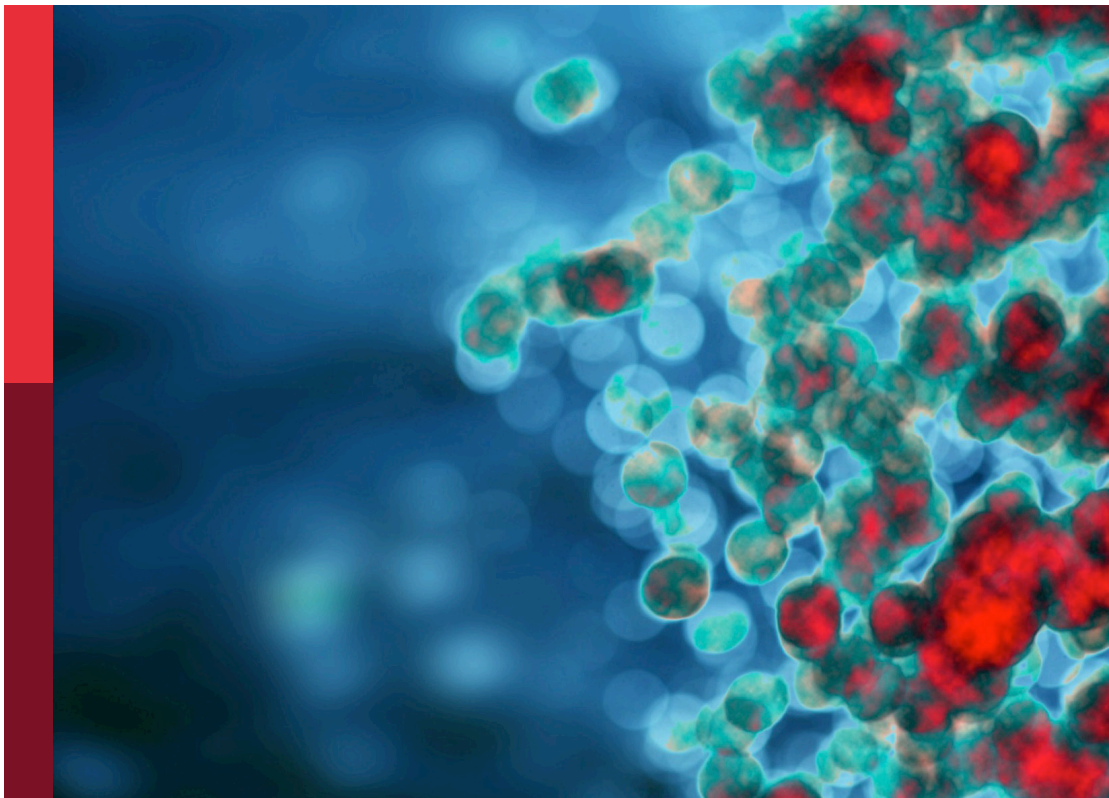
frontiers | Research Topics

Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies

Edited by
Nan Zhang, Meng Zhou, Hao Zhang, Jun Liu, and Ping Zheng

Published in
Frontiers in Immunology
Frontiers in Oncology



FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-3377-2
DOI 10.3389/978-2-8325-3377-2

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

Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies

Topic editors

Nan Zhang — Harbin Medical University, China
Meng Zhou — Wenzhou Medical University, China
Hao Zhang — Chongqing Medical University, China
Jun Liu — Department of Clinical Laboratory, Yuebei People's Hospital, China
Ping Zheng — The University of Melbourne, Australia

Citation

Zhang, N., Zhou, M., Zhang, H., Liu, J., Zheng, P., eds. (2023). *Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3377-2

Table of contents

05 Editorial: Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies
Ping Zheng and Jun Liu

08 LncRNA profiles from Notch signaling: Implications for clinical management and tumor microenvironment of colorectal cancer
Qin Dang, Zaoqu Liu, Yang Liu, Wenkang Wang, Weitang Yuan, Zhenqiang Sun, Lin Liu and Chengzeng Wang

24 An integrated model of acinar to ductal metaplasia-related N7-methyladenosine regulators predicts prognosis and immunotherapy in pancreatic carcinoma based on digital spatial profiling
Hao Yang, Julia Messina-Pacheco, Andrea Liliam Gomez Corredor, Alex Gregorieff, Jun-li Liu, Ali Nehme, Hamed S. Najafabadi, Yasser Riazalhosseini, Bo Gao and Zu-hua Gao

45 Transcriptional patterns reveal tumor histologic heterogeneity and immunotherapy response in lung adenocarcinoma
Mengxue Jiao, Hui Liu and Xuejun Liu

56 CDKN2A-mediated molecular subtypes characterize the hallmarks of tumor microenvironment and guide precision medicine in triple-negative breast cancer
Tianyi Cheng, Yingyi Wu, Zhiyu Liu, Yi Yu, Shixue Sun, Min Guo, Baoqing Sun and Chen Huang

78 mRNAsi-related metabolic risk score model identifies poor prognosis, immunoevasive contexture, and low chemotherapy response in colorectal cancer patients through machine learning
Meilin Weng, Ting Li, Jing Zhao, Miaomiao Guo, Wenling Zhao, Wenchao Gu, Caihong Sun, Ying Yue, Ziwen Zhong, Ke Nan, Qingwu Liao, Minli Sun, Di Zhou and Changhong Miao

97 FLT3LG and IFITM3P6 consolidate T cell activity in the bone marrow microenvironment and are prognostic factors in acute myelocytic leukemia
Haiyan Chen, Meng Wu, Hongping Xia, Songjie Du, Guoren Zhou, Guangfeng Long, Yanan Zhu, Xu Huang and Daheng Yang

115 Development and validation of a necroptosis-related gene prognostic score to predict prognosis and efficiency of immunotherapy in gastric cancer
Yan Xia, Rongzheng Zhang, Mingzhu Wang, Jiaqi Li, Jianming Dong, Kaitong He, Ting Guo, Xiaomei Ju, Jiaqiu Ru, Shuyun Zhang and Yihua Sun

134 The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma
Weichen Wang, Zhichao Lu, Maoyu Wang, Zongheng Liu, Bing Wu, Chengkai Yang, He Huan and Peipei Gong

150 Identification of clinical prognostic features of esophageal cancer based on m6A regulators
Huimei Wang, Yiping Zhang, Lin Chen, Yufeng Liu, Chen Xu, Dongxian Jiang, Qi Song, Haixing Wang, Liyan Wang, Yu Lin, Yuanmei Chen, Junqiang Chen, Yuanji Xu and Yingyong Hou

159 Immune depletion of the methylated phenotype of colon cancer is closely related to resistance to immune checkpoint inhibitors
Chengqian Zhong, Tingjiang Xie, Long Chen, Xuejing Zhong, Xinjing Li, Xiumei Cai, Kaihong Chen and Shiqian Lan

175 Immunoregulation and clinical significance of neutrophils/NETs-ANGPT2 in tumor microenvironment of gastric cancer
Shifeng Yang, Xiaoming Zou, Jiacheng Li, Hao Yang, Ange Zhang, Yanli Zhu, Lei Zhu and Lisha Zhang

192 Protective effect of tertiary lymphoid structures against hepatocellular carcinoma: New findings from a genetic perspective
Weili Jia, Qianyun Yao, Yanfang Wang, Zhenzhen Mao, Tianchen Zhang, Jianhui Li, Ye Nie, Xinjun Lei, Wen Shi and Wenjie Song

212 Regulatory T-cells-related signature for identifying a prognostic subtype of hepatocellular carcinoma with an exhausted tumor microenvironment
Genhao Zhang

230 Applications of machine learning in tumor-associated macrophages
Zhen Li, Qijun Yu, Qingyuan Zhu, Xiaojing Yang, Zhaobin Li and Jie Fu

246 A novel neutrophil extracellular trap signature to predict prognosis and immunotherapy response in head and neck squamous cell carcinoma
Qilin Li, Weimin Chen, Qiuhui Li, Jing Mao and Xin Chen

262 Machine learning reveals two heterogeneous subtypes to assist immune therapy based on lipid metabolism in lung adenocarcinoma
Xuyu Gu, Shiyou Wei, Zhixin Li and Huan Xu

277 Integrated machine learning methods identify FNDC3B as a potential prognostic biomarker and correlated with immune infiltrates in glioma
Xiao Wang, Yeping Huang, Shanshan Li and Hong Zhang

frontiers | Frontiers in Immunology

TYPE Editorial
PUBLISHED 15 August 2023
DOI 10.3389/fimmu.2023.1239648

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Massimo Guidoboni,
Scientific Institute of Romagna for the Study and Treatment of Tumors (IRCCS), Italy

*CORRESPONDENCE
Ping Zheng
zhengp@unimelb.edu.au

RECEIVED 13 June 2023
ACCEPTED 03 July 2023
PUBLISHED 15 August 2023

CITATION
Zheng P and Liu J (2023) Editorial: Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies. *Front. Immunol.* 14:1239648. doi: 10.3389/fimmu.2023.1239648

COPYRIGHT
© 2023 Zheng 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.

Editorial: Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies

Ping Zheng[1]* and Jun Liu[2]

[1]Department of Neurosurgery, Pudong New Area People's Hospital, Shanghai, China, [2]Yuebei People's Hospital, Shaoguan, China

KEYWORDS

tumor microenvironment, machine learning, tumor, disease, immune infiltrations

Editorial on the Research Topic
Unveiling the tumor microenvironment by machine learning to develop new immunotherapeutic strategies

A total of 35 paper are included in this series. We selected eight as representative ones:

Yang et al. reported pancreatic ductal adenocarcinoma (PDAC) with a high 7-methylguanosine (m7G) score were characterized by increased immune cell infiltration, increased genomic instability, higher response rate to combined immune checkpoint inhibitors (ICIs), and overall poor survival. Their findings indicate that the m7G score is associated with tumor invasiveness, immune cell infiltration, ICI treatment response, and overall patients' survival. They also identified FN1 and ITGB1 as core genes in the m7Gscore model, which affect immune cell infiltration and genomic instability not only in pancreatic cancer but also in pan-cancer. FN1 and ITGB1 can inhibit immune T cell activation by upregulation of macrophages and neutrophils, thereby leading to immune escape of pancreatic cancer cells and reducing the response rate of ICI treatment.

Wang H. et al. reported that most of the 23 m6A regulators were significantly differentially expressed in the Esophageal cancer (ESCA) tissues. LASSO regression analysis was used to perform a prognostic risk model that included seven m6A-related regulators (FMR1, RBMX, IGFBP1, IGFBP2, ALKBH5, RBM15B, and METTL14). Moreover, they found that this risk model was significantly correlated with biological functions, including base metabolism, DNA repair, and mismatch repair. A nomogram was constructed to predict the prognosis of patients with ESCA. The results of bioinformatics analysis were further validated in human ESCA and normal tissues by qRT-PCR.

Cheng et al. applied an unsupervised cluster analysis based on Cyclin-dependent kinase inhibitor 2A (CDKN2A)-correlated genes unveiled three subtypes, namely cold-immune subtype, IFN-γ activated subtype and FTL-dominant subtype. Subsequent analyses depicting hallmarks of tumor microenvironment (TME) among three subtypes suggested strong association between triple-negative breast cancer (TNBC) and CDKN2A. Given the fact that the most clinically heterogeneous TNBC always displayed the most severe

outcomes and lacked relevant drug targets, they further explored the potential of immunotherapy for TNBC by interfering CDKN2A and constructed the CDKN2A-derived prognostic model for TNBC patients by Lasso-Cox. The 21-gene–based prognostic model showed high accuracy and was verified in external independent validation cohort. Moreover, they proposed three drugs for TNBC patients based on our model *via* targeting epidermal growth factor receptor.

Wang W. et al. identified two subclusters based on cuproptosis-related signature (CRGs) in glioma. Patients in cluster2 had better clinical outcomes. The cuproptosis-signature was constructed based on CuproptosisScore. Patients with higher CuproptosisScore had higher WHO grades and worse prognosis, while patients with lower grades were more likely to develop IDH mutations or MGMT methylation. Univariate and Multivariate Cox regression analysis demonstrated CuproptosisScore was an independent prognostic factor. The accuracy of the signature in prognostic prediction was further confirmed in 11 external validation datasets. In groups with high-CuproptosisScore, PIK3CA, MUC16, NF1, TTN, TP53, PTEN, and EGFR showed high mutation frequency. IDH1, TP53, ATRX, CIC, and FUBP1 demonstrated high mutation frequency in low-CuproptosisScore group. The level of immune infiltration increased as CuproptosisScore increased. SubMap analysis revealed patients with high-CuproptosisScore may respond to anti-PD-1 therapy. The IC50 values of Bexarotene, Bicalutamide, Bortezomib, and Cytarabine were lower in the high-CuproptosisScore group than those in the low-CuproptosisScore group.

Xia et al. showed the high- Necroptosis-Related Gene Prognostic Score (NRGPS) group had significantly lower the overall survival (OS) than the low-NRGPS group in gastric cancer. Cox regression analyses showed that NRGPS was an independent prognostic variable. Tumor-mutation-burden (TMB), tumor microenvironment (TME), microsatellite instability (MSI), and Tumor Immune Dysfunction and Exclusion (TIDE) scoring were used as predictors of the immunotherapy response. The high-NRGPS group was characterized by a cancer-friendly immune microenvironment, a high TIDE score, and a low TMB, a low MSI all of which consistently demonstrated that the problems observed in the high-NRGPS group are associated with immune escape in gastric cancer GC.

Wang X. et al. detected Differentially expressed genes (DEGs) by the Wilcoxon test based on the TCGA-LGG dataset and the weighted gene co-expression network analysis (WGCNA) was implemented to identify the significant module associated with the expression level of FNDC3B. Furthermore, they investigated the correlation between FNDC3B with cancer immune infiltrates using TISIDB, ESTIMATE, and CIBERSORTx. Higher FNDC3B expression displayed a remarkably worse overall survival and the expression level of FNDC3B was an independent prognostic indicator for patients with glioma. Based on TCGA LGG dataset, a co-expression network was established and the hub genes were identified. FNDC3B expression was positively correlated to the tumor-infiltrating lymphocytes and immune infiltration score, and high FNDC3B expression was accompanied by the increased expression of B7-H3, PD-L1, TIM-3, PD-1, and CTLA-4. Moreover, expression of FNDC3B was significantly associated with infiltrating levels of several types of immune cells and most of their gene markers in glioma.

Xiao et al. applied the large-scale machine learning to find that SOX family can be divided into two distinct clusters in gliomas, with significant immune characteristics and genomic profiles. Among them, SOX10 was identified as an excellent immune regulator of macrophage in gliomas. High expression of SOX10 is related to shorter OS in LGG, HGG, and pan-cancer groups, but benefited from the immunotherapy. Single-cell sequencing proved SOX10 is

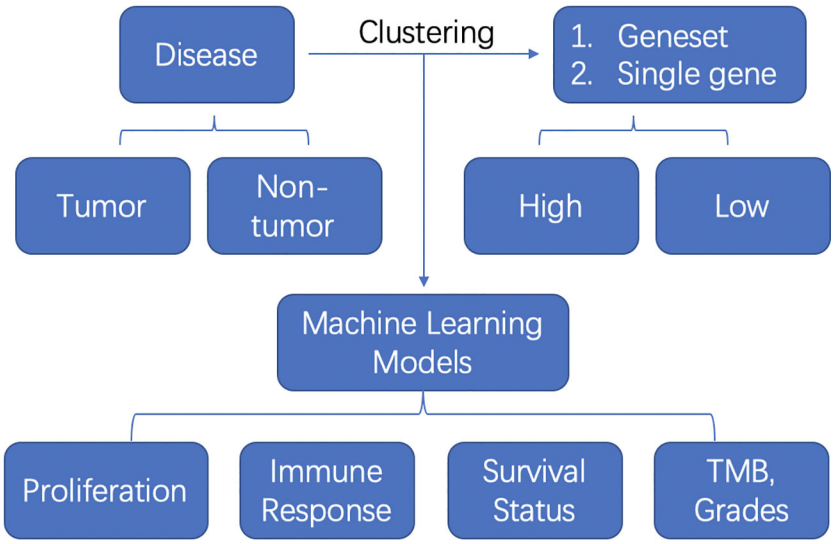


FIGURE 1
The schematic flow of machine learning model application in tumor or non-tumor diseases.

high in neurons, M1 macrophages, and neural stem cells. Macrophages are found to be elevated in the SOX10 high expression group. SOX10 has a positive correlation with macrophage cytokine production and negative regulation of macrophages' chemotaxis and migration.

Lu et al. created a cuproptosis-related lncRNA prognostic model based on the cuproptosis-related lncRNA score (CLS) by performing lasso regression. They identified ten cuproptosis-related genes and 13 correlated prognostic lncRNAs were collected for model construction. CLS was positively or negatively correlated with cancer-related pathways. In addition, cell cycle and immune related pathways were enriched. By performing tumor microenvironment (TME) analysis, they determined that T-cells were activated. High CLS had more tumor characteristics and may lead to higher invasiveness and treatment resistance. Three genes (TP53, CSMD1 and RB1) were found in high CLS samples with more mutational frequency. More amplification and deletion were detected in high CLS samples. In clinical application, a CLS-based nomogram was constructed. 5-Fluorouracil, gemcitabine and doxorubicin had better sensitivity in patients with high CLS. However, patients with low CLS had better immunotherapeutic sensitivity.

In summary, one study involved the single cell RNA-seq and the other study applied the digital spatial profiling. Other most studies were using the TCGA bulk-RNA seq data. The schematic flow can be summarized as follows (Figure 1): The authors first clustered the samples into two or three clusters based on the geneset or single gene medium expression. Then they established the diagnostic or prognosis models with machine learning methods and verified them in some cases with external datasets. At last, they compared the proliferation and invasiveness status, immune response with CIBERSORT, survival status and TMB, grades between or among these clusters. The future point would be adopting more advanced machine learning models, deep neural networks, transfer learning to deal with a big population data and compare the predictive ability of these machine learning methods.

Author contributions

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

Funding

This study was supported by The Featured Clinical Discipline Project of Shanghai Pudong (PW2022A-27) and Key Discipline Group Construction Project of Shanghai Pudong New area People's Hospital (PRYLH202301).

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.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 25 July 2022
DOI 10.3389/fimmu.2022.953405

Check for updates

OPEN ACCESS

EDITED BY
Jun Liu,
Yuebei People's Hospital, China

REVIEWED BY
Ganglei Li,
Zhejiang University, China
Yufeng Gao,
First Affiliated Hospital of Anhui Medical University, China

*CORRESPONDENCE
Chengzeng Wang
czw202112@zzu.edu.cn
Lin Liu
liulin@zzu.edu.cn
Zhenqiang Sun
fccsunzq@zzu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 26 May 2022
ACCEPTED 28 June 2022
PUBLISHED 25 July 2022

CITATION
Dang Q, Liu Z, Liu Y, Wang W, Yuan W, Sun Z, Liu L and Wang C (2022) LncRNA profiles from Notch signaling: Implications for clinical management and tumor microenvironment of colorectal cancer.
Front. Immunol. 13:953405.
doi: 10.3389/fimmu.2022.953405

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

LncRNA profiles from Notch signaling: Implications for clinical management and tumor microenvironment of colorectal cancer

Qin Dang[1,2†], Zaoqu Liu[3†], Yang Liu[4†], Wenkang Wang[5], Weitang Yuan[1], Zhenqiang Sun[1,2*], Lin Liu[2,6*] and Chengzeng Wang[2,6*]

[1]Department of Colorectal Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Henan Institute of Interconnected Intelligent Health Management, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [3]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [4]Department of Radiotherapy, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [5]Department of Breast Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [6]Department of Ultrasound, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

The interplay between long non-coding RNAs (lncRNAs) and the Notch pathway involves a variety of malignancies. However, Notch-derived lncRNAs and their latent clinical significance remain elusive in colorectal cancer (CRC). In this study, we introduced a framework that could screen Notch-derived lncRNAs (named "NLncer") and ultimately identified 24 NLncers. To further explore the clinical significance of these NLncers, we performed LASSO and Cox regression in TCGA-CRC cohort (n = 584) and then retained six lncRNAs tightly associated with prognosis. The final model (termed "NLncS") was subsequently tested in GSE38832 (n = 122), GSE39582 (n = 573), and an in-house clinical cohort (n = 115). Ultimately, our NLncS model could serve as an independent risk factor and afford a robust performance for assessing the prognosis of CRC patients. Additionally, patients with high NLncS risk scores were characterized by upregulation of immune pathways, strong immunogenicity, abundant CD8 + T-cell infiltration, and potentially higher response rates to CTLA4 blockers, which turned out to be suitable for immunotherapy. Aiming at globally observing the characteristics of high-risk patients, somatic mutation and methylation modification analysis provide us with evidence at the genomic and transcriptomic levels. To facilitate the clinical transformability, we mined deeply into the sensitive compounds targeting high-risk individuals and identified dasatinib as a candidate agent for patients with a high Notch risk score. In conclusion, our NLncS model is a promising biomarker for optimizing the clinical management of CRC patients.

KEYWORDS

colorectal cancer, LncRNA, notch, prognosis, immunotherapy, tumor microenvironment

Introduction

As one of the primary causes in tumor-related death, colorectal cancer (CRC) ranks the third most common malignant neoplasm worldwide (1, 2). CRC is often diagnosed at an advanced stage, accompanied by high postoperative recurrence and metastasis rate, which seriously threatens the health of the public. The survival of CRC patients was subjected to a series of clinical difficulties such as unresectable surgery, chemotherapy resistance, and radiotherapy side effects (3, 4). Notably, immunotherapy has shown spectacular achievements in the oncology treatment field (5, 6). However, with the rapid development, limitations of immunotherapy emerged. As proof, immunotherapy brings the desired curative effect when applied to the suitable patient subgroups (7, 8). To address this plight, researchers need to precisely identify individuals suitable for immunotherapy.

The Notch pathway, a signal transduction system ubiquitous in cellular organisms, determines the cell fate and function (9). Studies widely validated that the Notch pathway is involved in diversified aspects of oncology, including carcinogenesis, metastasis, stemness, metabolism, apoptosis, and angiogenesis (10–15). *NOTCH3* expression is closely associated with malignant phenotypes of CRC, including higher grade, the existence of lymph nodes, and distant metastasis (10). Additionally, several Notch signaling receptors could serve as therapeutic targets for breast cancer (16). Moreover, accumulating evidence has shown that the Notch pathway influences the immune system, both innate and adaptive, *via* dendritic cells and T cells (17, 18).

For decades, the identification of abundant long non-coding RNAs (lncRNAs) >200 bp has brought out their characterization as profound components in tumor biology. LncRNA molecules were commonly present in tumors as a double-edged sword of driving tumor development or inhibiting progression (19–22). Previous studies have confirmed that lncRNA could govern key ligands in the Notch pathway to achieve tumor control, such as breast cancer, renal cell cancer, and gastric cancer (23–25). Nevertheless, the comprehensive landscape of the Notch pathway-related lncRNA in CRC remains elusive, and we hope to endue Notch pathways with a novel character and clinical application through functional lncRNA analysis.

Herein, we constructed an integrated frame capable of identifying Notch-derived lncRNA drivers (termed "NLncer"). We characterized the prognostic value of the Notch-related lncRNA signature (NLncS) for CRC patients and distinguished high-risk subgroups that are suitable for immunotherapy. For verification, two independent cohorts and a clinical in-house dataset were enrolled. Subsequently, we probed the methylation levels of distinct individuals in order to find genome-level drivers that contribute to differences in outcomes. Further, based on cell line expression profiles and drug sensitivity results (CTRP and PRISM) and Connectivity Map (CMap) analysis, we recommend the anticancer drug dasatinib as a latent treatment for a CRC high-risk subgroup.

Materials and methods

Public dataset collection and procession

The workflow of this research is depicted in Figure 1. **A** total of 1,279 CRC patients from three independent public cohorts were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov) and Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), including TCGA-CRC, GSE39582, and GSE38832 (Supplementary Table 1). We transformed the RNA-seq raw read count of TCGA-CRC to transcripts per kilobase million (TPM). The GEO datasets were collected from the Affymetrix® Human Genome U133 Plus 2.0 Array (GPL570 platform) and processed by the robust multiarray averaging (RMA) algorithm with the *Affy* package. We obtained 19,526 protein-coding genes and 15,299 lncRNAs from TCGA database based on GENCODE (Homo sapiens GRCh38, https://www.gencodegenes.org/). All probes were mapped to the human genome (Hg38), and 3,439 lncRNAs were acquired by reannotating the probe sets of the GPL570 array (26). Given the batch effect, we combined the colonic adenocarcinoma (COAD) and rectum adenocarcinoma (READ) data into TCGA-CRC queues after *ComBat* algorithm-based processing. LncRNAs with an empty expression in more than half of the sample size in each cohort were excluded. Detailed baselines for three independent queues are pooled in Supplementary Table 1.

NLncer: screening the potential LncRNA drivers of the Notch pathway

Aiming at discovering potential Notch-associated lncRNA drivers, we developed a comprehensive pipeline by referring to previous findings (27, 28). In brief, we arranged the mRNAs in descending order to the pertinence of particular lncRNA adjusted for tumor purity. The *fgsea* R package was used to determine whether Notch pathway-related genes were enriched. Next, the Notch enrichment score (*NES*) of all lncRNAs was measured, and those with significant *NES* were identified as Nlncers. LncRNA *i* and mRNA *k* in *n* samples were termed as $Lnc(i) = (lnc_1, lnc_2, \ldots, lnc_n)$ and $M(k) = (m_1, m_2, \ldots, m_n)$ in the expression matrix, respectively. The tumor purity of *n* patients was quantified *via* the *ESTIMATE* R package and termed as $p_1, p_2, \ldots, p_n$. The first-order partial correlation coefficients (PCCs) of lncRNA i and mRNA k were determined by performing the following operation:

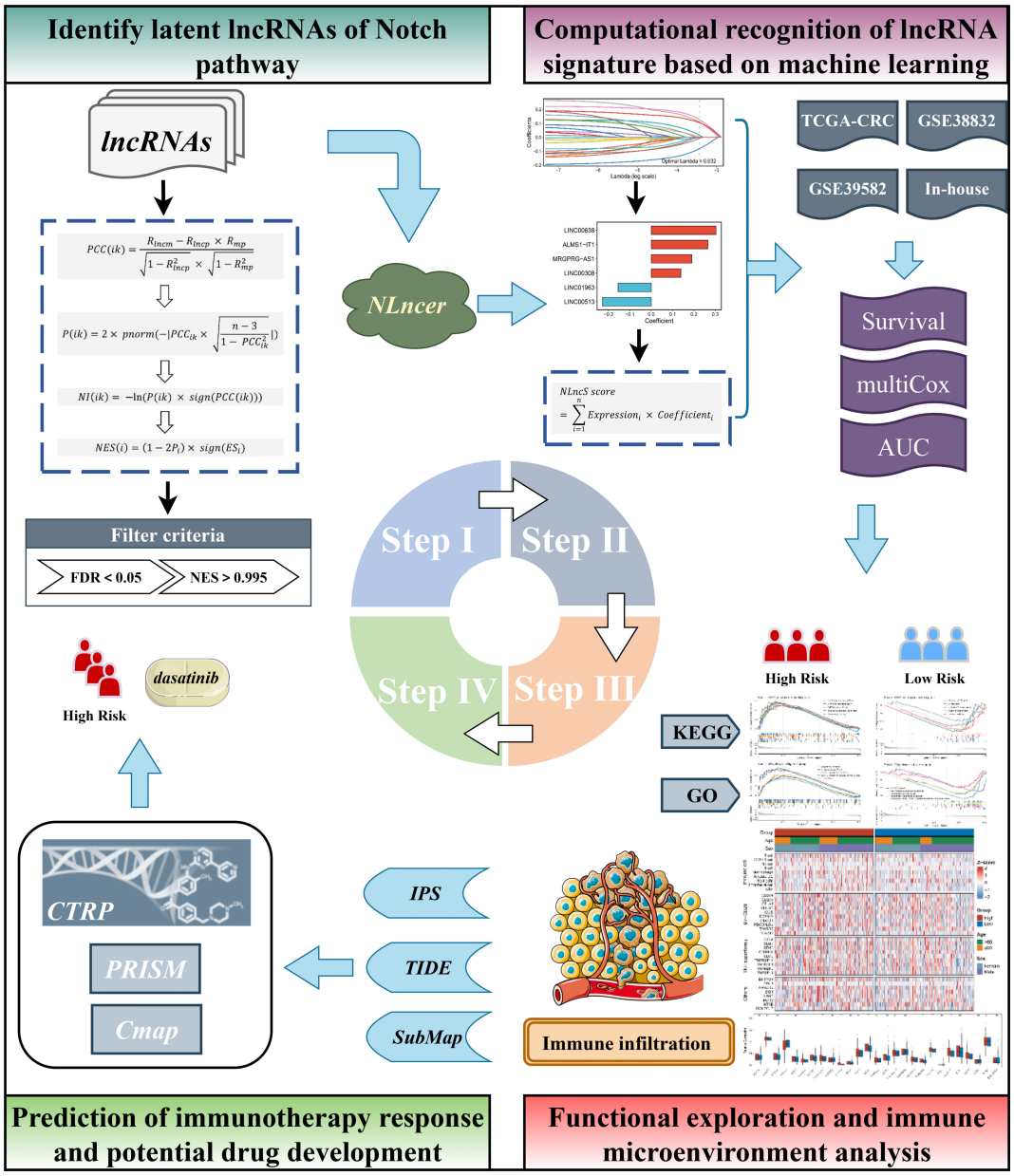


FIGURE 1
The workflow of the study.

$$PCC(ik) = \frac{R_{lncm} - R_{lncp} \times R_{mp}}{\sqrt{1 - R_{lncp}^2} \times \sqrt{1 - R_{mp}^2}}$$

R_{lncm}, R_{lncp}, and R_{mp} were named as the Pearson correlation coefficients of lncRNA i and mRNA k, lncRNA i and tumor purity p, and mRNA k and tumor purity p, respectively. Then, $P(ik)$, the P-value of $PCC(ik)$, was calculated:

$$P(ik) = 2 \times pnorm\left(-\left|PCC_{ik} \times \sqrt{\frac{n-3}{1 - PCC_{ik}^2}}\right|\right)$$

$pnorm$ is defined as the normal distribution function, and n is the sample size. The Notch index (NI) was measured:

$$NI(ik) = -\ln(P(ik) \times sign(PCC(ik)))$$

sign is a method that could realize the symbolic separation of functions. All mRNAs were sequenced according to descending *NI*, and further gene set concentration analysis (GSEA) was performed. The adjusted *P*-values and concentration scores (*ES*) of lncRNA *i* were evaluated by the *fgsea* R package and combined into a *NES*:

$$NES(i) = (1 - 2P_i) \times \ sign(ES_i)$$

Thereby, the *NES* was calculated to range from -1 to 1. Referring to previous findings (27, 28), lncRNAs with false discovery rate (FDR) <0.05 and *NES* absolute value >0.995 were filtrated as Notch-derived lncRNAs.

Signature generation

Given the comparability between disparate cohorts, lncRNA expression in three cohorts was converted to z-score before Notch-related lncRNA signatures (NLncS) were generated. According to the expression profile of Notch-derived lncRNAs, univariate Cox regression was calculated among TCGA-CRC, GSE39582, and GSE38832. Aware that the rigor of multiple test corrections and the small sample size may screen out some latent lncRNAs associated with survival, lncRNAs with unadjusted $P < 0.15$ and the same hazard ratio (HR) direction were adopted to construct the NLncS in combination with LASSO regression (29). The optimal lambda was obtained when the partial likelihood deviation achieved the minimum by adopting the 10-fold cross-validation algorithm. The lncRNAs with non-zero coefficients were involved in fitting signatures. The NLncS equation was computed with the following LASSO model weighting coefficient:

$$NLncS \ \ score = \sum_{i=1}^{n} Expression_i \ \times \ Coefficient_i$$

n refers to the total of significant lncRNAs, $Expression_i$ stands for the lncRNA *i* expression, and $Coefficient_i$ replaced the matching regression coefficient.

Clinical specimens and information collection

The ethics committee of the First Affiliated Hospital of Zhengzhou University provided consent to the study. One hundred fifteen pairs of CRC primary and normal tissues surgically resected in the First Affiliated Hospital of Zhengzhou University were included. The patient selection criteria were as follows: 1) no preoperative treatment, such as radiotherapy, chemotherapy, or targeted therapy was received; 2) no complication of any other tumors; 3) no autoimmune disease. Supplementary Table 1 provides the record of the baseline data. Fresh specimens were obtained and frozen in liquid nitrogen at -80°C for preservation. The clinical stage was in accordance with NCCN (2019) guidelines. Each individual signed informed consent. The relevant ethical review number is 2019-KY-423.

Quantitative real-time PCR

Tissue RNA was extracted by RNAiso Plus (Takara, China) reagent, and the quality was evaluated by NanoDrop One C (Waltham, USA). Complementary DNA (cDNA) was obtained by following the protocol of the Reverse Transcription Kit (Takara Bio, Japan). Then, quantitative real-time polymerase chain reaction (qRT-PCR) was performed using SYBR Assay I Low ROX (Eurogentec, USA) and SYBR® Green PCR Master Mix (Yeason, China). Each test was repeated three times. The expression level was quantized by $2^{-\Delta\Delta Ct}$ mode. GAPDH serves as an internal reference for normalization. The reader is referred to Supplementary Table 2 for primer sequence information.

Functional enrichment

For distinguishing the differences of distinct NLncS scores, the GSEA tool and *clusterProfiler* package were used to analyze the biological pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) (30). To get a standardized concentration, we chose the permutation to 1,000. Gene sets with adjusted *P*-value <0.05 were selected as significant.

Immune infiltration assessment

We adopted the *MCPcounter* package to explore immune cell type, stromal cell, and immune checkpoint (ICPs, recruited *B7-CD28* family, *TNF superfamily*, and others, Supplementary Table 5) abundance in CRC tissues (31). The correlation between these immune components and the NLncS model was further compared.

The mutation landscape and copy number variation of CRC

Somatic mutation and copy number variation (CNV) data were downloaded from TCGA-CRC portal and cBioportal website, respectively. The *TCGAbiolinks* R package was performed to get the raw mutation file. Mutations in different patient subpopulations were analyzed and visualized based on the *maftools* and *ComplexHeatmap* R package. CNV waterfall maps of the first 10 amplification (AMP) and homozygous

deletion (Homdel) chromosome fragments were visualized by the *ComplexHeatmap* package.

Estimation of methylation drivers

The raw methylation data were obtained from HumanMethylation450 array TCGA-CRC. The global methylation level (GML) of each TCGA-CRC sample was quantified based on the average beta value of a particular probe (32). The *MethylMix* package was adopted for integration of methylation and mRNA expression data. Methylation drivers, genes that were considerably inversely correlated with expression, were used to investigate the association with NLncS.

Evaluation of response to immunotherapy

Combining the immunophenoscore (IPS), tumor immune dysfunction and exclusion (TIDE, http://tide.dfci.harvard.edu/), and subclass mapping (SubMap) algorithms predicts responsiveness to immunotherapy (33–35). TIDE is an algorithm that integrates T-cell features to characterize immune evasion situations. The IPS Z-score was computed by assessing the scores of four immunophenotypes (antigenic presentation, effector cell, inhibitory cell, and ICP) (36). The higher IPS foreboded stronger immunogenicity of individuals. SubMap was adopted to judge the degree of similarity. The Bonferroni-corrected *P* was applied to indicate similarity, and the magnitude of the *P*-value is negatively correlated with similarity.

Prediction of therapeutic agents

The Cancer Therapeutics Response Portal datasets (CTRP) and PRISM databases store cancer cell line (CCL) sensitivity data with over 481 and 1448 compounds, respectively. In addition, both data provide the area under the curve (AUC) of dose–response as a measure of drug sensitivity. A lower AUC predicts stronger drug responsiveness. After filtering out compounds with ≥20% missing AUC values, the K-nearest neighbor (k-NN) imputation method was employed to interpolate missing variables for the remaining compounds.

Alternative compounds that target with high-risk groups

CMap (https://portals.broadinstitute.org/cmap/), a public online tool developed from the Broad Institute, was used to determine which drugs may have an effect on high-risk samples (37). Differential expression genes (DEGs) were obtained *via* the *limma* package. Subsequently, the top 150 upregulated and 150 downregulated genes were selected for CMap analysis. Each ranked list in CMap datasets was compared with DEGs to specify where the DEGs appeared, thus yielding a score of -100 to 100. Then the enrichment score was re-ranked; the top is strongly and positively correlated with the high risk, and the bottom is the opposite.

Statistical analysis

All work was performed in R software (4.1.0). Pearson's chi-square test was used for categorical variables. Continuous variables were compared adapting the Wilcoxon rank-sum test or Student's *t* test. Spearman analysis was used to analyze the correlation between groups. The *glmnet* package was adopted to fit LASSO regression. Cox regression and Kaplan–Meier (K-M) analysis were computed based on the *survival* package. The optimal cutoff value was chosen by the *survminer* package. Survival ROC and AUC were quantified *via* the *timeROC* package. $P < 0.05$ of two tails was judged statistically significant.

Results

The Notch pathway was significantly enriched and correlated with clinical features of CRC

For decades, the functional impact of the Notch pathway-based interplay with lncRNAs for neoplastic diseases has gradually become the focus of researchers' attention. In this context, we proposed whether a Notch pathway-derived global lncRNA signature could improve the outcomes and treatment efficacy and seek potential drugs for CRC patients. To address the question, we initially validated the Notch index of CRC primary and normal tissues in TCGA-CRC (for training dataset) cohort. Accordingly, the Notch signaling pathway was upregulated in tumor tissues (Figure 2A). Further, we utilized single-sample GSEA (ssGSEA) to evaluate the index of Notch pathway features for each sample. K-M analysis showed that the high Notch index predicted the adverse overall survival (OS, log-rank $P = 0.0001$) and relapse-free survival (RFS, log-rank $P = 0.0003$) of patients (Figures 2B, C). In addition, correlation analysis between the Notch index and clinical features showed that a high Notch index tended to be associated with a high clinical stage (such as AJCC stage, T, N, and M stage) and microsatellite stability (all $P < 0.05$, Figures 2D-H). Notably, the establishment of consensus molecular subtypes (CMSs) provides the most reliable classification system for CRC to date (38). We further explored the underlying link between the Notch index

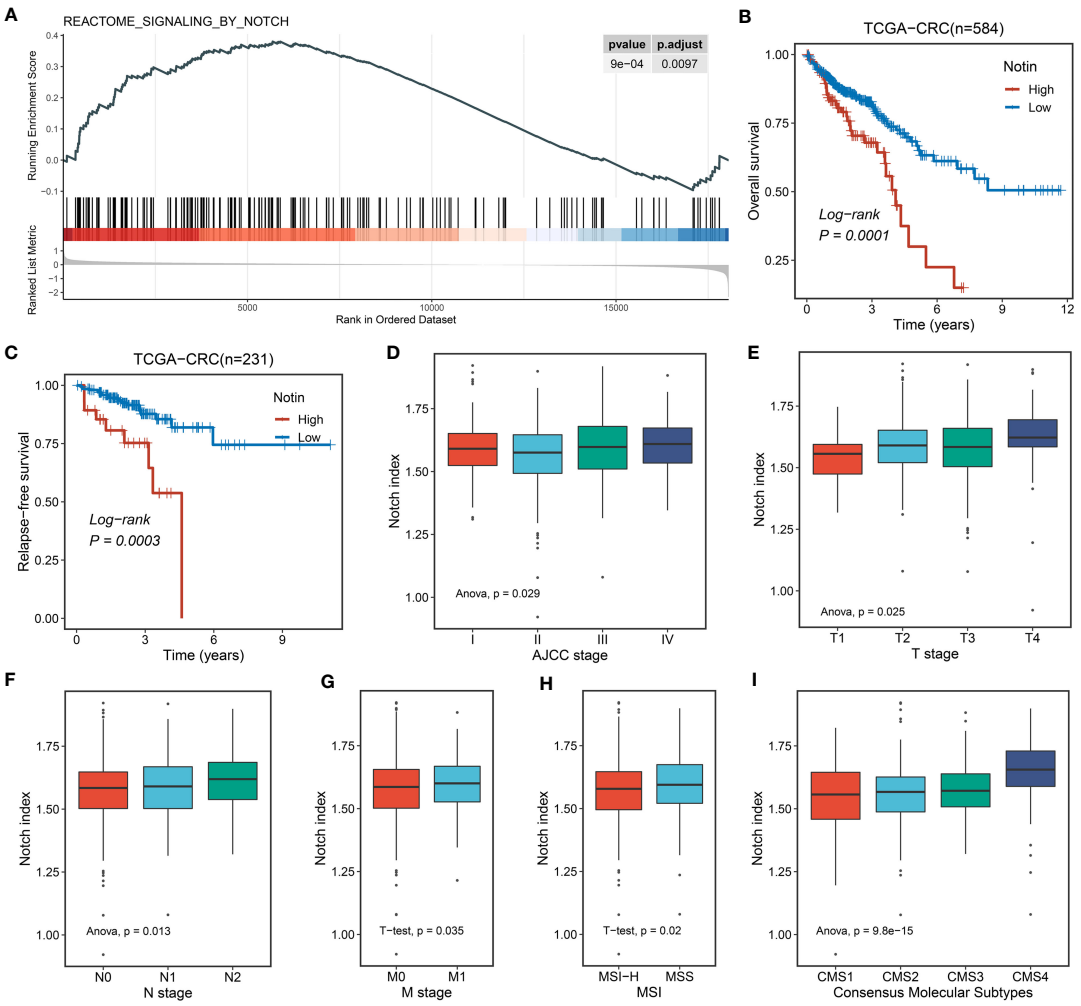


FIGURE 2
The potential biological significance of the Notch pathway in CRC. **(A)** Single-sample GSEA (ssGSEA) enrichment analysis of the Notch pathway in TCGA-CRC. **(B, C)** Kaplan–Meier curves of OS **(B)** and RFS **(C)** according to the Notch index (Notin) in TCGA-CRC. **(D-H)** Correlation between Notin and clinical features, such as AJCC stage **(D)**, T stage **(E)**, N stage **(F)**, M stage **(G)**, and microsatellite instability (MSI) state **(H)** of TCGA-CRC. **(I)** Correlation between Notin and classical consensus molecular subtypes (CMSs).

and CMS1–4 and found that CMS4 (mesenchymal subtype) had the highest Notch index ($P < 0.001$, Figure 2I). The CMS4 subtype prominently manifested an upregulation of epithelial–mesenchymal transition, angiogenesis, and lowest survival. Collectively, the above support that the Notch pathway has profound predictive value in CRC.

Nlncer to NLncS: Identifying Notch-derived LncRNA and generating a signature with the LASSO algorithm

For better consistency, we retained a total of 3,390 lncRNA molecules with relevant data in the three cohorts. Based on our previously developed framework of "TGFmitor" and the recognition of an immune-related lncRNA signature, similarly, we screened and obtained 24 candidate lncRNAs stably associated with the Notch pathway (27, 39, 40). According to the expression profiles of these 24 lncRNAs, we constructed the LASSO model *via* a 10-fold cross-validation. When the optimal lambda of 0.032 was selected, eight key lncRNA molecules with non-zero coefficients were identified (Figure 3A). Subsequently, after multivariate Cox regression, six lncRNAs stably associated with OS were used to construct the model, namely, LINC00638, ALMS1-IT1, MRGPRG-AS1, LINC00308, LINC01963, and LINC00513 (Figure 3B). Finally, a continuous risk score of the NLncS model was computed through a linear combination of regression coefficient-weighted expression values of the six lncRNAs. Next, we distinguish the high- and low-risk samples according to the median value. In TCGA-CRC, GSE38832, and

GSE39582, the high-risk group presented a prominently worse prognosis than the low-risk group (all Log-rank P <0.0001, Figures 3C, D, F). After incorporating available clinical characteristics including age, sex, stage (T, N, M, and AJCC), and microsatellite instability (MSI) state, multivariate Cox regression showed that NLncS were still statistically significant. This suggests that NLncS can be used as an independent prognostic factor for CRC (all P < 0.001, Figures 3E, G, Supplementary Figure 1A). We detected the recognition of NLncS and calculated the AUC at 1, 3, and 5 years of TCGA-CRC (0.780, 0.780, and 0.806), GSE38832 (0.701, 0.762, and 0.768), and GSE39582 (0.701, 0.762, and 0.768). We compared the NLncS model with three other lncRNA signatures. It was found that our NLncS model had the highest C-index and was significantly better than the other three lncRNA prediction models (*P < 0.05, ****P < 0.0001, Supplementary Figure 3, Supplementary Table 4) (41–43). Therefore, NLncS possessed the ability of robust prediction in CRC patients.

Clinical in-house cohort for NLncS validation *via* qRT-PCR

To further explore the potential for clinical translational applications of NLncS, we examined the expression of lncRNAs *via* qRT-PCR and calculated risk scores in a clinical in-house cohort. The results are plotted in Figure 4. Patients with high-risk scores had lower OS (Figure 4A) and disease-free survival (DFS, Figure 4B), which validates the prognostic predictive power of the model (P < 0.0001 for both). After inclusion of clinical and pathological features, multivariate analysis showed that risk score was an independent indicator in both OS (Figure 4C, P < 0.001) and DFS (Figure 4D, P = 0.006). What is more, the AUC calculation of the model showed that the 1-, 3-, and 5-year AUC values were 0.703, 0.916, and 0.837 for OS and 0.674, 0.911, and 0.770 for DFS (Figures 4E, F), respectively, showing good predictive power.

Immune-related mechanisms were upregulated in the high-risk group, showing potential immunologic properties

Further, we performed GSEA to explore the potential biological mechanisms that may cause differences on two groups. We found a significant enrichment of signaling pathways in the high-risk group such as immunity (humoral and cellular immunity) and regulation of cellular differentiation (Figures 5A, C), for instance, B-cell receptor and chemokine signaling, T-cell activation and phagocytosis, and cellular differentiation-related pathways such as the MAPK pathway. However, the downregulation of energy metabolism and redox reactions was predominantly related with the low-risk group, including glycolysis, tricarboxylic acid cycle (TCA cycle), aerobic respiration, and oxidative phosphorylation (Figures 5B, D). Hence, it is manifest that immune-related roles may be influencing factors contributing to NLncS predicting differences.

Tumor immune microenvironment landscape and immune checkpoint profiles of NLncS

Since immune-related pathways differ among subgroups of NLncS, we expanded our analysis of the tumor immune microenvironment (TIME) landscape consisting of nine immune cell types and 27 immune checkpoints to further distinguish immune features. The scenario is depicted in Figure 5E. Moreover, CD8+ T cells, macrophages, endothelial cells, and cancer-associated fibroblasts (CAFs) were significantly enriched in the high-risk group (Supplementary Figure 1B). Specifically, HHLA2 was overexpressed in the low-risk group, while ICOSLG, PDCD1, PDCD1LG2, VTCN1, CD40, TNFRSF9, TNFSF14, ENTPD1, HAVCR2, and LAG3 were considerably overexpressed in the high-risk group (Figure 5F). Thus viewed, both at the level of pathway enrichment, cell infiltration, and molecular expression, the high- and low-risk groups showed more or less imparities.

Somatic mutation landscape, CNVs, and latent methylation driver in CRC

Previously, it has been reported that tumors with a high mutation load were more likely to respond to clinical strategies based on immune checkpoint blockers (44–47). Indeed, based on the immunological background of NLncS, we next estimated the effects of somatic mutation on NLncS for seeking clinical benefit. Figure 6A depicts the mutational landscape of NLncS. APC and TP53 ranked first and second with 79% and 61% mutation frequencies, respectively, which supported that high mutation rates of APC and TP53 might be responsible for giving rise to CRC (Figure 6A). Given that CNV dominatingly consisted of amplification (AMP) and homozygous deletion (HOMDEL), we sequenced the genes by frequency of AMP and HOMDEL (Figure 6B). The results revealed that in the high-risk group, TTPAL, RAE1, R3HDML, PABPC1L, LINC01620, and LINC01430 were visibly amplified; dramatically deletions were RBFOX1, WWOX, and MACROD2. Interestingly, there was merely a slight difference in HOMDEL of the low-risk group. Visible amplifications were TTLL9, TM9SF4, POFUT1, PDRG1, MYLK2, and FOXS1.

Considering that the methylation modifications could alter the biological function of RNA molecules at the epigenetic stage, we performed a global methylation overview of CRC (48, 49). With the help of the *MethylMix* R package and Wilcox test, we found six

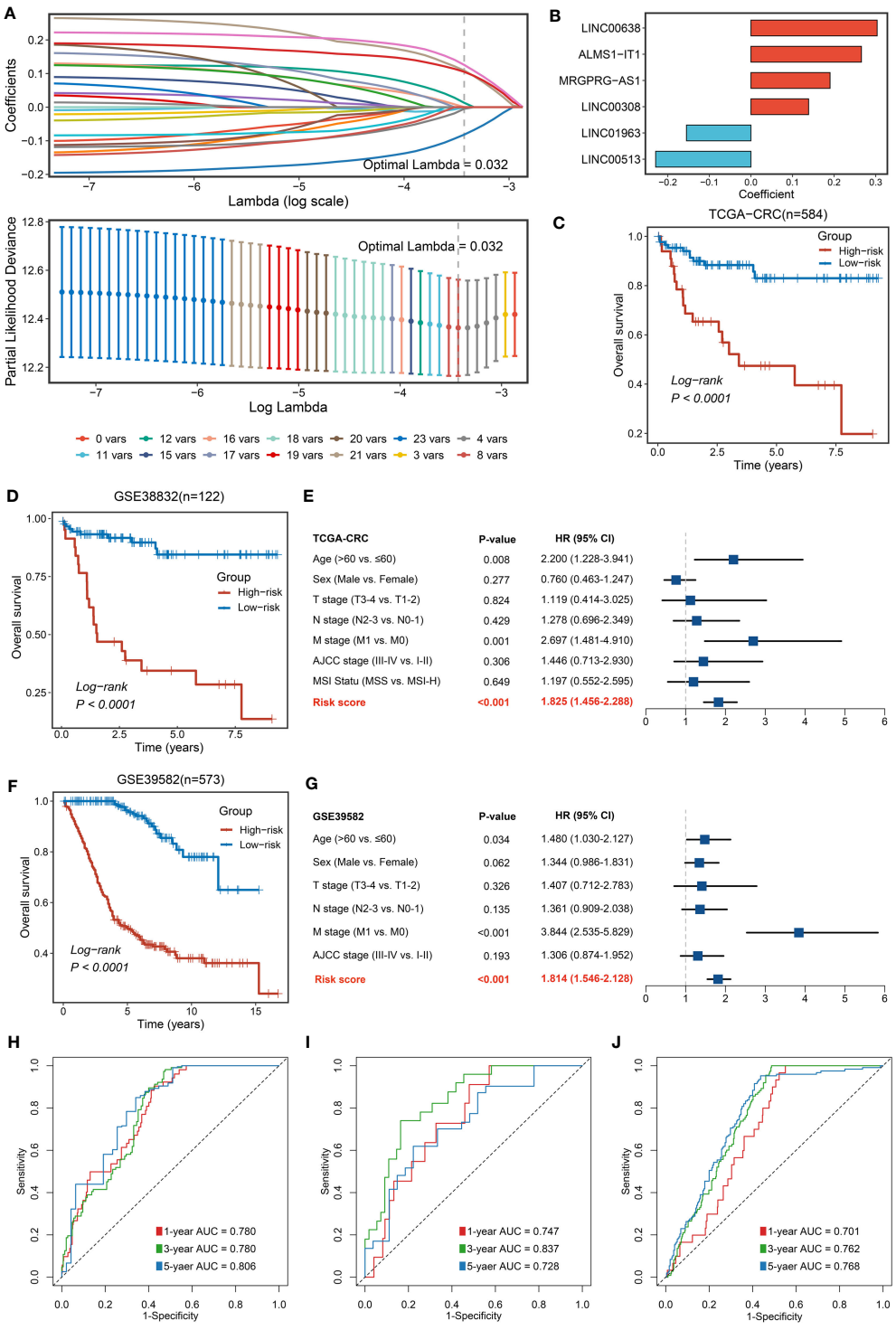


FIGURE 3
Construction and validation of NLncS. **(A)** Determination of the optimal lambda was obtained when the partial likelihood deviance reached the minimum value. **(B)** Univariate Cox regression was further performed to generate the key lncRNAs with non-zero coefficients and predictive values. LASSO coefficient profiles of the candidate lncRNAs for NLncS construction. **(C, D, F)** Kaplan–Meier curves of OS according to NLncS in TCGA-CRC **(C)**, GSE38832 **(D)**, and GSE39582 **(F)**. **(E, G)** Multivariable Cox regression analysis of NLncS in TCGA-CRC **(E)** and GSE39582 **(G)**. **(H-J)** Time-dependent ROC analysis for predicting OS at 1, 3, and 5 years in TCGA-CRC **(H)**, GSE38832 **(I)**, and GSE39582 **(J)**.

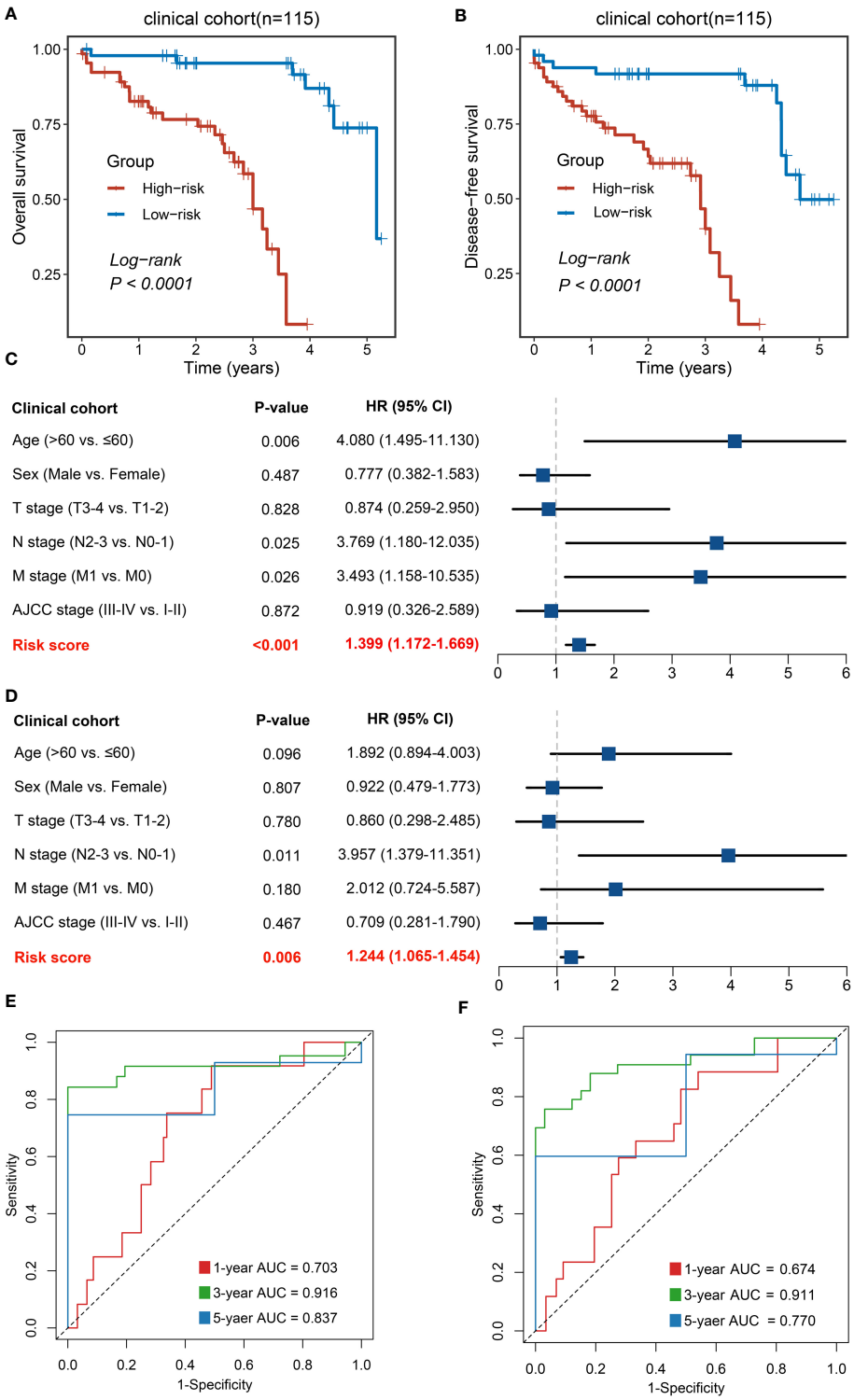


FIGURE 4
Validation of NLncS *via* qRT-PCR. **(A, B)** Kaplan–Meier curves of OS **(A)** and DFS **(B)** based on NLncS. **(C, D)** Multivariable Cox regression analysis of OS **(C)** and DFS **(D)**. **(E, F)** Time-dependent ROC analysis for predicting OS **(E)** and DFS **(F)** at 1, 3, and 5 years.

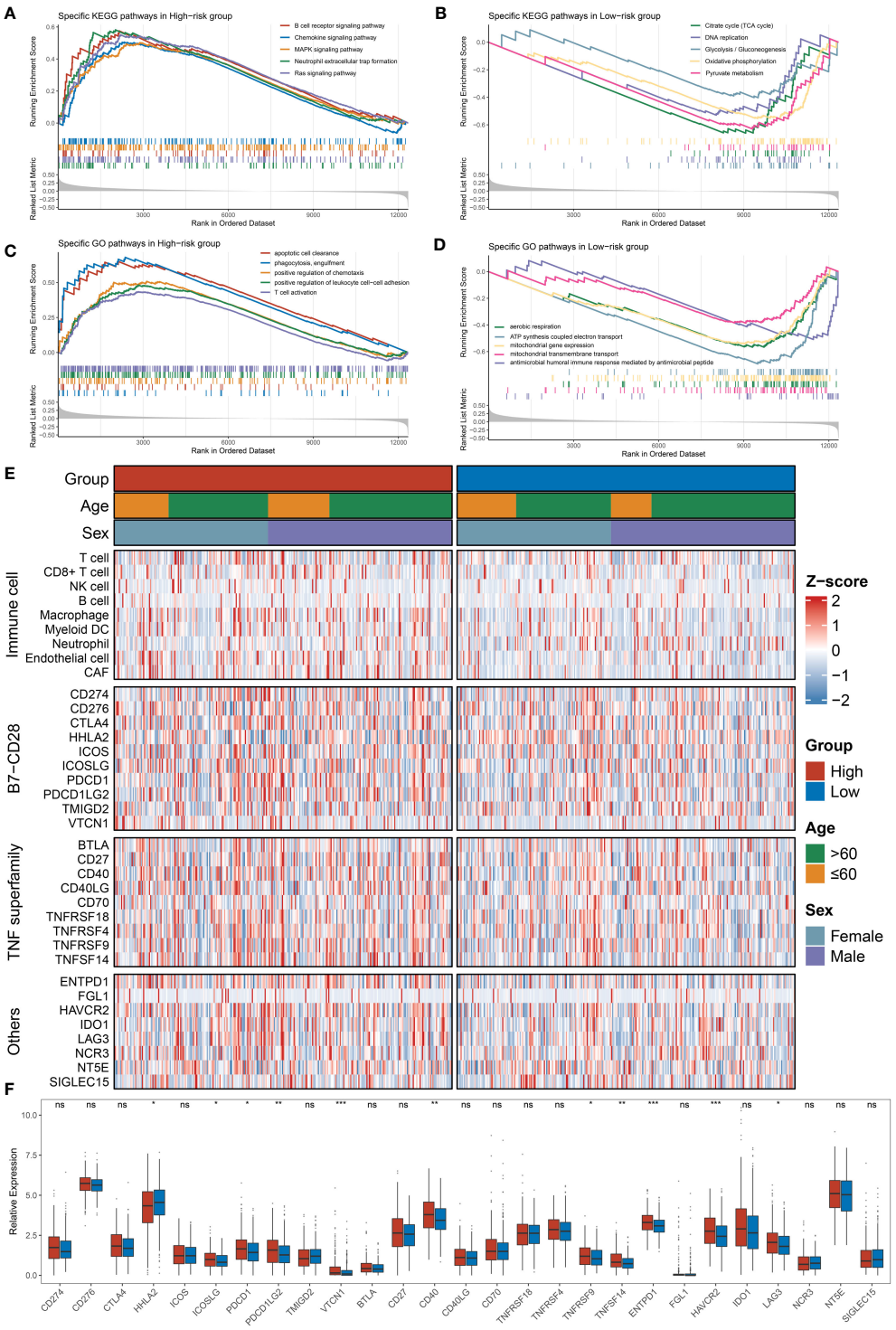


FIGURE 5
Functional enrichment analysis and immune infiltration analysis were performed in high- and low-risk groups. **(A, B)** The top five KEGG-enriched pathways and immune landscapes in the high- and low-risk groups. **(C, D)** The top five GO-enriched pathways and immune landscapes in the high- and low-risk groups. **(E)** The heatmaps of eight immune cell and 27 immune checkpoint profiles in the high- and low-risk groups. **(F)** The relative expression of 27 immune checkpoints in the high- and low-risk groups. (ns, none significance, P> 0.05; *P< 0.05; **P< 0.01; ***P< 0.001).

methylation driver genes whose methylation levels were significantly negatively correlated with expression in two groups, including CES3, FAM127B, MAP9, PARVB, RLN2, and ZNF548. Strikingly, the growing NLncS fraction was accompanied by increased methylation levels of CES3 and RLN2 but decreased expression levels (Supplementary Figure 2; Figures 6C, D). This suggested that CES3 and RLN2 might play an antitumor role as protective factors in the high-risk group, while methylation

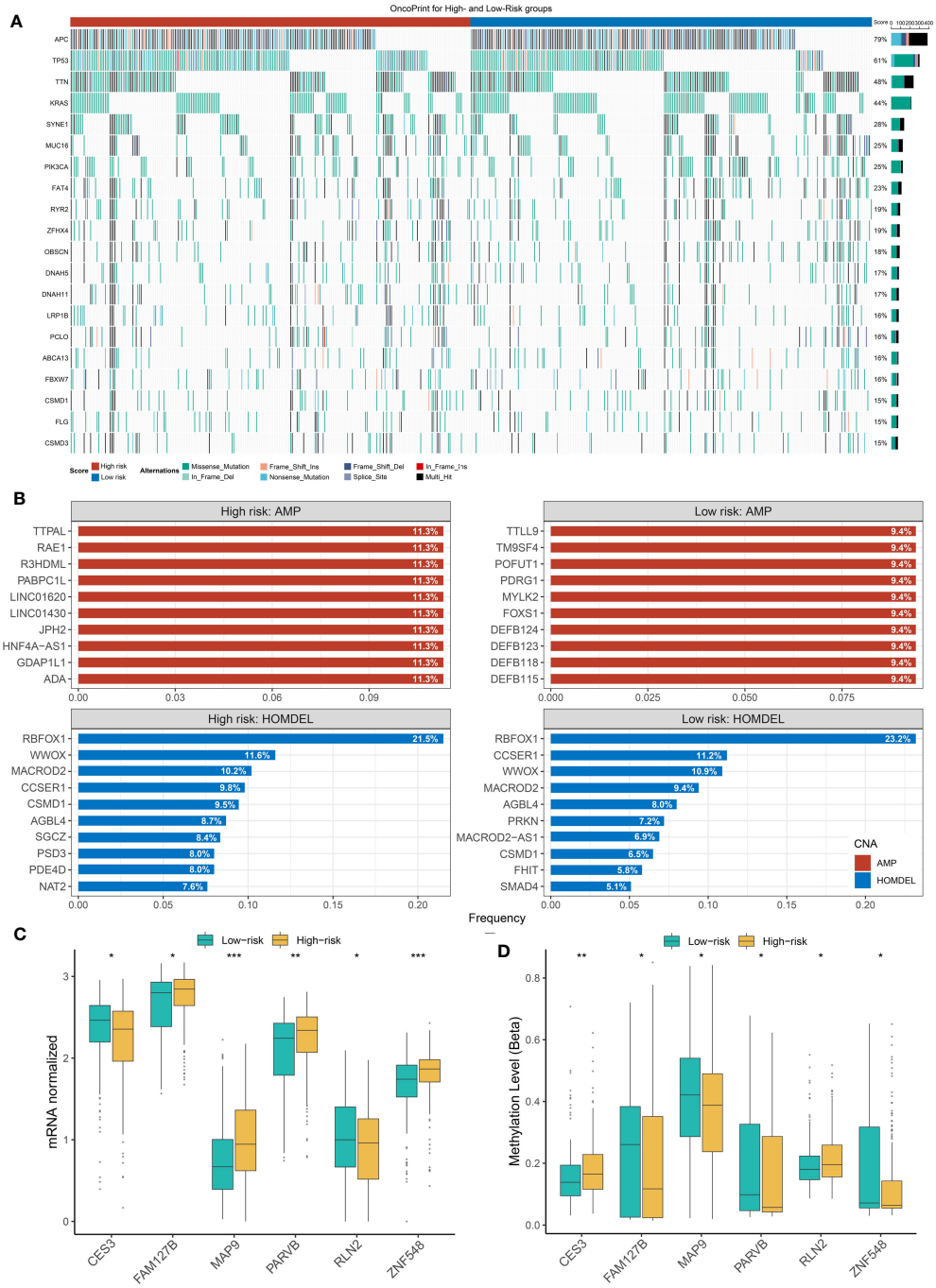


FIGURE 6
The Landscapes of frequently mutated genes (FMGs) and methylation level in two groups of NLncS. **(A)** The oncoplot depicts the discrepancies in FMGs of CRC among the three cohorts. **(B)** The top 10 genes were amplified and homozygously deleted in the high- and low-risk groups. **(C, D)** The expression and methylation level of the methylation drivers in CRC. (*P< 0.05; **P< 0.01; *** P< 0.001).

modification silenced the corresponding mRNA fragments resulting in CES3 and RLN2 reduced expression levels. Collectively, this suggested that the influence of mutational burden, CNV, and methylation level on NLncS scores was likely non-redundant.

Deducing response to immunotherapy

As we all predicted, immunogenicity predicts stimulation to the immune system (50). As an evaluation tool of tumor immunogenicity, IPS was adopted to indirectly judge the local immune activation status of the sample. Scores were calculated for four different immunophenotypes (antigen presentation, effector cells, suppressor cells, checkpoints). The Z-score was the integration of the four. As expected, we found a significantly high Z-score in high-risk patients, indicating greater immunogenicity (P <0.0001, Figure 7A). Guided by the TIDE network tool, the response rate to immunotherapy (50.7%) of the high-risk group was markedly higher (P = 0.001, Figure 7B). Further, according to the SubMap method, we compared the similarity in mRNA expression patterns between two groups of CRC patients and 47 patients who responded (R) or did not respond (NR) to immunotherapy. The result showed that high-risk patients were closer to those who responded to immunotherapy (Bonferroni corrected P = 0.032, Figure 7C). These results strongly indicate that high-risk individuals could benefit from immunotherapy.

Exploring insidious therapeutic agents for high-risk CRC individuals

Herein, referring to the workflow of Yang et al., we mined optional drugs for high-risk patients following the framework shown in Figure 7D (51). In order to obtain robust therapeutic drugs, we used two databases that store gene expression profile data and drug sensitivity data for CCLs, CTRP, and PRISM. After removing drugs with >20% missing values, we obtained 266 and 1285 compounds from CTRP and PRISM, respectively. We selected ridge regression to predict the drug response for each patient and calculated the AUC for each compound, with a lower AUC suggesting increased sensitivity. Then, we evaluated the appropriate drug in terms of the difference in response and correlation analysis (Figures 7E, F). For the difference analysis, we set a strict threshold Log2 FC >0.01 (CTRP) and Log2 FC >0.05 (PRISM), respectively, to identify compounds with lower AUC values, which may act as potential clinical therapies for high-risk individuals. We employed coefficients of r <−0.2 (CTRP) and r < −0.1 (PRISM) to constrain compounds that are more closely related to the NLncS score. In Figures 7E, F, a total of six compounds from CTRP and four compounds from PRISM were finally determined (all P < 0.05). Further, to more deeply explain the functional links between compounds and RNA molecules as well as disease states, we performed the CMap analysis to explore available drugs (37, 52). Surprisingly, we found that dasatinib, widely used as a first-line antineoplastic agent for leukemia, had a high negative correlation, indicating that it probably to exerts therapeutic effects for high-risk score patients (Figure 7G).

Discussion

As early as 1917, Morgan and colleagues first discovered the Notch gene in mutant flies, and then the Notch signaling pathway was gradually reported (53). Notch signaling is a classical pathway for tumorigenesis and disease progression by accommodating cell proliferation and differentiation, among others (54). Multiple Notch receptors exhibit oncogenic or tumor-suppressive effects in various cells (55). Many clinical trials have also assessed the anticancer efficacy of Notch inhibitors (56). Notably, the interplay of lncRNA molecules with Notch pathway-related molecules regulates a variety of malignant phenotypes of tumors (23–25). Therefore, this study is based on the lncRNA profiles from Notch signaling and aims to establish a model with the ability to predict prognosis and immune response and assist in the screening of sensitive drugs.

The symptoms of CRC are insidious and often lead to the majority of patients presenting at an advanced stage, which contributed to an unfavorable outcome (57). Presently, with the arrival of the era of individualized treatment, the stratification and refined management system for clinical patients urgently need to be improved, and CRC patients are no exception. However, current clinical measures are immature, which is reflected in the inaccuracy of disease prognosis evaluation.

Single biomarkers are also increasingly unable to meet the needs of practical applications, and more and more attention has been paid to synthesizing multiple data types to form appropriate scoring rules (58). We developed an evaluation system consisting of six key lncRNA molecules by LASSO algorithm and Cox regression analysis to perform a risk score that can be continuously quantified in CRC patients. The model showed good prognostic evaluation ability in three public cohorts and one clinical cohort, which can accurately distinguish the high-risk group to facilitate the layer management for clinical patients. This was also initially validated in an internal cohort of 115 clinical individuals.

Consistently, how to evaluate the efficacy of ICP therapy is a problem faced by the majority of oncologists (59–61). On the one hand, “magic drugs” will bring a resurrection effect to some patients, but on the other hand, improper use is likely to cause various side effects (62). Given this, we considered whether the NLncS scoring model could reflect the difference in immune response among distinct patients. We found that there was a significant difference in the pathways enriched. The high-risk group was primarily enriched in cellular or molecular pathways involved in immune processes, including B-cell receptor,

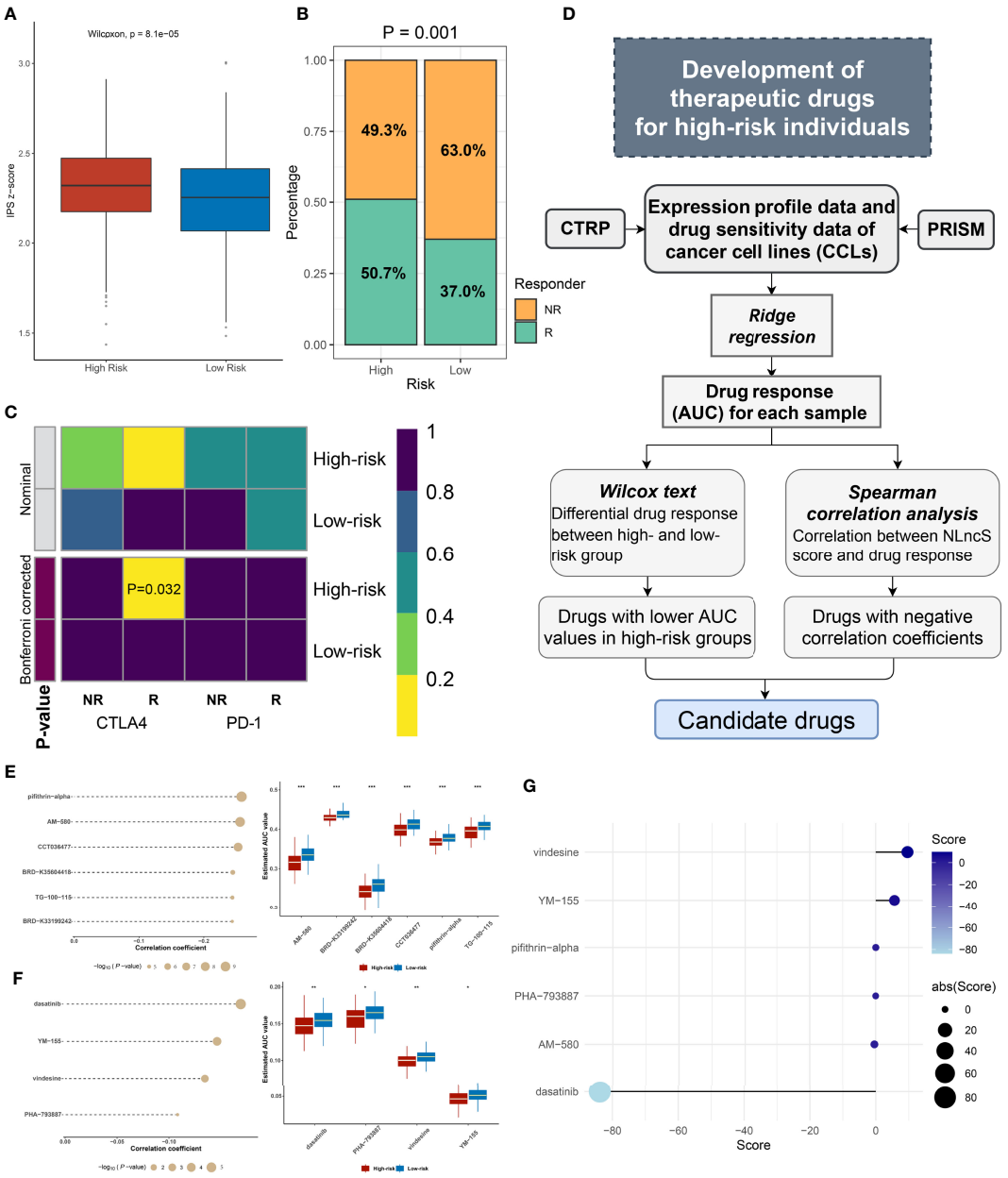


FIGURE 7
Efficacy evaluation of immunotherapy and potential drug screen for CRC patients. **(A)** The IPS z-score of the two groups. **(B)** The predicted efficacy of the two groups to immunotherapy in TCGA cohorts via TIDE method. **(C)** SubMap analysis of the two groups and 47 pretreated patients with comprehensive immunotherapy annotations in TCGA cohort. For SubMap analysis, a smaller P-value implied a higher similarity of paired expression profiles. **(D)** The framework of exploring insidious drugs for high-risk patients. **(E, F)** The candidate drugs screened by CTRP and PRISM datasets. The correlation analysis of drugs and high-risk groups in CTRP **(E** left) and PRISM **(F** left). The differential drug response between high- and low-risk groups in CTRP **(E** right) and PRISM **(F** right). **(G)** CMap analysis identified dasatinib as a potential compound able to target the high-risk patient. (*P< 0.05; **P< 0.01; ***P< 0.001).

chemokine pathways, T-cell activation, and phagocytosis pathways. The low-risk group was primarily downregulated among metabolic aspects (glycolysis, TCA cycle, and oxidative phosphorylation, etc.). Further, immune landscape and immune infiltration analysis suggested visible differences and abundance. Meanwhile, analyses of mutation and methylation levels showed that differences at the genome level were not redundant.

Growing studies have discussed two environments before immune escape occurs in the tumor: 1) ineffective infiltration of a large number of inactive T cells (63); 2) destruction of T cells

infiltrating the tumor by immunosuppressive molecules (64). Remarkably, Peng Jiang et al. raised a novel computational architecture, TIDE score, to synthesize these two environments (33). The project involved 189 studies with 33,197 specimens and was thought to be an alternative to a single predictor to analyze the efficacy of immune checkpoint suppression. As predicted, we found that the response rate to immunotherapy in high-risk (50.7%) patients was significantly better than that in low-risk (37.0%) patients (P = 0.001). An additional result of SubMap reported that the high-risk group appeared to be more sensitive to CTLA4-blocker-based therapy. These results validate our conjecture that the high-risk group is suitable for immunotherapy.

Dasatinib, as a first-line strategy for metastatic non-small cell lung cancer, can also be widely used as an adjuvant therapy for patients with pancreatic cancer, imatinib-resistant chronic myelogenous leukemia, etc. (65, 66). It is gratifying that we found that dasatinib is an excellent alternative drug for high-risk patients after drug sensitivity prediction and CMap analysis, which not only shows good sensitivity but also reflects powerful targeting ability. However, there is no evidence that dasatinib could be used as a first-line drug in CRC patients yet. Follow-up basic experimental and clinical studies are still needed before dasatinib is adopted to CRC patients. Nonetheless, we argued that dasatinib would possess a bright prospect in improving the prognosis of CRC patients.

Collectively, this study integrally delineated Notch-derived lncRNAs through a "NLncer" workflow and further constructed a systematic scoring system (termed "NLncS") for accurately and stably evaluating prognosis and immune efficacy in CRC. This work may contribute to interpret patient characteristics and directed therapy. In addition, dasatinib might get a preferential seat in the first line in CRC treatment.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethnics Committee of The First Affiliated Hospital of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

QD made the conceptualization. ZS, ZL, and YL were involved in the methodology. ZS, LL, CW, and WY provided the resources. QD, LZ, and YL analyzed the data. QD and YL prepared the original draft. QD and WW reviewed and edited the manuscript. ZS, LL, and CW supervised the study. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by The National Natural Science Foundation of China (81972663, 82173055, U2004112), The Excellent Youth Science Project of Henan Natural Science Foundation (212300410074), The Key Scientific Research Project of Henan Higher Education Institutions (20A310024), The Youth Talent Innovation Team Support Program of Zhengzhou University (32320290), The Provincial and Ministry co-constructed key projects of Henan Medical Science and Technology (SBGJ202102134), Key Scientific and Technological Research Projects of Henan Provincial Department of Science and Technology (212102310117), Henan Provincial Health Commission and Ministry of Health Co-construction Project, and Henan Provincial Health and Health Commission Joint Construction Project (LHGJ20200158).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.953405/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The multivariable Cox regression analysis of NLncS in GSE38832 **(A)**. And the differences in the abundance of cell infiltrates between high- and low-risk groups **(B)**.

SUPPLEMENTARY FIGURE 2
Correlation analysis between expression of identified methylation drivers and methylation levels.

SUPPLEMENTARY FIGURE 3
Comparison of predictive efficacy of NLncS model with three published studies.

References

1. Jian X, He H, Zhu J, Zhang Q, Zheng Z, Liang X, et al. Hsa_circ_001680 affects the proliferation and migration of CRC and mediates its chemoresistance by regulating BMI1 through miR-340. *Mol Canc* (2020) 19:20. doi: 10.1186/s12943-020-1134-8

2. Hernandez-Alejandro R, Ruffolo LI, Sasaki K, Tomiyama K, Orloff MS, Pineda-Solis K, et al. Recipient and donor outcomes after living-donor liver transplant for unresectable colorectal liver metastases. *JAMA Surg* (2022) 157:524–30. doi: 10.1001/jamasurg.2022.0300

3. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* (2019) 394:1467–80. doi: 10.1016/S0140-6736(19)32319-0

4. Liao T-T, Lin C-C, Jiang J-K, Yang S-H, Teng H-W, Yang M-H. Harnessing stemness and PD-L1 expression by AT-rich interaction domain-containing protein 3B in colorectal cancer. *Theranostics* (2020) 10:6095–112. doi: 10.7150/thno.44147

5. Luke JJ, Lemons JM, Karrison TG, Pitroda SP, Melotek JM, Zha Y, et al. Safety and clinical activity of pembrolizumab and multisite stereotactic body radiotherapy in patients with advanced solid tumors. *J Clin Oncol* (2018) 36:1611–8. doi: 10.1200/JCO.2017.76.2229

6. Fang H, Guo Z, Chen J, Lin L, Hu Y, Li Y, et al. Combination of epigenetic regulation with gene therapy-mediated immune checkpoint blockade induces anti-tumour effects and immune response *in vivo*. *Nat Commun* (2021) 12:6742. doi: 10.1038/s41467-021-27078-x

7. Roh W, Chen P-L, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* (2017) 9: eaah3560. doi: 10.1126/scitranslmed.aah3560

8. Zhang C, Zeng Z, Cui D, He S, Jiang Y, Li J, et al. Semiconducting polymer nano-PROTACs for activatable photo-immunometabolic cancer therapy. *Nat Commun* (2021) 12:2934. doi: 10.1038/s41467-021-23194-w

9. Zhou B, Lin W, Long Y, Yang Y, Zhang H, Wu K, et al. Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduct Target Ther* (2022) 7:95. doi: 10.1038/s41392-022-00934-y

10. Varga J, Nicolas A, Petrocelli V, Pesic M, Mahmoud A, Michels BE, et al. AKT-dependent NOTCH3 activation drives tumor progression in a model of mesenchymal colorectal cancer. *J Exp Med* (2020) 217:e20191515. doi: 10.1084/jem.20191515

11. Wieland E, Rodriguez-Vita J, Liebler SS, Mogler C, Moll I, Herberich SE, et al. Endothelial Notch1 activity facilitates metastasis. *Cancer Cell* (2017) 31:355–67. doi: 10.1016/j.ccell.2017.01.007

12. Jiang N, Zou C, Zhu Y, Luo Y, Chen L, Lei Y, et al. HIF-1α-regulated miR-1275 maintains stem cell-like phenotypes and promotes the progression of LUAD by simultaneously activating wnt/β-catenin and notch signaling. *Theranostics* (2020) 10:2553–70. doi: 10.7150/thno.41120

13. Ciofani M, Zúñiga-Pflücker JC. Notch promotes survival of pre-T cells at the beta-selection checkpoint by regulating cellular metabolism. *Nat Immunol* (2005) 6:881–8. doi: 10.1038/ni1234

14. Meurette O, Stylianou S, Rock R, Collu GM, Gilmore AP, Brennan K. Notch activation induces akt signaling *via* an autocrine loop to prevent apoptosis in breast epithelial cells. *Cancer Res* (2009) 69:5015–22. doi: 10.1158/0008-5472.CAN-08-3478

15. Dufraine J, Funahashi Y, Kitajewski J. Notch signaling regulates tumor angiogenesis by diverse mechanisms. *Oncogene* (2008) 27:5132–7. doi: 10.1038/onc.2008.227

16. Krishna BM, Jana S, Singhal J, Horne D, Awasthi S, Salgia R, et al. Notch signaling in breast cancer: From pathway analysis to therapy. *Cancer Lett* (2019) 461:123–31. doi: 10.1016/j.canlet.2019.07.012

17. Wang L, Yu S, Chan ER, Chen K-Y, Liu C, Che D, et al. Notch-regulated dendritic cells restrain inflammation-associated colorectal carcinogenesis. *Cancer Immunol Res* (2021) 9:348–61. doi: 10.1158/2326-6066.CIR-20-0428

18. Maekawa Y, Ishifune C, Tsukumo S-i, Hozumi K, Yagita H, Yasutomo K. Notch controls the survival of memory CD4+ T cells by regulating glucose uptake. *Nat Med* (2015) 21:55–61. doi: 10.1038/nm.3758

19. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov* (2011) 1:391–407. doi: 10.1158/2159-8290.CD-11-0209

20. Yang X, Zhang S, He C, Xue P, Zhang L, He Z, et al. METTL14 suppresses proliferation and metastasis of colorectal cancer by down-regulating oncogenic long non-coding RNA XIST. *Mol Canc* (2020) 19:46. doi: 10.1186/s12943-020-1146-4

21. Han M, Wang S, Fritah S, Wang X, Zhou W, Yang N, et al. Interfering with long non-coding RNA MIR22HG processing inhibits glioblastoma progression through suppression of wnt/β-catenin signalling. *Brain* (2020) 143:512–30. doi: 10.1093/brain/awz406

22. Qin G, Tu X, Li H, Cao P, Chen X, Song J, et al. Long noncoding RNA p53-stabilizing and activating RNA promotes p53 signaling by inhibiting heterogeneous nuclear ribonucleoprotein K deSUMOylation and suppresses hepatocellular carcinoma. *Hepatology* (2020) 71:112–29. doi: 10.1002/hep.30793

23. Tao S, Chen Q, Lin C, Dong H. Linc00514 promotes breast cancer metastasis and M2 polarization of tumor-associated macrophages *via* Jagged1-mediated notch signaling pathway. *J Exp Clin Cancer Res* (2020) 39:191. doi: 10.1186/s13046-020-01676-x

24. Hu G, Ma J, Zhang J, Chen Y, Liu H, Huang Y, et al. Hypoxia-induced lncHILAR promotes renal cancer metastasis *via* ceRNA for the miR-613/206/ 1-1-3p/Jagged-1/Notch/CXCR4 signaling pathway. *Mol Ther* (2021) 29:2979–94. doi: 10.1016/j.ymthe.2021.05.020

25. Mao X, Ji T, Liu A, Weng Y. ELK4-mediated lncRNA SNHG22 promotes gastric cancer progression through interacting with EZH2 and regulating miR-200c-3p/Notch1 axis. *Cell Death Dis* (2021) 12:957. doi: 10.1038/s41419-021-04228-z

26. Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* (2008) 24:2395–6. doi: 10.1093/bioinformatics/btn429

27. Liu Z, Lu T, Wang Y, Jiao D, Li Z, Wang L, et al. Establishment and experimental validation of an immune miRNA signature for assessing prognosis and immune landscape of patients with colorectal cancer. *J Cell Mol Med* (2021) 25:6874–86. doi: 10.1111/jcmm.16696

28. Li Y, Jiang T, Zhou W, Li J, Li X, Wang Q, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun* (2020) 11:1000. doi: 10.1038/s41467-020-14802-2

29. Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* (1997) 16:385–95. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

31. Liu Z, Lu T, Li J, Wang L, Xu K, Dang Q, et al. Clinical significance and inflammatory landscape of aNovel recurrence-associated immune signature in stage II/III colorectal cancer. *Front I Immunol* (2021) 12:702594. doi: 10.3389/fimmu.2021.702594

32. Jung H, Kim HS, Kim JY, Sun J-M, Ahn JS, Ahn M-J, et al. DNA Methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun* (2019) 10:4278. doi: 10.1038/s41467-019-12159-9

33. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1

34. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019

35. Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One* (2007) 2: e1195. doi: 10.1371/journal.pone.0001195

36. Yang Z, Yan G, Zheng L, Gu W, Liu F, Chen W, et al. As a potential predictor of prognosis and immunotherapy response for oral squamous cell carcinoma, is related to cell invasion, metastasis, and CD8+ T cell infiltration. *Oncoimmunology* (2021) 10:1938890. doi: 10.1080/2162402X.2021.1938890

37. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* (2006) 313:1929–35. doi: 10.1126/science.1132939

38. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* (2015) 21:1350–6. doi: 10.1038/nm.3967

39. Liu Z, Weng S, Xu H, Wang L, Liu L, Zhang Y, et al. Computational recognition and clinical verification of TGF-β-Derived miRNA signature with potential implications in prognosis and immunotherapy of intrahepatic cholangiocarcinoma. *Front Oncol* (2021) 11:757919. doi: 10.3389/fonc.2021.757919

40. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun* (2022) 13:816. doi: 10.1038/s41467-022-28421-6

41. Huang R, Zhou L, Chi Y, Wu H, Shi L. LncRNA profile study reveals a seven-lncRNA signature predicts the prognosis of patients with colorectal cancer. *Biomark Res* (2020) 8:8. doi: 10.1186/s40364-020-00187-3

42. Zhang Z, Liu Q, Wang P, Li J, He T, Ouyang Y, et al. Development and internal validation of a nine-lncRNA prognostic signature for prediction of overall survival in colorectal cancer patients. *PeerJ* (2018) 6:e6061. doi: 10.7717/peerj.6061

43. Zeng J-H, Liang L, He R-Q, Tang R-X, Cai X-Y, Chen J-Q, et al. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma. *Oncotarget* (2017) 8:16811–28. doi: 10.18632/oncotarget.15161

44. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348

45. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* (2015) 348:69–74. doi: 10.1126/science.aaa4971

46. Gubin MM, Schreiber RD. CANCER. the odds of immunotherapy success. *Science* (2015) 350:158–9. doi: 10.1126/science.aad4140

47. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* (2015) 350:207–11. doi: 10.1126/science.aad0095

48. Ma S, Chen C, Ji X, Liu J, Zhou Q, Wang G, et al. The interplay between m6A RNA methylation and noncoding RNA in cancer. *J Hematol Oncol* (2019) 12:121. doi: 10.1186/s13045-019-0805-7

49. Dang Q, Shao B, Zhou Q, Chen C, Guo Y, Wang G, et al. RNA -methyladenosine in cancer metastasis: Roles, mechanisms, and applications. *Front Oncol* (2021) 11:681781. doi: 10.3389/fonc.2021.681781

50. Nam G-H, Lee EJ, Kim YK, Hong Y, Choi Y, Ryu M-J, et al. Combined rho-kinase inhibition and immunogenic cell death triggers and propagates immunity against cancer. *Nat Commun* (2018) 9:2165. doi: 10.1038/s41467-018-04607-9

51. Yang C, Huang X, Li Y, Chen J, Lv Y, Dai S. Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: An *in silico* strategy towards precision oncology. *Brief Bioinform* (2021) 22:bbaa164. doi: 10.1093/bib/bbaa164

52. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* (2017) 171:1437–52. doi: 10.1016/j.cell.2017.10.049

53. Lino MM, Merlo A, Boulay JL. Notch signaling in glioblastoma: a developmental drug target? *BMC Med* (2010) 8:72. doi: 10.1186/1741-7015-8-72

54. Takebe N, Harris PJ, Warren RQ, Ivy SP. Targeting cancer stem cells by inhibiting wnt, notch, and hedgehog pathways. *Nat Rev Clin Oncol* (2011) 8:97–106. doi: 10.1038/nrclinonc.2010.196

55. Huang T, Zhou Y, Cheng AS, Yu J, To KF, Kang W. NOTCH receptors in gastric and other gastrointestinal cancers: oncogenes or tumor suppressors? *Mol Canc* (2016) 15:80. doi: 10.1186/s12943-016-0566-7

56. Chiang MY, Wang Q, Gormley AC, Stein SJ, Xu L, Shestova O, et al. High selective pressure for Notch1 mutations that induce myc in T-cell acute lymphoblastic leukemia. *Blood* (2016) 128:2229–40. doi: 10.1182/blood-2016-01-692855

57. Costea T, Hudita A, Ciolac OA, Galateanu B, Ginghina O, Costache M, et al. Chemoprevention of colorectal cancer by dietary compounds. *Int J Mol Sci* (2018) 19:3787. doi: 10.3390/ijms19123787

58. Wade JH, Alsop AT, Vertin NR, Yang H, Johnson MD, Bailey RC. Rapid, multiplexed phosphoprotein profiling using silicon photonic sensor arrays. *ACS Cent Sci* (2015) 1:374–82. doi: 10.1021/acscentsci.5b00250

59. Lin A, Qi C, Wei T, Li M, Cheng Q, Liu Z, et al. CAMOIP: a web server for comprehensive analysis on multi-omics of immunotherapy in pan-cancer. *Brief Bioinform* (2022) 23. doi: 10.1093/bib/bbac129

60. Li X, Dai Z, Wu X, Zhang N, Zhang H, Wang Z, et al. The comprehensive analysis identified an autophagy signature for the prognosis and the immunotherapy efficiency prediction in lung adenocarcinoma. *Front Immunol* (2022) 13:749241. doi: 10.3389/fimmu.2022.749241

61. Zhang J, Wang Z, Zhang X, Dai Z, Zhi-Peng W, Yu J, et al. Large-Scale single-cell and bulk sequencing analyses reveal the prognostic value and immune aspects of CD147 in pan-cancer. *Front Immunol* (2022) 13:810471. doi: 10.3389/fimmu.2022.810471

62. Georgantzoglou N, Kokkali S, Tsourouflis G, Theocharis S. Tumor microenvironment in adrenocortical carcinoma: Barrier to immunotherapy success? *Cancers (Basel)* (2021) 13:1798. doi: 10.3390/cancers13081798

63. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: From immunosurveillance to tumor escape. *Nat Immunol* (2002) 3:991–8. doi: 10.1038/ni1102-991

64. Hu-Lieskovan S, Robert L, Homet Moreno B, Ribas A. Combining targeted therapy with immunotherapy in BRAF-mutant melanoma: Promise and challenges. *J Clin Oncol* (2014) 32:2248–54. doi: 10.1200/JCO.2013.52.1377

65. Johnson FM, Bekele BN, Feng L, Wistuba I, Tang XM, Tran HT, et al. Phase II study of dasatinib in patients with advanced non-small-cell lung cancer. *J Clin Oncol* (2010) 28:4609–15. doi: 10.1200/JCO.2010.30.5474

66. Bartscht T, Rosien B, Rades D, Kaufmann R, Biersack H, Lehnert H, et al. Dasatinib blocks transcriptional and promigratory responses to transforming growth factor-beta in pancreatic adenocarcinoma cells through inhibition of smad signalling: implications for *in vivo* mode of action. *Mol Canc* (2015) 14:199. doi: 10.1186/s12943-015-0468-0

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 28 July 2022
DOI 10.3389/fimmu.2022.961457

Check for updates

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Guan Xu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Bo Lu,
Seventh Affiliated Hospital, Sun Yat-sen University, China

*CORRESPONDENCE
Zu-hua Gao
zuhua.gao@ubc.ca
Bo Gao
bo.gao@bjmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 04 June 2022
ACCEPTED 24 June 2022
PUBLISHED 28 July 2022

CITATION
Yang H, Messina-Pacheco J, Corredor ALG, Gregorieff A, Liu J-L, Nehme A, Najafabadi HS, Riazalhosseini Y, Gao B and Gao Z-H (2022) An integrated model of acinar to ductal metaplasia-related N7-methyladenosine regulators predicts prognosis and immunotherapy in pancreatic carcinoma based on digital spatial profiling.
Front. Immunol. 13:961457.
doi: 10.3389/fimmu.2022.961457

COPYRIGHT
© 2022 Yang, Messina-Pacheco, Corredor, Gregorieff, Liu, Nehme, Najafabadi, Riazalhosseini, Gao and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author (s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An integrated model of acinar to ductal metaplasia-related N7-methyladenosine regulators predicts prognosis and immunotherapy in pancreatic carcinoma based on digital spatial profiling

Hao Yang[1†], Julia Messina-Pacheco[2†], Andrea Liliam Gomez Corredor[2], Alex Gregorieff[2], Jun-li Liu[3], Ali Nehme[4,5], Hamed S. Najafabadi[4,5], Yasser Riazalhosseini[4,5], Bo Gao[6*] and Zu-hua Gao[7*]

[1]Department of General Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Pathology, McGill University and the Research Institute of McGill University Health Centre, Montreal, QC, Canada, [3]MeDic Program, The Research Institute of McGill University Health Centre, & Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, QC, Canada, [4]Department of Human Genetics, McGill University, Montreal, QC, Canada, [5]McGill University Genome Centre, Montreal, QC, Canada, [6]Department of General Surgery, Peking University People's Hospital, Beijing, China, [7]Department of Pathology and Laboratory Medicine, British Columbia (BC) Cancer Research Center, University of British Columbia, Vancouver, BC, Canada

Acinar-to-ductal metaplasia (ADM) is a recently recognized, yet less well-studied, precursor lesion of pancreatic ductal adenocarcinoma (PDAC) developed in the setting of chronic pancreatitis. Through digital spatial mRNA profiling, we compared ADM and adjacent PDAC tissues from patient samples to unveil the bridging genes during the malignant transformation of pancreatitis. By comparing the bridging genes with the 7-methylguanosine (m7G)-seq dataset, we screened 19 m7G methylation genes for a subsequent large sample analysis. We constructed the "m7G score" model based on the RNA-seq data for pancreatic cancer in The Cancer Genome Atlas (TCGA) database and The Gene Expression Omnibus (GEO) database. Tumors with a high m7G score were characterized by increased immune cell infiltration, increased genomic instability, higher response rate to combined immune checkpoint inhibitors (ICIs), and overall poor survival. These findings indicate that the m7G score is associated with tumor invasiveness, immune cell infiltration, ICI treatment response, and overall patients' survival. We also identified FN1 and ITGB1 as core genes in the m7Gscore model, which affect immune cell infiltration and

10.3389/fimmu.2022.961457

genomic instability not only in pancreatic cancer but also in pan-cancer. FN1 and ITGB1 can inhibit immune T cell activition by upregulation of macrophages and neutrophils, thereby leading to immune escape of pancreatic cancer cells and reducing the response rate of ICI treatment.

KEYWORDS

acinar to ductal metaplasia, N7-methyladenosine, pancreatic carcinoma, prognosis, immunotherapy, digital spatial profiling

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most important histological subtype of pancreatic cancer, accounting for approximately 90% of all pancreatic cancers. The 5-year survival rate of PDAC is less than 5%, and the median survival time after diagnosis is less than 6 months (1). The poor prognosis of PDAC has been attributed to multiple factors including late diagnosis, the lack of sensitive and specific biomarkers to detect PDAC, and the lack of effective measures to prevent its development and interrupt its progression (2). Studies have shown that in pancreatitis, pancreatic acinar cells lose their morphology and characteristics, undergo cell transdifferentiation and acquire ductal morphology and characteristics. This process is called acinar to ductal metaplasia (ADM) (3). ADM developed in the setting of acute pancreatitis is usually transient and reversible. However, persistent ADM in the setting of chronic or recurrent pancreatitis may progress to pancreatic intraepithelial neoplasias (PanIN) and eventually to invasive tumor (4, 5) (6, 7). Studies have found that when ADM occurs, a variety of signaling pathways in acinar cells are activated (Notch, Wnt, PI3K/AKT, etc.), which inhibits the transcription of specific genes in acinar cells (e.g. *Mist1, Cpa1, Amy2a*, etc), while duct cell genes (e,g, *Krt19, Sox9*, etc.) are upregulated (8, 9). Previous reports have demonstrated that suppression of transdifferentiation signals in these cells blocks subsequent PanIN and PDAC (10). Therefore, elucidating the key bridging molecules in the malignant process of ADM-related PDAC can not only help us find a novel mechanisms of PDAC pathogenesis, but also provide us with new therapeutic and preventive strategies against PDAC.

Epitranscriptomics provides insights into the biological and pathological roles of different RNA modifications. An emerging type of RNA methylation, 7-methylguanosine (m7G) modification, has been a research hotspot over the past two years. Studies have shown that m7G modification is one of the most common forms of base modification in post-transcriptional regulation (11), and is widely distributed in the 5' cap region of tRNA, rRNA, and eukaryotic mRNA (12). m7G methylation was found to play an important role in the development of a variety of cancers, including colon and lung cancer (13, 14). m7G-related epigenetic regulation can also affect the tumor immune microenvironment and the efficacy of immunotherapy (15). The m7g modification process is regulated by a collection of key genes including *mettl1, mettl3*, *Cdk1*, etc. (14, 16) However, the role of m7g modification and its underlying regulatory genes in the malignant progression of PDAC is still unclear.

In this study, we compared ADM and adjacent PDAC tissues from pancreatic cancer patients and identified high and low-expressed bridging genes during the malignant transformation of pancreatitis through digital spatial mRNA profiling (DSP) (17). There was a high degree of overlap between these bridging genes and the m7G methylation genes. After comparing the bridging genes with the m7G-seq dataset, we selected 21 m7G methylation genes for subsequent bioinformatics analysis. Based on these 21 m7G methylated genes, we constructed a model, the m7Gscore, and used it to classify potential molecules that are associated with different patterns of immune infiltration and genomic instability in PDAC. We also evaluated whether m7G score and m7G target genes could be used to predict patients' response to immune checkpoint inhibitors (ICIs). As the core genes of m7G score model, FN1 and ITGB1 are highly expressed not only in the stroma and epithelial cells of ADM and PDAC, but also in pan-cancer. FN1 and ITGB1 also affect Overall survival rate, immune cell infiltration, tumor mutation burden and microsatellite instability in pan-cancer. Finally, we concluded that FN1 and ITGB1 can also up-regulate macrophages and neutrophils and inhibit immune T cell activition in pancreatic cancer, leading to immune escape and reducing the response rate of ICIs treatment.

Materials and methods

Sample collection

The experimental design and analysis are shown in the flow chart (Figure 1A). With the approval of the institutional ethics review board of the McGill University Health Center, a total of 8

sets of PDAC tissue samples were obtained from 8 patients with a history of chronic pancreatitis who underwent surgical resection in McGill University Health Centre. In each case, formalin fixed paraffin embedded (FFPE) tissue blocks that contain normal acini, ADM tissue and PDAC on the same tissue section were selected. The clinical features of the eight patients are shown in Table 1. None of the patients with pancreatic cancer received any pre-operative treatments, including radiotherapy, chemotherapy, or biological treatments. All specimens were histopathologically diagnosed by two pathologists according to the WHO diagnostic criteria for PDAC (18).

NanoString Technologies' digital spatial mRNA profiling

We selected several regions of interest (ROIs) from each PDAC sample, including normal, ADM and PDAC ROIs. NanoString Technologies' newly developed GeoMx™ digital spatial profiling (DSP) technology allows for morphology-driven, high-plex spatial analysis of FFPE samples (17). Using the GeoMx Cancer Transcriptome Atlas, a panel of RNA probes designed for comprehensive profiling of the tumor, tumor microenvironment, and tumor immune status with 1833 RNA targets, we directly analyzed the *in situ* RNA expression of a total of 48 ROIs from 8 PDAC samples. Briefly, RNA probes coupled to unique photocleavable oligonucleotide tags are hybridized to slide-mounted FFPE tissue sections. Slides are then stained and visualized, and the oligonucleotides are then released from ROIs *via* UV exposure. The oligonucleotides are collected separately and quantified. Counts are then mapped back to each tissue location, yielding high quality, spatially resolved differential gene expression profiles. The flowchart of DSP technology was shown in Figure 1B.

Histology, immunohistochemistry and immunofluorescence

Formalin fixed tissue was processed, embedded in paraffin, and cut into 5 μm sections. Hematoxylin and Eosin (H&E) (Thermo Fisher Scientific, 7221, 7111) staining was performed according to the clinical laboratory standard. Two areas of normal acini, ADM and PDAC from each case were selected for the construction of a tissue microarray. Immunohistochemical (IHC) staining was performed using antibodies against Fibronectin 1 (FN1, 1:1000, Cell Signaling Technologies 26836) and integrin β 1(ITGB1, 1:1000, Cell Signaling Technologies 34971). Tissue sections were deparaffinized in xylene and rehydrated in graded ethanol. Antigen retrieval was performed by heating sections in boiling sodium citrate buffer (Sigma-Aldrich, C-9999) for 20 minutes. After blocking with 3% hydrogen peroxide and bovine serum albumin (BSA), the tissues were incubated with the primary antibody at 4°C overnight. After washing, the tissues were incubated with corresponding horseradish peroxidase (HRP)-conjugated secondary antibodies. The color was developed using diaminobenzidine (DAB) substrate (Sigma-Aldrich, D-7304) and slides were counterstained with hematoxylin. Images of three random areas from each section were captured at 400x, 500x, and 600x magnification for evaluation. Immunofluorescence staining was performed using primary antibodies against cytokeratin-19 (CK19; 1:500, DSHB, TROMAII), alpha-smooth muscle actin (SMA, 1:2000, Sigma-Aldrich A2547), FN1 (1:200), and ITGB1 (1:200). Corresponding Alexa Fluor dyes were used for fluorescent detection. DAPI was used for nuclear counter staining. Images were captured on the Zeiss LSM780 laser scanning confocal microscope.

Data retrieval and processing

We obtained the m1A dataset (1655 regulator genes) and m5C dataset (34 regulator genes) through the RMBase database (https://rna.sysu.edu.cn/rmbase June 2021) (19, 20). We acquired the m6A dataset (417 regulator genes) through the M6A2Target database (https://m6a2target.canceromics.org June 2021) (21). We obtained the m7G-seq dataset (2795 regulator genes) through the m7GHub database (https://www.xjtlu.edu.cn/biologicalsciences/m7ghub June 2021) (22). We intersected each group of methylated genes with the bridging genes to identify the proportion of each methylated gene in the bridging gene set. The bridging genes and the m7G-seq dataset were intersected to obtain 54 m7G methylation genes. Through the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/: accessed June 2021), we obtained the raw mRNA matrix data of PDAC in fragments per kilobase million (FPKM) format and the copy number data for pancreatic cancer. The raw data of the mRNA matrix were processed to remove duplicate samples. We also obtained the clinical data of pancreatic cancer patients through the TCGA database. To reduce statistical error, patients with survival times less than or equal to 90 days were excluded from the data. We downloaded the GSE21501 dataset from the Gene Expression Omnibus (GEO) database to obtain the mRNA matrix and clinical data of pancreatic cancer. The FPKM matrix of pancreatic cancer was converted to the TPM format and then merged with the GEO matrix, and some missing genes were removed through batch correction to expand the sample size for subsequent analysis. Similarly, patients with survival times less than or equal to 90 days were excluded from the GEO database. We download pan-cancer raw mRNA matrix data, clinical data and copy number data through UCSC database (Xena.ucsc.edu/December 2021). The clinicopathological characteristics of the pancreatic cancer patients in the TCGA database and the GEO database are shown in Table 2.

Cluster analysis

To investigate whether m7G methylation gene expression is associated with pancreatic cancer, we used the "Consensus

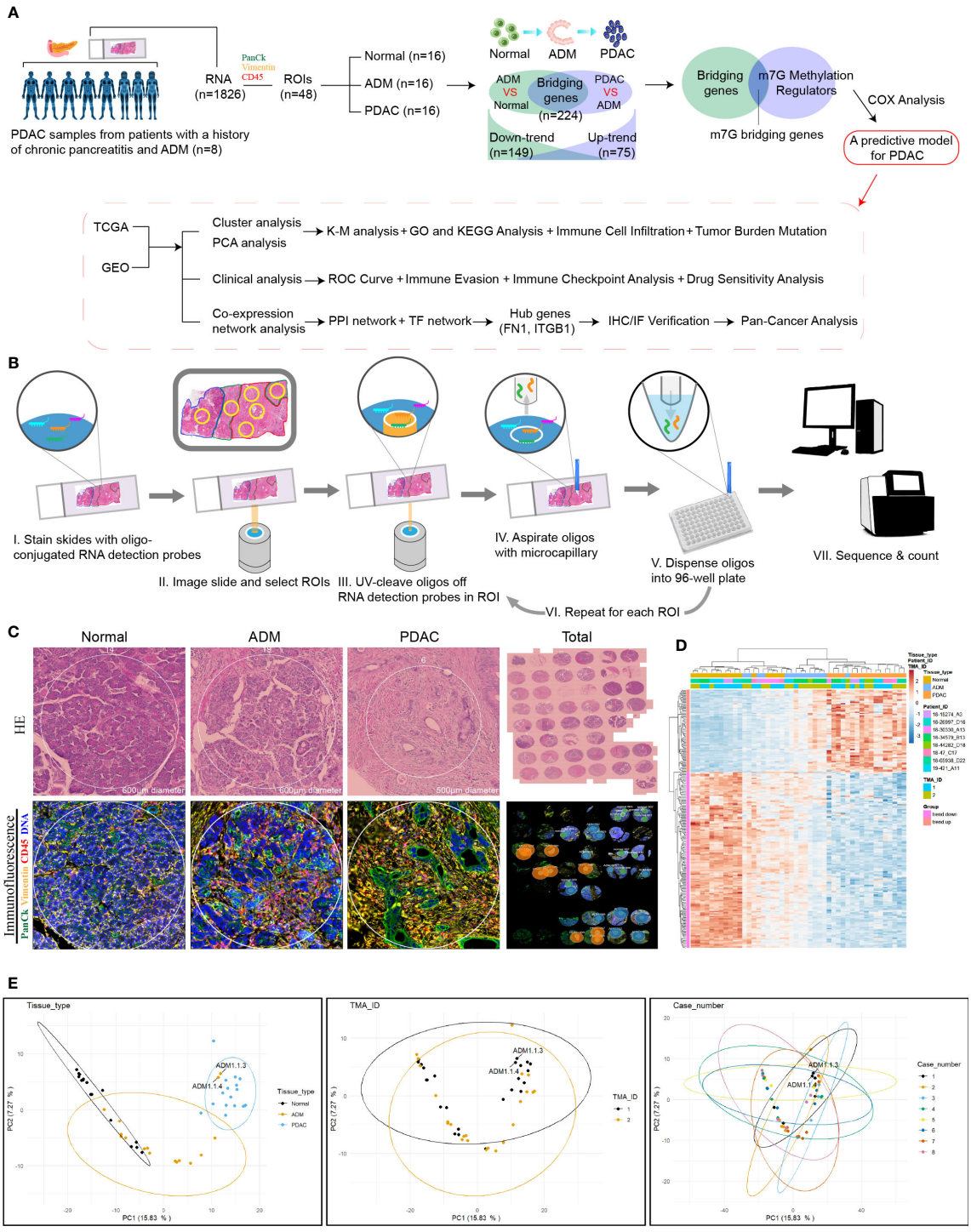


FIGURE 1
Digital spatial profiling of pancreatic cancer FFPE samples from patients with a history of chronic pancreatitis and ADM. **(A)** Flowchart of the study. **(B)** Schematic overview of the DSP workflow. **(C)** Representative HE staining and IHC images from each group. **(D)** Heatmap of the 1826 detected genes. Cluster analysis indicated the clusters marked in different colors. **(E)** PCA analysis of tissues, slides and patients.

TABLE 1 Patient characteristics.

CASE No.	Age	Sex	Grade	Stage
CASE 1	67	M	G2	pT3N1
CASE 2	80	M	G2	pT3N1
CASE 3	86	M	G2	pT2N0
CASE 4	66	F	G2	pT2N0
CASE 5	83	M	G2	pT2N1
CASE 6	67	F	G1	pT1cN0
CASE 7	83	F	G1-G3	pT3N1
CASE 8	66	M	G3	pT2N

ClusterPlus" package in R to classify pancreatic cancer data. When the clustering index k is between 2 and 9, k=3 is determined as the optimal number of subtypes. When k=3, the intergroup correlation is weak, while the intragroup correlation is strong. We calculated the survival curves of different clusters of pancreatic cancer using the Kaplan–Meier method and plotted them using the "survminer" package. The relationship between the expression of the m7G gene according to the pancreatic cancer classification and the clinical data of patients was shown with a heat map. The DEGs in the three clusters of PDAC were identified by using the VennDiagram package R language. Then intersected the DEGs, we obtained 907 intergenes. Using the single-sample gene set enrichment analysis (ssGSEA) algorithm, we obtained the scores of immune cells in different clusters of pancreatic cancers, and the scores were plotted as box plots using the "GSEABase" and "GSVA" packages in R. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway files were downloaded from the GSEA website (https://www.gsea-msigdb.org June 2021). Then, the enriched functional pathways in the pancreatic cancer classifications were plotted into a heat map using the "GSEABase" and "GSVA" packages in R.

Principal component analysis

We obtained the principal component analysis (PCA) scores for intergenes in the three m7G clusters of PDAC. PCA maps of different clusters of pancreatic cancer were then plotted using the "limma" and "ggplot2" packages in R. The m7G score of each sample was obtained *via* a PCA analysis of DEGs in the three m7G clusters (23). Based on the m7G score, we divided all the pancreatic cancer samples into high and low m7G score groups. A Sankey diagram was used to show the relationship among the three clusters of pancreatic cancer, the two m7G score types of pancreatic cancer and the overall survival rate. We analyzed the correlation between m7G scores and immune cells by using the ssGSEA algorithm. We analyzed the relationship between m7G clusters, m7G score, and tumor mutation load using the "ggpubr" and "reshape2" packages in R. Using the "limma" package, the expression levels of UBQLN4 in the high and low m7G score groups were displayed using box plots. We also used the "survival" and "survminer" packages to analyze the combined survival rate of the high tumor mutation burden

TABLE 2 Baseline characteristics of patients from TCGA and GEO database.

Clinical features	Total patients (317)		TCGA (185)		GSE21501 (132)	
	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)
Fustat						
Alive	121	38.2	85	45.9	36	27.3
Dead	166	52.3	100	54.1	66	50.0
Unknown	30	9.5	0	0	30	22.7
Stage T						
1-2	49	15.5	31	16.7	18	13.6
3-4	232	73.2	152	82.2	80	60.6
Unknown	36	11.3	2	1.1	34	25.8
Stage N						
N0	78	24.6	50	27.0	28	21.2
N1	203	64.1	130	70.3	73	55.3
Unknown	36	11.3	5	2.7	31	23.5

group, the low tumor mutation burden group, the high m7G score group, and the low m7G score group. Through the "maftools" package, we calculated the gene frequencies of the high and low m7G score groups, and selected the top 20 genes with the highest mutation frequency to draw a waterfall chart. The "plyr" and "ggpubr" packages were used to plot different clinical traits in the high and low m7G score groups as histograms and box plots. TCGA database data is used to draw NOMO diagram and ROC curve by using "rms" package, "regplot" package and "timeROC" R packages. The time gradient of ROC was 1,2 and 3 year. TIDE signature was an algorithm for calculating T cell dysfunction and rejection in various tumors. TIDE score was not only consistent with immune escape characteristics, but also can predict the effect of immune checkpoint treatment in patients with tumor (24). We downloaded pancreatic cancer-related TIDE score, Exclusion score and Dysfunction score data from TIDE database (http://tide.dfci.arvard.edu/ June 2021). We then analyzed the difference of the scores between the high- and low- m7G score groups.

The immune checkpoint treatment scoring data for pancreatic cancer were downloaded from The Cancer Immunome Database (TCIA) (https://tcia.at/ June 2021). We then analyzed the treatment of ctla4 and pd1 immune checkpoints in pancreatic cancer according to m7G scores.

Protein-protein interaction (PPI) networks and transcription factor regulatory networks

Through the STRING website (https://string-db.org/cgi/input.pl June 2021), the protein-protein interaction networks of the m7G methylation genes were constructed. The TSV format files of the m7G PPI networks were also downloaded. Based on the TSV files, we plotted histograms to visualize the core genes of the m7G PPI networks. From the DAVID website (https://david.ncifcrf.gov June 2021), we obtained the transcription factors associated with m7G methylation. We constructed a transcription factor regulatory networks map of the m7G methylation genes using Cytoscape software.

Statistical analysis

The copy number variation frequency of the m7G methylation gene was obtained by calculating the increases and reductions in the number of gene copies in the TCGA samples. The "RCircos" package in R was used to plot the circle diagram of gene copy number. Cox analysis and coexpression analysis were used to map the prognostic gene network.

Results

Bridging the gap between ADM and PDAC: Bridging genes identified by DSP

The experimental design and analysis are shown in Figure 1A. In order to identify the genes that bridge the gap between ADM and PDAC, we collected 8 samples from PDAC patients and selected 6 regions of interest (ROIs) from each sample, including 2 normal, 2 ADM and 2 PDAC ROIs. The results of hematoxylin and Eosin(H&E) staining and immunohistochemistry (IHC) showed that the selection of the ROIs of Normal, ADM and PDAC were correct. Microscopic examination of the PDAC samples showed histological evidence of conversion from normal pancreatic tissue to ADM and PDAC (Figure 1C). GeoMx™ digital spatial profiling (DSP) analysis of paired ADM and PDAC tissues (ADM vs Normal ∩ PDAC vs ADM) identified a total of 224 trend genes, among which 75 genes showed gradually increasing expression, and 149 genes showed gradually decreased expression. The gene expression heatmap is shown in Figure 1D. ADM samples 1.1.3 and 1.1.4 appeared in the cluster of PDAC samples, so they were removed from the study queue (Figure 1E). There was no confounding effect between the remaining patients and samples.

High expression of m7G methylation genes in pancreatic cancer

The proportions of m1A, m5C, m6A and m7G related regulators in bridging genes were 14.28%, 0.04%, 5.80% and 24.10%, respectively. Since m7G-related regulators account for a high proportion of bridging genes, we conducted more in-depth study on this type of methylation in PDAC (Figure 2A). The intersection of the 224 bridging genes and the 2795 m7G methylation genes yielded 54 bridging m7G methylation genes expressed in pancreatic cancer (Figure 2B). The 54 bridging m7G methylation genes in the TCGA dataset were subjected to univariate Cox analysis, and 21 prognostic-related bridging m7G genes were obtained (Figure 2C). The frequency of copy number variations in the 21 m7G methylation genes was observed with a histogram (Figure 2D). In most of the bridging m7G methylation genes, the frequency of copy number increases was higher than the frequency of deletions. Among these genes, the LY6E methylation gene had the most significant frequency of copy number increases. The MAP2K2 and OAZ1 methylation genes had the most significant frequency of copy number deletion. The gene copy number circle diagram shows that the m7G methylation genes were mainly concentrated on human chromosomes 8, 10, and 19 (Figure 2E). After merging the TCGA and GEO datasets, we obtained a total of 19 m7G methylation genes by removing some of the missing genes through batch correction. A co-expression analysis of the m7G

methylation genes was performed, and a prognostic network was plotted (Figure 2F). As illustrated in the figure, the vast majority of the m7G genes regulate one another and form a functional ensemble that jointly affects the progression of pancreatic cancer.

Three clusters of PDAC based on bridging m7G methylation genes

Through a cluster analysis of the bridging m7G methylation genes, we divided all the samples into three clusters (Figures 2G–I). The survival analysis of the three clusters of pancreatic cancer showed that cluster 2 had the lowest overall survival rate, while cluster 3 had the highest survival rate (Figure 2J). This indicates that the degree of malignancy of cluster 2 was relatively high, and the degree of malignancy of cluster 3 was relatively low. On the heat map (Figure 2L), the expression of most m7G methylation genes was significantly increased in cluster 2 and significantly decreased in cluster 3. Figure 4K shows that of the 23 types of immune cells, there were statistically significant differences in the expression of 20 types of immune cells among the three clusters of pancreatic cancer. These results indicate that bridging m7G methylation genes can regulate immune cell infiltration in pancreatic cancer. We performed GO and KEGG enrichment analyses between three clusters of pancreatic cancer (Figures S1A–F). Cluster 2 has the highest degree of malignancy, and its enrichment pathways are mainly concentrated in cell differentiation pathways, tumor microenvironment pathways and carcinogenic pathways, including regulation of cell morphogenesis involved in differentiation, ECM receptor interaction, pancreatic cancer, P53 signaling pathway and so on. The degree of malignancy of cluster 1 is weaker than that of cluster 2, and its pathways mainly focus on metabolism and tumor-related pathways, such as integrin mediated signaling pathway, movement in host environment and pathways in cancer. Cluster 3 has the lowest degree of malignancy. Moreover, 907 genes (intergenes) were overlapped among the three clusters of pancreatic cancer (Figure S1G). GO and KEGG analysis results showed that the intergenes were mainly enriched in ECM-receptor interaction, focal adhesion and Wnt signaling pathway (Figures S1H–K).

Risk stratification of PDAC based on PCA analysis and m7G scores

Through the PCA analysis of all the samples (Figure 3A), we found that there was basically no overlap in m7G scores among the three clusters of pancreatic cancer, and there was a good correlation within the clusters. This indicates that our m7G classification is very accurate. Figure 3B shows that the m7G score was the highest for cluster 2 and lowest for cluster 3. In addition, the m7G scores of the three clusters were significantly different. Through the PCA analysis, we obtained m7G scores and divided all the pancreatic cancer samples into high and low m7G score groups. We conducted ROC analysis on the m7Gscore model and found that its 1, 2, and 3-year AUC areas were all greater than 0.6 (Figure S1L). Sankey diagram showed that most of the cases in cluster 2 with the highest degree of malignancy belong to the high m7G score group, most of the cases in cluster 1 belong to the low m7G score group, and all the cases in cluster 3 with the lowest degree of malignancy belong to the low m7G score group (Figure 3C). Moreover, the survival rate of the high m7G score group was significantly lower than that of the low m7G score group (Figure 3D). This indicates that a high m7G score is associated with high risk, while a low m7G score often reflects a low risk. This is consistent with the results of previous studies. Cluster 2 pancreatic cancer has a high m7G score and low survival rate, while cluster 3 has a low m7G score and a high survival rate.

m7G score is associated with immune cell infiltration and tumor mutation burden

We performed a correlation analysis for m7G score and immune cell infiltration (Figure 3E). Among the 23 types of immune cells, the expression of 20 types of immune cells was statistically significantly associated with the m7G score: 18 types were positively correlated with the m7G score, and 2 types were negatively correlated with the m7G score. We also analyzed the relationship between m7G score and immune cell infiltration through multiple softwares. Multiple software results showed that m7G score was closely related to a variety of inflammatory related immune cells, including macrophage, neutrophil and cancer associated fibroblast (Figure 3F). Figure 3G shows that the expression level of UBQLN4 in the high m7G score group was significantly higher than that in the low m7G score group, indicating that the higher the m7G score was, the greater the genomic instability (25). Figure 3H shows that the m7G score was positively correlated with tumor mutation burden, with cluster 2 having the highest tumor mutation burden and cluster 3 having the lowest tumor mutation burden. The waterfall chart shows that the gene mutation frequency of the high m7G score group was significantly higher than that of the low m7G score group (Figure 3I, J). The genes with the highest mutation frequency in the high and low m7G score groups were KRAS, TP53, CDKN2A and SMAD4 (26, 27). Figure 3K shows that the survival rate of the high tumor mutation burden group was significantly higher than that of the low tumor mutation burden group. We also performed a joint analysis of the high and low tumor mutation burden groups and the high and low m7G score groups (Figure 3L), and the results showed significant differences ($p<0.001$), indicating that both the tumor mutation burden and m7G scores were correlated with patient prognosis.

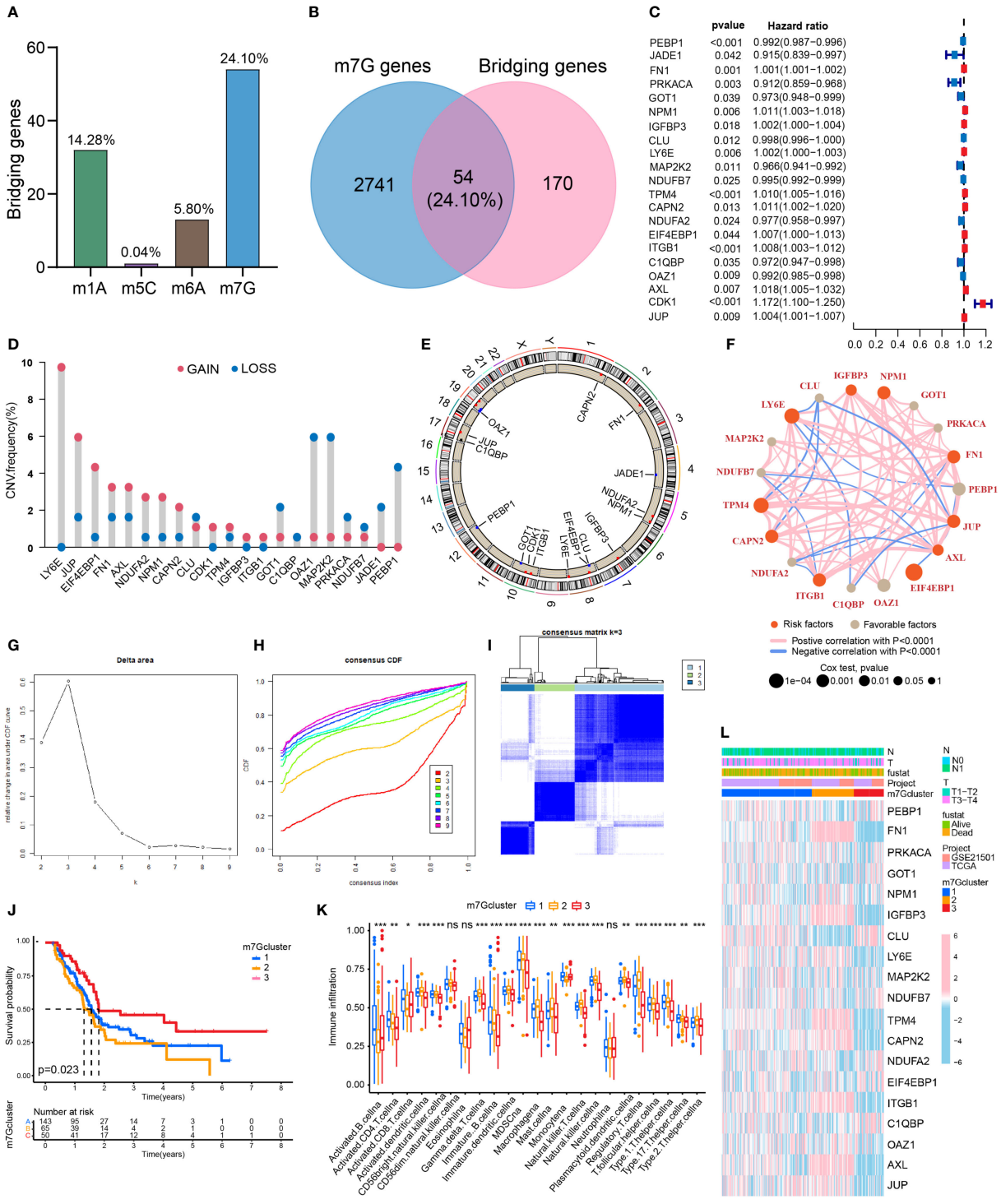


FIGURE 2
Classification of PDAC based on m7G methylation related bridging genes. **(A)** The proportion of four common methylation type genes in bridging genes. m7G methylation genes have the highest proportion. **(B)** Intersection of the bridging genes and the m7G methylation genes. **(C)** Univariate Cox analysis of the m7G methylation genes. **(D)** Diagram of the frequency of copy number variations in the m7G methylation genes. **(E)** Copy number circle diagram for the m7G methylation genes. **(F)** Prognostic network of the m7G methylation genes. **(G–I)** Classification of pancreatic cancers based on m7G methylation related bridging genes. Changes in the length and inclination of the CDF curve for k=2–9. Area under the cumulative distribution function curve for k=2–9. Division of the pancreatic cancer samples into three clusters. **(J)** Kaplan–Meier survival curves of the three clusters. **(K)** Immune cell infiltration of the three clusters. *p<0.05; **p<0.01; ***p<0.001; ns, no significance. **(L)** Heat map of pancreatic cancer classification, m7G methylation genes, and clinicopathological characteristics.

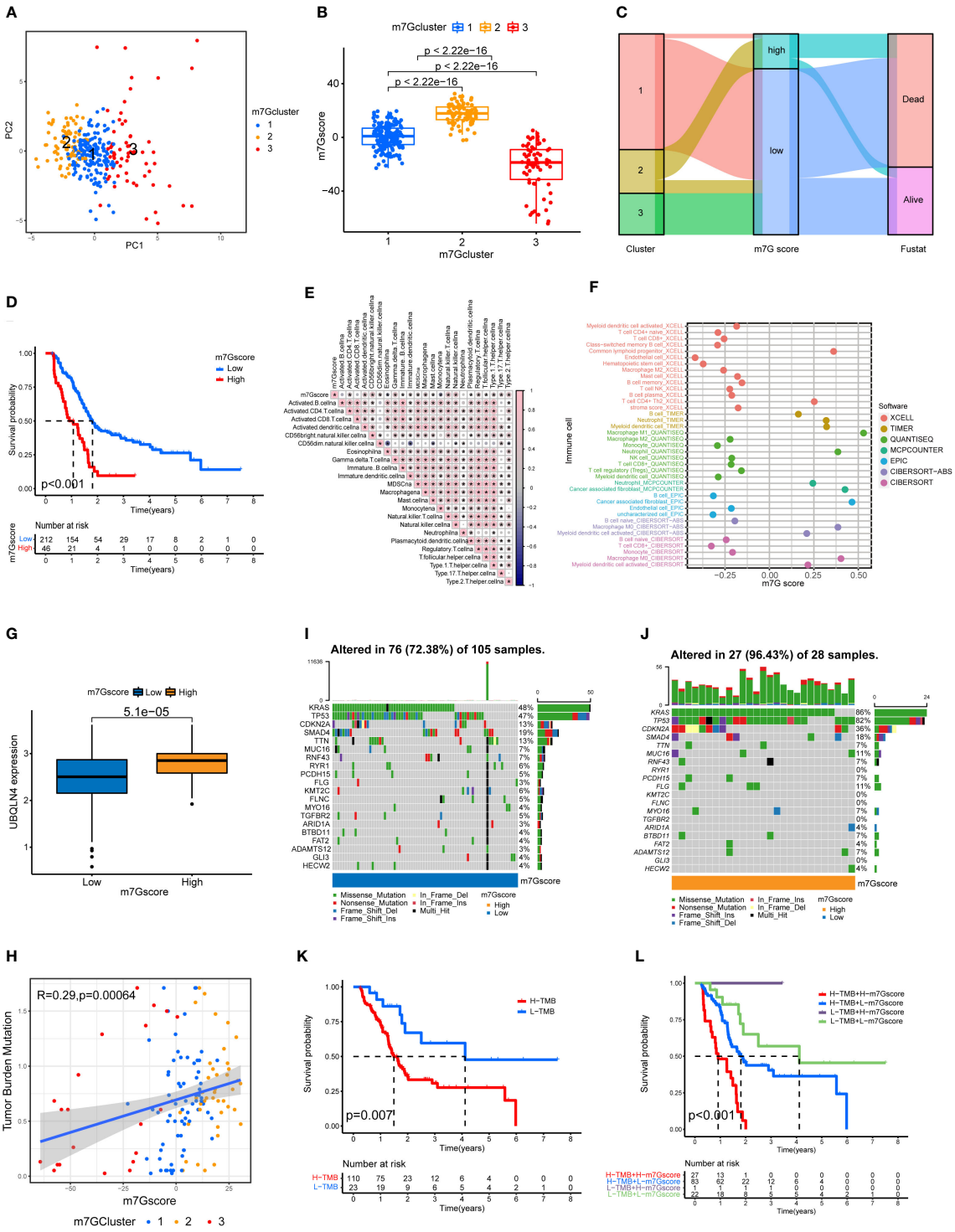


FIGURE 3
Immune cell infiltration and genomic instability in the m7G score. **(A)** PCA of the m7G methylation genes. **(B)** The m7G scores for the three clusters of pancreatic cancers. **(C)** The Sankey plots revealed the correlation results of m7G clusters, m7G scores and the future state of patients. **(D)** Kaplan–Meier survival curves of the high m7G score group and the low m7G score group. **(E)** Spearman correlation analysis of the relationship between m7G scores and immune cell types *p<0.05. **(F)** Correlation of m7G score and immune cell infiltration through multiple software. **(G)** UBQLN4 expression in the high m7G score group and the low m7G score group. **(H)** Correlation analysis of the relationship between m7G scores and tumor mutation burden among the three clusters of PDAC. **(I, J)** Waterfall plot of gene mutation frequencies for low and high m7G score groups. **(K)** Kaplan–Meier survival curves of the high tumor mutation burden group and the low tumor mutation burden group. **(L)** Combined survival analysis results for the high and low tumor mutation burden groups and the high and low m7G score groups.

m7G score is associated the clinical behavior of PDAC

We analyzed the patients' m7G scores and clinicopathological characteristics, including survival, T stage, and N stage. Figures 4A–C shows that a high m7G score was closely associated with a poor prognosis, local tumor invasion, and lymph node metastasis. This indicates that from the perspective of clinical pathological characteristics, a high m7G score represents a higher degree of malignancy. We scored various clinicopathological features and m7G scores and drew a NOMO map to predict the prognosis of patients. If the total score reaches 753, the probability that the patient's survival time was less than 1, 2, and 3 years is 0.0842%, 0.278%, and 0.327%, respectively (Figure 4D). Among multiple indicators, only the m7G score was statistically significant. The ROC curve results showed that the 1, 2, and 3-year AUC areas of the NOMO model were 0.713, 0.834, and 0.847, respectively (Figure 4E).

m7G score predicts tumor response to immune checkpoint inhibitor treatment

The results of TIDE signatures showed that the immune evasion mechanism of the high m7G score group was mainly immune rejection, and the immune evasion mechanism of the low m7G score group was mainly immune dysfunction. The TIDE score of the low m7G score group was significantly higher than that of the high m7G score group. (Figures 5A–C) The results of immunosuppressant monotherapy and combination therapy analysis showed that tumors with low m7G scores had a higher rate of response to single immune checkpoint treatment (ICI). Although the efficacy of PD-1 and CTLA-4 as single immune checkpoint treatment was lower in patients with high m7G scores than in those with low m7G scores, the efficacy of two-drug combination immune checkpoint therapy in patients with high m7G scores was higher than that of the low m7G score group (Figure 5D). The observation was in agreement with the results of most drug clinical trials, in which combination immune checkpoint therapy had better efficacy than monotherapy for advanced pancreatic cancer. We also conducted drug sensitivity analysis in patients with high and low m7G score groups, so as to predict potential effective drugs for pancreatic cancer patients (Figure 5E). A total of 9 drugs were screened for patients with high m7G score, and 25 drugs were selected for patients with low m7G score (Figure S2).

FN1 and ITGB1 were the core genes regulating m7G methylation

We obtained 6 m7G target genes by taking the overlapping genes between clusters intergenes, m7G genes, and bridging genes (Figure 6A). By constructing a PPI network (Figures 6B, C), we found that the m7G methylation genes were closely related to each other and could interact with one another to form a functional ensemble and jointly regulate the occurrence and development of tumors. Among these, FN1 and ITGB1 were the core genes in the PPI network and play a leading role in regulation. M7G score was positively correlated with the expression of 6 m7G target genes (Figure 6D). The expression levels of six m7G target genes in the high m7G score group were significantly higher than those in the low m7G score group (Figure 6E). M7G target genes and m7Gscore were closely correlated with multiple immune checkpoint genes (Figure 6F). Through GSVA analysis, it was found that m7G target genes and m7Gscore were closely related to multiple carcinogenic pathways (Figures 6G, H).

IHC and immunofluorescence results for FN1 and ITGB1

To validate the RNA results at protein level, we performed IHC and immunofluorescence analysis of the core genes (FN1, ITGB1) in the m7G methylation model (Figures 7A, B). Compared with the negative staining in normal pancreas acini, FN1 showed strong expression in the stroma of ADM and PDAC (Figure 7A), whereas ITGB1 showed strong expression in the epithelial cells of both ADM and PDAC (Figure 7B). The ADM epithelium was further validated by co-staining with the ductal marker CK19 and SMA highlighted activated stromal myofibroblasts and smooth muscle in vessel walls.

Analysis of FN1 in pan-cancer

Figure 8A shows the expression of FN1 in 33 cancers, among which FN1 most expressed in THCA. Figures 8B–E shows that FN1 can affect overall survival、disease free survival disease specific survival and progression free survival of patients in a variety of cancers (including PAAD). Moreover, FN1 can affect immune cell infiltration in pan-cancer (Figure 8F). FN1 is closely associated with tumor mutation burden and microsatellite instability in a variety of tumors (Figures 8G, H). We further analyzed the role of FN1 in the immune microenvironment of pancreatic cancer. Based on the ESTIMATE analysis, we found that the estimate score, stromal score and immune score of FN1 high-expression group were significantly higher than those of FN1 low-expression group (Figure 8I). FN1 is also closely associated with multiple immunotherapy pathways and classical process of the cancer immunity cycle (Figure 8J). Using CIBERSORT algorithm, we calculated the correlation between FN1 and immune cells (Figure S3A). The results showed that FN1 was positively correlated with macrophages

10.3389/fimmu.2022.961457

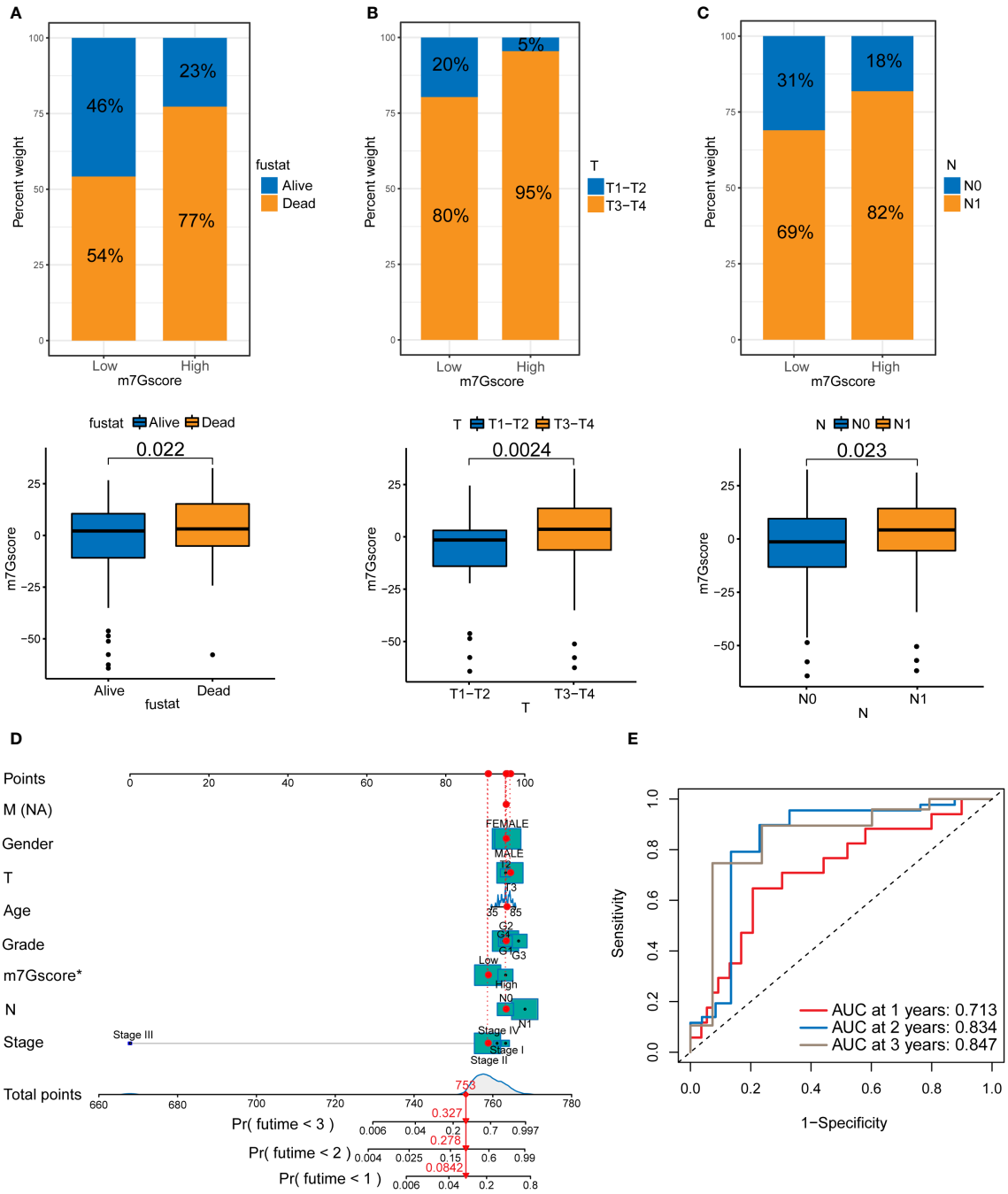


FIGURE 4
The relationship between m7G scores and prognosis in PDAC. **(A)** Survival in the high and low m7G score groups. **(B)** T staging of the high and low m7G score groups. **(C)** N staging of the high and low m7G score groups. **(D)** NOMO map associated with m7G score and clinical information. **(E)** Area under the curve (AUC) for time-dependent receiver operating characteristic curves demonstrating the prognostic performance of the NOMO model.

and neutrophils, and negatively correlated with various immune T cells (Figure 8K). Figure 8L showed that the response rate of PD1 and CTLA4 treatment in FN1 low-expression group was significantly higher than that in FN1 high-expression group, which was consistent with the results of m7G score model on the response rate of ICIs. We also calculated the response possibility of FN1 to immunotherapy by TIDE algorithm, and the results showed that the response of FN1 low-expression group was

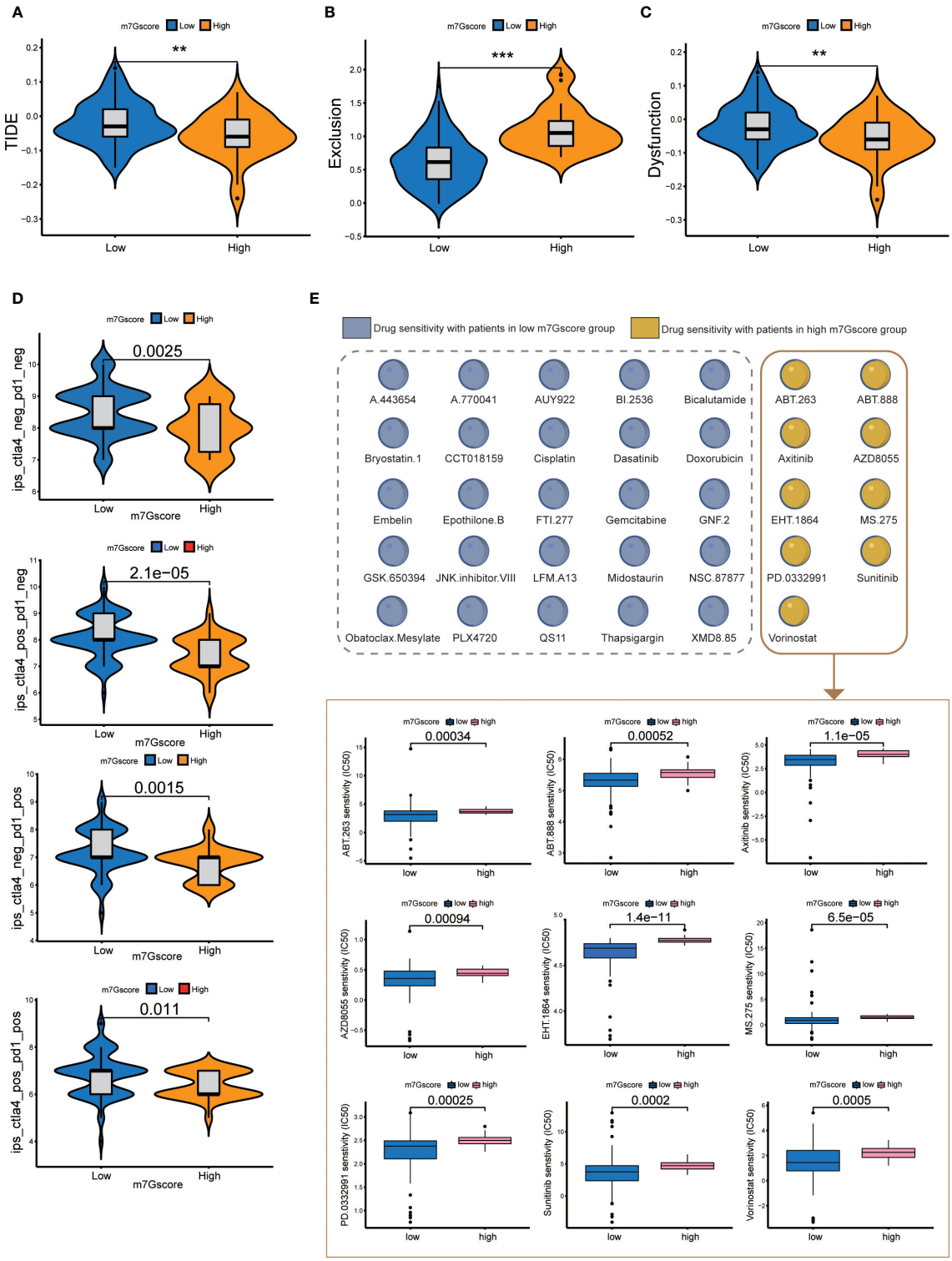


FIGURE 5
Immune checkpoint treatment and drug sensitivity. **(A–C)** TIDE score, Exclusion score and Dysfunction score of high and low m7G score groups. **p<0.01; ***p<0.001 **(D)** Immune checkpoint treatment in the high and low m7G score groups. **(E)** Drug sensitivity in the high and low m7G score groups.

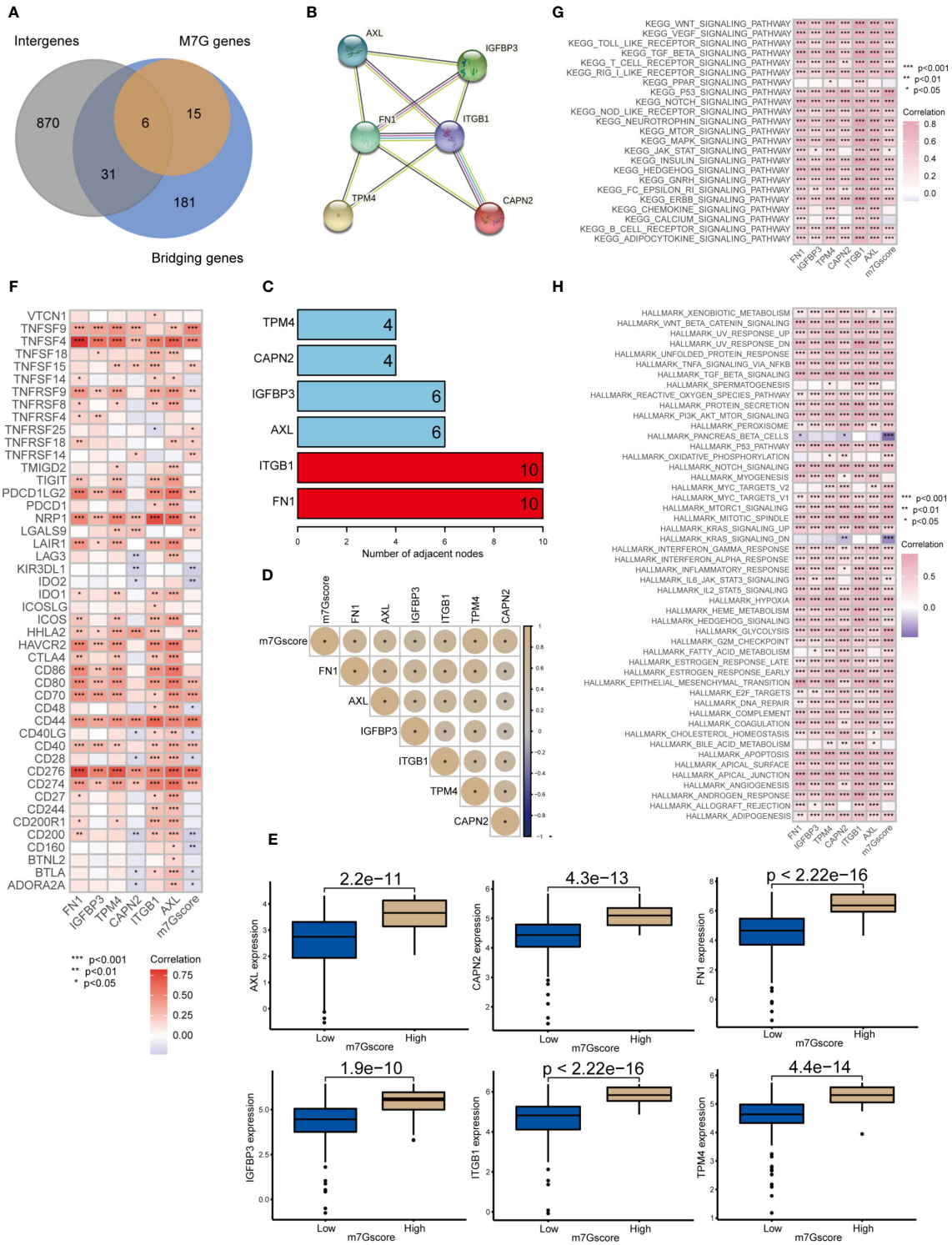


FIGURE 6
Network of m7G target genes in the m7G score model *p<0.05; **p<0.01; ***p<0.001. **(A)** Venn diagram of cluster intergenes, bridging genes and m7G target genes. **(B)** Core genes of the PPI network of m7G methylation genes. **(C)** PPI network of the m7G methylation genes. **(D)** Correlation analysis of m7Gscore and m7G target gene. **(E)** Expression of m7G target genes in high and low m7G groups. **(F)** Correlation analysis of m7G target genes and m7G score with immune checkpoint related genes. **(G, H)** GSVA analysis of m7G target genes and m7Gscore.

significantly higher than that of FN1 high-expression group (Figure 8M). Furthermore, the expression of TIDE score, exclusion score and MSI expr sig in the FN1 high-expression group was higher than that in low-expression group (Figure 8N). We analyzed the correlation of checkpoint genes between the high and low expressed FN1 groups, the result showed that multiple checkpoint genes were found to be highly expressed in the high-expression FN1 group (Figure S3B). These results were consistent with the results of m7G score model, suggesting that FN1 may cause immunotherapy resistance in pancreatic cancer patients through immune evasion.

Analysis of ITGB1 in pan-cancer

Figure 9A shows the expression of ITGB1 in 33 cancers, among which ITGB1 expressed highest in CHOL and PAAD. Figures 9B–E shows that ITGB1 can affect overall survival、disease free survival disease specific survival and progression free survival of patients in a variety of cancers (including PAAD). Moreover, ITGB1 can affect immune cell infiltration in pan-cancer (Figure 9F). ITGB1 is closely associated with tumor mutation burden and microsatellite instability in a variety of tumors (Figures 9G, H). ITGB1 can also affect the immune microenvironment in pancreatic cancer (Figure 9I). Based on the ESTIMATE analysis, we found that the estimate score, stromal score and immune score of ITGB1 high-expression group were significantly higher than those of ITGB1 low-expression group. ITGB1 is also closely associated with multiple immunotherapy pathways and classical process of the cancer immunity cycle (Figure 9J). Using CIBERSORT algorithm, we calculated the correlation between ITGB1 and immune cells (Figure S3C). The results showed that ITGB1 was positively correlated with macrophages and neutrophils, and negatively correlated with various immune T cells (Figure 9K). Figure 9L showed that the response rate of CTLA4 treatment in ITGB1 low-expression group was significantly higher than that of ITGB1 high-expression group, while there was no significant statistical difference to PD1 treatment between ITGB1 high- and low-expression group. We also calculated the response possibility of ITGB1 to immunotherapy by TIDE algorithm, and the results showed that the response of ITGB1 low-expression group was significantly higher than that of ITGB1 high-expression group (Figure 9M). In addition, the expression of TIDE score, exclusion score and MSI expr sig in the ITGB1 high-expression group was higher than that in low-expression group (Figure 9N). We analyzed the correlation of checkpoint genes between the high and low expressed ITGB1 groups, the result showed that multiple checkpoint genes were found to be highly expressed in the high-expression ITGB1 group (Figure S3D).

Discussion

Although ADM is a benign and reversible process in the setting of acute pancreatitis, long-term pancreas inflammatory stimulation can lock metaplastic cells into a duct-like state. Persistent ADM has been proven to be a precursor lesion for the development of PDAC (28, 29). Pancreatic cancer develops through a series of genetic events triggered by different driver gene mutations. These differences in the mutated driver genes lead to differences in the molecular phenotypes and biological behaviors of pancreatic cancer, ultimately resulting in different clinical outcomes (30). Studies have shown that the progression from ADM to PDAC is driven by complex malignant bridging genes and pathways (31). It is still unclear whether these bridging genes can continue to play a malignant driving role and how they function after the occurrence of PDAC. As a newly-developed spatial genomics technology, DSP can accurately detect the *in situ* expression of RNAs and proteins in both ADM and PDAC areas simultaneously. This technology avoids contamination during the process of laser microdissection and RNA preparation (32). In this study, we collected tissues from 8 PDAC patients with a history of chronic pancreatitis, and with histological manifestation of the malignant progression from ADM to PDAC on the same tissue section. Analysis using DSP technology on these human PDAC tissue samples identified 224 bridging genes in the progression from ADM to PDAC. Among these 224 genes, there was a significantly a higher degree of overlap with the m7G methylation genes. Therefore, we speculated that the development of pancreatic cancer might be closely related to the m7G methylation genes.

Cluster analysis is a powerful tool that can classify tumors into subtypes based on their genomic similarities and differences in association with patients' clinical parameters and outcome data. It can also facilitate comparative study of different subtypes and discovery of new tumor subtypes (33). Cluster 2 had the lowest overall survival rate, and Cluster3 had the highest overall survival rate. The expression of most m7G methylation genes in Cluster2 was significantly higher than that in cluster 3. Some classic oncopathways have also been significantly activated in Cluster2, such as TGF-beta, ERBB, Wnt and so on. These results suggest that m7G regulators may affect the malignant progression of pancreatic cancer through a variety of cancer pathways.

Malignant progression of tumor-related diseases is often accompanied by changes in cell morphology, such as ADM and epithelial to mesenchymal transition (34). Roland et al. found that PDAC is a process characterized by the extreme involvement of the ECM, and the changes in the ECM-receptor interaction pathway in PDAC are consistent with ECM remodeling (35). Functional analysis showed the most

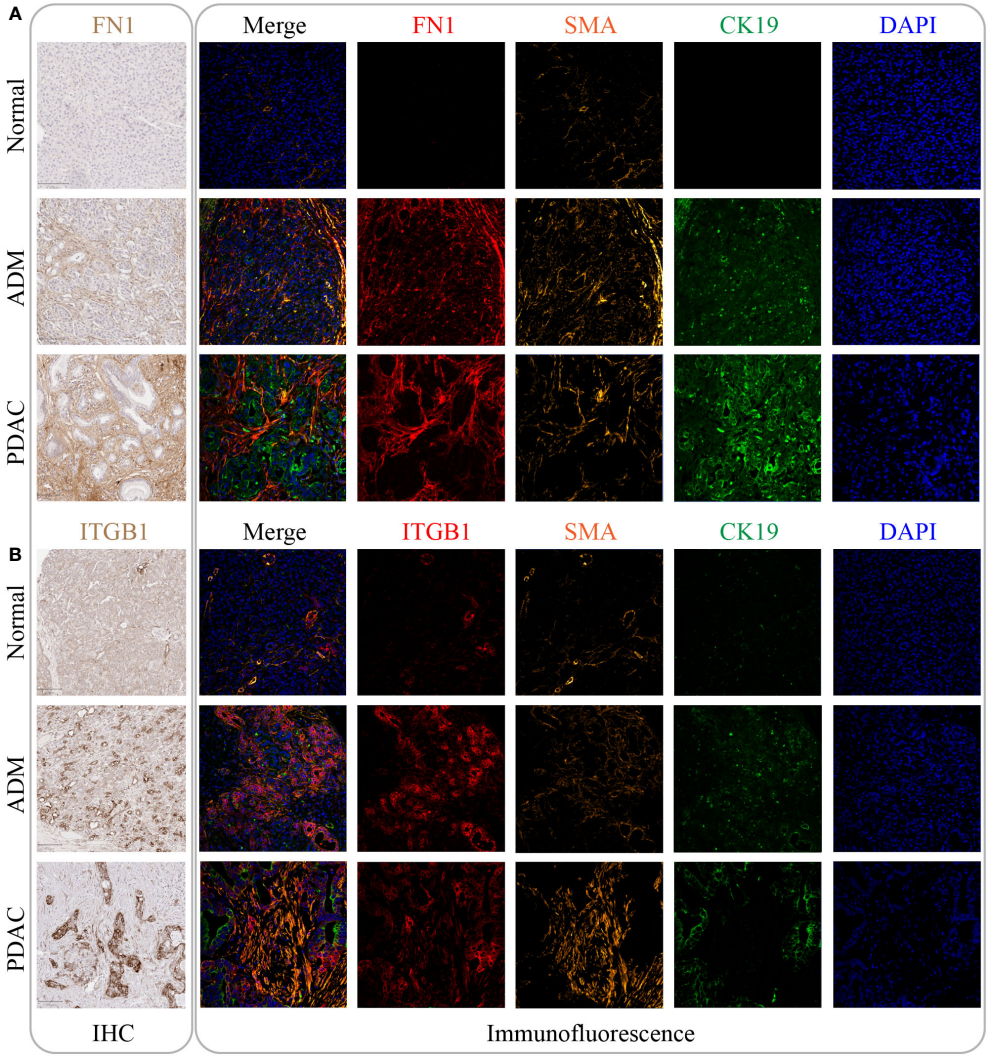


FIGURE 7
The expression of FN1 and ITGB1 in normal tissue, ADM tissue and PDAC tissue. **(A)** The expression of FN1 was gradually increased by IHC and immunofluorescence. Its expression location was mainly in the intercellular substance. **(B)** The expression of ITGB1 was gradually increased by IHC and immunofluorescence. Its expression location was mainly inside the cancer cell.

malignant Cluster2 was closely related to differentiation-related pathways, such as regulation of cell morphogenesis involved in differentiation. In this study, we found that several ECM and tumor microenvironment pathways were activated in cluster 2, such as ECM receptor interaction and cell substrate junction.

PDAC is characterized histologically by the presence of abundant desmoplastic stroma containing very small number of infiltrating lymphocytes, indicating an overall immunosuppressive microenvironment (36). It is well known that the degree of immune cell infiltration is closely related to the efficacy of immunotherapy and the prognosis of cancer patients. Studies have found that immune cell infiltration was regulated by a variety of epigenetic factors, including m6a and m5c methylation (37). But as a newer type of methylation, there are few immune-related studies on m7G methylation. The results of this study suggest that m7G methylation regulators may affect immune cell infiltration, which further affect the immunotherapy response and patient's prognosis.

In this study, the PCA algorithm showed that m7G scores were negatively correlated with the overall survival rate. The Sankey chart showed that most of the cases in Cluster2 with the worst prognosis belong to the high m7G score group. These results support the notion that m7G regulators play an important role in the malignant progression of pancreatic cancer.

Through the TIDE score, we can intuitively understand the immune escape mechanism of the high and low m7G score groups. Studies have shown that in some tumors, although the

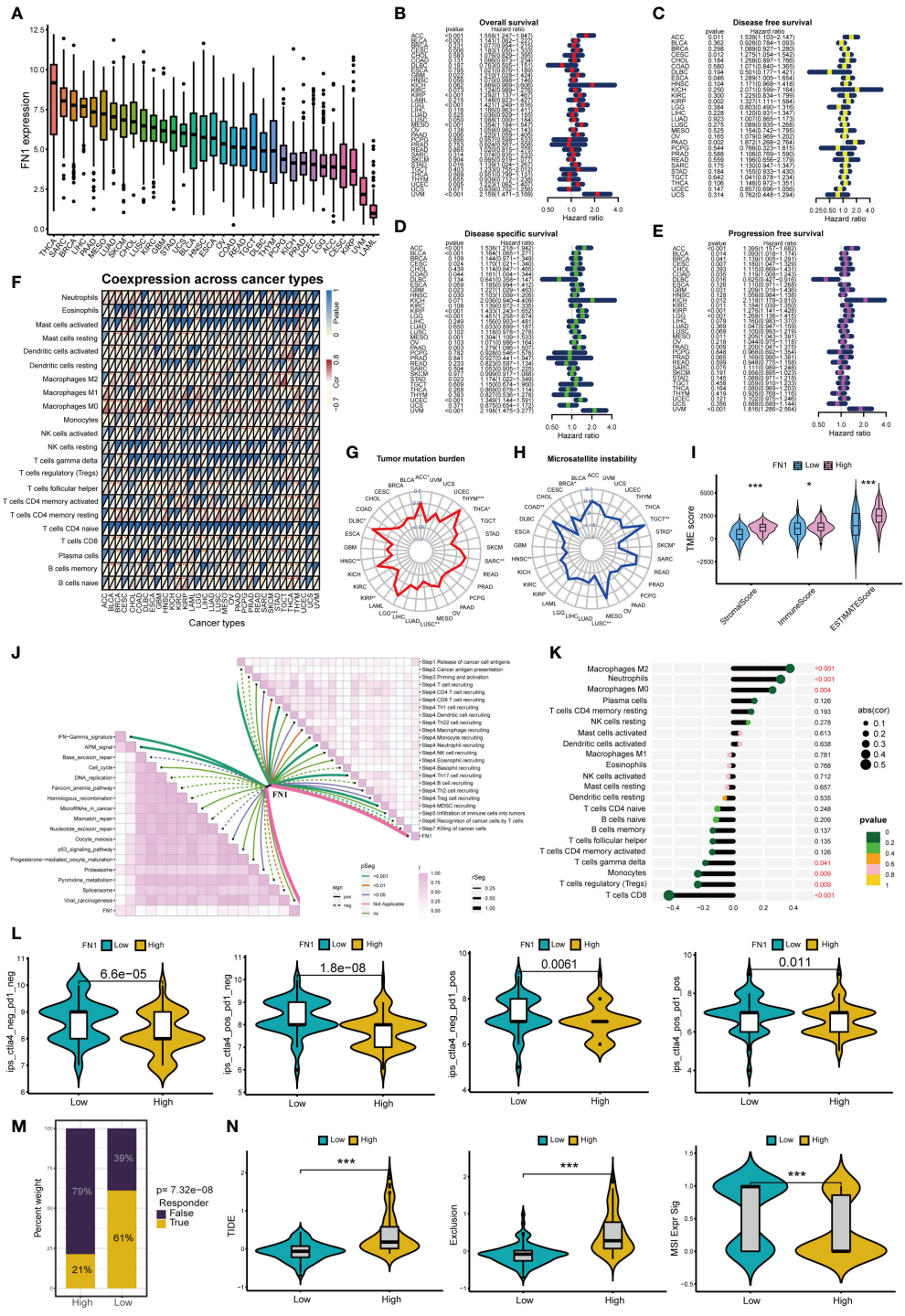


FIGURE 8
Analysis of FN1 in cancers. *p<0.05; **p<0.01; ***p<0.001. **(A)** Expression of FN1 in 33 cancers. **(B)** Overall survival of FN1 in pan-cancer. **(C)** Disease free survival of FN1 in pan-cancer. **(D)** Disease special survival of FN1 in pan-cancer. **(E)** Progression free survival of FN1 in pan-cancer. **(F)** Co-expression analysis of FN1 and immune cells in pan-cancer. **(G)** Tumor mutation burden of FN1 in pan-cancer. **(H)** Microsatellite instability of FN1 in pan-cancer. **(H) (I)** ESTIMATE analysis of FN1 high- and low-expression group. **(J)** Correlation between FN1, immunotherapy pathway and cancer immunity cycle. **(K)** Correlation analysis between FN1 and CIBERSORT immune cells. **(L)** Response rate of ICIs in FN1 high- and low-expression group. **(M)** Response of immunotherapy treatment in FN1 high- and low-expression group. **(N)** TIDE analysis of FN1 high- and low-expression group.

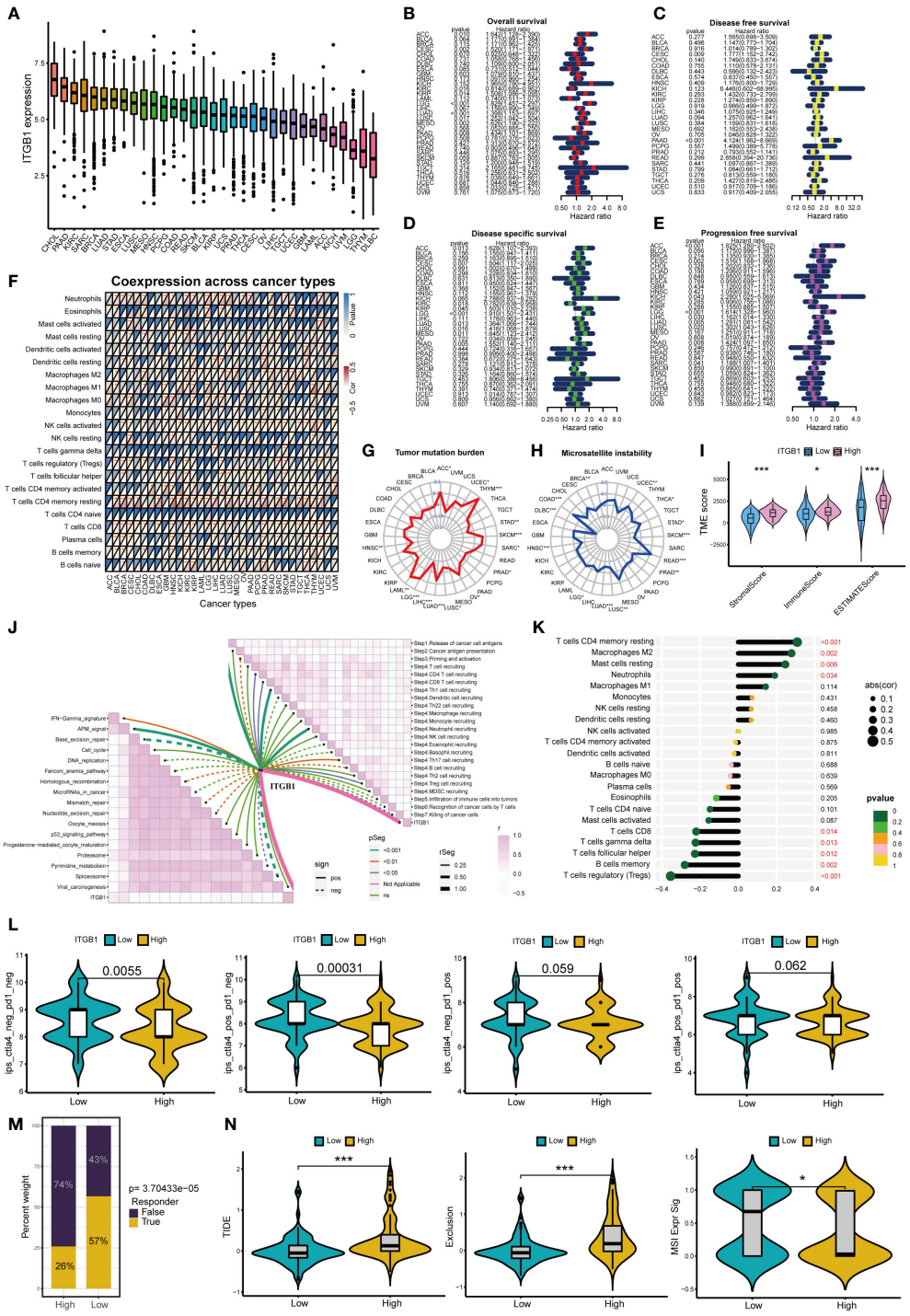


FIGURE 9
Analysis of ITGB1 in cancers *p<0.05; **p<0.01; ***p<0.001. **(A)** Expression of ITGB1 in 33 cancers. **(B)** Overall survival of ITGB1 in pan-cancer. **(C)** Disease free survival of ITGB1 in pan-cancer. **(D)** Disease special survival of ITGB1 in pan-cancer. **(E)** Progression free survival of ITGB1 in pan-cancer. **(F)** Co-expression analysis of ITGB1 and immune cells in pan-cancer. **(G)** Tumor mutation burden of ITGB1 in pan-cancer. **(H)** Microsatellite instability of ITGB1 in pan-cancer. **(I)** ESTIMATE analysis of ITGB1 high- and low-expression group. **(J)** Correlation between ITGB1, immunotherapy pathway and cancer immunity cycle. **(K)** Correlation analysis between ITGB1 and CIBERSORT immune cells. **(L)** Response rate of ICIs in ITGB1 high- and low-expression group. **(M)** Response of immunotherapy treatment in ITGB1 high- and low-expression group. **(N)** TIDE analysis of ITGB1 high- and low-expression group.

degree of cytotoxic T cell infiltration is high, these T cells are often in a state of dysfunction. In other tumors, immunosuppressive factors can eliminate T cells infiltrating the tumor tissue (38). The TIDE score results showed that the immune escape mechanism of the low m7G score group was mainly by dysfunction, while the high m7G score group was mainly by immune exclusion.

In a phase I clinical trial of 207 patients with different types of advanced cancer who received ICI monotherapy, Brahmer et al. found that drug efficacy was relatively poor in patients with advanced pancreatic cancer (39). Another randomized phase II trial of 65 patients with metastatic pancreatic cancer who failed first-line treatment with 5-FU or gemcitabine showed that the disease control rate of the combined drug treatment was significantly better than that of the monotherapy (40). We performed immune checkpoint assessments of the different m7G scores groups. The results showed that tumors with low m7G scores had a higher rate of response to ICI monotherapy. ICI monotherapy had poor efficacy in the high m7G score group, although the efficacy of combination therapy was relatively good, which is consistent with the results of multiple clinical studies. Studies have shown that the tumor microenvironment contributes to ICI resistance. A nonimmunogenic tumor microenvironment could potentially inhibit the immune response and prevent the accumulation of immune lymphocytes in tumor tissues (41), thereby affecting the efficacy of ICI treatment and leading to the development of drug resistance. On the other hand, long-term pancreatic cancer survivors have high-quality neoantigens in the tumor microenvironment. Therefore, it is conceivable that targeting these neoantigens may improve the effectiveness of ICIs in the treatment of pancreatic cancer (42).

We constructed a PPI network for the 6 m7G target genes. In the PPI network, the core genes were FN1 and ITGB1. FN1 is a glycoprotein that is mainly involved in the processes of cell adhesion and migration. Studies have reported that FN1 expression is upregulated in a variety of tumors and is negatively correlated with patients' prognosis. FN1 overexpression can be used as a molecular marker for the invasive phenotype of PDAC (43). Tsukamoto et al. found that alcohol consumption could induce pancreatitis in mice, increase FN1 expression and promote PDAC carcinogenesis (44). In a TGF-β treatment-induced PDAC model, Yuzuru et al. found that upregulation of FN1 was a hallmark of the ductal growth of PDAC (45). It has been shown that stromal cells are capable of inducing epithelial-mesenchymal transformation, an event that is closely associated with the progression of inflammation to tumors (46).Margareta et al. demonstrated that FN1 functions in epithelial misplacement (AEM) and adenomas with early carcinoma (AEC) transformation in colon cancer (47), suggesting that FN1 plays a role in the inflammatory transformation of cells to colon cancer. Proteomics study showed that abundant FN1 is present in extracellular vesicles (EVs) of PDAC and that high expression of FN1 reduces the sensitivity of PDAC to gemcitabine treatment (48). A member of the integrin family, ITGB1 was also reported to play an important role in PDAC carcinogenesis and biological behavior. ITGB1 signaling has been shown to promote the proliferation and metastatic ability of pancreatic carcinoma *in situ* in mice by stimulating the production of inflammatory cytokines (49). ITGB1 also influences the malignant progression of epithelioid-like ovarian cancer by regulating the production of the inflammatory factors IL-6, TGF-β1 and SDF-1 (50). A study by Oklahoma University suggested that ZIP4 could increase the resistance of pancreatic cancer patients to gemcitabine by upregulating the expression of ITGB1, which was associated with a poor prognosis (51). Another study from MD Anderson Cancer Center showed that GAL3 regulates the production of inflammatory cytokines in ITGB1. Inhibition of this pathway can reduce the growth and metastasis of pancreatic cancer in mice (49). These literature data and our PPI analysis results all suggest that the group of FN1 and ITGB1 genes interact with one another and exert their functions as a coordinated network. In this study, both immunohistochemical and immunofluorescence analysis showed that (1) FN1 protein was highly expressed in the stroma of ADM and PDAC lesions, and (2) ITGB1 protein was highly expressed in the epithelium of ADM and PDAC. This result suggests that increased expression of FN1 and ITGB1 is associated with the metaplastic transdifferentiation of normal pancreatic acinar cells to ductal cells through ADM and eventually, the development of PDAC.

FN1 and ITGB1 not only play a role in pancreatic cancer, but also be closely associated with overall survival, immune cell infiltration, tumor mutation burden and microsatellite instability in pan-cancer. This suggests that m7G score model and m7G target genes may be independent prognostic factors for a variety of tumors. The expression of FN1 and ITGB1 was positively correlated with macrophages and neutrophils, and negatively correlated with immune-related T cells. Studies have shown that tumor-associated macrophages (TAM) play an important role in tumor immune evasion (52). Various mediators in the tumor microenvironment mediate the recruitment of myeloid-derived suppressor cells (MDSC) and monocytes, and polarize macrophages through different signaling pathways, thereby promoting the formation of the immunosuppressive myeloid microenvironment. Meanwhile, tumor-associated neutrophils (TAN) are also an important part of the immunosuppressive myeloid microenvironment (53). Neutrophils in tumor microenvironment can inhibit the immune function of T cells, which leads to the failure of ICIs treatment (54). Both CTLA4 and PD1 ICIs treatment can

activate immune checkpoint molecules expressed on the surface of T cells, thereby reactivating T cells to play anti-tumor role (55, 56). When T cells are depleted, tumors are more likely to form immunosuppressive microenvironments that help tumor cells evade immune surveillance (57). Moreover, the higher expression levels of FN1 and ITGB1, the lower response rate of patients to ICIs treatment. TIDE score, exclusion score and MSI score in high-expression group were also significantly higher than those in low-expression group. This also indicates that the high-expression group is more prone to immune evasion. Therefore, we conclude that FN1 and ITGB1 can lead to immune evasion in pancreatic cancer and reduce the response rate of ICIs by up-regulating the activity of macrophages and neutrophils, and down-regulating expression of immune T cells.

In summary, we used spatial genomics technology DSP to identify the bridging genes in the transition from normal parenchyma to ADM to PDAC. We found that these bridging genes highly overlapped with m7G methylation genes. The integrated model of ADM-Related m7G regulators was able to predict genomic instability, immune checkpoint treatment effectiveness, and overall survival in patients with pancreatic cancer. Once validated in large clinical trials, m7G score could be used to classify PDAC into different groups with different patterns of immune infiltration, genomic instability, and ICI response rate. M7G target genes have the potential to become novel diagnostic biomarkers or therapeutic targets of PDAC.

Author contributors

HY: bioinformatics analysis, manuscript writing, and graphical visualization. JM-P: prepare tissue and coordinating the DSP study, conduct immunohistochemistry, and manuscript editing. AL: experimental design and supervision. AG: experimental supervision. J-lL: provided input to the conception and experimental design. YR: conducted the DSP study and preliminary data analysis. AN: performed initial bioinformatics analysis of the data. HN: bioinformatics analysis supervision. BG: conceptual design, bioinformatics analysis, manuscript revision, and providing fund. Z-hG: conceptual design, selection of experimental tissue, analyze histology and immunohistochemistry data, and manuscript revision. All authors contributed to the article and approved the submitted version.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found in the GEO database, accession number: GSE208536.

Ethics Statement

This study was approved by the Institutional Research Ethics Board of the McGill University Health Center (MP-37-2018-4399, MP-37-2018-3171, MP-37-2020-5723). Written informed consent was waived by the Ethics board for this retrospective study using archived tissue only.

Funding

This work was supported by the National Natural Science Foundation of China [Grant No. 81800573], Peking University People's Hospital Research and Development Funds [Grant No. RDY2018-06 and RS2021-08] and Academic Rising Star Program of Peking University People's Hospital [Grant No. RS2021-08].

Conflict of Interest

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

Publisher's Note

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

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.961457/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The GO and KEGG enrichment analysis of the three PDAC clusters. **(A–C)** GO enrichment analysis of the three clusters. **(D–F)** KEGG enrichment analysis of the three clusters. **(G)** The differentially expressed genes among the three types were intersected, resulting in a total of 907 genes. **(H, I)** GO enrichment analysis of the 907 genes. **(J, K)** KEGG enrichment analysis of the 907 genes. **(L)** Area under the curve (AUC) for time-dependent receiver operating characteristic curves of the m7G score model.

SUPPLEMENTARY FIGURE 2
The 25 drugs were selected for patients with low m7G score.

SUPPLEMENTARY FIGURE 3
CIBERSORT and checkpoint analysis of core genes. **(A)** CIBERSORT analysis of FN1.**(B)** checkpoint analysis of FN1. **(C)** CIBERSORT analysis of ITGB1. **(D)** checkpoint analysis of ITGB1.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590

2. Ghaneh P, Costello E, Neoptolemos JP. Biology and management of pancreatic cancer. *Gut* (2007) 56(8):1134–52. doi: 10.1136/Q26 gut.2006.103333

3. Ding L, Liou GY, Schmitt DM, Storz P, Zhang JS, Billadeau DD. Glycogen synthase kinase-3beta ablation limits pancreatitis-induced acinar-to-ductal metaplasia. *J Pathol* (2017) 243(1):65–77. doi: 10.1002/path.4928

4. Perkail S, Andricovich J, Kai Y, Tzatsos A. BAP1 is a haploinsufficient tumor suppressor linking chronic pancreatitis to pancreatic cancer in mice. *Nat Commun* (2020) 11(1):3018. doi: 10.1038/s41467-020-16589-8

5. Takahashi R, Macchini M, Sunagawa M, Jiang Z, Tanaka T, Valenti G, et al. Interleukin-1β-induced pancreatitis promotes pancreatic ductal adenocarcinoma *via* b lymphocyte-mediated immune suppression. *Gut* (2021) 70(2):330–41. doi: 10.1136/gutjnl-2019-319912

6. Pitarresi JR, Liu X, Sharma SM, Cuitino MC, Kladney RD, Mace TA, et al. Stromal ETS2 regulates chemokine production and immune cell recruitment during acinar-to-ductal metaplasia. *Neoplasia* (2016) 18(9):541–52. doi: 10.1016/j.neo.2016.07.006

7. Hessmann E, Zhang JS, Chen NM, Hasselluhn M, Liou GY, Storz P, et al. NFATc4 regulates sox9 gene expression in acinar cell plasticity and pancreatic cancer initiation. *Stem Cells Int* (2016) 2016:5272498. doi: 10.1155/2016/5272498

8. Wong CH, Li YJ, Chen YC. Therapeutic potential of targeting acinar cell reprogramming in pancreatic cancer. *World J Gastroenterol* (2016) 22(31):7046–57. doi: 10.3748/wjg.v22.i31.7046

9. Zechner D, Kroemer T, Albert AC, Schonrogge M, Radecke T, Vollmar B. Analysis of Axin2 expression and function in murine models for pancreatic cancer. *Cell Biosci* (2016) 6(1):49. doi: 10.1186/s13578-016-0116-4

10. Palagani V, Bozko P, El Khatib M, Belahmer H, Giese N, Sipos B, et al. Combined inhibition of notch and JAK/STAT is superior to monotherapies and impairs pancreatic cancer progression. *Carcinogenesis* (2014) 35(4):859–66. doi: 10.1093/carcin/bgt394

11. Zhao Y, Kong L, Pei Z, Li F, Li C, Sun X, et al. m7G methyltransferase METTL1 promotes post-ischemic angiogenesis *via* promoting VEGFA mRNA translation. *Front Cell Dev Biol* (2021) 9:642080. doi: 10.3389/fcell.2021.642080

12. Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT. Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* (2017) 14(9):1138–52. doi: 10.1080/15476286.2016.1259781

13. Liu Y, Zhang Y, Chi Q, Wang Z, Sun B. Methyltransferase-like 1 (METTL1) served as a tumor suppressor in colon cancer by activating 7-methyguanosine (m7G) regulated let-7e miRNA/HMGA2 axis. *Life Sci* (2020) 249:117480. doi: 10.1016/j.lfs.2020.117480

14. Ma J, Han H, Huang Y, Yang C, Zheng S, Cai T, et al. METTL1/WDR4-mediated m(7)G tRNA modifications and m(7)G codon usage promote mRNA translation and lung cancer progression. *Mol Ther* (2021) 29(12):3422–35. doi: 10.1016/j.ymthe.2021.08.005

15. Zhang M, Song J, Yuan W, Zhang W, Sun Z. Roles of RNA methylation on tumor immunity and clinical implications. *Front Immunol* (2021) 12:641507. doi: 10.3389/fimmu.2021.641507

16. Aregger M, Kaskar A, Varshney D, Fernandez-Sanchez ME, Inesta-Vaquera FA, Weidlich S, et al. CDK1-cyclin b1 activates rnmt, coordinating mrna cap methylation with g1 phase transcription. *Mol Cell* (2016) 61(5):734–46. doi: 10.1016/j.molcel.2016.02.008

17. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* (2020) 38(5):586–99. doi: 10.1038/s41587-020-0472-9

18. Nagtegaal I, Odze R, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology* (2020) 76(2):182–8. doi: 10.1111/his.13975

19. Sun W, Li J, Liu S, Wu J, Zhou H, Qu L, et al. RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res* (2016) 44:D259–265. doi: 10.1093/nar/gkv1036

20. Xuan J, Sun W, Lin P, Zhou K, Liu S, Zheng L, et al. RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res* (2018) 46:D327–34. doi: 10.1093/nar/gkx934

21. Deng S, Zhang H, Zhu K, Li X, Ye Y, Li R, et al. M6A2Target: a comprehensive database for targets of m6A writers, erasers and readers. *Briefings Bioinf* (2021) 22(3):bbaa055. doi: 10.1093/bib/bbaa055

22. Song B, Tang Y, Chen K, Wei Z, Rong R, Lu Z, et al. m7GHub: deciphering the location, regulation and pathogenesis of internal mRNA N7-methylguanosine (m7G) sites in human. *Bioinf (Oxford England)* (2020) 36(11):3528–36. doi: 10.1093/bioinformatics/btaa178

23. Zhou Z, Zhang J, Xu C, Yang J, Zhang Y, Liu M, et al. An integrated model of N6-methyladenosine regulators to predict tumor aggressiveness and immune evasion in pancreatic cancer. *EBioMedicine* (2021) 65:103271. doi: 10.1016/j.ebiom.2021.103271

24. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

25. Jachimowicz R, Beleggia F, Isensee J, Velpula B, Goergens J, Bustos M, et al. UBQLN4 represses homologous recombination and is overexpressed in aggressive tumors. *Cell* (2019) 176(3):505–519.e522. doi: 10.1016/j.cell.2018.11.024

26. Li Y, He Y, Peng J, Su Z, Li Z, Zhang B, et al. Mutant kras co-opts a proto-oncogenic enhancer network in inflammation-induced metaplastic progenitor cells to initiate pancreatic cancer. *Nat Cancer* (2021) 2(1):49–65. doi: 10.1038/s43018-020-00134-z

27. Canon J, Rex K, Saiki A, Mohr C, Cooke K, Bagal D, et al. The clinical KRAS (G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* (2019) 575 (7781):217–23. doi: 10.1038/s41586-019-1694-1

28. Chen N, Neesse A, Dyck M, Steuber B, Koenig A, Lubeseder-Martellato C, et al. Context-dependent epigenetic regulation of nuclear factor of activated t cells 1 in pancreatic plasticity. *Gastroenterology* (2017) 152(6):1507–1520.e1515. doi: 10.1053/j.gastro.2017.01.043

29. Barkin J, Freeman M, Barkin J. Is it acute pancreatitis or recurrent acute pancreatitis leading to chronic pancreatitis that increases pancreatic cancer risk? *Gastroenterology* (2018) 155(4):1279–80. doi: 10.1053/j.gastro.2018.09.023

30. Rozenblum E, Schutte M, Goggins M, Hahn SA, Panzer S, Zahurak M, et al. Tumor-suppressive pathways in pancreatic carcinoma. *Cancer Res* (1997) 57(9):1731–4.

31. Storz P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat Rev Gastroenterol Hepatol* (2017) 14(5):296–304. doi: 10.1038/nrgastro.2017.12

32. Zollinger DR, Lingle SE, Sorg K, Beechem JM, Merritt CR. Geomx rna assay: high multiplex, digital, spatial analysis of rna in ffpe tissue. *Methods Mol Biol* (2020) 2148:331–45. doi: 10.1007/978-1-0716-0623-0_21

33. Jonckheere N, Auwercx J, Hadj Bachir E, Coppin L, Boukrout N, Vincent A, et al. Unsupervised hierarchical clustering of pancreatic adenocarcinoma dataset from tcga defines a mucin expression profile that impacts overall survival. *Cancers (Basel)* (2020) 12(11):3309. doi: 10.3390/cancers12113309

34. Chuvin N, Vincent DF, Pommier RM, Alcaraz LB, Gout J, Caligaris C, et al. Acinar-to-ductal metaplasia induced by transforming growth factor beta facilitates kras(g12d)-driven pancreatic tumorigenesis. *Cell Mol Gastroenterol Hepatol* (2017) 4(2):263–82. doi: 10.1016/j.jcmgh.2017.05.005

35. Ansari D, Carvajo M, Bauden M, Andersson R. Pancreatic cancer stroma: controversies and current insights. *Scand J Gastroenterol* (2017) 52:641–6. doi: 10.1080/00365521.2017.1293726

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

37. Wu XR, Chen Z, Liu Y, Chen ZZ, Tang F, Chen ZZ, et al. Prognostic signature and immune efficacy of m(1) a-, m(5) c- and m(6) a-related regulators in cutaneous melanoma. *J Cell Mol Med* (2021) 25(17):8405–18. doi: 10.1111/jcmm.16800

38. Spranger S, Gajewski TF. Tumor-intrinsic oncogene pathways mediating immune avoidance. *Oncoimmunology* (2016) 5(3):e1086862. doi: 10.1080/2162402X.2015.1086862

39. Brahmer J, Tykodi S, Chow L, Hwu W, Topalian S, Hwu P, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *New Engl J Med* (2012) 366(26):2455–65. doi: 10.1056/NEJMoa1200694

40. O'Reilly E, Oh D, Dhani N, Renouf D, Lee M, Sun W, et al. Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: a phase 2 randomized clinical trial. *JAMA Oncol* (2019) 5 (10):1431–8. doi: 10.1001/jamaoncol.2019.1588

41. von Bernstorff W, Voss M, Freichel S, Schmid A, Vogel I, Jöhnk C, et al. Systemic and local immunosuppression in pancreatic cancer patients. *Clin Cancer Res Off J Am Assoc Cancer Res* (2001) 7:925s–32s.

42. Balachandran V, Łuksza M, Zhao J, Makarov V, Moral J, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* (2017) 551(7681):512–6. doi: 10.1038/nature24462

43. Filippini D, Agosto S, Delfino P, Simbolo M, Piro G, Rusev B, et al. Immunoevolution of mouse pancreatic organoid isografts from preinvasive to metastatic disease. *Sci Rep* (2019) 9(1):12286. doi: 10.1038/s41598-019-48663-7

44. Asahina K, Balog S, Hwang E, Moon E, Wan E, Skrypek K, et al. Moderate alcohol intake promotes pancreatic ductal adenocarcinoma development in mice expressing oncogenic kras. *Am J Physiol Gastrointest Liver Physiol* (2020) 318(2): G265–76. doi: 10.1152/ajpgi.00218.2019

45. Yamaguchi T, Ikehara S, Akimoto Y, Nakanishi H, Kume M, Yamamoto K, et al. TGF-β signaling promotes tube-structure-forming growth in pancreatic duct adenocarcinoma. *Sci Rep* (2019) 9(1):11247. doi: 10.1038/s41598-019-47101-y

46. Murtaugh L. Keefe mjarop: regeneration and repair of the exocrine pancreas. *Annu Rev Physiol* (2015) 77:229–49. doi: 10.1146/annurev-physiol-021014-071727.

47. Žlajpah M, Boštjančič E, Tepeš B. Zidar njc: expression of extracellular matrix-related genes and their regulatory micrornas in problematic colorectal polyps. *Cancers (Basel)* (2020) 12(12):3715. doi: 10.3390/cancers12123715

48. Xavier C, Castro I, Caires H, Ferreira D, Cavadas B, Pereira L, et al. Vasconcelos mjcl: chitinase 3-like-1 and fibronectin in the cargo of extracellular vesicles shed by human macrophages influence pancreatic cancer cellular response to gemcitabine. *Cancer Lett* (2021) 501:210–23. doi: 10.1016/j.canlet.2020.11.013

49. Zhao W, Ajani J, Sushovan G, Ochi N, Hwang R, Hafley M, et al. Galectin-3 mediates tumor cell-stroma interactions by activating pancreatic stellate cells to produce cytokines *via* integrin signaling. *Gastroenterology* (2018) 154(5):1524–37.e1526. doi: 10.1053/j.gastro.2017.12.014

50. Yang Z, Zhou X, Liu Y, Gong C, Wei X, Zhang T, et al. Gao QJA-caimc: Activation of integrin β1 mediates the increased malignant potential of ovarian cancer cells exerted by inflammatory cytokines. *Anticancer Agents Med Chem* (2014) 14(7):955–62. doi: 10.2174/1871520614666140613123108

51. Liu M, Zhang Y, Yang J, Cui X, Zhou Z, Zhan H, et al. Zip4 increases expression of transcription factor zeb1 to promote integrin α3β1 signaling and inhibit expression of the gemcitabine transporter ent1 in pancreatic cancer cells. *Gastroenterology* (2020) 158(3):679–92.e671. doi: 10.1053/j.gastro

52. Qiu Y, Chen T, Hu R, Zhu R, Li C, Ruan Y, et al. Next frontier in tumor immunotherapy: macrophage-mediated immune evasion. *biomark Res* (2021) 9 (1):72. doi: 10.1186/s40364-021-00327-3

53. Shan Z, Zhao Y, Zhang J, Yan Z, Wang T, Mao F, et al. Fasl pd-l2 identifies a novel immunosuppressive neutrophil population in human gastric cancer that promotes disease progression. *Adv Sci (Weinheim Baden-Wurttemberg Germany)* (2022) 9(5):e2103543. doi: 10.1002/advs.202103543

54. Kargl J, Zhu X, Zhang H, Yang G, Friesen T, Shipley M, et al. Neutrophil content predicts lymphocyte depletion and anti-PD1 treatment failure in NSCLC. *JCI Insight* (2019) 4(24):3309. doi: 10.1172/jci.insight.130850

55. Pitt J, Vétizou M, Daillère R, Roberti M, Yamazaki T, Routy B, et al. Resistance mechanisms to immune-checkpoint blockade in cancer: tumor-intrinsic and -extrinsic factors. *Immunity* (2016) 44(6):1255–69. doi: 10.1016/j.immuni.2016.06.001

56. Garner H, de Visser K. Immune crosstalk in cancer progression and metastatic spread: a complex conversation. *Nat Rev Immunol* (2020) 20(8):483–97. doi: 10.1038/s41577-019-0271-z

57. Burger M, Cruz A, Crossland G, Gaglia G, Ritch C, Blatt S, et al. Antigen dominance hierarchies shape TCF1 progenitor CD8 T cell phenotypes in tumors. *Cell* (2021) 184(19):4996–5014.e4926. doi: 10.1016/j.cell.2021.08.020

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 August 2022
DOI 10.3389/fimmu.2022.957751

Check for updates

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Fan Zhang,
University of New South Wales, Australia
Yanglan Gan,
Donghua University, China

*CORRESPONDENCE
Hui Liu
hliu@njtech.edu.cn
Xuejun Liu
xjliu@njtech.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 31 May 2022
ACCEPTED 27 June 2022
PUBLISHED 08 August 2022

CITATION
Jiao M, Liu H and Liu X (2022) Transcriptional patterns reveal tumor histologic heterogeneity and immunotherapy response in lung adenocarcinoma.
Front. Immunol. 13:957751.
doi: 10.3389/fimmu.2022.957751

COPYRIGHT
© 2022 Jiao, 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.

Transcriptional patterns reveal tumor histologic heterogeneity and immunotherapy response in lung adenocarcinoma

Mengxue Jiao, Hui Liu* and Xuejun Liu*

School of Computer Science and Technology, Nanjing Tech University, Nanjing, China

Tumoral heterogeneity has proven to be a leading cause of difference in prognosis and acquired drug resistance. High intratumor heterogeneity often means poor clinical response and prognosis. Histopathological subtypes suggest tumor heterogeneity evolved during the progression of lung adenocarcinoma, but the exploration of its molecular mechanisms remains limited. In this work, we first verified that transcriptional patterns of a set of differentially expressed genes profoundly revealed the histologic progression of lung adenocarcinoma. Next, a predictive model based on the transcriptional patterns was established to accurately distinguish histologic subtypes. Two crucial genes were identified and used to construct a tumor heterogeneous scoring model (L2SITH) to stratify patients, and we found that patients with low heterogeneity score had better prognosis. Low L2SITH scores implied low tumor purity and beneficial tumor microenvironment. Moreover, L2SITH effectively identified cohorts with better responses to anti–PD-1 immunotherapy.

KEYWORDS

intratumor heterogeneity, immune microenvironment, histologic progression, prognostic model, transcriptional pattern

1 Introduction

Non–small cell lung cancer (NSCLC) is one of the malignant tumors over the world (1), and its morbidity and mortality increased gradually in recent years. The 5-year survival rate of patients with NSCLC is only 18%. Lung adenocarcinoma (LUAD) is the main subtype, accounting for 40% of patients with NSCLC (2). LUAD has distinctive histological stages during its progression. Pathologists have categorized difference of histological phenotype, referred to as histologic subtypes, including lepidic, papillary, acinar, and solid (3). With the histologic progression from lepidic to solid, LUADs become increasingly aggressive and metastatic.

The immune system was supposed to drive intertumor and intratumor heterogeneity by exerting different selective pressures to different regions of the solid tumor (4). Prior studies have shown significant difference of stroma and immune infiltrating cells in different intratumor regions, as well as interpatient tumors (5, 6). Accordingly, the tumor microenvironment has been shown to play an important role in tumor growth, angiogenesis, immune evasion, and metastasis (7). Substantial evidence suggests that tumor heterogeneity increases the likelihood that cancer cells survive conventional chemotherapy and targeted anticancer drugs (8–10). In addition, tumor heterogeneity affects the efficacy of immunotherapies, especially immune checkpoint inhibitors (11–13). However, the exploration of molecular mechanism underlying the histologic heterogeneity remained nascent (14).

In fact, the potential molecular mechanism of histologic heterogeneity of LUAD is multifaceted. The genomic aberrations, epigenetic modifications, small-molecule RNA, malfunction of transcriptional regulations, and environmental factors may lead to phenotypic differences (15–17). Studies have found strong associations between the histologic heterogeneity and prognosis of LUAD, but the investigation of molecular signature underlying each histologic subtype is scarce (18). In this paper, we set about to find molecular determinations of tumor heterogeneity from the perspective of histologic subtypes. The immunogenetic transcriptional patterns showed strong link to the histologic progression and tumor microenvironment. We established a heterogeneous scoring model (L2SITH) based on the molecular signatures to stratify patients and found that patients in the low-scored group had better prognosis, which was more predictive than the stratification based on histologic subtypes. In contrast, histologic subtypes did not showed significant prognostic value for patients with LUAD. Moreover, L2SITH effectively identified cohorts with better responses to anti–PD-1 immunotherapy.

2 Materials and methods

2.1 Data sources

Two LUAD cohorts from The Cancer Genome Atlas (TCGA) (N = 246) and Gene Expression Omnibus (GEO) (GSE58772, N = 48) were included in our study. The RNA-seq and clinical data were obtained from the Genomic Data Commons (https://gdc.cancer.gov/) and GEO database (https://www.ncbi.nlm. nih.gov/geo/), respectively.

Intratumoral heterogeneity of these samples was annotated as lepidic (L), papillary (P), acinar (A), or solid (S) subtypes, according to the most popular histopathological classification standard (3). The annotations of the TCGA cohort were from (19), which included lepidic (N = 10), papillary (N = 50), acinar (N = 69), and solid (N = 58) patients. The GEO GSE58772 cohort included lepidic (N = 10), papillary (N = 18), acinar (N = 10). and solid (N = 10) samples (20).

The immune subtypes of 455 LUAD samples were marked by Thorsson et al. (21), including C1 (wound healing, N = 82), C2 (Interferon (IFN)-dominant, N = 147), C3 (inflammatory, N = 178), C4 (lymphocyte depleted, N = 20), and C6 (Transforming Growth Factor (TGF)-dominant, N = 28).

The cohort treated with the PD-1 inhibitor Nivolumab was obtained from GEO (GSE126044, N = 16). This cohort included 11 non-responders (11 PD cases) and five responders (one SD case and four partial response (PR) cases).

2.2 Differential expression and enrichment analysis

The *DESeq2* R package (22) was used to conduct differential analysis between normal and two most representative subtypes, including lepidic vs. normal, solid vs. normal, and lepidic vs. solid. The differentially expressed genes were chosen using the filtering criterion of absolute $|\ log_2FoldChange\ | > 1$ and p.adj < 0.05. The differential expression genes overlapped with immune-related genes from InnateDB were filtered out for further analysis. Out of the 1,952 immune genes, we got 96 immune-related differentially expressed genes.

The *clusterProfiler* R package (23) was applied for Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The *GOPlot* R package (24) was used to calculate the z-score with the filtering threshold of P.adj < 0.05 to select statistically significant pathways.

2.3 Histologic subtype clustering analysis

To visualize the molecular signature difference among histologic subtypes, ComplexHeatmap package (25) was used to draw the heatmap of expression profiles. The *tSNE* (26) and *UMAP* (27) tools were used to perform dimensionality reduction and clustering of expression profile of genes underlying the tumor heterogeneity. The *ggplot2* package (28) was used to display the clusters.

2.4 MLP model for histologic subtype classification

A multilayer perceptron (MLP) model was constructed using the *Neuralnet* (29) package. The input layer included 38 nodes corresponding to differentially expressed immune-related genes. The two hidden layers include 11 and 9 nodes, respectively. There were three nodes in the output layer for classification of histologic subtypes. The performance of the classification model was evaluated by ROC curve and Receiver Operating Characteristic (ROC)-Area Under Curve (AUC)

values. The sklearn and matplotlib tools were used to calculate the ROC-AUC values and plot ROC curves.

2.5 Random forest for immune subtype classification

The transcriptomic data of filtered genes were transformed to *z*-scores and then fed into a random forest model to predict immune subtypes (C1, C2, C3, C4, or C6). The samples were split to training and test set by 7:3 ratio. The random forest model included 50 decision trees with a maximum tree depth of 5 and a maximum number of leaf nodes of 50. The sklearn and matplotlib tools were used to calculate the ROC-AUC values and plot ROC curves.

2.6 Tumor purity and immune microenvironment analysis

The *ESTIMATE* (30) tool was used to calculate the tumor purity, stromal, and immune scores. The *t*-test was used for statistical significance. The *corrplot* (31) package was used to plot heatmap, in which $p < 0.05$ was denoted by *, $p < 0.01$ by **, and $p < 0.001$ by ***.

2.7 Establishment of L2SITH score model

We performed univariate and multivariate Cox regression analysis regarding the set of genes related to histologic subtype. The genes significantly related to prognosis were used to construct the heterogeneity score model L2SITH. The *survival* tepallison2010survival package was used to run survival analysis for high- and low-scored group, and the Kaplan-Meier (K-M) curves were plotted by *survminer* (32) package. The pRRophetic (33) package was used for drug proposal based on Cancer Cell Line Encyclopedia drug sensitivity dataset.

3 Results

3.1 Transcriptional patterns reflect histologic progression

Prior studies have reported the histologic heterogeneity of LUAD, which mainly included four subtypes: lepidic, papillary, acinar, and solid during tumor progression (3). From lepidic to solid, tumor aggressiveness and metastasis increase. For simplicity, we mainly focused on lepidic and solid patterns, regarding papillary and acinar as intermediate state. On the basis of the differential expression analysis between normal, lepidic, and solid samples, 96 differentially expressed genes related to immunologic function were filtered out, as shown in Figure 1. After dimensionality reduction and visualization by tSNE and UMAP tools, the transcriptional patterns clearly distinguished the histologic subtypes, as shown in Figure 2. The enriched GO annotations of 96 differentially expressed immune genes were shown in Figure S1. The 38 significantly upregulated immune genes showed positive correlation (Figures S2, S3).

As shown in Figure 3, these screened genes showed distinctive transcriptional patterns among different histologic subtypes. Interestingly, their expression levels increased significantly from lepidic to solid pattern. This may implied the activation of the immune response during tumor progression so that more and more immune cells infiltrated into the tumor. Our analysis preliminarily verified that transcriptional patterns of immunologic genes reflected the histologic heterogeneity in LUAD.

3.2 Transcriptional patterns accurately predict histologic subtypes

As molecular mechanism underlies the cell phenotype, we supposed transcriptional patterns should be predictive of phenotypic label. For this purpose, a MLP model was trained to classify histologic subtypes. We split the 187 samples into three types according to histologic subtypes: lepidic (N = 10), papillary and acinar (N = 119), and solid (N = 58). This was a typical multi-class classification task, based on the molecular signature of the set of immune-related genes. Expectedly, the MLP model achieved extremely high performance. The ROC-AUC was close to 1, and the accuracy rate reaches 97%, as shown in Figure 4A. In particular, 8 of 10 lepidic samples were correctly classified, 117 of 119 papillary and acinar samples were correctly classified, and all 58 solid samples were correctly classified. Moreover, we found that, as the histologic pattern progressed, a higher accuracy of the model is achieved. This implied that progressive tumor tended to develop distinctive molecular signatures that dominate the histological morphology and cell phenotype.

To verify the generalization, we verified the classification model on a GEO cohort (N = 48, where L = 10, P&A = 28, and S = 10). We were pleased to find that the model can completely distinguish the three types of histologic subtypes correctly, as shown in Figure 4B. The independent test set fully validated the outstanding potential of immunogenetic molecular signature in differentiating tumor histologic heterogeneity.

3.3 Establishment of heterogeneity scoring model L2SITH

We further performed Cox regression and survival analysis to screen genes significantly related to prognosis and obtained two key genes KIR2DL4 and SLC7A7. It has been reported that KIR2DL4 was highly associated with cancer development (34), and its lower expression level means better prognosis (Figure 5A). SLC7A7 was a suppressor gene inhibiting the progression of LUAD (35), and its

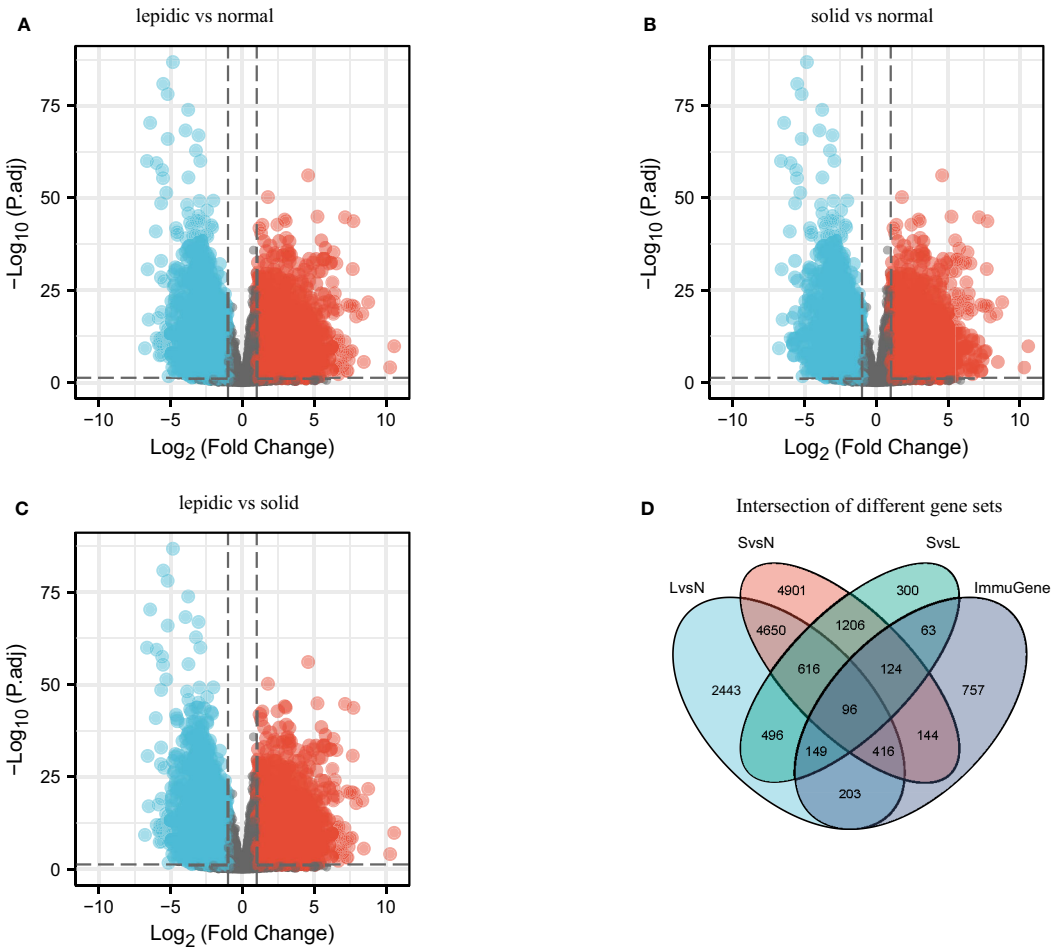


FIGURE 1
Differential analysis between normal, lepidic, and solid histologic subtypes. **(A)** Differential expression genes between lepidic and normal samples (LvsN). **(B)** Differential expression genes between solid and normal samples (SvsN). **(C)** Differential expression genes between lepidic and solid samples (LvsS). **(D)** Intersection of differentially expressed genes and immune genes (ImmuGene).

higher expression level yielded to better prognosis (Figure 5B). Therefore, we established the heterogeneity score model L2SITH using these two genes, using the regression coefficients and the transcriptional levels of these two genes. Its formal definition is L2SITH = −0.0003708 * SLC7A7 (express) + 0.0047614 * KIR2DL4 (express). Using the L2SITH score, we divided the 62 lepidic and solid patients into high- and low-scored groups. The KM survival curves showed that the low-scored group had significantly high overall survival rate, as shown in Figure 5C. However, if the patients were divided into groups by histologic subtype (lepidic vs. solid), then the overall survival has no significant difference (Figure 5D). This suggested that the L2SITH model captured molecular factors underlying tumor progression, thereby acquired better prognostic power than the histologic heterogeneity. To verify this point, we further divided the solid samples (N = 52) using L2SITH scores and showed that the low-scored subgroup still had better overall survival, as shown in Figure 5E.

3.4 L2SITH revealed beneficial tumor microenvironment

To validate the association between transcriptional patterns and tumor microenvironment, we adopted the immune subtypes established by Thorsson et al. (21) for analysis. We divided the 890 samples by 7:3 for training and test and built a random forest model to predict the immune subtypes from molecular signature. On the training and test set, the accuracy reached 0.867 and 0.813, respectively. The detail of each immune subtypes was shown in Figures 6A, B, and the ROC curve was shown in Figure 6C. This demonstrated that the transcriptional patterns of the key immune-related genes can significantly differentiate immune subtypes of LUAD.

We used ESTIMATE tool to evaluate the immune infiltration level of the high- and low-scored group, to explore the significance of the L2SITH in the stratification of tumor

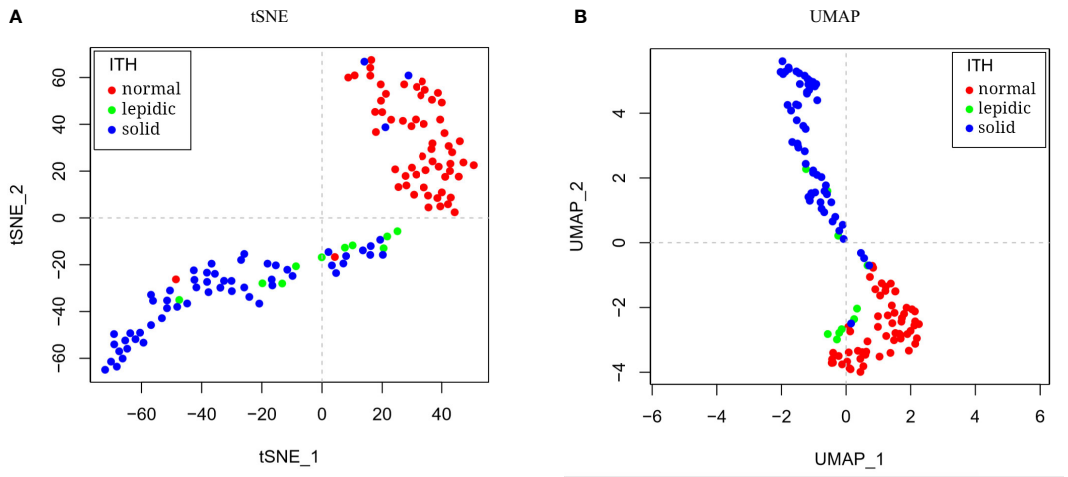


FIGURE 2
Dimension reduction and visualization of histologic subtypes. The LUAD samples showed separate distribution by **(A)** tSNE and **(B)** UMAP dimension reduction on transcriptional profiles.

microenvironment. As shown in Figure 7, we found that low-scored L2SITH group had higher stromal level, immune infiltration, and ESTIMATE scores than high-scored group. Conversely, when samples were grouped by histologic subtype, we observed a different trend, that is, the solid samples had higher stromal level, immune infiltration, and ESTIMATE scores than lepidic samples. The results indicated that the patients with low L2SITH scores had low tumor purity and beneficial tumor microenvironment to immunotherapy.

3.5 L2SITH accurately predicts clinical immunotherapy response

To validate the effectiveness of the L2SITH score model in predicting response to immunotherapy, we analyzed a cohort of patients with NSCLC received Nivolumab PD-1 inhibitor treatment. Using the L2SITH to divide the sample into high- and low-scored groups, we observed significant difference in clinical response (Figures 8A, B). Clearly, the low-scored group

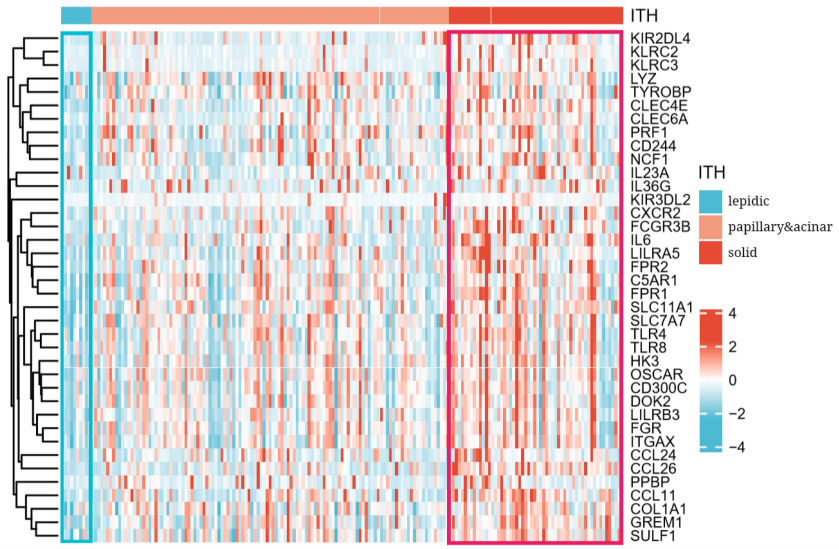


FIGURE 3
Heatmap of transcriptional patterns reflected histologic progression from lepidic to solid subtype.

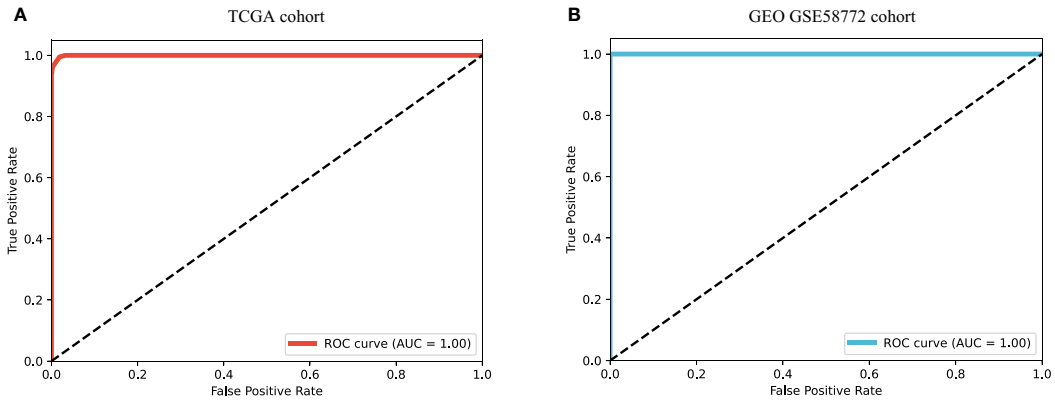


FIGURE 4
RCO curves of MLP model for histologic subtype prediction using transcriptional pattern. **(A)** ROC curve on TCGA cohort. **(B)** ROC curve on GEO GSE58772 cohort.

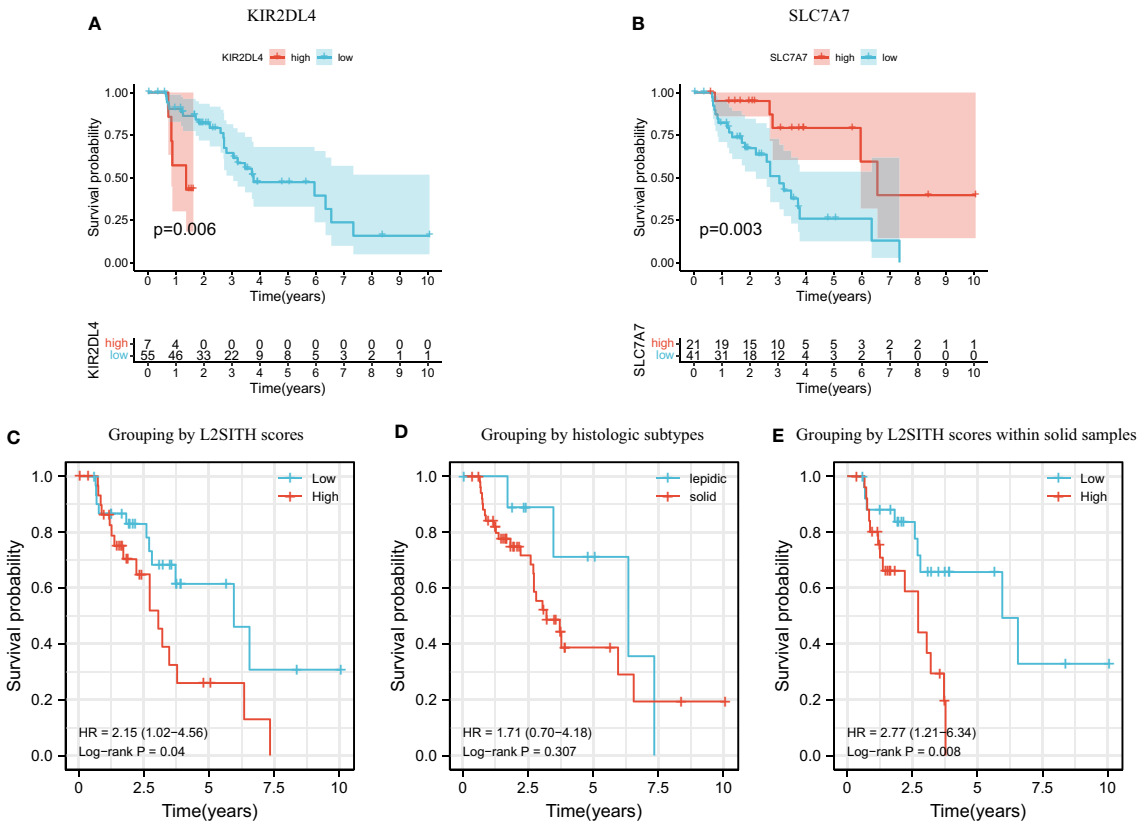


FIGURE 5
Survival analysis of two key genes, L2SITH heterogeneity score, and histologic phenotype. Panels **(A, B)** showed the K-M survival curves of patients grouped by KIR2DL4 and SLC7A7 gene, respectively. **(C)** K-M survival curve of patients grouped by L2SITH heterogeneity score. **(D)** K-M survival curve of patients grouped by lepidic and solid subtypes. **(E)** K-M survival curve of the solid group patients divided by L2SITH heterogeneity score.

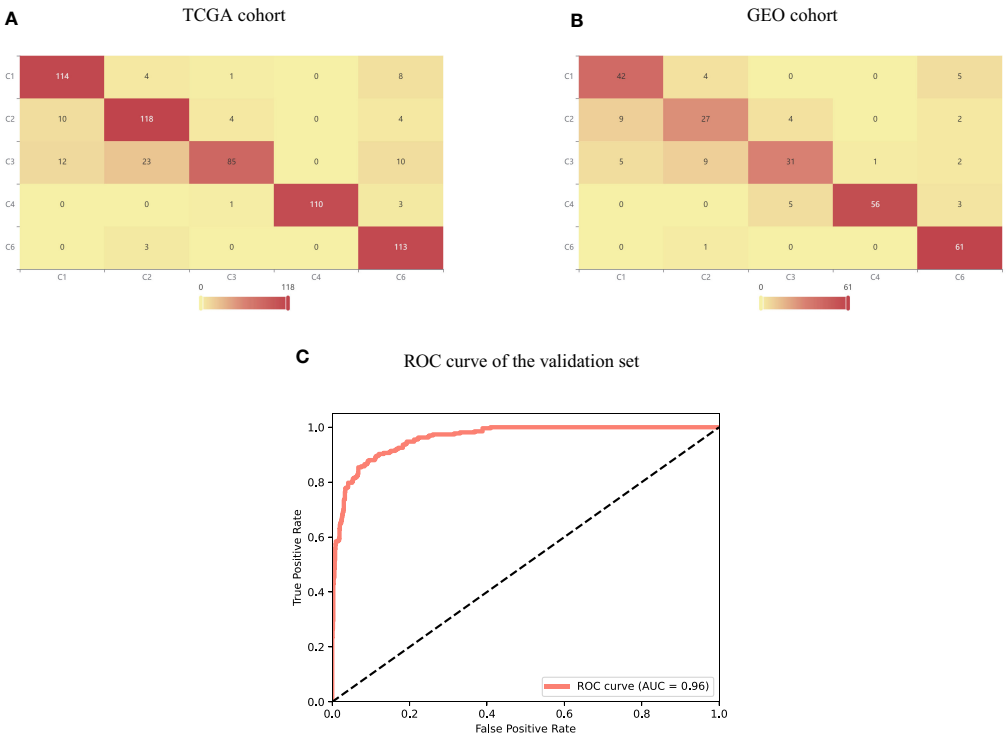


FIGURE 6
Performance of immune subtypes prediction based on transcriptional pattern. **(A)** Immune subtype prediction results of TCGA cohort. **(B)** Immune subtype prediction results of GEO cohort. **(C)** ROC curve of immune subtype prediction on GEO cohort.

had notably better OS and PFS than high-scored group. Furthermore, we found that the patients who responded to Nivolumab treatment were all in the low-scored group, as shown in Figure 8C. The clinical response to immunotherapy in the high-scored group was weak, and these patients suffered progressive disease (Figure 8D). This confirmed the reliable performance of L2SITH in predicting the immune response of NSCLC.

In addition to the prognostic value, we further utilized the L2SITH scores to screen potential beneficial drugs. Three approved drugs, vinorelbine, gemcitabine, and etoposide, were predicted to have higher potential sensitivity to the low-scored group stratified by L2SITH (Figure 9). Vinorelbine is an anti-mitotic chemotherapy drug that was approved in 1990s for the treatment of non-NSCLC (29). According to a recent study, vinorelbine is a suitable choice for elderly patients with NSCLC and also a partner drug with immunochemotherapy (36). Gemcitabine is a nucleoside metabolism inhibitor approved by FDA in combination with cisplatin for the treatment of NSCLC (37). Low-dose gemcitabine treatment is sufficient to inhibit tumor growth with few side effects *in vivo*. Gemcitabine can also activate antitumor immune response in patients with normal immune system (38). Etoposide, a coccine toxin derivative, has also been shown to be useful in the treatment of small cell lung tumors (39). Etoposide plus cisplatin chemotherapy improved the efficacy and safety of small cell lung cancer (40).

4 Discussion

At present, the prognosis of patients with cancer depends mainly on the clinical staging and pathological grading. In recent years, more and more studies have shown that intratumor heterogeneity is an important factor of clinical treatment efficacy and prognosis. Tumors with high heterogeneity tend to be more aggressive and indicate poor prognosis. In fact, interpatient and intratumor heterogeneity is prevalent in both lymphoma and solid tumor. Apart from the molecular feature difference, the heterogeneity is also reflected in immune microenvironment, such as different immune infiltration level and tumor purity.

This study tried to reveal the underlying molecular features of histologic heterogeneity in LUAD. Through in-depth exploration of the transcriptional profiles of immune-related genes, we constructed a machine learning model to predict histologic subtypes. In addition, the transcriptional profiles were highly predictive of tumor immune subtypes. From these

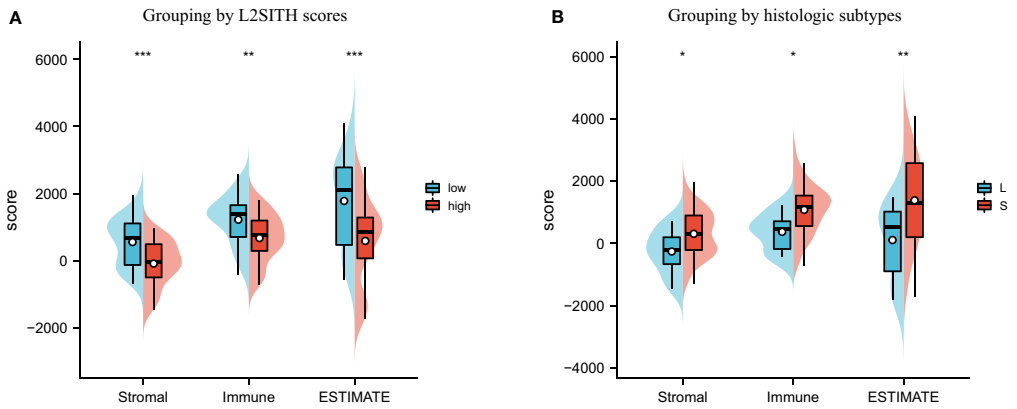


FIGURE 7
Tumor purity estimation of LUAD samples grouped by L2SITH score and histologic subtype. **(A)** Stromal, immune, and ESTIMATE scores of high- and low-scored groups. **(B)** Stromal, immune, and ESTIMATE scores of lepidic and solid histologic subtypes. The symbol * indicates statistical significance at the 0.05 level ($p<0.05$), ** indicates the level of significance 0.01 ($p<0.01$), *** indicates the level of significance 0.001 ($p<0.001$).

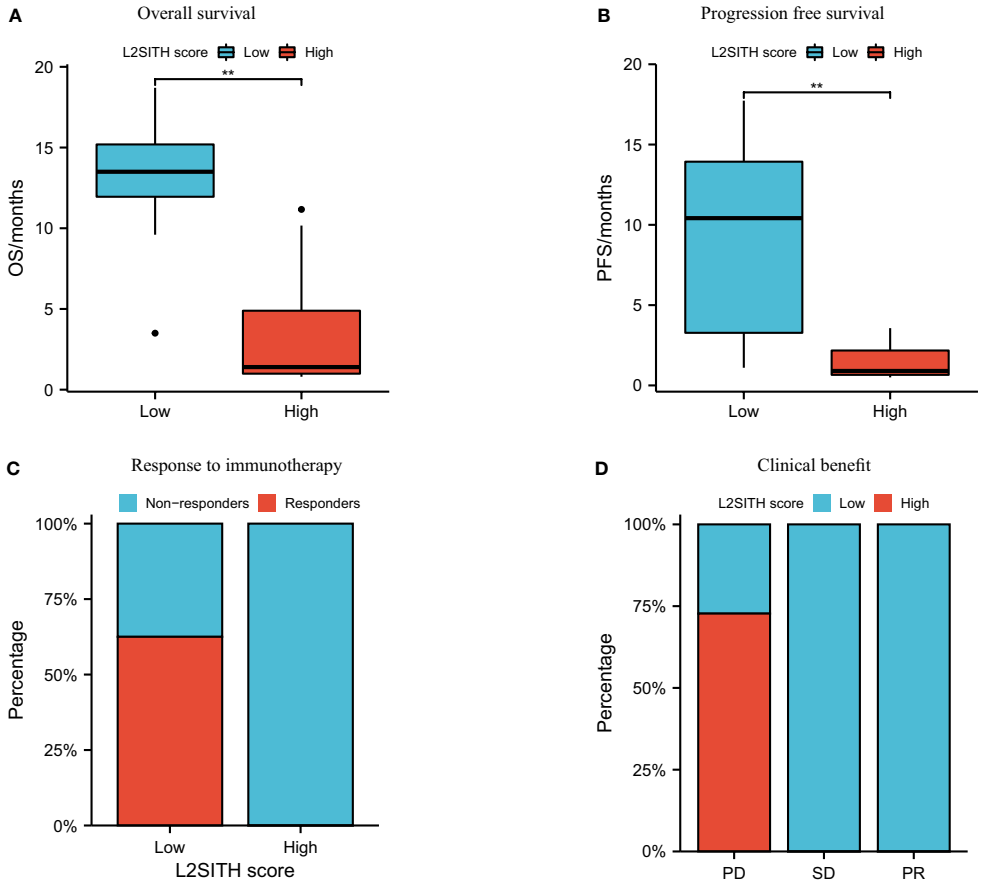


FIGURE 8
Survival and clinical response to anti–PD-1 inhibitor (Nivolumab) of high- and low-L2SITH score groups. **(A)** Overall survival of patients in high- and low-scored groups. **(B)** Progression-free survival of patients in high- and low-scored groups. **(C)** Percentage of patients with response to Nivolumab high- and low-scored groups. **(D)** Percentage of high- and low-scored patients with clinical response from Nivolumab, including progressive disease (PD), stable disease (SD), and partial response (PR). The symbol ** indicates significant difference at the 0.01 level ($p<0.01$).

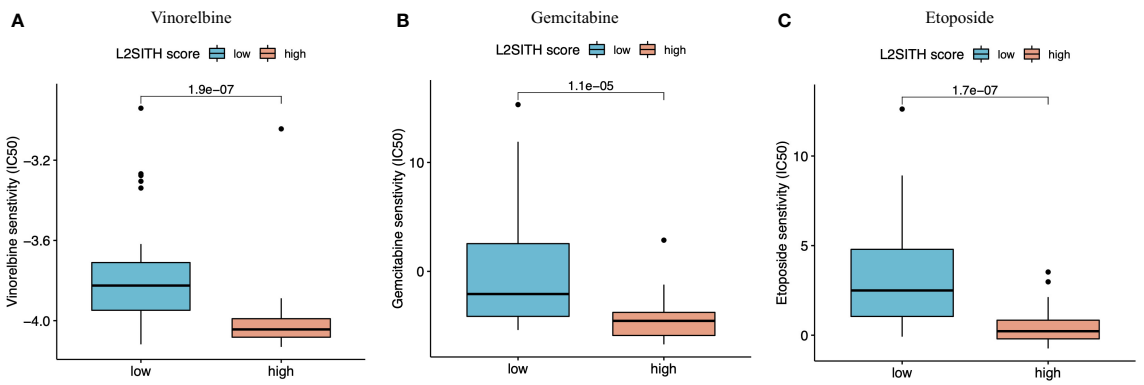


FIGURE 9
Potential beneficial drugs to patients with LUAD grouped by L2SITH score. **(A)** Vinorelbine, **(B)** gemcitabine, and **(C)** etoposide were inferred to benefit the low-scored LUAD patients.

results, the molecular signature reflects the essential causes of phenotypic differences.

We found significant correlation between the transcriptional pattern of immune genes and the histologic progression (Figure S4). During the early stage, most immune genes were not activated and showed low transcriptional levels. With the histologic progression, immune genes were activated along with the tumor immune infiltration increased. This was reflected by the significant positive correlation between the expression of immune genes and histologic subtypes from lepidic to solid. We also explored the correlation of the differentially expressed immune genes with mutation type, but no significant association was found (Figure S5).

From the perspective of histologic progression, the lepidic cells were more differentiated, whereas solid cells were poorly differentiated and aggressive, but we did not observed significant difference in overall survival between lepidic and solid groups. This motivated us to find the molecular factors that really lead to differences in histologic subtypes and survival. Upon the transcriptional profiles, we developed L2SITH, a simple but predictive two-gene score model. The patient stratified by L2SITH scores showed statistically significant differences in overall survival. Within the solid group, L2SITH score model was still effective in distinguishing patients with better prognosis. On the other hand, this suggested that patients grouped in the same histologic subtypes still have significantly different molecular mechanisms. Rather than phenotypic differences, molecular mechanisms are the real reasons for tumor progression. In fact, a prior study (14) has explored the impact of epigenetic factors on histological heterogeneity. It concluded that the LUAD histological progression from lepidic to solid was mainly caused by epigenetic and transcriptional factors. We were exactly inspired by this prior study and focused on the transcriptomic mechanism that actually drove the histological progression but has not been explored before. In addition, we are also interested in the small-molecule RNAs that are potentially associated to histological subtypes and plan to explore such association analysis in the near future work.

Finally, we tried to explore the association of genomic mutations with the histologic progression in LUAD. However, the mutation landscape of driver genes cannot reveal the histologic progression. Moreover, we inspected the most highly mutated genes in LUAD but found no consistent trends of histologic progression driven by highly mutated gene (Figures S6, S7). These were consistent with the conclusion that the histologic progression is not dominated by genetic mutations (14). Further, we explored the relationship between L2SITH grouping and genomic mutations. The samples within each histologic subtype were split into high- and low-scored subgroups, and the mutations of each subgroup were shown in Figures S8–S10. Although our L2SITH model cannot distinguish the mutations, we found that, as the sample size increased, the number of genes with significant differences between high- and low-scored subgroups increased.

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

MJ and HL conceived the main idea and the framework of the manuscript. MJ performed the experiments. XL and HL helped to improve the idea. MJ drafted the manuscript and HL

revised the manuscript. HL supervised the study and provided funding. All authors read and commented on the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China under grant no. 62072058.

Conflict of interests

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.957751/full#supplementary-material

References

1. Nasim F, Sabath BF, Eapen GA. Lung cancer. *Med Clinics North America* (2019) 103:463–73. doi: 10.1016/j.mcna.2018.12.006

2. Justilien V, Ali SA, Jamieson L, Yin N, Cox AD, Der CJ, et al. Ect2-dependent rrna synthesis is required for kras-trp53-driven lung adenocarcinoma. *Cancer Cell* (2017) 31:256–69. doi: 10.1016/j.ccell.2016.12.010

3. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JH, Beasley MB, et al. The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10:1243–60. doi: 10.1097/JTO.0000000000000630

4. Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N, Werb Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat Cell Biol* (2018) 20:1349–60. doi: 10.1038/s41556-018-0236-7

5. Bremnes RM, Al-Shibli K, Donnem T, Sirera R, Al-Saad S, Andersen S, et al. The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer. *J Thorac Oncol* (2011) 6:824–33. doi: 10.1097/JTO.0b013e3182037b76

6. Khoury T, Nagrale V, Opyrchal M, Peng X, Wang D, Yao S. Prognostic significance of stromal versus intratumoral infiltrating lymphocytes in different subtypes of breast cancer treated with cytotoxic neoadjuvant chemotherapy. *Appl Immunohistochem Mol Morphol: AIMM* (2018) 26:523. doi: 10.1097/PAI.0000000000000466

7. Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* (2018) 19:1–14. doi: 10.1186/s13059-018-1593-z

8. Sun XX, Yu Q. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol Sin* (2015) 36:1219–27. doi: 10.1038/aps.2015.92

9. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* (2018) 15:81–94. doi: 10.1038/nrclinonc.2017.166

10. Lim ZF, Ma PC. Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *J Hematol Oncol* (2019) 12:134. doi: 10.1186/s13045-019-0818-2

11. Kalbasi A, Ribas A. Tumour-intrinsic resistance to immune checkpoint blockade. *Nat Rev Immunol* (2020) 20:25–39. doi: 10.1038/s41577-019-0218-4

12. Zhou M, Zhang Z, Bao S, Hou P, Yan C, Su J, et al. Computational recognition of lncrna signature of tumor-infiltrating b lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. *Briefings Bioinf* (2020) 22(3):bbaa047. doi: 10.1093/bib/bbaa047.

13. Sun J, Yan C, Xu D, Zhang Z, Li K, Li X, et al. Immuno-genomic characterisation of high-grade serous ovarian cancer reveals immune evasion mechanisms and identifies an immunological subtype with a favourable prognosis and improved therapeutic efficacy. *Br J Cancer* (2022) 126:1570–80. doi: 10.1038/s41416-021-01692-4

14. Tavernari D, Battistello E, Dheilly E, Petruzzella AS, Mina M, Sordet-Dessimoz J, et al. Nongenetic evolution drives lung adenocarcinoma spatial heterogeneity and progression. *Cancer Discovery* (2021) 11:1490–507. doi: 10.1158/2159-8290.CD-20-1274

15. Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, et al. A potential signature of eight long non-coding rnas predicts survival in patients with non-small cell lung cancer. *J Transl Med* (2015) 13:231. doi: 10.1186/s12967-015-0556-3

16. Marino FZ, Bianco R, Accardo M, Ronchi A, Cozzolino I, Morgillo F, et al. Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications. *Int J Med Sci* (2019) 16:981. doi: 10.7150/ijms.34739

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

18. Caso R, Sanchez-Vega F, Tan KS, Mastrogiacomo B, Zhou J, Jones GD, et al. The underlying tumor genomics of predominant histologic subtypes in lung adenocarcinoma. *J Thorac Oncol* (2020) 15:1844–56. doi: 10.1016/j.jtho.2020.08.005

19. Collisson EA, Campbell JD, Sougnez C, Campbell JD, Wilkerson MD, Danilova L, et al. Never-smoker NEs. comprehensive molecular profiling of lung adenocarcinoma. *Nature* (2014) 511:543–50. doi: 10.1038/nature13385

20. Zabeck H, Dienemann H, Hoffmann H, Pfannschmidt J, Warth A, Schnabel PA, et al. Molecular signatures in iaslc/ats/ers classified growth patterns of lung adenocarcinoma. *PLoS One* (2018) 13:e0206132. doi: 10.1371/journal.pone.0206132

21. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Yang THO, et al. The immune landscape of cancer. *Immunity* (2018) 48:812–30. doi: 10.1016/j.immuni.2018.03.023

22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol* (2014) 15:1–21. doi: 10.1186/s13059-014-0550-8

23. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: an r package for comparing biological themes among gene clusters. *Omic: J Integr Biol* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

24. Walter W, Sánchez-Cabo F, Ricote M. Goplot: an r package for visually combining expression data with functional analysis. *Bioinformatics* (2015) 31:2912–4. doi: 10.1093/bioinformatics/btv300

25. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016) 32:2847–9. doi: 10.1093/bioinformatics/btw313

26. van der Maaten L, Hinton G. Visualizing data using t-sne. *J Mach Learn Res* (2008) 9:2579–605.

27. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv Prepr arXiv:1802.03426* (2018) 3 (29):861. doi: 10.21105/joss.00861

28. Wickham H. Ggplot2. *WIREs Comput. Stat. 3.* (2011) 2:180–5. doi: 10.1002/wics.147

29. Günther F, Fritsch S. neuralnet: Training of neural networks. *The R Journal* (2020) 2(1):30–8. doi: 10.32614/RJ-2010-006

30. Duan N. Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* (1983) 78:605–10. doi: 10.1080/01621459.1983.10478017

31. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corrplot'. *Statistician* (2017) 56:e24.

32. Kassambara A, Kosinski M, Biecek P, Fabian S. *Survminer: Drawing survival curves using 'ggplot2'. r package version 0.3 1*. (2017).

33. Paul G, Nancy C, Huang RS, Barbour JD. Prrophetic: An r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* (2014) 9:e107468–e107468. doi: 10.1371/journal.pone.0107468

34. Ding XF, Chen J, Ma HL, Liang Y, Wang YF, Zhang HT, et al. Kir2dl4 promotes the proliferation of rcc cell associated with pi3k/akt signaling activation. *Life Sci* (2022) 293:120320. doi: 10.1016/j.lfs.2022.120320

35. Dai W, Feng J, Hu X, Chen Y, Gu Q, Gong W, et al. Slc7a7 is a prognostic biomarker correlated with immune infiltrates in non-small cell lung cancer. *Cancer Cell Int* (2021) 21:1–14. doi: 10.1186/s12935-021-01781-7

36. Nobili S, Lavacchi D, Perrone G, Vicini G, Tassi R, Landini I, et al. Vinorelbine in non-small cell lung cancer: Real-world data from a single-institution experience. *Oncol Res* (2020) 28:237–48. doi: 10.3727/096504019X15755437099308

37. Toschi L, Finocchiaro G, Bartolini S, Gioia V, Cappuzzo F. Role of gemcitabine in cancer therapy. *Future Oncol* (2005) 1:7–17. doi: 10.1517/14796694.1.1.7

38. Zhang X, Wang D, Li Z, Jiao D, Jin L, Cong J, et al. Low-dose gemcitabine treatment enhances immunogenicity and natural killer cell-driven tumor immunity in lung cancer. *Front Immunol* (2020) 11:331. doi: 10.3389/fimmu.2020.00331

39. Kobayashi K, Ratain MJ. (1994). Pharmacodynamics and long-term toxicity of etoposide. *Cancer Chemother Pharmacol*, 34 (Suppl):S64–S68. doi: 10.1007/BF00684866

40. Wang Z, Mai S, Lv P, Xu L, Wang Y. Etoposide plus cisplatin chemotherapy improves the efficacy and safety of small cell lung cancer. *Am J Trans Res* (2021) 13:12825–33.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fimmu.2022.970950

Check for updates

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Lixin Cheng,
Jinan University, China
Yubao Lu,
Third Affiliated Hospital of Sun Yat-sen University, China
Xiang Pan,
Jiangnan University, China

*CORRESPONDENCE
Baoqing Sun
sunbaoqing@vip.163.com
Chen Huang
chuang@must.edu.mo

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 16 June 2022
ACCEPTED 22 July 2022
PUBLISHED 16 August 2022

CITATION
Cheng T, Wu Y, Liu Z, Yu Y, Sun S, Guo M, Sun B and Huang C (2022) CDKN2A-mediated nolecular subtypes characterize the hallmarks of tumor microenvironment and guide precision medicine in triple-negative breast cancer.
Front. Immunol. 13:970950.
doi: 10.3389/fimmu.2022.970950

COPYRIGHT
© 2022 Cheng, Wu, Liu, Yu, Sun, Guo, Sun 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.

CDKN2A-mediated molecular subtypes characterize the hallmarks of tumor microenvironment and guide precision medicine in triple-negative breast cancer

Tianyi Cheng[1†], Yingyi Wu[1,2†], Zhiyu Liu[1], Yi Yu[1], Shixue Sun[3], Min Guo[4], Baoqing Sun[5*] and Chen Huang[1,2*]

[1]Faculty of Chinese Medicine, State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macao, Macao SAR, China, [2]Dr. Neher's Biophysics Laboratory for Innovative Drug Discovery, Macau University of Science and Technology, Macao, Macao SAR, China, [3]Guangzhou Municipal and Guangdong Provincial Key Laboratory of Molecular Target and Clinical Pharmacology, The NMPA and State Key Laboratory of Respiratory Disease, School of Pharmaceutical Sciences and the Fifth Affiliated Hospital, Guangzhou Medical University, Guangzhou, China, [4]Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China, [5]Department of Allergy and Clinical Immunology, Department of Laboratory, National Center for Respiratory Medicine, National Clinical Research Center for Respiratory Disease, State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Health, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

Currently, breast cancer (BRCA) has become the most common cancer in the world, whose pathological mechanism is complex. Among its subtypes, triple-negative breast cancer (TNBC) has the worst prognosis. With the increasing number of diagnosed TNBC patients, the urgent need of novel biomarkers is also rising. Cyclin-dependent kinase inhibitor 2A (CDKN2A) has recently emerged as a key regulator associated with ferroptosis and cuproptosis (FAC) and has exhibited a significant effect on BRCA, but its detailed mechanism remains elusive. Herein, we conducted the first converge comprehensive landscape analysis of FAC-related gene CDKN2A in BRCA and disclosed its prognostic value in BRCA. Then, an unsupervised cluster analysis based on CDKN2A-correlated genes unveiled three subtypes, namely cold-immune subtype, IFN-γ activated subtype and FTL-dominant subtype. Subsequent analyses depicting hallmarks of tumor microenvironment (TME) among three subtypes suggested strong association between TNBC and CDKN2A. Given the fact that the most clinically heterogeneous TNBC always displayed the most severe outcomes and lacked relevant drug targets, we further explored the potential of immunotherapy for TNBC by interfering CDKN2A and constructed the CDKN2A-derived prognostic model for TNBC patients by Lasso-Cox. The 21-gene–based prognostic model showed high accuracy and was verified in external independent validation cohort. Moreover, we proposed three drugs for TNBC patients based on our model *via* targeting epidermal growth factor

receptor. In summary, our study indicated the potential of CDKN2A as a pioneering prognostic predictor for TNBC and provided a rationale of immunotherapy for TNBC, and offered fresh perspectives and orientations for cancer treatment *via* inducing ferroptosis and cuproptosis to develop novel anti-cancer treatment strategies.

KEYWORDS

cuproptosis, immunotherapy, tumor microenvironment, triple-negative breast cancer, cyclin-dependent kinase inhibitor 2A

Introduction

Nowadays, breast cancer (BRCA) is the worldwide leading cause of cancer incidences and the second leading cause of cancer-related death (1). Triple-negative breast cancer (TNBC), which is distinguished by the absent expression of human epidermal growth factor receptor 2 (Her2) and estrogen receptor/progesterone receptor (ER/PR), is the most invasive subtype with the highest mortality rate accounting for approximately 15% of all BRCA (2). The mortality rate is up to 40% within 5 years after the first diagnosis and distant metastasis will occur in approximately 46% of TNBC patients (3). Hence, the incidence and mortality of TNBC make it necessary to explore reliable predictive biomarkers, construct more promising prognostic models and develop novel drugs that target at the known molecular pathways.

Cyclin-dependent kinase inhibitor 2A (CDKN2A), a cyclin-dependent kinase inhibitor gene, that encodes the p16 protein involved in the regulation of cell cycle pathways, is known as a tumor suppressor (4). CDKN2A can inactivate the retinoblastoma protein by binding to and inactivating the cyclin D-cyclin-dependent kinase 4 complex (5). The expression of this gene is verified to cause cell cycle arrested in the G1 phase, inhibit cell proliferation, promote tumor cell apoptosis, and increase tumor cell chemotherapy sensitivity (6). Recent studies have pointed out that CDKN2A is correlated with ferroptosis (7) and cuproptosis (8) (FAC), which are both novel types of regulated cell death that their occurrence was ion-dependent. Ferroptosis indicates an oxidative cell death resulting from the deterioration of antioxidant function and accretion of lipid reactive oxygen species (ROS) (9). Recent attention has been brought to a brand-new cell death mode identified as cuproptosis, which indicates that the excess copper can trigger proteotoxic stress and death in cells through the combination with lipoylated components of the tricarboxylic acid (TCA) cycle (8). These ion-dependent cell death different from apoptosis, necrosis, and autophagy can contribute to a burgeoning field that promising cancer drugs are designed based on the induction of ferroptosis (10) and cuproptosis (8).

Additionally, researchers have discovered that the malfunctioning of CDKN2A in BRCA has promoted the discovery of many CDK inhibitors (11). The role of CDKN2A in BRCA cannot be ignored and needs further investigations. However, the investigations about the role of CDKN2A in BRCA are limited, we are thereby unable to comprehensively elaborate the biological function of CDKN2A. Hence, we conducted the first converge comprehensive landscape analysis of FAC-related gene CDKN2A in BRCA, including expression, prognostic values, DNA methylation, tumor microenvironment (TME) analysis, and drug sensitivity of CDKN2A in BRCA. Immediately afterward, unsupervised cluster analysis revealed the difference in immunological analysis and FAC status of CDKN2A-associated genes among 3 groups, namely cold-immune subtype, IFN-γ activated subtype, and FTL-dominant subtype, groundbreakingly laying a foundation for the application of immunotherapy and FAC regulators in BRCA. Given the strong association between TNBC and CDKN2A, as well as the fact that the most clinically heterogeneous TNBC always displayed the most severe outcomes and lacked relevant drug targets, we further explored the potential of immunotherapy for TNBC by regulating CDKN2A and constructed the CDKN2A-derived prognostic model for TNBC patients by machine learning, aiming to predict the prognosis of TNBC patients and provide the guidance on their long-term disease outlook and design of treatment strategies.

Materials and methods

Data collection

BRCA data was downloaded from the UCSC Xena data mining platform (http://xena.ucsc.edu/), which included the messenger RNA (mRNA) expression matrix from The Cancer Genome Atlas (TCGA), as well as the clinical information of 33 cancer types. Gene expression profile of BRCA patients and their corresponding clinical information were obtained. Specimens without survival information were excluded during this study and FPKM values of RNA-Seq were log2 transformed. In total,

1072 BRCA patients and included 185 TNBC samples were retained for subsequent analysis.

We extracted and exhibited detailed clinical information of 1072 BRCA patients as shown in Table 1: age, sex, pathological stage, estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor 2 (HER2) status, T/N/M stage, and adjuvant chemotherapies. Additionally, gene expression array GSE58812 containing 107 TNBC patients and their corresponding clinical data was retrieved from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) to externally validate the CDKN2A-derived prognostic model constructed in our study. Gene expression array GSE173839 containing clinical information of 100 BRCA patients was used to evaluate the status of immunotherapy response between high and low CDKN2A expression groups in TNBC patients.

TABLE 1 Clinical pathological characteristics of extracted BRCA patients.

Characteristic	Group	No. of cases (%)
Age (years)	<60	570 (53.17%)
	≥60	500 (46.64%)
	Unknown	2 (0.18%)
Sex	Female	1059 (98.78%)
	Male	12 (1.12%)
	Unknown	1 (0.09%)
Pathological Stage	Stage I	176 (16.42%)
	Stage II	607 (56.62%)
	Stage III	245 (22.85%)
	Stage IV	20 (1.86%)
	Stage X	12 (1.12%)
	Unknown	12 (1.12%)
Pathological T	T1	274 (25.56%)
	T2	621 (57.93%)
	T3	133 (12.40%)
	T4	40 (3.73%)
	TX	3 (0.27%)
	Unknown	1 (0.09%)
Pathological N	N0	502 (46.83%)
	N1	355 (33.11%)
	N2	118 (11.01%)
	N3	76 (7.09%)
	NX	20 (1.86%)
	Unknown	1 (0.09%)
Metastasis	M0	896 (83.58%)
	M1	22 (2.05%)
	MX	153 (14.27%)
	Unknown	1 (0.09%)
ER	Positive	789 (73.60%)
	Negative	232 (21.64%)
	Unknown	51 (4.75%)
PR	Positive	684 (63.80%)
	Negative	334 (31.15%)
	Unknown	54 (5.04%)
HER2	Positive	162 (15.11%)
	Negative	547 (51.02%)
	Unknown	363 (33.86%)
Adjuvant therapy	No	1056 (98.51%)
	Yes	13 (1.21%)
	Unknown	3 (0.27%)
OS Status	Living	921 (85.91%)
	Dead	150 (13.99%)
	Unknown	1 (0.09%)

Landscape analysis of CDKN2A in BRCA

To comprehensively investigate the biological role of CDKN2A in BRCA, we started with the pan-cancer analysis of the CDKN2A expressions *via* Tumor Immune Estimation Resource (TIMER) (https://cistrome.shinyapps.io/timer/) database (12). Then, the expression levels of CDKN2A in BRCA, as well as normal tissues were validated in Gene expression profiling interactive analysis (GEPIA) (http://gepia.cancer-pku.cn/) sequencing expression (13). Subsequently, Human Protein Atlas (HPA) (https://www.proteinatlas.org/) provided the immunohistochemical images of CDKN2A expression in BRCA samples (14). To further explore the biological role of CDKN2A in BRCA, UALCAN (http://ualcan.path.uab.edu/index.html) (15) and MEXPRESS (http://mexpress.be) (16) were utilized to analyze the DNA promoter methylation status of CDKN2A in BRCA and the effects of methylation of CDKN2A on clinical stages in BRCA. Besides, cBio Cancer Genomics Portal (cBioPortal) (http://cbioportal.org/) and Catalogue of Somatic Mutations In Cancer (COSMIC) (https://cancer.sanger.ac.uk) were conducted to analyze the mutation status of CDKN2A in BRCA (17, 18). TIMER and Tumor and Immune System Interaction Database (TISIDB) (http://cis.hku.hk/TISIDB) database (19) were used to comprehensively explore the relationship between CDKN2A expression and immune infiltration. Moreover, CellMiner (20) was performed to evaluate the relationship between CDKN2A and drug sensitivity, looking for the targeted therapies for BRCA patients.

Unsupervised clustering of CDKN2A-associated differentially expressed genes

Based on the integration of CDKN2A strongly associated genes and differential genes, we used the R package "ConsensusClusterPlus" (https://bioconductor.org/packages/ConsensusClusterPlus/), to perform an unsupervised cluster analysis. After 1,000 iterations of the consensus clustering algorithm, the number of optimal clusters was confirmed according to the Item-Consensus plot, cumulative distribution

function curves, and the k-means clustering algorithm. Three unsupervised clusters (cold-immune cluster, IFN-γ activated cluster, and FTL-dominant cluster) were selected for subsequent analysis.

Profiling analysis of tumor microenvironment

As a generic computational method of calculating cell fractions from gene expression data, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) (21), was separately used to analyze the proportions of 22 infiltrating immune cells in high and low CDKN2A expression groups in BRCA, as well as between 3 groups from the unsupervised cluster. The ESTIMATE approach calculated the stromal, immune, and ESTIMATE scores of three groups, predicting the level of stromal cells and infiltrating immune, which constructed the basis of tumor purity (22). The immunotherapy-related pathways, immune checkpoints, and 122 immunomodulators, including chemokines, receptors, MHCs, and immune stimulators were obtained from previous studies (23–25). Single-Sample Gene Set Enrichment Analysis (ssGSEA) was performed to derive the enrichment score of all steps *via* the R package "GSVA" (26).

Evaluation of immunotherapy response sensitivity in TNBC patients with different CDKN2A expression

Immunophenoscore (IPS) is a generic machine learning-based algorithm for quantifying tumor immunogenicity, which is measured grounded in the gene expression of representative cell types, including immunomodulators, immunosuppressive cells, MHC molecules, and effector cells (27). The IPS from The Cancer Immunome Atlas (TCIA) (https://tcia.at/) was calculated in the high and low CDKN2A expression groups in TNBC patients. In general, the higher IPS indicates a better immunotherapy response. Subsequently, to further predict the sensitivity to immunotherapy response, GSE173839 was used to evaluate the status of immunotherapy response between high and low CDKN2A expression subpopulations in TNBC patients.

Exploration of functional annotation of CDKN2A-associated genes

Gene Ontology (GO) and Kyoto Encyclopedia of Genes, Genomes (KEGG) functional enrichment analysis confirmed the functions of CDKN2A-associated genes *via* the R language "cluster Profiler" package (https://guangchuangyu.github.io/software/clusterProfiler/). Additionally, gene sets "h.all.v7.5.1.symbols.gmt" were obtained from Molecular Signatures Database (MSigDB) (https://software.broadinstitute.org/gsea/downloads.jsp) and were used for calculations of 50 hallmark tumor-related pathways. Moreover, oxidative stress caused by the accumulation of lethal ROS is the recognized process of ferroptosis. But, due to the definition of cuproptosis being relatively avant-garde, the controversy of whether cuproptosis is a form of cell death independent of other cell death modes or not still exists. Therefore, based on the GO enrichment analysis and previous studies, we collected 7 pathways implicated in ferroptosis and cuproptosis *via* literature retrieval and vicariously evaluated their activities in BRCA by ssGSEA analysis, including fatty acids degradation (28), inflammatory response (29), oxidative stress (9), positive regulation of MAPK cascade (30), regulation of mitochondrial membrane potential (9), TCA cycle (8) and VEGF signaling pathway (31). Moreover, Metascape (32) was used to analyze the functional annotation of 413 survival-related differentially expressed genes (SDEGs), aiming to reveal the biological mechanism of the influence of CDKN2A on the survival status of TNBC patients. Besides, Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/) database (33) was utilized to gather and construct data about protein-protein interaction (PPI) of CDKN2A and genes constructing the model.

Weighted gene co-expression network analysis based on RNA-seq data

After deleting the outliers in the gene expression matrix, the TCGA-A7-A0DC-01A sample and the TCGA-A2-A3XV-01A sample were expelled. Based on a scale-free topology with $R2 = 0.85$, the adjacency matrix was defined by using soft thresholding with power $\beta = 6$, to identify and build the different co-expression gene modules in BRCA samples. Then, the CDKN2A-derived genes were clustered based on a topological overlap matrix (TOM)-based dissimilarity measure, and the cluster dendrogram of all these genes was constructed by R package "WGCNA" (34). Every identified co-expression module was labeled with a different color. Then, we conducted principal component analysis (PCA) of each module, extracted and summarized the gene co-expression based on the eigengene external traits that included TNBC and the status of ferroptosis and cuproptosis (substituted by the scores of oxidative stress, regulation of mitochondrial membrane potential and TCA cycle). We further calculated the correlation between each eigengene external trait and each module with the biweight midcorrelation (bicor) that could offer robust correlations with minor weight given to outlier measures (35, 36). Subsequently, we selected genes in modules that possessed the strongest relationship with TNBC and FAC as the input for the Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis.

Construction and validation of the CDKN2A-derived prediction model

Based on the genes from WGCNA, we sequentially developed univariate Cox, LASSO regression *via* the R package "glmnet" to construct the CDKN2A-derived prognostic model (37). The risk score was calculated *via* the following formula:

$$Risk\ \ Score\ = \sum_{i=1}^{n} Coef_i \times exp_i$$

The $Coef_i$ represented the risk coefficients of each gene weighted by LASSO-Cox model, and exp_i indicated the expression of each gene in our study. Then, the Kaplan–Meier survival analysis was developed to evaluate the difference in survival between low and high risk-score groups through R package "survival". Subsequently, we used the time-dependent receiver operating characteristic (ROC) curve to appraise the performance of the CDKN2A-derived model. Further, to test whether risk score could be an independent prognostic predictor of TNBC patients, univariate Cox and multivariate Cox regression analyses were conducted with risk score, sex, age, metastasis status, tumor stage and pathological status as variables. Ultimately, external validation of the CDKN2A-derived prognostic model was performed *via* the clinical data of 107 TNBC patients contained in GSE58812.

Potential drug prediction based on the CDKN2A-derived model

Drug sensitivity data of diverse cell lines and corresponding gene-expression data from three databases were used to perform the drug sensitivity analysis based on the CDKN2A-derived signature, including GDSC (Genomics of Drug Sensitivity in Cancer), PRISM (Profiling Relative Inhibition Simultaneously in Mixtures) and CTRP (Cancer Therapeutics Response Portal) (38) (39). AUC values functioned as a measure of drug sensitivity and drugs with missing AUC values more than 80% were excluded. Based on the different drug reactions of high and low groups, drugs with Padj value less than 0.05 were screened out. The compound overlapping in the outcomes of PRISM, CTRP, and GDSC analyses may serve as a potential treatment for the certain subpopulation.

Statistical analysis

Correlations were analyzed *via* Pearson correlation except for the part of WGCNA using bicor. Statistical analyses were conducted using Kruskal–Wallis, Wilcoxon, chi-square test, and Tuckey's honestly significant difference and differences were considered significant at P value < 0.05.

Results

Landscape analysis of CDKN2A hints at prognostic value and its association with drug sensitivity in BRCA

Figure 1 illustrates the flow chart of the present study. In the comparisons of multiple cancers with corresponding normal tissues, CDKN2A exhibited a significant overexpression in bladder urothelial carcinoma (BLCA), BRCA, cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), head and neck cancer (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC) (Figure 2A). GEPIA database was used to further confirm that CDKN2A was notably upregulated in BRCA (Figure 2B). Immunohistochemistry outcomes from the HPA database illustrated that the protein level of CDKN2A was significantly increased in BRCA tissue (Figure 2C). As a fundamental constituent element of epigenetics, DNA methylation modification plays a vital role in silencing the expressions of methylated genes. Our data indicated that CDKN2A was notably hypermethylated in BRCA (Figure 2D), especially in the luminal subtype and TNBC subtype (Supplementary Figure S1C). BRCA patients with CDKN2A hypermethylation possessed a relatively undesirable clinical outcome (P = 0.0527833983) (Supplementary Figure S1D), which needs further investigations. Additionally, the hypermethylation of CDKN2A was positively correlated with the tumor progression, verifying that CDKN2A may be a crucial impact factor in the stage and grade of BRCA (Supplementary Figures S1A, B). Besides, the COSMIC database was conducted to evaluate the mutation type of CDKN2A. (Supplementary Figures S2C, D). Missense substitutions notably occupied the largest portion, accounting for 38.59%. Next, nonsense substitutions occurred in 30.54%. Frameshift deletions occupied 9.81% of the samples and frameshift insertions occupied 6.40% of the samples. Moreover, our results indicated that the substitution mutations chiefly occurred at C > T (44.24%), G > A (20.45%), G > T (14.91%), and C > A (5.5%). Additionally, cBioPortal database revealed that the mutation frequency of CDKN2A in BRCA was 0.5% and the mutation of CDKN2A had no effects on the prognosis of BRCA patients (P > 0.05) (Supplementary Figures S2A, B).

To further explore the molecular characteristics of CDKN2A in BRCA, we performed the relationship between CDKN2A and tumor-infiltrating immune cells using the CIBERSORT

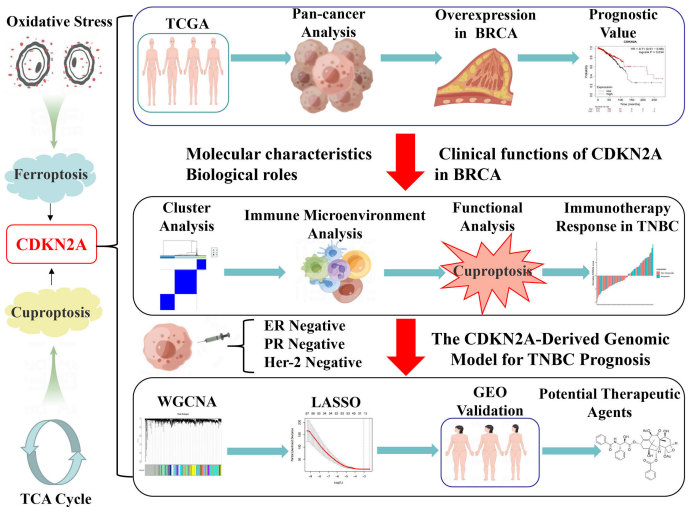


FIGURE 1
The flow chart of our study.

algorithm. Our results demonstrated that BRCA patients with high expression of CDKN2A exhibited an increased infiltration level of most immune cells in BRCA, including activated dendritic cells, M0 macrophages, activated NK cells, activated memory CD4 + T cells, CD8 + T cells, follicular helper T cells, and regulatory T cells (Figure 2H). Moreover, the expression of CDKN2A was positively associated the expression of multiple immune checkpoints, including CTLA4, PDCD1, PVR, TIGIT, and so on in BRCA (Figure 2I). Our data from Timer 2.0 and TISIDB database also indicated that CDKN2A expression could significantly affect the immune infiltration status and immune microenvironment of BRCA. The expression of CDKN2A was respectively correlated with the infiltration abundances of macrophages in basal-like BRCA, myeloid dendritic cells and CD8+ T cells in luminal A (Supplementary Figure S3A). Besides, CDKN2A expression was also correlated with CCL5, CCL7, CCL8, CXCL16, etc (Supplementary Figure S3B).

Survival analysis was further indicative that BRCA patients with high expression of CDKN2A had better overall survival (OS) than those with low CDKN2A expression (Figure 2E). The prognostic values of CDKN2A expression in different subtypes of BRCA also presented a significant differentiation (Figure 2F). The genetic alterations caused by the heterogeneity of BRCA may also affect the responses to target agents (40). Improved reliable biomarkers for targeted treatment are needed. Consequently, the relationship between CDKN2A expression and drug sensitivity was conducted, exploring the clinical roles of CDKN2A. Our data indicated that the expression of CDKN2A was positively connected with sensitivity to acetalax (Figure 2G). Otherwise, CDKN2A presented a negative correlation to sensitivity to mitoxantrone, O-6-Benzylguanine, bleomycin, valrubicin, and mitomycin. In conclusion, CDKN2A has the potential of being a predictive marker of the aforementioned agents.

Characterizations of CDKN2A-mediated genes reveal linkage of CDKN2A to TME and prognostic value in TNBC

To depict the crosstalk between CDKN2A and TME in BRCA, two strategies were initially proposed to detect CDKN2A-medicated genes. Concretely, 737 CDKN2A-correlated genes were obtained using Pearson correlation analysis (|corrcoef| > 0.4). Then we divided 1072 TCGA-BRCA patients into four quartiles ranked by their expression of CDKN2A and identified 228 differentially expressed genes (DEGs) between the two quartiles groups with the highest and lowest expression (|logFC| > 1, adjust P < 0.05). Combined with 737 CDKN2A-correlated genes, a total of 885 CDKN2A-mediated genes were figured out. Subsequent functional analysis indicated that these genes might involve in mitotic cell cycle phase transition, double-strand break repair *via* break-induced replication, DNA replication, positive regulation of ubiquitin protein ligase activity, etc (Figure 3A, Supplementary Figure S4A). Afterwards, an unsupervised cluster analysis was conducted based on CDKN2A-medicated genes *via* R package "ConsensusClusterPlus". As a result, 1072 BRCA patients were divided into 3 subgroups with optimal stability of the classification. Through CIRBERSORT analysis, Subgroup 1 was closely associated with an increased infiltration of resting dendritic cells, M2 Macrophages, resting mast cells, eosinophils,

10.3389/fimmu.2022.970950

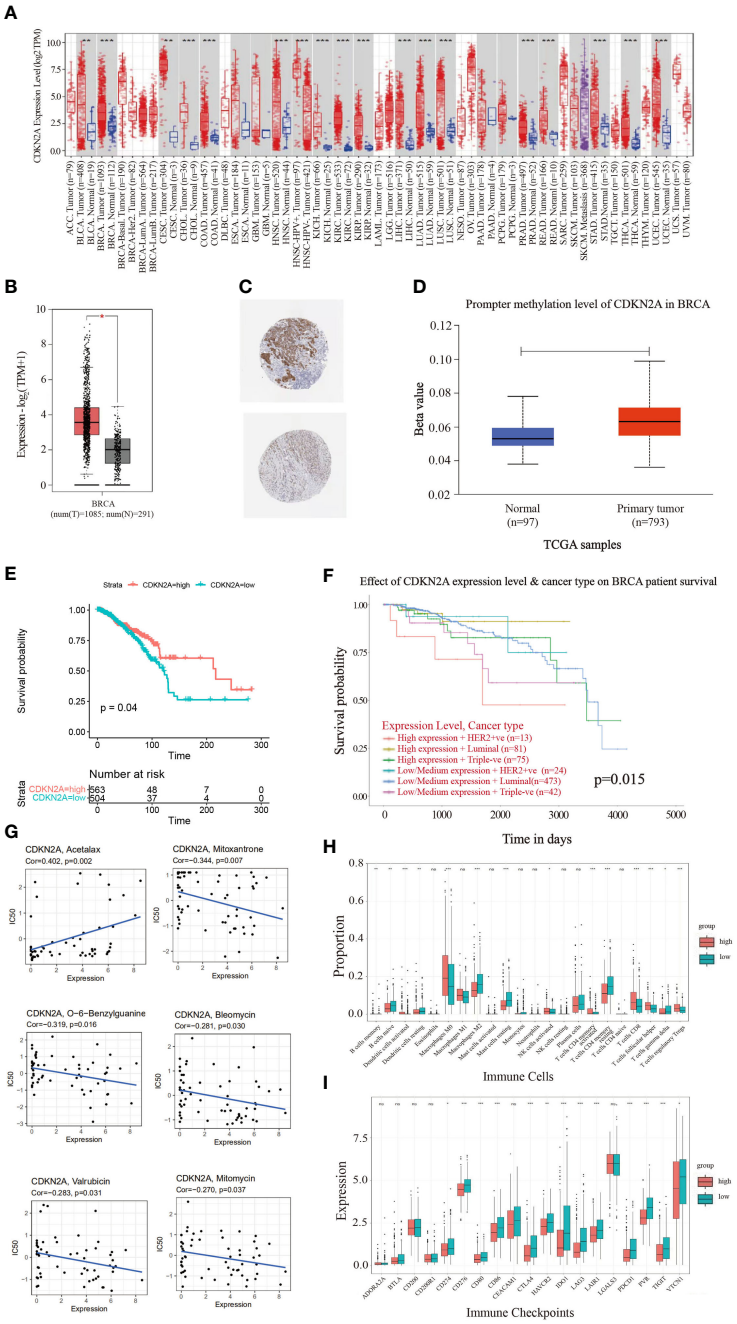


FIGURE 2
The landscape analysis of overexpressed CDKN2A in BRCA. **(A)** The difference in expression of CDKN2A between various malignant cancer types from the cancer genome map (TCGA) database across TIMER database. CDKN2A was upregulated in bladder urothelial Carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), Prostate adenocarcinoma (PRAD), Rectum adenocarcinoma (PEAD), Stomach adenocarcinoma (STAD), Thyroid carcinoma (THCA) and Uterine Corpus Endometrial Carcinoma (UCEC). (*P < 0.05. **P < 0.01. ***P < 0.001). **(B)** CDKN2A was significantly upregulated in BRCA by GEPIA database. **(C)** Representative immunohistochemical images of CDKN2A in BRCA tissues. **(D)** Promoter methylation levels of CDKN2A in normal tissues and primary BRCA tissues in the UALCAN database. **(E)** The Kaplan-Meier curves of OS for low and high expression of BRCA patients. **(F)** The prognostic values of CDKN2A in different BRCA subtypes. **(G)** Scatter plots depict the relationship between CDKN2A expression and drug sensitivity in BRCA. **(H)** Comparison of infiltration of immune cells between high and low CDKN2A expression groups in BRCA. **(I)** Comparison of immune checkpoints expression between high and low CDKN2A expression groups in BRCA. The P values of the figure are shown as follows: *P < 0.05. **P < 0.01. ***P < 0.001. ns (not significiant, P > 0.05).

monocytes, and resting memory CD4+T cells (Figure 3C), equivalent to the phenotype with immunosuppressive characteristic (41), which, thereby, was defined as cold-immune subtype. Similarly, subgroup 2 was defined as IFN-γ activated subtype due to its elevated infiltration of M0 and M1 Macrophages, activated dendritic cell, CD8 T cells, follicular helper T cells and activated memory CD4+T cells (Figure 3C), which correspond to the active-immune phenotype (42, 43). Subtype 3 was characterized by the highest expression of FTL, namely FTL-dominant subtype. Also, three subgroups were found to present the conspicuous discrepancy of expression differences of immune checkpoints genes (Figure 3D). In particular, IFN-γ activated subtype exhibited an elevated expression level of multiple immune checkpoints, including CD274, CTLA4, PDCD1, PVR, TIGIT and VTCN1, etc. Moreover, IFN-γ activated subtype correlated the relatively highest scores of certain immunotherapy-related pathways, especially in IFN-γ pathway (Supplementary Figure S4B), which was another reason for naming it. Furthermore, the heatmap depicted each BRCA patient with a corresponding enrichment of 122 immunomodulators among three groups, including chemokines, receptors, MHCs and immune stimulators (Figure 3B). As demonstrated in the chart, cold-immune subtype could be insinuated as an immunologically "cold" phenotype. Notably, IFN-γ activated subtype relatively possessed the highest immune activity. These findings were consistent with results of ESTIMATE analysis (Figure 3F).

Notably, IFN-γ activated cluster was significantly associated with numerous pathways, such as MYC targets, inflammatory response, IL6/JAK/STAT3 signaling pathway and IFN-γ response, which was consistent with our way of naming it (Figure 3E). Since CDKN2A is the ferroptosis and cuproptosis-related gene, we collected seven pathways implicated in ferroptosis and cuproptosis *via* literature retrieval and outcomes of GO enrichment analysis, aiming to vicariously evaluate their activities in patients with BRCA by ssGSEA. The results demonstrated that ssGSEA scores of those FAC-related pathways significantly differed among three subgroups (Figure 3G). Notably, the FTL-dominant subtype possessed the relatively highest scores of oxidative stress, demonstrating its elevated activity in ferroptosis.

Next, we investigated the correlation between molecular subtyping, immunological subtyping, and our unsupervised subtyping in BRCA. Unexpectedly, we found that the majority of patients (approximately 98.4%) of basal-like subtype were part of IFN-γ activated subtype (Figure 4A). More subtly, IFN-γ activated subtype chiefly belonged to the C2 subtype that was dominated by IFN-γ (Figures 4B, C). Our results also demonstrated that CDKN2A more significantly overexpressed in TNBC patients than non-TNBC patients (Supplementary Figure S5A). Notably, our analysis demonstrated the CDKN2A expression was relatively the highest in our S2 subtype (Figure 4D), which suggested the close correlation between CDKN2A and TNBC (basal-like) subtype. This point was supported by subsequent survival analysis, which indicated that TNBC patients with low expression of CDKN2A exhibited an undesirable clinical outcome (Figure 4E). Moreover, CDKN2A also exhibited four methylation sites with statistical significance among the molecular subtypes (Figure 4H). Subsequently, we aimed at exploring the underlying biological mechanism behind the survival difference between the two groups through difference analysis and function annotation analysis. We further identified 413 survival-related differentially expressed genes (SDEGs) between two groups with high and low CDKN2A expression in TNBC based on the best cut results of survival analysis ($|logfc| > 0.5$, adjust $P < 0.05$). Our results showed that there were 294 up-regulated SDEGs in high CDKN2A expression groups, which were associated with the positive regulation of transforming growth factor beta receptor signaling pathway, cellular response to metal ion, regulation of actin cytoskeleton, signaling by Rho GTPases, Miro GTPases and RHOBTb3, etc (Figure 4G). And 119 down-regulated SDEGs indicated in low CDKN2A expression group correlated with the regulation of production of molecular mediator of the immune response, mitochondrion organization and cytokine signaling in the immune system, etc (Figure 4G). The above enriched functional pathways may be the reason for the significant difference in survival between the two groups of TNBC patients. Further, we tried to explore the potential interplay between CDKN2A and TME implicated in TNBC. The IPS score was used to assess the impact of CDKN2A expression on TNBC immunity. The results showed that the low CDKN2A expression was positively correlated with the decreased IPS (Figure 4F), which indicates that low expression of CDKN2A might be unresponsible for immunotherapy, probably linking to inhibition of T cell infiltration and suppression of immunogenicity (Supplementary Figure S5B). GSE173839 further effectively verified that high expression of CDKN2A had a better immunotherapy response (Figures 4I, J). Taken together, our analyses suggested that CDKN2A might influence the progression and prognosis of TNBC and affect the effectiveness of immunotherapy in TNBC through TME, implying the potential of CDKN2A as a pioneering prognostic predictor for TNBC.

The CDKN2A-derived prognostic model by machine learning for TNBC patients

To further explore the relationship between CDKN2A, FAC, and TNBC, on the basis of TCGA-BRCA cohort, a co-expression network and modules of differentially expressed CDKN2A-derived genes were constructed *via* the WGCNA. Overall, the brown and green module had the strongest correlation with TNBC *via* the Kruskal-Wallis test and Tuckey's honestly significant difference, and simultaneously possessed the most

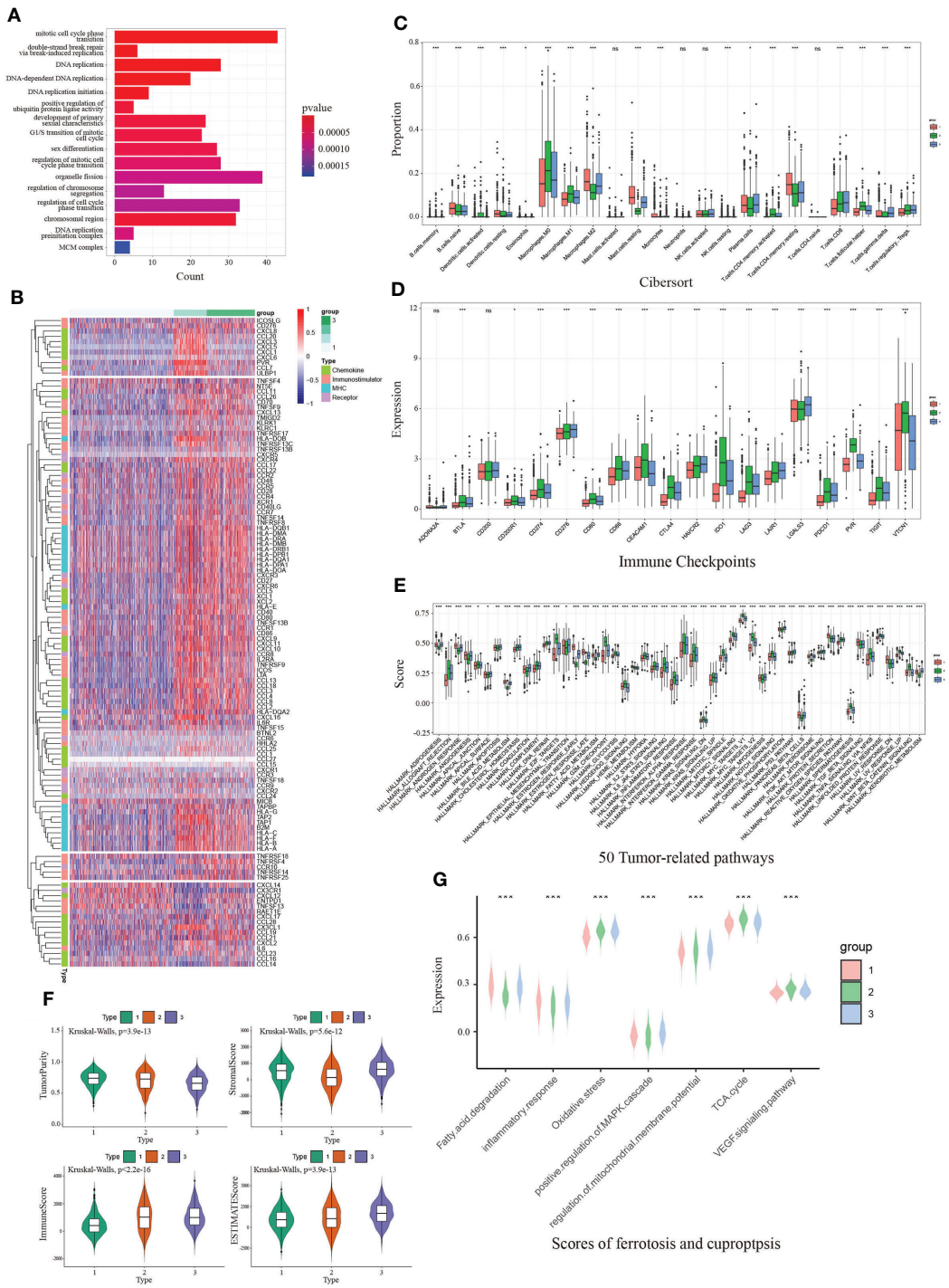


FIGURE 3
The immunological and functional analysis of CDKN2A among 3 groups from unsupervised clustering in BRCA. **(A)** The GO enrichment analysis revealed the function of CDKN2A-mediated genes. **(B)** The heatmap depicted each BRCA patient with a difference of a corresponding enrichment of 122 immunomodulators. **(C)** Comparison of infiltration of immune cells between 3 groups. **(D)** Comparison of immune checkpoints expression between 3 groups in BRCA. **(E)** Comparison of 50 tumor-related pathways between 3 groups in BRCA. **(F)** Comparison of estimate score, immune score, stromal score, and tumor purity between 3 groups in BRCA. **(G)** Comparison of scores of ferroptosis and cuproptosis between 3 groups in BRCA. The P values of the figure are shown as follows: *P < 0.05. **P < 0.01. ***P < 0.001. ns (not significiant, P > 0.05).

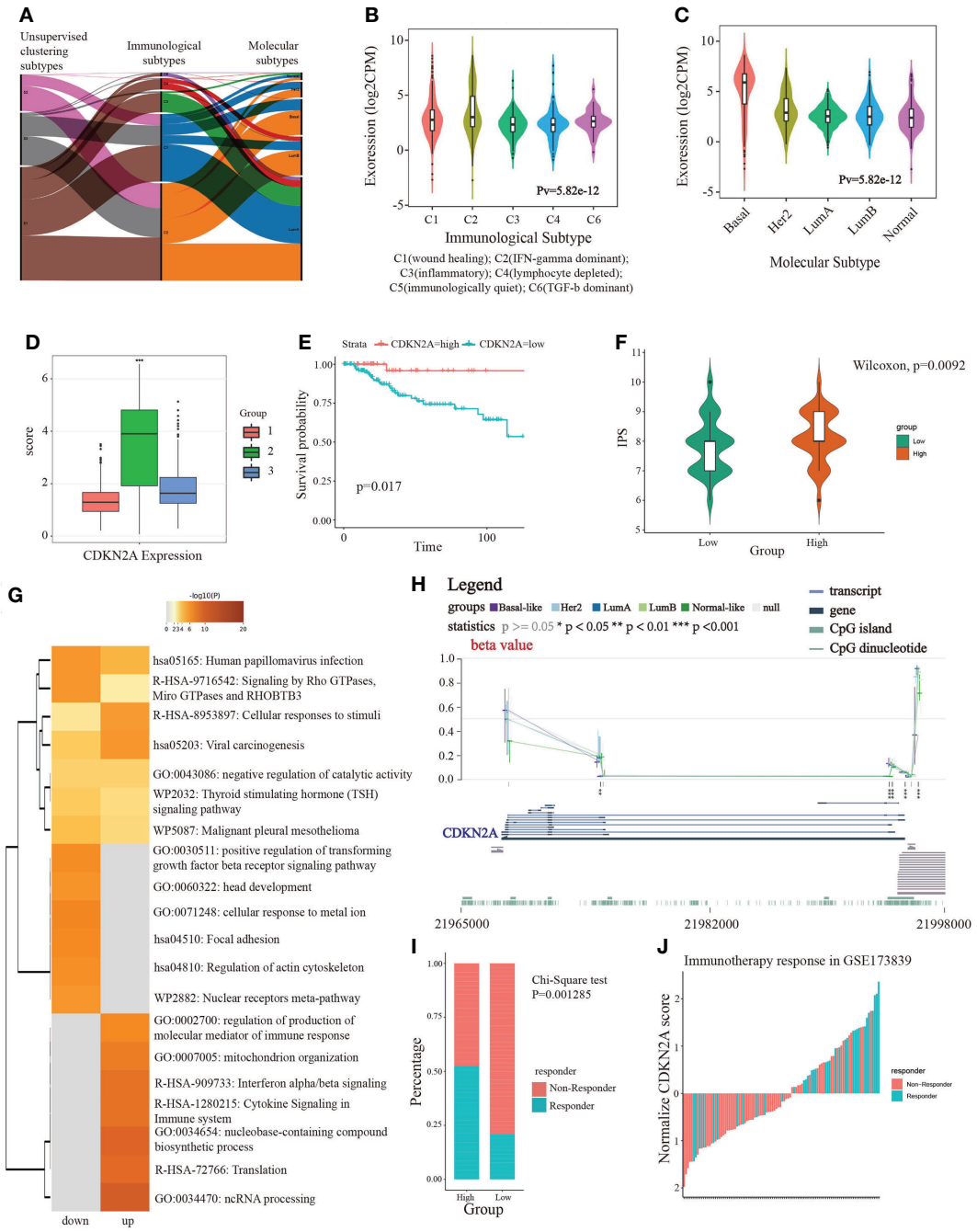


FIGURE 4
The linkage of CDKN2A to immunotherapy and TNBC. **(A)** The correlation between molecular subtypes, immunological subtypes, and our unsupervised subtypes in BRCA. **(B)** The relationship between CDKN2A expression and immunological subtypes of BRCA. **(C)** The relationship between CDKN2A expression and molecular subtypes of BRCA. **(D)** The comparison between CDKN2A expression and unsupervised subtypes of BRCA. **(E)** The survival value of CDKN2A in TNBC. **(F)** The comparison between IPS score and high and low CDKN2A expression subpopulations in TNBC. **(G)** The function annotation analysis of up-regulated and down-regulated SDEGs in high and low CDKN2A expression subpopulations. **(H)** The comparison between methylation status of CDKN2A and molecular subtypes of BRCA. **(I, J)** The correlation between immunotherapy response status and CDKN2A expression in TNBC *via* chi-square test. The P values of the figure are shown as follows: *P < 0.05. **P < 0.01. ***P < 0.001.

outstanding connection with the FAC activity (Figure 5A). A total of 1,924 CDKN2A-derived genes in these two modules were selected for further study. Subsequently, the univariate Cox regression analysis was conducted to gain 106 genes associated with prognosis (Table 2). Then, LASSO regression further screened out 21 prognostic genes for constructing the risk predictive model (Table 3, Figure 5B). On the foundation of 21 genes, the formula of risk scores is as follows:

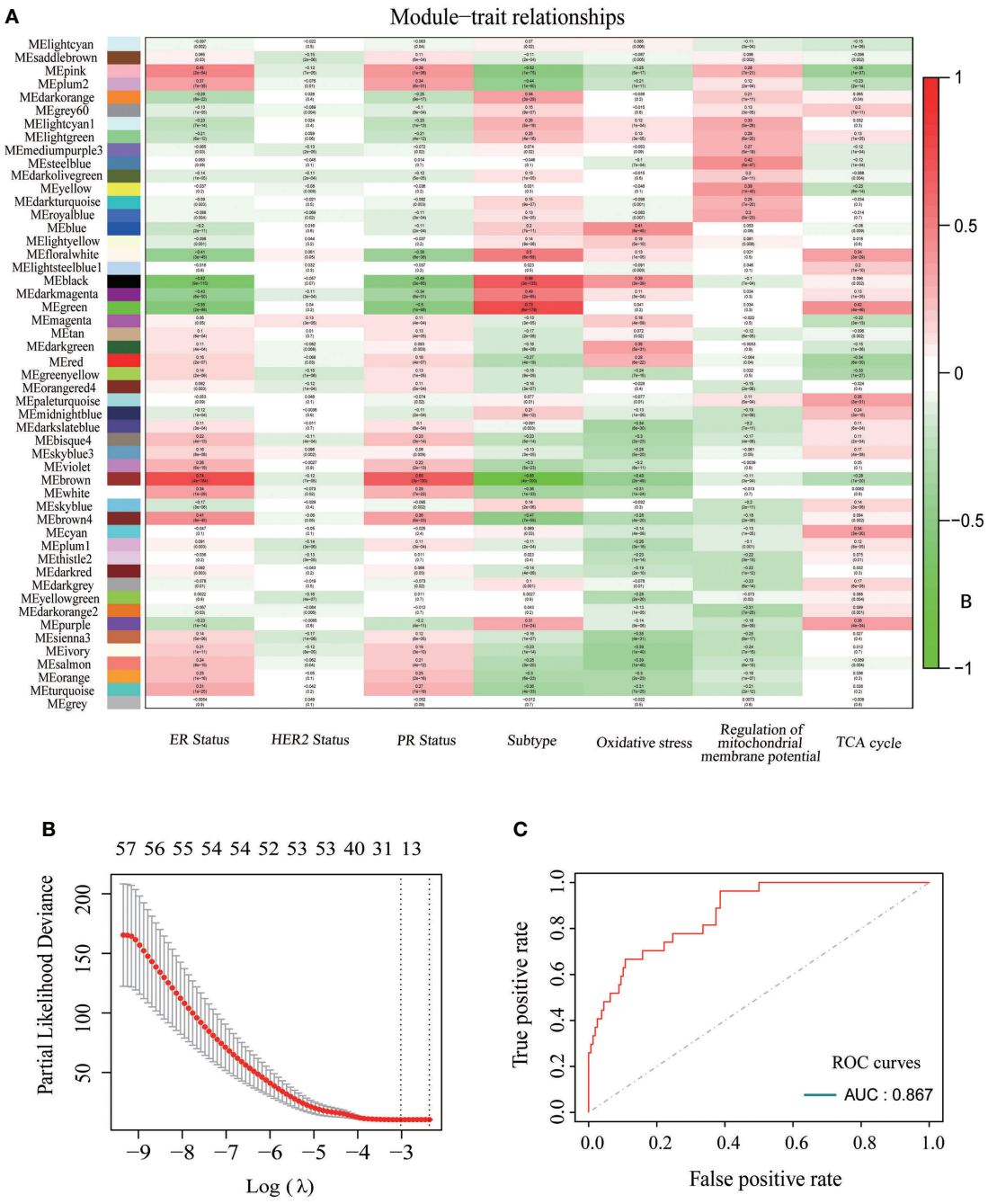


FIGURE 5
The construction of CDKN2A-derived prognostic model of TNBC. **(A)** The relationships between each module and ER status, HER status, PR status, BRCA subtypes, oxidative stress, regulation of mitochondrial membrane potential and TCA cycle. **(B)** 21 modeling genes determined by lasso algorithm. **(C)** The ROC curves and AUC value of CDKN2A-derived model.

10.3389/fimmu.2022.970950

TABLE 2 The 106 prognostic genes obtained by the univariate Cox regression analysis.

	gene	HR	z	P value
1	PIGA	0.397277864	-2.198591033	0.027907015
2	PDK1	0.357375189	-2.671129231	0.007559654
3	DLGAP5	0.567091502	-2.062635139	0.039147307
4	ASF1A	0.359053378	-2.653642374	0.007962817
5	ST6GALNAC6	1.898724967	2.019711022	0.043413371
6	SUMO2	0.346718485	-2.365117868	0.018024333
7	BRIP1	0.453575362	-2.119749975	0.034027136
8	AC131097.2	3.937831563	2.632738899	0.008469943
9	CENPF	0.586505608	-2.30669878	0.021071618
10	PTPN2	0.378693961	-2.177849625	0.029417234
11	CHEK2	0.422240769	-2.07745616	0.037759477
12	PAK1IP1	0.483556649	-2.245128753	0.024759868
13	NUS1	0.295640654	-2.890349779	0.003848134
14	C15ORF59	2.956081195	2.237525193	0.025252035
15	GTSE1	0.451487549	-2.438868384	0.014733333
16	TRIM59	0.310058943	-3.060009496	0.0022133
17	FAM111B	0.59996083	-2.269372812	0.023245664
18	ASPM	0.518452041	-2.349979602	0.01877444
19	MCM6	0.523212965	-2.251137167	0.024376851
20	TOM1L2	2.161448416	2.143470398	0.032075346
21	NEIL3	0.462511653	-2.160366585	0.030744302
22	HELLS	0.321344977	-2.789331014	0.005281705
23	ZDHHC1	2.628482692	2.440752437	0.014656698
24	GJC3	1.44465959	2.114886157	0.034439651
25	E2F8	0.425974425	-2.553875056	0.010653148
26	GRIA1	1723.544976	3.621990675	0.000292345
27	KIF11	0.55847202	-2.247271125	0.024622705
28	EXO1	0.460937898	-3.030924746	0.00243806
29	EZH2	0.572485292	-1.983570228	0.047303771
30	YES1	0.601630765	-2.039299095	0.041420186
31	FOXM1	0.650871311	-2.133764473	0.032862065
32	TYMS	0.628826297	-2.080108988	0.037515537
33	RAD51AP1	0.60276431	-2.043099948	0.041042545
34	CENPU	0.493079906	-2.724145067	0.006446818
35	RAPGEF3	3.967600937	2.640190769	0.008285937
36	DUSP4	0.565981807	-2.369869937	0.017794344
37	CENPQ	0.447055144	-2.271846928	0.023095757
38	ZNF883	1.55594112	2.089772472	0.036638243
39	LRRC8D	0.605125564	-2.132304254	0.032981843
40	CNIH2	0.65503232	-2.144605916	0.031984369
41	CEP55	0.562523781	-2.39416595	0.01665821
42	CCDC160	0.325162054	-2.138272672	0.032494619
43	KIF14	0.510799562	-2.092932584	0.036355173
44	ZWILCH	0.445430957	-2.106543951	0.03515713
45	FAM219A	1.893507046	1.964846926	0.049431957
46	KIF18A	0.449222812	-2.402609976	0.016278539
47	TMPO	0.553419185	-2.117824503	0.034189933
48	NFIA	0.638052834	-2.189504267	0.028560209

(Continued)

10.3389/fimmu.2022.970950

TABLE 2 Continued

	gene	HR	z	P value
49	TPCN1	3.290932788	2.415741287	0.015703214
50	RHNO1	0.514295412	-2.204128211	0.027515329
51	CTSF	1.472449338	2.072960013	0.038176001
52	FAM72C	3.066644433	2.822432329	0.004766088
53	SEPT3	0.55369899	-3.504183774	0.000458009
54	APBA2	1.714725532	2.073801231	0.038097775
55	FUT8	0.474158219	-2.111769715	0.034706206
56	LRGUK	0.062432238	-2.110568263	0.034809438
57	ADCY6	2.067368208	2.081207676	0.037414901
58	VWA2	0.497605963	-2.143156172	0.03210056
59	TTC39C	0.508246458	-2.015928299	0.043807474
60	CYB5D2	2.662637199	2.878174729	0.003999835
61	EXOC6	0.291966726	-2.558775588	0.010504153
62	FAM228B	4.003772856	2.542647066	0.011001629
63	FYB2	0.004631401	-1.992285256	0.046339768
64	SPACA9	2.189328115	2.467635909	0.013600858
65	ARNT2	0.680364619	-2.106494914	0.035161384
66	KRT37	20.62888681	2.558603819	0.010509343
67	AGBL2	0.009647128	-2.17119562	0.029916388
68	AGR2	0.591598511	-2.312899108	0.020728187
69	CCNG2	0.505625171	-2.100827065	0.03565615
70	DNAH5	2.989128412	1.980489126	0.047648594
71	CFAP99	9925778.946	3.182566191	0.001459761
72	C16ORF71	3.729277111	2.109173849	0.034929578
73	FOLH1	0.560549119	-2.027960378	0.042564292
74	C11ORF70	2.694865042	2.529973668	0.011407109
75	LYPD6B	0.427966493	-2.203336375	0.027571049
76	TEX9	0.227190102	-2.109559499	0.034896316
77	NCCRP1	1.287452778	2.89136402	0.003835735
78	SLC1A4	0.569212551	-2.344642453	0.019045334
79	PSD3	0.467450303	-2.303773658	0.021235353
80	KITLG	0.589293262	-2.217628814	0.026580152
81	NT5DC2	1.82691064	2.549310104	0.010793628
82	HMGCL	2.643566798	2.3086158	0.02096491
83	AK8	5.224451486	3.075228928	0.00210341
84	TRERF1	0.510437127	-2.029071436	0.042451015
85	PLPPR3	1.450251013	2.269292578	0.02325054
86	PER2	0.403209642	-2.331594507	0.019722033
87	CFAP45	3.034070634	3.486607251	0.000489189
88	TRIM3	2.974918176	2.62253278	0.008727887
89	ZNF587B	0.19192526	-3.234985541	0.001216489
90	KIAA0040	0.620949694	-2.094736682	0.036194406
91	KCNK6	1.592776419	2.606639009	0.00914357
92	ZNF92	0.359485724	-2.793031961	0.005221653
93	PATZ1	0.401765187	-2.695232839	0.007033946
94	FRY	0.338592705	-2.016077604	0.043791862
95	RHOB	1.713754168	2.231692989	0.025635261
96	ZNF586	0.383365946	-2.064031138	0.039014764

(Continued)

TABLE 2 Continued

	gene	HR	z	P value
97	ZNF703	1.9295092	2.970538785	0.002972779
98	AC008560.1	0.217653792	-2.008129249	0.044629559
99	LRRC46	8.348516913	2.84673037	0.004417076
100	ERMARD	3.155353	2.806137256	0.005013933
101	IKBKB	1.812822308	1.965508385	0.049355426
102	OSCP1	2.410070524	2.698798257	0.006959035
103	AC096887.1	16.86023895	2.189770024	0.02854092
104	INAVA	0.647831638	-2.038272415	0.041522697
105	CASD1	0.407790687	-2.164321144	0.030439711
106	ST8SIA6	0.494039492	-2.094576191	0.036208684

$$Risk\ Score = AC131097.2\ expression \times 0.603906061605295 + TRIM59\ expression \times (-0.250311439654858) + GRIA1\ expression \times 4.75432106611303 + EXO1\ expression \times (-0.151147624875987) + RAPGEF3\ expression \times 0.0694637045689866 + FAM72C\ expression \times 0.687109193015145 + SEPT3\ expression \times (-0.15730852851611) + FAM228B\ expression \times 0.0245662769232489 + AGBL2\ expression \times (-0.255478956488232) + AGR2\ expression \times (-0.0563974779408671) + CFAP99\ expression \times 7.67515956697156 + CFAP300\ expression \times 0.271605260819289 + LYPD6B\ expression \times (-0.0224796251790447) + NCCRP1\ expression \times 0.0265870429310081 + NT5DC2\ expression \times 0.196412885888264 + AK8\ expression \times 0.224051779836318 + CFAP45\ expression \times 0.226574196279908 + ZNF587B\ expression \times (-0.00393169189489084) + ZNF703\ expression \times 0.213684395875983 + LRRC46\ expression \times 0.202045367075346 + EMARD\ expression \times 0.0329920921209145.$$

The AUC was 0.867 and the survival analysis indicated that TNBC patients with a high-risk score possessed a prognosis with misery than those with a low-risk score ($P < 0.0001$) (Figures 5C, 6A, B). Additionally, univariate and multivariate Cox regression analyses were both used to assess whether the 21 CDKN2A-derived genes signature was an independent prognostic factor for other features, including age, sex, metastasis status, tumor stage, and so on. As the forest plots shown, univariate and multivariate Cox regression analyses both indicated that risk score, age, sex, metastasis status, tumor stage, and pathological status were the independent prognostic factors (Figures 6C, D). All results indicated that the 21 CDKN2A-derived genes signature was an independent prognostic factor for TNBC patients.

To further assess the robustness of the CDKN2A-derived genomic model, an independent GEO dataset was used for validation. Reassembly, our scoring system indicated that TNBC patients in the low-risk subgroup had better survival than those in the high-risk subgroup ($P = 0.013$) (Figure 6E). The AUC for OS was 0.874 at 1.5 years, 0.577 at 3 months, 0.622 at 4.5 years, and 0.617 at 6 years in the GSE58812 cohort (Figure 6F).

Moreover, 46 TNBC-specific differentially expressed genes (TDEGs) were screened by two subpopulations comparison in TCGA training data, and then the enrichment analysis of TDEGs was conducted. GO functional annotations described those TDEGs mainly involved in response to interferon-gamma, virus receptor binding, several chemokines receptor binding, and so on (Supplementary Figure S6A). The analysis of the KEGG pathway revealed enrichment of COVID-19, pertussis, Kaposi sarcoma-associated herpesvirus infection, IL-17 signaling pathway, staphylococcus aureus infection, and so on (Supplementary Figure S6B). Additionally, our PPI network also exhibited the correlation between CDKN2A and the genes constructing the model (Supplementary Figure S6C).

Potential therapeutic agents for TNBC patients based on the CDKN2A-derived model

Profiles of gene expression and drug sensitivity were obtained from the PRISM, CTRP, and GDSC dataset, which was used to build the predictive signature of drug response for TNBC. We obtained a total of 1995 drugs from the three databases, as well as 12 compounds shared among 3 datasets (Figure 7A). After removing the drugs whose missing AUC value exceeded 80% and was regarded as NA value, we obtained 174 drugs and 270 cell lines in GDSC, 355 drugs and 638 cell lines in CTRP, as well as 1444 drugs and 462 cell lines in PRISM. The procedure in detail is shown in (Figure 7B). We separated the TNBC patients into high and low risk-score subpopulations pursuant to the CDKN2A-derived prognostic model. The difference in AUC estimates of lapatinib was compared *via* the Wilcoxon rank-sum test. Our data demonstrated that the high risk-score group had higher AUC estimates (Figure 7C). After confirming the reliability of the calculation method, we made some modifications to the analysis of Yang et al. (44). In our study, we started with the differential drug response analysis between low risk-score group and high risk-

TABLE 3 The 21 prognostic genes for constructing the risk predictive model.

Symbol	Name	Category	Ensembl Version	Description and Functional Summary
TRIM59	Tripartite Motif Containing 59	Protein Coding	ENSG00000213186	Activating ubiquitin protein ligase and Acting upstream of or within negative regulation of I-kappaB kinase/NF-kappaB signaling.
GRIA1	Glutamate Ionotropic Receptor AMPA Type Subunit 1	Protein Coding	ENSG00000155511	Ionotropic glutamate receptor. This gene belongs to a family of alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) receptors. It can alternatively splice transcript variants encoding different isoforms.
EXO1	Exonuclease 1	Protein Coding	ENSG00000174371	Encoding a protein with 5' to 3' exonuclease activity and being essential for male and female meiosis.
RAPGEF3	Rap Guanine Nucleotide Exchange Factor 3	Protein Coding	ENSG00000079337	Enabling guanyl-nucleotide exchange factor activity and protein domain specific binding activity.
FAM72C	Family With Sequence Similarity 72 Member C	Protein Coding	ENSG00000263513	A neuronal progenitor cell (NPC) self-renewal supporting protein expressed under physiological conditions at low levels in other tissues.
SEPT3	Neuronal-specific septin-3	Protein coding	ENSG00000224883	Playing a role in cytokinesis.
FAM228B	Family With Sequence Similarity 228 Member B	Protein Coding	ENSG00000219626	FAM228B is a Protein Coding gene. An important paralog of this gene is ENSG00000276087.
AGBL2	AGBL Carboxypeptidase 2	Protein Coding	ENSG00000165923	Enabling metallocarboxypeptidase activity and involved in protein side chain deglutamylation.
AGR2	Anterior Gradient 2, Protein Disulphide Isomerase Family Member	Protein Coding	ENSG00000106541	Encoding a member of the disulfide isomerase (PDI) family of endoplasmic reticulum proteins that catalyze protein folding and thiol-disulfide interchange reactions.
CFAP99	Cilia And Flagella Associated Protein 99	Protein Coding	ENSG00000206113	Predicted to be located in motile cilium.
CFAP300	Cilia and Flagella-associated Protein 300	Protein Coding	ENSG00000137691.13	Playing a role in axonemal structure organization and motility.
LYPD6B	LY6/PLAUR Domain Containing 6B	Protein Coding	ENSG00000150556	Enabling acetylcholine receptor regulator activity and predicted to be located in extracellular region and plasma membrane.
NCCRP1	NCCRP1, F-Box Associated Domain Containing	Protein Coding	ENSG00000188505	Predicted to contribute to ubiquitin protein ligase activity and be involved in positive regulation of cell population proliferation.
NT5DC2	5'-Nucleotidase Domain Containing 2	Protein Coding	ENSG00000168268	Predicted to enable 5'-nucleotidase activity and be involved in dephosphorylation.
AK8	Adenylate Kinase 8	Protein Coding	ENSG00000165695	Enabling AMP binding activity and nucleobase-containing compound kinase activity.
CFAP45	Cilia And Flagella Associated Protein 45	Protein Coding	ENSG00000213085	Enabling AMP binding activity and involved in establishment of left/right asymmetry and flagellated sperm motility.
ZNF587B	Zinc Finger Protein 587B	Protein Coding	ENSG00000269343	Enabling DNA-binding transcription repressor activity, RNA polymerase II-specific and RNA polymerase II transcription regulatory region sequence-specific DNA binding activity.
ZNF703	Zinc Finger Protein 703	Protein Coding	ENSG00000183779	Enabling DNA-binding transcription factor binding activity.
LRRC46	Leucine Rich Repeat Containing 46	Protein Coding	ENSG00000141294	LRRC46 is a Protein Coding gene. Diseases associated with LRRC46 include Ciliary Dyskinesia, Primary, 13. An important paralog of this gene is LRGUK.
EMARD	Not Available	Not Available	Not Available	Not Available
AC131097.2	Not Available	Not Available	Not Available	Not Available

score group by the median split. Next, agents with correlation coefficients ($P < 0.05$) were identified according to Pearson rank correlation analysis between the risk-score of the CDKN2A-derived model and AUC values. Three overlapping drugs were eventually found in these three databases, including afatinib, erlotinib and lapatinib (Figures 7D–F). Then, we analyzed the target gene expression difference of three mentioned-above potential drugs between high risk-score and low risk-score subpopulations. Notably, EGFR is the co-target gene of the three candidate drugs. The expression level of EGFR was significantly upregulated in the low risk-score subpopulation (Figure 7G). In summary, our outcomes indicated that afatinib, erlotinib and lapatinib could be designated as the potential drugs for low risk-score TNBC patients by targeting EGFR.

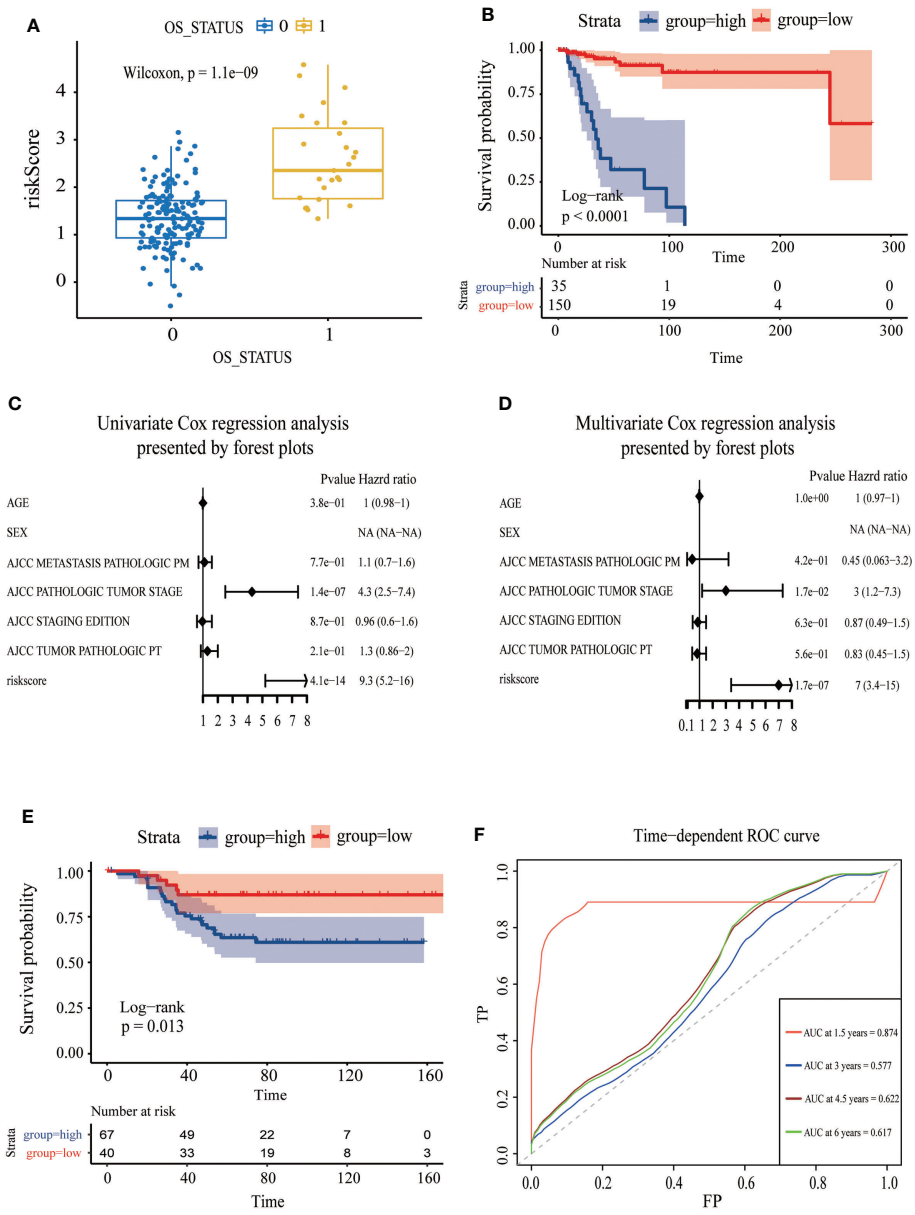


FIGURE 6
Assessment of the independent prognostic value and validation of CDKN2A-derived prognostic model of TNBC. **(A)** The correlation between OS status and risk score. **(B)** The survival curve of high and low risk score in TNBC. **(C)** Univariate Cox regression analysis of CDKN2A-derived prognostic model. **(D)** Multivariate Cox regression analysis of CDKN2A-derived prognostic model. **(E)** The survival curve verified by the external validation set. **(F)** The time-dependent ROC curve and AUC values respectively at 1.5 years, 3 years, 4.5 years, and 6 years verified by an external validation set.

Discussion

The rapidly increasing number of diagnosed BRCA patients results in the urgent need for new biomarkers that can elucidate breast carcinogenesis and predict the immune prognosis (45). In view of problems, such as small sample size, multiple BRCA subtypes, and complex mechanisms of BRCA, previous studies disputed that heterogeneity existed in the expression of CDKN2A in BRCA (46–48) and did not obtain the final verdict of CDKN2A's effects on BRCA.

The landscape analysis based on the multiple data indicated that CDKN2A had an overexpression and critical values of prognosis in BRCA, hinting at its clinical property as a prognostic biomarker. Additionally, its upregulation was strikingly correlated to DNA hypermethylation. Genetic and epigenetic alterations are both involved in the procession of

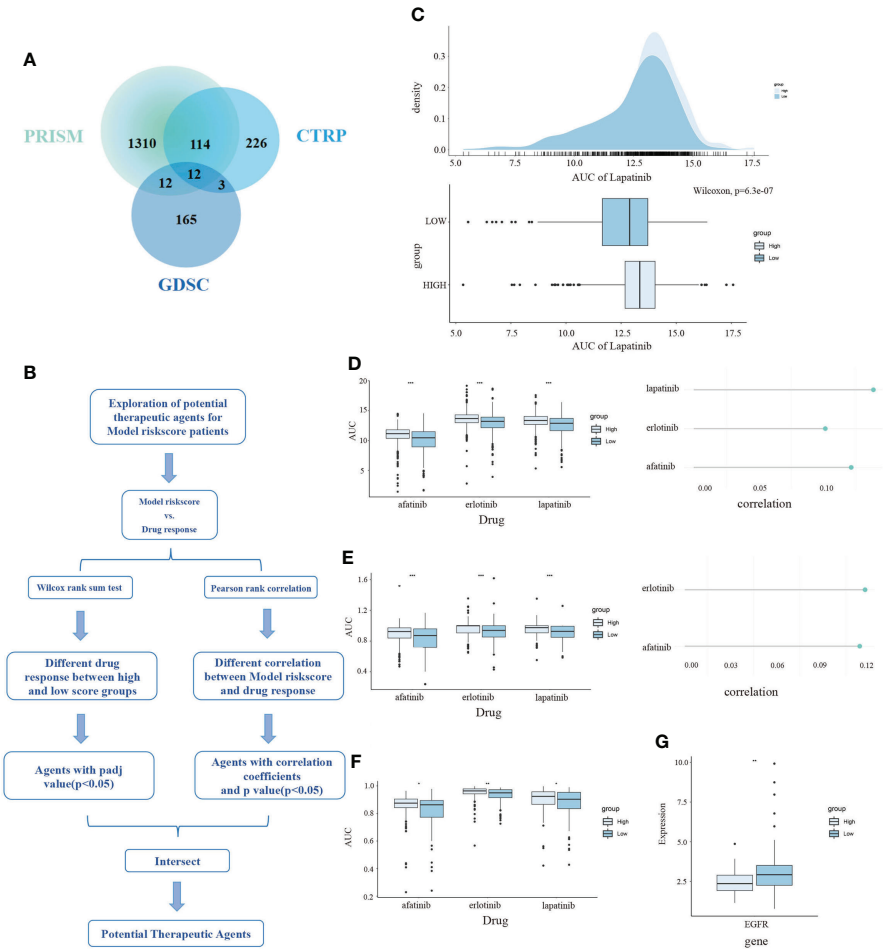


FIGURE 7
The exploration of potential targeted drugs based on the CDKN2A-derived model for TNBC patients. **(A)** The shared drug between PRISM, GDSC and CTRP. **(B)** The flow chart of exploring potential therapeutic agents. **(C)** The AUC of lapatinib in high and low risk score subpopulations of TNBC patients. **(D–F)** The AUC of three selected drugs in high and low risk score subpopulations of TNBC patients. **(G)** The relationship between AUC values and targets of three drugs. The P values of the figure are shown as follows: *P < 0.05. **P < 0.01. ***P < 0.001.

breast carcinogenesis. The promoter hypermethylation level is commonly associated with transcriptional gene silencing (49). Our results were indicative that DNA hypermethylation of CDKN2A promoted breast carcinogenesis and had a significant association with subtypes of BRCA. Especially in the Luminal and TNBC subtypes, the hypermethylation of CDKN2A was more significant (P < 0.05). Lubecka et al. (50) indicated that the administration of sulforaphane and clofarabine could inhibit the tumor cell growth in breast tissues *via* reactivating methylation-silenced CDKN2A. Thus, inhibiting the hypermethylation levels of CDKN2A could be a potential therapeutic method of BRCA, especially for patients with Luminal and TNBC subtypes. In view of the uncommon phenomenon that CDKN2A showed hypermethylation in BRCA but exhibited high expression, we searched more literatures. Smith et al. (51) reviewed several cases about promoter DNA hypermethylation promoting target gene transcription and they postulated a context-dependent model whereby epigenetic contributions to transcriptional regulation occur in a more complex and dynamic manner, which needs further investigation. The analysis of CDKN2A and drug sensitivity in BRCA expanded clinical applications of CDKN2A. A case-control study (52) indicated the risk of BRCA had a 1.82-fold increase in women with high sensitivity to bleomycin, which reversely confirms our results that CDKN2A upregulation could reduce sensitivity to bleomycin, resulting in a positive prognosis. Herein, regulating the expression of CDKN2A might alter the drug sensitivity and affect the therapeutic results.

As numerous evidence robustly supported, the overload of copper is thought to induce neurotoxicity in neurodegenerative disorders (Parkinson's disease and Alzheimer's disease) and hepatocerebral (Wilson's disease) over several decades (53).

However, the previous regulation of ferroptosis to trigger tumor cell death (54) gave us inspiring hints that it is highly viable to regulate the certain copper levels in a suitable concentration to induce the cuproptosis and tumor cell death (55). Ferroptosis indicates an oxidative cell death resulting from the deterioration of antioxidant function and accretion of lipid reactive oxygen species (9). Excess copper can trigger proteotoxic stress and death in cells through the combination with lipoylated components of the tricarboxylic acid (TCA) cycle (8). The stimulation of inflammatory conditions will lead to elevated serum copper levels and trigger oxidative stress, thereby activating the inflammatory response (56). In reverse, inflammation could also accelerate the cytotoxicity mediated by copper *via* overexpressing six-transmembrane epithelial antigens of prostate 4 (STEAP4) (29). Disulfiram/copper was investigated to induce cytotoxic and anti-tumor effects on nasopharyngeal carcinoma cells through p53-mediated ferroptosis and ROS/MAPK pathways (30). Fatty acids degradation can tremendously alter the microbial sensitivity to copper, thus induce copper toxicity (28). Copper also could trigger the expression of GPER, VEGF, and HIF-1α *via* activating EGFR/ERK/c-fos transduction pathway, affecting the angiogenesis and tumor progression in BRCA and LIHC (31). Our study groundbreakingly and vicariously evaluated the activities of ferroptosis and cuproptosis for patients with BRCA based on the seven above-mentioned pathways. Our three subtypes obtained from unsupervised cluster analysis not only exhibited distinct activities in multiple tumor-related pathways but also had critical significance in scores of FAC. Ke et al. (57) indicated that FTL could function as a prognostic and diagnostic ferroptosis regulator in hepatocellular carcinoma *via* random forest analysis, which was consistent with our results that the FTL-dominant cluster possessed a strong connection with ferroptosis. Their resemble conclusion that higher infiltrating immune cells, including Gamma delta T cells and activated CD8+ T cells, emerged in the high FTL expression group, was also confirmed in our study. Hence, regulating pathways involved in CDKN2A-associated genes or designing novel metal-based anticancer agents to induce ferroptosis and cuproptosis may guide us to develop new anti-cancer treatment strategies for BRCA, especially for the patients in the FTL-dominant subtype.

As previous studies reported, the dynamical characteristics of the TME, chemokines, immune checkpoints, and tumor immune infiltration have a clear underlying role in tumorigenesis and progression (58, 59). Surgery, endocrine therapy, and chemotherapy remain the fundamental cornerstones of BRCA, nevertheless, immunotherapy has gradually become one of the neoadjuvant combination therapy strategies (60). Our further analysis proved that CDKN2A overexpression was correlated to the increased immune cells, enhanced immune checkpoints, and elevated chemokines, indicating that CDKN2A might be applied as a potential immunotherapeutic therapy. The profile from CIBERTSORT in our unsupervised groups is highly in line with current studies on immune cell infiltration. Group 1, namely the "cold immune subtype", showed relatively high levels of naïve B cells, resting memory CD4+ T, and M2 macrophages. Gunderson et al. (22) reported that patients with overexpression of naïve B cells had a sign of misery prognosis, verifying its carcinogenic effect. The higher ratio of resting memory CD4+ T cells in our cold immune subtype was consistent with the hints that resting memory CD4+ T cells predicted an undesirable clinical outcome (61). As an anti-inflammatory and pro-tumor factor, M2 macrophage, was widely recognized as a promoter of metastatic progression and poor prognosis in BRCA (62). However, Spear et al. demonstrated that the infiltration of memory B cells could serve as an immunostimulatory factor and supported the adaptive antitumor immunotherapy (63), which was consistent with our analyses of INF-γ activated subtype. As was mentioned above, our unsupervised groups were correlated with the TME and gave us potential immune therapeutic opportunities by respectively modulating corresponding immune cells in each group.

Additionally, our results from unsupervised clusters analysis were consistent with prior investigations that TNBC was more likely to harbor immunogenicity and more suitable for immunotherapy than other molecular subtypes (64). Moreover, current clinical investigations are paying attention to making non-responders convert to responders or deepening those occurred responses. The previous study reported that the loss of CDKN2A significantly made non-small cell lung cancer patients experience disease progression after immune checkpoint blockade therapy (65). Horn et al. also demonstrated that the frequent loss of the CDKN2A could trigger the susceptibility to IFN-γ resistance *via* JAK2 gene deletion in melanoma (66), which was in line with our conclusion that high expression of CDKN2A potentially benefited from immunotherapy. However, the immunotherapy response of CDKN2A in TNBC has not been reported. Our IPS and verification of external BRCA cohorts (GSE173839) comprehensively suggested that the expression of CDKN2A could modulate the response to immunotherapy to TNBC, and TNBC with high CDKN2A expression patients have higher immunogenicity and benefit from immunotherapy.

Because the overexpression of CDKN2A was indicative of desirable clinical outcomes for TNBC patients, we further conducted WGCNA analysis to determine CDKN2A-derived genes that were chiefly associated with TNBC and pathways of FAC. Determining genes and utilizing cox and lasso analysis, we established a CDKN2A-derived prognostic model, consisting of TRIM59, EXO1, AGR2, ZNF703, and other 17 genes. According to immunohistochemistry, Liu et al. (67) found that TRIM59 levels were notably higher in the TNBC subtype and promoted the malignant behavior *via* regulating the AKT pathway, leading to the undesirable prognosis. Previously, RT-qPCR also proved

the overexpression of EXO1 in BRCA cells MDA-MB231, and the elevated EXO1 might be utilized as an indicator of poor BRCA prognosis (68). A clinical observation study (69) *via* the cross-sectional method indicated that AGR2 expression is positively associated with the incidence of distant metastases in BRCA and upregulated AGR2 was a poor prognosis predictor. Current research reported that ZNF703 expressed in approximately 34.2% of TNBC *via* immunohistochemistry and the knockdown of ZNF703 triggered a powerful inhibition of TNBC cell proliferation and cell cycle, along with the downregulation of cyclin D1, CDK4, CDK6, and E2F1 (70). Remaining genes were firstly explored to have effects on the prognosis of TNBC patients. Deeper studies of the biological roles of these genes in TNBC are warranted and clinical investigations of this signature need to be further tested.

In terms of the high heterogeneity of TNBC, it's incredibly difficult to find new curative targets and develop novel targeted therapy. DNA microarray analysis conducted by Nielsen et al. (71) indicated that overexpression of EGFR existed in 60% of TNBC samples, which was consistent with our results. The study of Livasy et al. (72) also validified that approximately 70% of TNBC samples significantly expressed elevated EGFR. Hence, it is inferred that EGFR may be a promising curative target in TNBC, especially for TNBC patients with low risk-score according to our model. As the irreversible ErbB family blocker, afatinib (AFT) was approved by the FDA to treat the advanced EGFR mutation-positive NSCLC (73). The investigations of AFT treatment in BRCA are undergoing. In an open-label, multicenter, and phase II clinical trial, Hickish et al. (74) reported that for metastatic BRCA patients whose prior HER2-targeted therapy had undesirably failed, AFT alone and combined with paclitaxel or vinorelbine could enhance the objective response. Our data demonstrated AFT may have a good therapeutic effect on TNBC. Coherent with our outcomes, Wang et al. (75) developed AFT/2-BP@PLGA@MD, a poly(d,l-lactide-glycolide) (PLGA)-based intelligent bionic nanoplatform, which was covered under a cancer cell membrane to block PD-1 and PD-L1. AFT/2-BP@PLGA@MD nanoparticles integrated the targeted therapy of AFT and immunotherapy, exhibiting enhanced inhibition of the growth of TNBC. As a dual inhibitor of EGFR and HER2, lapatinib could also induce inhibition of p-Akt and CIP2A and trigger apoptosis in TNBC cell lines (76). LHNPs, human serum albumin nanoparticles loaded with lapatinib, were developed by the advanced nanoparticle albumin-bound technology, and could inhibit the brain metastasis from TNBC ascribed to the downregulation of metastasis-related proteins (77). Collectively, based on the model, we proposed three drugs that may be applicable to TNBC patients with low risk-score. Previous ssGSEA results also presented the CDKN2A-associated genes also correlated to the EGFR activity, indicating that CDKN2A may function as a promising predictive biomarker for anti-EGFR therapy in TNBC. Moreover, drawing support from advanced nanoparticle technology, we put forward the perspective that developing novel nanoparticles combined with immunotherapy and targeted therapy to achieve a better prognosis for TNBC patients.

Conclusion

In summary, our study comprehensively analyzed the biological role and prognostic values of CDKN2A in BRCA. Given the strong association between CDKN2A and FAC, we indicated that regulating pathways involved in CDKN2A-associated genes or designing novel metal-based anticancer agents to induce ferroptosis and cuproptosis may guide us to develop new anti-cancer treatment strategies. Besides, we substantively found that CDKN2A may serve as the pioneering prognostic predictor for TNBC. TNBC patients with high CDKN2A expression possess the higher immunogenicity and benefit from immunotherapy. The CDKN2A-derived model we established can also guide the prognosis of TNBC patients. To further guide the treatment, we also provided three drugs for precision medicine of TNBC *via* targeting EGFR and indicated that CDKN2A may function as a promising predictive biomarker for anti-EGFR therapy in TNBC. Therefore, this investigation provides a rationale and offers fresh perspectives and orientations for TNBC treatment.

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

TC and YW conceived the presented idea and analyzed the data under the supervision of CH. ZL, YY, SS and MG collected the data. TC and YW took the lead in drafting the manuscript with input from all authors. BS revised the manuscript. SS and BS interpreted results from a clinical point of view. All authors read and approved the final manuscript.

Funding

This work was funded by The Science and Technology Development Fund, Macau SAR 462 (File no. 0020/2021/A, 001/2020/ALC, SKL-QRCM (MUST)-2020-2022), and Dr. Neher's Biophysics Laboratory for Innovative Drug Discovery, State Key Laboratory of Quality Research in Chinese Medicine, Macau University of Science and Technology, Macau, China (Grant no. FRG-21-032-SKL).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.970950/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Dawson SJ, Provenzano E, Caldas C. Triple negative breast cancers: clinical and prognostic implications. *Eur J Cancer.* (2009) 45 Suppl 1:27–40. doi: 10.1016/S0959-8049(09)70013-9

3. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* (2007) 13(15 Pt 1):4429–34. doi: 10.1158/1078-0432.CCR-06-3045

4. Chan SH, Chiang J, Ngeow J. CDKN2A germline alterations and the relevance of genotype-phenotype associations in cancer predisposition. *Hereditary Cancer Clin practice.* (2021) 19(1):21. doi: 10.1186/s13053-021-00178-x

5. Liggett WH Jr., Sidransky D. Role of the p16 tumor suppressor gene in cancer. *J Clin Oncol* (1998) 16(3):1197–206. doi: 10.1200/JCO.1998.16.3.1197

6. Luan Y, Zhang W, Xie J, Mao J. CDKN2A inhibits cell proliferation and invasion in cervical cancer through LDHA-mediated AKT/mTOR pathway. *Clin Trans* (2021) 23(2):222–8. doi: 10.1007/s12094-020-02409-4

7. Shi J, Wu P, Sheng L, Sun W, Zhang H. Ferroptosis-related gene signature predicts the prognosis of papillary thyroid carcinoma. *Cancer Cell Int* (2021) 21(1):669. doi: 10.1186/s12935-021-02389-7

8. Tsvetkov P, Coy S, Petrova B, Dreishpoon M, Verma A, Abdusamad M, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. *Science* (2022) 375(6586):1254–61. doi: 10.1126/science.abf0529

9. Li J, Cao F, Yin HL, Huang ZJ, Lin ZT, Mao N, et al. Ferroptosis: past, present and future. *Cell Death Dis* (2020) 11(2):88. doi: 10.1038/s41419-020-2298-2

10. Chen D, Tavana O, Chu B, Erber L, Chen Y, Baer R, et al. NRF2 is a major target of ARF in p53-independent tumor suppression. *Mol Cell* (2017) 68(1):224–32.e4. doi: 10.1016/j.molcel.2017.09.009

11. Aftab A, Shahzad S, Hussain HMJ, Khan R, Irum S, Tabassum S. CDKN2A/P16INK4A variants association with breast cancer and their in-silico analysis. *Breast Cancer.* (2019) 26(1):11–28. doi: 10.1007/s12282-018-0894-0

12. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* (2017) 77(21):e108–e10. doi: 10.1158/0008-5472.CAN-17-0307

13. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45(W1):W98–w102. doi: 10.1093/nar/gkx247

14. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci Publ Protein Society* (2018) 27(1):233–44. doi: 10.1002/pro.3307

15. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia (New York NY)* (2017) 19(8):649–58. doi: 10.1016/j.neo.2017.05.002

16. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* (2015) 16:636. doi: 10.1186/s12864-015-1847-z

17. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signaling* (2013) 6(269):pl1. doi: 10.1126/scisignal.2004088

18. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* (2019) 47(D1):D941–D7. doi: 10.1093/nar/gky1015

19. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinf (Oxford England)* (2019) 35(20):4200–2. doi: 10.1093/bioinformatics/btz210

20. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* (2012) 72(14):3499–511. doi: 10.1158/0008-5472.CAN-12-1370

21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337

22. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

23. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460

24. Sokolov A, Carlin DE, Paull EO, Baertsch R, Stuart JM. Pathway-based genomics prediction using generalized elastic net. *PLoS Comput Biol* (2016) 12(3):e1004790. doi: 10.1371/journal.pcbi.1004790

25. Hu J, Yu A, Othmane B, Qiu D, Li H, Li C, et al. Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. *Theranostics* (2021) 11(7):3089–108. doi: 10.7150/thno.53649

26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

27. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019

28. Avery SV, Howlett NG, Radice S. Copper toxicity towards saccharomyces cerevisiae: dependence on plasma membrane fatty acid composition. *Appl Environ Microbiol* (1996) 62(11):3960–6. doi: 10.1128/aem.62.11.3960-3966.1996

29. Jiang C, Wu B, Xue M, Lin J, Hu Z, Nie X, et al. Inflammation accelerates copper-mediated cytotoxicity through induction of six-transmembrane epithelial antigens of prostate 4 expression. *Immunol Cell Biol* (2021) 99(4):392–402. doi: 10.1111/imcb.12427

30. Li Y, Chen F, Chen J, Chan S, He Y, Liu W, et al. Disulfiram/Copper induces antitumor activity against both nasopharyngeal cancer cells and cancer-associated fibroblasts through ROS/MAPK and ferroptosis pathways. *Cancers (Basel)* (2020) 12(1):138. doi: 10.3390/cancers12010138

31. Rigiracciolo DC, Scarpelli A, Lappano R, Pisano A, Santolla MF, De Marco P, et al. Copper activates HIF-1alpha/GPER/VEGF signalling in cancer cells. *Oncotarget* (2015) 6(33):34158–77. doi: 10.18632/oncotarget.5779

32. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6

33. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49(D1):D605–D12. doi: 10.1093/nar/gkaa10749

34. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

35. Langfelder P, Horvath S. Fast r functions for robust correlations and hierarchical clustering. *J Stat Software* (2012) 46(11):11. doi: 10.18637/jss.v046.i11

36. Dill CD, Dammer EB, Griffen TL, Seyfried NT, Lillard JWJr. A network approach reveals driver genes associated with survival of patients with triple-negative breast cancer. *iScience* (2021) 24(5):102451. doi: 10.1016/j.isci.2021.102451

37. Wang H, Lengerich BJ, Aragam B, Xing EP. Precision lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* (2019) 35(7):1181–7. doi: 10.1093/bioinformatics/bty750

38. Hua X, Zhao W, Pesatori AC, Consonni D, Caporaso NE, Zhang T, et al. Genetic and epigenetic intratumor heterogeneity impacts prognosis of lung adenocarcinoma. *Nat Commun* (2020) 11(1):2459. doi: 10.1038/s41467-020-16295-5

39. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41(Database issue): D955–61. doi: 10.1093/nar/gks1111

40. Lyons TG. Targeted therapies for triple-negative breast cancer. *Curr Treat Options Oncol* (2019) 20(11):82. doi: 10.1007/s11864-019-0682-x

41. Chen YP, Wang YQ, Lv JW, Li YQ, Chua MLK, Le QT, et al. Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy. *Ann Oncol* (2019) 30(1):68–75. doi: 10.1093/annonc/mdy470

42. Sacher AG, St Paul M, Paige CJ, Ohashi PS. Cytotoxic CD4(+) T cells in bladder cancer-a new license to kill. *Cancer Cell* (2020) 38(1):28–30. doi: 10.1016/j.ccell.2020.06.013

43. Pitt JM, Andre F, Amigorena S, Soria JC, Eggermont A, Kroemer G, et al. Dendritic cell-derived exosomes for cancer therapy. *J Clin Invest* (2016) 126 (4):1224–32. doi: 10.1172/JCI81137

44. Chu G, Shan W, Ji X, Wang Y, Niu H. Multi-omics analysis of novel signature for immunotherapy response and tumor microenvironment regulation patterns in urothelial cancer. *Front Cell Dev Biol* (2021) 9:764125. doi: 10.3389/fcell.2021.764125

45. Cheng T, Chen P, Chen J, Deng Y, Huang C. Landscape analysis of matrix metalloproteinases unveils key prognostic markers for patients with breast cancer. *Front Genet* (2021) 12:809600. doi: 10.3389/fgene.2021.809600

46. Bartels S, van Luttikhuizen JL, Christgen M, Magel L, Luft A, Hanzelmann S, et al. CDKN2A loss and PIK3CA mutation in myoepithelial-like metaplastic breast cancer. *J Pathol* (2018) 245(3):373–83. doi: 10.1002/path.5091

47. Witkiewicz AK, Rivadeneira DB, Ertel A, Kline J, Hyslop T, Schwartz GF, et al. Association of RB/p16-pathway perturbations with DCIS recurrence: dependence on tumor versus tissue microenvironment. *Am J Pathol* (2011) 179 (3):1171–8. doi: 10.1016/j.ajpath.2011.05.043

48. Herschkowitz JI, He X, Fan C, Perou CM. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal b breast carcinomas. *Breast Cancer Res* (2008) 10(5):R75. doi: 10.1186/bcr2142

49. Lewis CM, Cler LR, Bu DW, Zochbauer-Muller S, Milchgrub S, Naftalis EZ, et al. Promoter hypermethylation in benign breast epithelium in relation to predicted breast cancer risk. *Clin Cancer Res* (2005) 11(1):166–72. doi: 10.1158/1078-0432.166.11.1

50. Lubecka K, Kaufman-Szymczyk A, Fabianowska-Majewska K. Inhibition of breast cancer cell growth by the combination of clofarabine and sulforaphane involves epigenetically mediated CDKN2A upregulation. *Nucleosides Nucleotides Nucleic Acids* (2018) 37(5):280–9. doi: 10.1080/15257770.2018.1453075

51. Smith J, Sen S, Weeks RJ, Eccles MR, Chatterjee A. Promoter DNA hypermethylation and paradoxical gene activation. *Trends Cancer* (2020) 6 (5):392–406. doi: 10.1016/j.trecan.2020.02.007

52. Hu M, Han D, Sun S, Yan Y, Zhang J, Zhou Y. Bleomycin-induced mutagen sensitivity, passive smoking, and risk of breast cancer in Chinese women: a case-control study. *Cancer Causes Control* (2013) 24(4):629–36. doi: 10.1007/s10552-012-0137-1

53. Pal A. Copper toxicity induced hepatocerebral and neurodegenerative diseases: an urgent need for prognostic biomarkers. *Neurotoxicology* (2014) 40:97–101. doi: 10.1016/j.neuro.2013.12.001

54. Liang C, Zhang X, Yang M, Dong X. Recent progress in ferroptosis inducers for cancer therapy. *Adv Mater* (2019) 31(51):e1904197. doi: 10.1002/adma.201904197

55. Ge EJ, Bush AI, Casini A, Cobine PA, Cross JR, DeNicola GM, et al. Connecting copper and cancer: from transition metal signalling to metalloplasia. *Nat Rev Cancer* (2022) 22(2):102–13. doi: 10.1038/s41568-021-00417-2

56. Pereira TC, Campos MM, Bogo MR. Copper toxicology, oxidative stress and inflammation using zebrafish as experimental model. *J Appl Toxicol* (2016) 36 (7):876–85. doi: 10.1002/jat.3303

57. Ke S, Wang C, Su Z, Lin S, Wu G. Integrated analysis reveals critical ferroptosis regulators and FTL contribute to cancer progression in hepatocellular carcinoma. *Front Genet* (2022) 13:897683. doi: 10.3389/fgene.2022.897683

58. Neal JT, Li X, Zhu J, Giangarra V, Grzeskowiak CL, Ju J, et al. Organoid modeling of the tumor immune microenvironment. *Cell* (2018) 175(7):1972–88 e16. doi: 10.1016/j.cell.2018.11.021

59. Kaderbhai C, Tharin Z, Ghiringhelli F. The role of molecular profiling to predict the response to immune checkpoint inhibitors in lung cancer. *Cancers (Basel).* (2019) 11(2):201. doi: 10.3390/cancers11020201

60. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet* (2021) 397(10286):1750–69. doi: 10.1016/S0140-6736(20)32381-3

61. Gu J, Zhang J, Huang W, Tao T, Huang Y, Yang L, et al. Activating miRNA-mRNA network in gemcitabine-resistant pancreatic cancer cell associates with alteration of memory CD4(+) T cells. *Ann Transl Med* (2020) 8(6):279. doi: 10.21037/atm.2020.03.53

62. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. *Clin Cancer Res* (2019) 25(16):5002–14. doi: 10.1158/1078-0432.CCR-18-3524

63. Spear S, Candido JB, McDermott JR, Ghirelli C, Maniati E, Beers SA, et al. Discrepancies in the tumor microenvironment of spontaneous and orthotopic murine models of pancreatic cancer uncover a new immunostimulatory phenotype for b cells. *Front Immunol* (2019) 10:542. doi: 10.3389/fimmu.2019.00542

64. Emens LA. Breast cancer immunotherapy: Facts and hopes. *Clin Cancer Res* (2018) 24(3):511–20. doi: 10.1158/1078-0432.CCR-16-3001

65. Gutiontov SI, Turchan WT, Spurr LF, Rouhani SJ, Chervin CS, Steinhardt G, et al. CDKN2A loss-of-function predicts immunotherapy resistance in non-small cell lung cancer. *Sci Rep* (2021) 11(1):20059. doi: 10.1038/s41598-021-99524-1

66. Horn S, Leonardelli S, Sucker A, Schadendorf D, Griewank KG, Paschen A. Tumor CDKN2A-associated JAK2 loss and susceptibility to immunotherapy resistance. *J Natl Cancer Inst* (2018) 110(6):677–81. doi: 10.1093/jnci/djx271

67. Liu Y, Dong Y, Zhao L, Su L, Diao K, Mi X. TRIM59 overexpression correlates with poor prognosis and contributes to breast cancer progression through AKT signaling pathway. *Mol Carcinog* (2018) 57(12):1792–802. doi: 10.1002/mc.22897

68. Liu J, Zhang J. Elevated EXO1 expression is associated with breast carcinogenesis and poor prognosis. *Ann Transl Med* (2021) 9(2):135. doi: 10.21037/atm-20-7922

69. Kereh DS, Pieter J, Hamdani W, Haryasena H, Sampepajung D, Prihantono P. Correlation of AGR2 expression with the incidence of metastasis in luminal breast cancer. *Breast Dis* (2021) 40(S1):S103–S7. doi: 10.3233/BD-219015

70. Zhang X, Mu X, Huang O, Wang Z, Chen J, Chen D, et al. ZNF703 promotes triple-negative breast cancer cells through cell-cycle signaling and associated with poor prognosis. *BMC Cancer* (2022) 22(1):226. doi: 10.1186/s12885-022-09286-w

71. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* (2004) 10(16):5367–74. doi: 10.1158/1078-0432.CCR-04-0220

72. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* (2006) 19(2):264–71. doi: 10.1038/modpathol.3800528

73. Wecker H, Waller CF. Afatinib. *Recent Results Cancer Res* (2018) 211:199–215. doi: 10.1007/978-3-319-91442-8_14

74. Hickish T, Mehta A, Liu MC, Huang CS, Arora RS, Chang YC, et al. Afatinib alone and in combination with vinorelbine or paclitaxel, in patients with HER2-positive breast cancer who failed or progressed on prior trastuzumab and/or lapatinib (LUX-breast 2): an open-label, multicenter, phase II trial. *Breast Cancer Res Treat* (2022) 192(3):593–602. doi: 10.1007/s10549-021-06449-4

75. Wang X, Zhu X, Li B, Wei X, Chen Y, Zhang Y, et al. Intelligent biomimetic nanoplatform for systemic treatment of metastatic triple-negative breast cancer *via* enhanced EGFR-targeted therapy and immunotherapy. *ACS Appl Mater Interfaces* (2022) 14:23152–63. doi: 10.1021/acsami.2c02925

76. Liu CY, Hu MH, Hsu CJ, Huang CT, Wang DS, Tsai WC, et al. Lapatinib inhibits CIP2A/PP2A/p-akt signaling and induces apoptosis in triple negative breast cancer cells. *Oncotarget* (2016) 7(8):9135–49. doi: 10.18632/oncotarget.7035

77. Wan X, Zheng X, Pang X, Pang Z, Zhao J, Zhang Z, et al. Lapatinib-loaded human serum albumin nanoparticles for the prevention and treatment of triple-negative breast cancer metastasis to the brain. *Oncotarget* (2016) 7(23):34038–51. doi: 10.18632/oncotarget.8697

10.3389/fimmu.2022.970950

Glossary

BRCA	breast cancer
BLCA	bladder urothelial carcinoma
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	cholangiocarcinoma
COAD	colon adenocarcinoma
HNSC	head and neck cancer
KICH	kidney chromophobe
KIRC	kidney renal clear cell carcinoma
KIRP	kidney renal papillary cell carcinoma
LIHC	liver hepatocellular carcinoma
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
PRAD	prostate adenocarcinoma
READ	rectum adenocarcinoma
STAD	stomach adenocarcinoma
THCA	thyroid carcinoma
UCEC	uterine corpus endometrial carcinoma
TNBC	triple-negative breast cancer
CDKN2A	cyclin-dependent kinase inhibitor 2A
FAC	ferroptosis and cuproptosis
TME	tumor microenvironment
Her2	human epidermal growth factor receptor 2
ER/PR	estrogen receptor/progesterone receptor
ROS	reactive oxygen species
TCA	tricarboxylic acid
mRNA	the messenger RNA
IPS	Immunophenoscore
MAPK	mitogen-activated protein kinase
VEGF	vascular endothelial growth factor
PCA	principal component analysis
Bicor	biweight midcorrelation
ROC	receiver operating characteristic
GDSC	Genomics of Drug Sensitivity in Cancer
PRISM	Profiling Relative Inhibition Simultaneously in Mixtures
CTRP	Cancer Therapeutics Response Portal
OS	overall survival
DEGs	differentially expressed genes
TDEGs	TNBC-specific differentially expressed genes
STEAP4	six-transmembrane epithelial antigens of prostate 4
GPER	G protein-coupled estrogen receptor
AFT	afatinib
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
TIMER	Tumor Immune Estimation Resource
GEPIA	Gene expression profiling interactive analysis
HPA	Human Protein Atlas
TISIDB	Tumor and Immune System Interaction Database
CIBERSORT	Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts
ssGSEA	Single-Sample Gene Set Enrichment Analysis
TCIA	The Cancer Immunome Atlas
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes
Genomes	
MSigDB	Molecular Signatures Database
LASSO	Least Absolute Shrinkage and Selection Operator.

(Continued)

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fimmu.2022.950782

Check for updates

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Fangfang Duan,
Shenzhen Campus of Sun Yat-sen University, China
Ping Zheng,
The University of Melbourne, Australia
Zhenhuan Zhao,
University of Virginia, United States

*CORRESPONDENCE
Changhong Miao
15111230032@fudan.edu.cn
Di Zhou
judy612542@163.com
Minli Sun
sunminli@sina.com

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

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 23 May 2022
ACCEPTED 01 August 2022
PUBLISHED 23 August 2022

CITATION
Weng M, Li T, Zhao J, Guo M, Zhao W, Gu W, Sun C, Yue Y, Zhong Z, Nan K, Liao Q, Sun M, Zhou D and Miao C (2022) mRNAsi-related metabolic risk score model identifies poor prognosis, immunoevasive contexture, and low chemotherapy response in colorectal cancer patients through machine learning. *Front. Immunol.* 13:950782. doi: 10.3389/fimmu.2022.950782

COPYRIGHT
© 2022 Weng, Li, Zhao, Guo, Zhao, Gu, Sun, Yue, Zhong, Nan, Liao, Sun, Zhou and Miao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

mRNAsi-related metabolic risk score model identifies poor prognosis, immunoevasive contexture, and low chemotherapy response in colorectal cancer patients through machine learning

Meilin Weng[1,2†], Ting Li[1,2†], Jing Zhao[3†], Miaomiao Guo[1,2], Wenling Zhao[1,2], Wenchao Gu[4,5], Caihong Sun[1,2], Ying Yue[1,2], Ziwen Zhong[1,2], Ke Nan[1,2], Qingwu Liao[1,2], Minli Sun[1,2*], Di Zhou[1,2*] and Changhong Miao[1,2*]

[1]Department of Anesthesiology, Zhongshan hospital, Fudan University, Shanghai, China, [2]Shanghai Key Laboratory of Perioperative Stress and Protection, Zhongshan hospital, Fudan University, Shanghai, China, [3]Department of Pathology, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China, [4]Department of Diagnostic and Interventional Radiology, University of Tsukuba, Ibaraki, Japan, [5]Department of Diagnostic Radiology and Nuclear Medicine, Gunma University Graduate School of Medicine, Maebashi, Japan

Colorectal cancer (CRC) is one of the most fatal cancers of the digestive system. Although cancer stem cells and metabolic reprogramming have an important effect on tumor progression and drug resistance, their combined effect on CRC prognosis remains unclear. Therefore, we generated a 21-gene mRNA stemness index-related metabolic risk score model, which was examined in The Cancer Genome Atlas and Gene Expression Omnibus databases (1323 patients) and validated using the Zhongshan Hospital cohort (200 patients). The high-risk group showed more immune infiltrations; higher levels of immunosuppressive checkpoints, such as CD274, tumor mutation burden, and resistance to chemotherapeutics; potentially better response to immune therapy; worse prognosis; and advanced stage of tumor node metastasis than the low-risk group. The combination of risk score and clinical characteristics was effective in predicting overall survival. Zhongshan cohort validated that high-risk score group correlated with malignant progression, worse prognosis, inferior adjuvant chemotherapy responsiveness of CRC, and shaped an immunoevasive contexture. This tool may provide a more accurate risk stratification in CRC and screening of patients with CRC responsive to immunotherapy.

KEYWORDS

colorectal cancer, mRNAsi, stemness, risk score model, immunotherapy, metabolism, immune evasion, Machine learning

Introduction

Colorectal cancer (CRC) is one of the deadliest cancers of the digestive system (1, 2). Although there is an increasing number of potential therapeutic approaches for CRC, such as surgery, chemotherapy, radiotherapy, and molecular targeted therapy, the clinical prognosis remains unsatisfactory, especially for patients with distant metastasis of CRC (3, 4). Therefore, accurate medical treatment is essential for the eradication of malignancy. At the same time, due to the high molecular heterogeneity of CRC, most existing biomarkers lack strong predictive accuracy (5). Hence, it has become an urgent problem to find a powerful index to predict and evaluate the clinical prognosis and therapeutic effect to achieve accurate clinical intervention.

Cancer stem cells (CSCs) play a crucial part in the progression, recurrence, and drug resistance of solid malignant tumors (6). Furthermore, CSCs promoted immunosuppression, immune escape, tumor metastasis, and therapeutic resistance by interacting with immune cells (7). For example, in a co-transplantation environment, CSCs can promote the polarization of $CD14^+$ peripheral monocytes into immunosuppressive M2 macrophages and the generation of tumorigenic myeloid cells, followed by the acceleration of tumor growth in immunocompromised mice (8). CSCs also drive the recruitment and polarization of TH17 cells and Treg cells by secreting CCL1, CCL2, CCL5, and TGF-β, resulting in an immunosuppressive environment (7). In recent years, the mRNA expression-based stemness index (mRNAsi) developed by a machine learning algorithm has been used to quantify the stemness characteristics of tumors (9) such as esophageal cancer (10), gastric cancer (11), hepatocellular carcinoma (12), and glioma (13). However, the risk score model for stemness features associated with immunological properties in CRC remains uninvestigated.

Metabolic reprogramming, a hallmark of cancer, is another important factor leading to antitumor immunity and immune escape. For example, excessive glycolysis in tumor cells produces a large amount of lactate, which leads to acidification of the microenvironment and, consequently, inhibits the proliferation and function of cytotoxic T cells. Other studies also showed that inhibition of mTOR or the glycolysis pathway regulated T-cell differentiation into naïve and memory phenotypes (14). Furthermore, when CAR-T cells were expanded *in vitro*, inhibition of AKT improved their metabolism and promoted their differentiation to the memory phenotype, thus improving the progression of acute lymphoblastic leukemia (15). Therefore, further elucidation of the effects of tumor stemness and metabolic characteristics on the immune microenvironment may provide significant clinical benefits.

In this study, we generated a new risk prediction model using mRNAsi and metabolism-related genes using CRC expression data retrieved from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases (n = 1323). We also determined the association between the risk score and several functional and clinical features of patients with CRC. Clinical prognosis, tumor microenvironment and immunophenotype, response to chemotherapy and immunotherapy, and genomic variation between two risk score groups were evaluated comprehensively. We then validated the mRNAsi-related metabolic risk score model using the Zhongshan Hospital cohort (n = 200). Our data aimed to provide new insights into the screening of patients more likely to benefit from immunotherapy, and to improve individualized treatments for CRC patients.

Methods

Data collection and processing

The expression profile data of colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) of 591 patients, and their clinicopathological annotation were retrieved from the TCGA GDC website (https://portal.gdc.cancer.gov/). TPM values were converted from FPKM. Furthermore, tumor mutation burden (TMB) was obtained by analyzing the copy number variation (CNV) and somatic mutation data using the maftools package of R.

In addition, CRC gene expression data of GSE17536 (16–18) and GSE39582 (19) and the clinicopathological features of the patients were also downloaded from the GEO database. GSE17536 included 177 CRC tissue samples, and GSE39582 included 555 CRC tissue samples. Subsequently, the TCGA and GEO data were merged (n = 1323), and the limma (20) R package and sva (21) R package were used to combine and eliminate any batch effect.

Analysis of mRNAsi and differentially expressed genes

Based on the relative expression data provided by Zheng et al. (22), the mRNAsi of each sample was determined conforming to the gene expression matrix by the single-sample gene-set enrichment analysis (ssGSEA) method using the R-package GSVA. According to the mRNAsi of each sample obtained, combined with the survival status of the patients, the best cut-off value of mRNAsi was set, and the patients with CRC were distributed into the high-mRNAsi group and low-mRNAsi group.

DEGs between the high-mRNAsi groups and low-mRNAsi groups in patients with CRC were analyzed using the "limma" R package. DEGs were defined as genes with Log2 (fold change) > 1.0 and $P < 0.05$. Metabolism-related gene sets were copied from

the Molecular Signature Database (MSigDB) V7.0 (22). Finally, the overlap between DEGs and metabolism-related genes resulted in the identification of metabolism-related DEGs.

Weighted gene co-expression network analysis

WGCNA was achieved using the WGCNA package in R, which aims to determine the correlation between genes by building important modules. First, a scale-free gene co-expression network was constructed according to the weight of the correlation coefficient, and a hierarchical clustering tree was established depending on the adjacency matrix of the network. The module significance (MS) was then calculated to judge the correlation between the mRNAsi value and different modules. The genes in each module were recorded and defined as module characteristic genes. Modules with maximum and minimum MS values were regarded as positive and negative modules, respectively. After selecting the modules of interest according to the MS values, all gene expressions in the modules were identified as genes highly correlated with mRNAsi.

Construction of mRNAsi-related metabolic risk score model

By integrating the results of metabolism-related DEG analysis and WGCNA, mRNAsi-related metabolic genes were finally obtained. Significantly differentially expressed mRNAsi-related metabolic genes were included in the model, dimensionality reduction analysis was performed using the minimum absolute contraction and selection operator (least absolute shrinkage and selection operator, LASSO) algorithm, and the characteristic genes related to prognosis were obtained. Using the normalized gene expression value weighted by the penalty coefficient obtained by LASSO Cox analysis, a risk score formula was established, and patients were divided into high-risk group and low-risk group according to the median risk score.

$$Risk\ score = \sum_i Coefficient\ (hub\ gene_i) \times mRNA\ Expression\ (hub\ gene_i)$$

Functional and pathway enrichment analyses

Gene Ontology (GO) analysis is a widely used method for functional enrichment studies and generates data related to biological processes (BP), molecular functions (MF), and cellular components (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database for systematic analysis of gene function, linking genome information with more orderly biological function information. The clusterProfiler package of R (23) was used for GO analysis and KEGG pathway enrichment in the mRNAsi-related metabolic risk score model. FDR < 0.05 was regarded as significant.

To investigate differences in BP between different groups, we employed gene-set enrichment analysis (GSEA) (22). The "h.all.v7.2.symbols.gmt" gene set was copied from the MSigDB for the GSEA. $P < 0.05$ was considered significant.

Molecular network analysis

The STRING database (https://cn.string-db.org) (24) was used to construct a protein-protein interaction (PPI) network. Genes with scores greater than 0.4 were chosen to build a network model, which was visualized using Cytoscape (v3.7.2) (25). Then, eight hub genes were selected using the CytoHubba plug-in (26) in the Cytoscape software. Furthermore, we use the GOSemSim package in R (27) to judge the GO semantic similarity of the eight genes (28).

Information regarding miRNA-mRNA interactions from the miRTarBase database was downloaded before analyzing the basic statistics. Based on the core mRNA obtained by PPI analysis, the miRTarBase database was used to predict the miRNAs that may be regulated and to further predict the related lncRNAs. Cytoscape software was used to visually display the results of ceRNA analysis.

Analysis of tumor immune infiltrating cells

An ssGSEA algorithm was deployed to measure the relative number of tumor-infiltrating immune cells in patients with CRC (29). The enrichment score calculated by ssGSEA using the GSVA R package (30) indicates the in level of each immune cell type in each sample. In addition, depending on the gene expression profile, the ESTIMATE R package (31) was used to quantify the level of immune infiltration of tumor samples, and the immune score of each tumor sample was obtained. The differences in the immune infiltration characteristics of CRC patients between the high-risk score group and low-risk score groups were evaluated.

Analysis of drug sensitivity and immunotherapy response

The Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) is an open database for molecular therapy and mutation exploration in cancer. The pRRophetic package of R (32) was used to download the cell

line gene mutation data and the IC50 values of different anticancer drugs from GDSC (33) and to analyze the correlation between patients with high and low-risk scores and different anticancer drug sensitivities.

In addition, we used online tumor immune dysfunction and exclusion (TIDE) scores (34) to examine immunotherapy sensitivity and compare the scores of tumor immunotherapy markers, such as CD8 and CD274, between the high-risk score groups and the low-risk scoring groups. The response of immune-checkpoint blockade was predicted.

CNV analysis

To analyze the changes in copy number in different risk score groups of patients with TCGA-CRC, we used the TCGAbiolinks package of R to obtain the masked copy number segment data of the patients. The downloaded CNV fragments were analyzed using GISTIC 2.0, with default settings in GenePattern5. Finally, the analysis results of GISTIC 2.0 were visualized through the maftools package of R.

Establishment of a prognostic model

Univariate and multivariate Cox analyses were used to predict the overall survival (OS) of patients with CRC. The clinicopathological features were then incorporated into the risk score model to construct a clinical predictive nomogram. To quantify the differential performance of the nomogram, Harrell's consistency index (C-index) was estimated. A calibration curve was produced and the capability of the nomogram was evaluated.

Patients and CRC tissue samples

The Zhongshan Hospital cohort included 200 patients who underwent CRC surgery between January 2008 and December 2014. The patients' baseline characteristics included sex, age, adjuvant chemotherapy, tumor location, tumor histology, tumor differentiation, nerve invasion, surgical margin positivity, and stage of tumor node metastasis (TNM). Tumor staging was performed according to the 7th edition of the American Joint Commission on Cancer (AJCC) TNM Classification (35). Conforming to the National Comprehensive Cancer Network guidelines and patient wishes, patients with stage III-IV TNM were treated with ACT after surgery. OS was described as the time from the date of diagnosis to death or last follow-up. Disease-free survival (DFS) was described as the time from the date of diagnosis to relapse or last follow-up. The follow-up period ended on December 31, 2020. Clinical data validation was approved by the ethics committee of the Zhongshan Hospital (B2022-068R2).

RNA separation and quantitative reverse transcription PCR

The mRNA expression of 21 mRNAsi-related metabolic genes was measured by qRT-PCR in Zhongshan cohort. Total RNA was obtained using the TRIzol reagent (Invitrogen, Waltham, MA, USA). cDNA was obtained by reverse transcription using the PrimeScript RT kit (Takara). The expression of candidate genes and the housekeeping gene GAPDH was evaluated by quantitative reverse transcription PCR using the ABI 7900HT real-time PCR system (Applied Biosystems, Carlsbad, CA, USA). Relative transcription levels were calculated using the ΔΔCt method (36). The primer sequences used are listed in Supplementary Table 7.

Immunohistochemical staining

We randomly selected 20 cases from 200 Zhongshan patients for IHC, including 10 cases in high risk group and 10 cases in low risk group. Paraffin-embedded tissues were stained with antibodies. The staining score was decided by two experienced pathologists at the Zhongshan Hospital. Six high-power fields (HPFs, ×200 magnification) were randomly counted by two independent pathologists (each with three fields), and the densities of $CD8^+$T cells, $Foxp3^+$Tregs, $CD19^+$B cells, CD11c dendritic cells, immunosuppressive checkpoints (PD-1, PD-L1) and effector molecules (GZMB, PRF1) were recorded. Immunohistochemistry antibodies are listed in Supplementary Table 8.

Statistical analysis

All data processing and analyses were accomplished using the R software (version 3.6.2) and SPSS (version 25; IBM, Armonk, USA). For the comparison of two groups of continuous variables, the statistical significance of normally distributed variables was calculated using an independent *t*-test, and the difference between non-normally distributed variables was measured using the Mann–Whitney U test. Chi-square test or Fisher's exact test was used to analyze the significant differences between the two groups of classified variables. The survival package in R was conducted for survival analysis. The receiver-operating characteristic (ROC) curve was drawn by the pROC package of R (37) and the area under the curve (AUC) was calculated to evaluate the performance of the risk score model. Univariate and multivariate Cox analyses were used to determine independent prognostic factors. All statistical P values were bilateral, and $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ were regarded as statistically significant.

Results

Relationship between colorectal cancer stemness characteristics and clinical features

A flowchart of this study is shown in Figure 1A. To explore the role of mRNAsi on the progression of CRC, including COAD and READ, the gene expression matrices of GSE17536 and GSE39582 datasets and TCGA database were downloaded (Supplementary Figures 1A, B). The data from the two databases were then merged (n = 1323) and cleaned from any batch effect (Supplementary Figures 1C, D).

First, to explore the correlation between mRNAsi and clinical characteristics, we determined CRC mRNAsi using the ssGSEA algorithm. Then, according to the optimal mRNAsi cut-off value, the patients with CRC were separated into high-mRNAsi and low-mRNAsi groups. The relationship between CRC stemness characteristics and clinical characteristics is shown in Figures 1B-E. No significant correlation between mRNAsi and age ($P = 0.56$) or gender ($P = 0.54$) was observed (Figures 1B, C). However, higher mRNAsi were associated with staging of TNM (stage 2 vs. stage 3, $P = 0.025$; stage 2 vs. stage 4, $P = 0.02$; Figure 1D). Furthermore, patients with high mRNAsi showed a significant increase in OS compared to those with low mRNAsi (log-rank $P < 0.001$, Figure 1E).

Identification of mRNAsi-related metabolic genes in patients with CRC

To determine the role of the mRNAsi in metabolic processes in CRC, DEGs between the high-mRNAsi and low-mRNAsi groups were identified and intersected with a metabolic gene set

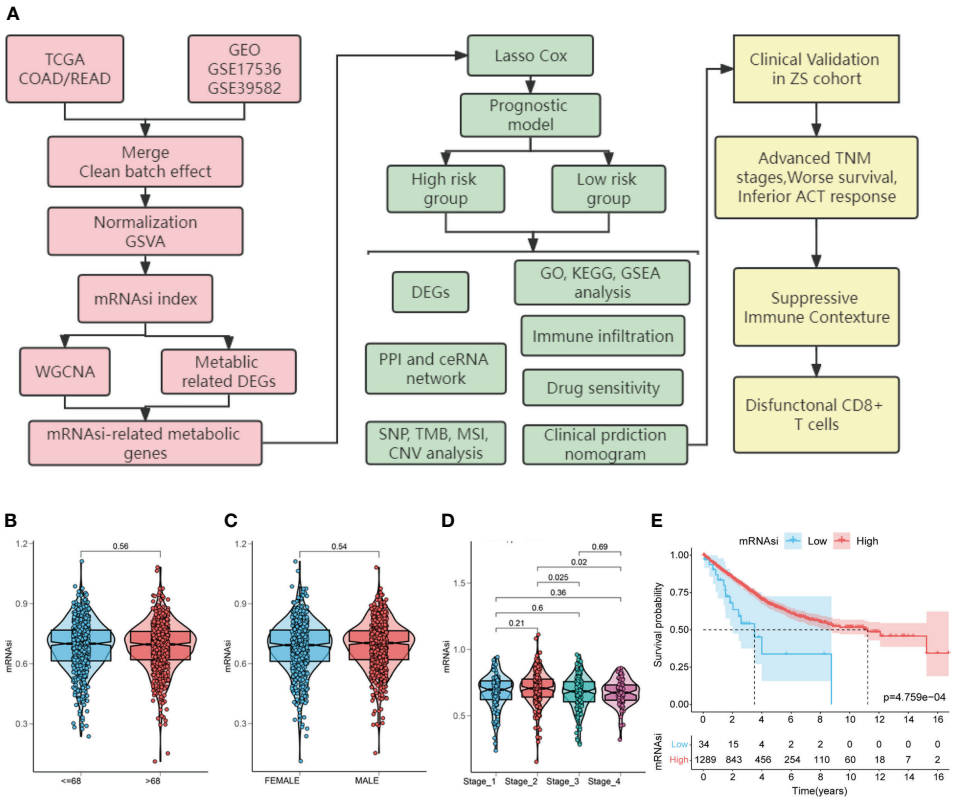


FIGURE 1
Study flow chart and the relationship between colorectal cancer (CRC) stemness characteristics and clinical features. **(A)**. Flow chart for construction and validation of mRNAsi-related metabolic risk score model in CRC. TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus databases; mRNAsi, mRNA expression-based stemness index; DEGs, differentially expressed genes; WGCNA, weighted gene co-expression network analysis; GO: gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene-set enrichment analysis; PPI, protein-protein interaction; ceRNA, competing endogenous RNAs; SNP, single nucleotide polymorphism; TMB, tumor mutation burden; MSI, microsatellite instability; CNV, copy number variation; TNM, tumor node metastasis; ACT, adjuvant chemotherapy. **(B–E)**. Relationship between CRC stemness characteristics and clinical features. Analysis of the correlation of mRNAsi with age **(B)**, gender **(C)**, TNM stage **(D)** and overall survival **(E)** in patients with CRC.

(2752 genes). One hundred and twenty-six genes were obtained and labeled as metabolism-related DEGs, of which 108 genes were significantly upregulated, and 18 genes were significantly downregulated (Figures 2A, B and Supplementary Figure 3A). WGCNA was used to identify modules closely related to mRNAsi-related genes. A total of 22 co-expression modules were identified, with the black module showing the strongest correlation with mRNAsi in CRC (Figures 2C, D). All genes in the black module were intersected with the metabolism-related DEGs, and 83 mRNAsi-related metabolic genes were obtained for further analysis, as shown in the Venn diagram (Figures 2E, F and Supplementary Figure 3B).

Construction of mRNAsi-related metabolic risk score model

To quantitatively evaluate the predictive value of identified mRNAsi-related metabolic genes in the clinical prognosis of CRC, we constructed a risk score model based on these genes. First, the expression characteristics of the 83 mRNAsi-related metabolic genes were included in the LASSO Cox analysis and 21 genes with the optimal predictive value were selected (Figures 3A, B). Simultaneously, a risk score formula was established based on the normalized expression of important

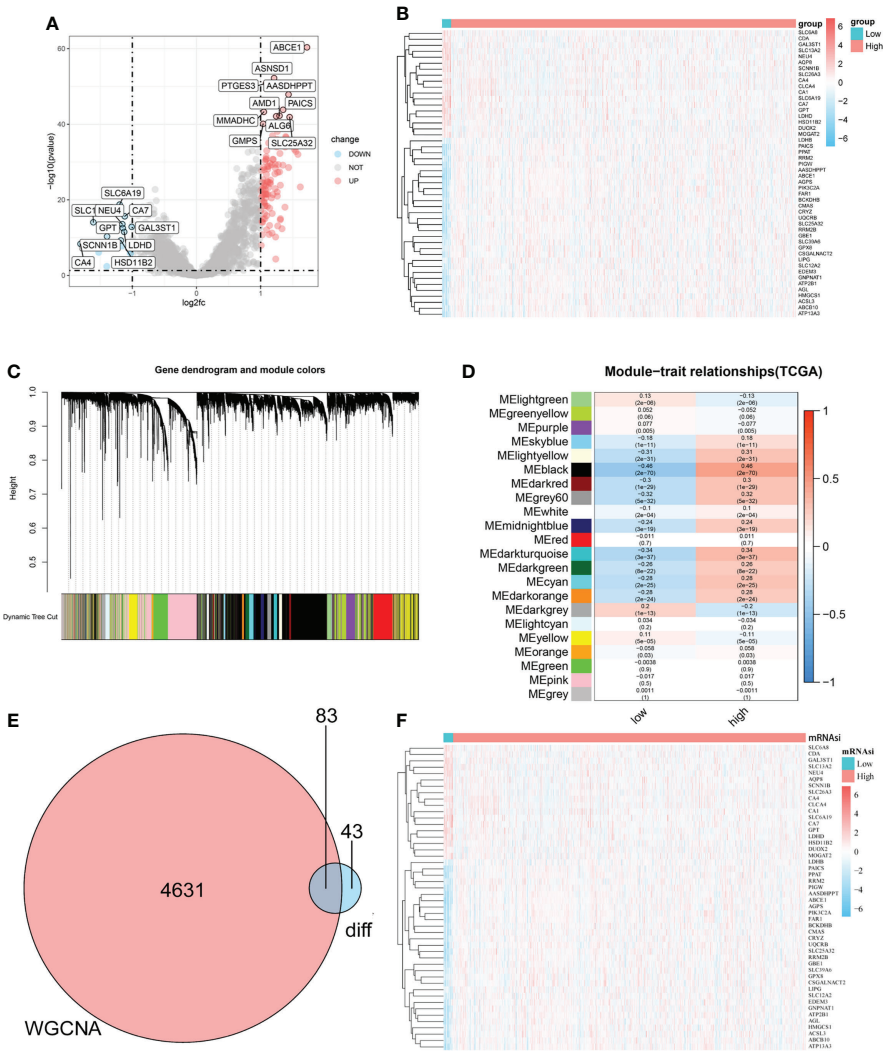


FIGURE 2
Identification of mRNAsi-related metabolic genes in patients with CRC. **(A–B)**. Volcano plot and heatmap showing the expression of metabolism-related DEGs in patients with CRC. **(C)**. Genes with similar expression patterns were merged in the same module to create a hierarchical cluster tree. **(D)**. Correlations and significant differences between different gene modules and mRNAsi and P values are displayed in the module. **(E)**. All genes in the black module, which were more closely related to mRNAsi, were overlapped with metabolism-related genes and 83 candidate genes were obtained, which were defined as mRNAsi-related metabolic genes. **(F)**. The heatmap shows the expression of 83 significantly differentially expressed mRNAsi-related metabolic genes in the CRC and normal tissues.

characteristic genes weighted by the penalty coefficient calculated by LASSO Cox analysis, and a risk score for each sample was calculated. An example of the formula used to calculate the risk score is given below.

$$\text{Risk score} = (-0.1053) \times \text{PTGES3} + (-0.1874) \times \text{PAICS} + (-0.0133) \times \text{GNPNAT1} + (0.02893) \times \text{PGM3} + (0.08862) \times \text{MTHFD2} + (-0.0043) \times \text{DCK} + (0.06428) \times \text{MTAP} + (0.22468) \times \text{SLC25A36} + (0.00308) \times \text{GBE1} + (-0.0679) \times \text{RRM2} + (0.0029) \times \text{KCTD3} + (-0.0436) \times \text{ACADSB} + (-0.0339) \times \text{ABCD3} + (-0.0137) \times \text{BCKDHB} + (-0.0525) \times \text{PHOSPHO2} + (-0.1185) \times \text{FUT4} + (0.00089) \times \text{EDEM3} + (-0.0684) \times \text{NEU4} + (0.6165) \times \text{SLC16A1} + (-0.0162) \times \text{ELOVL7} + (0.04312) \times \text{SLC6A8}$$

Then we performed the time-dependent ROC curve analysis and found that the model had appropriate accuracy in predicting OS in patients with CRC, and the AUC of 1-year, 2-year and 3-year OS was 0.647, 0.644, and 0.672, respectively (Figure 3C). Kaplan–Meier analysis showed a reduction in OS in patients with high-risk scores (log-rank $P < 0.001$; Figure 3D). In addition, there was a significant negative correlation between mRNAsi and risk scores (Rho = -0.2, $P < 0.001$, Figure 3E). The distribution of the risk score, survival status, and expression pattern of characteristic genes is shown in Figure 3F.

GSEA, GO, KEGG analyses of DEGs between high-risk and low-risk patients in mRNAsi-related metabolic risk score model

To analyze the impact of mRNAsi-related metabolic risk score models on the occurrence and development of CRC, we

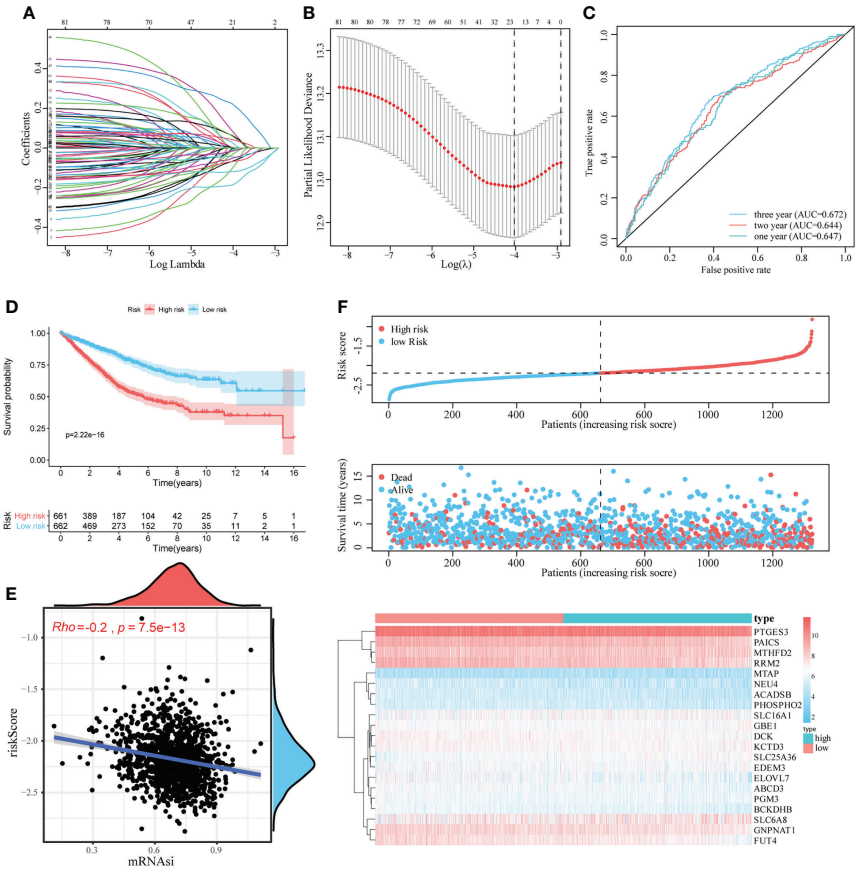


FIGURE 3
Construction and evaluation of the mRNAsi-related metabolic risk score model. **(A, B)**. LASSO Cox analysis identified 21 genes most associated with OS in the TCGA dataset. **(C)**. Time-dependent ROC curve analysis of risk score. **(D)**. The effect of the risk score assessed by Kaplan–Meier curve on the overall survival rate of patients with CRC. **(E)**. Spearman rank correlation analysis was used to analyze the relationship between mRNAsi and risk score. **(F)**. The risk score distribution, survival status, and heatmap of characteristic gene expression in patients with CRC.

used the median LASSO Cox risk score of CRC cases from TCGA dataset and divided the CRC cases into high-risk and low- risk score groups. There were 242 DEGs between high-risk and low-risk patients (Log2 (fold change) > 1.0 and $P < 0.05$), of which 195 were significantly upregulated, and 47 were significantly downregulated (Figures 4A, B). The correlation between the risk score with the clinical characteristics of CRC patients in the TCGA and GEO database is shown in Supplementary Table 1. The functional annotation of the GO showed the DEGs were closely related to several BP, including the organization of the extracellular matrix, the organization of the extracellular structure and ossