07***	1.549	0.719–3.336	0.263
DEPDC1B	1.07	1.014–1.124	0.01*	1.330	1.080–1.637	0.007**

values indicate* p < *0.05, ** values indicate* p < *0.01, * values indicate* p < *0.001.*

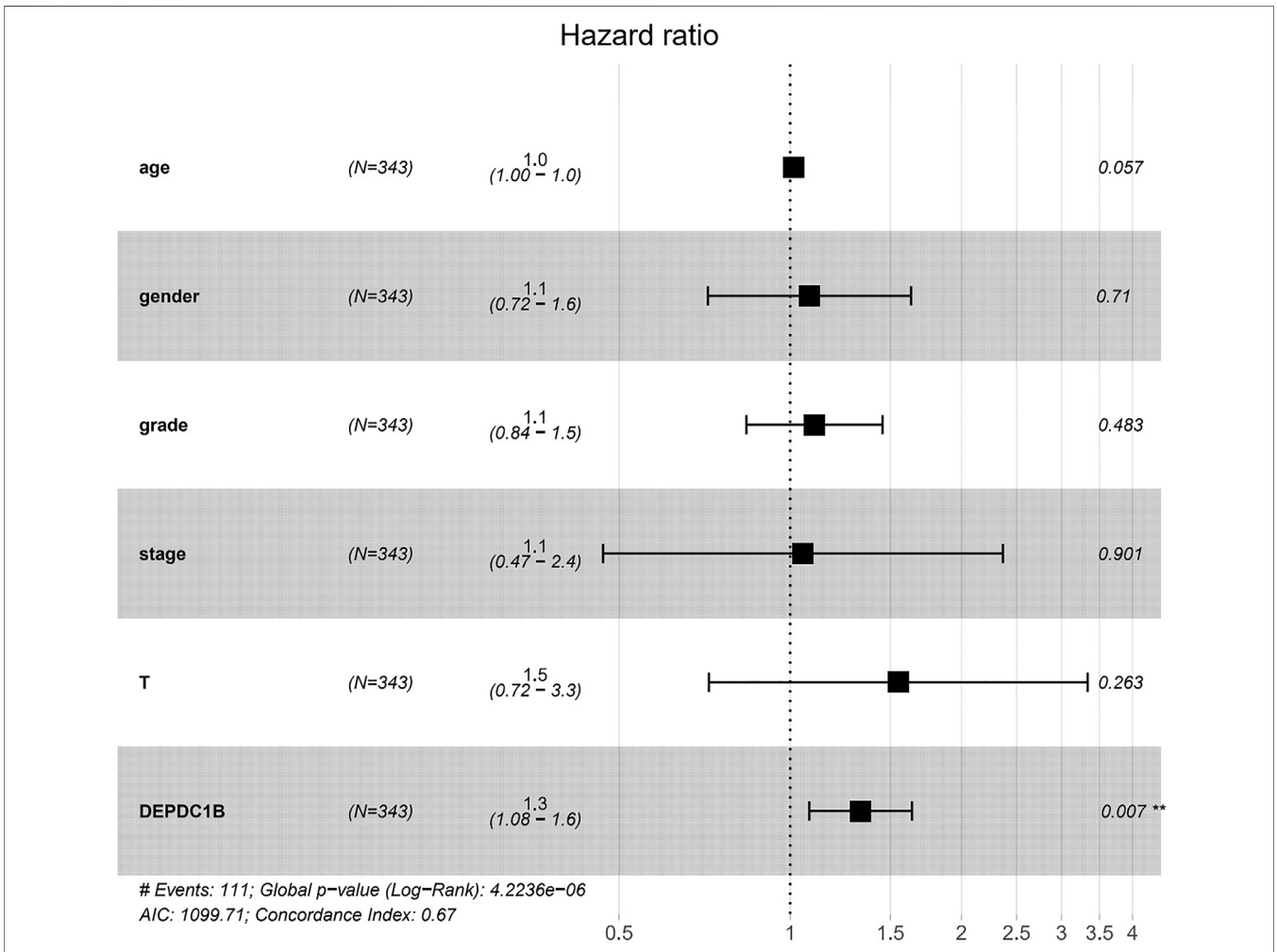


FIGURE 3 | DEPDC1B is an independent risk factor for OS among LIHC patients using multivariate Cox analysis. ***p* = 0.007.

between the expression of DEPDC1B and the clinicopathological characteristics of LIHC patients using the Kaplan-Meier plotter tool. DEPDC1B expression was negatively associated with OS in female and male patients (p = 0.019, p = 4.42e-04, respectively; **Table 2**; **Supplementary Figure S1**), and patients from white and Asian races (p = 0.014, p = 8.9e-08, respectively). High DEPDC1B mRNA expression has a significant correlation with low OS in stage II, IV, and grade I, II, III patients (p = 0.018, p = 0.0067, p = 0.0021, p = 0.0043, p = 0.022, respectively). Also, similar significant results were found between the expression of DEPDC1B and Relapse Free Survival (RFS) in gender, race stage, and grade (**Table 2**; **Supplementary Figure S1**; all p < 0.05). Interestingly, in patients with risk factors (especially alcohol consumption and hepatitis virus), the expression of DEPDC1B had no significant impact on the OS of patients with LIHC (p = 0.059, p = 0.31, respectively). These results suggest that high DEPDC1B expression significantly

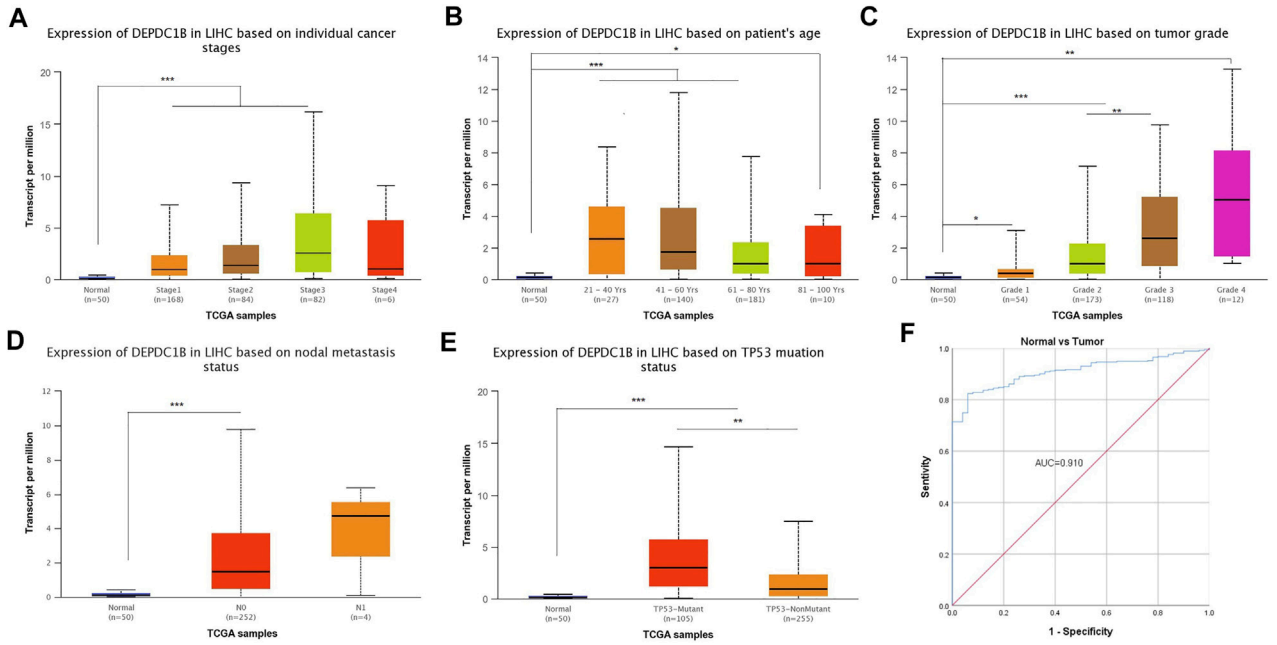


FIGURE 4 | Diagnostical potential of DEPDC1B expression in LIHC. Association between DEPDC1B expression and clinical characters: **(A)** cancer stage. Normal vs Stage 1, p = 1.62e-12; Normal *vs.* Stage 2, p = 3.44e-06; Normal *vs.* Stage 3, p = 2.59e-07; Stage 1 *vs.* Stage 3, p = 2.15e-03. **(B)** patient's age. Normal *vs.* Age (21–40 Years), p = 9.70e-05; Normal *vs.* Age (41–60 Years), p = 6.92e-11; Normal *vs.* Age (61–80 Years), p = 6.81e-10, Normal *vs.* Age (81–100 Years), p = r0.046. **(C)** tumor grade. Normal vs Grade 1, p = 4.80e-02; Normal vs Grade 2, p = 2.85e-12; Normal vs Grade 3, p = 8.09e-12; Normal *vs.* Grade 4, p = 1.30e-03. Grade 2 *vs.* Grade 3, p = 3.30e-03. **(D)** LIHC nodal metastasis status. Normal *vs.* N0, p < 1E-12. **(E)** TP53 mutation status. Normal *vs.* TP53-Mutant, p < 1E-12; Normal *vs.* TP53-NonMutant, p = 1.99e-11; TP53-Mutant *vs.* TP53-NonMutant, p = 1.94e-03. **(F)** Diagnosis value of DEPDC1B using ROC analysis. AUC: area under ROC curve. Student's t test was used to generate a p-value. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

affects the OS and RFS of LIHC patients exhibiting most clinical characteristics.

High DEPDC1B Expression is an Independent Risk Factor for OS Among LIHC Patients

To investigate whether DEPDC1B is an independent risk factor for OS in LIHC patients, univariate and multivariate Cox analyses were performed using an R script. In the univariate Cox analysis, tumor stage, T classification, and DEPDC1B expression were all independent risk factors for OS (p = 1.12e-06, 5.82e-07, and 0.01, respectively; **Table 3**). In the multivariate Cox analysis, only the expression of DEPDC1B was found to be an independent risk factor for OS (p = 0.007, HR = 1.33, 95% CI: 1.08–1.64; **Table 3**; **Figure 3**). These findings indicate that the expression of DEPDC1B expression could be an independent risk factor for the OS of LIHC patients.

Diagnostic Value of DEPDC1B Expression in LIHC Patients

Figures 4A–E indicates the observance of significant differences in the expression of DEPDC1B and the clinical characteristics (Normal *vs.* Stages, Normal *vs.* age, Normal *vs.* grade, Normal *vs.* nodal metastasis, and Normal *vs.* TP53 mutation status, all p < 0.001) in LIHC patients. The results above indicated that the expression of DEPDC1B may be a potential diagnostic biomarker for LIHC. To further elucidate the diagnostic value of DEPDC1B in LIHC patients, ROC curves were generated using SPSS 26.0. The AUC was 0.91, which strongly suggested that the level of DEPDC1B mRNA expression might be a strong diagnostic biomarker in LIHC (**Figure 4F**).

KEGG Pathway Analysis

To further explore the possible mechanism of DEPDC1B in LIHC, the KEGG analysis was performed to clarify the DEPDC1B-associated signaling pathways. The analysis revealed that 139/178 gene sets are upregulated in the high DEPDC1B expression phenotype, and 39/178 gene sets are upregulated in the low DEPDC1B expression phenotype (**Table 4**). Gene sets differentially enriched in the high expression of DEPDC1B phenotype included several familiar signaling pathways, such as the MAPK signaling, the regulation of actin cytoskeleton, p53 signaling, and the Wnt signaling pathways (**Table 4**; **Figures 5A–D**). Furthermore, the high expression of DEPDC1B may be associated with various cancers (**Figures 5E–L**).

Validation of DEPDC1B Protein Expression Level

To evaluate the protein level of DEPDC1B, immunohistochemistry was analyzed using the Human Protein

TABLE 4 | KEGG pathways in the high DEPDC1B expression phenotype.

Gene set name	NES	NOM p-val	FDR q-val
KEGG_CELL_CYCLE	2.24	0.00	0.00
KEGG_OOCYTE_MEIOSIS	2.20	0.00	0.00
KEGG_BASE_EXCISION_REPAIR	2.11	0.00	0.00
KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	2.09	0.00	0.00
KEGG_RNA_DEGRADATION	2.08	0.00	0.00
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	2.07	0.00	0.00
KEGG_NUCLEOTIDE_EXCISION_REPAIR	2.02	0.00	0.00
KEGG_P53_SIGNALING_PATHWAY	2.02	0.00	0.00
KEGG_SPLICEOSOME	2.02	0.00	0.00
KEGG_ENDOCYTOSIS	2.00	0.00	0.00
KEGG_PATHWAYS_IN_CANCER	2.00	0.00	0.00
KEGG_DNA_REPLICATION	2.00	0.00	0.00
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY	1.99	0.00	0.00
KEGG_PYRIMIDINE_METABOLISM	1.98	0.00	0.00
KEGG_PURINE_METABOLISM	1.98	0.00	0.00
KEGG_INOSITOL_PHOSPHATE_METABOLISM	1.96	0.00	0.00
KEGG_ADHERENS_JUNCTION	1.96	0.00	0.00
KEGG_COLORECTAL_CANCER	1.95	0.00	0.00
KEGG_BASAL_TRANSCRIPTION_FACTORS	1.94	0.00	0.00
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	1.93	0.00	0.01
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	1.93	0.00	0.00
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	1.93	0.00	0.00
KEGG_HOMOLOGOUS_RECOMBINATION	1.92	0.00	0.01
KEGG_GNRH_SIGNALING_PATHWAY	1.92	0.00	0.00
KEGG_SMALL_CELL_LUNG_CANCER	1.92	0.00	0.00
KEGG_PANCREATIC_CANCER	1.92	0.00	0.00
KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	1.92	0.00	0.00
KEGG_CHRONIC_MYELOID_LEUKEMIA	1.91	0.00	0.01
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	1.90	0.00	0.01
KEGG_N_GLYCAN_BIOSYNTHESIS	1.88	0.00	0.01
KEGG_WNT_SIGNALING_PATHWAY	1.88	0.00	0.01
KEGG_RENAL_CELL_CARCINOMA	1.87	0.00	0.01
KEGG_ERBB_SIGNALING_PATHWAY	1.86	0.00	0.01
KEGG_NOTCH_SIGNALING_PATHWAY	1.86	0.00	0.01
KEGG_BLADDER_CANCER	1.85	0.00	0.01
KEGG_MISMATCH_REPAIR	1.85	0.00	0.01
KEGG_INSULIN_SIGNALING_PATHWAY	1.85	0.00	0.01
KEGG_LONG_TERM_POTENTIATION	1.83	0.00	0.01
KEGG_THYROID_CANCER	1.82	0.00	0.01
KEGG_MAPK_SIGNALING_PATHWAY	1.82	0.00	0.01
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	1.81	0.00	0.01
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	1.80	0.00	0.01
KEGG_MELANOGENESIS	1.79	0.00	0.01
KEGG_NON_SMALL_CELL_LUNG_CANCER	1.79	0.00	0.01
KEGG_ACUTE_MYELOID_LEUKEMIA	1.79	0.00	0.01
KEGG_TIGHT_JUNCTION	1.79	0.00	0.01
KEGG_MTOR_SIGNALING_PATHWAY	1.78	0.00	0.01
KEGG_GLIOMA	1.78	0.00	0.01
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	1.78	0.00	0.01
KEGG_TGF_BETA_SIGNALING_PATHWAY	1.78	0.00	0.01
KEGG_VEGF_SIGNALING_PATHWAY	1.77	0.00	0.01
KEGG_PROSTATE_CANCER	1.76	0.01	0.02
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.76	0.00	0.02
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	1.76	0.00	0.02
KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION	1.76	0.00	0.02
KEGG_APOPTOSIS	1.73	0.00	0.02
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	1.73	0.00	0.02
KEGG_REGULATION_OF_AUTOPHAGY	1.73	0.01	0.02
KEGG_VIBRIO_CHOLERAE_INFECTION	1.73	0.01	0.02
KEGG_LONG_TERM_DEPRESSION	1.73	0.00	0.02
KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	1.72	0.01	0.02
KEGG_RNA_POLYMERASE	1.72	0.02	0.02
KEGG_DORSO_VENTRAL_AXIS_FORMATION	1.72	0.01	0.02
KEGG_ENDOMETRIAL_CANCER	1.72	0.01	0.02

(Continued on following page)

TABLE 4 | (*Continued*) KEGG pathways in the high DEPDC1B expression phenotype.

Gene set name	NES	NOM p-val	FDR q-val
KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	1.68	0.01	0.03
KEGG_GAP_JUNCTION	1.67	0.01	0.03
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	1.66	0.03	0.03
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.65	0.01	0.04
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	1.65	0.00	0.04
KEGG_SELENOAMINO_ACID_METABOLISM	1.65	0.01	0.04
KEGG_AXON_GUIDANCE	1.65	0.01	0.03
KEGG_ETHER_LIPID_METABOLISM	1.64	0.01	0.04
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.64	0.02	0.04
KEGG_HEDGEHOG_SIGNALING_PATHWAY	1.63	0.01	0.04
KEGG_CHEMOKINE_SIGNALING_PATHWAY	1.63	0.02	0.04

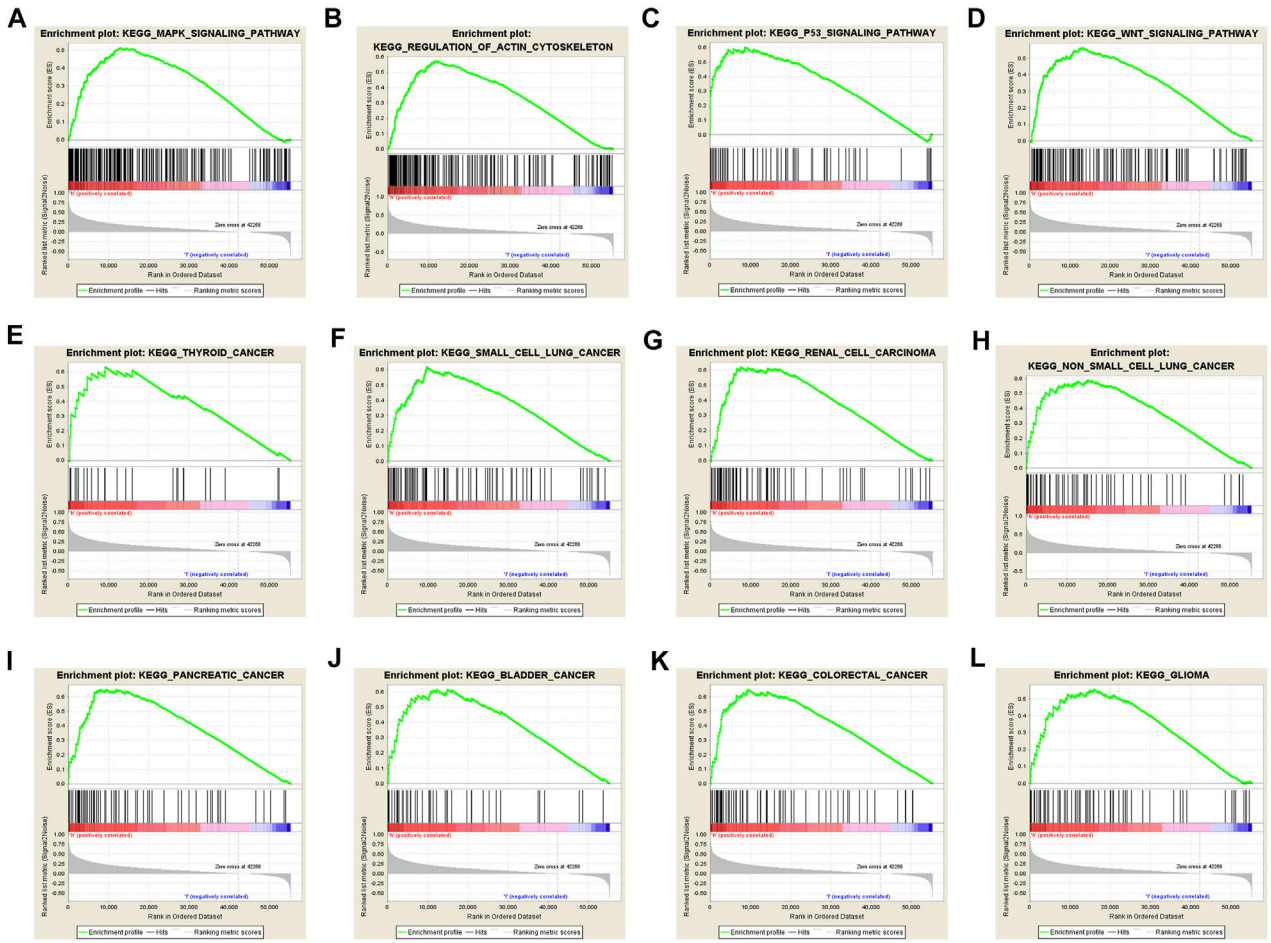


FIGURE 5 | Gene sets enriched in the high DEPDC1B expression phenotype using GSEA 3.0. Panel **(A–D)** showing some verified signaling pathways: **(A)** MAPK signaling pathway **(B)** regulation of actin cytoskeleton **(C)** p53 signaling pathway **(D)** Wnt signaling pathway; Panel **(E–L)** showing some cancer-associated pathways **(E)** Thyroid cancer **(F)** small cell lung cancer **(G)** renal cell carcinoma **(H)** non-small cell lung cancer **(I)** pancreatic cancer **(J)** bladder cancer **(K)** colorectal cancer **(L)** glioma.

Atlas database. As indicated in **Figure 6A**, the DEPDC1B protein was strongly expressed in liver cancer, compared with that in other cancers using HPA072558 antibody (Atlas Antibodies Sigma-Aldrich) (**Figure 6A**). In addition, the pattern of DEPDC1B expression in LIHC tissues is shown in **Figure 6**. (strong: **Figures 6B,C**; medium: **Figures 6D,E**).

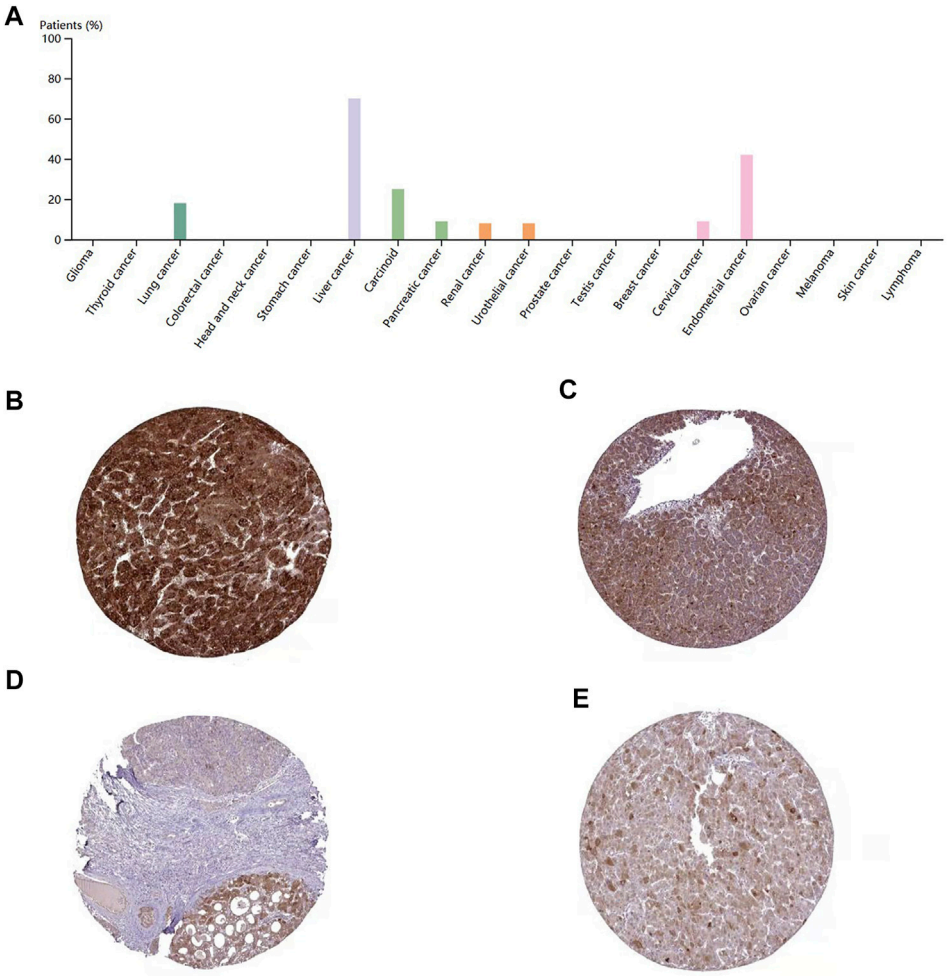


FIGURE 6 | immunohistochemistry of DEPDC1B in various types of cancer **(A)** and LIHC tissues **(B–E)** using HPA072558 antibody. **(A)** Strong cytoplasmic positivity was displayed in several hepatocellular carcinomas and single cases of carcinoma and urothelial cancer. Several endometrial cancers and a few other cancer tissues showed moderate immunoreactivity. The remaining cancer tissues were weakly stained or negative. Tumor cells staining: high expression **(B,C)** and medium expression **(D,E)**; **(B)**Patient id: 3,477, male, age 67; **(C)**: Patient id: 5,032, female, age 58; **(D)** Patient id: 3,196, male, age 65; **(E)** Patient id: 4,823, female, age 25. DEPDC1B was mainly stained in the cytoplasmic/membranous LIHC cells using HPA072558 antibody (Atlas Antibodies Sigma-Aldrich).

High Expression of DEPDC1B mRNA in LIHC Tissues

To confirm the expression of DEPDC1B mRNA in LIHC, we performed qPCR in five pairs of matched LIHC tissues and their adjacent noncancerous tissues, and as shown in **Figure 7**, the expression of DEPDC1B mRNA was upregulated in LIHC cancer tissues, compared with that in the corresponding noncancerous tissues ($p < 0.01$).

DISCUSSION

Recently, systemic therapies for LIHC patients are rapidly changing (Villanueva, 2019). Compared with less than 15% of the 5-years survival rate of patients detected at later stages, the early diagnosis of LIHC could result in a survival rate of more than 50%. However, since the value of current diagnostic biomarkers in LIHC is greatly limited (Sun and Zhang, 2020), it becomes paramount to identify novel biomarkers for the treatment of LIHC patients as such biomarkers may help to improve the 5-years survival rate of LIHC patients and further help to establish personalized treatments for each patient. Herein, the current study presented the value of DEPDC1B as a potential biomarker in LIHC.

In 2007, Boudreau et al. reported that DEPDC1B was highly expressed in the placenta and testis with only little expressions in the small intestine and the heart (Boudreau et al., 2007). In the current study, using Human Protein Atlas database, we further confirmed that DEPDC1B was highly expressed in the placenta and testis, with only little expressions in the heart and the small intestine (**Supplementary Figure S2**). Furthermore, DEPDC1B is highly expressed in the 1) gallbladder and the thymus at the RNA level and 2) gallbladder, lymph node, and tonsil at protein level (**Supplementary Figure S3**) indicated that DEPDC1B might play

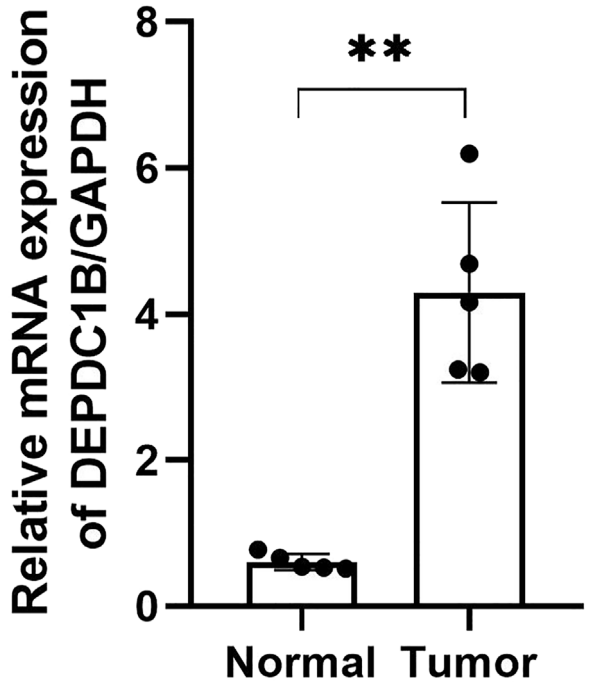


FIGURE 7 | DEPDC1B mRNA expression is frequently upregulated in LIHC tissues. T: LIHC tissues; N: noncancerous tissues. The results were shown with Mean ± SD. N = 5, $^{**}p < 0.01$.

an important role in the gallbladder, albeit the detailed mechanism remains unclear.

DEPDC1B plays an important role in the regulation of cell mitosis, transcription, and tumorigenesis (Martemyanov et al., 2003; Marchesi et al., 2014). Increasing evidence suggests that DEPDC1B is associated with various types of human cancers. However, the expression pattern and roles of DEPDC1B in LIHC remain unclear. To evaluate the role of DEPDC1B in LIHC, we used various online tools to observe the expression of DEPDC1B in LIHC. As indicated in **Figure 1**, the expression of DEPDC1B was higher in LIHC tissues than in normal tissues, and this is consistent with the results in non-small cell lung cancer, oral cancer, malignant melanoma (Xu et al., 2019), bladder cancer (Lai et al., 2020), glioblastoma, and pancreatic cancer (Liu et al., 2020). Using qPCR assay, we confirmed the higher DEPDC1B mRNA expression levels in LIHC than that in the normal tissues (**Figure 7**). These corresponding results further verified the expression of DEPDC1B mRNA is upregulated in LIHC tissues. Notably, some studies demonstrated that the overexpression of DEPDC1B could be used as a prognostic biomarker to predict the outcomes of patients with prostate and non-small cell lung cancers (Yang et al., 2014; Bai et al., 2017). Consistent with these previous reports, the present study indicated that the expression of DEPDC1B was negatively associated with patient prognosis (**Figure 2**), suggesting that the expression of DEPDC1B may be an invaluable prognostic biomarker for this disease. In addition, the univariate and multivariate Cox analyses further indicated the expression of DEPDC1B mRNA may be a useful biomarker in the prognosis of LIHC cancer (**Figure 3**). As aforementioned, this study was mainly focused on early-stage/grade LIHC and comparisons with normal tissues. As indicated in **Figure 4**, significant differences were observed in the DEPDC1B expression between the normal and tumor stage/grade cells. These exciting results further verified that DEPDC1B may be used as a valuable diagnostic biomarker of early-stage/grade LIHC. Generally, these multifaceted results strongly suggest that DEPDC1B may be an early diagnostic and prognostic biomarker in LIHC.

DEPDC1B either directly or indirectly affects the prognosis of patients with LIHC, however, the present understanding of the oncogenic function of DEPDC1B in LIHC progression remains unclarified. Increasing studies provide possible mechanistic explanations for the relationship between high DEPDC1B expression and poor prognosis in other types of tumors. Some studies indicated that the downregulation of DEPDC1B expression could suppress cell proliferation by promoting apoptosis in malignant melanoma (Xu et al., 2019), bladder cancer (Lai et al., 2020), and glioblastoma (Chen et al., 2020). In non-small cell lung cancer, DEPDC1B could enhance cell migration and invasion through the activation of Wnt/β-catenin signaling, and this biological effect could be inhibited by the depletion of LEF1 or TCF4 (Yang et al., 2014). In oral carcinoma, however, due to disruption of HPV E2, which is a viral tumor suppressor and is known to downregulate the expression of DEPDC1B, highly expressed DEPDC1B could interact with RAC1 and result in cell invasion/metastasis (Ahuja and Singh, 2016). RAC1 which is a major component of Rho GTPase signaling, and is known to regulate actin cytoskeleton (participate in the early stage of autophagosome formation), could regulate the cell cycle, cellular growth, and mediate cell proliferation *via* NF-κB activation (Ehrlich et al., 2002; Bauer et al., 2007; Bosco et al., 2010; Saci et al., 2011). Furthermore, LC3 which is a famous autophagy marker could interact with SOS1 to inhibit the GEF activity and block the activation of RAC1, which could also be inhibited by exogenous expression of DEPDC1B. Thus, the RAC1 signaling pathway is mutually regulated by autophagy and DEPDC1B. A similar study further proved that DEPDC1B may affect the prognosis of patients with prostate cancer through the regulation of autophagy (Bai et al., 2017). DEPDC1B could regulate RAC1 activity by increasing GTP loading in RAC1 instead of affecting Rho A activities in normal or cancer cells (Su et al., 2014). More recently, Liu et al. reported that long noncoding RNA lncNB1 could interact with ribosomal protein RPL35 to enhance the synthesis of E2F1 protein, leading to DEPDC1B gene transcription. The GTPase-activating protein DEPDC1B then induces ERK protein phosphorylation and the stabilization of the N-Myc protein in neuroblastoma cells. Conversely, the downregulation of lncNB1 mitigates the clonogenic capacity of neuroblastoma cells *in vitro* and leads to tumor regression *in vivo* (Liu et al., 2019). The study strongly suggests an additional pathway of DEPDC1B to induce carcinoma.

CONCLUSION

In summary, based on the Oncomine and TCGA databases, we have observed that high DEPDC1B expression is associated with poor prognosis in LIHC, suggesting that DEPDC1B could be a valuable diagnostic and prognostic marker in LIHC.

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 Institute Research Ethics Committee at The Fourth Hospital of Hebei Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XF and DH designed the studies and finished most of the work. JW and DH drafted the article. JW, LB, and FG contributed to data collection and analyses. YL performed qPCR assay. All authors read and approved the final article.

FUNDING

The present study was supported by Natural Science Foundation of Hebei Province (H20206360, H2021307016), Projects from Health and Family Planning Commission of Hebei Province (Grant Nos. 20190738, 20190007, 20200745, and 20200573).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.681809/full#supplementary-material

REFERENCES

Ahuja, P., and Singh, K. (2016). In Silico Approach for SAR Analysis of the Predicted Model of DEPDC1B: A Novel Target for Oral Cancer. *Adv. Bioinformatics* 2016, 3136024. doi:10.1155/2016/3136024

Bai, S., Chen, T., Du, T., Chen, X., Lai, Y., Ma, X., et al. (2017). High Levels of DEPDC1B Predict Shorter Biochemical Recurrence-free S-urvival of P-atients with P-rostate C-ancer. *Oncol. Lett.* 14, 6801–6808. doi:10.3892/ol.2017.7027

Ballon, D. R., Flanary, P. L., Gladue, D. P., Konopka, J. B., Dohlman, H. G., and Thorner, J. (2006). DEP-domain-mediated Regulation of GPCR Signaling Responses. *Cell* 126, 1079–1093. doi:10.1016/j.cell.2006.07.030

Bauer, N. N., Chen, Y.-W., Samant, R. S., Shevde, L. A., and Fodstad, O. (2007). Rac1 Activity Regulates Proliferation of Aggressive Metastatic Melanoma. *Exp. Cell Res.* 313, 3832–3839. doi:10.1016/j.yexcr.2007.08.017

Bosco, E. E., Nakai, Y., Hennigan, R. F., Ratner, N., and Zheng, Y. (2010). NF2-deficient Cells Depend on the Rac1-Canonical Wnt Signaling Pathway to Promote the Loss of Contact Inhibition of Proliferation. *Oncogene* 29, 2540–2549. doi:10.1038/onc.2010.20

Boudreau, H., Broustas, C., Gokhale, P., Kumar, D., Mewani, R., Rone, J., et al. (2007). Expression of BRCC3, a Novel Cell Cycle Regulated Molecule, Is Associated with Increased Phospho-ERK and Cell Proliferation. *Int. J. Mol. Med.* 19, 29–39. doi:10.3892/ijmm.19.1.29

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

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., and Hamm, H. E. (2006). DEP Domains: More Than Just Membrane Anchors. *Developmental Cell* 11, 436–438. doi:10.1016/j.devcel.2006.09.011

Chen, X., Guo, Z.-Q., Cao, D., Chen, Y., and Chen, J. (2020). Knockdown of DEPDC1B Inhibits the Development of Glioblastoma. *Cancer Cell Int* 20, 310. doi:10.1186/s12935-020-01404-7

Ehrlich, J. S., Hansen, M. D. H., and Nelson, W. J. (2002). Spatio-temporal Regulation of Rac1 Localization and Lamellipodia Dynamics during Epithelial Cell-Cell Adhesion. *Developmental Cell* 3, 259–270. doi:10.1016/s1534-5807(02)00216-2

El–Serag, H. B., and Rudolph, K. L. (2007). Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis. *Gastroenterology* 132, 2557–2576. doi:10.1053/j.gastro.2007.04.061

Kim, K. M., Sinn, D. H., Jung, S.-H., Gwak, G.-Y., Paik, Y.-H., Choi, M. S., et al. (2016). The Recommended Treatment Algorithms of the BCLC and HKLC Staging Systems: Does Following These Always Improve Survival Rates for HCC Patients. *Liver Int.* 36, 1490–1497. doi:10.1111/liv.13107

Lai, C.-H., Xu, K., Zhou, J., Wang, M., Zhang, W., Liu, X., et al. (2020). DEPDC1B Is a Tumor Promotor in Development of Bladder Cancer through Targeting SHC1. *Cell Death Dis* 11, 986. doi:10.1038/s41419-020-03190-6

Li, Z., Wang, Q., Peng, S., Yao, K., Chen, J., Tao, Y., et al. (2020). The Metastatic Promoter DEPDC1B Induces Epithelial-mesenchymal Transition and Promotes Prostate Cancer Cell Proliferation *via* Rac1-PAK1 Signaling. *Clin. Translational Med.* 10, e191. doi:10.1002/ctm2.191

Liu, P. Y., Tee, A. E., Milazzo, G., Hannan, K. M., Maag, J., Mondal, S., et al. (2019). The Long Noncoding RNA lncNB1 Promotes Tumorigenesis by Interacting with Ribosomal Protein RPL35. *Nat. Commun.* 10, 5026. doi:10.1038/s41467-019-12971-3

Liu, X., Li, T., Huang, X., Wu, W., Li, J., Wei, L., et al. (2020). DEPDC1B Promotes Migration and Invasion in Pancreatic Ductal Adenocarcinoma by Activating the Akt/GSK3β/Snail Pathway. *Oncol. Lett.* 20, 1. doi:10.3892/ol.2020.12009

Lu, F., Shah, P. A., Rao, A., Gifford-Hollingsworth, C., Chen, A., Trey, G., et al. (2020). Liver Cancer-specific Serine Protease Inhibitor Kazal Is a Potentially Novel Biomarker for the Early Detection of Hepatocellular Carcinoma. *Clin. Translational Gastroenterol.* 11, e00271. doi:10.14309/ctg.0000000000000271

Marchesi, S., Montani, F., Deflorian, G., D'Antuono, R., Cuomo, A., Bologna, S., et al. (2014). DEPDC1B Coordinates De-adhesion Events and Cell-Cycle Progression at Mitosis. *Developmental Cell* 31, 420–433. doi:10.1016/j.devcel.2014.09.009

Martemyanov, K. A., Lishko, P. V., Calero, N., Keresztes, G., Sokolov, M., Strissel, K. J., et al. (2003). The DEP Domain Determines Subcellular Targeting of the GTPase Activating Protein RGS9*In Vivo*. *J. Neurosci.* 23, 10175–10181. doi:10.1523/jneurosci.23-32-10175.2003

Mishra, N. K., Southekal, S., and Guda, C. (2019). Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma. *Front. Genet.* 10, 624. doi:10.3389/fgene.2019.00624

Nagy, Á., Lánczky, A., Menyhárt, O., and Győrffy, B. (2018). Validation of miRNA Prognostic Power in Hepatocellular Carcinoma Using Expression Data of Independent Datasets. *Sci. Rep.* 8, 9227. doi:10.1038/s41598-018-27521-y

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

Saci, A., Cantley, L. C., and Carpenter, C. L. (2011). Rac1 Regulates the Activity of mTORC1 and mTORC2 and Controls Cellular Size. *Mol. Cell* 42, 50–61. doi:10.1016/j.molcel.2011.03.017

Schulze, K., Nault, J.-C., and Villanueva, A. (2016). Genetic Profiling of Hepatocellular Carcinoma Using Next-Generation Sequencing. *J. Hepatol.* 65, 1031–1042. doi:10.1016/j.jhep.2016.05.035

Sokol, S. (2000). A Role for Wnts in Morpho-Genesis and Tissue Polarity. *Nat. Cell Biol* 2, E124–E125. doi:10.1038/35017136

Su, Y.-F., Liang, C.-Y., Huang, C.-Y., Peng, C.-Y., Chen, C. C., Lin, M.-C., et al. (2014). A Putative Novel Protein, DEPDC1B, Is Overexpressed in Oral Cancer Patients, and Enhanced anchorage-independent Growth in Oral Cancer Cells that Is Mediated by Rac1 and ERK. *J. Biomed. Sci.* 21, 67. doi:10.1186/s12929-014-0067-1

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Sun, Y., and Zhang, Z. (2020). In SilicoIdentification of Crucial Genes and Specific Pathways in Hepatocellular Cancer. *Genet. Test. Mol. Biomarkers* 24, 296–308. doi:10.1089/gtmb.2019.0242

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

Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380, 1450–1462. doi:10.1056/NEJMra1713263

Wang, X. W., Hussain, S. P., Huo, T.-I., Wu, C.-G., Forgues, M., Hofseth, L. J., et al. (2002). Molecular Pathogenesis of Human Hepatocellular Carcinoma. *Toxicology* 181-182, 43–47. doi:10.1016/s0300-483x(02)00253-6

Wharton, K. A., Jr. (2003). Runnin' with the Dvl: Proteins that Associate with Dsh/Dvl and Their Significance to Wnt Signal Transduction. *Developmental Biol.* 253, 1–17. doi:10.1006/dbio.2002.0869

Xu, Y., Sun, W., Zheng, B., Liu, X., Luo, Z., Kong, Y., et al. (2019). DEPDC1B Knockdown Inhibits the Development of Malignant Melanoma through Suppressing Cell Proliferation and Inducing Cell Apoptosis. *Exp. Cell Res.* 379, 48–54. doi:10.1016/j.yexcr.2019.03.021

Yang, Y., Liu, L., Cai, J., Wu, J., Guan, H., Zhu, X., et al. (2014). DEPDC1B Enhances Migration and Invasion of Non-small Cell Lung Cancer Cells *via* Activating Wnt/β-Catenin Signaling. *Biochem. Biophysical Res. Commun.* 450, 899–905. doi:10.1016/j.bbrc.2014.06.076

Zhang, S., Shi, W., Hu, W., Ma, D., Yan, D., Yu, K., et al. (2020). DEP Domain-Containing Protein 1B (DEPDC1B) Promotes Migration and Invasion in Pancreatic Cancer through the Rac1/PAK1-LIMK1-Cofilin1 Signaling Pathway. *Ott* Vol. 13, 1481–1496. doi:10.2147/OTT.S229055

Zheng, J., Wong, H. C., Mao, J., Nguyen, J. T., Srinivas, S., Zhang, W., et al. (2000). Structural Basis of the Recognition of the Dishevelled DEP Domain in the Wnt Signaling Pathway. *Nat. Struct. Biol.* 7, 1178–1184. doi:10.1038/82047

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

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

Copyright © 2022 Fan, Wen, Bao, Gao, Li 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 Genetics

ORIGINAL RESEARCH
published: 18 February 2022
doi: 10.3389/fgene.2022.708003

A Prognostic Model for Predicting Tumor Mutation Burden and Tumor-Infiltrating Immune Cells in Bladder Urothelial Carcinoma

Chengbang Wang[1,2,3,4,5†], *Shaohua Chen*[1,2,3,4,5†], *Songheng Li*[1] and *Hua Mi*[1*]

[1]Department of Urology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, [2]Institute of Urology and Nephrology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China, [3]Guangxi Collaborative Innovation Center for Genomic and Personalized Medicine, Nanning, China, [4]Guangxi Key Laboratory for Genomic and Personalized Medicine, Nanning, China, [5]Guangxi Key Laboratory of Colleges and Universities, Nanning, China

Tremendous progress has been made in development of immunotherapeutic approaches for treatment of bladder urothelial carcinoma (BLCA). However, efficacy and safety of these approaches remain unsatisfactory, necessitating further investigations for identification of indicators for predicting prognosis and efficacy. In this study, we downloaded transcriptomic and clinical data of BLCA patients from The Cancer Genome Atlas (TCGA) database, and identified differentially expressed genes (DEGs) between tumor and normal tissues. We incorporated these DEGs in an intersection analysis with immune-related genes (IRGs) obtained from the Immunology Database and Analysis Portal (ImmPort) database, and identified immune-related DEGs. These genes were subjected to Cox and least absolute shrinkage and selection operator (LASSO) regression analyses, then a prognostic model containing AHNAK, OAS1, NGF, PPY and SCG2 genes was constructed, for prediction of prognosis of BLCA and efficacy of immunotherapy. Finally, we explored the relationship between the prognostic model and tumor mutational burden (TMB), abundance of tumor-infiltrating immune cells (TICs) and immunotherapeutic targets, and found that patients with higher risk score (RS) had poorer prognosis and significantly lower levels of TMB. Patients in the low-RS group exhibited higher numbers of lymphoid cells, whereas those in the high-RS group exhibited higher proportions of myeloid cells. However, patients with high-RS tended to respond better to immunotherapy relative to those in the low-RS group. The constructed prognostic model provides a new tool for predicting prognosis of BLCA patients and efficacy of immunotherapy, offering a feasible option for management of the disease.

Keywords: tumor mutational burden, bladder urothelial carcinoma, tumor-infiltrating immune cells, immunotherapy, prognosis

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Kshipra Chauhan,
Independent researcher, Ghaziabad, India
Chel Hun Choi,
Samsung, South Korea

***Correspondence:**
Hua Mi
mihua2019@163.com

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

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

Received: *11 May 2021*
Accepted: *21 January 2022*
Published: *18 February 2022*

Citation:
Wang C, Chen S, Li S and Mi H (2022) A Prognostic Model for Predicting Tumor Mutation Burden and Tumor-Infiltrating Immune Cells in Bladder Urothelial Carcinoma. Front. Genet. 13:708003. doi: 10.3389/fgene.2022.708003

INTRODUCTION

Bladder urothelial carcinoma (BLCA) is a common malignancy that negatively affects human health. Approximately 81,400 new cases of bladder cancer, with 17,980 deaths, were reported in the United States in 2020 alone (Siegel et al., 2020). To date, transurethral resection of bladder tumors (TURBT) remains the standard treatment for non-muscle invasive bladder cancer

(NMIBC) (Flaig et al., 2020). However, incidence of tumor recurrence after TURBT reportedly ranges from 50 to 70%, with 30% of the patients progressing to muscle-invasive bladder cancer (MIBC) (Kamat et al., 2016). This highly malignant condition, which is characterized by a high rate of postoperative distant metastasis, negatively affects patients' life quality (Flaig et al., 2020).

The rapid advancement in immunotherapy in recent years has generated a variety of immune checkpoint inhibitors (ICIs) which have been applied for treatment of BLCA since 2016 (Özdemir et al., 2018). In fact, clinical trials have demonstrated their safety and efficacy over second-line therapy (Bellmunt et al., 2017; Powles et al., 2018). Over the past few years, multiple types of ICIs have been approved for clinical treatment of MIBC or metastatic bladder cancer, with encouraging results reported from the clinical trials. Nevertheless, many patients have not benefited from these immunotherapies, as evidenced by low objective remission rates (ORR) of only 15–25% and complete remission rates (CRR) below 10%, as well as incidence of serious treatment-related adverse events (TRAEs) in a subset of patients (Sharma et al., 2016; Balar et al., 2017a; Balar et al., 2017b; Bellmunt et al., 2017). Therefore, there is need to identify novel indicators for predicting efficacy of immunotherapy to enhance efficacy and safety (Havel et al., 2019). Tumor mutational burden (TMB) refers to the total number of substitutions and insertions or deletions per one million bases in exons of genes in a tumor tissue (Gibney et al., 2016). Previous studies have shown that TMBs are correlated with ICIs response rates and survival times of patients with melanoma (Nathanson et al., 2017), as well as breast cancer (Barroso-Sousa et al., 2020), and non-small cell lung cancers (Rizvi et al., 2015; Ettinger et al., 2019). A phase II clinical trial, comprising 310 patients with locally advanced and metastatic BLCA, found that TMB was associated with patients prognosis and could predict treatment responses of atezolizumab (Rosenberg et al., 2016). To date, however, the relationship between TMB and immune response in bladder cancer is still unclear.

In the present study, we performed intersection analysis of differentially expressed genes (DEGs), between tumor tissues of bladder cancer alongside normal controls from The Cancer Genome Atlas (TCGA) database, and immune-related genes (IRGs) from the Immunology Database and Analysis Portal (ImmPort) database, to obtain immune-related DEGs. We performed Cox regression as well as least absolute shrinkage and selection operator (LASSO) regression analyses to screen the identified genes, then constructed a prognostic model for predicting survival times and efficacy of immunotherapy of BLCA. Finally, we explored the relationship between risk score (RS) and abundance of tumor-infiltrating immune cells (TICs), as well as the correlation between RS and immunotherapy targets. Our findings are expected to provide novel insights to guide future development of effective therapies for treatment of BLCA.

MATERIALS AND METHODS

Data Acquisition and Processing

We employed the HTSeq-counts workflow to download transcriptome data for 433 cases, including 414 tumor and 19 normal samples, from TCGA database of BLCA project (https://portal.gdc.cancer.gov/; accessed on 28 March 2021). Corresponding clinical data for the patients, including age at diagnosis, gender, tumor grade, tumor stage, survival time and survival status, were also downloaded using the bcr.xml format in TCGA database through GDC portal (https://portal.gdc.cancer.gov/; accessed on 28 March 2021).

Screening for Immune-Related DEGs

DEGs between normal and tumor BLCA tissues were identified using "edgeR" and "limma" packages implemented in R software, based on a false discovery rate (FDR) < 0.05 and |log2FC| >1. A volcano plot of the DEGs was then generated using "ggplot2". IRGs were obtained from ImmPort database (http://www.immport.org/; accessed on 28 March 2021). An intersection analysis of DEGs and IRGs was performed to obtain immune-related DEGs, which were then visualized using the "venn" and "ggplot" packages in R.

Functional Enrichment Analysis

The "org.Hs.eg.db" package was used to acquire Entrez-IDs for each immune-related DEGs. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were then conducted using cutoff criteria of $p < 0.05$ and q-value < 0.05. Results were visualized by the "clusterProfiler", "enrichplot" and "ggplot2" packages in R.

Construction of a BLCA Prognostic Prediction Model

We used perl scripts to merge clinic characteristics and immune-related DEGs, and generate a matrix showing the survival times, survival status, and levels of gene for each sample. Thereafter, we employed the createDataPartition function in the "caret" package to randomly divide all samples into two groups, namely a training (containing 2/3 of the BLCA samples) and test (containing 1/3 of the BLCA samples) set for cross-validation. Univariate Cox regression analysis was performed using the "survival" package to further obtain immune-related DEGs associated with prognosis of BLCA in the training set, at a threshold $p < 0.001$. We then used the "glmnet" package to perform LASSO regression, and multivariate Cox regression analyses was used to further screen the genes after LASSO regression. The resulting genes were used to construct a BLCA prognostic prediction model. Next, we divided all BLCA samples into high- and low-RS groups, based on median of RS values, then performed Kaplan-Meier survival analysis of both groups and generate time-dependent receiver operating characteristic (ROC) curves to cross-validate the predictive power of the constructed model using the training, test and the combined (containing all of the BLCA samples) sets.

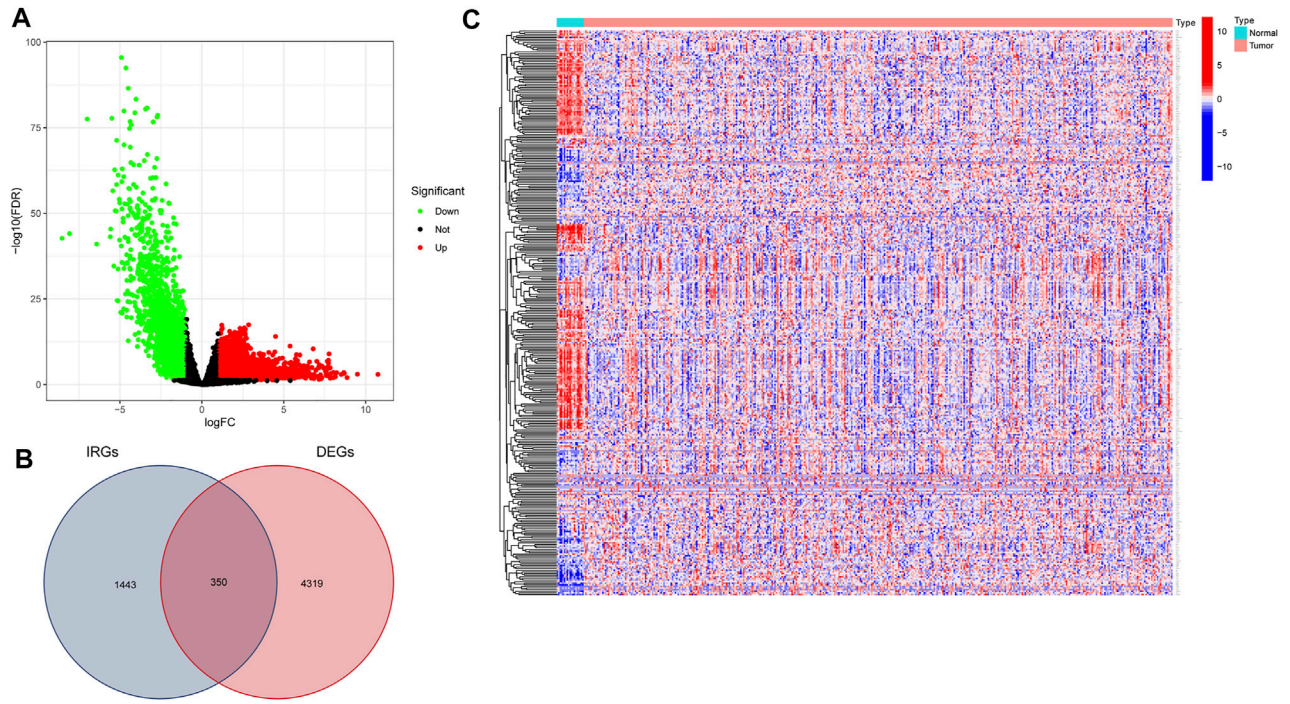


FIGURE 1 | Screening for immune-related differentially expressed genes (DEGs) in bladder urothelial carcinoma (BLCA). **(A)** Volcano plot showing DEGs between tumor and normal tissue based on The Cancer Genome Atlas (TCGA) database of BLCA samples. **(B)** Venn plot showing the intersection analysis of DEGs with immune-related genes (IRGs) from the Immunology Database and Analysis Portal (ImmPort) database. **(C)** Heat map of immune-related DEGs.

Identification of Independent Risk Factors Affecting BLCA Prognosis.

The clinical characteristics of all BLCA samples were merged with RS scores to generate a matrix. Univariate Cox regression analysis was performed on clinical characteristics and RS using the "survival" package, followed by multivariate Cox regression analysis to obtain independent risk factors associated with BLCA prognosis. Correlation between clinical characteristics and RS was conducted by the package "limma" and visualized by "ggpubr".

Analysis on TICs

We calculated relative abundance of 22 TIC subtypes in tumor samples using the CIBERSORT algorithm (Newman et al., 2015; Chen et al., 2018), then generated bar plots to present the relative proportion of TICs in each sample. The relationship between each TIC subtype and RS was visualized using the "vioplot" package.

Correlation Between TMB and RS

We downloaded "Masked Somatic Mutation" data, processed by VarScan2, from the TCGA database (https://portal.gdc.cancer.gov/; accessed on 28 March 2021), then applied the perl script to integrate RS and TMB data of BLCA samples. Tumor mutation profiles in high- and low-RS groups were visualized using the "maftools" package, while further analysis and visualization of the high- and low-TMB levels of survival analysis as well as correlation between TMB and RS were performed using "limma", "survival", "survminer", and "ggpubr" packages.

Correlation Between Immunotherapy and RS

The relationship between RS with immunotherapy targets and effects was analyzed using "limma" and "ggpubr" packages in R.

RESULTS

Identification of Immune-Related DEGs

We downloaded data for 433 BLCA samples from the TCGA database, including 414 cancer and 19 normal samples, respectively. Screening for DEGs, based on |log2FC|>1 and FDR<0.05 thresholds, revealed a total of 4,669 DEGs, of which 2,726 and 1943 genes were up-regulated and down-regulated, respectively (**Figure 1A**) Intersection analysis between the 1793 IRGs from the ImmPort database and DEGs resulted in 350 immune-related DEGs (**Figure 1B**). Then, we divided the samples into the normal and the tumor type, thereby visualizing relative expression levels of these genes in BLCA patients from TCGA database with heat maps (**Figure 1C**). GO analysis showed that the immune-related DEGs were mainly enriched in the following pathways: chemokine-related pathway (cellular

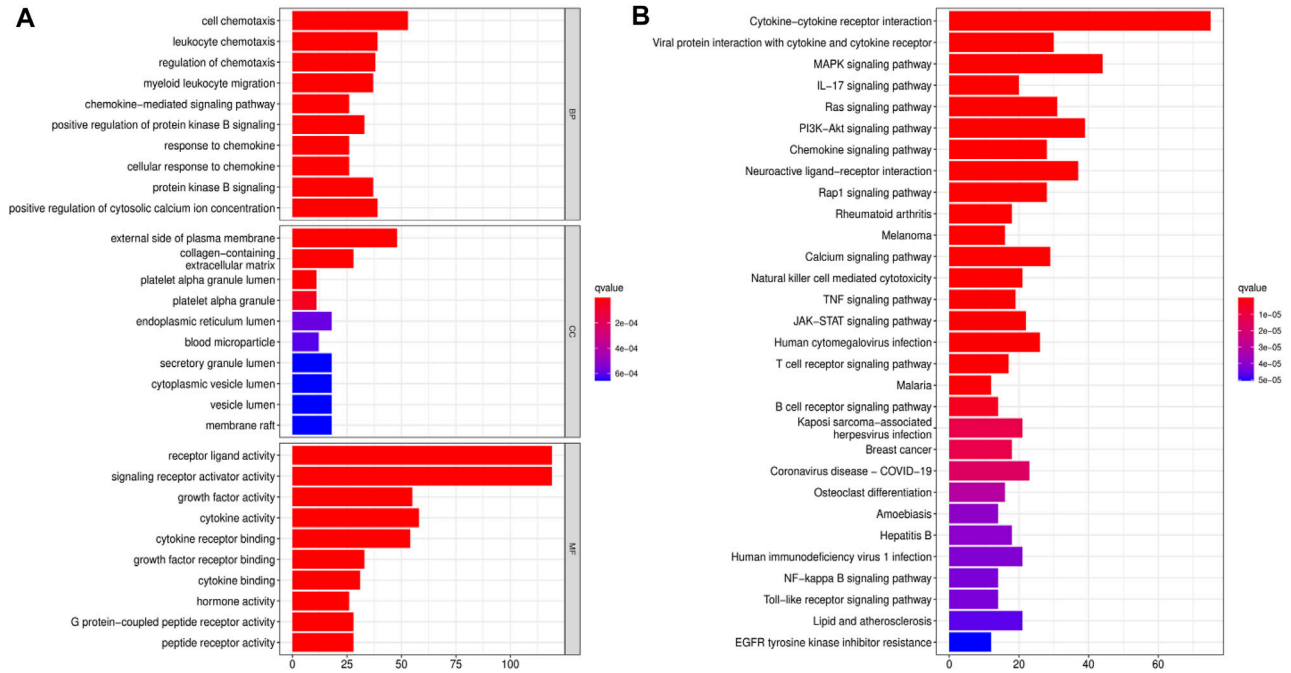


FIGURE 2 | Functional enrichment analyses for immune-related differentially expressed genes (DEGs). **(A)** Gene Ontology (GO) enrichment analysis for immune-related DEGs. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for immune-related DEGs.

TABLE 1 | Comparison of clinical characteristics between the training and test sets.

Clinical characteristics		Number	Test (%)	Training (%)	*p*-value
Age	≤65	159	46 (34.85)	113 (42.97)	0.149
	>65	236	86 (65.15)	150 (57.03)	
Gender	FEMALE	104	36 (27.27)	68 (25.86)	0.8567
	MALE	291	96 (72.73)	195 (74.14)	
Grade	High Grade	374	128 (96.97)	246 (93.54)	0.4254
	Low Grade	18	4 (3.03)	14 (5.32)	
	unknow	3	0 (0)	3 (1.14)	
Stage	Stage I-II	125	43 (32.58)	82 (31.18)	0.9059
	Stage III-IV	268	89 (67.42)	179 (68.06)	
	unknow	2	0 (0)	2 (0.76)	
T	T1-2	116	38 (28.79)	78 (29.66)	0.9786
	T3-4	247	79 (59.85)	168 (63.88)	
	unknow	32	15 (11.36)	17 (6.46)	
M	M0	189	58 (43.94)	131 (49.81)	0.7877
	M1	10	4 (3.03)	6 (2.28)	
	unknow	196	70 (53.03)	126 (47.91)	
N	N0	228	78 (59.09)	150 (57.03)	0.5096
	N1-3	126	38 (28.79)	88 (33.46)	
	unknow	41	16 (12.12)	25 (9.51)	

response to chemokine and chemokine-mediated signaling pathway), biofilm lumen metabolic pathway (external side of plasma membrane and cytoplasmic vesicle lumen) and cytokine ligand receptor activity pathway (growth factor activity and signaling receptor activator activity) (**Figure 2A**). On the other hand, KEGG pathway enrichment analysis demonstrated that these genes were mainly enriched in cytokine-cytokine receptor interaction and NK cell-mediated cytotoxicity (**Figure 2B**). Thus, the 350 shared genes obtained by the intersection analysis were uniformly enriched in immune-related activities.

TABLE 2 | Cox regression analysis for screening of immune-related differentially expressed genes (DEGs) affecting the prognosis of bladder urothelial carcinoma (BLCA).

Gene id	Univariate cox analysis		Multivariate cox analysis	
	HR (95%CI)	*p*-value	HR (95%CI)	*p*-value
CXCL12	1.20 (1.08–1.33)	6.90E-04		
LRP1	1.46 (1.23–1.73)	1.50E-05		
PDGFRA	1.27 (1.11–1.45)	4.44E-04		
AHNAK	1.52 (1.24–1.86)	5.89E-05	1.46 (1.19–1.78)	2.23E-04
OAS1	0.78 (0.68–0.9)	7.96E-04	0.86 (0.74–1.00)	5.54E-02
EDNRA	1.29 (1.11–1.5)	7.05E-04		
NGF	1.54 (1.23–1.92)	1.62E-04	1.23 (0.96–1.57)	1.02E-01
PPY	1.75 (1.32–2.31)	8.52E-05	1.82 (1.37–2.41)	3.24E-05
SCG2	1.26 (1.12–1.43)	1.25E-04	1.21 (1.05–1.39)	8.75E-03
TGFB3	1.23 (1.09–1.39)	7.71E-04		
FGFR1	1.22 (1.09–1.36)	4.98E-04		
NRP2	1.31 (1.15–1.5)	5.34E-05		
PTGER3	1.23 (1.09–1.39)	6.41E-04		

Construction of a BLCA Prognostic Model

Next, we randomly divided all samples into two groups (ratio of 2: 1) for cross-validation, naming training and test sets. As shown in **Table 1**, no significant differences were observed in the varying clinical characteristics between the training and test sets after

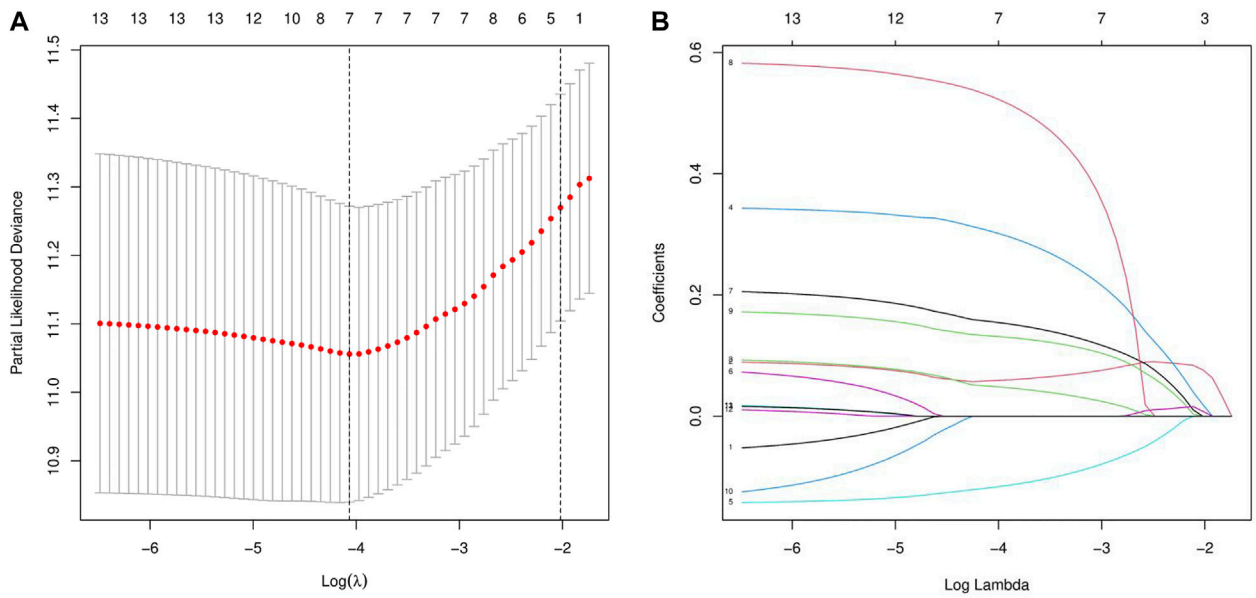


FIGURE 3 | LASSO regression analysis for the screening of immune-related differentially expressed genes (DEGs). **(A, B)** least absolute shrinkage and selection operator (LASSO) regression complexity was controlled by lambda *via* the R package "glmnet".

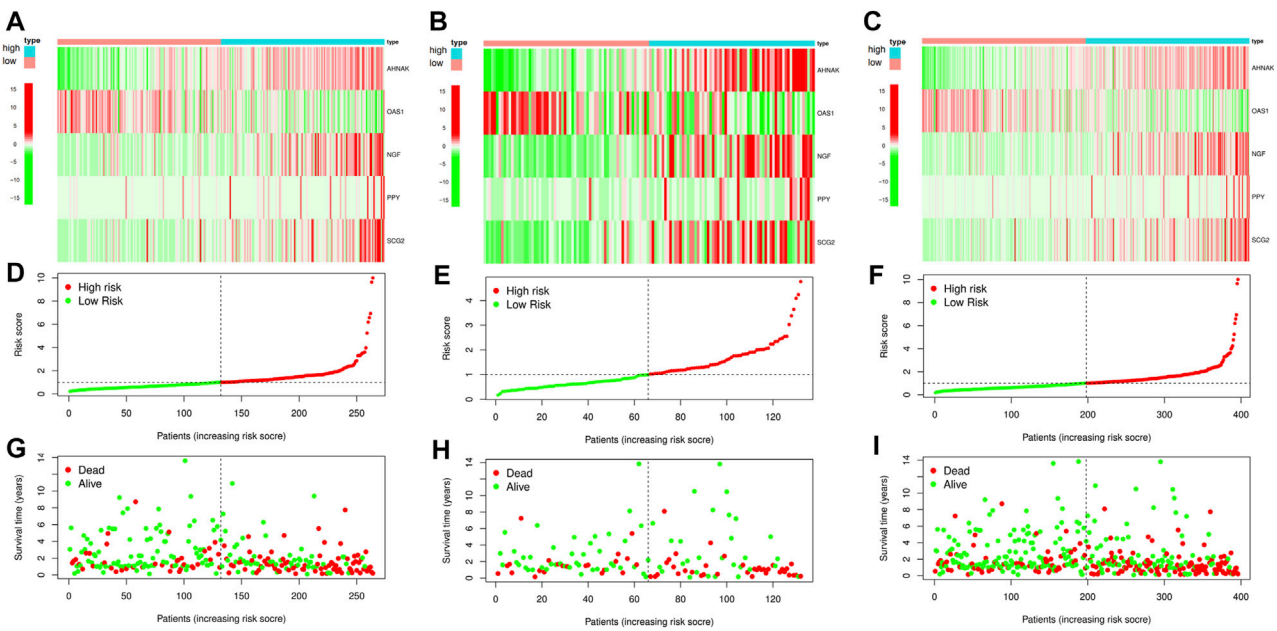


FIGURE 4 | Construction and cross-validation of bladder urothelial carcinoma (BLCA) prognostic model. **(A–C)** The heat map showing the expression levels of the key immune-related differentially expressed genes (DEGs) in the prognostic model for each sample in the training set, test set and combined set, respectively. **(D–F)** Distribution of risk scores (RSs) for each sample in three different datasets. **(G–I)** Distribution of survival states for each sample in three different datasets.

random grouping, indicating that they can be used as independent datasets. Univariate Cox regression analysis of 350 immune-related DEGs in the training set revealed 13 genes that were significantly associated with patients' prognosis ($p < 0.001$) (**Table 2**). Further screening of the genes using LASSO regression (**Figures 3A,B**) and multivariate Cox regression (**Table 2**) analyses revealed 5 immune-related DEGs, which we subsequently used to construct a BLCA prognostic model. We divided the samples into high- and low-RS groups, with values below the median

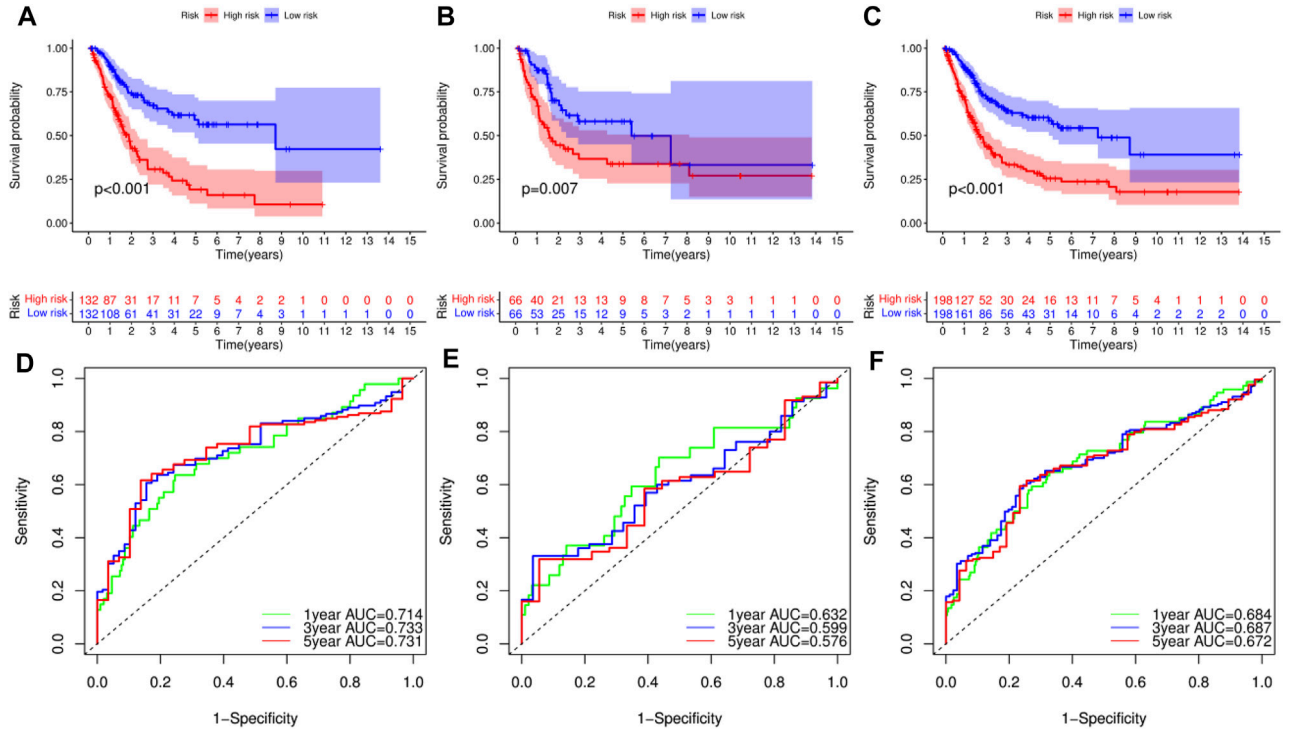


FIGURE 5 | Survival analysis of bladder urothelial carcinoma (BLCA) in different datasets and receiver operating characteristic (ROC) curve analysis for prognostic model. **(A–C)** Survival analysis of BLCA patients with high- and low-risk scores (RSs) in the training set, test set and combined set, respectively. **(D–F)** ROC curve analysis for the prognostic model at 1-, 3- and 5-years in the training set, test set and combined set, respectively.

classified into the low-RS group and vice versa. Heat maps for BLCA samples in the training, test and combined sets revealed that OAS1 was downregulated while the other 4 genes were upregulated in the high-RS group (**Figures 4A–C**), suggesting that individual genes in the prognostic model may play diametrically opposed roles in BLCA. Profiles of RS distribution (**Figures 4D–F**) and RS-related patient survival status (**Figures 4G–I**) revealed that an increase in RS significantly increased mortality risk of patients.

Cross-Validation of the Prognostic Model

Survival analysis of the training set, used to validate the constructed model, revealed that patients in the low-RS group had longer overall survival (OS) time than those in the high-RS group ($p < 0.01$, **Figure 5A**). Survival analysis for the test and combined sets revealed poor prognosis in the high-RS than the low-RS group ($p < 0.01$) (**Figures 5B,C**). Meanwhile, ROC curves revealed that the prognostic model could efficiently predict clinical outcomes, as evidenced by area under the curve (AUC) values of 0.714, 0.733 and 0.731 for 1-, 3- and 5-years survival, respectively (**Figure 5D**). Similar results were obtained in the ROC curves analyses for the test and combined sets (**Figures 5E,F**). Overall, these results indicated that the constructed prognostic model was highly stable and reliable in predicting prognosis of BLCA patients, and guarantees good sensitivity and specificity.

RS was an Independent Risk Factor for Patient Prognosis

To identify the independent risk factors for BCLA development in patients, we used a univariate Cox regression to analyze the relationship between patients' clinical characteristics and RS in combination with survival time. Results indicated that age, clinical stage and RS were risk factors associated with prognosis of BLCA patients in combined dataset (**Table 3**). On the other hand, multivariate Cox regression analysis showed that age, clinical stage and RS were independent risk factors associated with prognosis of BLCA (**Table 3**). Notably, the Cox regression analysis results of the training set and the test set invariably present the similar results, indicating that RS were independent risk factors in these two datasets (**Supplementary Tables S1, S2**). Thereafter, we correlated RS with clinical characteristics, including age, gender, tumor grade and clinic stage, and found that RS was closely associated with various clinical characteristics ($p < 0.05$) (**Figures 6A–G**). Meanwhile, tumor grade and tumor stage were positively correlated with RS. These results indicated that RS was related to progression and metastasis of BLCA, implying its potential in predicting prognosis.

TMB Profile

Next, we analyzed somatic mutation profiles of 405 BLCA patients downloaded from the TCGA database, after dividing

TABLE 3 | Cox regression analysis of clinical characteristics and risk score (RS) affecting patients' prognosis in combined set.

Variable	Univariate cox analysis		Multivariate cox analysis	
	HR (95%CI)	*p*-value	HR (95%CI)	*p*-value
Age	1.033 (1.017–1.049)	4.6938E-05	1.029 (1.013–1.045)	0.000371
Gender	0.873 (0.628–1.213)	0.41752846		
Grade	2.801 (0.692–11.330)	0.14836518		
Stage	1.744 (1.435–2.119)	2.1573E-08	1.699 (1.396–2.068)	1.25E-07
RiskScore	1.059 (1.039–1.080)	4.808E-09	1.057 (1.036–1.078)	4.08E-08

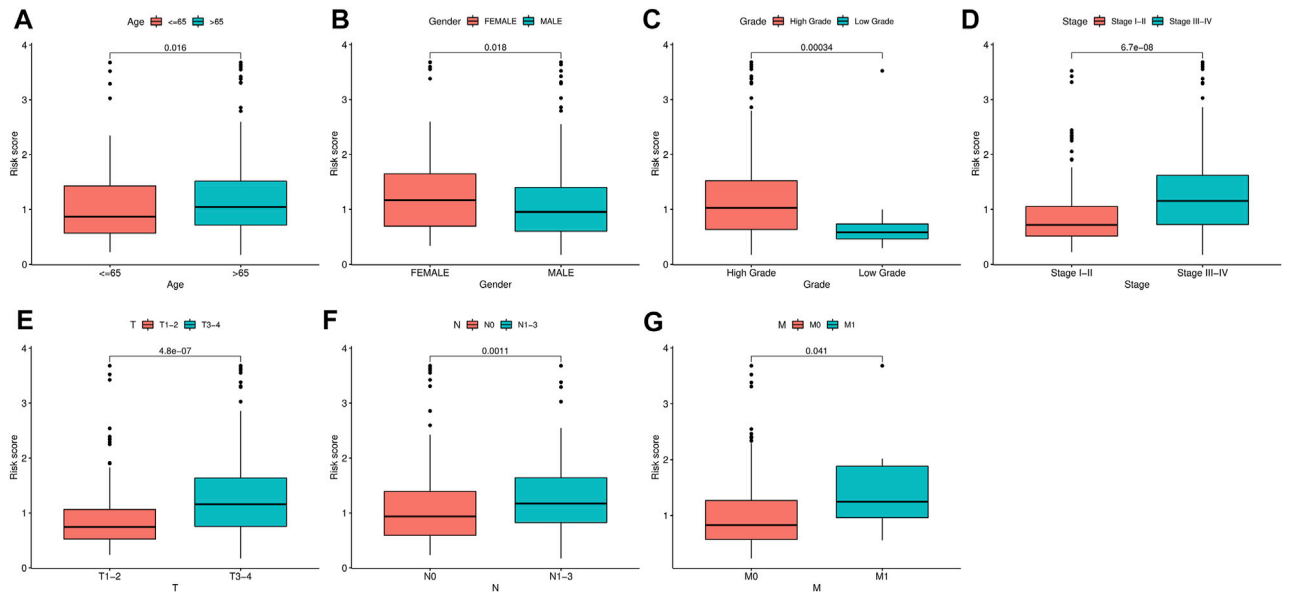


FIGURE 6 | Correlation analysis between clinical characteristics and risk scores (RSs). **(A–G)** The relationship between RS and gender, tumor stage, tumor grade, T classification, and M classification of bladder urothelial carcinoma (BLCA) patients.

them into high- and low-TMB groups based on median TMB value, and combining them with patients' survival time (**Figure 7A**). Results showed that patients in high-TMB group had significantly longer survival times than those in the low-TMB group, suggesting that TMB was associated with prognosis of BLCA. Similarly, patients in the low-RS group had significantly higher TMB levels than those in the high-RS group ($p < 0.01$) (**Figure 7B**), indicating that the prognostic model of BLCA was correlated with TMB. Waterfall plots were used to display detailed mutation information in each sample, with various color annotations used to distinguish between different mutation types (**Figures 7C,D**).

TICs and Immunotherapy Analysis

We adopted the CIBERSORT algorithm to further analyze relative abundance of various TIC subtypes in tumor samples and revealed that the relationship between prognostic models and TICs. A total of 22 TIC types in tumor samples were compared using the Wilcoxon rank sum test, revealing differences in proportions between the high- and low-RS groups, and the results visualized using violin plot (**Figure 8A**). Summarily, patients in the low-RS group exhibited significantly higher relative abundance of plasma cells, CD8 T cells, follicular helper T cells, regulatory T cells and activated dendritic cells, while their resting memory CD4 T, macrophages M0 and neutrophils were lower in the high-RS group. We then divided the samples into high- and low-groups, based on the median levels of each TIC, and combined them with survival times. Results indicated that lower proportions of resting mast cells and neutrophils in tumor samples were correlated with longer survival times of patients ($p < 0.05$) (**Figures 8B,C**). Expression analysis of immunotherapy targets in the high- and low-RS groups revealed upregulation of programmed cell death ligand 1 (PD-L1), programmed death 1 (PD-1), and cytotoxic T lymphocyte-associated protein 4 (CTLA4) in the high-RS group relative to the low-RS group (**Figures 9A–C**), indicating that patients in high-RS group were more likely to benefit from immunotherapy.

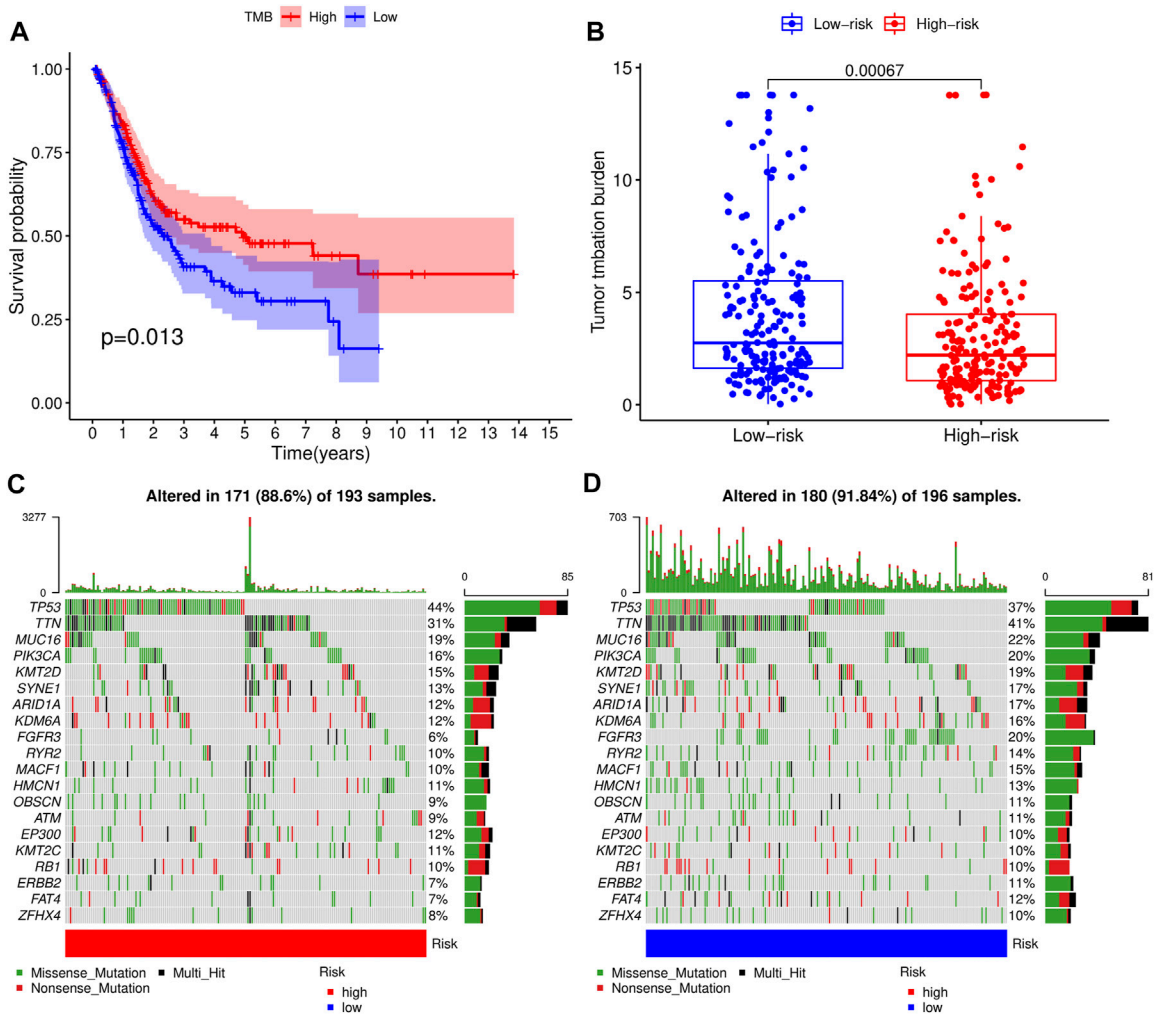


FIGURE 7 | Correlation analysis of tumor mutational load (TMB) and risk scores (RSs). **(A)** Survival analysis for patients with high- and low-TMB. **(B)** Difference of TMB level between high- and low-RS groups. **(C)** Waterfall plot of the top 20 TMB-related genes in the high-RS group. **(D)** Waterfall plot of the top 20 TMB-related genes in the low-RS group.

DISCUSSION

BLCA, the ninth most common cancer in the world that poses a major challenge to global public health (Cumberbatch et al., 2018), is characterized by high aggressiveness and a high recurrence rate. Traditional pathological cystoscopy and urine cytology have been used as for clinical diagnosis of BLCA (Tran et al., 2021). To date, cisplatin-based chemotherapy remains the first-line treatment therapy for locally advanced or metastatic BLCA (von der Maase et al., 2005). However, the strategy is not effective since it requires patients to have renal function reserve and good physical condition, and has also been associated with various adverse effects (Chen et al., 2021b). In recent years, immunotherapy has rapidly developed, giving rise to a variety of ICIs that have consequently been used for treatment of advanced BLCA (Chism, 2017; Siefker-Radtke and Curti, 2018). Although previous clinical studies have demonstrated the strategy's safety and efficacy, nothing is known regarding indicators for assessment of ICI efficacy. Despite researchers identifying various markers associated with immunotherapy response, such as TMB, TICs, and immune gene signatures, limitations and discrepancies among studies have constrained their application (Gibney et al., 2016; Chan et al., 2019; Samstein et al., 2019).

In the present study, we screened BLCA samples and identified immune-related DEGs, and used them to construct a novel prognostic model. We systematically investigated the relationship between our prognostic model with TMB, TICs and immunotherapeutic targets, and validated its reliability in a test set. The model integrated five immune-related DEGs, namely AHNAK, OAS1, NGF, PPY and SCG2, of which AHNAK, OAS1 and NGF were cell cycle-related. Specifically, AHNAK enhances transcriptional activity of receptor-regulated Smads (R-Smads), causing cell cycle arrest, and participates in cell growth regulation by potentiating transforming growth factor β (TGFβ) signaling (Lee et al., 2014). Moreover, it functions as a

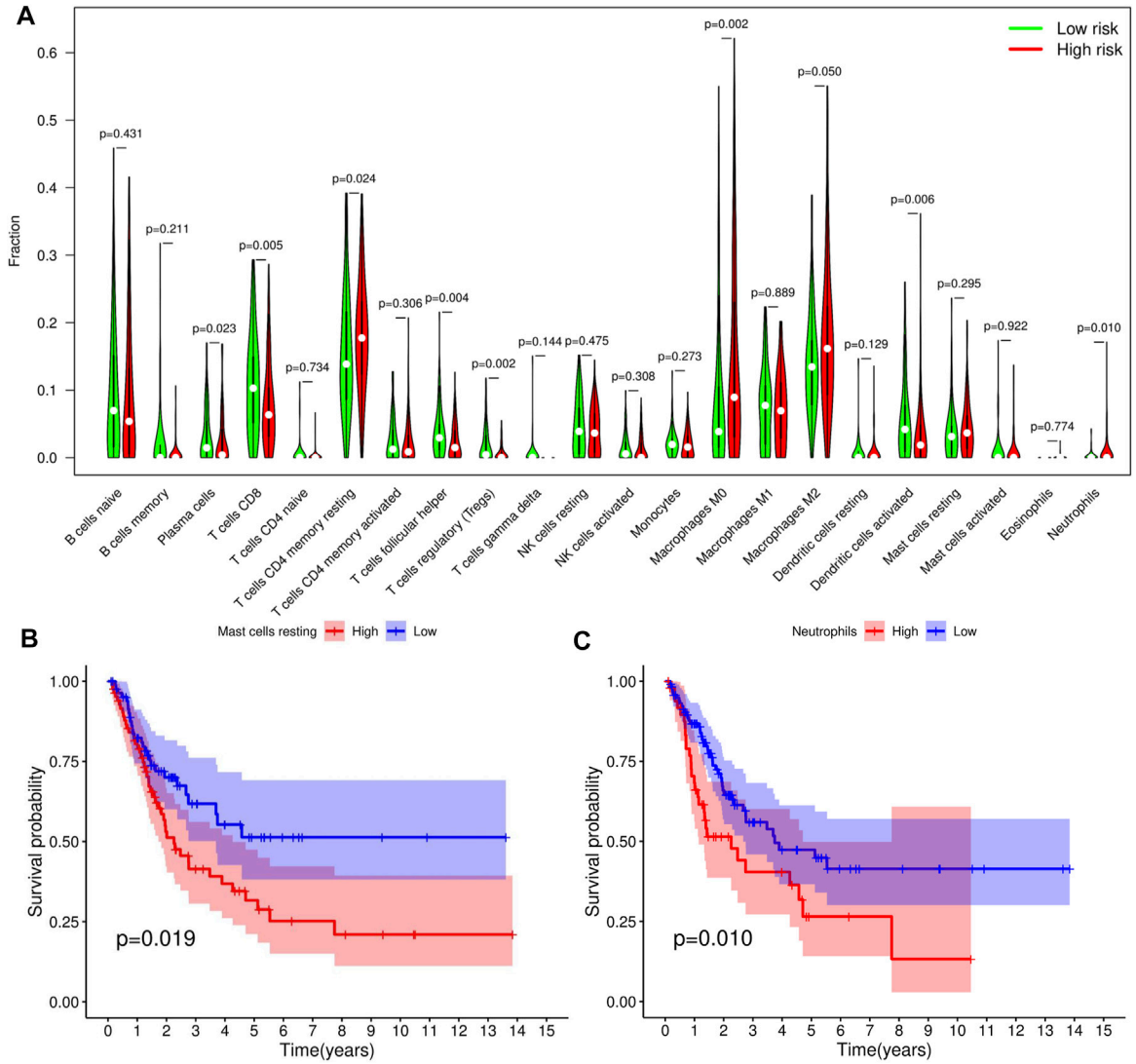


FIGURE 8 | Tumor-infiltrating immune cell (TICs) profile of bladder urothelial carcinoma (BLCA). **(A)** Violin plot showing the difference in the proportion of 22 TIC subtypes in BLCA for the high- and low-risk score (RS) groups. **(B,C)** Two TICs influencing patients' survival outcome.

tumor suppressor and has also been shown to play an adjuvant role during diagnosis of BLCA (Lee et al., 2018). The 2′-5′ oligoadenylate synthetases (OAS) are interferon-inducible enzymes that recognize viral double-stranded RNA (Kristiansen et al., 2011). Previous studies have shown that OAS1 could suppress accumulation of excess Poly (ADP-ribose) (PAR) in response to DNA damage, thereby inhibiting programmed cell death due to energy depletion and/or activation of PAR (Kondratova et al., 2020). Moreover, Qu and colleagues once proposed a IRG-based prognostic index containing AHNAK and OAS1 that could assess immune status and prognosis with BLCA (Qu et al., 2021). Besides, NGF is a member of the neurotrophic factor family that protects peripheral nerve cells (Rocco et al., 2018). Previous studies have shown that NGF ameliorates the inhibitory effect of the neurotrophin (NTR) receptor on cell-cycle protein expression in cancer cells of BLCA (Khwaja and Djakiew, 2003). However, the roles played by PPY and SCG2 in BLCA remain unclear, necessitating further studies. In the present study, we analyzed the relationship between clinical characteristics of BLCA and RS, and found that patients in low-RS group had better survival outcomes than those in the high-RS group. Notably, older and male patients exhibited significantly higher levels of RS than younger and female ones, in keeping with the findings of Shariat et al. who demonstrated that BLCA was a highly prevalent in the middle-aged and elderly population, with a median age of diagnosis around 70 (Shariat et al., 2010).

Since TICs have been shown to play a crucial role in cancer development and metastasis (Chen et al., 2021a), we further investigated the relationship between TICs and RS. Our results indicated that patients in the low-RS group predominantly exhibited higher numbers of lymphoid cells, whereas those in the

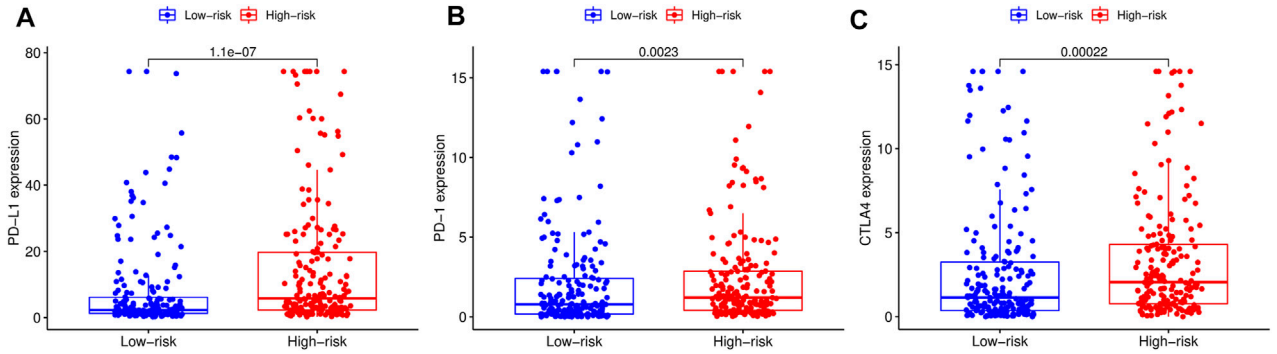


FIGURE 9 | Comparison of the expression levels of immunotherapeutic targets of high- and low-risk score (RS) groups. **(A)** Comparison of programmed cell death ligand 1 (PD-L1) expression levels between high- and low-RS groups. **(B)** Comparison of programmed death 1 (PD-1) expression levels between high- and low-RS groups. **(C)** Comparison of cytotoxic T lymphocyte-associated protein 4 (CTLA4) expression levels between high- and low-RS groups.

high-RS group had higher proportions of myeloid cells. Previous studies have shown that high neutrophil/lymphocyte ratio (NLR) are associated with poor prognosis of several malignancies (Templeton et al., 2014). Survival analysis revealed a significant correlation between lower levels of neutrophils and resting mast cell infiltration with better survival outcomes. Notably, neutrophils play a double-edged role in bladder cancer, stimulating anti-tumor immune responses by releasing IFN-γ (Xiang et al., 2020) or inducing inflammation and production of growth factors and neutrophil elastase to favor tumor growth (Galdiero et al., 2018). Application of ICIs in recent years has generated encouraging efficacy in management of various solid tumors, including bladder cancer (Nadal and Bellmunt, 2019). However, this therapy has been found to have clinical benefits in less than half of patients with advanced BLCA (Crispen and Kusmartsev, 2020). In the present study, we explored the relationship between RS and immunotherapeutic targets, and found significant upregulation of PD-L1, PD1 and CTLA4 in patients in the high-RS relative to those in the low-RS group. These results suggested that better outcomes may be achieved with ICIs in patients with high-RS.

CONCLUSION

In summary, we constructed and validated a reliable model for predicting prognosis of BLCA patients, based on transcriptomic and clinical data obtained from public databases. In addition, we elaborated the correlation between the prognostic model and TMB and TICs using correlation analysis. The constructed model provides a new assessment tool for predicting prognosis of BLCA patients and efficacy of immunotherapy, therefor offering an alternative for management of the disease.

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

SC and CW wrote the manuscript. CW and SL analyzed data. HM was responsible for the acquisition and interpretation of data. SC revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China, Grant/Award Number. 81860142.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.708003/full#supplementary-material

REFERENCES

Balar, A. V., Castellano, D., O'Donnell, P. H., Grivas, P., Vuky, J., Powles, T., et al. (2017a). First-line Pembrolizumab in Cisplatin-Ineligible Patients with Locally Advanced and Unresectable or Metastatic Urothelial Cancer (KEYNOTE-052): a Multicentre, Single-Arm, Phase 2 Study. *Lancet Oncol.* 18 (11), 1483–1492. doi:10.1016/s1470-2045(17)30616-2

Balar, A. V., Galsky, M. D., Rosenberg, J. E., Powles, T., Petrylak, D. P., Bellmunt, J., et al. (2017b). Atezolizumab as First-Line Treatment in Cisplatin-Ineligible Patients with Locally Advanced and Metastatic Urothelial Carcinoma: a Single-Arm, Multicentre, Phase 2 Trial. *The Lancet* 389 (10064), 67–76. doi:10.1016/s0140-6736(16)32455-2

Barroso-Sousa, R., Jain, E., Cohen, O., Kim, D., Buendia-Buendia, J., Winer, E., et al. (2020). Prevalence and Mutational Determinants of High Tumor Mutation burden in Breast Cancer. *Ann. Oncol.* 31 (3), 387–394. doi:10.1016/j.annonc.2019.11.010

Bellmunt, J., de Wit, R., Vaughn, D. J., Fradet, Y., Lee, J.-L., Fong, L., et al. (2017). Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N. Engl. J. Med.* 376 (11), 1015–1026. doi:10.1056/NEJMoa1613683

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

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, S., Sun, Y., Zhu, X., and Mo, Z. (2021a). Prediction of Survival Outcome in Lower-Grade Glioma Using a Prognostic Signature with 33 Immune-Related Gene Pairs. *Ijgm* Vol. 14, 8149–8160. doi:10.2147/ijgm.S338135

Chen, S., Wang, C., Su, X., Dai, X., Li, S., and Mo, Z. (2021b). KCNN4 Is a Potential Prognostic Marker and Critical Factor Affecting the Immune Status of the Tumor Microenvironment in Kidney Renal clear Cell Carcinoma. *Transl Androl. Urol.* 10 (6), 2454–2470. doi:10.21037/tau-21-332

Chism, D. D. (2017). Urothelial Carcinoma of the Bladder and the Rise of Immunotherapy. *J. Natl. Compr. Canc Netw.* 15 (10), 1277–1284. doi:10.6004/jnccn.2017.7036

Crispen, P. L., and Kusmartsev, S. (2020). Mechanisms of Immune Evasion in Bladder Cancer. *Cancer Immunol. Immunother.* 69 (1), 3–14. doi:10.1007/s00262-019-02443-4

Cumberbatch, M. G. K., Jubber, I., Black, P. C., Esperto, F., Figueroa, J. D., Kamat, A. M., et al. (2018). Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018. *Eur. Urol.* 74 (6), 784–795. doi:10.1016/j.eururo.2018.09.001

Ettinger, D. S., Wood, D. E., Aggarwal, C., Aisner, D. L., Akerley, W., Bauman, J. R., et al. (2019). NCCN Guidelines Insights: Non-small Cell Lung Cancer, Version 1.2020. *J. Natl. Compr. Canc Netw.* 17 (12), 1464–1472. doi:10.6004/jnccn.2019.0059

Flaig, T. W., Spiess, P. E., Agarwal, N., Bangs, R., Boorjian, S. A., Buyyounouski, M. K., et al. (2020). Bladder Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 18 (3), 329–354. doi:10.6004/jnccn.2020.0011

Galdiero, M. R., Marone, G., and Mantovani, A. (2018). Cancer Inflammation and Cytokines. *Cold Spring Harb Perspect. Biol.* 10 (8), a028662. doi:10.1101/cshperspect.a028662

Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive Biomarkers for Checkpoint Inhibitor-Based Immunotherapy. *Lancet Oncol.* 17 (12), e542–e551. doi:10.1016/s1470-2045(16)30406-5

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The Evolving Landscape of Biomarkers for Checkpoint Inhibitor Immunotherapy. *Nat. Rev. Cancer* 19 (3), 133–150. doi:10.1038/s41568-019-0116-x

Kamat, A. M., Hahn, N. M., Efstathiou, J. A., Lerner, S. P., Malmström, P.-U., Choi, W., et al. (2016). Bladder Cancer. *The Lancet* 388 (10061), 2796–2810. doi:10.1016/s0140-6736(16)30512-8

Khwaja, F., and Djakiew, D. (2003). Inhibition of Cell-Cycle Effectors of Proliferation in Bladder Tumor Epithelial Cells by the p75NTR Tumor Suppressor. *Mol. Carcinog.* 36 (3), 153–160. doi:10.1002/mc.10106

Kondratova, A. A., Cheon, H., Dong, B., Holvey-Bates, E. G., Hasipek, M., Taran, I., et al. (2020). Suppressing PAR Ylation by 2′,5′-oligoadenylate Synthetase 1 Inhibits DNA Damage-induced Cell Death. *Embo j* 39 (11), e101573. doi:10.15252/embj.2019101573

Kristiansen, H., Gad, H. H., Eskildsen-Larsen, S., Despres, P., and Hartmann, R. (2011). The Oligoadenylate Synthetase Family: an Ancient Protein Family with Multiple Antiviral Activities. *J. Interferon Cytokine Res.* 31 (1), 41–47. doi:10.1089/jir.2010.0107

Lee, H., Kim, K., Woo, J., Park, J., Kim, H., Lee, K. E., et al. (2018). Quantitative Proteomic Analysis Identifies AHNAK (Neuroblast Differentiation-Associated Protein AHNAK) as a Novel Candidate Biomarker for Bladder Urothelial Carcinoma Diagnosis by Liquid-Based Cytology. *Mol. Cell Proteomics* 17 (9), 1788–1802. doi:10.1074/mcp.RA118.000562

Lee, I. H., Sohn, M., Lim, H. J., Yoon, S., Oh, H., Shin, S., et al. (2014). Ahnak Functions as a Tumor Suppressor via Modulation of TGFβ/Smad Signaling Pathway. *Oncogene* 33 (38), 4675–4684. doi:10.1038/onc.2014.69

Nadal, R., and Bellmunt, J. (2019). Management of Metastatic Bladder Cancer. *Cancer Treat. Rev.* 76, 10–21. doi:10.1016/j.ctrv.2019.04.002

Nathanson, T., Ahuja, A., Rubinsteyn, A., Aksoy, B. A., Hellmann, M. D., Miao, D., et al. (2017). Somatic Mutations and Neoepitope Homology in Melanomas Treated with CTLA-4 Blockade. *Cancer Immunol. Res.* 5 (1), 84–91. doi:10.1158/2326-6066.Cir-16-0019

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

Özdemir, B. C., Siefker-Radtke, A. O., Campbell, M. T., and Subudhi, S. K. (2018). Current and Future Applications of Novel Immunotherapies in Urological Oncology: A Critical Review of the Literature. *Eur. Urol. Focus* 4 (3), 442–454. doi:10.1016/j.euf.2017.10.001

Powles, T., Durán, I., van der Heijden, M. S., Loriot, Y., Vogelzang, N. J., De Giorgi, U., et al. (2018). Atezolizumab versus Chemotherapy in Patients with Platinum-Treated Locally Advanced or Metastatic Urothelial Carcinoma (IMvigor211): a Multicentre, Open-Label, Phase 3 Randomised Controlled Trial. *The Lancet* 391 (10122), 748–757. doi:10.1016/s0140-6736(17)33297-x

Qu, G., Liu, Z., Yang, G., Xu, Y., Xiang, M., and Tang, C. (2021). Development of a Prognostic index and Screening of Prognosis Related Genes Based on an Immunogenomic Landscape Analysis of Bladder Cancer. *Aging* 13 (8), 12099–12112. doi:10.18632/aging.202917

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

Rocco, M. L., Soligo, M., Manni, L., and Aloe, L. (2018). Nerve Growth Factor: Early Studies and Recent Clinical Trials. *Cn* 16 (10), 1455–1465. doi:10.2174/1570159x16666180412092859

Rosenberg, J. E., Hoffman-Censits, J., Powles, T., van der Heijden, M. S., Balar, A. V., Necchi, A., et al. (2016). Atezolizumab in Patients with Locally Advanced and Metastatic Urothelial Carcinoma Who Have Progressed Following Treatment with Platinum-Based Chemotherapy: a Single-Arm, Multicentre, Phase 2 Trial. *The Lancet* 387 (10031), 1909–1920. doi:10.1016/s0140-6736(16)00561-4

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. *Nat. Genet.* 51 (2), 202–206. doi:10.1038/s41588-018-0312-8

Shariat, S. F., Sfakianos, J. P., Droller, M. J., Karakiewicz, P. I., Meryn, S., and Bochner, B. H. (2010). The Effect of Age and Gender on Bladder Cancer: a Critical Review of the Literature. *BJU Int.* 105 (3), 300–308. doi:10.1111/j.1464-410X.2009.09076.x

Sharma, P., Callahan, M. K., Bono, P., Kim, J., Spiliopoulou, P., Calvo, E., et al. (2016). Nivolumab Monotherapy in Recurrent Metastatic Urothelial Carcinoma (CheckMate 032): a Multicentre, Open-Label, Two-Stage, Multi-Arm, Phase 1/2 Trial. *Lancet Oncol.* 17 (11), 1590–1598. doi:10.1016/s1470-2045(16)30496-x

Siefker-Radtke, A., and Curti, B. (2018). Immunotherapy in Metastatic Urothelial Carcinoma: Focus on Immune Checkpoint Inhibition. *Nat. Rev. Urol.* 15 (2), 112–124. doi:10.1038/nrurol.2017.190

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

Templeton, A. J., McNamara, M. G., Šeruga, B., Vera-Badillo, F. E., Aneja, P., Ocaña, A., et al. (2014). Prognostic Role of Neutrophil-To-Lymphocyte Ratio in Solid Tumors: a Systematic Review and Meta-Analysis. *J. Natl. Cancer Inst.* 106 (6), dju124. doi:10.1093/jnci/dju124

Tran, L., Xiao, J.-F., Agarwal, N., Duex, J. E., and Theodorescu, D. (2021). Advances in Bladder Cancer Biology and Therapy. *Nat. Rev. Cancer* 21 (2), 104–121. doi:10.1038/s41568-020-00313-1

von der Maase, H., Sengelov, L., Roberts, J. T., Ricci, S., Dogliotti, L., Oliver, T., et al. (2005). Long-term Survival Results of a Randomized Trial Comparing Gemcitabine Plus Cisplatin, with Methotrexate, Vinblastine, Doxorubicin, Plus Cisplatin in Patients with Bladder Cancer. *Jco* 23 (21), 4602–4608. doi:10.1200/jco.2005.07.757

Xiang, Z., Zhou, Q., Zeng, H., Wang, Z., Zhang, H., Liu, Z., et al. (2020). Intratumoral CCR5+ Neutrophils Identify Immunogenic Subtype Muscle-Invasive Bladder Cancer with Favorable Prognosis and Therapeutic Responses. *Oncoimmunology* 9 (1), 1802176. doi:10.1080/2162402x.2020.1802176

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, orclaim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fgene.2021.767834

DNTTIP1 is a Prognostic Biomarker Correlated With Immune Infiltrates in Hepatocellular Carcinoma: A Study Based on The Cancer Genome Atlas Data

Changyu Zhu[1,2], *Rongsheng Tong*[2,3], *Xiaolei Jiang*[4], *Hua Xiao*[2,3], *Jianmei Guan*[5] and *Jianchen Shu*[6]*

[1]Department of Pharmacy, Sichuan Academy of Medical Science & Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, [2]Personalized Drug Therapy Key Laboratory of Sichuan Province, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, [3]Department of Pharmacy, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China, [4]Department of Pharmacy, Gansu Provincial Hospital of TCM, Lanzhou, China, [5]Central Sterile Supply Department, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, China, [6]Department of Orthopedics, People's Hospital of Leshan, Leshan, China

OPEN ACCESS

Edited by:
Liuyang Wang,
Duke University, United States

Reviewed by:
Raghavendran Partha,
GRAIL Inc., United States
Anqi Zhu,
23andMe, United States

***Correspondence:**
Jianchen Shu
sjc841631@163.com

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

Received: *31 August 2021*
Accepted: *08 November 2021*
Published: *21 February 2022*

Citation:
Zhu C, Tong R, Jiang X, Xiao H, Guan J and Shu J (2022) DNTTIP1 is a Prognostic Biomarker Correlated With Immune Infiltrates in Hepatocellular Carcinoma: A Study Based on The Cancer Genome Atlas Data. Front. Genet. 12:767834. doi: 10.3389/fgene.2021.767834

Deoxynucleotidyltransferase terminal-interacting protein 1 (DNTTIP1) is involved in the deacetylation of p53 in regulating cell cycle and is associated with cancers at the molecular level. In this study, we evaluated the prognostic value in hepatocellular carcinoma (HCC) based on data from The Cancer Genome Atlas (TCGA) database. Kruskal–Wallis test, Wilcoxon signed-rank test, and logistic regression were used to evaluate the relationship between DNTTIP1 expression and clinicopathological features. Cox regression and the Kaplan–Meier method were adopted to evaluate prognosis-related factors. Gene set enrichment analysis (GSEA) was performed to identify the key pathways related to DNTTIP1. The correlations between DNTTIP1 and cancer immune infiltrates were investigated by single-sample Gene Set Enrichment Analysis (ssGSEA). DNTTIP1 was found to be upregulated with amplification in tumor tissues in multiple HCC cohorts. High DNTTIP1 expression was associated with poorer overall survival (OS) and disease-free survival (DFS). GSEA suggested that DNTTIP1 regulates the cell cycle mitotic, G1/S, and G2/M phases and Fc fragment of IgE receptor I (FCERI)–mediated NF-κB and MAPK pathway and Fc fragment of IgG receptor (FCGR) activation pathways. Notably, ssGSEA indicated that DNTTIP1 expression was positively correlated with infiltrating levels of Th2 cells, Tfh, NK CD56 bright cells, aDCs, T helper cells, Th1 cells, and macrophages. These findings suggest that DNTTIP1 is correlated with prognosis and immune infiltration in HCC, which lays a foundation for further study of the immune-regulatory role of DNTTIP1 in HCC.

Keywords: DNTTIP1, HCC, tumor infiltration, prognosis, immune infiltrating

INTRODUCTION

Hepatocellular carcinoma (HCC) is a serious medical problem that ranks sixth among the most common malignancies and is a third leading cause of cancer-related death (Forner et al., 2018). It has been reported that 80% of HCC cases occur in eastern Asia and sub-Saharan Africa where exposure to chronic hepatitis B (CHB) and aflatoxin B1 (AFB1) is the main risk factor (Forner et al., 2018). Over the past 10 years, treatment of HCC has considerably evolved. Today, HCC patients diagnosed at any stage of the disease can benefit from effective treatment that substantially improves their survival. However, several areas still need to be urgently improved. Deep sequencing in HCC has shown that the aggregation of driver and passenger gene alternations in the somatic genome and epigenetic modifications result in HCC (Schulze et al., 2016), which explains its huge molecular heterogeneity. Several candidate biomarkers are being studied in HCC; however, significant challenges exist largely stemming from HCC molecular heterogeneity (Sia et al., 2017; Sia and Llovet, 2017). There has been little progress made on the development of clinically useful biomarkers for early detection of HCC over the last 2 decades (Sengupta and Parikh, 2017). Furthermore, the molecular mechanisms underlying tumorigenesis and progression of HCC remain poorly understood (Zhao et al., 2019). Consequently, there is a critical gap in the current treatment and understanding of HCC due to absence of specific markers for tumor type or disease stage. Hence, the investigation of effective prognostic biomarkers is a pivotal area among several considerations within the research of HCC.

Deoxynucleotidyltransferase terminal-interacting protein 1 (DNTTIP1), that was first reported in 2001, is highly homologous to the transcription factor p65 (Yamashita et al., 2001). DNTTIP1 can enhance DNA polymerase activity (Motea and Berdis, 2010) and form a complex with histone deacetylase (HDAC) (Bantscheff et al., 2011); hence, it has a close relationship with cancer progression. For example, a recent study reported that DNTTIP1 expression could be a specific biomarker present in acute myelocytic leukemia (AML) (Zhuang et al., 2020). Previous research proved that DNTTIP1–HDAC interaction could deacetylate p53 and promote tumor growth in oral squamous cell carcinomas (OSCCs) (Sawai et al., 2018). Former studies have also shown the overexpression of individual HDACs and p53 in HCC (subtypes) and their impact on HCC progression (Ler et al., 2015; Liao et al., 2017; Freese et al., 2019). Based on previous studies, DNTTIP1 might be a promising biomarker of HCC. However, the relationship between DNTTIP1 and HCC has not yet been revealed.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

The RNA-seq data of 371 HCC and 50 normal tissues and patient clinical information were downloaded from the Liver Hepatocellular Carcinoma (LIHC) Project of The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) (Blum et al., 2018) until 12 July 2020. Then, RNA-seq data in FPKM format were transferred to TPM (transcripts per million reads) format, retained, and further analyzed.

Differentially Expressed Gene Analysis

We used the unpaired Student's t-test within the DESeq2 R package (3.6.3) (Love et al., 2014) to compare the expression data (HTseq-Counts) between high- and low-expression groups according to the median DNTTIP1 expression level. The thresholds for the DEGs were |log2-fold change (FC)| >2.0 and adjusted $p < 0.05$.

Enrichment Analysis

Metascape (3.0) (http://metascape.org), a user-friendly, well-maintained, free, gene list online analysis tool for gene analysis and annotation (Zhou et al., 2019), was adopted to perform Gene Ontology (GO) analysis. ClusterProfiler package in R (3.6.3) (Yu et al., 2012) was used to perform Gene Set Enrichment Analysis (GSEA) and detect the correlation between DNTTIP1 and the pathway. As a computational method, GSEA determines whether a priori defined set of genes have statistical significance and concordant differences in two biological states. The samples were divided into high- and low-expression groups according to the median expression level of DNTTIP1. DEseq was used to compare the different expressions between different groups. Gene set permutations were performed with 1,000 times random combinations for each analysis. In the whole process, the expression level of DNTTIP1 was regarded as a phenotype. Additionally, the adjusted P and normalized enrichment score (NES) were utilized to sort the enriched pathways in each phenotype (Subramanian et al., 2005). c2.cp.v7.0.symbols.gmt [Curated] in MSigDB collections was selected as a reference gene set. Gene sets with a false discovery rate (FDR) < 0.25 and adjusted $p < 0.05$ were considered significantly enriched.

Immune Infiltration Analysis by Single-Sample Gene Set Enrichment Analysis

SSGSEA classifies marker gene sets in a single sample with common biologic functions, chromosomal localization, and physiological regulation. In this study, the ssGSEA method was realized by the GSVA package (Hänzelmann et al., 2013) in R to analyze the immune infiltration for 24 types of immune cells and correlation between DNTTIP1 and every immunocyte in HCC samples according to the published literature (Bindea et al., 2013). The relative enrichment score of signature genes was quantified from the gene expression profile for each tumor sample. Spearman's correlation was adopted to analyze the correlation between DNTTIP1 and 24 types of immune cells, and the Wilcoxon rank-sum test was adopted to analyze the infiltration of immune cells between the high-expression groups of DNTTIP1.

Protein–Protein Interaction Network

The Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org) (Szklarczyk et al., 2019) was applied to predict the PPI network. The combined score

TABLE 1 | Demographic and clinicopathological parameters of patients with hepatocellular carcinoma in TCGA–LIHC.

Characters	Level	Low expression of DNTTIP1	High expression of DNTTIP1	*p*
N		186	185	
T stage (%)	T1	100 (54.6%)	81 (43.8%)	0.146
	T2	43 (23.5%)	51 (27.6%)	
	T3	36 (19.7%)	44 (23.8%)	
	T4	4 (2.2%)	9 (4.9%)	
N stage (%)	N0	125 (99.2%)	127 (97.7%)	0.622
	N1	1 (0.8%)	3 (2.3%)	
M stage (%)	M0	126 (97.7%)	140 (99.3%)	0.351
	M1	3 (2.3%)	1 (0.7%)	
Pathologic stage (%)	Stage I	94 (53.7%)	77 (44.8%)	0.141
	Stage II	41 (23.4%)	45 (26.2%)	
	Stage III	36 (20.6%)	49 (28.5%)	
	Stage IV	4 (2.3%)	1 (0.6%)	
Histologic grade (%)	G1	35 (19.0%)	20 (11.0%)	<0.001
	G2	100 (54.3%)	77 (42.3%)	
	G3	47 (25.5%)	75 (41.2%)	
	G4	2 (1.1%)	10 (5.5%)	
Gender (%)	Female	65 (34.9%)	56 (30.3%)	0.395
	Male	121 (65.1%)	129 (69.7%)	
Fibrosis Ishak score (%)	0	52 (42.6%)	22 (24.4%)	0.023
	1/2	16 (13.1%)	15 (16.7%)	
	3/4	11 (9.0%)	17 (18.9%)	
	5/6	43 (35.2%)	36 (40.0%)	
Vascular invasion (%)	No	116 (71.2%)	90 (59.2%)	0.035
	Yes	47 (28.8%)	62 (40.8%)	
Tumor status (%)	Tumor-free	103 (59.2%)	98 (55.1%)	0.499
	With tumor	71 (40.8%)	80 (44.9%)	
TP53 status (%)	Mut	28 (15.3%)	74 (42.3%)	<0.001
	WT	155 (84.7%)	101 (57.7%)	
Age (median [IQR])		61.00 [53.00, 68.75]	61.00 [51.00, 69.00]	0.929
AFP(ng/ml) (median [IQR])		7.50 [3.00, 46.00]	35.50 [7.00,1779.50]	<0.001
Prothrombin time (median [IQR])		1.10 [1.00, 10.00]	1.10 [1.00, 1.30]	0.022

threshold of interaction was 0.4. Furthermore, we extracted the hub genes in this PPI network by using MCODE to identify crucial subnetworks and visualize the PPI network by using Cytoscape (version 3.7.2).

Statistical Analysis

The statistical data acquired from TCGA were merged and processed by R 3.6.3. The Wilcoxon rank-sum test and Wilcoxon signed-rank test were used for comparing the expression levels of DNTTIP1 between HCC and the control group. Kruskal–Wallis test, Wilcoxon rank-sum test, Wilcoxon signed-rank test, and Spearman's correlation were used to analyze the relation between DNTTIP1 expression and grade of clinicopathological factors. Normal and adjusted Pearson's κ^2 test, Fisher's exact test, and univariate logistic regression were used to analyze whether the grade of clinicopathological factors affects DNTTIP1 expression. Spearman's correlation and the Wilcoxon rank-sum test were adopted to analyze the infiltration of immunocytes between the high- and low-expression groups of DNTTIP1. Comparison of multiple groups was performed using a nonparametric Kruskal–Wallis test followed by a post hoc Dunn's test with Bonferroni correction for pairwise comparisons. Univariate Cox regression analysis and multivariate Cox regression analysis were used to evaluate the influence of DNTTIP1 expression and other clinicopathological factors (age and gender) on survival. The significant variables in the univariate analysis ($p < 0.1$) were included into the multivariate analysis (Boeck et al., 2013; Fischer-Rasokat et al., 2021). The Kaplan–Meier curve was drawn to evaluate the prognostic value of DNTTIP1. Hazard risk (HR) of individual factors was estimated by measuring the HR with a 95% confidence interval (CI).

Receiver operating characteristic (ROC) analysis was performed by the pROC package (Robin et al., 2011). The calculated area under the curve (AUC) value ranges, which were from 0.5 to 1.0, indicated the discrimination ability of 50%–100%. We constructed a nomogram by the rms R package based on the results of the multivariate analysis. The predicted survival probability for 1, 3, and 5 years is visualized in the nomogram, which includes a calibration plot as well as significant clinical characteristics. All statistical tests were considered significant when two-tailed $p \leq 0.05$.

RESULTS

Clinical Characteristics

The clinical data of 371 HCC patients included patient age, gender, T stage, N stage, M stage, pathologic stage, histologic grade, fibrosis Ishak score, vascular invasion, tumor status, TP53

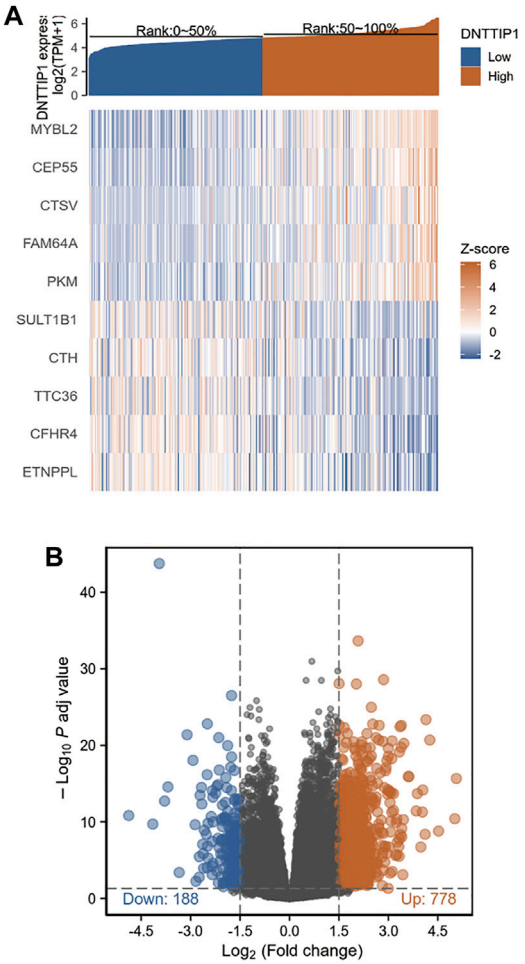


FIGURE 1 | Results of differentially expressed gene (DEG) analysis. **(A)** Heat map of the 10 differentially expressed genes. **(B)** Volcano plot of differentially expressed RNAs.

status (%), age, alpha-fetoprotein (AFP) (ng/ml), and prothrombin time (**Table 1**). A total of 250 males and 121 females with a mean age of 61 years were analyzed in the present study, including 184 white patients and 175 non-white patients. The chi squared test showed that DNTTIP1 was significantly correlated with fibrosis Ishak score (p = 0.023), vacuum invasion (p = 0.035), and TP53 status (p < 0.001). Fisher's exact test showed that DNTTIP1 was significantly correlated with histologic grade (p < 0.001). Wilcoxon rank sum test showed that DNTTIP1 was significantly correlated with weight (p = 0.047), AFP (ng/ml) (p < 0.001), and prothrombin time (p = 0.022). DNTTIP1 expression was not significantly correlated with other clinicopathological features.

Identification of Differentially Expressed Genes in Hepatocellular Carcinoma

Based on the cutoff criteria (|logFC| <1.5 and adjusted p < 0.05), we identified a total of 966 DEGs (778 upregulated and 188 downregulated) after using the DESeq2 package in R (Love et al., 2014) to analyze the HTSeq-count data from TCGA. DEG expressions were illustrated by a heat map and volcano plot (**Figures 1A,B**). DEGs included 812 differentially expressed RNAs (651 upregulated and 161 downregulated), which contained 560 mRNAs (462 upregulated and 98 downregulated) and 252 lncRNAs (189 upregulated and 63 downregulated) (**Supplementary Figures S1A,B**).

Functional Enrichment Analysis of Differentially Expressed Genes

We used Metascape to perform GO enrichment analyses of the functions of DNTTIP1-associated DEGs in HCC. The GO results displayed that DNTTIP1-associated DEGs had significant regulation on immunoglobulin complex, neuron fate commitment, epithelial cell differentiation, neuron fate commitment, antimicrobial humoral response, regionalization, channel activity, DNA-binding transcription activator activity, RNA polymerase II-specific, galactosylceramide biosynthetic process, and regulation of neural retina development. Moreover, galactosylceramide biosynthetic process, skeletal system development, receptor regulator activity, inner ear morphogenesis, anchored component of membrane, dense core granule, an integral component of postsynaptic membrane, anion transmembrane transport, water homeostasis, trigeminal nerve development, apical part of the cell, and morphogenesis of a branching structure were also involved in regulating DNTTIP1-interacting genes (**Figure 2A**). A network of DNTTIP1 and its potential coexpression genes in DNTTIP1-related DEGs are shown in **Figure 2B**. DNTTIP1-related signaling pathways identified by GSEA are shown in **Figures 2C–H**.

As many pathways contribute to tumor formation, high DNTTIP1 expression associated with poor survival may be related to active signaling pathways in HCC. We performed GSEA of differences between low- and high–DNTTIP1 expression data sets to identify the key signaling pathways associated with DNTTIP1. A total of 476 pathways showed significant differences (FDR<0.05, adjusted p < 0.05) in the enrichment of the MSigDB collection (c2.cp.v7.0.symbols). The most significantly enriched signaling pathways based on their NES are shown in Table S1. In particular, DNTTIP1 was related to the cell cycle mitotic, G1/S, and G2/M phases in the cell cycle and Fc fragment of IgE receptor I (FCERI)–mediated NF-κB and MAPK pathway and Fc fragment of IgG receptor (FCGR) activation (**Figures 2C–H**).

The Correlation Between DNTTIP1 Expression and Immune Infiltration

We employed Spearman's correlation to show the association between the expression level (TPM) of DNTTIP1 and immune cell infiltration level quantified by ssGSEA in the HCC tumor microenvironment. Th2 cells were significantly positively correlated with DNTTIP1 expression (Spearman R = 0.394, p < 0.001) (**Figure 3A**). Other immune cell subsets, including

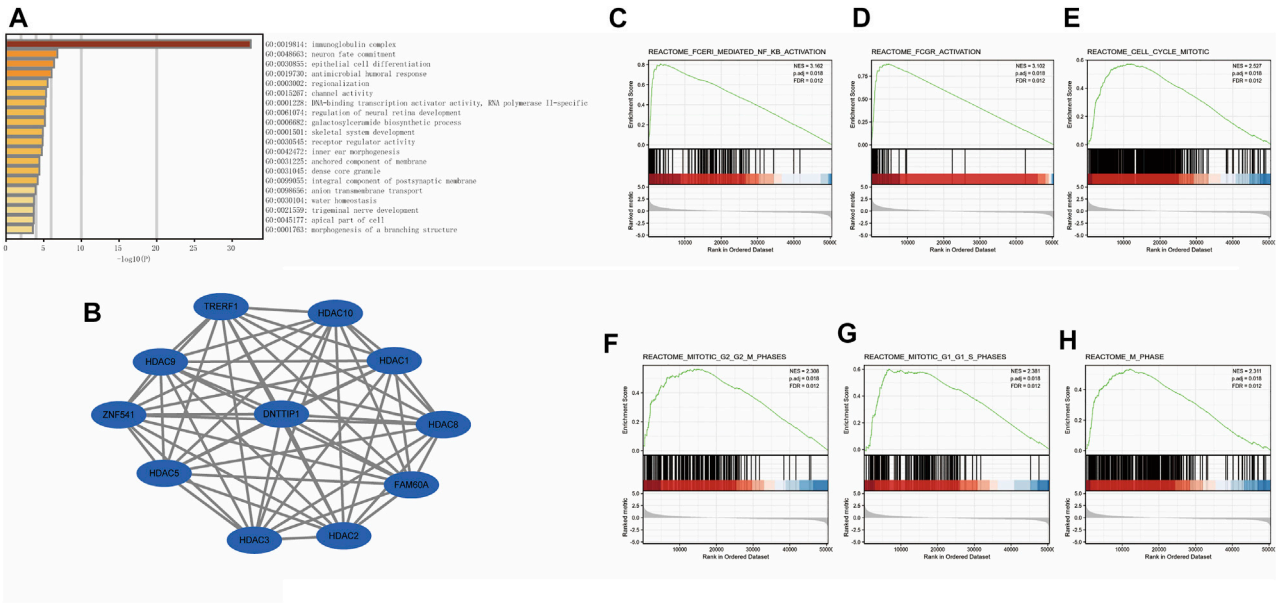


FIGURE 2 | Enrichment analysis of DNTTIP1 in HCC. **(A)** Top 20 biological process enrichment related to DNTTIP1-related genes with enrichment heatmap. **(B)** Network of DNTTIP1 and its potential coexpression genes in DNTTIP1-related DEGs. **(C–H)** Results of enrichment analysis from GSEA.

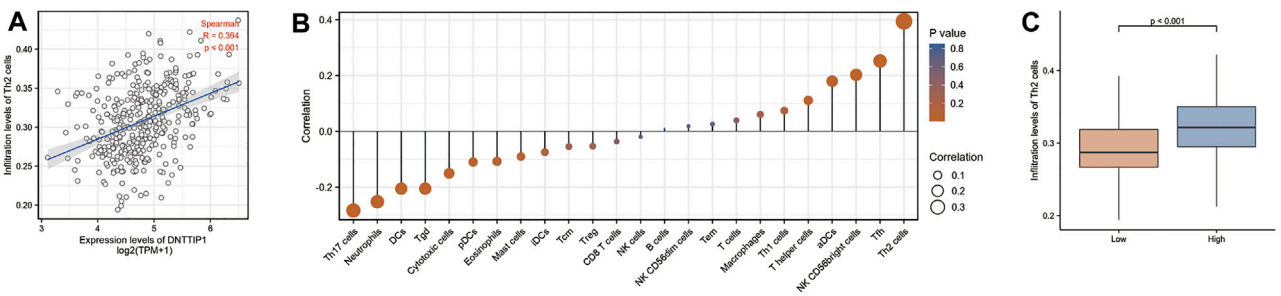


FIGURE 3 | Results of analysis between DNTTIP1 expression and immune infiltration. **(A)** Th2 cells were significantly positively correlated with DNTTIP1 expression. **(B)** Correlation between the relative abundances of 24 immune cells and DNTTIP1 expression level. The color of dots shows the absolute value of Spearman R. **(C)** Th2 cell infiltration level in different DNTTIP1 expression groups.

Tfh, NK CD56 bright cells, aDCs, T helper cells, Th1 cells, and macrophages, were also correlated with DNTTIP1 expression (**Figure 3B**). The Th2 cell infiltration level in the DNTTIP1 high-expression group was significantly different from that of the low-expression group ($p < 0.001$) (**Figure 3C**).

Associations Between DNTTIP1 Expression and Clinico-Pathologic Variables

The Kruskal–Wallis rank-sum test revealed that the expression of DNTTIP1 was significantly correlated with the T stage of HCC ($p = 0.049$). Pathologic stage ($p = 0.015$), histologic grade ($p < 0.001$), fibrosis Ishak score ($p = 0.008$) (**Figures 4A–D**), and Wilcoxon rank-sum test revealed that expression of DNTTIP1 was significantly correlated with AFP($p < 0.001$), vascular invasion ($p = 0.037$), TP53 status ($p < 0.001$), and prothrombin time ($p < 0.001$) (**Figures 4E–H**). Bonferroni correction was applied to the p value of Dunn's test to correct for multiple comparisons within the T stage of HCC. Pathologic stage, histologic grade, and fibrosis Ishak score are shown in **Supplementary Figures S2A–D**. Logistic regression analysis showed that DNTTIP1 was significantly correlated with T stage ($p = 0.038$), histologic grade ($p < 0.001$), fibrosis Ishak score ($p = 0.007$), vascular invasion ($p = 0.026$), and TP53 status ($p < 0.001$) (**Table 2**).

DNTTIP1 is highly expressed in LIHC samples when compared with normal tissues (**Figures 5A,B**). The area under the curve (AUC) of DNTTIP1 was 0.905, which indicated that DNTTIP1 might be a potential diagnostic molecule (**Figure 5C**). In many other cancer types such as adrenocortical carcinoma

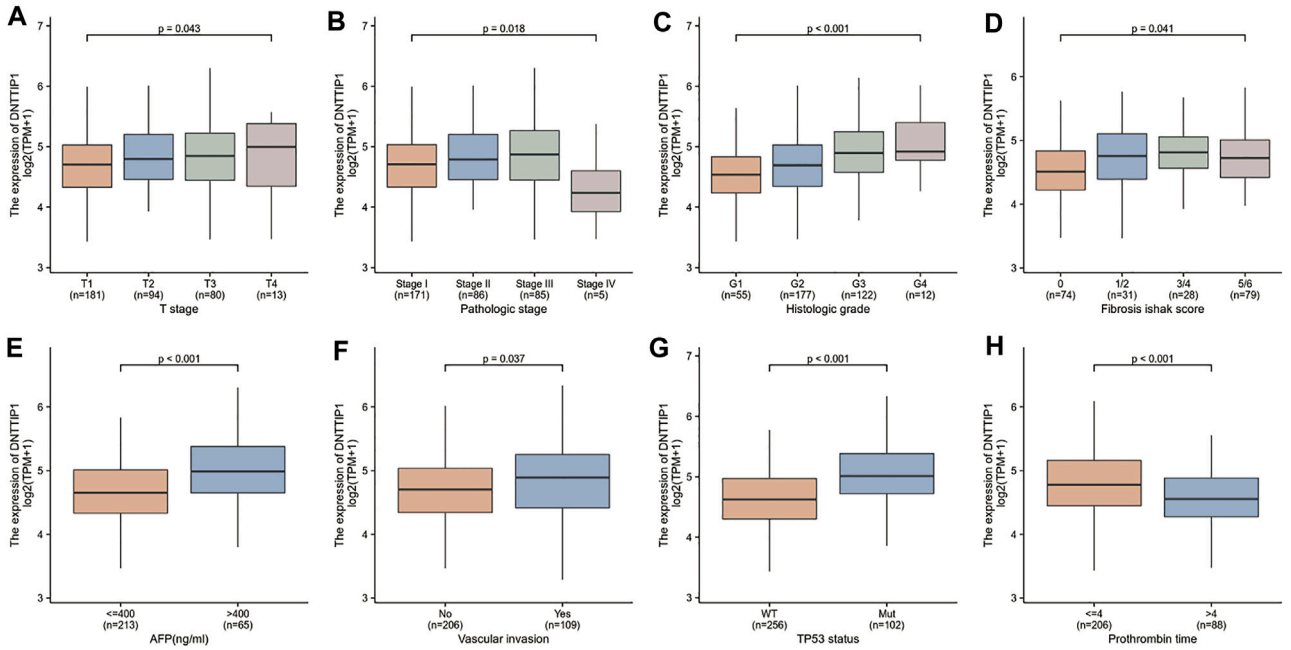


FIGURE 4 | Association between the DNTTIP1 expression and different clinicopathological characteristics. **(A)** Association between the DNTTIP1 expression and T stage of HCC, **(B)** pathologic stage, **(C)** histologic grade, **(D)** fibrosis Ishak score, **(E)** AFP, **(F)** vascular invasion, **(G)** TP53 status, and **(H)** prothrombin time.

TABLE 2 | DNTTIP1 expression associated with clinicopathological characteristics (logistic regression).

Characteristics	Odds ratio in DNTTIP1 expression	Odds ratio (OR)	*p*-value
T stage (T2 and T3 and T4 vs. T1)	368	1.55 (1.03–2.34)	0.038
N stage (N1 vs. N0)	256	2.95 (0.37–60.13)	0.351
M stage (M1 vs. M0)	270	0.30 (0.01–2.38)	0.300
Pathologic stage (Stage II and Stage III and Stage IV vs. Stage I)	347	1.43 (0.94–2.19)	0.096
Histologic grade (G3 and G4 vs. G1 and G2)	366	2.41 (1.56–3.76)	<0.001
Residual tumor (R1 and R2 vs. R0)	342	1.05 (0.40–2.76)	0.919
Fibrosis Ishak score (1/2 and 3/4 and 5/6 vs. 0)	212	2.30 (1.27–4.24)	0.007
Adjacent hepatic tissue inflammation (mild and severe vs. none)	234	1.53 (0.91–2.58)	0.113
Vascular invasion (yes vs. no)	315	1.70 (1.07–2.73)	0.026
Tumor status (with tumor vs. tumor-free)	352	1.18 (0.78–1.81)	0.433
TP53 status (Mut vs. WT)	358	4.06 (2.48–6.79)	<0.001

(ACC), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma, and endocervical adenocarcinoma (CESC), DNTTIP1 was also significantly overexpressed when compared with normal tissues (**Figure 5D**).

In the Cox regression model, variables with $p < 0.1$ in univariate Cox regression were included in multivariate Cox regression. The variables that met this threshold were T stage ($p < 0.001$), M stage ($p = 0.018$), pathologic stage ($p < 0.001$), tumor status ($p < 0.001$), and DNTTIP1 ($p < 0.001$) (**Table 3**). Furthermore, multivariate Cox regression showed that tumor status ($p = 0.001$) and DNTTIP1 ($p = 0.027$) were independent prognostic factors for overall survival ($p < 0.05$).

The Kaplan–Meier survival curve drawn by the survminer package in R was used to evaluate the prognostic value of DNTTIP1 in overall survival of HCC (**Figure 6A**). Gene expression values were divided into high- and low-expression groups according to the median value. High expression of DNTTIP1 was associated with poor overall survival (HR = 1.94 (1.36–2.76), $p < 0.001$). The lower part of these figures is shown in the risk table which records the number of people still under follow-up at each time point. The prognosis data are derived from an article published in Cell (Liu et al., 2018).

Based on multivariate Cox regression, a nomogram that integrated DNTTIP1 and independent clinical risk factors (tumor status and DNTTIP1) was drawn by using the rms package in R as a tool for clinicians to predict the prognosis of HCC patients (**Supplementary Figure S3A**). The C-index in the nomogram defines the accuracy of the model. C-index values are generally between 0.5–1. The main contents of the nomogram are as follows: the variables of the prediction model (scales are

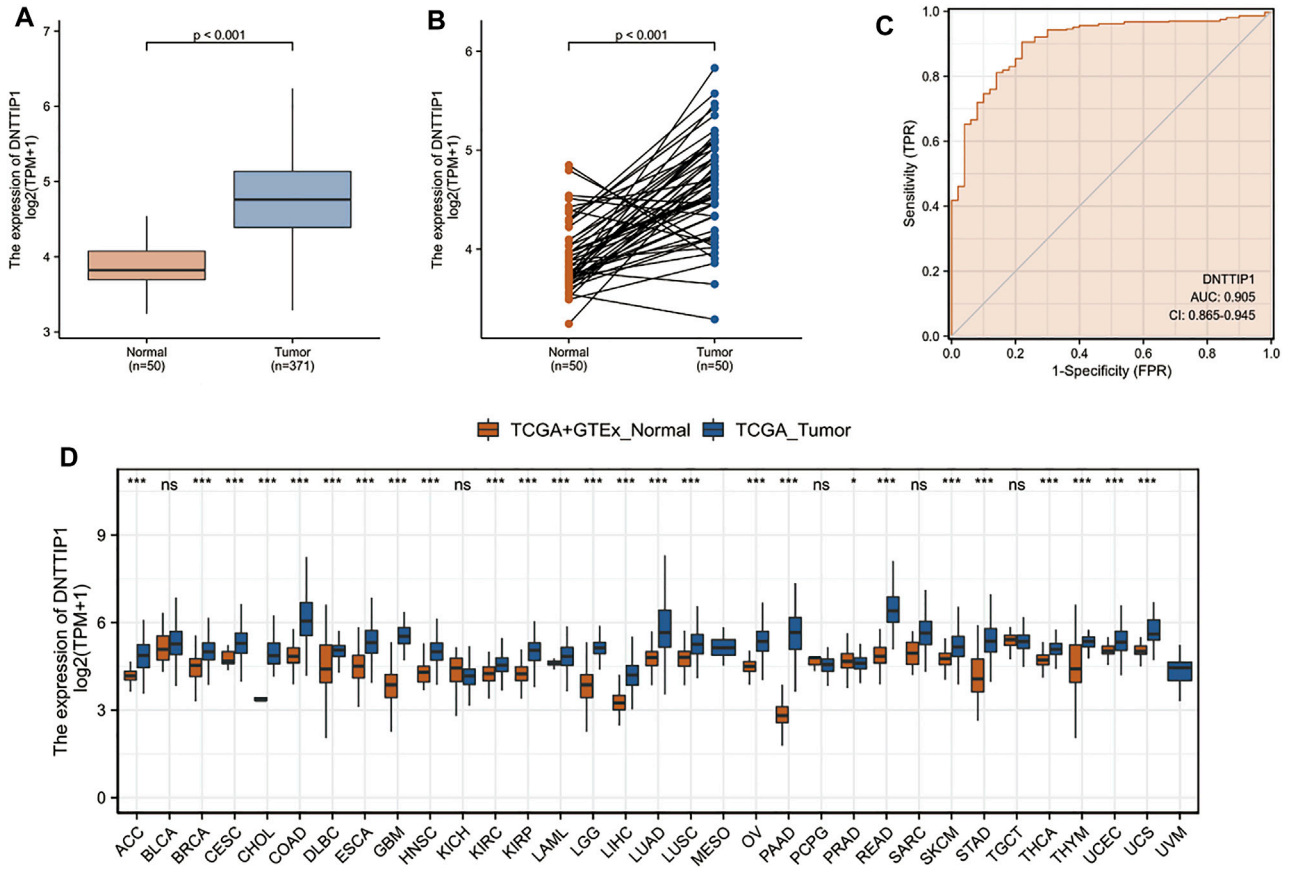


FIGURE 5 | Prognostic value of DNTTIP1 in LIHC and other cancer types. **(A,B)** DNTTIP1 is highly expressed in LIHC samples when compared with normal tissues. **(C)** ROC curve indicates that DNTTIP1 is a potential diagnostic marker. **(D)** DNTTIP1 is significantly overexpressed in other cancer types when compared with normal tissues. Data were processed TPM format RNA-seq data originated from TCGA and GTEx downloaded from UCSC XENA (https://xenabrowser.net/datapages/).

TABLE 3 | Univariate and multivariate analyses of clinicopathological parameters in patients with hepatocellular carcinoma in TCGA–LIHC.

Characteristics	Total (*N*)	HR (95%CI) univariate analysis	*p*-value univariate analysis	HR (95%CI) multivariate analysis	*p*-value multivariate analysis
T stage (T2 and T3 and T4 vs. T1)	367	2.109 (1.469–3.028)	<0.001	0.890 (0.120–6.620)	0.909
N stage (N1 vs. N0)	256	2.004 (0.491–8.181)	0.333		
M stage (M1 vs. M0)	270	4.032 (1.267–12.831)	0.018	1.681 (0.400–7.063)	0.478
Pathologic stage (Stage II and Stage III and Stage IV vs. Stage I)	346	2.074 (1.418–3.032)	<0.001	2.422 (0.314–18.705)	0.396
Histologic grade (G3 and G4 vs. G1 and G2)	365	1.120 (0.781–1.606)	0.539		
Residual tumor (R1 and R2 vs. R0)	341	1.571 (0.795–3.104)	0.194		
Age (>60 vs. ≤60)	370	1.248 (0.880–1.768)	0.214		
Gender (male vs. female)	370	0.816 (0.573–1.163)	0.260		
AFP (ng/ml) (>400 vs. ≤400)	277	1.056 (0.646–1.727)	0.827		
Albumin (g/dl) (≥3.5 vs. <3.5)	296	0.921 (0.565–1.503)	0.743		
Prothrombin time (>4 vs. ≤4)	293	1.330 (0.877–2.015)	0.179		
Fibrosis Ishak score (1/2 and 3/4 & 5/6 vs. 0)	211	0.779 (0.470–1.293)	0.334		
Vascular invasion (yes vs. no)	314	1.348 (0.890–2.042)	0.159		
Tumor status (with tumor vs. tumor-free)	351	2.361 (1.620–3.441)	<0.001	2.299 (1.400–3.774)	0.001
TP53 status (Mut vs. WT)	357	1.434 (0.972–2.115)	0.069	1.393 (0.822–2.359)	0.218
DNTTIP1 (high vs. low)	370	1.941 (1.363–2.764)	<0.001	1.735 (1.064–2.829)	0.027

HR, hazard ratio; CI, confidence interval; AFP, alpha-fetoprotein.

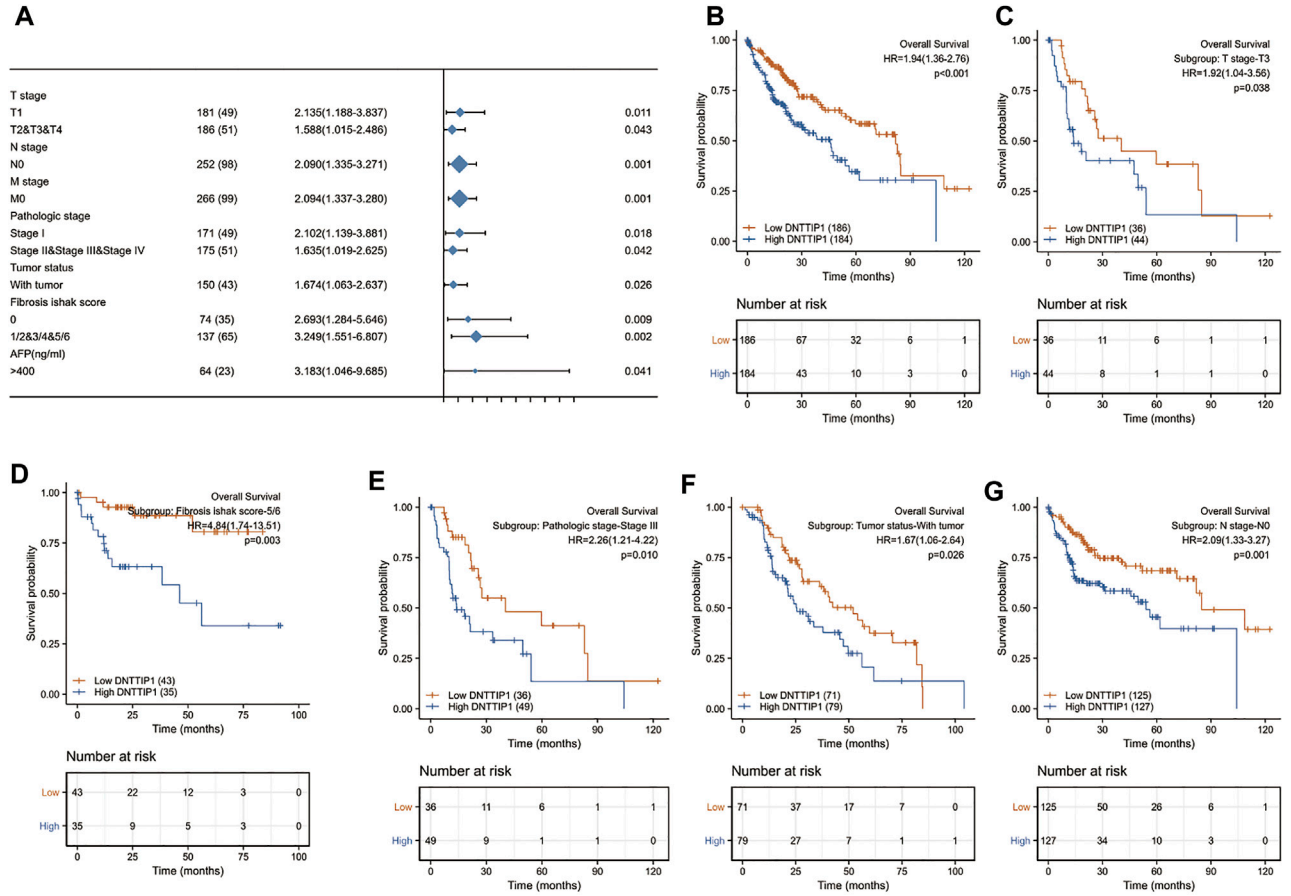


FIGURE 6 | Prognostic value of DNTTIP1 in different analyses. **(A)** Prognostic value of DNTTIP1 in overall survival of HCC. **(B)** Forest plot of the prognostic value of DNTTIP1 in overall survival in different subgroups of HCC in TCGA–LIHC. **(C–G)** High expression of DNTTIP1 was associated with poor overall survival in different subgroups.

marked on the line segment corresponding to each variable, which represent the range of the variable, while the length of the line segment reflects the contribution of the factor to the outcome events); score (the individual score of each variable under different values and the total score of the sum of the individual scores corresponding to the value of all variables); prediction probability (survival probability at different time nodes). Predictions made by the calibration curve conformed well to the ideal line (the 45-degree line), which indicated that the prediction was consistent with the observation (**Supplementary Figure S3B**).

The prognostic values of DNTTIP1 in relation to the overall survival under different subgroups of LIHC in TCGA are shown in **Table 2**. The expression of DNTTIP1 had significant effect in T1 (HR = 2.135 (1.188–3.837), $p = 0.011$), T2, and T3 and T4 (HR = 1.588 (1.015–2.486, $p = 0.043$) subgroups of T stage, N0 (HR = 2.090 (1.335–3.271, $p = 0.001$) subgroups of N stage, and M0 (HR = 2.094 (1.337–3.280), $p = 0.001$) of M stage (**Figure 6B**). High expression of DNTTIP1 was associated with poor overall survival in the T3 subgroup of T stage (HR = 1.92 (1.04–3.56), $p = 0.038$), 5/6 subgroup of fibrosis Ishak score (HR = 4.84 (1.74–13.51), $p = 0.003$), Stage III subgroup of pathologic stage (HR = 2.26 (1.21–4.22), $p = 0.010$), with tumor subgroup of tumor status (HR = 1.92 (1.04–3.56), $p = 0.038$), and N0 subgroup of N stage (HR = 2.09 (1.33–3.27), $p = 0.001$) (**Figures 6C–G**).

DISCUSSION

Clinicians have been perplexed for a long time by the early detection of HCC that adopts AFP as an indicator for HCC screening at the early stage. More than 30% of HCC patients are AFP negative (Luo et al., 2018). Therefore, new markers for HCC are needed to improve early diagnosis.

Previous *in vitro* and *in vivo* studies have shown that DNTTIP1 interacts with histone deacetylase 1 (HDAC1) to control the status of p53 acetylation (Sawai et al., 2018), thus regulating the expression of several p53 target genes that participate in cell cycle arrest (Narayanan et al., 2003; Zhang et al., 2004) and promote tumor cell migration and invasion in oral squamous cell carcinomas (OSCCs). Former bioinformatics analysis indicated that DNTTIP1 could predict the survival of patients with acute myelocytic leukemia (Zhuang et al., 2020). Yet, the role of DNTTIP1 in HCC progression needs to be further investigated.

In this study, bioinformatics analysis of sequencing data from TCGA was performed to gain a deeper understanding of the potential function of DNTTIP1 in HCC and guide future research in HCC. RNA-seq normalization was realized to remove the technical deviation of sequencing data: sequencing depth and gene length. We transferred count data to TPM. This RNA-seq normalization step not only eliminated the influence of single sequencing depth and gene length but also TPM quantified gene expression rate so that the data can be compared among cells with different basal expressions.

Elevated DNTTIP1 expression in HCC was associated with advanced clinicopathological features (AFP, fibrosis Ishak score, histologic grade, pathological stage, TP53 status, and vascular invasion), poor prognosis, and survival time. Furthermore, in univariate and multivariate Cox regression analyses, we found that after removing confounding factors, DNTTIP1 was still an independent prognostic factor, which showed a higher prognostic value than many other clinical variables, including AFP. Our results suggested that DNTTIP1 is a potential prognostic and diagnostic marker deserving further clinical validation. The function of DNTTIP1 in HCC was further investigated in GSEA using TCGA data.

The PPI network indicates that DNTTIP1 can interact with several histone deacetylase (HDAC) family members other than HDAC1. HDACs are key enzymes that maintain the acetylation balance of nucleosomes in basic units of chromosomes. Their catalytic histone deacetylation is closely related to the inhibition of gene transcription (Sanaei and Kavoosi, 2019). GSEA showed that cell cycle mitotic, G1/S, and G2/M phases and Fc fragment of IgE receptor I (FCERI)–mediated NF-κB and MAPK pathway and Fc fragment of IgG receptor (FCGR) activation in HCC were enriched in the DNTTIP1 high-expression phenotype. These findings indicated that DNTTIP1 might participate in the regulation of cell cycle and immune response in the tumorigenesis of HCCs.

SSGSEA and Spearman's correlation were adopted to uncover connections between DNTTIP1 expression and immune infiltration levels in HCC. Our results demonstrated that DNTTIP1 expression was significantly positively correlated with Th2 cells. Furthermore, there was a strong-to-moderate correlation between Tfh, NK CD56 bright cells, aDCs, T helper cells, Th1 cells, macrophages, and DNTTIP1 expression. Our results suggest a possible mechanism where DNTTIP1 regulates the balance of Th1/Th2 in HCC. The Th2 cells produce IL-4 and IL-10 and inhibit the host immune system, hence having a role in promoting tumor growth (Ko et al., 2001; Zhao et al., 2015). This indicates that overexpression of DNTTIP1 promotes Th2 cell immune response and infiltration in tumor progression. Th1/Th2 balance can be regulated to inhibit tumor progression. A global Th1/Th2-like cytokine shift (a decrease in Th1 and an increase in Th2 cytokines) can be induced to promote HCC metastasis (Budhu et al., 2006).

On the other hand, there was an inverse correlation between Th17 immune cells, neutrophils, DCs, and DNTTIP1. Th17 immune cells were associated with both good and bad prognoses (Guéry and Hugues, 2015). It is reported that Th17 cells were correlated with a bad prognosis of HCC (Zhang et al., 2009). However, Th17 cells can also drive antitumor immune responses by recruiting immune cells into tumors, activating effector $CD8^+$ T cells, or even directly by converting toward Th1 phenotype and producing IFN-γ (Lee et al., 2009; Nistala et al., 2010). The downregulation of Th17 caused by DNTTIP1 overexpression may affect h1/h2 balance, leading to a bad prognosis. Neutrophils have an active role in regulating the immune system; they can promote or inhibit the establishment of a permissive tumor microenvironment (Shaul and Fridlender, 2017). Many studies have shown that tumor-associated neutrophils not only promote tumor growth (Wang et al., 2020) but also antitumor effects on tumors and can regulate their different phenotypes through tumor signal transduction. Our result indicates that the antitumor function of neutrophils may be hindered when DNTTIP1 is overexpressed. Most DC subsets have been found in tumors where they play a major role in cancer immune surveillance by coordinating adaptive immunity against tumor antigens. DCs are critical for autoimmunity and tissue inflammation and have prominent roles in cellular and humoral immune response and protection from infectious diseases or tumors (Vatner and Janssen, 2019). Due to the role of DCs in initiating antitumor immunity, there is a negative selective pressure hampering the accumulation of DCs by tumor-secreted mediators that inhibit dendropoiesis, promote DC apoptosis (Pirtskhalaishvili et al., 2000), and accelerate DC turnover. The suppression of DCs is normally found in tumors and may facilitate HCC progression. All findings according to ssGSEA support that DNTTIP1 has a role in regulating and recruiting immune infiltrating cells in HCC. However, more trials are needed to accurately understand the relationship between DNTTIP1 and Th1/Th2 balance *in vivo*.

In addition, the Kaplan–Meier survival curves with high HR for poor OS and PFS when DNTTIP1 was highly expressed in HCC showed the correlation between high-level expression of DNTTIP1 and poor prognosis of HCC, thus suggesting that DNTTIP1 was a prognostic biomarker in HCC.

The Cox HR model suggested that DNTTIP1 was strongly associated with OS in patients. Moreover, we developed a DNTTIP1-related nomogram to predict the 1-, 3-, and 5-year survival probabilities in HCC patients. The results were confirmed by calibration plots and log-rank tests.

Moreover, Kaplan–Meier survival analysis was performed in HCC patients according to their DNTTIP1 expression levels, stratified by clinicopathological characteristics. A high level of DNTTIP1 expression was associated with poor prognosis of HCC patients with a fibrosis Ishak score of 5/6, thus indicating that the high association between DNTTIP1 expression level and survival may be influenced by the degree of hepatic fibrosis.

Although our investigation of the relationship between DNTTIP1 and HCC furthered our understanding of the vital role of DNTTIP1 in HCC, some limitations remained. First, cell experiments and clinical samples should be used to verify the correlation between DNTTIP1 mRNA and protein expression. In the present study, we only used mRNA levels to predict protein expression (Fortelny et al., 2017). Second, clinical factors such as the details of patient treatment should be sufficiently considered to clarify the specific role of DNTTIP1 in the development of HCC.

Third, while multicenter research based on public databases intends to overcome the shortage of single-center studies, retrospective studies have two major shortages. One is missing variables. In our study, to clarify the specific role of DNTTIP1 in the development of HCC comprehensively, more clinical factors should be taken into consideration such as the detail of treatments for every single patient involved. However, the information of treatments was often inconsistent or even lacking in public databases; the other is sample size imbalance. We have a smaller number of healthy samples in our control group than that of HCC patients in our study; the sample size imbalance may lead to statistical bias. Therefore, future prospective studies are needed to reduce analysis bias. Finally, we cannot illustrate the expression of DNTTIP1 from the protein level and also cannot evaluate the direct mechanisms of DNTTIP1 involved in HCC progression. Consequently, further studies are needed to clarify the direct mechanisms of DNTTIP1 in HCC.

In summary, DNTTIP1 has an important role in the regulation of cell cycle and immune response in the tumorigenesis of HCC. Further studies are needed to clarify the biological mechanisms of DNTTIP1 in HCC. In addition, additional experiments are needed to evaluate the relationship between DNTTIP1 expression and clinical features, HCC stage, and prognosis using additional clinical data, which might facilitate the identification of new markers for evaluating tumor stage, aiding drug development, and improving treatment efficiency.

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

CZ and JS conceived and designed the study; RT, XJ, HX, CZ, and JG analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

We are grateful for support of Key Research and Development Program of Science and Technology Department of Sichuan Province (2019YFS0514) and National Key R and D Program of China (2020YFC2005500).

ACKNOWLEDGMENTS

We are grateful to all co-authors of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.767834/full#supplementary-material

Supplementary Figure 1 | Volcano plot of differentially expressed RNAs. **(A)** Volcano plot of differentially expressed mRNAs. **(B)** Volcano plot of differentially expressed lncRNAs.

Supplementary Figure 2 | Association between DNTTIP1 expression and different clinicopathological characteristics after *post hoc* Dunn's test with Bonferroni correction for pairwise comparisons. **(A)** Association between DNTTIP1 expression and the T stage of HCC, **(B)** pathologic stage, **(C)** histologic grade, and **(D)** Fibrosis Ishak score.

Supplementary Figure 3 | Nomogram and calibration curve. **(A)** Nomogram that integrated DNTTIP1 and independent clinical risk factors between tumor status and DNTTIP1. **(B)** Predictions made by the calibration curve were found to conform well to the ideal line.

REFERENCES

Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A.-M., et al. (2011). Chemoproteomics Profiling of HDAC Inhibitors Reveals Selective Targeting of HDAC Complexes. *Nat. Biotechnol.* 29, 255–265. doi:10.1038/nbt.1759

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

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Boeck, S., Wittwer, C., Heinemann, V., Haas, M., Kern, C., Stieber, P., et al. (2013). Cytokeratin 19-fragments (CYFRA 21-1) as a Novel Serum Biomarker for Response and Survival in Patients with Advanced Pancreatic Cancer. *Br. J. Cancer* 108, 1684–1694. doi:10.1038/bjc.2013.158

Budhu, A., Forgues, M., Ye, Q.-H., Jia, H.-L., He, P., Zanetti, K. A., et al. (2006). Prediction of Venous Metastases, Recurrence, and Prognosis in Hepatocellular Carcinoma Based on a Unique Immune Response Signature of the Liver Microenvironment. *Cancer cell* 10, 99–111. doi:10.1016/j.ccr.2006.06.016

Fischer-Rasokat, U., Renker, M., Liebetrau, C., Weferling, M., Rolf, A., Hain, A., et al. (2021). Long-Term Survival in Patients with or without Implantable Cardioverter Defibrillator after Transcatheter Aortic Valve Implantation. *JCM* 10, 2929. doi:10.3390/jcm10132929

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular Carcinoma. *The Lancet* 391, 1301–1314. doi:10.1016/S0140-6736(18)30010-2

Fortelny, N., Overall, C. M., Pavlidis, P., and Freue, G. V. C. (2017). Can We Predict Protein from mRNA Levels? *Nature* 547, E19–E20. doi:10.1038/nature22293

Freese, K., Seitz, T., Dietrich, P., Lee, S. M. L., Thasler, W. E., Bosserhoff, A., et al. (2019). Histone Deacetylase Expressions in Hepatocellular Carcinoma and Functional Effects of Histone Deacetylase Inhibitors on Liver Cancer Cells *In Vitro*. *Cancers* 11, 1587. doi:10.3390/cancers11101587

Guéry, L., and Hugues, S. (2015). Th17 Cell Plasticity and Functions in Cancer Immunity. *Biomed. Res. Int.* 2015, 1–11. doi:10.1155/2015/314620

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

Kavoosi, F., and Sanaei, M. (2019). Histone Deacetylases and Histone Deacetylase Inhibitors: Molecular Mechanisms of Action in Various Cancers. *Adv. Biomed. Res.* 8, 63. doi:10.4103/abr.abr_142_19

Ko, K. S., Lee, M., Koh, J. J., and Kim, S. W. (2001). Combined Administration of Plasmids Encoding IL-4 and IL-10 Prevents the Development of Autoimmune

Diabetes in Nonobese Diabetic Mice. *Mol. Ther.* 4, 313–316. doi:10.1006/mthe.2001.0459

Lee, Y. K., Turner, H., Maynard, C. L., Oliver, J. R., Chen, D., Elson, C. O., et al. (2009). Late Developmental Plasticity in the T Helper 17 Lineage. *Immunity* 30, 92–107. doi:10.1016/j.immuni.2008.11.005

Ler, S. Y., Leung, C. H. W., Khin, L. W., Lu, G.-D., Salto-Tellez, M., Hartman, M., et al. (2015). HDAC1 and HDAC2 Independently Predict Mortality in Hepatocellular Carcinoma by a Competing Risk Regression Model in a Southeast Asian Population. *Oncol. Rep.* 34, 2238–2250. doi:10.3892/or.2015.4263

Liao, P., Zeng, S. X., Zhou, X., Chen, T., Zhou, F., Cao, B., et al. (2017). Mutant P53 Gains its Function via C-Myc Activation upon CDK4 Phosphorylation at Serine 249 and Consequent PIN1 Binding. *Mol. Cel* 68, 1134–1146.e6. doi:10.1016/j.molcel.2017.11.006

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, 400–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. doi:10.1186/s13059-014-0550-8

Luo, P., Yin, P., Hua, R., Tan, Y., Li, Z., Qiu, G., et al. (2018). A Large-scale, Multicenter Serum Metabolite Biomarker Identification Study for the Early Detection of Hepatocellular Carcinoma, *Hepatology*, 67, 662–675. doi:10.1002/hep.29561

Motea, E. A., and Berdis, A. J. (20101804). Terminal Deoxynucleotidyl Transferase: the story of a Misguided DNA Polymerase. *Biochim. Biophys. Acta (Bba) - Proteins Proteomics* 1804, 1151–1166. doi:10.1016/j.bbapap.2009.06.030

Narayanan, B. A., Narayanan, N. K., Re, G. G., and Nixon, D. W. (2003). Differential Expression of Genes Induced by Resveratrol in LNCaP Cells: P53-Mediated Molecular Targets. *Int. J. Cancer* 104, 204–212. doi:10.1002/ijc.10932

Nistala, K., Adams, S., Cambrook, H., Ursu, S., Olivito, B., de Jager, W., et al. (2010). Th17 Plasticity in Human Autoimmune Arthritis Is Driven by the Inflammatory Environment. *Proc. Natl. Acad. Sci.* 107, 14751–14756. doi:10.1073/pnas.1003852107

Pirtskhalaishvili, G., Shurin, G. V., Esche, C., Cai, Q., Salup, R. R., Bykovskaia, S. N., et al. (2000). Cytokine-mediated protection of Human Dendritic Cells from Prostate Cancer-Induced Apoptosis Is Regulated by the Bcl-2 Family of Proteins. *Br. J. Cancer* 83, 506–513. doi:10.1054/bjoc.2000.1289

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77

Sawai, Y., Kasamatsu, A., Nakashima, D., Fushimi, K., Kasama, H., Iyoda, M., et al. (2018). Critical Role of Deoxynucleotidyl Transferase Terminal Interacting Protein 1 in Oral Cancer. *Lab. Invest.* 98, 980–988. doi:10.1038/s41374-018-0070-3

Schulze, K., Nault, J.-C., and Villanueva, A. (2016). Genetic Profiling of Hepatocellular Carcinoma Using Next-Generation Sequencing. *J. Hepatol.* 65, 1031–1042. doi:10.1016/j.jhep.2016.05.035

Sengupta, S., and Parikh, N. D. (2017). Biomarker Development for Hepatocellular Carcinoma Early Detection: Current and Future Perspectives. *Hepatic Oncol.* 4, 111–122. doi:10.2217/hep-2017-0019

Shaul, M. E., and Fridlender, Z. G. (2017). Neutrophils as Active Regulators of the Immune System in the Tumor Microenvironment. *J. Leukoc. Biol.* 102, 343–349. doi:10.1189/jlb.5MR1216-508R

Sia, D., and Llovet, J. M. (2017). Translating '-omics' Results into Precision Medicine for Hepatocellular Carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 14, 571–572. doi:10.1038/nrgastro.2017.103

Sia, D., Villanueva, A., Friedman, S. L., and Llovet, J. M. (2017). Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology* 152, 745–761. doi:10.1053/j.gastro.2016.11.048

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

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

Vatner, R. E., and Janssen, E. M. (2019). STING, DCs and the Link between Innate and Adaptive Tumor Immunity. *Mol. Immunol.* 110, 13–23. doi:10.1016/j.molimm.2017.12.001

Wang, Y., Yao, R., Zhang, D., Chen, R., Ren, Z., and Zhang, L. (2020). Circulating Neutrophils Predict Poor Survival for HCC and Promote HCC Progression through P53 and STAT3 Signaling Pathway. *J. Cancer* 11, 3736–3744. doi:10.7150/jca.42953

Yamashita, N., Shimazaki, N., Ibe, S., Kaneko, R., Tanabe, A., Toyomoto, T., et al. (2001). Terminal Deoxynucleotidyltransferase Directly Interacts with a Novel Nuclear Protein that Is Homologous to P65. *Genes Cells* 6, 641–652. doi:10.1046/j.1365-2443.2001.00449.x

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

Zhang, J.-P., Yan, J., Xu, J., Pang, X.-H., Chen, M.-S., Li, L., et al. (2009). Increased Intratumoral IL-17-producing Cells Correlate with Poor Survival in Hepatocellular Carcinoma Patients. *J. Hepatol.* 50, 980–989. doi:10.1016/j.jhep.2008.12.033

Zhang, S., Cao, H. J., Davis, F. B., Tang, H.-Y., Davis, P. J., and Lin, H.-Y. (2004). Oestrogen Inhibits Resveratrol-Induced post-translational Modification of P53 and Apoptosis in Breast Cancer Cells. *Br. J. Cancer* 91, 178–185. doi:10.1038/sj.bjc.6601902

Zhao, P., Bu, X., Wei, X., Sun, W., Xie, X., Li, C., et al. (2015). Dendritic Cell Immunotherapy Combined with Cytokine-Induced Killer Cells Promotes Skewing toward Th2 Cytokine Profile in Patients with Metastatic Non-small Cell Lung Cancer. *Int. Immunopharmacology* 25, 450–456. doi:10.1016/j.intimp.2015.02.010

Zhao, X., Liu, J., Ge, S., Chen, C., Li, S., Wu, X., et al. (2019). Saikosaponin A Inhibits Breast Cancer by Regulating Th1/Th2 Balance. *Front. Pharmacol.* 10, 624. doi:10.3389/fphar.2019.00624

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

Zhuang, H., Chen, Y., Sheng, X., Hong, L., Gao, R., and Zhuang, X. (2020). Searching for a Signature Involving 10 Genes to Predict the Survival of Patients with Acute Myelocytic Leukemia through a Combined Multi-Omics Analysis. *PeerJ* 8, e9437. doi:10.7717/peerj.9437

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fgene.2022.655169

Establishment and Validation of a 5 m6A RNA Methylation Regulatory Gene Prognostic Model in Low-Grade Glioma

*Zhiqun Bai, Xuemei Wang and Zhen Zhang**

Department of Ultrosonic Diagnosis, The First Affiliated Hospital of China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Longxiang Xie,
Henan University, China

Reviewed by:
Sven R. Kantelhardt,
Johannes Gutenberg University Mainz, Germany
Xianfang Wang,
Henan Institute of Science and Technology, China
Dianshuang Zhou,
Harbin Institute of Technology, China

****Correspondence:***
Zhen Zhang
zhen2017zhang@qq.com

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

Received: *16 April 2021*
Accepted: *07 February 2022*
Published: *25 February 2022*

Citation:
Bai Z, Wang X and Zhang Z (2022) Establishment and Validation of a 5 m6A RNA Methylation Regulatory Gene Prognostic Model in Low-Grade Glioma. Front. Genet. 13:655169. doi: 10.3389/fgene.2022.655169

Background: The prognosis of low-grade glioma (LGG) is different from that of other intracranial tumors. Although many markers of LGG have been established, few are used in clinical practice. M6A methylation significantly affects the biological behavior of LGG tumors. Therefore, establishment of an LGG prognostic model based on m6A methylation regulatory genes is of great interest.

Methods: Data from 495 patients from The Cancer Genome Atlas (TCGA) and 172 patients from the Chinese Glioma Genome Atlas (CGGA) were analyzed. Univariate Cox analysis was used to identify methylation regulatory genes with prognostic significance. LASSO Cox regression was used to identify prognostic genes. Receiver operating characteristic (ROC) and Kaplan–Meier curves were used to verify the accuracy of the model. Gene Set Enrichment Analysis (GSEA) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to identify cellular pathways that were significantly associated with the prognosis of LGG.

Results: A glioma prognostic model based on five methylation regulatory genes was established. Compared with low-risk patients, patients identified as high risk had a poorer prognosis. There was a high degree of consistency between the internal training and internal validation CGGA cohorts and the external validation TCGA cohort. Furthermore, KEGG and GSEA analyses showed that the focal adhesion and cell cycle pathways were significantly upregulated in high-risk patients. This signature could be used to distinguish among patients with different immune checkpoint gene expression levels, which may inform immune checkpoint inhibitor (ICI) immunotherapy.

Conclusion: We comprehensively evaluated m6A methylation regulatory genes in LGG and constructed a prognostic model based on m6A methylation, which may improve prognostic prediction for patients with LGG.

Keywords: LASSO, M6A RNA methylation, LGG, prognostic model, ICI

INTRODUCTION

Low-grade gliomas (II and III of the World Health Organization) are the most common primary malignant tumors in the brain and are mainly localized to the cerebral hemispheres (Ostrom et al., 2013). Neurosurgery, radiotherapy, and chemotherapy are common treatment strategies, but recurrence and drug resistance rates are high (Brat et al., 2015; Hayes et al., 2018). Some patients will quickly develop high-grade glioblastoma, resulting in a very poor prognosis (Zhang et al., 2016). Therefore, it is urgent to identify a sensitive and accurate biomarker to predict the prognosis of patients with LGG.

N6-methyladenosine (m6A) is the most studied RNA modification (Desrosiers et al., 1974), which plays an important role in the posttranscriptional regulation of RNA in eukaryotes (Zhao et al., 2017). The m6A modifications can occur on both mRNA and ncRNA (Alarcón et al., 2015). When m6a modifications occur in mRNA, they play multiple roles in mRNA processing and metabolism, including splicing, stability, nuclear export, and translation (Lee et al., 2020). In addition, m6A modification is also found in most ncRNAs, including miRNA, lncRNA, and circRNA, which participated in multiple roles in chromatin remodeling, transcription, posttranscriptional modifications, and signal transduction (Yang et al., 2020). M6A is dynamically and reversibly regulated by an m6A regulator, including methyltransferases (writers), demethylases (erasers), and binding proteins (readers) to add, remove, or recognize m6A-modified sites, respectively, thereby altering important biological processes accordingly (Guo et al., 2021). The m6A methyltransferases (writers) mediate the process of methylation modification of RNA, which mainly includes methyltransferase-like 3 (METTL3), methyltransferase-like 14 (METTL14), and Wilms' tumor 1-associating protein (WTAP) (Meyer and Jaffrey, 2017). The m6A demethylases (erasers) including obesity-associated protein (FTO) and alkB homolog 5 (ALKBH5), which mediate the process of decreasing m6A modifications of RNA, are the key to the reversibility of the m6A modification process (Meyer and Jaffrey, 2017). The m6A binding proteins (readers) are able to specifically recognize m6A-modified RNAs and participated in the regulation of RNA splicing, turnover, export, and translation and m6A readers mainly including YTH domain family YTHDF1-3, YTHDC1-2, insulin-like growth factor 2, mRNA-binding proteins IGF2BP1-3, heterogeneous nuclear ribonucleoprotein A2B1 (HNRNPA2B1), heterogeneous nuclear ribonucleoprotein C (HNRNPC), and embryonic lethal abnormal vision *Drosophila* like 1 (ELAVL1) (Meyer and Jaffrey, 2017). M6A, as a modification in RNA, plays an important role in bioprocesses such as self-renewal and differentiation of embryonic stem cells and hematopoietic stem cell, tissue development, circadian rhythm, heat shock or DNA damage response, and sex determination, although it does not change base pairing and coding (Huang et al., 2020). To investigate the specific mechanisms of m6a in cells or tissues, the expression of m6A regulatory genes, the global m6A abundance in RNA, and m6A modification site and gene need to be detected (Wang and Jia, 2020). LC-MS/MS, 2D-TLC, and dot blot can be used to detect the global m6A abundance in RNA. MeRIP-qRT-PCR, MeRIP-seq, SCARLET, SELECT, the m6A-sensitive deoxyribozyme method, the m6A-sensitive base-pairing method, and m6A-sensitive HRM analysis can be used to detect the m6A modification site and gene (Wang and Jia, 2020). Meanwhile, several databases, including RMBase (Xuan et al., 2018), MeT-DB (Liu H et al., 2018), CVm6A (Han et al., 2019), RNAmod (Liu and Gregory, 2019), SRAMP (Zhou et al., 2016), REPIC (Liu et al., 2020), and M6ADD (Zhou et al., 2021) were constructed to organize and integrate existing resources in order to better explore the mechanism of m6A.

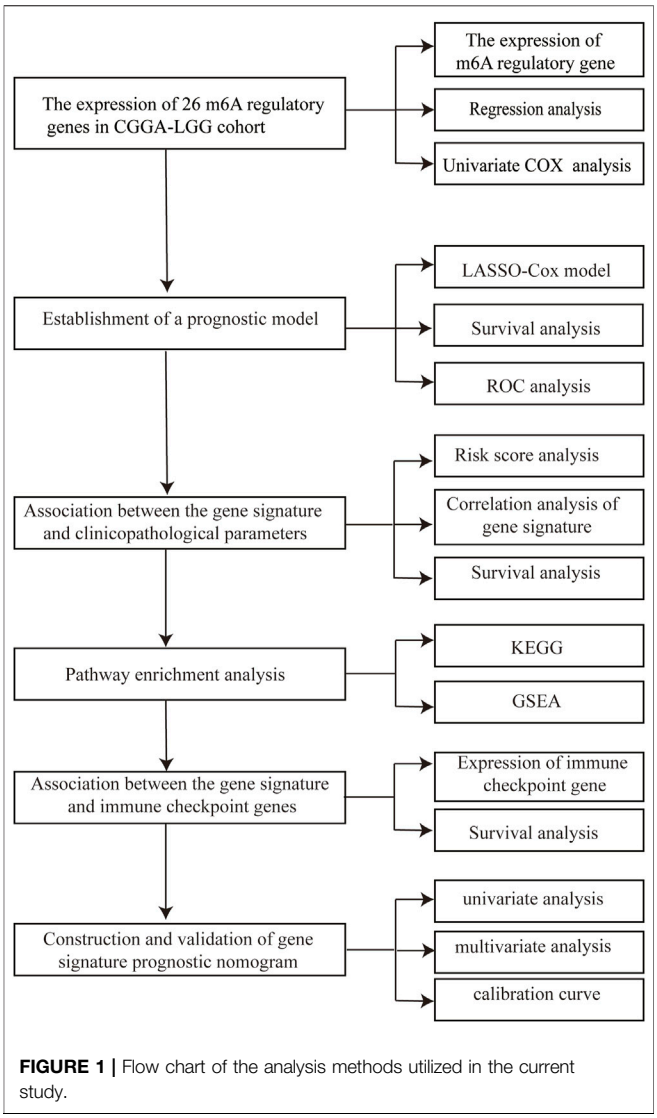


FIGURE 1 | Flow chart of the analysis methods utilized in the current study.

Numerous studies have shown that the global abundance of m6A and the expression levels of m6A regulatory genes, which are frequently dysregulated in various types of cancer, are critical for cancer development, progression, and metastasis, as well as drug resistance and cancer recurrence (Huang et al., 2020). Decreases or increases in global m6A abundance have recently been reported in several cancer types that may be associated with cancer progression and clinical outcomes (Huang et al., 2020).

TABLE 1 | Two low-grade glioma datasets used in this study.

Datasets	Platform	Component of samples	Use
TCGA LGG mRNA-seq	IlluminaHiseq_RNAseq	534 lower-grade glioma	Internal training and validation set for prognositc gene signature
CGGA mRNA-seq_325	Illumina Hiseq 2000	182 lower-grade glioma	External validation set for prognositc gene signature

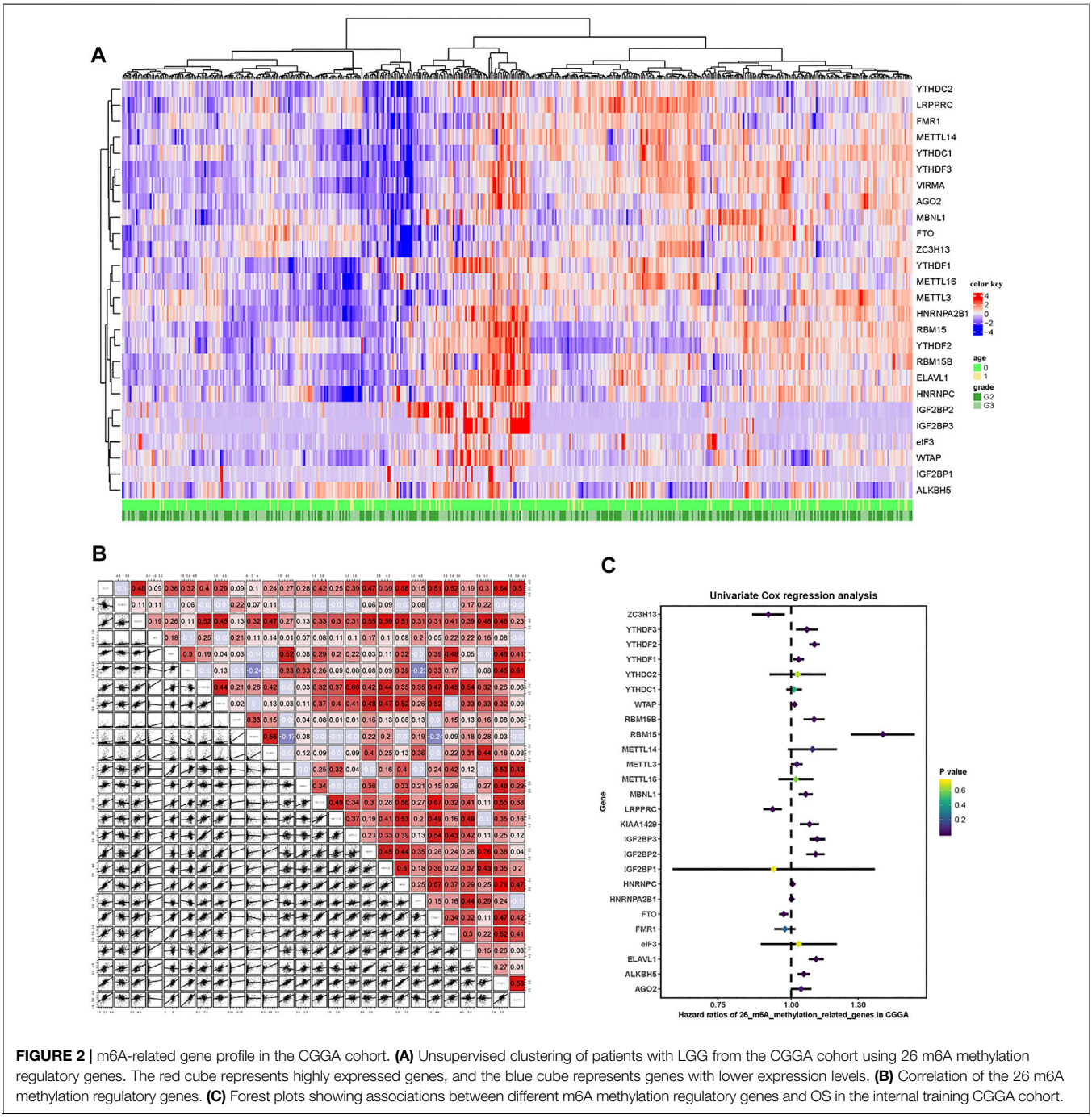


FIGURE 2 | m6A-related gene profile in the CGGA cohort. **(A)** Unsupervised clustering of patients with LGG from the CGGA cohort using 26 m6A methylation regulatory genes. The red cube represents highly expressed genes, and the blue cube represents genes with lower expression levels. **(B)** Correlation of the 26 m6A methylation regulatory genes. **(C)** Forest plots showing associations between different m6A methylation regulatory genes and OS in the internal training CGGA cohort.

The global m6A abundance was aberrantly upregulated in gastric cancer (Wang J et al., 2020), while aberrantly downregulated in bladder cancer (Gu et al., 2019). In addition, abnormalities in m6A regulatory genes can lead to a range of diseases, including cancer, neurological disorders, embryonic developmental delays, and obesity (Jiang et al., 2021). Studies have shown that M6A

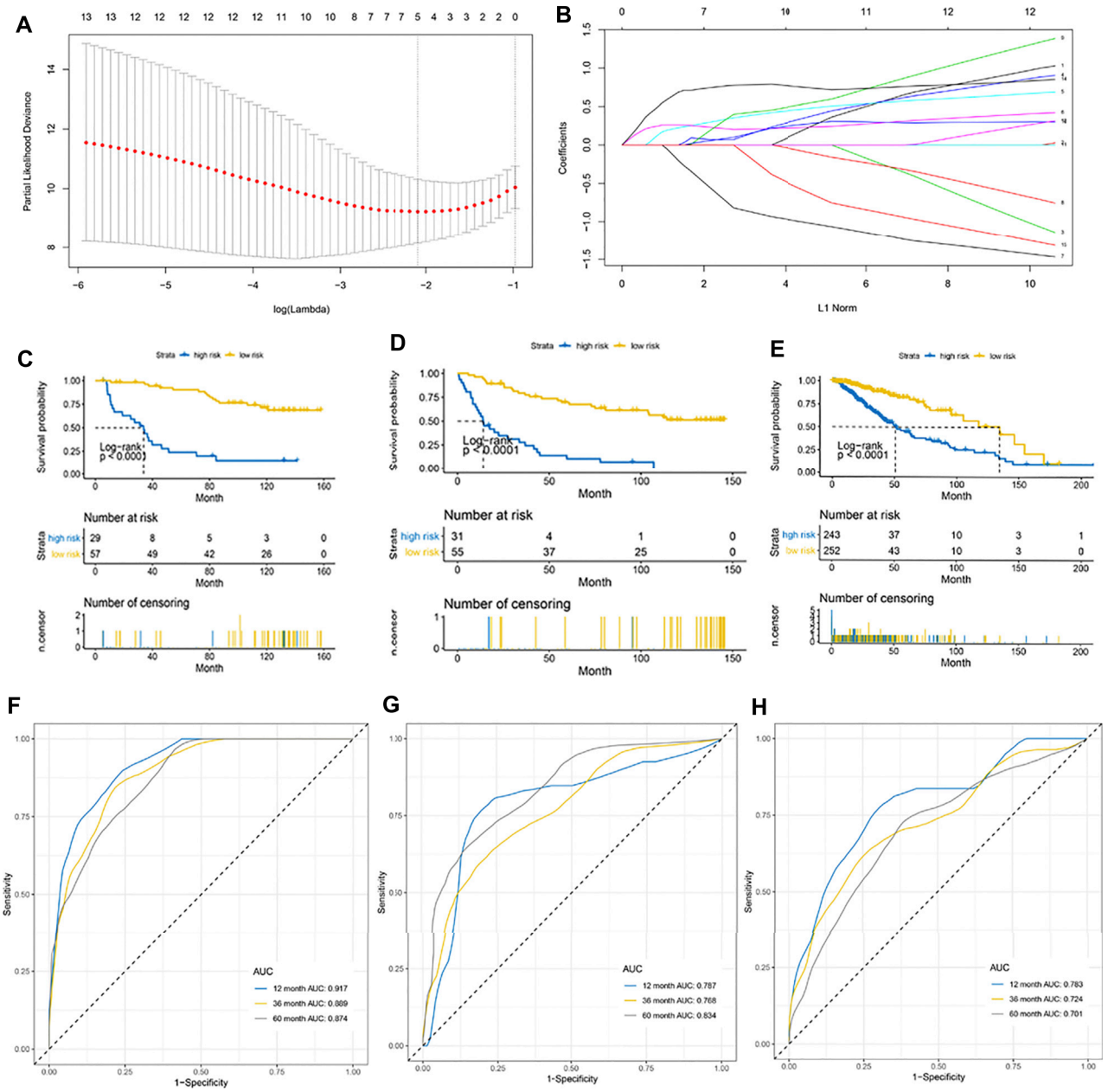


FIGURE 3 | Establishment of a 5-gene prognostic model. **(A)** LASSO coefficient profiles of the fractions of immune cells. The minimum lambda value was reached when the number of genes was 5. **(B)** Parameter selection for tuning by 10-fold cross validation in the LASSO model. **(C–E)** Kaplan–Meier curve for patients with high and low risk in the internal training CGGA cohort, internal validation CGGA cohort, and external validation TCGA cohort, respectively. **(F–H)** Risk score measured using time-dependent receiver operating characteristic (ROC) curves in the internal training cohort, internal validation CGGA cohort, and external validation TCGA cohort at 1, 3, and 5 years, respectively.

modifications plays important roles in various tumors and are involved in tumor proliferation, carcinogenesis, and migration (Meyer et al., 2012; Ke et al., 2015). For example, in glioblastoma multiforme (GBM), METTL3 suppresses the proliferation and self-renewal of glioblastoma stem cells by enhancing m6A modification of ADAM19 and decreasing its expression, which suppressed the progression of GMB (Visvanathan et al., 2018). METTL3 and FTO3 play an oncogenic role in acute myeloid leukemia (Vu et al., 2017). Some studies have indicated that genes that express methylation enzymes, including YTHDC2, RBM15B, METTL16, YTHDF3, IGF2BP3, RBM15, METTL14, ZC3H13, YTHDF1, YTHDF2, ALKBH5, HNRNPA2B1, ALKBH3, IGF2BP1, HNRNPC, YTHDC2, METTL3, WTAP, YTHDC1, IGF2BP2, and FTO (Ma et al., 2021), may be important in

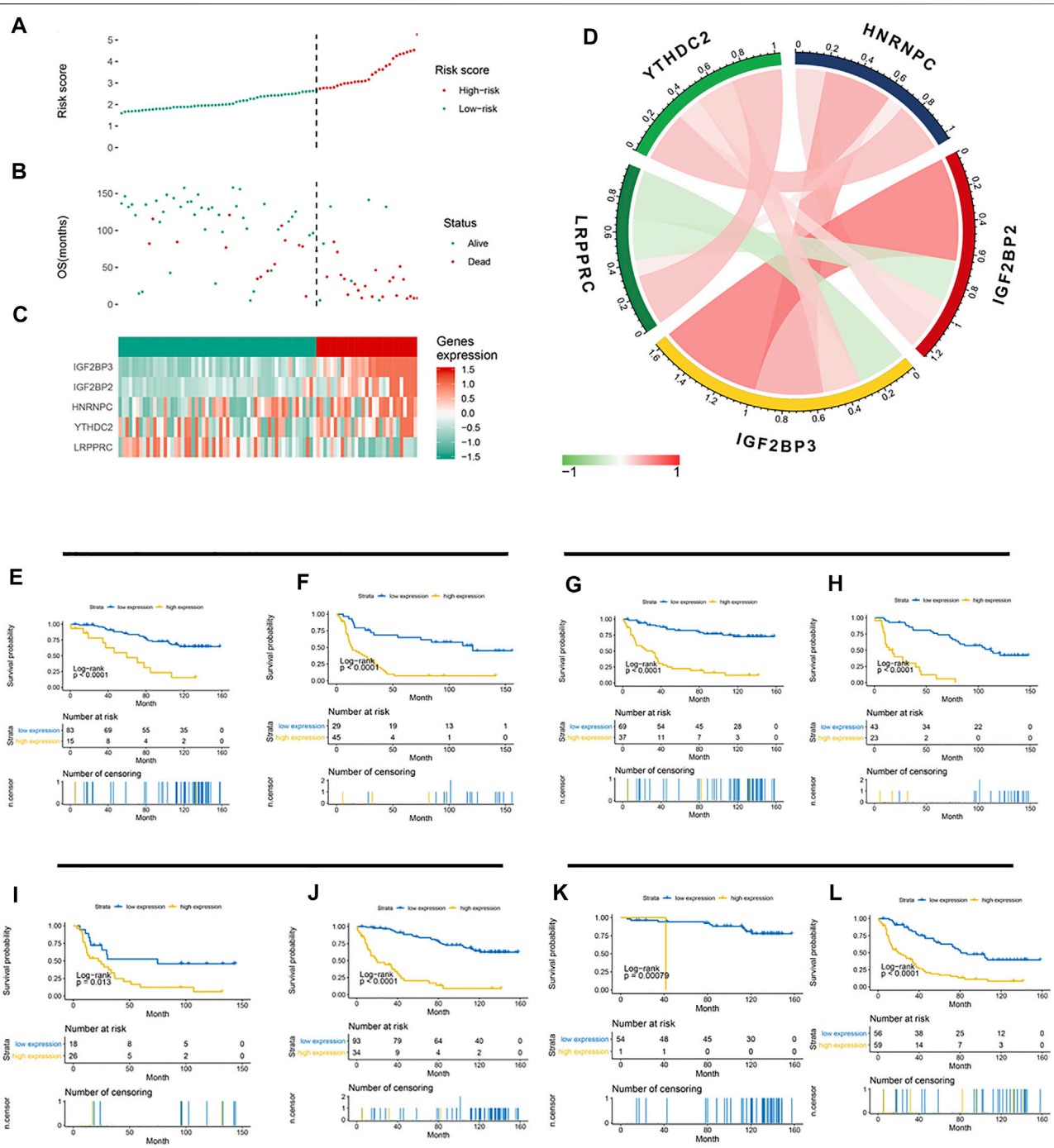


FIGURE 4 | Association between the 5-gene signature and clinicopathological parameters. **(A–C)** 5-gene signature-based risk score in the CGGA cohort. **(A)** Risk score plot based on the 5-gene signature. **(B)** Live/dead state corresponding to the risk score in the upper panel. **(C)** Z-score-transformed expression value of each gene in the 5-gene signature. **(D)** Correlation analysis of the 5 methylation regulatory genes in the signature. **(E–L)** Kaplan–Meier curve showed significant statistical differences in overall survival between the high- and the low-risk groups regardless of gender **(E,F)**, WHO grade **(G,H)**, IDH mutation status **(I,J)**, and co-mutation state of chromosomes 1p and 19q **(K,L)**.

LGG (Tu et al., 2020; Fang et al., 2021; Kowalski-Chauvel et al., 2020; Yarmishyn et al., 2020). However, the expression status of a single m6A regulatory gene is not sufficient to describe a patient's status and outcome.

In this study, we systematically characterized the expression levels of a group of m6A methylation regulator genes in patients with LGG and constructed a prognostic model for LGG. As a result, we established a framework to quantify prognosis using an

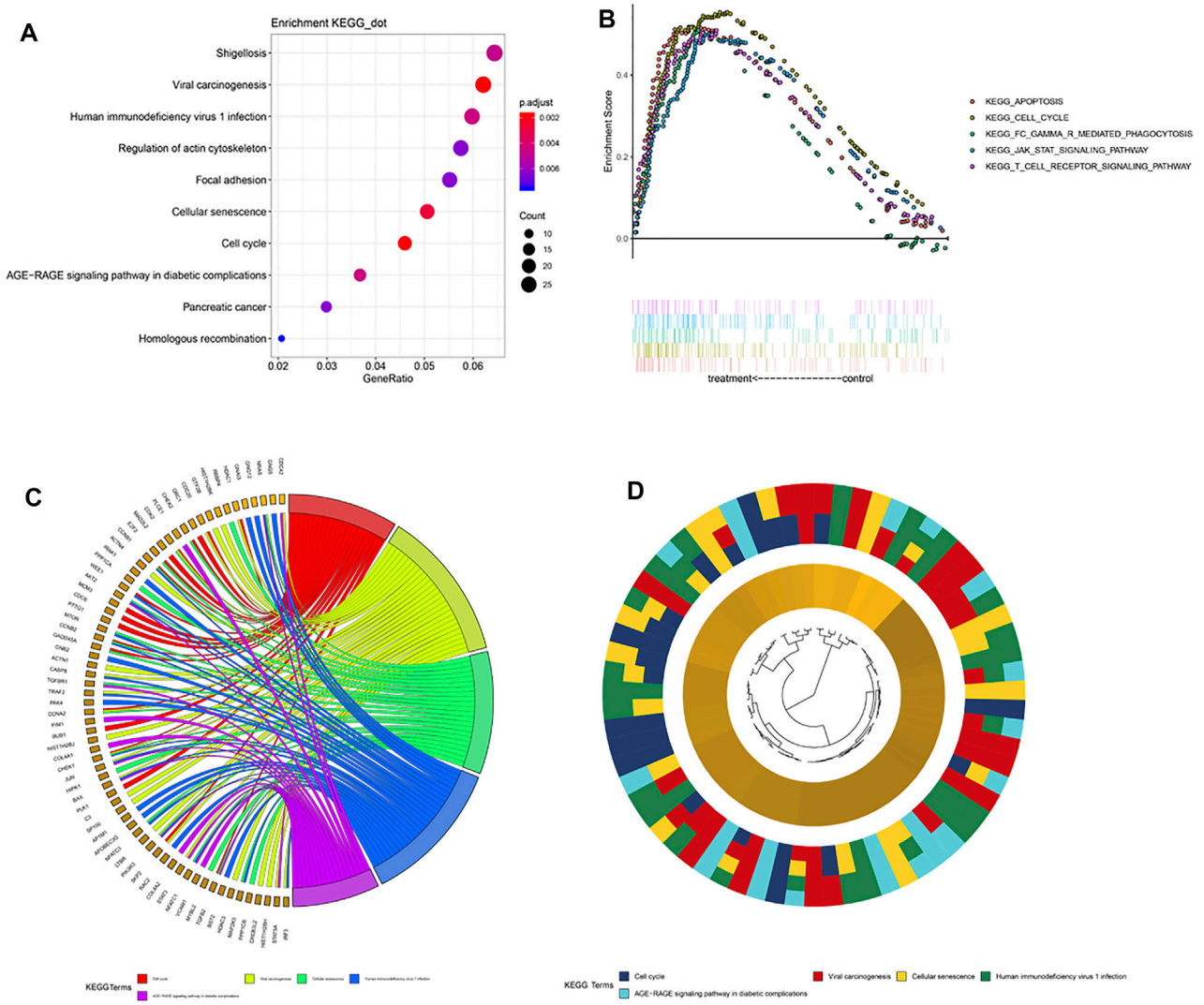


FIGURE 5 | Pathway enrichment analysis. **(A)** Spearman correlation for PRI top 1,000 genes was used for KEGG analysis. These genes were enriched in KEGG pathways "cell cycle" and "focal adhesion." **(B)** GSEA terms significantly enriched in the CGGA cohort. "KEGG_CELL_CYCLE," "KEGG_APOPTOSIS," "KEGG_JAK_STAT_SIGNALING_PATHWAY," and "KEGG_T_CELL_RECEPTOR_SINGALING_PATHWAY" were significantly enriched. **(C)** Hierarchical clustering of gene expression profiles for each KEGG pathway. **(D)** Chord plots show the relationship between genes and the KEGG pathway.

integrated analysis of the expression status of 5 m6A methylation regulatory genes, which resulted in a robust approach to prediction of overall survival (OS). The main flow of the article can be found in **Figure 1**. This approach, using a novel gene expression signature, is promising as a predictor OS of LGG.

MATERIALS AND METHODS

Data Preprocessing

Gene expression profiles and survival data for patients with LGG were downloaded from the CGGA (http://www.cgga.org.cn/) database. Overall, 495 samples from TCGA and 172 samples from the CGGA were analyzed (**Table 1**). Perl and R in R software were utilized to evaluate the samples.

Identifying the m6A RNA Methylation Regulatory Gene in LGG

The expression level of a total of 26 m6A RNA methylation regulatory genes in LGG samples and normal samples in TCGA and CGTA datasets were detected by the R package "limma" and visualized by the R package "pheatmap." The correlations between 26 m6A RNA methylation regulatory genes were detected by the R package "corrplot."

Establishment of the LASSO Cox Signature

A total of 172 patients from the CGGA database were randomly assigned in a 1:1 ratio to a training set and a validation set. The hazard ratio of OS of 26 m6A regulatory genes in the internal training set was calculated using the univariate Cox proportional hazard regression

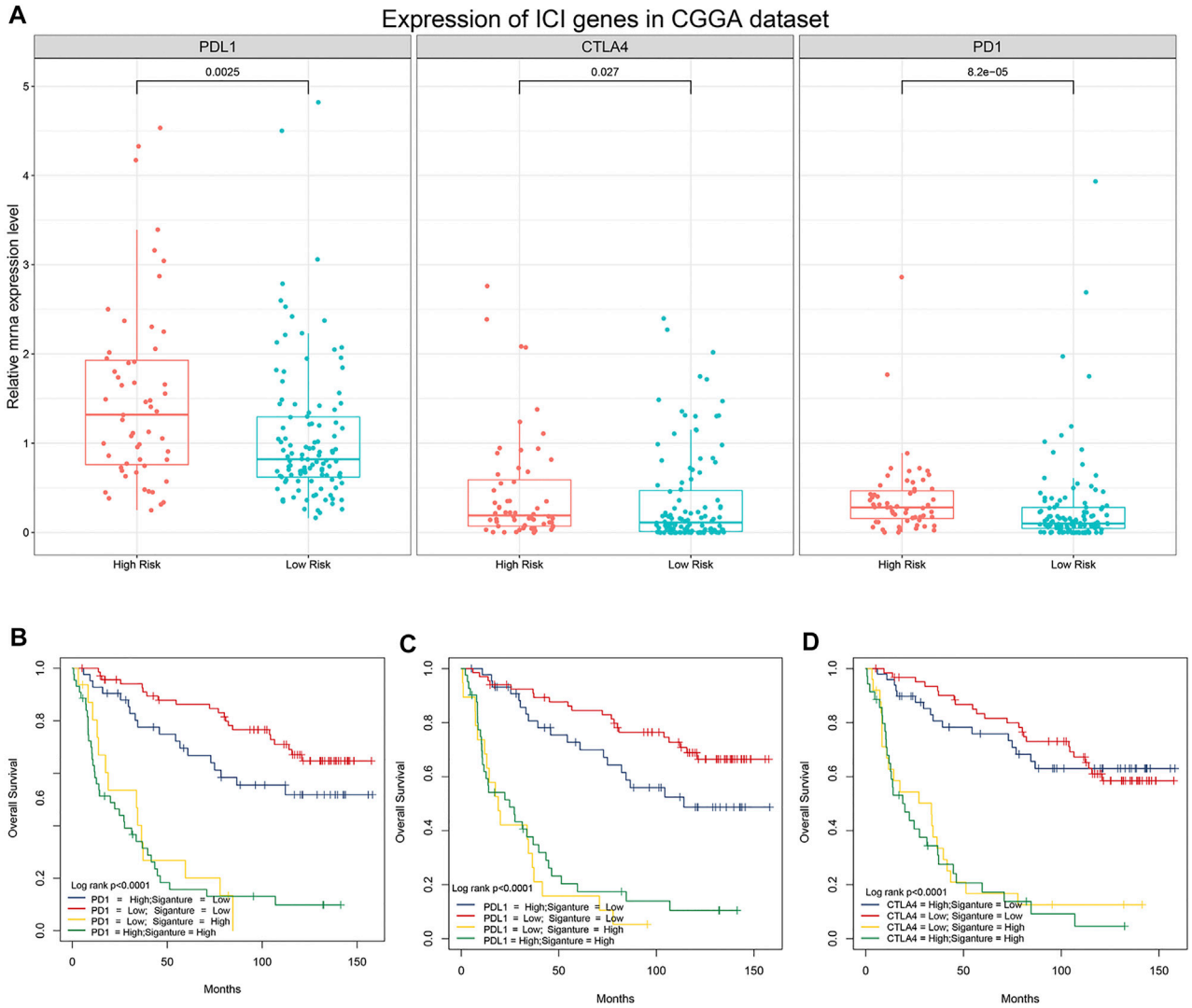


FIGURE 6 | Association between the 5-gene signature and immune checkpoint genes. **(A)** Comparison of the expression pattern of immune checkpoint genes (PD-1, PD-L1, and CTLA-4) between patients with different risk scores in the CGGA analysis. **(B)** Kaplan–Meier survival curves of overall survival among four patient groups stratified by the risk score and PD-1 **(B)**, PD-L1 **(C),** and CTLA-4 **(D)**.

model, and genes with $p < 0.05$ were considered statistically significant and included in subsequent analyses. Then, m6A-related genes that were identified as significant in the univariate analysis in the internal training set were included in the penalized Cox regression model with the least absolute shrinkage and selection operator (LASSO) Cox regression model for 10-fold cross validations to select the most significant genes. Finally, an m6A-related prognostic gene signature was constructed based on a linear combination of the regression coefficient derived from the LASSO Cox regression model coefficients multiplied by the mRNA expression level (Tibshirani,1997).

Risk Score Evaluation and Survival Analysis

The risk score was calculated as follows: $\sum Coefi \times Expi$, where Coefi is the coefficient of prognostic biomarker and Expi is the expression of the corresponding prognostic biomarker. According to the median value of the risk score, LGG patients in the CGGA cohort were divided into low-risk and high-risk groups. The overall Kaplan–Meier survival curves of low-risk and high-risk patients were analyzed by the R package "survival" and "survminer."

ROC Curve

The receiver operating characteristic curve (ROC) analysis was used to assess the accuracy of the diagnostic gene signature. The R software package "pROC" was used to generate an ROC curve (Robin et al., 2011). ROC area under curve (AUC) was calculated to evaluate the performance of each signature, and AUC>0.7 was considered as a diagnostic gene signature.

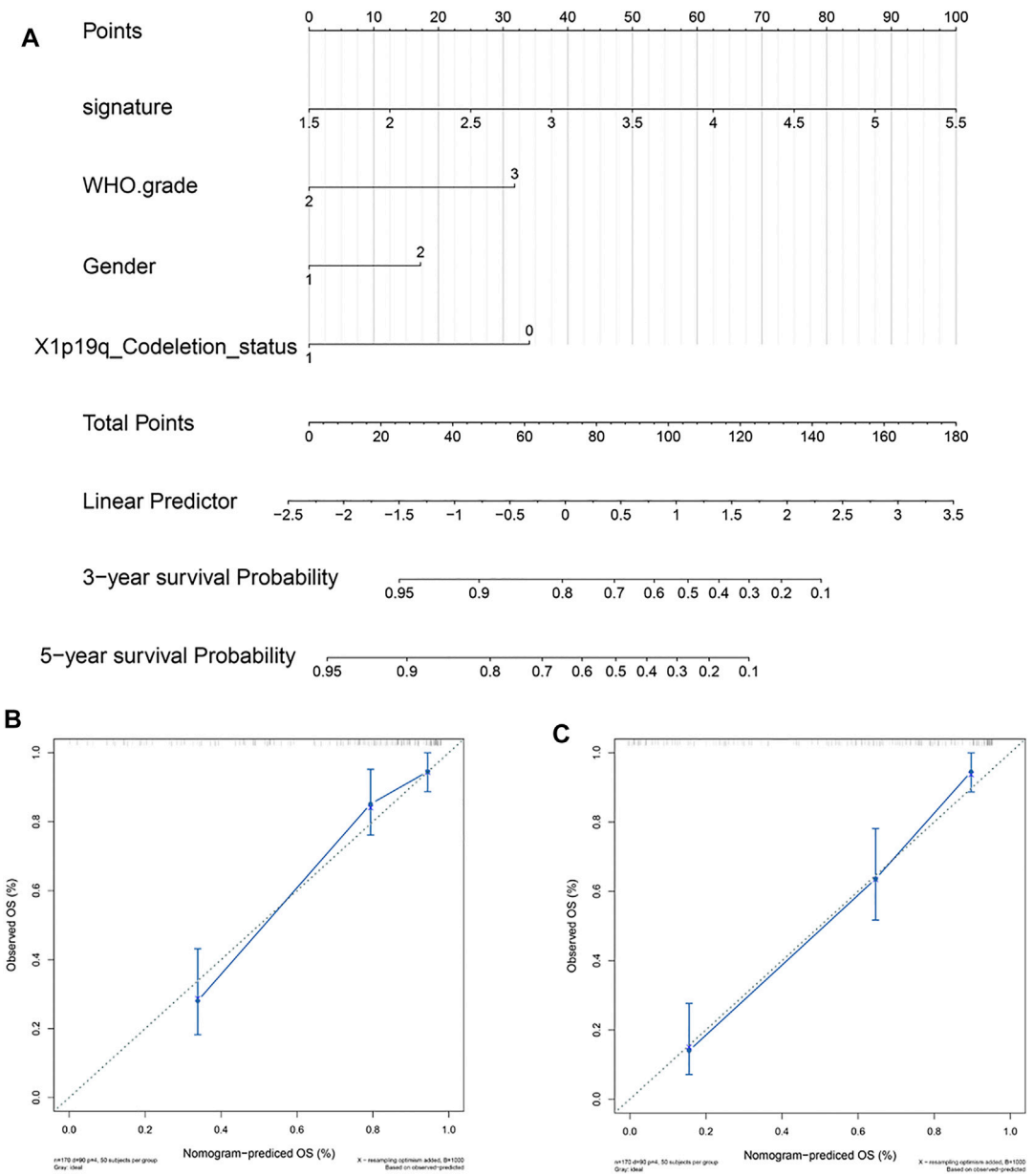


FIGURE 7 | Construction and validation of a 5-gene signature prognostic nomogram. **(A)** Prognostic nomogram based on the 5-gene signature, WHO grade, gender, and X1p19q codeletion status. **(B,C)** Calibration curve at 3 and 5 years.

Pathway Enrichment Analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis led to identification of enrichment pathways, which help to determine the biological pathways to which the identified genes belonged (Yu et al., 2012). The KEGG analysis is performed by the R package "clusterProfiler."

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis v2.2.2 (http://www.broadinstitute.org/gsea) was used to investigate the biological differences among patients with different expression patterns of the 5-gene signature. In addition, C2:CP KEGG gene sets from the Molecular Signatures database (MSigDB) were used as the reference gene sets.

Construction and Validation a Predictive Nomogram

The gene signatures were used to construct a nomogram by the "rms" R package, and the accuracy of the nomogram was evaluated by the calibration curve (1,000 bootstrap resamples).

Statistical Analysis

Continuous variables were analyzed using Student's t-tests or non-parametric tests. Categorical variables were analyzed using chi-squared tests or Fisher's exact tests. The R package copy was used for univariate and multivariate analyses. All data were analyzed using SPSS or R software (http://www.r-project.org/). $p < 0.05$ was considered statistically significant.

RESULTS

m6A Regulatory Gene Profile in the CGGA Cohort

A total of 26 m6A RNA methylation regulator genes were included in this study to evaluate the methylation status of tumor tissues. We constructed a profile containing 26 m6A regulatory genes in patients with LGG. The m6A regulatory genes and corresponding clinical parameters are shown in **Figure 2A**. As shown by the heatmap, the expression levels of the m6A regulatory genes differed among the patients. To better understand the relationships among m6A regulatory genes in LGG, we performed a regression analysis of the m6A regulatory genes (**Figure 2B**). The results showed that YTHDF2 was highly correlated with RBM15. In addition, there was a strong correlation between VIRMA and YTHDF3. These results indicated that these pairs of genes may be involved in the same biological functions. To determine which m6A regulatory genes had prognostic value, we conducted a univariate COX analysis. This analysis identified 16 of the 26 methylated genes as prognostic indicators (**Figure 2C**).

Establishment of a Prognostic Model

A total of 172 patients in the CGGA cohort were randomly divided in a 1:1 ratio into an internal training cohort and a validation cohort. The 16 m6A regulatory genes with prognostic value were included in the LASSO Cox model in the internal training CGGA cohort to screen for characteristic variables (**Figures 3A,B**). As shown in **Figure 2A**, the minimum lambda value was reached with inclusion of 5 genes. Therefore, we selected 5 genes to construct a prognostic model. Finally, we constructed the following formula for prognosis of OS in patients with LGG: (formula= (0.033 ×HNRNPC) + (0.237 × IGF2BP2) + (0.260 × IGF2BP3) − (0.271 × LRPPRC) + (0.713 × YTHDF2)). According to the cutoff value (2.03) obtained using the survminer package, the patients in the training cohort were divided into high- and low-PRI groups. Patients with high risk had more events and a worse prognosis (**Figure 3C**). Application of the cutoff value to the internal validation CCGA cohorts and external validation TCGA cohorts produced the same result as that observed in the training group (**Figures 3D,E**). To evaluate the accuracy of this 5-gene signature for determination of prognosis, we generated an ROC curve. In the internal training cohort, the accuracy of the 5-gene signature was investigated at 2, 3, and 5 years, which resulted in AUC values of 0.917, 0.889, and 0.874, respectively (**Figure 3F**). In the internal validation cohort, the AUC values were 0.787, 0.768, and 0.834, respectively (**Figure 3G**). In the external validation TCGA cohort, the AUC values at 2, 3, and 5 years were 0.783, 0.721, and 0.701, respectively (**Figure 3H**). In summary, the 5-gene profile was an excellent prognostic indicator for patients with LGG.

Association Between the 5-Gene Signature and Clinicopathological Parameters

Based on the LASSO Cox model, the prognostic risk score for each patient in the CGGA cohort was calculated according to the factor value and expression level. According to the median risk score, patients were divided into a high-risk group and a low-risk group (**Figures 4A–C**). The high-risk group had higher levels of IGF2BP2, IGF2BP3, HNRNPC, and YTHDC2 and lower levels of LRPPRC. More patients survived in the low-risk group than in the high-risk group. Further correlation analysis found a negative correlation between LRPPRC and IGF2BP2 and between LRPPRC and IGF2BP2 (**Figure 4D**). We further evaluated the stability of 5-gene signature in different groups. The results showed that the 5-gene signature was an excellent prognostic indicator regardless of gender (**Figures 4E,F**), grade (**Figures 4G,H**), IDH mutation status (**Figures 4I,J**), and the co-mutation state of chromosome 1p and 19q (**Figures 4K,L**).

Pathway Enrichment Analysis

To elucidate the differences in biological characteristics between the high- and low-risk groups using the 5-gene signature, we performed Spearman's correlation analysis and selected the first 1,000 genes. Then, the clusterProfiler package in R software was used to perform KEGG enrichment analysis. These genes are significantly enriched in the focal adhesion and cell cycle pathways (**Figure 5A**). Specifically enriched genes for each KEGG term are summarized in **Figures 5C,D**. Next, we divided the patients into high- and low-risk groups according to their risk scores for the GSEA analysis, which showed enrichment of the 5-gene signature in cell cycle pathways (**Figure 5B**), which indicated that the cell cycle pathway may be a critical factor in poor prognosis in patients with LGG. In addition, expression levels of genes in the T-cell receptor signaling pathway were abnormally high in patients with LGG, which indicated that immune status may have differed with level of risk. Therefore, we further explored the expression levels of immune checkpoint (IC) genes in the different patient risk groups, which may provide additional information regarding personalization of treatment.

Association Between the 5-Gene Signature and Immune Checkpoint Genes

Previous studies have shown that the expression levels of immune checkpoint (IC) genes are associated with immunotherapy efficacy. We compared the expression patterns of IC genes (PD-1, PD-L1, and CTLA-4) in different

risk groups of patients with LGG. As shown in **Figure 6A**, patients in the high-risk group had higher expression levels of IC genes. We further evaluated whether our research model for IC genes was generalizable to patients with similar expression levels of IC genes. As shown in **Figure 6B**, the overall survival of patients with low risk scores and high PD-1 expression was significantly better than that of patients with high risk scores and high PD-1 expression. Furthermore, the survival of patients with low risk scores and low PD-1 expression was longer than that of patients with high risk scores and low PD-1 expression. Similar results were observed for PD-L1 and CTLA-4 (**Figures 6C,D**). Stratification of patients with low risk scores according to IC gene expression showed that IC gene expression was significantly associated with survival of patients with low risk scores. However, there was no differences in survival of patients with high risk scores when stratified based on IC gene expression. In addition, patients with low risk scores and low IC gene expression tended to have much higher survival rates than patients in the other three groups. These results suggested that our 5-gene model may be a predictor of ICI immunotherapy efficacy.

Construction and Validation of a 5-Gene Signature Prognostic Nomogram

To provide patients with more accurate prognostic predictions, we incorporated the 5-gene signature and clinical parameters that had prognostic value in the univariate analysis and performed a multivariate analysis in the CGGA cohort. The results showed that the 5-gene signature, WHO grade, gender, and X1p19q codeletion status were stable predictors (**Supplement Tables S1,2**). These results were used to construct a nomogram to predict the prognosis of patients (**Figure 7A**). The calibration curve shows that the nomogram had stable predictive values at 3 and 5 years (**Figures 7B,C**).

DISCUSSION

Low-grade glioma (LGG) is a common invasive brain tumor in adults, and it includes World Health Organization (WHO) grade II and III diffuse gliomas (Ostrom et al., 2013). Although some progress has been made in the treatment of LGG with advances in neurosurgery, chemotherapy, and radiotherapy, a considerable number of patients experience recurrent and malignant glioblastoma multiforme (Chaichana et al., 2010; Okita et al., 2015; Deng et al., 2019; Fukuya et al., 2019; Chen et al., 2020; Mathur et al., 2020), resulting in decreased quality of life and shortened lifespan. The heterogeneity of prognosis for patients with LGG highlights the need to develop effective biomarkers for early stratification and preventive treatment of high-risk patients with poor prognoses.

M6A methylation, as the most widespread internal modification of mRNA, has been shown to play an important role in many cell types (Desrosiers et al., 1974; Dominissini et al., 2012; Fu et al., 2014; Wang et al., 2014; Ma et al., 2019; Sun et al., 2019). Many studies have shown that m6A regulatory genes can be used as markers to predict the prognosis of many kinds of cancers. (Liu J et al., 2018; Chen et al., 2019; Zhao and Cui 2019; Qu et al., 2020; Wang Q et al., 2020). Many previous studies have evaluated individual methylation regulatory genes, which may not be an accurate reflection of overall regulation of methylation. In this study, we evaluated multiple methylation regulatory genes in patients with LGG and selected a group of methylation regulatory genes that had prognostic value for patients with LGG through univariate Cox analysis and LASSO Cox modeling. Finally, a 5-gene signature was constructed (HNRNPC, IGF2BP2, IGF2BP3, LRPPRC, and YTHDF2) with good prognostic value and consistency between the internal validation and the external validation TCGA cohorts. To determine the pathways most closely associated with poor prognosis of patients with LGG, we performed a correlation analysis and selected the 1,000 genes most related to risk score. Furthermore, the KEGG enrichment analysis showed that differences in survival may have been associated with the cell cycle and focal adhesion pathways, which provides novel potential targets for treatment of LGG.

Advances in immune checkpoint inhibitor therapies have resulted in effective treatment of many cancers (Havel et al., 2019; Doroshow et al., 2019; Zhao et al., 2019; Gedeon et al., 2020). These advances have highlighted the importance of screening patients to determine whether they are good candidates for ICI treatment. The expression levels of independent immune checkpoint genes such as PD-L1 cannot be used as an independent predictor of ICI response (Gridelli et al., 2017; Lupo et al., 2018). By comparing the survival distribution of patients following stratification based on our 5-gene signature and immune checkpoint gene expression levels, we were able to show that our 5-gene profile correlated well with IC gene expression and risk, which indicated that our 5-gene signature can be used as a biomarker for knowing whether a patient is a good candidate for immunotherapy.

In summary, our study showed that m6A methylation regulatory genes could be used to classify patients with LGG into high- or low-risk subgroups exhibiting significantly different OS. Furthermore, this risk score may also be a marker for predicting the efficacy of ICI immunotherapy.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ZB designed research and analyzed the data. XW provided clinical relevant opinions and helped in writing the discussion part. ZZ supervised all works. All authors wrote, read, and approved the final manuscript.

FUNDING

This research was supported by the Natural Science Foundation of Liaoning Province, China (2019-ZD-0772), and the National Natural Science Foundation of China (Nos. 81471809 and 81971639).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.655169/full#supplementary-material

REFERENCES

Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015). N6-methyladenosine marks Primary microRNAs for Processing. *Nature* 519 (7544), 482–485. doi:10.1038/nature14281

Brat, D. J., Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* 372 (26), 2481–2498. doi:10.1056/NEJMoa1402121

Chaichana, K. L., McGirt, M. J., Laterra, J., Olivi, A., and Quiñones-Hinojosa, A. (2010). Recurrence and Malignant Degeneration after Resection of Adult Hemispheric Low-Grade Gliomas. *Jns* 112 (1), 10–17. doi:10.3171/2008.10.jns08608

Chen, M., Nie, Z.-y., Wen, X.-h., Gao, Y.-h., Cao, H., and Zhang, S.-f. (2019). m6A RNA Methylation Regulators Can Contribute to Malignant Progression and Impact the Prognosis of Bladder Cancer. *Biosci. Rep.* 39 (12). doi:10.1042/bsr20192892

Chen, P.-Y., Li, X.-D., Ma, W.-N., Li, H., Li, M.-M., Yang, X.-Y., et al. (2020). Comprehensive Transcriptomic Analysis and Experimental Validation Identify lncRNA HOXA-AS2/miR-184/COL6A2 as the Critical ceRNA Regulation Involved in Low-Grade Glioma Recurrence. *Ott* 13, 4999–5016. doi:10.2147/ott.S245896

Deng, T., Gong, Y.-Z., Wang, X.-K., Liao, X.-W., Huang, K.-T., Zhu, G.-Z., et al. (2019). Use of Genome-Scale Integrated Analysis to Identify Key Genes and Potential Molecular Mechanisms in Recurrence of Lower-Grade Brain Glioma. *Med. Sci. Monit.* 25, 3716–3727. doi:10.12659/msm.913602

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

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

Doroshow, D. B., Sanmamed, M. F., Hastings, K., Politi, K., Rimm, D. L., Chen, L., et al. (2019). Immunotherapy in Non-small Cell Lung Cancer: Facts and Hopes. *Clin. Cancer Res.* 25 (15), 4592–4602. doi:10.1158/1078-0432.Ccr-18-1538

Fang, R., Chen, X., Zhang, S., Shi, H., Ye, Y., Shi, H., et al. (2021). EGFR/SRC/ERK-stabilized YTHDF2 Promotes Cholesterol Dysregulation and Invasive Growth of Glioblastoma. *Nat. Commun.* 12 (1), 177. doi:10.1038/s41467-020-20379-7

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene Expression Regulation Mediated through Reversible m6A RNA Methylation. *Nat. Rev. Genet.* 15 (5), 293–306. doi:10.1038/nrg3724

Fukuya, Y., Ikuta, S., Maruyama, T., Nitta, M., Saito, T., Tsuzuki, S., et al. (2019). Tumor Recurrence Patterns after Surgical Resection of Intracranial Low-Grade Gliomas. *J. Neurooncol.* 144 (3), 519–528. doi:10.1007/s11060-019-03250-8

Gedeon, P. C., Champion, C. D., Rhodin, K. E., Woroniecka, K., Kemeny, H. R., Bramall, A. N., et al. (2020). Checkpoint Inhibitor Immunotherapy for Glioblastoma: Current Progress, Challenges and Future Outlook. *Expert Rev. Clin. Pharmacol.* 13 (10), 1147–1158. doi:10.1080/17512433.2020.1817737

Gridelli, C., Ardizzoni, A., Barberis, M., Cappuzzo, F., Casaluce, F., Danesi, R., et al. (2017). Predictive Biomarkers of Immunotherapy for Non-small Cell Lung Cancer: Results from an Experts Panel Meeting of the Italian Association of Thoracic Oncology. *Transl. Lung Cancer Res.* 6 (3), 373–386. doi:10.21037/tlcr.2017.05.09

Gu, C., Wang, Z., Zhou, N., Li, G., Kou, Y., Luo, Y., et al. (2019). Mettl14 Inhibits Bladder TIC Self-Renewal and Bladder Tumorigenesis through N6-Methyladenosine of Notch1. *Mol. Cancer* 18 (1), 168. doi:10.1186/s12943-019-1084-1

Guo, J., Zheng, J., Zhang, H., and Tong, J. (2021). RNA m6A Methylation Regulators in Ovarian Cancer. *Cancer Cel Int* 21 (1), 1–13. doi:10.1186/s12935-021-02318-8

Han, Y., Feng, J., Xia, L., Dong, X., Zhang, X., Zhang, S., et al. (2019). CVm6A: A Visualization and Exploration Database for m6As in Cell Lines. *Cells* 8 (2), 168. doi:10.3390/cells8020168

Havel, J. J., Chowell, D., and Chan, T. A. (2019). The Evolving Landscape of Biomarkers for Checkpoint Inhibitor Immunotherapy. *Nat. Rev. Cancer* 19 (3), 133–150. doi:10.1038/s41568-019-0116-x

Hayes, J., Yu, Y., Jalbert, L. E., Mazor, T., Jones, L. E., Wood, M. D., et al. (2018). Genomic Analysis of the Origins and Evolution of Multicentric Diffuse Lower-Grade Gliomas. *Neurons Oncol.* 20 (5), 632–641. doi:10.1093/neuonc/nox205

Huang, H., Weng, H., and Chen, J. (2020). m6A Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. *Cancer Cell* 37 (3), 270–288. doi:10.1016/j.ccell.2020.02.004

Jiang, X., LiuLiu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., et al. (2021). The Role of m6A Modification in the Biological Functions and Diseases. *Sig Transduct Target. Ther.* 6 (1), 1–16. doi:10.1038/s41392-020-00450-x

Ke, S., Alemu, E. A., Mertens, C., Gantman, E. C., Fak, J. J., Mele, A., et al. (2015). A Majority of m6A Residues Are in the Last Exons, Allowing the Potential for 3′ UTR Regulation. *Genes Dev.* 29 (19), 2037–2053. doi:10.1101/gad.269415.115

Kowalski-Chauvel, A., Lacore, M. G., Arnauduc, F., Delmas, C., Toulas, C., Cohen-Jonathan-Moyal, E., et al. (2020). The m6A RNA Demethylase ALKBH5 Promotes Radioresistance and Invasion Capability of Glioma Stem Cells. *Cancers (Basel)* 13 (1). doi:10.3390/cancers13010040

Lee, Y., Choe, J., Park, O. H., and Kim, Y. K. (2020). Molecular Mechanisms Driving mRNA Degradation by m6A Modification. *Trends Genet.* 36 (3), 177–188. doi:10.1016/j.tig.2019.12.007

Liu, H., Wang, H., Wei, Z., Zhang, S., Hua, G., Zhang, S.-W., et al. (2018). MeT-DB V2.0: Elucidating Context-specific Functions of N6-Methyl-Adenosine Methyltranscriptome. *Nucleic Acids Re* 46 (D1), D281–D287. doi:10.1093/nar/gkx1080

Liu, J., Eckert, M. A., Harada, B. T., Liu, S.-M., Lu, Z., Yu, K., et al. (2018). m6A mRNA Methylation Regulates AKT Activity to Promote the Proliferation and Tumorigenicity of Endometrial cancerA mRNA Methylation Regulates AKT Activity to Promote the Proliferation and Tumorigenicity of Endometrial Cancer. *Nat. Cel Biol* 20 (9), 1074–1083. doi:10.1038/s41556-018-0174-4

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

Liu, S., Zhu, A., He, C., and Chen, M. (2020). REPIC: a Database for Exploring the N6-Methyladenosine Methylome. *Genome Biol.* 21 (1), 1–13. doi:10.1186/s13059-020-02012-4

Lupo, A., Alifano, M., Wislez, M., Boulle, G., Velut, Y., Biton, J., et al. (2018). Biomarqueurs prédictifs de l'immunothérapie anti-PD1/PD-L1 dans le cancer broncho-pulmonaire non à petites cellules. *Revue de Pneumologie Clinique* 74 (5), 339–350. doi:10.1016/j.pneumo.2018.09.010

Ma, L., Liu, Z., Zhang, H., Ning, W., Qu, Y., Yu, C., et al. (2021). The Landscape and Prognostic Value of m6A Methylation-Related Genes in Low Grade Glioma. *Arch. Microbiol. Immunol.* 05 (2): 214–231. doi:10.26502/ami.93650059

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

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 (11), 1580–1590. doi:10.1093/neuonc/noaa059

Meyer, K. D., and Jaffrey, S. R. (2017). Rethinking m6A Readers, Writers, and Erasers. *Annu. Rev. Cel Dev. Biol.* 33, 319–342. doi:10.1146/annurev-cellbio-100616-060758

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

Okita, Y., Narita, Y., Miyahara, R., Miyakita, Y., Ohno, M., and Shibui, S. (2015). Health-related Quality of Life in Long-Term Survivors with Grade II Gliomas: the Contribution of Disease Recurrence and Karnofsky Performance Status. *Jpn. J. Clin. Oncol.* 45 (10), 906–913. doi:10.1093/jjco/hyv115

Ostrom, Q. T., Gittleman, H., Farah, P., Ondracek, A., Chen, Y., Wolinsky, Y., et al. (2013). CBTRUS Statistical Report: Primary Brain and central Nervous System Tumors Diagnosed in the United States in 2006-2010. *Neuro-Oncology* 15 (Suppl. 2Suppl 2), ii1–ii56. doi:10.1093/neuonc/not151

Qu, N., Qin, S., Zhang, X., Bo, X., Liu, Z., Tan, C., et al. (2020). Multiple m6A RNA Methylation Modulators Promote the Malignant Progression of Hepatocellular Carcinoma and Affect its Clinical Prognosis. *BMC Cancer* 20 (1), 165. doi:10.1186/s12885-020-6638-5

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77

Sun, T., Wu, R., and Ming, L. (2019). The Role of m6A RNA Methylation in Cancer. *Biomed. Pharmacother.* 112, 108613. doi:10.1016/j.biopha.2019.108613

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Statist. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

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

Visvanathan, A., Patil, V., Arora, A., Hegde, A. S., Arivazhagan, A., Santosh, V., et al. (2018). Essential Role of METTL3-Mediated m6A Modification in Glioma Stem-like Cells Maintenance and Radioresistance. *Oncogene* 37 (4), 522–533. doi:10.1038/onc.2017.351

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

Wang, J., Zhang, C., He, W., and Gou, X. (2020). Effect of m6A RNA Methylation Regulators on Malignant Progression and Prognosis in Renal Clear Cell Carcinoma. *Front. Oncol.* 10, 3. doi:10.3389/fonc.2020.00003

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, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent Regulation of Messenger RNA Stability. *Nature* 505 (7481), 117–120. doi:10.1038/nature12730

Wang, Y., and Jia, G. (2020). Detection Methods of Epitranscriptomic Mark N6-Methyladenosine. *Essays Biochem.* 64 (6), 967–979. doi:10.1042/ebc20200039

Xuan, J.-J., Sun, W.-J., Lin, P.-H., Zhou, K.-R., Liu, S., Zheng, L.-L., et al. (2018). RMBase v2.0: Deciphering the Map of RNA Modifications from Epitranscriptome Sequencing Data. *Nucleic Acids Res.* 46 (D1), D327–D334. doi:10.1093/nar/gkx934

Yang, X., Hu, X., Liu, J., Wang, R., Zhang, C., Han, F., et al. (2020). N6-methyladenine Modification in Noncoding RNAs and its Function in Cancer. *Biomark Res.* 8 (1), 1–12. doi:10.1186/s40364-020-00244-x

Yarmishyn, A. A., Yang, Y.-P., Lu, K.-H., Chen, Y.-C., Chien, Y., Chou, S.-J., et al. (2020). Musashi-1 Promotes Cancer Stem Cell Properties of Glioblastoma Cells via Upregulation of YTHDF1. *Cancer Cel Int* 20 (1), 597. doi:10.1186/s12935-020-01696-9

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.-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

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

Zhao, J., Chen, A. X., Gartrell, R. D., Silverman, A. M., Aparicio, L., Chu, T., et al. (2019). Immune and Genomic Correlates of Response to Anti-PD-1 Immunotherapy in Glioblastoma. *Nat. Med.* 25 (3), 462–469. doi:10.1038/s41591-019-0349-y

Zhao, X., and Cui, L. (2019). Development and Validation of a m6A RNA Methylation Regulators-Based Signature for Predicting the Prognosis of Head and Neck Squamous Cell Carcinoma. *Am. J. Cancer Res.* 9 (10), 2156–2169.

Zhou, D., Wang, H., Bi, F., Xing, J., Gu, Y., Wang, C., et al. (2021). M6ADD: a Comprehensive Database of m6A Modifications in Diseases. *RNA Biol.* 18, 2354–2362. doi:10.1080/15476286.2021.1913302

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

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

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

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fgene.2021.699503

Identification and Validation of a Pyroptosis-Related Prognostic Model for Gastric Cancer

Chaowei Liang[1], *Jiaxin Fan*[1], *Chaojie Liang*[2]* *and Jiansheng Guo*[1]*

[1]Department of Gastrointestinal Surgery, First Hospital of Shanxi Medical University, Taiyuan, China, [2]Department of Hepatobiliary and Pancreatic Surgery, First Hospital of Shanxi Medical University, Taiyuan, China

Pyroptosis is an inflammatory form of programmed cell death triggered by caspase-1/4/5/11 that plays an important role in the occurrence and development of gastric cancer (GC). We investigated the prognostic value of pyroptosis-related genes in GC. The "LIMMA" R package and univariate Cox analysis were used to find pyroptosis-related genes with differential expression and prognostic value in the TCGA cohort and the identified genes were analyzed for GO enrichment and KEGG pathways. The selected genes were then included in a multivariate Cox proportional hazard regression analysis, and a ten genes prognostic model (BIRC2, CD274, IRGM, ANXA2, GBP5, TXNIP, POP1, GBP1, DHX9, and TLR2) was established. To evaluate the predictive value of the risk score on prognosis, patients were divided into high-risk and low-risk groups according to the median risk score, and survival analysis was carried out. Compared with the low-risk group, the OS of GC patients in the high-risk group was significantly worse. Additionally, these results were verified in the GSE84437 and GSE66229 datasets. Finally, through the combination of prognostic gene characteristics and clinicopathological features, a nomogram was established to predict individual survival probability. The results show that the genetic risk characteristics related to clinical features can be used as independent prognostic indicators for patients with GC. In summary, the pyroptosis-related risk signals proposed in this study can potentially predict the prognosis of patients with GC. In addition, we also found significant infiltration of dendritic cells, macrophages, and neutrophils in tissues of high-risk patients.

Keywords: gastric cancer, pyroptosis, prognostic signature, nomogram, the cancer genome atlas

OPEN ACCESS

Edited by:
Liuyang Wang,
Duke University, United States

Reviewed by:
Emil Bulatov,
Kazan Federal University, Russia
Raghavendran Partha,
GRAIL Inc., United States

***Correspondence:**
Chaojie Liang
liangchaojie8@126.com
Jiansheng Guo
sdyygjs2017@126.com

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

Received: *23 April 2021*
Accepted: *11 October 2021*
Published: *25 February 2022*

Citation:
Liang C, Fan J, Liang C and Guo J (2022) Identification and Validation of a Pyroptosis-Related Prognostic Model for Gastric Cancer. Front. Genet. 12:699503. doi: 10.3389/fgene.2021.699503

INTRODUCTION

Gastric cancer (GC) is the fifth most common cancer and the third most common cause of cancer death in the world, with more than one million new gastric cancer patients each year, of which about 73% die. The median survival time of advanced gastric cancer is less than 12 months (Smyth et al., 2020). Gastric cancer is the third most common cancer in Asia, after breast and lung cancer (Sung et al., 2021). The occurrence of gastric cancer is a multi-factor, multi-step process, and a variety of mechanisms affect its occurrence and development (Chia and Tan, 2016), making the prognosis evaluation full of challenges. Therefore, it is very important to find an effective prediction model.

Pyroptosis is an inflammatory form of programmed cell death triggered by caspase-1/4/5/11 (Bergsbaken et al., 2009). Caspase-1 and-11 trigger pyroptosis through the cleavage of Gasdermin D (Kayagaki et al., 2015; Shi et al., 2015). The characteristics of pyroptosis are pore formation, cell

swelling, plasma membrane rupture, and release of intracellular contents (Fang et al., 2020). Pyroptosis may affect all stages of cancer development and has therefore become a new topic in cancer research (Ruan et al., 2020). Recent studies have shown that polyphyllin VI induces the transition of A549 and H1299 cells from apoptosis to pyroptosis by activating caspase-1, which induces the ROS/NF-κB/NLRP3/GSDMD signal axis, resulting in cell death (Teng et al., 2020). GSDME can convert TNF-α, chemotherapy, or caspase-3-mediated apoptosis into pyroptosis (Wang et al., 2017; Yu et al., 2019; Wang et al., 2017; Yu et al., 2019). Caspase-3-dependent apoptosis and pyroptosis can promote the clearance of stressed, injured, transformed, or infected cells, which plays a very important role in the development and treatment of tumors (Jiang et al., 2020). PD-L1 converts tumor cell apoptosis induced by TNF-α into pyroptosis, resulting in tumor necrosis (Hou et al., 2020a). Mutant BRAF and MEK inhibitors regulate the tumor immune microenvironment through pyrogenesis (Erkes et al., 2020). Thanks to the existing research results, it is known that pyroptosis is involved in tumorigenesis and anti-tumoral processes, but studies on its specific function in GC are missing. It is not clear whether pyroptosis-related genes are linked to the prognosis of patients with gastric cancer.

We systematically analyzed the differentially expressed genes related to pyroptosis in gastric cancer in TCGA database samples. Univariate analysis was used to screen the genes related to prognosis, and then the resulting genes were analyzed by multivariate Cox proportional hazard regression analysis to establish a prognostic model. A prognostic nomogram containing prognostic gene markers was established to predict overall survival. Since previous studies showed a significant correlation between pyroptosis and the immune microenvironment in GC (Shao et al., 2021), we also studied the infiltration of immune cells in tumor cells.

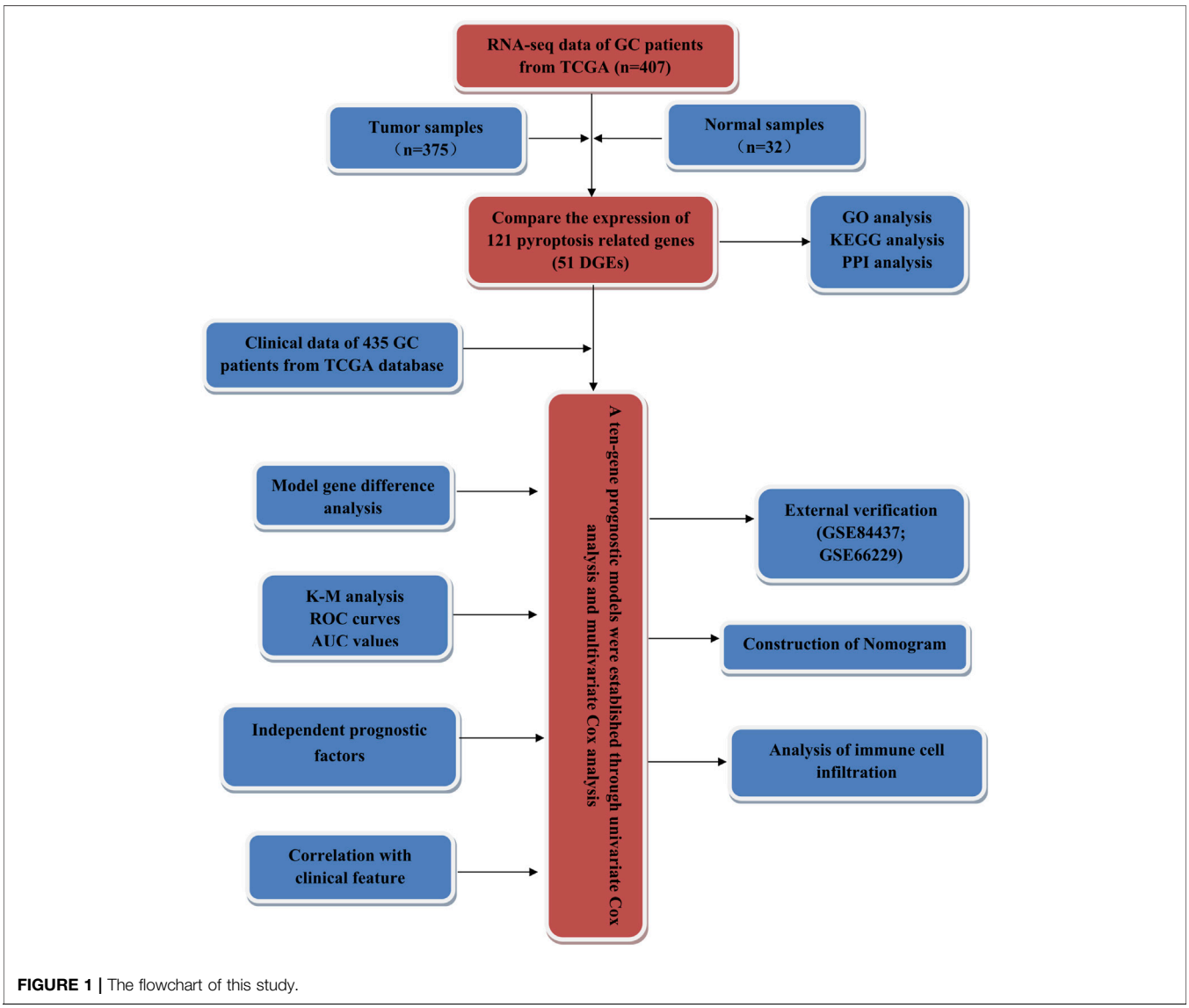


FIGURE 1 | The flowchart of this study.

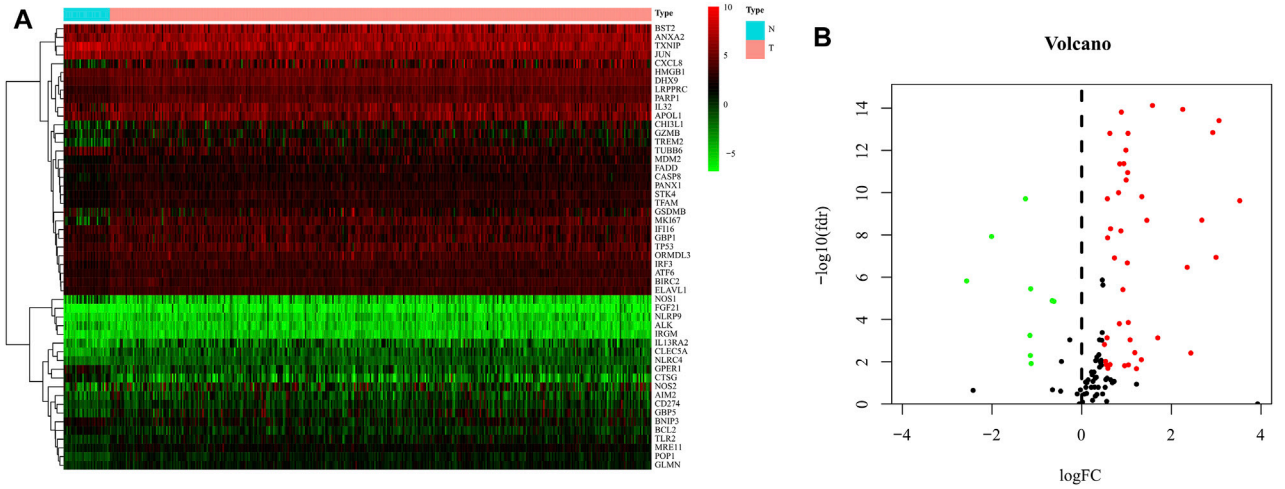


FIGURE 2 | Identification of differential expressed genes (DEGs) in GC and normal tissues. **(A)** Heat map of 51 DEGs in TCGA. Red: upregulation; Green: downregulation. The abscissa represents the type, N normal; T Tumor; ordinate represents the gene. **(B)** Volcano plots of the distributions of 51 DEGs. The abscissa represents logFC and the ordinate represents -log10 (FDR).

MATERIALS AND METHODS

Data Acquisition

Transcriptome sequencing data, survival information, and clinical information of TCGA gastric cancer data set obtained were obtained from the TCGA website (https://portal.gdc.cancer.gov/). Two external validation datasets, GSE84437 and GSE66229, were downloaded from the Gene Expression Omnibus database (GEO). The R (version 4.0.2) software was used to standardize and process the data. Pyroptosis-related genes were searched through THE HUMAN GENE DATABASE (https://www.genecards.org/) with the keyword "Pyroptosis." Therefore, 121 genes related to pyroptosis were included in the analysis and provided in **Supplementary Material S1**. First of all, the differentially expressed pyroptosis-related genes in TCGA gastric cancer tissues were identified by using the "LIMMA" R software package: false detection rate (FDR) < 0.05.

GO and KEGG Functional Enrichment

Gene Ontology (GO) enrichment analysis of 51 differentially expressed pyroptosis-related genes was carried out with the R software package "clusterprofiler," The selected background gene set is compiled by others, which is the human genome annotation in Carlson M's "org.Hs.eg.db," including biological process (BP), cellular component (CC), and molecular function (MF). The filter conditions were pvalueCutoff = 0.05, qvalueCutoff = 0.05. The same tool was used to analyze the pathways enrichment according to the Kyoto Encyclopedia of Gene and Genome (KEGG).

PPI Network

Protein-protein interaction (PPI) networks in differentially expressed pyroptosis-related genes were constructed using the STRING database and visualized with the Cytoscape software. The Molecular Complex Detection (MCODE) of the Cytoscape plugin was used to detect the important modules in the PPI network, and GO and KEGG analyses were carried out to further study its molecular function in gastric cancer.

Establishment and Verification of Prognostic Model

Univariate Cox proportional hazard regression analysis was used to screen pyroptosis-related genes significantly associated with overall survival (OS) in TCGA gastric cancer data set. Then multivariate Cox proportional hazard regression analysis was performed to establish a prognostic model. We use the following formula to calculate the risk score for each patient: risk score = $e^{\text{sum (each gene's expression}\times\text{corresponding coefficient)}}$. Patients were divided into high-risk and low-risk groups according to the median value of the risk score.

To determine the role of the risk score in predicting the clinical prognosis of patients with gastric cancer, we analyzed the difference in survival time between the high-risk group and low-risk group by the Kaplan-Meier method. To verify whether our prognostic model is also applicable to other datasets, we selected GSE84437 and GSE66229 gastric cancer datasets from the GEO database for external verification and calculated the risk score using the same formula as the TCGA cohort in these two datasets. Univariate and multivariate Cox regression analyses were used to study whether pyroptosis-related risk index could be an independent predictor of OS in the TCGA dataset of patients with gastric cancer. Risk score, age, sex, tumor subtype, pathological stage, and histological grade were regarded as covariates.

Copy number changes and mutations of key genes were investigated using the online database cBioPortal (Gao et al., 2013), and protein expression data were retrieved from the human protein map (HPA) database.

The Construction of Nomogram

According to age, staging, grading, T, N, M, and risk score, the nomogram was constructed with "RMS" and "survival" software

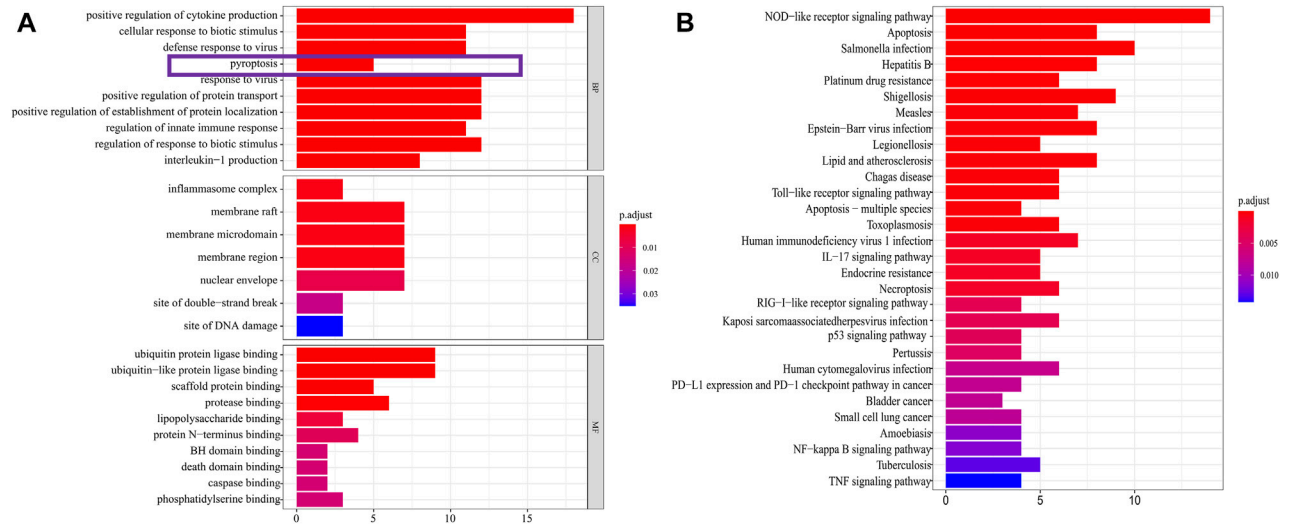


FIGURE 3 | DEGs enrichment analysis. **(A)** The results of GO enrichment in the TCGA cohort. "BP" stands for "biological process," "CC" stands for "cellular component" and "MF" stands for "molecular function". The abscissa represents the gene ratio. **(B)** The results of KEGG enrichment in the TCGA cohort. The abscissa represents the gene ratio.

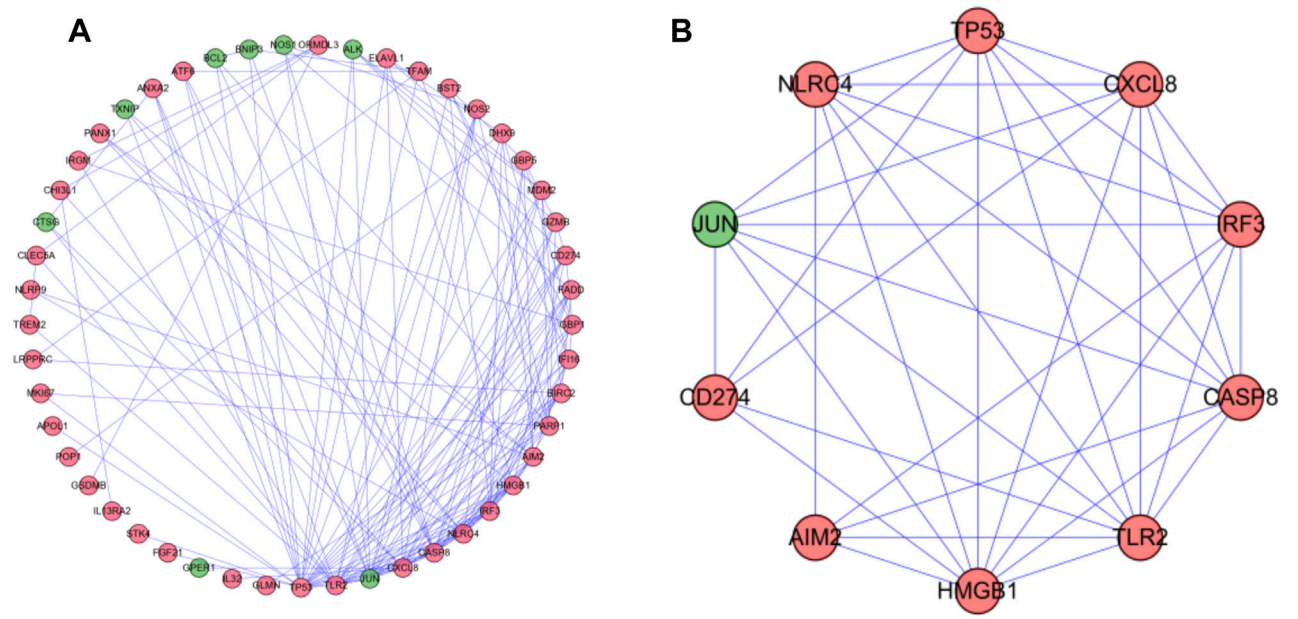


FIGURE 4 | PPI network and modules analysis. **(A)** PPI network for DEGs. **(B)** Key module in PPI network. Red: upregulation; Green: downregulation.

packages in R. The evaluation of the consistency between actual and predicted survival is achieved by generating a calibration curve. Finally, the ROC curve of the nomogram changing with time is generated, and the AUC value is calculated.

Analysis of the Correlation Between Risk Score Model and Immune Cell Infiltration

We used tumor immune estimation resources (TIMER), which is a reliable resource for comprehensive analysis of tumor-infiltrating immune cells, to explore the relationship between prognostic models and immune cell infiltration. The TIMER algorithm can help users estimate the composition of six tumor-infiltrating immune cell subpopulations (B cells, CD4+T cells, CD8+T cells, macrophages, neutrophils, and dendritic cells). The level of immune infiltration in patients with gastric cancer was obtained from the TIMER website (http://cistrome.dfci.harvard.edu/TIMER/), and the correlation between 6 tumor-infiltrating immune cell types and our prognostic model was analyzed in R.

TABLE 1 | Genes included in the prognostic gene signature.

Gene symbol	Coefficient	HR	*p* Value
BIRC2	−0.400605792	0.669914095	0.099458
CD274	−0.323210286	0.723821629	0.140321
IRGM	2.352594533	10.51281021	0.001412
ANXA2	0.267383737	1.306541719	0.036579
GBP5	−0.252281345	0.777026095	0.107473
TXNIP	0.142945075	1.153666434	0.116587
POP1	−0.479593899	0.619034731	0.057659
GBP1	0.301409059	1.351762179	0.074925
DHX9	0.539214896	1.714660147	0.050588
TLR2	0.295854165	1.344274101	0.03618

Statistical Analysis

Further statistical analyses were performed with the R software (v4.0.2). Wilcoxon ranked sum (Mann Whitney) test was used to screen the differentially expressed pyroptosis-related genes between tumor and adjacent tissues. The independent predictors of OS and the relationship between risk score and clinical information and prognosis were determined by univariate and multivariate Cox regression analysis. The difference of OS between the high-risk group and low-risk group was compared by the Kaplan-Meier method, and the *p*-value was calculated by logarithmic rank-sum test. T-test was used to compare the differences in risk scores among different

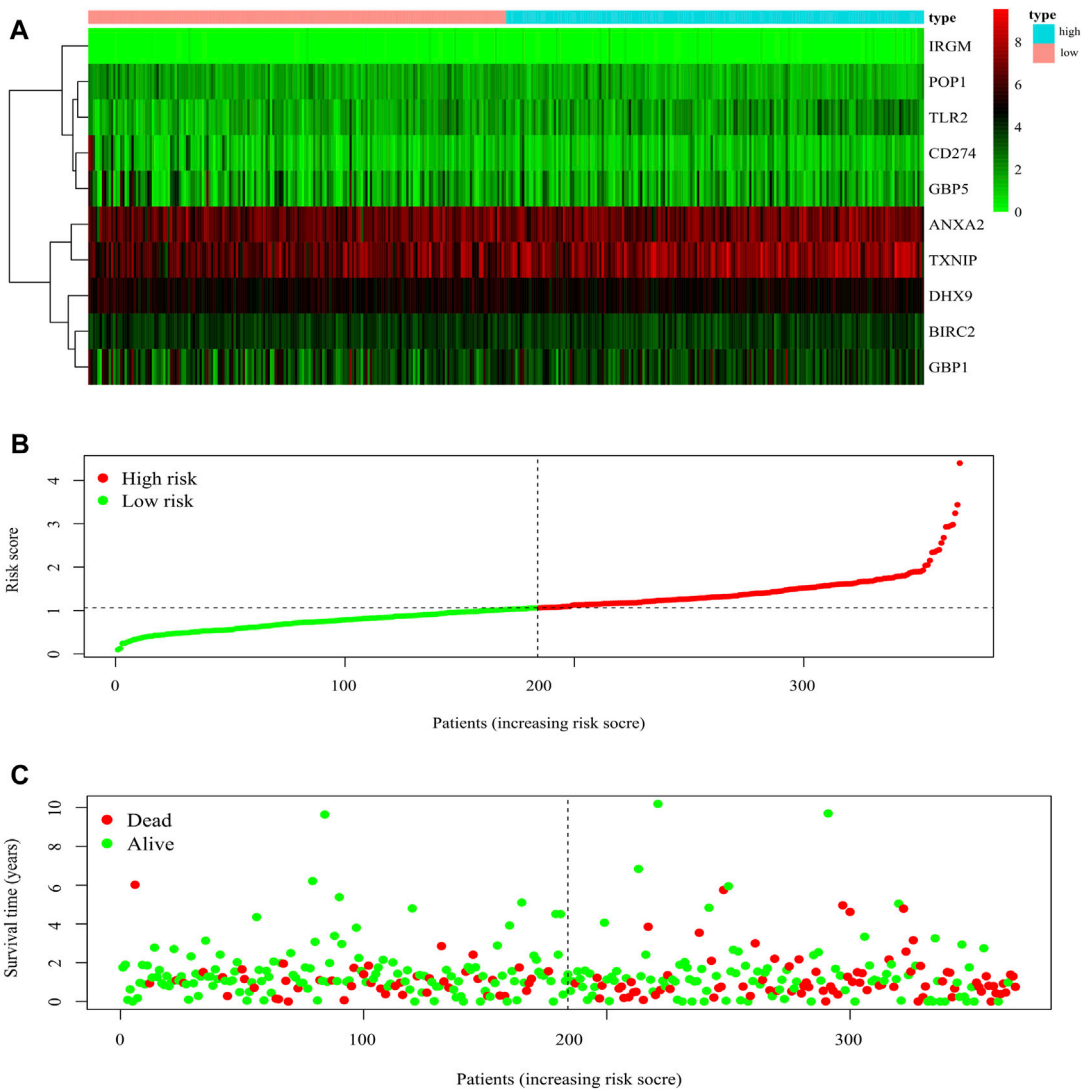


FIGURE 5 | Characteristics of prognostic gene signatures. **(A)** The heatmap of the ten DEGs expression profiles in high- and low-risk GC patients. The abscissa represents risk types, the ordinate represents the gene. **(B)** Distribution of risk scores of high- and low-risk GC patients. The abscissa represents the patients (increasing risk score), the ordinate represents the risk score. **(C)** The scatter plot shows the correlation between survival time and risk score. The abscissa represents the patients (increasing risk score), the ordinate represents survival time (years).

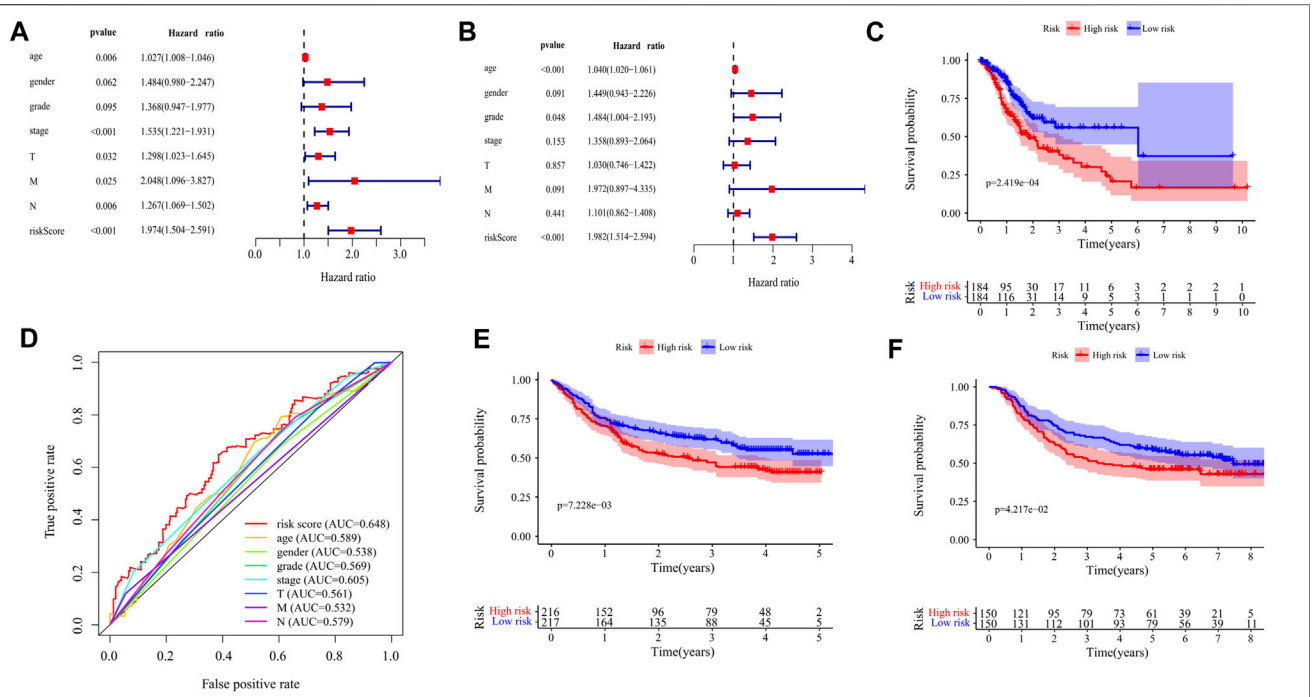


FIGURE 6 | Validation of the prognostic signature of ten DEGs. **(A)** The Forest plot reflects the univariate Cox analysis of the relationship between the clinical features, risk score, and OS of GC patients. Both stage and risk score significantly affect the prognosis of GC patients ($p < 0.001$). **(B)** The Forest plot reflects the multivariate Cox analysis of the relationship between the clinical features, risk score, and OS of GC patients. Age and risk score are independent prognostic risk factors for GC ($p < 0.001$). **(C)** The Kaplan-Meier Survival curve shows that the OS of high-risk GC patients is significantly lower than that of low-risk patients. The abscissa represents time (years), the ordinate represents survival probability. **(D)** The 1-year time-dependent ROC curve shows that the prediction accuracy of the risk score is higher than other clinical features (AUC = 0.648). The abscissa represents false positive rate, the ordinate represents true positive rate. **(E)** Kaplan-Meier Survival Curve of patients with GC in high-risk and low-risk groups in GSE84437 (p = 7.228e−03). The abscissa represents time (years), the ordinate represents survival probability. **(F)** Kaplan-Meier Survival Curve of patients with GC in high-risk and low-risk groups in GSE66229 (p = 4.217e−02). The abscissa represents time (years), the ordinate represents survival probability.

clinical feature groups. $p < 0.05$ was considered to be statistically significant. We conducted our study as described in the flowchart (**Figure 1**).

RESULT

Identification and Enrichment Analysis of Differentially Expressed Genes

The expression data of 375 cases of gastric cancer tissues and 32 normal gastric tissues, for a total of 407 cases, were downloaded from the TCGA database, together with the expression data, survival status, and clinical data of 435 cases of gastric cancer patients. After extracting the patient information including survival status and survival time among these 435 GC patients, and intersecting them with the 407 patients for which RNA-seq data were available, we obtained 368 GC patients that presented RNA-seq data and clinical data and used it as a training set to build the model. 121 pyroptosis-related genes were obtained from the human gene database. R-Package "Limma" was used to screen pyroptosis-related genes in gastric cancer. The screening criteria were logFC >0.5, FDR <0.05. The results showed that 51 genes were identified as differentially expressed in gastric cancer (**Figure 2**). GO analysis showed that these genes were mainly enriched in the basic biological processes (BP) of positive regulation of cytokine (Mirgayazova et al., 2019) production, cellular response to biotic stimulus, defense response to viruses, and pyroptosis (**Figure 3A**). KEGG analysis showed that these genes were mainly related to NOD–like receptor signaling pathway, apoptosis, and *Salmonella* infection (**Figure 3B**).

PPI Network Analysis

To further understand the role of differential genes in the GC process, we use the STRING database and Cytoscape software to construct a PPI network, which is composed of 49 nodes and 155 edges (**Figure 4A**). Then, we used the MODE plugin Cytoscape to identify the key modules from the PPI network. The modules included nine up-regulated differentially expressed genes (DEGs) and one down-regulated DEGs (**Figure 4B**). Enrichment analysis showed that they were related to positive regulation of cytokine production, hepatitis, and NOD-like receptor signaling pathway.

Construction and Verification of Prognostic Genes Related to Pyroptosis

The 51 differentially expressed genes screened above were included in univariate COX analysis, and 47 genes related to prognosis were screened (**Supplementary Material S2**). The 47 genes were included in multivariate Cox regression analysis to construct prognostic

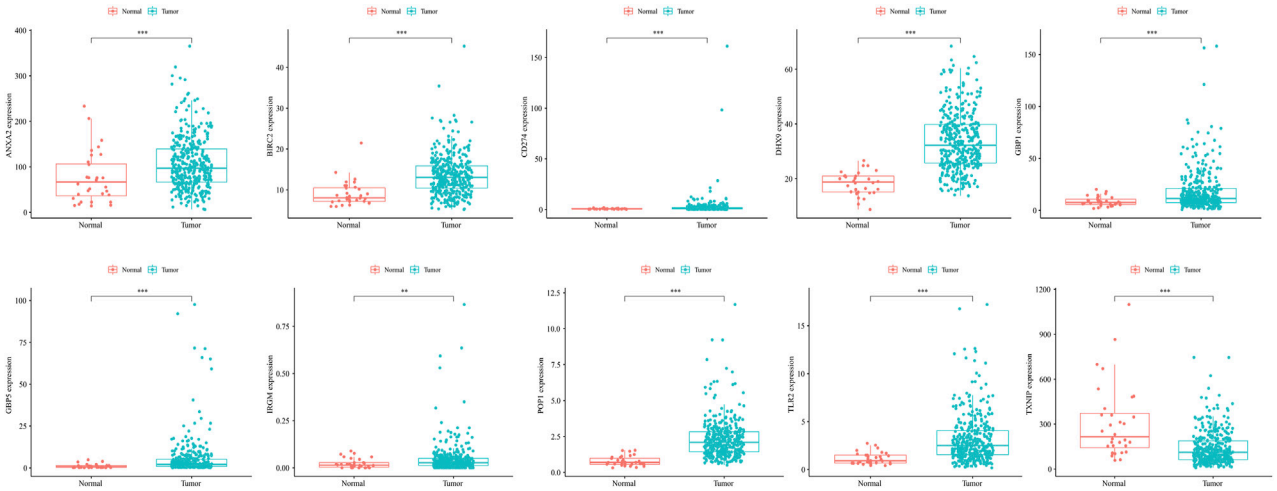


FIGURE 7 | The ten genes were differentially expressed between normal samples and tumor samples (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). The abscissa represents tissues, the ordinate represents gene expression.

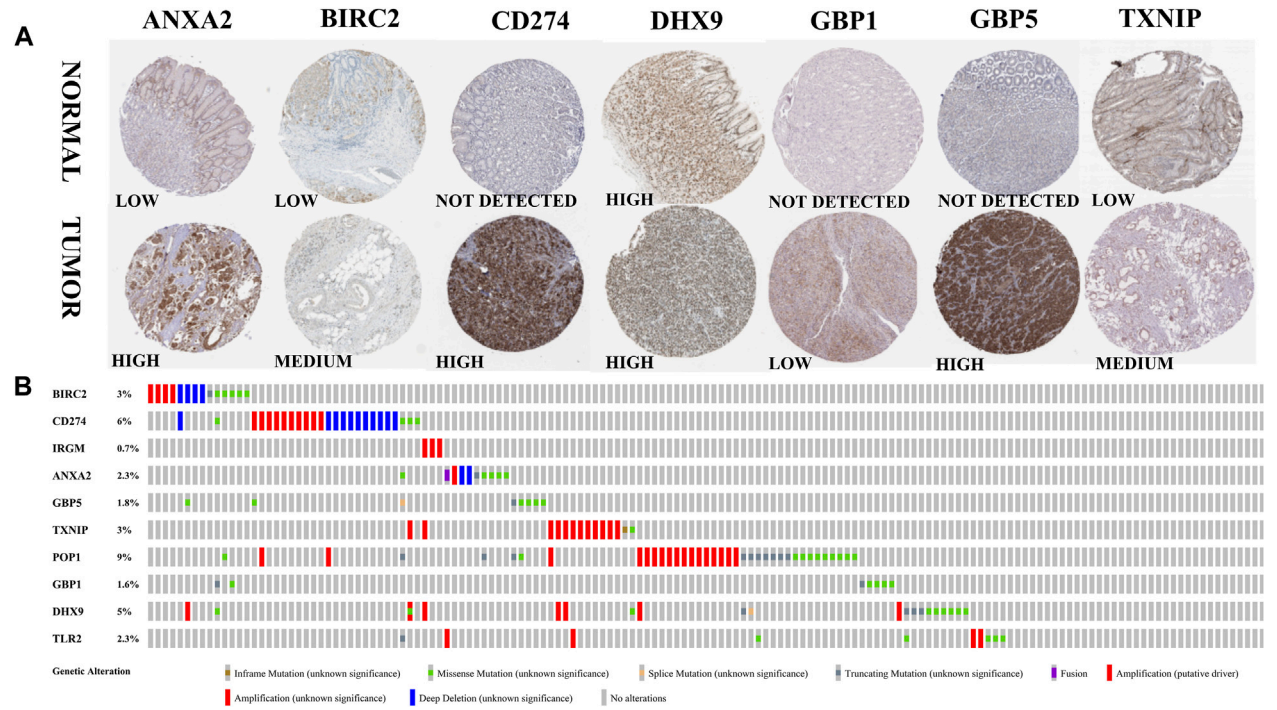


FIGURE 8 | Expression and alteration of the ten prognosis-related RBP genes. **(A)** The representative protein expression of the seven genes in GC and normal tissue. Data were from the Human Protein Atlas (http://www.proteinatlas.org) online database. **(B)** The expression alteration profiles of the ten genes in the TCGA GC RNA-seq dataset.

markers, and 10 genes (BIRC2, CD274, IRGM, ANXA2, GBP5, TXNIP, POP1, GBP1, DHX9, and TLR2) related to prognosis were obtained. The coefficients of each gene are shown in **Table 1**. The risk score was calculated as follows: risk score = (−0.400605791695688*BIRC2 expression) + (−0.323210285810829*CD274 expression) + (2.35259453336258*IRGM expression) + (0.267383737123388*ANXA2 expression) + (−0.252281344553063*GBP5 expression) + (0.142945074684017*TXNIP expression) + (−0.479593898864018*POP1 expression) + (0.301409058780011*GBP1 expression) + (0.539214896060124*DHX9 expression) + (0.295854165226579*TLR2 expression). The risk score of each gastric cancer patient was calculated according to the

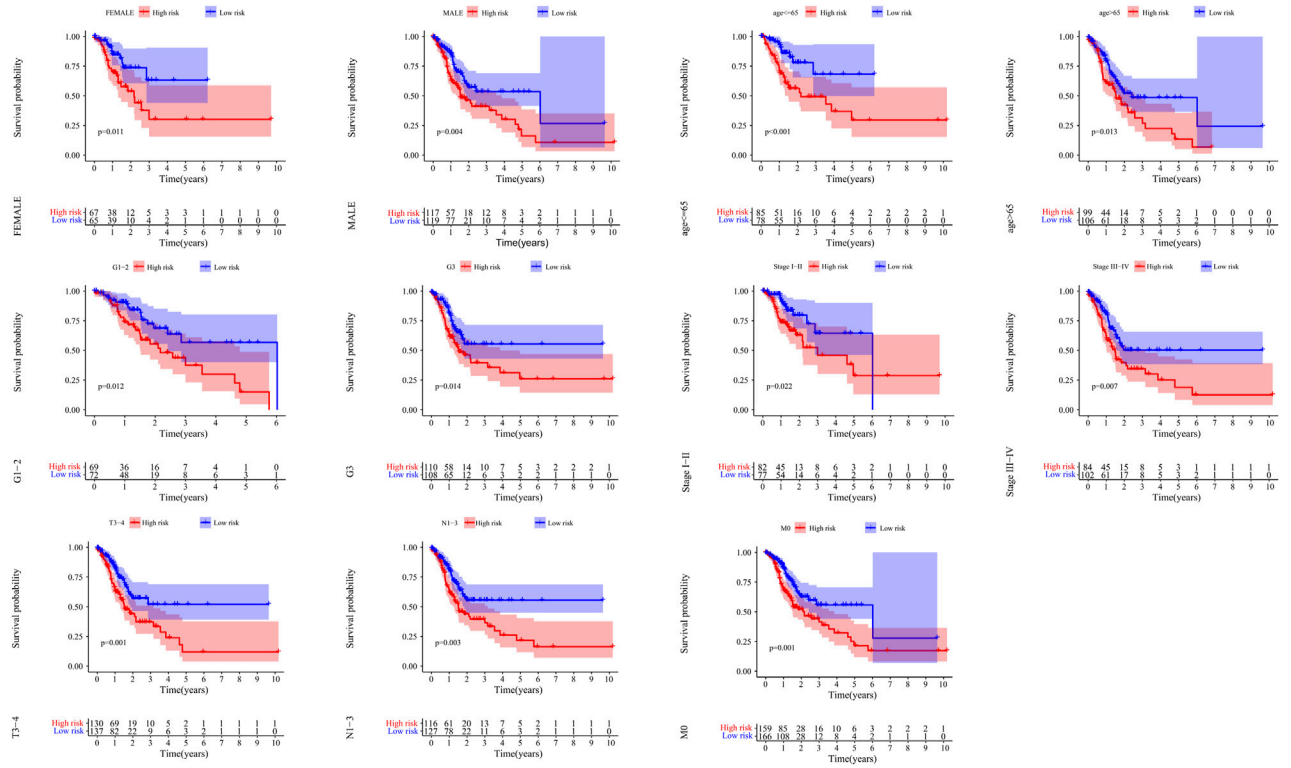


FIGURE 9 | The correlation between the clinical features and risk score of GC patients. The abscissa represents time (years), the ordinate represents survival probability.

expression level of the ten genes, and the patients were divided into high-risk and low-risk according to the median risk score. The gene expression profiles of the high-risk group and low-risk group are shown by the heatmap (**Figure 5A**). **Figure 5B** shows the risk score distribution of patients with gastric cancer, which increases gradually from left to right and divides the patients into two groups. **Figure 5C** shows the distribution of survival status and survival time of patients with different risk scores.

Validation of the 10-Gene Signature

The prognostic value of the risk score was evaluated by univariate and multivariate analysis. Univariate analysis showed that there was a significant correlation between risk score and overall survival (OS) (HR = 1.974,95%CI = 1.504–2.591, $p < 0.001$) (**Figure 6A**). Multivariate analysis showed that risk score was an independent prognostic index (HR = 1.982,95%CI = 1.514–2.594, $p < 0.001$) (**Figure 6B**). Kaplan-Meier cumulative curve shows that patients with a low-risk score have a longer survival time than patients with a high-risk score (**Figure 6C**). The AUC of risk score was higher than that of sex, age, pathological grade, and TNM stage, which proved that the Cox model was better than other single indexes in predicting the prognosis (**Figure 6D**). To verify the predictive value of this prognostic model, we used the same formula to calculate the risk scores of patients in the GSE84437 and GSE66229 datasets. The OS of the high-risk group was significantly lower than that of the low-risk group (GSE84437: $p = 7.228e{-}03$; GSE66229: $p = 4.217e{-}02$), which is consistent with the results in the TCGA cohort (**Figures 6E,F**).

Expression and Alteration of the Ten Prognosis-Related RBP Genes

The expression of model genes in GC was observed by analyzing the difference of model genes between normal samples and tumor samples (**Figure 7**). The ten genes were differentially expressed between normal samples and tumor samples (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). We further analyzed the expression of these model genes through the HPA database. **Figure 8A** shows the immunohistochemical results of seven key RBPs in GC and normal tissues. IRGM, POP1, and TLR2 are not included in the database. By using the cBioPortal online database, we found that the main alteration of 10 RBP genes identified in CG patients was amplification (**Figure 8B**).

The Correlation Between the Clinical Features and Risk Score of GC Patients

Kaplan-Meier curves showed that patients with low risk had a better prognosis in >65 years old, ≤ 65 years old, male, female,

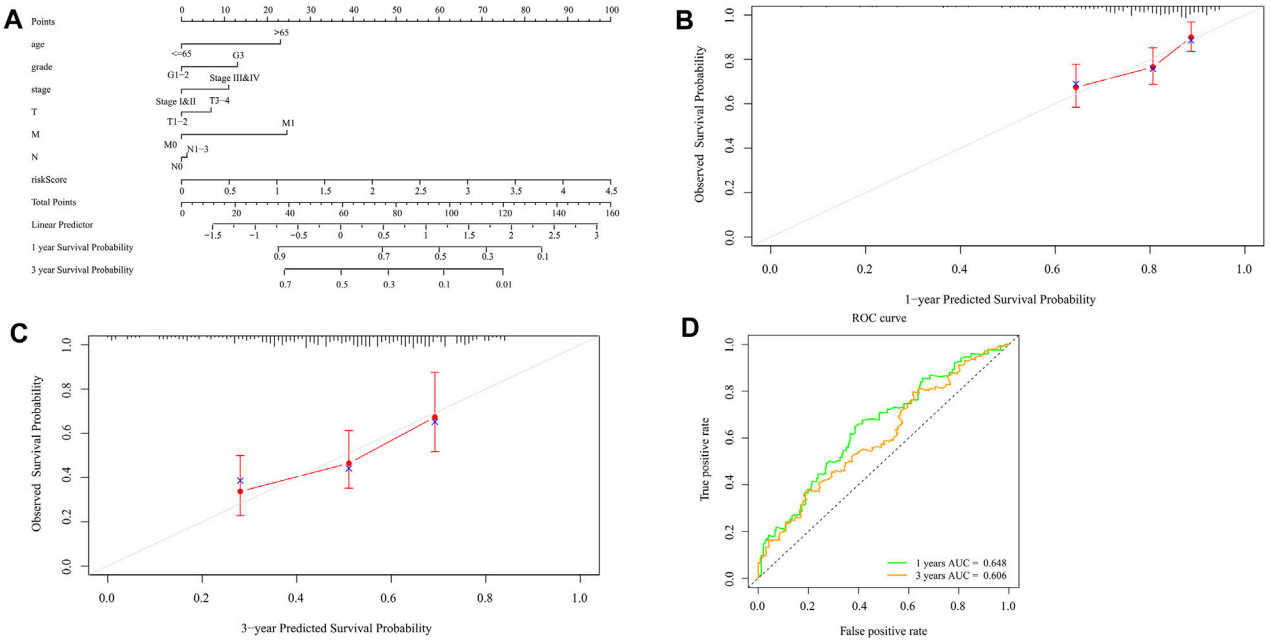


FIGURE 10 | Construction and validation of the nomogram. **(A)** Scores of each item of GC patients were calculated according to the nomogram, and the total scores obtained after addition can predict the 1- and 3-year survival probability. **(B,C)** The 1- and 3-year calibration curves of the nomogram **(D)** The ROC curves of 1- and 3-year nomogram (AUC = 0.648 for 1 year, AUC = 0.606 for 3 years). The abscissa represents false positive rate, the ordinate represents true positive rate.

G1-2, G3, M0, N1-3, StageI-II, StageIII-IV and T3-4 ($p < 0.05$) (**Figure 9**).

Construction and Verification of Nomogram

Nomograms can be used to help clinical interpretation of predictive signals, and can easily determine the survival rate of patients with gastric cancer. By combining the characteristics of ten pyroptosis-related prognostic genes with clinicopathological features, a nomogram for predicting individual survival probability was established (**Figure 10A**), and the possibility of 1-year and 3-year OS was predicted. When the calibration curve is closer to the diagonal, it is proved that the prediction result is more accurate (**Figures 10B,C**). The 1-year and 3-year ROC curves (**Figure 10D**) also show that the predictive ability of the nomogram is good (1-year AUC = 0.648,3-year AUC = 0.606).

Analysis of Immune Cell Infiltration

By exploring the relationship between the risk score model and immune cell infiltration, we found that dendritic cells, macrophages, and neutrophils were positively correlated with the risk score. However, there was no significant correlation between B cells, CD8+T cells, and CD4+T cells and the risk score (**Figure 11**).

DISCUSSION

The prognosis of gastric cancer, one of the most common malignant tumors in the world, is still not optimistic. Surgery is the more reliable treatment at present, but further treatment of patients after surgery and conservative treatment of advanced patients have limited benefits. Therefore, to achieve early diagnosis and find treatment targets, it is particularly important to explore the pathogenesis of gastric cancer and establish effective prognostic criteria, which may help patients to develop personalized treatment plans.

In this study, we analyzed the transcriptome information and clinical data of GC patients in the TCGA database. Univariate Cox analysis showed that 47 pyroptosis genes were associated with prognosis. Through multivariate Cox analysis, we established a prognostic model based on ten pyroptosis genes, including BIRC2, CD274, IRGM, ANXA2, GBP5, TXNIP, POP1, GBP1, DHX9, and TLR2. The risk score of each patient can be calculated from the mRNA expression level and risk coefficient of these ten genes. This is an independent risk factor that can affect the prognosis and can predict that the high-risk CG patients have a worse prognosis than the low-risk patients. Through the ROC curve analysis of the survival rate of patients with GC, it is found that the prognostic index has good sensitivity and specificity (AUC = 0.648), which can be used as a reliable predictor of the prognosis of patients with GC. Moreover, we could successfully verify our results in two independent datasets, GSE84437 and GSE66229. We also constructed a nomogram to predict the 1-and 3-year survival rates of patients for clinical application and to obtain more accurate prediction results. The nomogram has better prediction accuracy than the correction curve and ROC curve. Predicted results and the actual results are in good agreement.

BIRC2 is a closely related member of the inhibitor of apoptosis (IAP) family, which plays a key role in nuclear factor JB (NFjB)

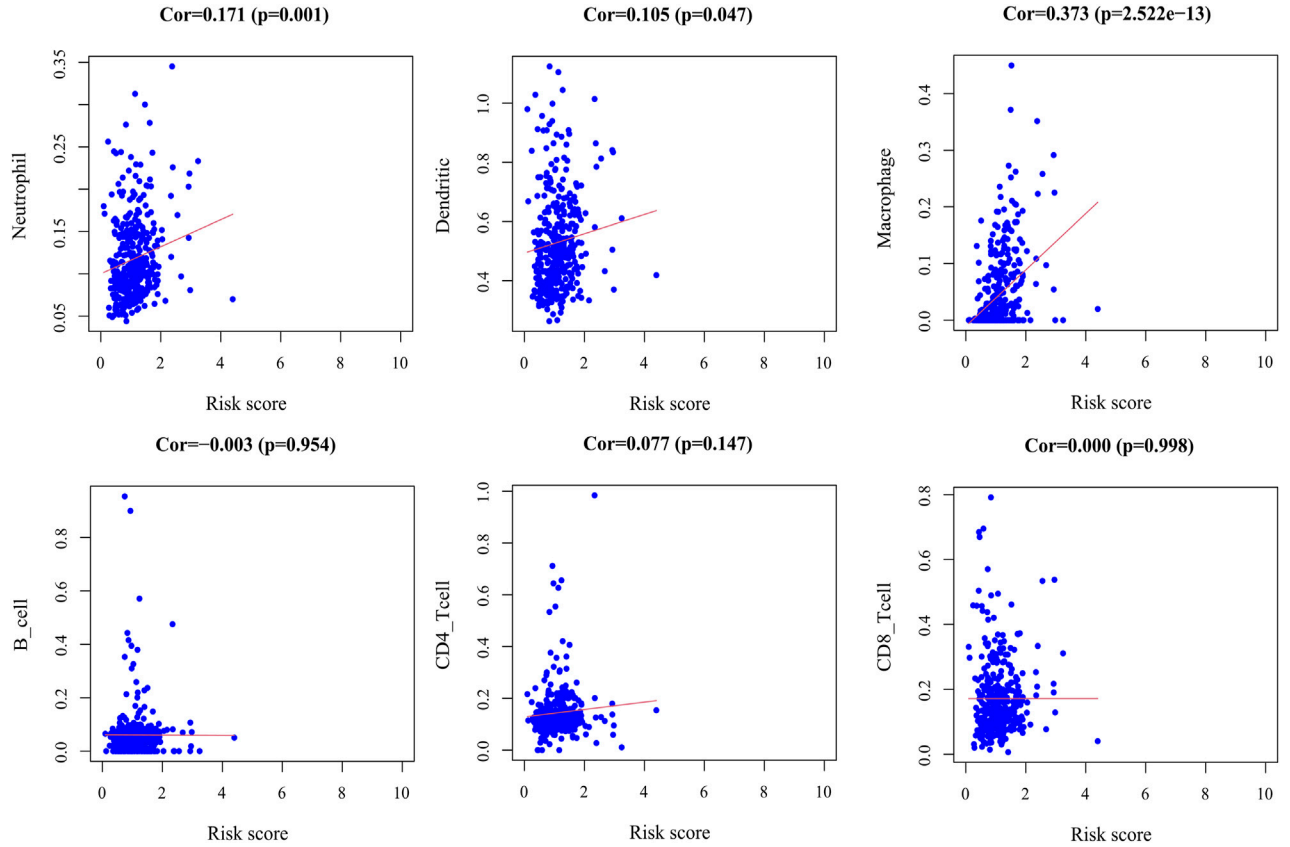


FIGURE 11 | Correlation plot between risk score and immune cells infiltration. The abscissa represents risk score, the ordinate represents infiltration abundances of immune cells.

signal transduction and apoptosis (Yamato et al., 2015). BIRC2 is an effective negative regulator of LTR-dependent HIV-1 transcription (Pache et al., 2015). When its expression is absent, it can inhibit the growth of breast cancer or melanoma through an immune-mediated mechanism (Samanta et al., 2020). On the contrary, the overexpression of BIRC2 promotes the metastasis of gastric cancer cells (Chen et al., 2020a). CD274 (Hou et al., 2020a) is a key molecule of tumor immune checkpoint mechanisms, is one of the main targets of immunotherapy (Fabrizio et al., 2018), and plays an important role in tumor immune escape (Huang et al., 2019). Cancer cells expressing CD274 may affect regulatory T cells in the tumor microenvironment (Masugi et al., 2017). PD-L1 switches TNFα-induced apoptosis to pyroptosis in cancer cells, resulting in tumor necrosis (Hou et al., 2020b). IRGM is a human protein of the immune-associated GTPase family that promotes autophagy during inflammation and infection (Song et al., 2015). It has been reported that IRGM plays an important role in non-small cell lung cancer (Wang et al., 2018) and liver cancer by regulating autophagy (Chen et al., 2021). In addition, IRGM interacts with NLRP3 and ASC and blocks the assembly of inflammatory bodies by blocking the oligomerization of NLRP3 and ASC (Mehto et al., 2019). ANXA2 is a 36 kDa calcium-dependent phospholipid-binding cytoskeletal protein. When ANXA2 is silenced, the ability of proliferation, invasion, and migration of gastric cancer cells is weakened (Xie et al., 2019), but when it is overexpressed, it can promote the migration, invasion, and metastasis of esophageal cancer cells *in vitro* and *in vivo* by activating the MYC-HIF1a-VEGF cascade pathway (Ma et al., 2018). GBP5 and GBP1 are interferon-inducible GTPases belonging to the guanylate binding protein (GBP) family, which can promote antibacterial immunity and cell death. They support the activation of caspase-1-containing inflammasome complexes or caspase-4, which triggers pyroptosis (Fisch et al., 2019). Studies by Jing Zhao et al. (2019) have shown that GBP1 can promote survival or carcinogenesis in prostate cancer. TXNIP is the only known α-arrestin protein family that binds to Trx (Schröder et al., 2020). The expression of TXNIP in tumors is very low, and it may play an inhibitory role in many kinds of cancers such as liver cancer, breast cancer and lung cancer (Chen et al., 2020b). POP1 is an intact membrane protein that regulates the formation of tight junctions (Williams et al., 2011). It inhibits the assembly of ASC-dependent inflammatory bodies by preventing the nucleation of inflammatory bodies, thereby interfering with the activation of caspase-1, the release of IL-1b and IL-18, pyroptosis and the release of ASC particles (de Almeida et al., 2015). DHX9 is a member of RNA helicase DExH subgroup,

which plays an important role in several aspects of RNA metabolism (Palombo et al., 2020). The expression of DHX9 is up-regulated in cervical cancer tissue, which promotes the movement and angiogenesis of cervical cancer cells. Moreover, DHX9 plays an important role in promoting the metastasis of colorectal cancer (Hou et al., 2021). TLR2 belongs to the Toll-like receptor family and is a key regulator of innate and acquired immune responses (Cao et al., 2019).

We found that dendritic cells, macrophages, and neutrophils infiltration into CG tumors were positively correlated with the risk score. Macrophages create an inflammatory environment that is mutagenic and promotes growth at the beginning of tumor formation. With tumor development, macrophages stimulate angiogenesis, enhance the migration and invasion of tumor cells, and inhibit anti-tumor immunity (Qian and Pollard, 2010). It has been confirmed that macrophages play an important role in the occurrence and development of gastric cancer (Chen et al., 2017; Zheng et al., 2017; Eissmann et al., 2019; Gambardella et al., 2020). Dendritic cells are powerful antigen-presenting cells that can stimulate immature resting T cells and initiate the initial immune response (Wu et al., 2004). The fusion vaccine of allogeneic dendritic cells and tumor cells can be used to enhance the effect of immunotherapy in patients with gastric cancer (Li et al., 2015). Neutrophils are the first responders to inflammation and infection. Tumor-associated neutrophils can promote tumor inflammation by promoting angiogenesis, extracellular matrix remodeling, metastasis, and immunosuppression (Mollinedo, 2019). It has been confirmed that tumor-associated neutrophils can promote the progression and metastasis of gastric cancer in many ways (Jaillon et al., 2020). Shao et al. (2021) established the gastric cancer score related to pyroptosis, which proved its significant correlation with the immune microenvironment, and further confirmed our research results. It provides support for the development of immunotherapy strategies related to pyroptosis in the future.

This study has some limitations. The conclusion of this study is based on bioinformatic analyses and lacks further verification *in vivo* and *in vitro*. The samples come from a retrospective study, so it is necessary to conduct a more in-depth prospective clinical study of the signature and nomogram.

In summary, we have established a new prognostic model for gastric cancer and pyroptosis. The score generated by the risk signature of the markers is an independent risk factor for predicting OS. According to the established nomogram, the 1- and 3-year survival rate of patients with gastric cancer can be predicted, which provides a reference for the formulation of personalized treatment for patients with gastric cancer, and provides an important basis for further study of the relationship between gastric cancer and pyroptosis.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) and THE HUMAN GENE DATABASE (https://www.genecards.org/).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CwL and CjL: conception and design. CwL and JF: acquisition, analysis, and interpretation of data. CwL and CjL: figures drawing. CwL and JG: writing and revision of the manuscript. JG: study supervision. All authors read and approved the final manuscript.

FUNDING

This research was supported by the Shanxi Province Applied Basic Research Program (201901D211480) and the Shanxi Province Surface Fund Project (201901D111349).

ACKNOWLEDGMENTS

We would like to thank everyone who took part in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.699503/full#supplementary-material

REFERENCES

Bergsbaken, T., Fink, S. L., and Cookson, B. T. (2009). Pyroptosis: Host Cell Death and Inflammation. *Nat. Rev. Microbiol.* 7 (2), 99–109. doi:10.1038/nrmicro2070

Cao, D., Wu, Y., Jia, Z., Zhao, D., Zhang, Y., Zhou, T., et al. (2019). 18β-Glycyrrhetinic Acid Inhibited Mitochondrial Energy Metabolism and Gastric Carcinogenesis through Methylation-Regulated TLR2 Signaling Pathway. *Carcinogenesis* 40 (2), 234–245. doi:10.1093/carcin/bgy150

Chen, P., Guo, H., Wu, X., Li, J., Duan, X., Ba, Q., et al. (2020). Epigenetic Silencing of microRNA-204 by *Helicobacter P* Augments the NF-Kb Signaling Pathway in Gastric Cancer Development and Progression. *Carcinogenesis* 41 (4), 430–441. doi:10.1093/carcin/bgz143

Chen, X., Yan, Y.-L., Zeng, S.-S., Gong, Z.-C., and Xu, Z.-J. (2021). Caveolin-1 Promotes Radioresistance via IRGM-Regulated Autophagy in Lung Cancer. *Ann. Transl Med.* 9 (1), 47. doi:10.21037/atm-20-3293

Chen, Y., Ning, J., Cao, W., Wang, S., Du, T., Jiang, J., et al. (2020). Research Progress of TXNIP as a Tumor Suppressor Gene Participating in the Metabolic

Reprogramming and Oxidative Stress of Cancer Cells in Various Cancers. *Front. Oncol.* 10, 568574. doi:10.3389/fonc.2020.568574

Chen, Y., Zhang, S., Wang, Q., and Zhang, X. (2017). Tumor-Recruited M2 Macrophages Promote Gastric and Breast Cancer Metastasis via M2 Macrophage-Secreted CHI3L1 Protein. *J. Hematol. Oncol.* 10 (1), 36. doi:10.1186/s13045-017-0408-0

Chia, N.-Y., and Tan, P. (2016). Molecular Classification of Gastric Cancer. *Ann. Oncol.* 27 (5), 763–769. doi:10.1093/annonc/mdw040

de Almeida, L., Khare, S., Misharin, A. V., Patel, R., Ratsimandresey, R. A., Wallin, M. C., et al. (2015). The PYRIN Domain-Only Protein POP1 Inhibits Inflammasome Assembly and Ameliorates Inflammatory Disease. *Immunity* 43 (2), 264–276. doi:10.1016/j.immuni.2015.07.018

Eissmann, M. F., Dijkstra, C., Jarnicki, A., Phesse, T., Brunnberg, J., Poh, A. R., et al. (2019). IL-33-mediated Mast Cell Activation Promotes Gastric Cancer through Macrophage Mobilization. *Nat. Commun.* 10 (1), 2735. doi:10.1038/s41467-019-10676-1

Erkes, D. A., Cai, W., Sanchez, I. M., Purwin, T. J., Rogers, C., Field, C. O., et al. (2020). Mutant BRAF and MEK Inhibitors Regulate the Tumor Immune Microenvironment via Pyroptosis. *Cancer Discov.* 10 (2), 254–269. doi:10.1158/2159-8290.cd-19-0672

Fabrizio, F. P., Trombetta, D., Rossi, A., Sparaneo, A., Castellana, S., and Muscarella, L. A. (2018). Gene Code CD274/PD-L1: from Molecular Basis toward Cancer Immunotherapy. *Ther. Adv. Med. Oncol.* 10, 1758835918815598. doi:10.1177/1758835918815598

Fang, Y., Tian, S., Pan, Y., Li, W., Wang, Q., Tang, Y., et al. (2020). Pyroptosis: A New Frontier in Cancer. *Biomed. Pharmacother.* 121, 109595. doi:10.1016/j.biopha.2019.109595

Fisch, D., Bando, H., Clough, B., Hornung, V., Yamamoto, M., Shenoy, A. R., et al. (2019). Human GBP1 Is a Microbe-Specific Gatekeeper of Macrophage Apoptosis and Pyroptosis. *EMBO J.* 38 (13), e100926. doi:10.15252/embj.2018100926

Gambardella, V., Castillo, J., Tarazona, N., Gimeno-Valiente, F., Martínez-Ciarpaglini, C., Cabeza-Segura, M., et al. (2020). The Role of Tumor-Associated Macrophages in Gastric Cancer Development and Their Potential as a Therapeutic Target. *Cancer Treat. Rev.* 86, 102015. doi:10.1016/j.ctrv.2020.102015

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

Hou, J., Zhao, R., Xia, W., Chang, C.-W., You, Y., Hsu, J.-M., et al. (2020). PD-L1-Mediated Gasdermin C Expression Switches Apoptosis to Pyroptosis in Cancer Cells and Facilitates Tumour Necrosis. *Nat. Cel Biol.* 22 (10), 1264–1275. doi:10.1038/s41556-020-0575-z

Hou, J., Zhao, R., Xia, W., Chang, C.-W., You, Y., Hsu, J.-M., et al. (2020). PD-L1-Mediated Gasdermin C Expression Switches Apoptosis to Pyroptosis in Cancer Cells and Facilitates Tumour Necrosis. *Nat. Cel Biol.* 22 (10), 1264–1275. doi:10.1038/s41556-020-0575-z

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, X., Zhang, Q., Lou, Y., Wang, J., Zhao, X., Wang, L., et al. (2019). USP22 Deubiquitinates CD274 to Suppress Anticancer Immunity. *Cancer Immunol. Res.* 7 (10), 1580–1590. doi:10.1158/2326-6066.cir-18-0910

Jaillon, S., Ponzetta, A., Di Mitri, D., Santoni, A., Bonecchi, R., and Mantovani, A. (2020). Neutrophil Diversity and Plasticity in Tumour Progression and Therapy. *Nat. Rev. Cancer* 20 (9), 485–503. doi:10.1038/s41568-020-0281-y

Jiang, M., Qi, L., Li, L., and Li, Y. (2020). The Caspase-3/GSDME Signal Pathway as a Switch between Apoptosis and Pyroptosis in Cancer. *Cell Death Discov.* 6, 112. doi:10.1038/s41420-020-00349-0

Kayagaki, N., Stowe, I. B., Lee, B. L., O'Rourke, K., Anderson, K., Warming, S., et al. (2015). Caspase-11 Cleaves Gasdermin D for Non-Canonical Inflammasome Signalling. *Nature* 526 (7575), 666–671. doi:10.1038/nature15541

Li, C., Liang, S., Zhang, C., Liu, Y., Yang, M., Zhang, J., et al. (2015). Allogenic Dendritic Cell and Tumor Cell Fused Vaccine for Targeted Imaging and Enhanced Immunotherapeutic Efficacy of Gastric Cancer. *Biomaterials* 54, 177–187. doi:10.1016/j.biomaterials.2015.03.024

Ma, S., Lu, C.-C., Yang, L.-Y., Wang, J.-J., Wang, B.-S., Cai, H.-Q., et al. (2018). ANXA2 Promotes Esophageal Cancer Progression by Activating MYC-Hif1a-VEGF Axis. *J. Exp. Clin. Cancer Res.* 37 (1), 183. doi:10.1186/s13046-018-0851-y

Masugi, Y., Nishihara, R., Yang, J., Mima, K., da Silva, A., Shi, Y., et al. (2017). Tumour CD274 (PD-L1) Expression and T Cells in Colorectal Cancer. *Gut* 66 (8), 1463–1473. doi:10.1136/gutjnl-2016-311421

Mehto, S., Jena, K. K., Nath, P., Chauhan, S., Kolapalli, S. P., Das, S. K., et al. (2019). The Crohn's Disease Risk Factor IRGM Limits NLRP3 Inflammasome Activation by Impeding its Assembly and by Mediating its Selective Autophagy. *Mol. Cel* 73 (3), 429–445. doi:10.1016/j.molcel.2018.11.018

Mirgayazova, R., Khadiullina, R., Mingaleeva, R., Chasov, V., Gomzikova, M., Garanina, E., et al. (2019). Novel Isatin-Based Activator of P53 Transcriptional Functions in Tumor Cells. *Mol. Biol. Res. Commun.* 8 (3), 119–128. doi:10.22099/mbrc.2019.34179.1419

Mollinedo, F. (2019). Neutrophil Degranulation, Plasticity, and Cancer Metastasis. *Trends Immunol.* 40 (3), 228–242. doi:10.1016/j.it.2019.01.006

Pache, L., Dutra, M. S., Spivak, A. M., Marlett, J. M., Murry, J. P., Hwang, Y., et al. (2015). BIRC2/cIAP1 Is a Negative Regulator of HIV-1 Transcription and Can Be Targeted by Smac Mimetics to Promote Reversal of Viral Latency. *Cell Host & Microbe* 18 (3), 345–353. doi:10.1016/j.chom.2015.08.009

Palombo, R., Verdile, V., and Paronetto, M. P. (2020). Poison-Exon Inclusion in DHX9 Reduces its Expression and Sensitizes Ewing Sarcoma Cells to Chemotherapeutic Treatment. *Cells* 9 (2), 328. doi:10.3390/cells9020328

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

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

Samanta, D., Huang, T. Y.-T., Shah, R., Yang, Y., Pan, F., and Semenza, G. L. (2020). BIRC2 Expression Impairs Anti-Cancer Immunity and Immunotherapy Efficacy. *Cel Rep.* 32 (8), 108073. doi:10.1016/j.celrep.2020.108073

Schröder, J., Schumacher, U., and Böckelmann, L. C. (2020). Thioredoxin Interacting Protein (TXNIP) Is Differentially Expressed in Human Tumor Samples but Is Absent in Human Tumor Cell Line Xenografts: Implications for its Use as an Immunosurveillance Marker. *Cancers (Basel)* 12 (10), 3028. doi:10.3390/cancers12103028

Shao, W., Yang, Z., Fu, Y., Zheng, L., Liu, F., Chai, L., et al. (2021). The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. *Front. Cel Dev. Biol.* 9, 676485. doi:10.3389/fcell.2021.676485

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by Inflammatory Caspases Determines Pyroptotic Cell Death. *Nature* 526 (7575), 660–665. doi:10.1038/nature15514

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

Song, Z., Guo, C., Zhu, L., Shen, P., Wang, H., Guo, C., et al. (2015). Elevated Expression of Immunity-Related GTPase Family M in Gastric Cancer. *Tumor Biol.* 36 (7), 5591–5596. doi:10.1007/s13277-015-3229-1

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. clinicians* 71 (3), 209–249. doi:10.3322/caac.21660

Teng, J. F., Mei, Q. B., Zhou, X. G., Tang, Y., Xiong, R., Qiu, W. Q., et al. (2020). Polyphyllin VI Induces Caspase-1-Mediated Pyroptosis via the Induction of ROS/NF-κB/NLRP3/GSDMD Signal Axis in Non-Small Cell Lung Cancer. *Cancers (Basel)* 12 (1), 193. doi:10.3390/cancers12010193

Wang, L.-L., Jin, X.-H., Cai, M.-Y., Li, H.-G., Chen, J.-W., Wang, F.-W., et al. (2018). AGBL2 Promotes Cancer Cell Growth through IRGM-Regulated Autophagy and Enhanced Aurora A Activity in Hepatocellular Carcinoma. *Cancer Lett.* 414, 71–80. doi:10.1016/j.canlet.2017.11.003

Wang, Y., Gao, W., Shi, X., Ding, J., Liu, W., He, H., et al. (2017). Chemotherapy Drugs Induce Pyroptosis through Caspase-3 Cleavage of a Gasdermin. *Nature* 547 (7661), 99–103. doi:10.1038/nature22393

Williams, C. S., Zhang, B., Smith, J. J., Jayagopal, A., Barrett, C. W., Pino, C., et al. (2011). BVES Regulates EMT in Human Corneal and colon Cancer Cells and Is

Silenced via Promoter Methylation in Human Colorectal Carcinoma. *J. Clin. Invest.* 121 (10), 4056–4069. doi:10.1172/jci44228

Wu, Y., Wang, L., and Zhang, Y. (2004). Dendritic Cells as Vectors for Immunotherapy of Tumor and its Application for Gastric Cancer Therapy. *Cell Mol Immunol* 1 (5), 351–356.

Xie, R., Liu, J., Yu, X., Li, C., Wang, Y., Yang, W., et al. (2019). ANXA2 Silencing Inhibits Proliferation, Invasion, and Migration in Gastric Cancer Cells. *J. Oncol.* 2019, 4035460. doi:10.1155/2019/4035460

Yamato, A., Soda, M., Ueno, T., Kojima, S., Sonehara, K., Kawazu, M., et al. (2015). Oncogenic Activity of BIRC 2 and BIRC 3 Mutants Independent of Nuclear factor-κB-Activating Potential. *Cancer Sci.* 106 (9), 1137–1142. doi:10.1111/cas.12726

Yu, J., Li, S., Qi, J., Chen, Z., Wu, Y., Guo, J., et al. (2019). Cleavage of GSDME by Caspase-3 Determines Lobaplatin-Induced Pyroptosis in Colon Cancer Cells. *Cell Death Dis.* 10 (3), 193. doi:10.1038/s41419-019-1441-4

Zhao, J., Li, X., Liu, L., Cao, J., Goscinski, M. A., Fan, H., et al. (2019). Oncogenic Role of Guanylate Binding Protein 1 in Human Prostate Cancer. *Front. Oncol.* 9, 1494. doi:10.3389/fonc.2019.01494

Zheng, P., Chen, L., Yuan, X., Luo, Q., Liu, Y., Xie, G., et al. (2017). Exosomal Transfer of Tumor-Associated Macrophage-Derived miR-21 Confers Cisplatin Resistance in Gastric Cancer Cells. *J. Exp. Clin. Cancer Res.* 36 (1), 53. doi:10.1186/s13046-017-0528-y

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

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

Copyright © 2022 Liang, Fan, Liang 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fgene.2022.767166

A Novel Inflammatory Response–Related Gene Signature Improves High-Risk Survival Prediction in Patients With Head and Neck Squamous Cell Carcinoma

Yanxun Han[1,2,3†], *Zhao Ding*[1,2,3†], *Bangjie Chen*[4], *Yuchen Liu*[1,2,3] *and Yehai Liu*[1*]

[1]Department of Otorhinolaryngology, Head and Neck Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Anhui Medical University, Hefei, China, [3]Graduate School of Anhui Medical University, Hefei, China, [4]The First Affiliated Hospital of Anhui Medical University, Hefei, China

OPEN ACCESS

Edited by:
Xiangqian Guo,
Henan University, China

Reviewed by:
Daiwei Wan,
The First Affiliated Hospital of Soochow University, China
Qing Tang,
University of Washington, United States
Zhihui Liu,
Stanford University, United States

***Correspondence:**
Yehai Liu
liuyehai@ahmu.edu.cn

[†]These authors have contributed equally to this work

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

Received: 30 August 2021
Accepted: 07 March 2022
Published: 11 April 2022

Citation:
Han Y, Ding Z, Chen B, Liu Y and Liu Y (2022) A Novel Inflammatory Response–Related Gene Signature Improves High-Risk Survival Prediction in Patients With Head and Neck Squamous Cell Carcinoma. Front. Genet. 13:767166. doi: 10.3389/fgene.2022.767166

Background: Head and neck squamous cell carcinoma (HNSCC) is a highly prevalent and malignant tumor that is difficult to effectively prognosticate outcomes. Recent reports have suggested that inflammation is strongly related to tumor progression, and several biomarkers linked to inflammation have been demonstrated to be useful for making a prognosis. The goal of this research was to explore the relevance between the inflammatory-related genes and HNSCC prognosis.

Methods: The clinical information and gene expression data of patients with HNSCC were acquired from publicly available data sources. A multigene prognostic signature model was constructed in The Cancer Genome Atlas and verified in the Gene Expression Omnibus database. According to the risk score calculated for each patient, they were divided into low- and high-risk groups based on the median. The Kaplan–Meier survival curve and receiver operating characteristic curve were applied to determine the prognostic value of the risk model. Further analysis identified the independent prognostic factors, and a prognostic nomogram was built. The relationship between tumor immune infiltration status and risk scores was investigated using Spearman correlation analysis. Finally, to confirm the expression of genes in HNSCC, quantitative real-time polymerase chain reaction (qRT-PCR) was performed.

Results: A prognostic model consisting of 14 inflammatory-related genes was constructed. The samples with a high risk had an apparently shorter overall survival than those with a low risk. Independent prognostic analysis found that risk scores were a separate prognostic factor in HNSCC patients. Immune infiltration analysis suggested that the abundance of B cells, CD8 T cells, M2 macrophages, myeloid dendritic cells, and monocytes in the low-risk group was higher, while that of M0, M1 macrophages, and resting NK cells was obviously higher in the high-risk group. The risk scores were related to chemotherapeutic sensitivity and the expression of several immune checkpoint genes. Moreover, CCL22 and IL10 were significantly higher in HNSCC tissues, as determined by qRT-PCR.

Conclusion: Taken together, we constructed a novel inflammatory response–related gene signature, which may be used to estimate outcomes for patients with HNSCC and may be developed into a powerful tool for forecasting the efficacy of immunotherapeutic and chemotherapeutic drugs for HNSCC.

Keywords: head and neck squamous cell carcinoma, inflammatory response, prognosis, chemotherapy sensitivity, immune cell infiltration

INTRODUCTION

Head and neck cancers (HNCs) are a group of heterogeneous diseases and include cancers occurring in the tongue, mouth, pharynx, and larynx and other oral cavities (Haddad and Shin, 2008). HNC ranks sixth in terms of the global cancer incidence rate and accounts for about 4% of all cancers. In the United States in 2020, it was estimated that HNC accounted for about 65,630 new oral cavity, pharynx, and larynx cancer cases and 14,500 deaths (Ferlay et al., 2019; Siegel et al., 2020). According to histopathologic classification, about 90% of HNCs are squamous cell carcinoma (Ma et al., 2020). The specific etiology of head and neck squamous cell carcinoma (HNSCC) is unknown. Previous studies have suggested that genetic factors, smoking, alcohol consumption, and viral infections [e.g., Epstein–Barr virus (EBV) and human papillomavirus (HPV)] were the most common risk factors (Mehanna et al., 2013; Maasland et al., 2014; Turunen et al., 2017). Surgery, radiation, chemotherapy, and molecular targeted medicines are all options for HNSCC treatment. Although there has been an improvement in the therapeutic modalities available, the clinical prognosis has not evolved significantly. Due to the high incidence of local recurrence and distant metastases, the 5-year survival rate of HNSCC patients is lower than 50% (Chow, 2020). Identification of novel molecular biomarkers is significant for improving the prognosis of these patients because of the multiple anatomical locations, molecular heterogeneity, and diverse etiology of HNSCC.

Since inflammation may contribute to the development of various cancers, the biological processes between inflammation and carcinogenesis have been the focus of research (Greten and Grivennikov, 2019). Chronic inflammation induces tumor angiogenesis and DNA damage, and promotes tumor proliferation and metastasis while preventing apoptosis (Grivennikov et al., 2010). Some studies have revealed a partial association between cancer and inflammatory markers by analyzing peripheral blood–related parameters. For example, many inflammatory response features like leukocytosis, hypoproteinemia, and hyperfibrinogenemia were confirmed in HNSCC patients (Cai et al., 2020; Zhou et al., 2019). In recent years, the use of systemic hematology inflammatory markers as prognostic factors in malignancy has attracted increasing attention. Neutrophil, lymphocyte, monocyte, and platelet counts, both individually and expressed as ratios, showed significant overall survival (OS) prognostic ability and were independent of existing prognostic factors for HNSCC (Zhou et al., 2020; Rassouli et al., 2015). Mcmillan reported that the Glasgow Prognostic Score, consisting of peripheral blood inflammatory markers, was an independent prognostic factor in cancer patients (Mcmillan, 2013). This spurred the researchers to further study the association between the comprehensive inflammation score and cancer prognosis.

In addition to blood parameters, some inflammatory response-related genes have also been used to assess distant metastasis of cancer (Clatot et al., 2014). The C-C motif chemokine ligand 22 (CCL22) and C-C chemokine receptor 7 (CCR7) have been reported to be highly expressed in tongue squamous cell carcinoma (TSCC) and associated with poor prognosis such as lymph node metastasis (Kimura et al., 2021; Qi Wang et al., 2021). Gene expression levels of coagulation factor III (F3) contribute to the stratification and prognosis prediction of prostate cancer patients (Peng et al., 2016). Interleukin 10 receptor subunit alpha (IL10RA) regulates tumor immune responses and is implicated in the pathogenesis of colorectal cancer. Interleukin 2 receptor subunit beta (IL2RB) polymorphism is associated with lung cancer susceptibility and involved in lung cancer progression (Zadka et al., 2018; Jia et al., 2019). Nevertheless, the association between the inflammatory response–related genes and the outcomes of HNSCC needs to be further investigated.

In this study, the mRNA profile and relevant clinical characteristics of patients with HNSCC were assessed through public databases. Univariate and multivariate Cox regression algorithms were employed to screen 14 genes in The Cancer Genome Atlas (TCGA) database, namely, CCL22, CCR7, CD48 molecule (CD48), F3, heparin-binding EGF-like growth factor (HBEGF), interleukin 10 (IL10), IL10RA, IL2RB, LCK proto-oncogene, Src family tyrosine kinase (LCK), phosphoinositide-3-kinase regulatory subunit 5 (PIK3R5), presenilin 1 (PSEN1), TIMP metallopeptidase inhibitor 1 (TIMP1), TNF alpha–induced protein 6 (TNFAIP6), and TNF receptor superfamily member 1B (TNFRSF1B). Then, the prognostic characteristics were established, and the Gene Expression Omnibus (GEO) database was used for validation. We then evaluated the genes' prognostic prediction efficiency in patients with HNSCC, as well as their relationships to the immune infiltrate types and the chemotherapeutic sensitivity. Finally, we performed a quantitative real-time polymerase chain reaction (qRT-PCR) analysis to preliminarily verify the mRNA expression levels of these genes in HNSCC.

METHODS

Data Collection

Clinical information related to HNSCC and the RNA sequencing data of gene expression were obtained from TCGA (https://

portal.gdc.cancer.gov/). The GSE41613 chip was acquired from the GEO (http://www.ncbi.nlm.nih.gov/geo/). Through the Molecular Signatures database (MSigDB), we obtained 200 inflammatory response-related genes. The clinicopathological data, including survival information, age, sex, grade, and clinical stage, were collected. All data used in our articles are publicly available in public databases, and patients without survival information were excluded from further analysis. This work is in accordance with the Helsinki Declaration.

Building and Validating a Prognostic Inflammatory Response–Related Gene Signature

The Akaike information criterion (AIC) (Akaike, 1974) was founded and developed by Japanese statistician Hirotugu Akaike and is a standard to measure the goodness of fit of statistical models. The smaller the AIC value, the better the model. By univariate Cox analysis, we preliminarily identified the inflammatory response–related genes correlated with prognosis in TCGA database. Subsequently, the significant genes obtained from univariate analysis were applied to multivariable Cox algorithms, and the prognostic feature model was established based on the minimum AIC. The signature risk score was generated in light of each gene's expression level and its relevant coefficient. The detailed equation was: risk scores = (gene-1 expression × corresponding coefficient) + (gene-2 expression × corresponding coefficient) + ... + gene-n expression × corresponding coefficient). The patients were then assigned to high- and low-risk groups based on their median risk score, and the Kaplan–Meier (K-M) curve was utilized to examine the OS differences between the two groups. To evaluate the predictive efficacy of the prognostic risk scoring model, the time-dependent receiver operating characteristics (ROC) curve was drawn, and the area under the curve (AUC) was calculated. Additionally, we compared the AUC of Wu et al. (2021) and Jiang et al. (2021) models. A univariate analysis and a multivariate analysis were performed to explore the independent prognostic significance of age, sex, grade, clinical stage, and risk score. After the above process, the best potential prognostic factors were selected and converted into visual nomograms of 1-, 3-, and 5-year OS. To estimate the nomogram's clinical utility, decision curve analysis (DCA) was utilized. Finally, the above genes and prognostic signatures were tested and verified in the GEO data set.

Investigation of Tumor Immune Infiltration

To comprehend the relevance among prognostic risk models and immune cell infiltration levels, we performed seven noncontroversial methods for calculating the immune infiltration status of HNSCC patients, including XCELL (Aran, 2020), TIMER (Li et al., 2020), QUANTISEQ (Plattner et al., 2020), MCPCOUNTER (Dienstmann et al., 2019), EPIC (Racle et al., 2017), CIBERSORT-ABS (Tamminga et al., 2020), and CIBERSORT (Chen et al., 2018). The obtained correlation coefficients were plotted as lollipop diagrams. Furthermore, to study the correlations between immune checkpoint expression levels and prognostic risk scores, the level of immune checkpoint genes among the two groups was compared by the Wilcoxon test.

Chemotherapy Sensitivity Analysis

The half inhibitory concentration (IC50) represents the level of drug required to inhibit tumor cells by 50%. The lower the IC50, the more sensitive the drug. In the TCGA cohort, to assess whether this risk model can guide chemotherapy in patients with HNSCC, we calculated the IC50s of commonly used chemotherapeutics with the "pRRophetic" package. The American Joint Committee on Cancer (AJCC) guidelines suggest anticancer drugs, such as cisplatin, docetaxel, lapatinib, methotrexate, and rapamycin, for use in the treatment of HNSCC. The Wilcoxon measure was used to analyze the discrepancies of IC50s among the high- and low-risk groups, and the outcomes were plotted as a box chart.

Verification by Quantitative Real-Time Polymerase Chain Reaction

Ten pairs of HNSCC tissues and adjacent noncancerous tissues from the First Affiliated Hospital of Anhui Medical University, approved by the ethics committee, were used in this study. Each participant filled out an informed consent form. The samples were from HNSCC patients who had undergone surgical resection but had not previously received chemotherapy or radiation. The total RNA was extracted from the tissue using TRIzol technology and stored in liquid nitrogen until needed. The Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) was used to reverse transcribe the extracted RNA into cDNA for further analysis, while the concentration of the cDNA was measured to ensure that it meets the standard of use. **Supplementary Table S1** shows the primer sequences for PCR. As an internal control, GAPDH was used, and the samples were recorded. Transcriptome levels were collected and differences were compared using the Wilcoxon test.

Statistical Analysis

The prognostic genes were selected by univariate Cox regression analysis. The K-M method was performed to analyze the distinctions of the OS between high- and low-risk groups, and the statistical significance was obtained by log-rank test. Multivariate Cox regression analyses were used to identify the independent risk factors for the outcomes. The prognostic performance of risk scores on survival prediction was evaluated by ROC curve, and the AUC value was calculated. Spearman correlation analysis was used to test the relevance among the risk score of prognostic models and the immune cell levels. IC50 between the two groups was compared by the Wilcoxon test. The differences in gene expressions between the HNSCC tissues and adjacent non-cancer tissues were tested by Wilcoxon test. All graphics and statistical

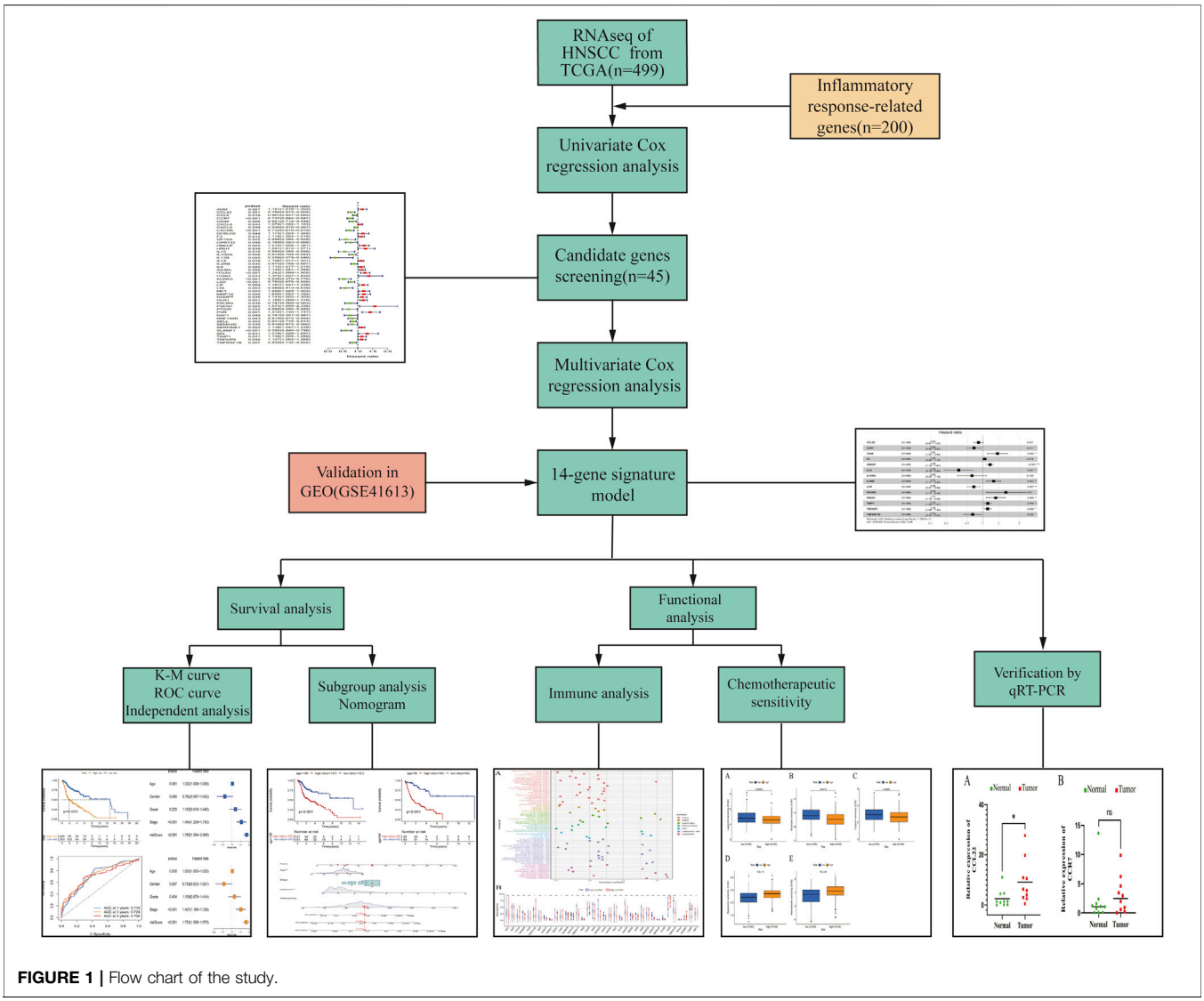


FIGURE 1 | Flow chart of the study.

analyses were completed in R (version 4.1.0) and GraphPad Prism (version 9.0.0).

RESULTS

Distinguishing Prognostic Inflammatory Response–Related Genes in Head and Neck Squamous Cell Carcinoma

We summarized the detailed flow chart of **Figure 1**. The study population included TCGA and GSE41613 data sets. To investigate the potential roles of prognostic inflammatory response-related genes in HNSCC, a total of 200 inflammatory response–related genes were obtained through the MSigDB (**Supplementary Table S2**). The univariate Cox proportional hazards regression method suggested that 45 of these genes were related to OS and were preserved as prognostic indicators. Twenty-two genes revealed positive coefficients that signified that a higher expression of these genes was accompanied by shorter survival, and another 23 genes had negative coefficients, which showed that high levels of these genes were associated with longer OS (**Table 1**).

Establishment and Verification of Prognostic Inflammatory Response–Related Signature

The relevant data of the above 45 genes were further processed by multivariate analysis. Fourteen genes with a minimum AIC value of 2319.65 were selected and estimated by the R software to develop a prognostic risk model (**Figure 2A**). The corresponding coefficients are shown in **Supplementary Table S3**. CD48, F3, HBEGF, IL2RB, PIK3R5, PSEN1, TIMP1, and TNFAIP6 were considered as risk factors with hazard ratio (HR) values greater than 1, whereas the remaining six genes (CCL22, CCR7, IL10, IL10RA, LCK, and TNFRSF1B) were considered protective

TABLE 1 | The results of the association between 45 inflammatory response–related genes expression and OS from univariate cox regression analysis.

Gene-ID	HR	95%CI	*p*-value
ADM	1.151	1.016–1.303	0.027
CCL22	0.780	0.673–0.905	0.001
CCL5	0.901	0.827–0.982	0.018
CCR7	0.737	0.642–0.847	0.000
CD48	0.821	0.712–0.946	0.006
CXCL8	1.079	1.002–1.163	0.044
CXCL9	0.936	0.879–0.997	0.039
CXCR6	0.732	0.610–0.878	0.001
DCBLD2	1.173	1.004–1.369	0.044
F3	1.116	1.024–1.216	0.012
GP1BA	0.598	0.385–0.928	0.022
GPR132	0.769	0.593–0.998	0.048
HBEGF	1.210	1.059–1.381	0.005
HRH1	1.261	1.013–1.571	0.038
IL10	0.565	0.356–0.896	0.015
IL10RA	0.815	0.704–0.942	0.006
IL12B	0.258	0.079–0.846	0.025
IL1A	1.106	1.017–1.201	0.018
IL2RB	0.870	0.768–0.987	0.030
IL6	1.112	1.017–1.215	0.020
INHBA	1.149	1.051–1.256	0.002
ITGA5	1.242	1.099–1.405	0.001
ITGB3	1.372	1.027–1.835	0.033
KCNA3	0.536	0.370–0.776	0.001
LCK	0.780	0.679–0.896	0.000
LIF	1.181	1.041–1.339	0.009
LTA	0.589	0.413–0.839	0.003
MET	1.202	1.029–1.403	0.020
MMP14	1.206	1.023–1.422	0.026
NAMPT	1.143	1.003–1.303	0.045
OLR1	1.185	1.068–1.316	0.001
PIK3R5	0.747	0.568–0.983	0.038
PSEN1	1.573	1.059–2.336	0.025
PTGIR	0.688	0.480–0.986	0.042
PVR	1.415	1.145–1.747	0.001
RAF1	0.741	0.551–0.997	0.048
RNF144B	0.819	0.675–0.994	0.043
SELL	0.841	0.749–0.944	0.003
SEMA4D	0.816	0.673–0.990	0.039
SERPINE1	1.139	1.047–1.239	0.002
SLAMF1	0.592	0.440–0.796	0.001
SRI	1.316	1.026–1.687	0.031
TIMP1	1.138	1.005–1.288	0.041
TNFAIP6	1.127	1.002–1.268	0.046
TNFRSF1B	0.835	0.732–0.952	0.007

factors, with HR values less than 1. The risk score was measured in the light of the expression values of the genes as follows: score = (−0.184 × CCL22 expression level) + (−0.386 × CCR7 expression level) + (0.649 × CD48 expression level) + (0.083 × F3 expression level) + (0.322 × HBEGF expression level) + (−1.055 × IL10 expression level) + (−0.474 × IL10RA expression level) + (0.479 × IL2RB expression level) + (−0.392 × LCK expression level) + (1.014 × PIK3R5 expression level) + (0.570 × PSEN1 expression level) + (0.223 × TIMP1 expression level) + (0.215 × TNFAIP6 expression level) + (−0.442 × TNFRSF1B expression level). Using the median risk score as the cutoff point, all patients in the TCGA group (n = 499) were separated into a high-risk group (n = 249) and a low-risk group (n = 250). The survival curve illustrated that the patients with low risk had an obviously longer OS than those in the high-risk group ($p < 0.001$; **Figure 2B**). To evaluate the model's performance, the ROC curve was generated, and the AUC was 0.715 at 1 year, 0.724 at 3 years, and 0.700 at 5 years (**Figure 2C**). To further test the stability of the signature, we evaluated it in the GEO cohort. Consistent with the conclusions of the TCGA cohort, the patients with low risk tended to have a longer survival time (**Figure 2D**). Furthermore, a similar result was obtained in the ROC curves, and the AUCs at 1, 3, and 5 years of the inflammatory response–related signature were 0.733, 0.754, and 0.745, respectively (**Figure 2E**). In addition, to verify the optimality of the risk model, we not only drew the 1-, 3-, and 5-year ROC curves but also compared the 3-year ROC curve with Wu et al. (2021) and Jiang et al. (2021) models (**Figure 2F**). The results demonstrated that the inflammatory response–related signature had better performance.

Association Between Prognostic Models and Clinicopathological Characteristics

To investigate the relationship between the prognostic model and clinical features, we first examined the distribution of risk values among clinical features. As shown in **Figures 3A–C**, risk scores did not differ substantially when age, sex, and grade were considered. By contrast, stages III–IV had a much higher risk score than stages I–II (p = 0.019) (**Figure 3D**). These findings suggested that higher risk scores were associated with higher malignancy in HNSCC.

Furthermore, to investigate the prognostic value of the model in HNSCC patients after stratification according to clinicopathological variables, we divided the patients into subsets based on their age, gender, grade, and stage to plot the K-M survival curve. The samples were classified into eight subgroups: younger (≤65 years) and older (>65 years), male and female, earlier grade and advanced grade, and earlier stage and advanced stage. We chose the previous cutoff value, and patients in each subgroup were further assigned to high- and low-risk groups. For all the different subgroups, those in the high-risk group had an obviously shorter OS than their low-risk counterparts (**Figures 4A–H**). The results suggested that the risk model can anticipate the outcomes of patients with HNSCC without considering clinicopathological subgroups. In sum, the risk signature exerts critical roles in determining the prognosis of patients with HNSCC.

Independence of the Prognostic Model

To clarify whether the signature was an independent prognostic variable for OS, the risk score and other clinical features (age, gender, grade, and clinical stage) were analyzed. Univariate analysis suggested that the risk score was greatly related to OS (TCGA data set: HR = 1.789, 95% CI = 1.594–2.009, $p < 0.001$; GEO data set: HR = 1.468, 95% CI = 1.246–1.729, $p < 0.001$; **Figures 5A,B**). Further multivariate Cox analysis excluded other confounding factors, and the results demonstrated that the risk score was still an independent prognostic tool for the OS in the TCGA data set (HR = 1.753, 95% CI = 1.556–1.975, $p < 0.001$) and the GEO data set (HR = 1.628, 95% CI = 1.351–1.962, $p < 0.001$; **Figures**

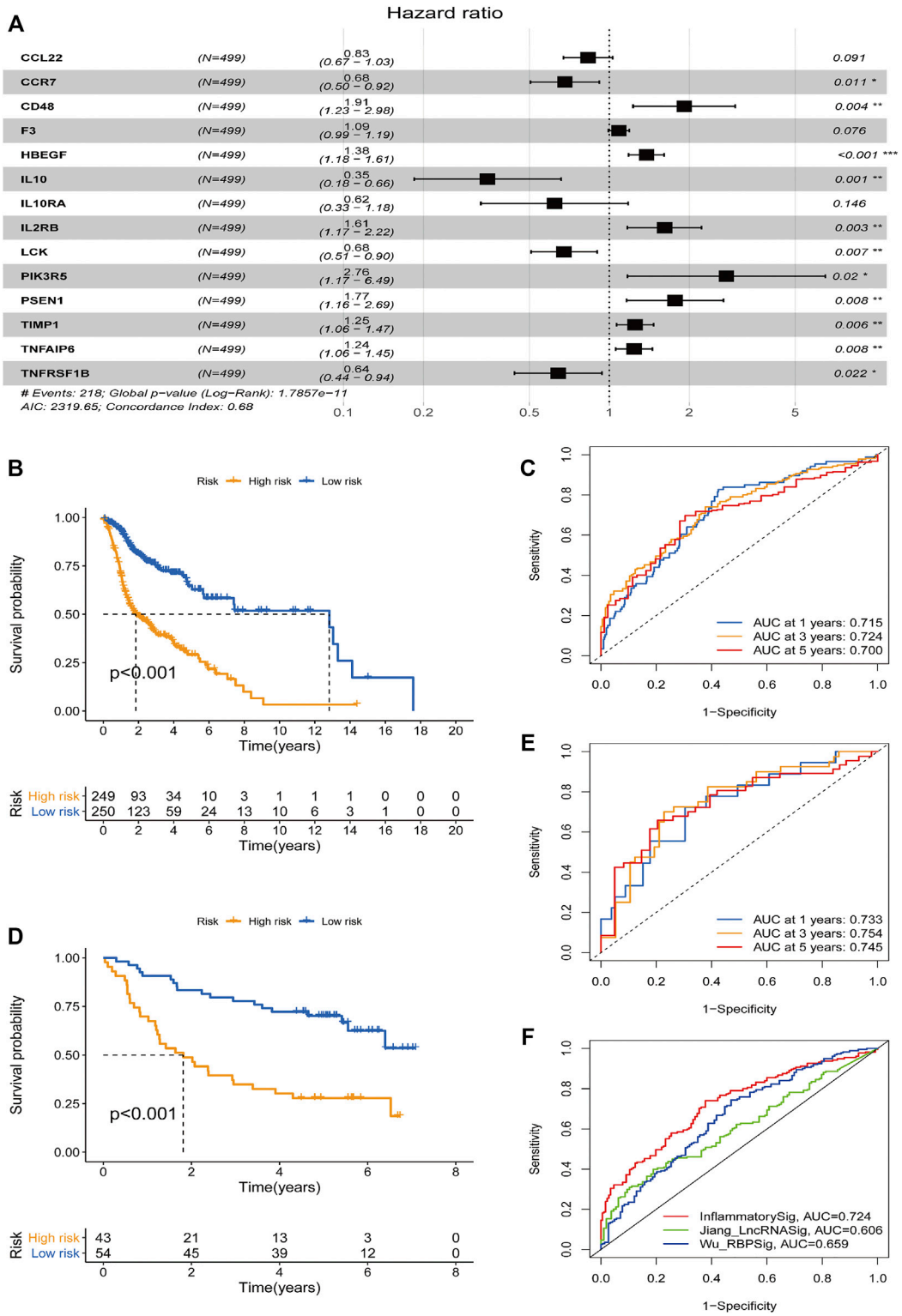


FIGURE 2 | Construction and prognostic analysis of the 14-gene signature model. **(A)** Forest plot showing 14 prognostic inflammatory response-related genes in head and neck squamous cell carcinoma based on a minimum Akaike information criterion (2319.65) value multivariate Cox results. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. HR < 1 means protective factors, HR > 1 means risk factors. Kaplan–Meier curve analysis of the inflammatory response–related gene signature in The Cancer Genome Atlas (TCGA) data set **(B)** and the Gene Expression Omnibus (GEO) data set **(D)**. Receiver operating characteristics (ROC) curve analysis of the inflammatory response–related gene signature of 1-, 3-, and 5-year in the TCGA data set **(C)** and GEO data set **(E)**. **(F)** A comparison of the 3-year ROC curve with Wu's and Jiang's models.

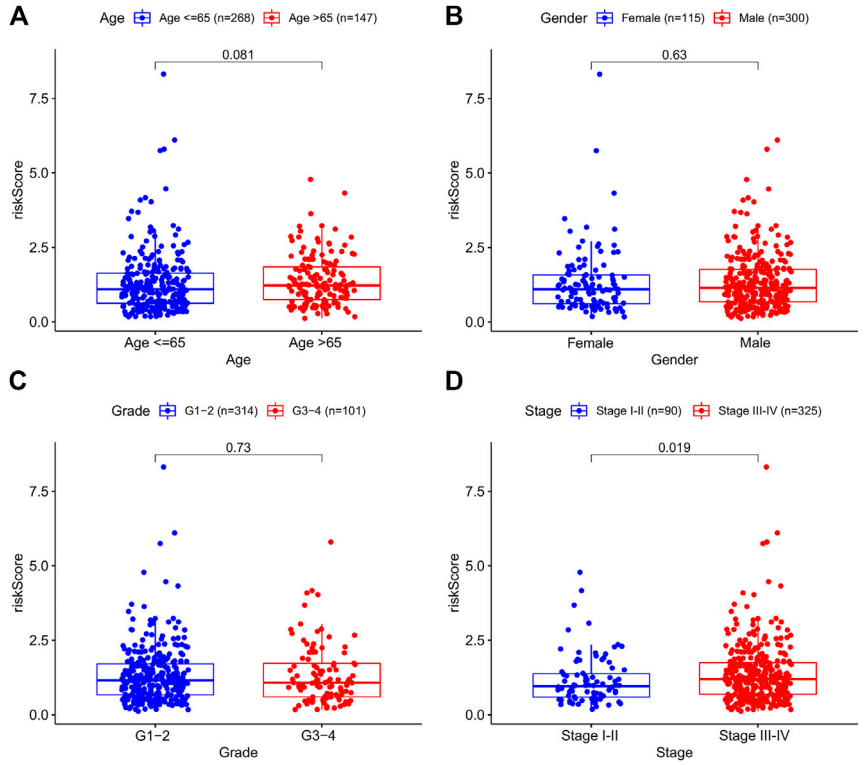


FIGURE 3 | The correlations between the risk model and clinical factors. **(A)** Age. **(B)** Gender. **(C)** Grade. **(D)** Stage.

5C,D). In addition, age and clinical stage were also confirmed as independent predictors for OS (**Figure 5**).

Nomogram Construction and Verification

A nomogram with an integrated prognostic model and other clinical parameters was developed to predict the OS of HNSCC patients and to assist doctors in providing better therapy (**Figure 6A**). Then, a calibration curve was created to determine the nomogram's reliability (**Figure 6B**). Moreover, DCA was performed to analyze the net benefits of none, all, age, AJCC stage, risk score, and the nomogram to verify the clinical utility of the nomogram. As shown in **Figure 6C**, the nomogram had a better potential for clinical utility compared with the other five groups. Thus, the risk model provided effective prognosis predictive value and delivered some net benefits that may be useful for patients with HNSCC.

Association of Risk Score With Immune Infiltrating Cells

To investigate the efficiency of inflammatory response–related genes on the status of the tumor microenvironment, seven algorithms were applied to investigate the correlations between the level of immune cells and the model-predicted risk score. After Spearman correlation analysis was conducted, we found that the risk score was positively linked to M0, M1 macrophages, and resting NK cells infiltrating, while being negatively associated with B cells, CD8 T cells, M2 macrophages, myeloid dendritic cells, and monocytes infiltrating in the HNSCC samples (**Figure 7A**; **Supplementary Table S4**). We then detected the expression levels of immune checkpoints among samples with different risk scores, and differences were further found in the expressions of some immune checkpoint genes between the two groups (**Figure 7B**). The outcomes revealed that these genes might be prospective indicators for the regulation of immune activity in the immune microenvironment and may play an important role in checkpoint inhibitor-based immunotherapies.

Correlation Between the Prognostic Signature and Chemosensitivity

Chemotherapy is an important part of the comprehensive treatment of HNSCC. Therefore, we sought to determine whether there is a correlation between risk scores and patient sensitivity to commonly used chemotherapeutic agents. The results showed that drugs such as lapatinib ($p < 0.001$; **Figure 8A**), docetaxel ($p < 0.001$; **Figure 8B**), and cisplatin ($p < 0.001$; **Figure 8C**) had lower IC50s in the high-risk group, while rapamycin ($p < 0.001$; **Figure 8D**) and methotrexate ($p < 0.001$; **Figure 8E**) were lower in the low-risk group. In short, high-risk groups were more sensitive to lapatinib, docetaxel, and cisplatin, while low-risk groups responded better to rapamycin and methotrexate. Therefore, this model shows great potential for predicting chemosensitivity and may help clinicians choose the best chemotherapy regimen.

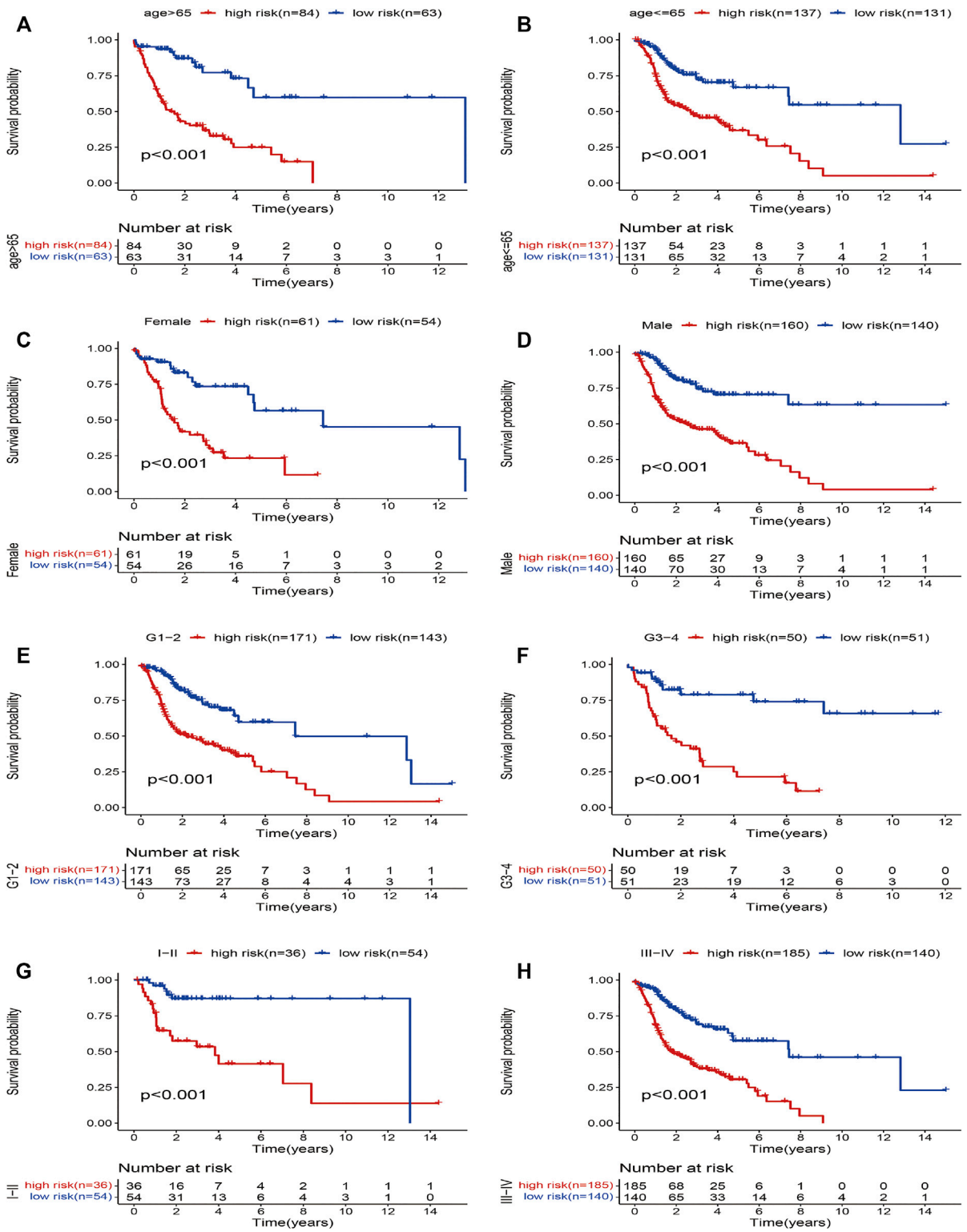


FIGURE 4 | Kaplan–Meier survival curves for the high- and low-risk groups stratified by clinical factors. **(A)** age >65 years, **(B)** age ≤65 years, **(C)** gender (female), **(D)** gender (male), **(E)** grades 1–2, **(F)** grades 3–4, **(G)** stages I–II, and **(H)** stages III–IV.

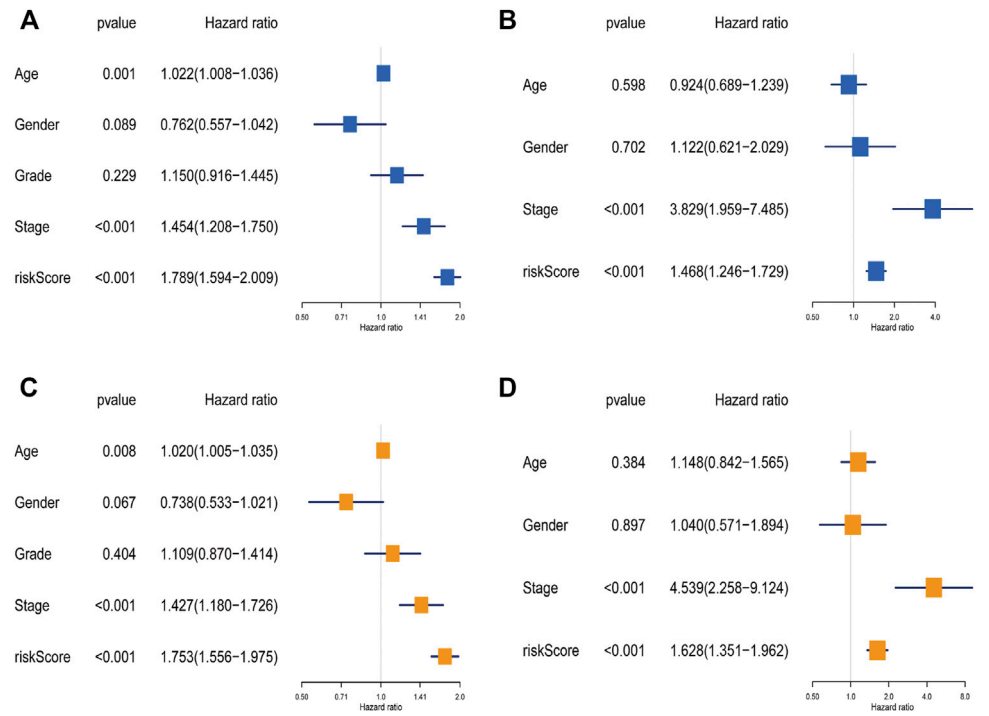


FIGURE 5 | The independence identification of the risk model. Univariate cox regression analysis and multivariate cox regression analysis were performed in The Cancer Genome Atlas data set **(A,C)** and the Gene Expression Omnibus data set **(B,D)**.

Expression Levels of C-C Motif Chemokine Ligand 22 and Interleukin 10 was Upregulated in Quantitative Real-Time Polymerase Chain Reaction

qRT-PCR was used to examine the relative mRNA expression levels of these genes in 10 pairs of HNSCC and adjacent non-cancer tissues. As shown in **Figure 9**, the expression of CCL22 and IL10 was much higher in HNSCC tissues than in adjacent tissues. However, no significant differences were observed in the mRNA expression levels of the remaining 12 genes between HNSCC tissues and normal tissues (**Supplementary Figure S1**). This may be related to the small number of samples we tested.

DISCUSSION

With the progress of medical technology, various treatments for HNSCC have been fully developed. However, due to the lack of reliable biomarkers, we are often unable to effectively predict the therapeutic efficacy of HNSCC treatments. A growing number of researchers have shown that inflammation response plays a significant role in the pathophysiology and progression of HNSCC and is associated with patient outcomes (Rassouli et al., 2015; Zhou et al., 2020). Serum indicators linked with inflammation, such as the neutrophil/lymphocyte ratio, platelet/lymphocyte ratio, and lymphocyte/monocyte ratio, have been shown to be effective at forecasting the prognosis of HNSCC in recent years (Zhou et al., 2019; Valdes et al., 2020). However, inflammatory response-related genes, as prognostic predictors for HNSCC, have not been developed. According to the existing literature, an RNA-binding protein-based signature (Wu et al., 2021) and a three-lncRNA signature (Jiang et al., 2021) predicted the 3-year OS for HNSCC with AUCs of 0.659 and 0.606, respectively. The inflammatory response-related gene signature built in the present study shows better validity than the above gene signatures. Moreover, another important advantage of our research is that we evaluated the value of the risk models in predicting immune cell infiltration and chemotherapy sensitivity, further increasing the model's clinical utility and providing potential biomarkers for clinical therapeutics.

In our research, we employed RNA expression data and 200 inflammatory response-related genes to systematically analyze OS in patients with HNSCC. In the TCGA data set, we built a prognostic signature containing 14 inflammatory response-related genes to establish risk stratification and predict clinical outcomes. Patients with HNSCC were compartmentalized into low-risk and high-risk groups using the median risk score. We found that the low-risk group was significantly correlated to a longer OS, which was effectively validated in the GEO database. Further analysis revealed that the correlation between the risk models and OS in HNSCC patients was independent of clinicopathological subgroups. In addition, independent analysis showed that the prognostic risk model was an independent prognostic factor for HNSCC. Then,

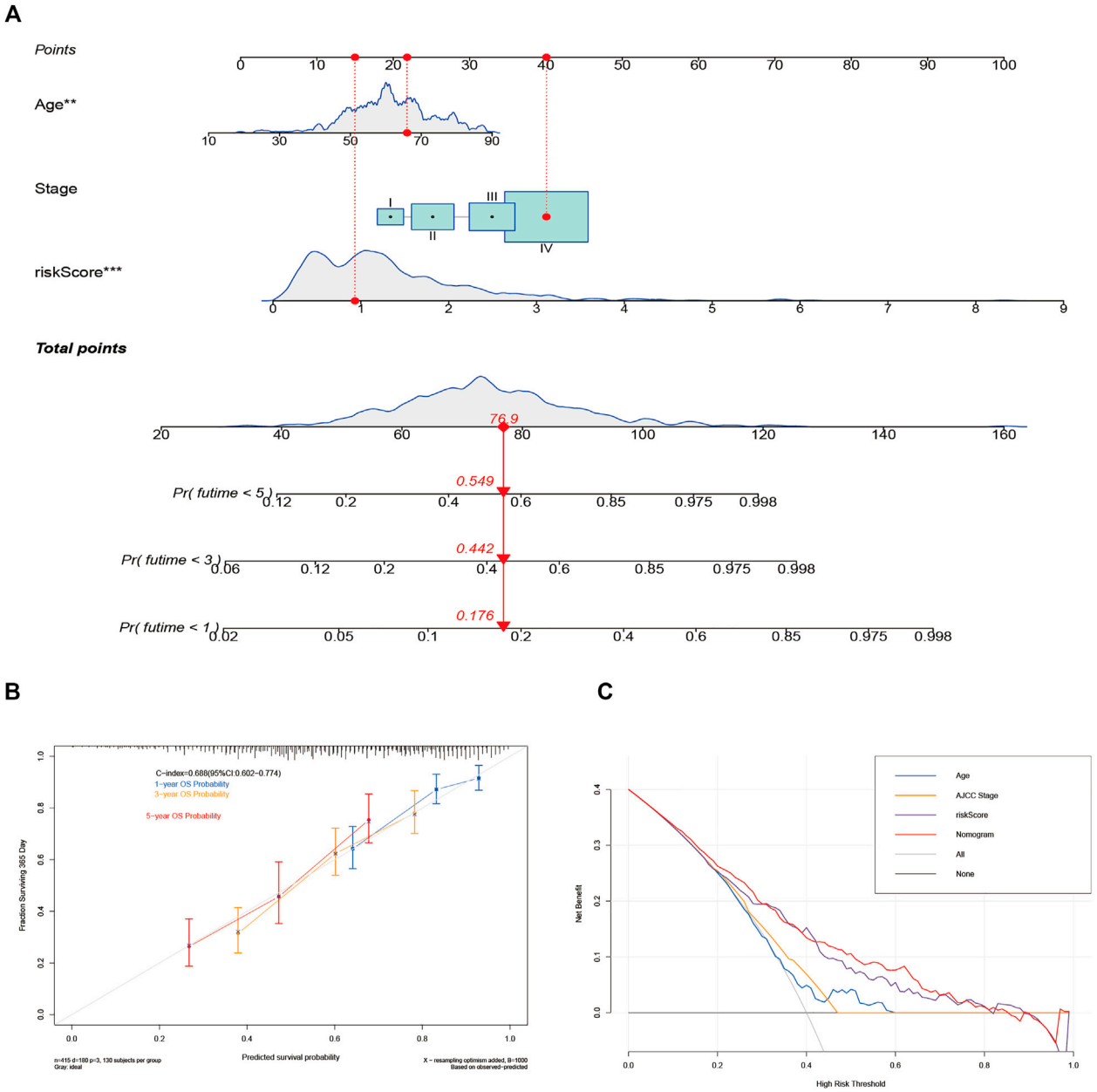


FIGURE 6 | Construction and validation of the nomogram for predicting the overall survival (OS) of patients with head and neck squamous cell carcinoma. **(A)** The nomogram for predicting the OS of patients at 1, 3, and 5 years. **(B)** Calibration curves of the nomogram for OS prediction at 1, 3, and 5 years. **(C)** Decision curve analysis for the prediction of 1-, 3-, and 5-year OS.

by combining the inflammatory prognostic signature with other prognostic variables, we developed a simple-to-use prediction nomogram model, which was helpful in predicting 1-, 3-, and 5-year OS and may prove beneficial in the clinical treatment of HNSCC patients.

The inflammatory response–related gene signature consisted of 14 genes (CCL22, CCR7, CD48, F3, HBEGF, IL10, IL10RA, IL2RB, LCK, PIK3R5, PSEN1, TIMP1, TNFAIP6, and TNFRSF1B). Previous research has shown that these 14 genes play a role in cancer progression. CCL22 is one of the several Cys-Cys (CC) cytokine genes, mainly produced by M2-like tumor-associated macrophages in the tumor microenvironment. CCL22 plays an important role in regulating regulatory T-cell migration by binding to its receptor CCR4 (Rohrle et al., 2020). Studies have found that high expression of CCL22 in the tumor microenvironment is closely related to lymph node recurrence in TSCC and poor prognosis of cervical cancer (Kimura et al., 2021; Wang et al., 2019). Our study also confirmed this, as we found that CCL22 mRNA levels were significantly upregulated in HNSCC tissues by qRT-PCR. Found in a variety of lymphoid

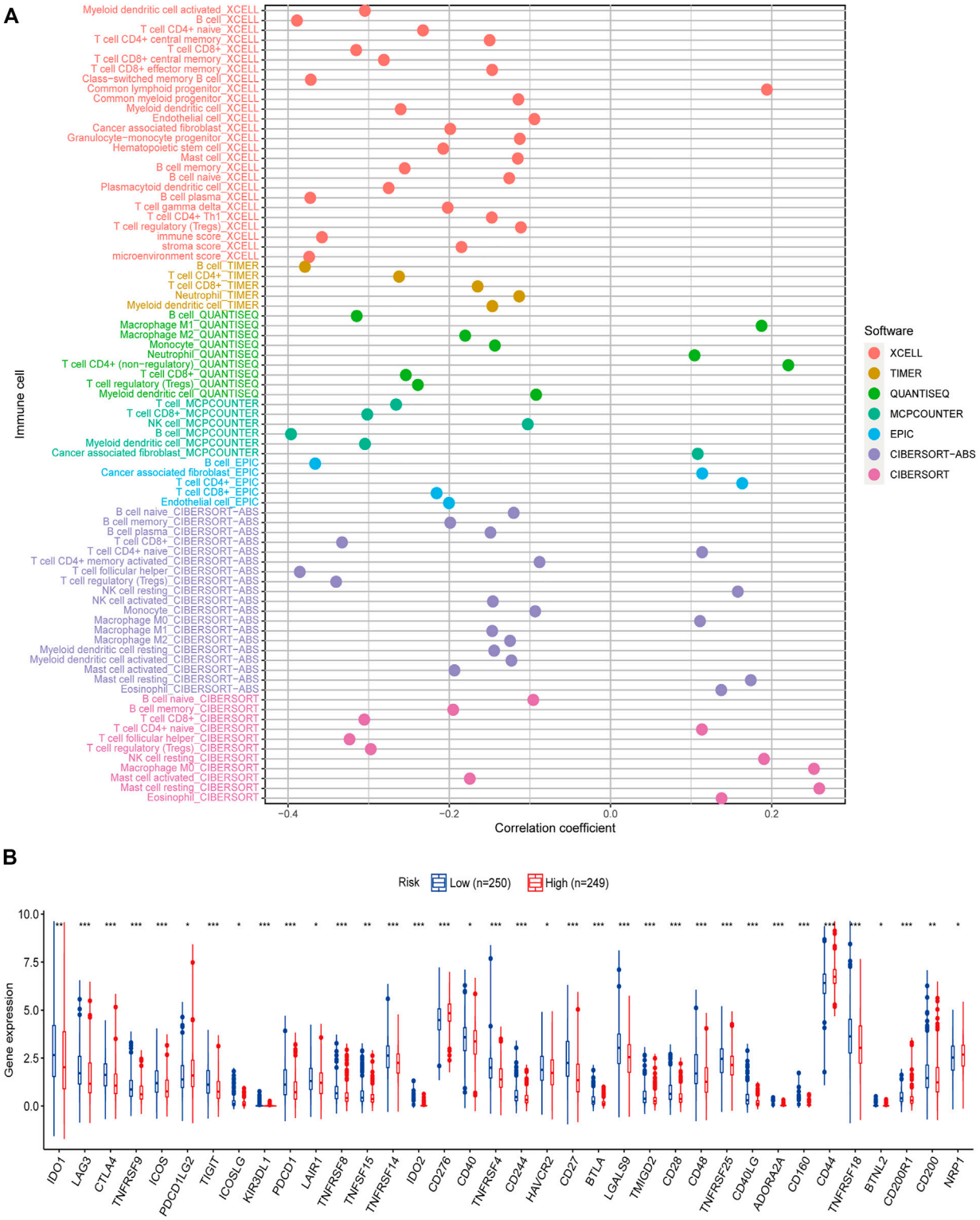


FIGURE 7 | Correlation of tumor-infiltrating cells and immune checkpoints with risk scores. **(A)** A lollipop diagram shows that high-risk score was positively related to M0, M1 macrophages, and resting NK cells infiltrating, whereas it was negatively associated with B cells, CD8 T cells, M2 macrophages, myeloid dendritic cells, and monocytes infiltrating in head and neck squamous cell carcinoma patients. **(B)** Expression of immune checkpoints among high- ($n = 249$) and low-risk groups ($n = 250$). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

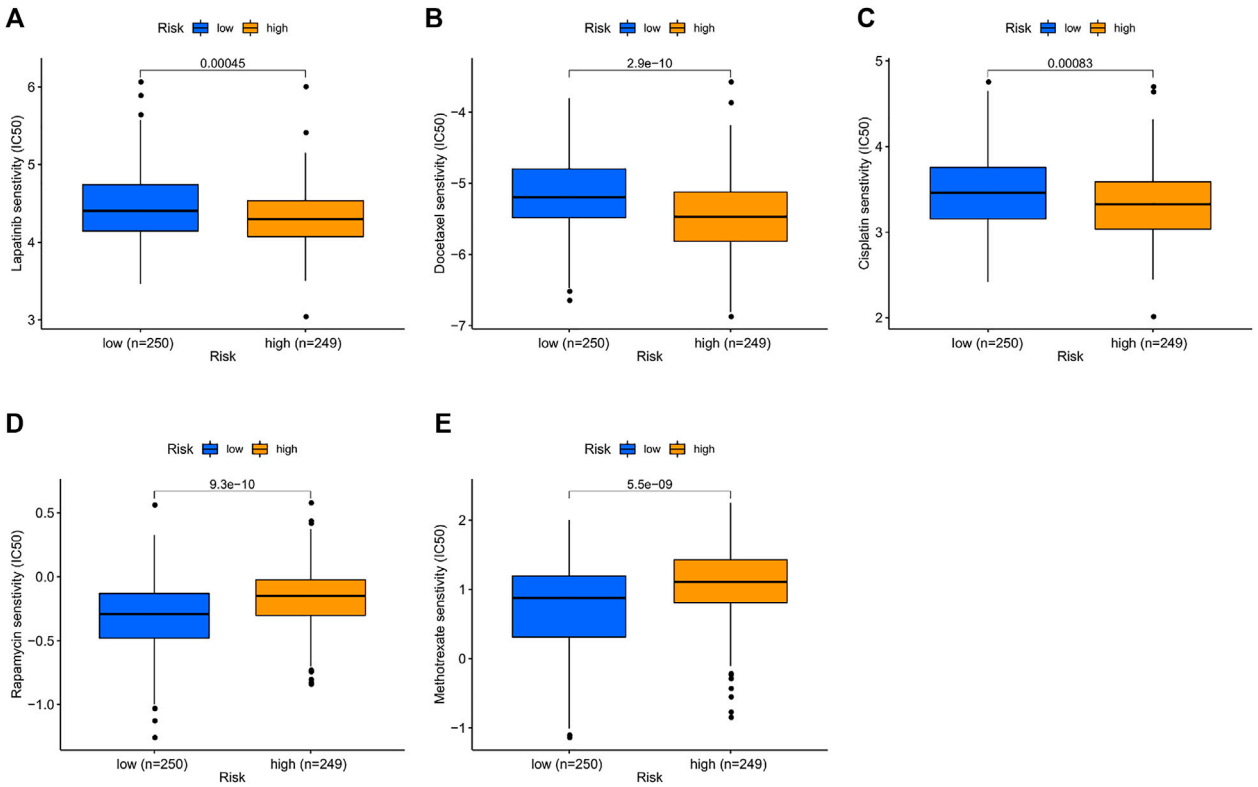


FIGURE 8 | Prediction sensitivity of chemotherapy benefit in high- (*n* = 249) and low-risk groups (*n* = 250). **(A)** Lapatinib. **(B)** Docetaxel. **(C)** Cisplatin. **(D)** Rapamycin. **(E)** Methotrexate.

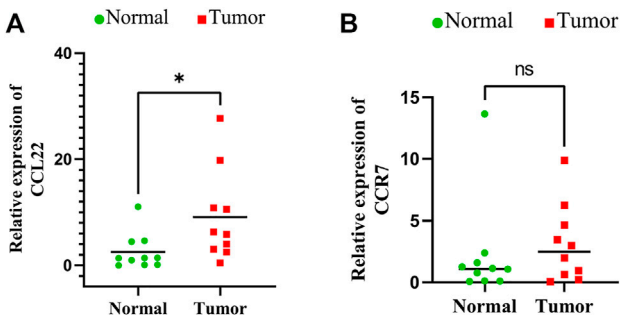


FIGURE 9 | Expressions of signature genes CCL22 and IL10 were upregulated in head and neck squamous cell carcinoma. **(A)** qPCR analysis of CCL22 mRNA levels in tumor and adjacent normal tissues. **(B)** qPCR analysis of IL10 mRNA levels in tumor and adjacent normal tissues.

organs, CCR7 has been shown to activate B and T lymphocytes, regulate the movement of memory T cells to inflamed tissues, and promote dendritic cell growth. Furthermore, CCR7 is required for the growth of lymph nodes and the construction of their follicular functional structures, as well as for the guidance of cells to lymphoid tissue (Fuss et al., 2021). Therefore, like CCL22, CRR7 plays a crucial role in the migration of tumor cells to the lymphatic system and makes a crucial contribution to the metastasis and expansion of TSCC (Qi Wang et al., 2021). As a co-stimulatory and adhesion molecule, CD48 is mainly synthesized by hematopoietic cells, especially antigen-presenting cells. CD48 interacts with CD2 and participates in various innate immune responses, playing a key role in regulating immune activation or suppression (Mcardel et al., 2016). In addition, the expression of CD48 has been shown to be upregulated throughout the development of glioblastoma, and patients with high CD48 expression have a poorer prognosis (Zou et al., 2019). HBEGF can activate growth factor activity and heparin-binding activity, which is involved in multiple processes, such as promoting wound healing, and plays a role in the epidermal growth factor receptor pathway (Higashiyama et al., 1991). HBEGF is abundantly produced in ovarian and breast cancers, and its high expression can promote the tumor cell or macrophage paracrine invasion loop, thereby promoting tumor infiltration, invasion, and metastasis (Zhou et al., 2014). F3, also known as tissue factor (TF), is a protein encoded by a gene in cells responsible for initiating the blood clotting process. It promotes the hypercoagulable state in cancer patients by acting as a high-affinity receptor for coagulation factor VII (Liu et al., 2015). Furthermore, F3 promotes tumorigenesis, angiogenesis, and tumor cell migration and metastasis, all of which contribute to tumor expansion. According to previous studies, F3 gene expression levels were significantly elevated in patients with advanced lung cancer and were associated with poorer survival (Régina et al., 2009). IL10 is an interleukin mainly produced by

monocytes, with multiple effects on innate and adaptive immunity and inflammatory processes. It inhibits the production of Th1 cytokines, MHC class II antigens, and co-stimulatory molecules in macrophages and also helps B cells survive, proliferate, and produce antibodies. Importantly, it is involved in the control of tumor cell proliferation and invasion through the JAK/STAT signaling pathway (Béguelin et al., 2015). In this study, the expression level of IL10 was significantly increased in HNSCC tissues, suggesting that IL10 may be a potential prognostic factor in HNSCC. As a receptor for IL10, IL10RA can regulate tumor immune responses and is highly expressed in HNSCC tissues (Juncheng Wang et al., 2021). However, in our study, its expression did not differ between tumors and adjacent tissues, which may be related to our small sample size. IL2RB is one of the receptors of IL2 that is involved in T-cell–mediated immune responses, and its binding to IL2 promotes the growth of NK92 cells and its anticancer function (Jounaidi et al., 2017). LCK is an important factor regulating the motility of oral cancer cells and is associated with the invasion and metastasis of oral cancer cells. LCK inhibitors could be used to regulate glioma cell movement, cancer invasion, and stem cell gene expression (Zepecki et al., 2018). This suggests that LCK may become a new target gene for cancer therapy. PIK3R5 gene has predictive potential for lymph node metastasis in breast cancer, while PSEN1 downregulation promotes the chemoradiotherapy resistance of esophageal carcinoma (Meng et al., 2016; Paula et al., 2017). TIMP1 and TNFRSF1B genes have been confirmed to take part in the pathogenesis of colorectal cancer and have prognostic value (Yu et al., 2014; Huang et al., 2019). Downregulation of TNFAIP6 can inhibit the proliferation, invasion, and metastasis of gastric cancer (GC). Previous research demonstrated that increased TNFAIP6 expression is linked to the level of infiltration, lymph node metastases, and poor outcome in GC patients, implying that TNFAIP6 may be a workable therapeutic target for GC (Zhang et al., 2021). Our findings demonstrate these genes can be applied to forecast HNSCC outcomes, although their mechanisms need to be further explored.

It has been increasingly recognized that the tumor microenvironment is closely correlated with the occurrence and progression of cancers. The results of existing studies have shown that immune component imbalance is a significant factor causing the short survival of patients with cancer (Li et al., 2017). To further understand the association among risk scores and tumor infiltrating immune cells, we performed seven established algorithms for immune infiltration analysis. Due to the heterogeneity of these methods, internal comparisons were rarely utilized. Through integration analysis, our study showed that the abundance of B cells, CD8 T cells, M2 macrophages, myeloid dendritic cells, and monocytes in the low-risk group was higher, and the high-risk group had higher M0, M1 macrophages, and resting NK cells. Yang et al. (2019) reported that in cervical cancer patients, CD8 T-cell infiltration tended to have a good prognosis, whereas Ali et al. (2016) suggested that increased levels of M0 macrophages led to a poor prognosis in breast cancer. The results of these studies are consistent with ours.

Checkpoint inhibitor immunotherapies significantly improved the prognosis of patients with advanced cancers (Hellmann et al., 2018). We found a meaningful distinction in the expression of immune checkpoints among the two groups in the present study, which suggested that the sensitivity for immunotherapy may also be different. Additionally, our study demonstrated that the high-risk patients were susceptible to chemotherapy drugs, for example, cisplatin, docetaxel, and lapatinib. Therefore, the results of our study may help foresee the immune checkpoint levels and potentially guide immunotherapy and chemotherapeutic decisions. Finally, we carried out qRT-PCR analysis and concluded that mRNA expression levels of CCL22 and IL10 were higher in HNSCC cancer tissues than in normal tissues.

However, our research has some limitations, which does not affect the value of this study. First, we presented bioinformatic evidence by constructing and evaluating 14 inflammatory response-related gene signatures based on traditional statistical methods that can accurately predict the prognosis of HNSCC. Although many reports have proved the practicability of these methods, it is still necessary to develop higher-level statistical techniques to improve the accuracy and stability of the model. Additionally, our outcomes were only verified using the GEO data sets and qRT-PCR, and further validation in additional clinical cases is required. Second, we only analyzed the correlation between inflammatory genes and clinical prognosis, and further experimental studies are required to explore the specific mechanism of these genes in relation to the prognosis of HNSCC. Finally, tobacco, alcohol, HPV, and EBV infections, which are important risk factors for HNSCC, were not included in the analysis. We will continue to investigate this field in the future.

CONCLUSION

In summary, we identified 14 prognostic inflammatory response-related genes in HNSCC and developed a new prognostic signature model that was confirmed to be independently related to OS. Furthermore, this signature revealed the immune cell infiltration patterns and chemotherapeutic sensitivity in HNSCC, which can assist in predicting the prognosis and guiding the treatment.

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

The five authors worked together to complete the article. YL, as the corresponding author, guided the design of the entire study, the writing of the article, and the revision of the manuscript. YH and ZD were responsible for the design of the research, collection

of data, the drafting of the manuscript, experimental verification, and other scattered work; they contributed equally to this work as did the first author. BC and YL were responsible for the data curation and analysis. 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/fgene.2022.767166/full#supplementary-material

REFERENCES

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi:10.1109/tac.1974.1100705

Ali, H. R., Chlon, L., Pharoah, P. D. P., Markowetz, F., and Caldas, C. (2016). Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *Plos Med.* 13, e1002194. doi:10.1371/journal.pmed.1002194

Aran, D. (2020). Cell-Type Enrichment Analysis of Bulk Transcriptomes Using xCell. *Methods Mol. Biol.* 2120, 263–276. doi:10.1007/978-1-0716-0327-7_19

Béguelin, W., Sawh, S., Chambwe, N., Chan, F. C., Jiang, Y., Choo, J.-W., et al. (2015). IL10 Receptor Is a Novel Therapeutic Target in DLBCLs. *Leukemia* 29, 1684–1694. doi:10.1038/leu.2015.57

Cai, H., Zhang, Z. H., Zhou, Y. J., Liu, J., Chen, H. Q., and Lin, R. Y. (2020). The Prognostic Value of Preoperative Plasma Fibrinogen and Neutrophil-To-Lymphocyte Ratio in Patients with Laryngeal Squamous Cell Carcinoma. *Ear Nose Throat J.* 100, 731–736. doi:10.1177/0145561320920746

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chow, L. Q. M. (2020). Head and Neck Cancer. *N. Engl. J. Med.* 382, 60–72. doi:10.1056/nejmra1715715

Clatot, F., Gouérant, S., Mareschal, S., Cornic, M., Berghian, A., Choussy, O., et al. (2014). The Gene Expression Profile of Inflammatory, Hypoxic and Metabolic Genes Predicts the Metastatic Spread of Human Head and Neck Squamous Cell Carcinoma. *Oral Oncol.* 50, 200–207. doi:10.1016/j.oraloncology.2013.12.009

Dienstmann, R., Villacampa, G., Sveen, A., Mason, M. J., Niedzwiecki, D., Nesbakken, A., et al. (2019). Relative Contribution of Clinicopathological Variables, Genomic Markers, Transcriptomic Subtyping and Microenvironment Features for Outcome Prediction in Stage II/III Colorectal Cancer. *Ann. Oncol.* 30, 1622–1629. doi:10.1093/annonc/mdz287

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., et al. (2019). Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods. *Int. J. Cancer* 144, 1941–1953. doi:10.1002/ijc.31937

Fuss, C. T., Other, K., Heinze, B., Landwehr, L. S., Wiegering, A., Kalogirou, C., et al. (2021). Expression of the Chemokine Receptor CCR7 in the Normal Adrenal Gland and Adrenal Tumors and its Correlation with Clinical Outcome in Adrenocortical Carcinoma. *Cancers (Basel)* 13, 5693. doi:10.3390/cancers13225693

Greten, F. R., and Grivennikov, S. I. (2019). Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* 51, 27–41. doi:10.1016/j.immuni.2019.06.025

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

Haddad, R. I., and Shin, D. M. (2008). Recent Advances in Head and Neck Cancer. *N. Engl. J. Med.* 359, 1143–1154. doi:10.1056/nejmra0707975

Hellmann, M. D., Nathanson, T., Rizvi, H., Creelan, B. C., Sanchez-Vega, F., Ahuja, A., et al. (2018). Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-small-cell Lung Cancer. *Cancer Cell* 33, 843–852. doi:10.1016/j.ccell.2018.03.018

Higashiyama, S., Abraham, J. A., Miller, J., Fiddes, J. C., and Klagsbrun, M. (1991). A Heparin-Binding Growth Factor Secreted by Macrophage-like Cells that Is Related to EGF. *Science* 251, 936–939. doi:10.1126/science.1840698

Huang, R., Wang, K., Gao, L., and Gao, W. (2019). TIMP1 Is A Potential Key Gene Associated with the Pathogenesis and Prognosis of Ulcerative Colitis-Associated Colorectal Cancer. *OncoTargets Ther.* 12, 8895–8904. doi:10.2147/ott.s222608

Jia, Z., Zhang, Z., Yang, Q., Deng, C., Li, D., and Ren, L. (2019). Effect of IL2RA and IL2RB Gene Polymorphisms on Lung Cancer Risk. *Int. Immunopharmacology* 74, 105716. doi:10.1016/j.intimp.2019.105716

Jiang, H., Ma, B., Xu, W., Luo, Y., Wang, X., Wen, S., et al. (2021). A Novel Three-lncRNA Signature Predicts the Overall Survival of HNSCC Patients. *Ann. Surg. Oncol.* 28, 3396–3406. doi:10.1245/s10434-020-09210-1

Jounaidi, Y., Cotten, J. F., Miller, K. W., and Forman, S. A. (2017). Tethering IL2 to its Receptor IL2Rβ Enhances Antitumor Activity and Expansion of Natural Killer NK92 Cells. *Cancer Res.* 77, 5938–5951. doi:10.1158/0008-5472.can-17-1007

Juncheng Wang, J., Tian, Y., Zhu, G., Li, Z., Wu, Z., Wei, G., et al. (2021). Establishment and Validation of Immune Microenvironmental Gene Signatures for Predicting Prognosis in Patients with Head and Neck Squamous Cell Carcinoma. *Int. immunopharmacology* 97, 107817. doi:10.1016/j.intimp.2021.107817

Kimura, S., Noguchi, H., Nanbu, U., and Nakayama, T. (2021). Macrophage CCL22 Expression Promotes Lymphangiogenesis in Patients with Tongue Squamous Cell Carcinoma via IL-4/STAT6 in the Tumor Microenvironment. *Oncol. Lett.* 21, 383. doi:10.3892/ol.2021.12644

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, 1529–1537. doi:10.1001/jamaoncol.2017.1609

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

Liu, Y., Hu, H., Zhang, C., Wang, Z., Li, M., Zhang, W., et al. (2015). Methylation Associated Genes Contribute to the Favorable Prognosis of Gliomas with Isocitrate Dehydrogenase 1 Mutation. *Am. J. Cancer Res.* 5, 2745–2755.

Ma, H., Chang, H., Yang, W., Lu, Y., Hu, J., and Jin, S. (2020). A Novel IFNα-Induced Long Noncoding RNA Negatively Regulates Immunosuppression by Interrupting H3K27 Acetylation in Head and Neck Squamous Cell Carcinoma. *Mol. Cancer* 19, 4. doi:10.1186/s12943-019-1123-y

Maasland, D. H., Van Den Brandt, P. A., Kremer, B., Goldbohm, R. A., and Schouten, L. J. (2014). Alcohol Consumption, Cigarette Smoking and the Risk of Subtypes of Head-Neck Cancer: Results from the Netherlands Cohort Study. *BMC Cancer* 14, 187. doi:10.1186/1471-2407-14-187

Mcardel, S. L., Terhorst, C., and Sharpe, A. H. (2016). Roles of CD48 in Regulating Immunity and Tolerance. *Clin. Immunol.* 164, 10–20. doi:10.1016/j.clim.2016.01.008

Mcmillan, D. C. (2013). The Systemic Inflammation-Based Glasgow Prognostic Score: a Decade of Experience in Patients with Cancer. *Cancer Treat. Rev.* 39, 534–540. doi:10.1016/j.ctrv.2012.08.003

Mehanna, H., Beech, T., Nicholson, T., El-Hariry, I., Mcconkey, C., Paleri, V., et al. (2013). Prevalence of Human Papillomavirus in Oropharyngeal and Nonoropharyngeal Head and Neck Cancer-Systematic Review and Meta-Analysis of Trends by Time and Region. *Head Neck* 35, 747–755. doi:10.1002/hed.22015

Meng, F., Qian, L., Lv, L., Ding, B., Zhou, G., Cheng, X., et al. (2016). miR-193a-3p Regulation of Chemoradiation Resistance in Oesophageal Cancer Cells via the PSEN1 Gene. *Gene* 579, 139–145. doi:10.1016/j.gene.2015.12.060

Paula, L. M., De Moraes, L. H. F., Do Canto, A. L., Dos Santos, L., Martin, A. A., Rogatto, S. R., et al. (2017). Analysis of Molecular Markers as Predictive Factors of Lymph Node Involvement in Breast Carcinoma. *Oncol. Lett.* 13, 488–496. doi:10.3892/ol.2016.5438

Peng, Z., Andersson, K., Lindholm, J., Dethlefsen, O., Pramana, S., Pawitan, Y., et al. (2016). Improving the Prediction of Prostate Cancer Overall Survival by Supplementing Readily Available Clinical Data with Gene Expression Levels of IGFBP3 and F3 in Formalin-Fixed Paraffin Embedded Core Needle Biopsy Material. *PLoS One* 11, e0145545. doi:10.1371/journal.pone.0145545

Plattner, C., Finotello, F., and Rieder, D. (2020). Deconvoluting Tumor-Infiltrating Immune Cells from RNA-Seq Data Using quanTIseq. *Methods Enzymol.* 636, 261–285. doi:10.1016/bs.mie.2019.05.056

Qi Wang, Q., Zou, H., Wang, Y., Shang, J., Yang, L., and Shen, J. (2021). CCR7-CCL21 axis Promotes the Cervical Lymph Node Metastasis of Tongue Squamous Cell Carcinoma by Up-Regulating MUC1. *J. Craniomaxillofac. Surg.* 49, 562–569. doi:10.1016/j.jcms.2021.02.027

Racle, J., De Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *Elife* 6, e26476. doi:10.7554/eLife.26476

Rassouli, A., Saliba, J., Castano, R., Hier, M., and Zeitouni, A. G. (2015). Systemic Inflammatory Markers as Independent Prognosticators of Head and Neck Squamous Cell Carcinoma. *Head Neck* 37, 103–110. doi:10.1002/hed.23567

Régina, S., Valentin, J.-B., Lachot, S., Lemarie´, E., Rollin, J., and Gruel, Y. (2009). Increased Tissue Factor Expression Is Associated with Reduced Survival in Non-small Cell Lung Cancer and with Mutations of TP53 and PTEN. *Clin. Chem.* 55, 1834–1842. doi:10.1373/clinchem.2009.123695

Röhrle, N., Knott, M. M. L., and Anz, D. (2020). CCL22 Signaling in the Tumor Environment. *Adv. Exp. Med. Biol.* 1231, 79–96. doi:10.1007/978-3-030-36667-4_8

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

Tamminga, M., Hiltermann, T. J. N., Schuuring, E., Timens, W., Fehrmann, R. S., and Groen, H. J. (2020). Immune Microenvironment Composition in Non-small Cell Lung Cancer and its Association with Survival. *Clin. Transl Immunol.* 9, e1142. doi:10.1002/cti2.1142

Turunen, A., Rautava, J., Grénman, R., Syrjänen, K., and Syrjänen, S. (2017). Epstein-Barr Virus (EBV)-encoded Small RNAs (EBERs) Associated with Poor Prognosis of Head and Neck Carcinomas. *Oncotarget* 8, 27328–27338. doi:10.18632/oncotarget.16033

Valdes, M., Villeda, J., Mithoowani, H., Pitre, T., and Chasen, M. (2020). Inflammatory Markers as Prognostic Factors of Recurrence in Advanced-Stage Squamous Cell Carcinoma of the Head and Neck. *Curr. Oncol.* 27, 135–141. doi:10.3747/co.27.5731

Wang, Q., Schmoeckel, E., Kost, B. P., Kuhn, C., Vattai, A., Vilsmaier, T., et al. (2019). Higher CCL22+ Cell Infiltration Is Associated with Poor Prognosis in Cervical Cancer Patients. *Cancers (Basel)* 11, 2004. doi:10.3390/cancers11122004

Wu, Z. H., Yue, J. X., Zhou, T., and Xiao, H. J. (2021). Integrated Analysis of the Prognostic Values of RNA -binding Proteins in Head and Neck Squamous Cell Carcinoma. *Biofactors* 47, 478–488. doi:10.1002/biof.1722

Yang, S., Wu, Y., Deng, Y., Zhou, L., Yang, P., Zheng, Y., et al. (2019). Identification of a Prognostic Immune Signature for Cervical Cancer to Predict Survival and Response to Immune Checkpoint Inhibitors. *Oncoimmunology* 8, e1659094. doi:10.1080/2162402x.2019.1659094

Yu, Y., Zheng, S., Zhang, S., Jin, W., Liu, H., Jin, M., et al. (2014). Polymorphisms of Inflammation-Related Genes and Colorectal Cancer Risk: a Population-Based Case-Control Study in China. *Int. J. Immunogenet.* 41, 289–297. doi:10.1111/iji.12119

Zadka, Ł., Kulus, M. J., Kurnol, K., Piotrowska, A., Glatzel-Plucińska, N., Jurek, T., et al. (2018). The Expression of IL10RA in Colorectal Cancer and its Correlation with the Proliferation index and the Clinical Stage of the Disease. *Cytokine* 110, 116–125. doi:10.1016/j.cyto.2018.04.030

Zepecki, J. P., Snyder, K. M., Moreno, M. M., Fajardo, E., Fiser, A., Ness, J., et al. (2018). Regulation of Human Glioma Cell Migration, Tumor Growth, and Stemness Gene Expression Using a Lck Targeted Inhibitor. *Oncogene* 38, 1734–1750. doi:10.1038/s41388-018-0546-z

Zhang, X., Xue, J., Yang, H., Zhou, T., and Zu, G. (2021). TNFAIP6 Promotes Invasion and Metastasis of Gastric Cancer and Indicates Poor Prognosis of Patients. *Tissue and Cell* 68, 101455. doi:10.1016/j.tice.2020.101455

Zhou, Z. N., Sharma, V. P., Beaty, B. T., Roh-Johnson, M., Peterson, E. A., Van Rooijen, N., et al. (2014). Autocrine HBEGF Expression Promotes Breast Cancer Intravasation, Metastasis and Macrophage-independent Invasion *In Vivo*. *Oncogene* 33, 3784–3793. doi:10.1038/onc.2013.363

Zhou, T., Yu, S.-T., Chen, W.-Z., Xie, R., and Yu, J.-C. (2019). Pretreatment Albumin Globulin Ratio Has a superior Prognostic Value in Laryngeal Squamous Cell Carcinoma Patients: a Comparison Study. *J. Cancer* 10, 594–601. doi:10.7150/jca.28817

Zhou, S., Yuan, H., Wang, J., Hu, X., Liu, F., Zhang, Y., et al. (2020). Prognostic Value of Systemic Inflammatory Marker in Patients with Head and Neck Squamous Cell Carcinoma Undergoing Surgical Resection. *Future Oncol.* 16, 559–571. doi:10.2217/fon-2020-0010

Zou, C., Zhu, C., Guan, G., Guo, Q., Liu, T., Shen, S., et al. (2019). CD48 Is a Key Molecule of Immunomodulation Affecting Prognosis in Glioma. *OncoTargets Ther.* 12, 4181–4193. doi:10.2147/ott.s198762

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

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

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership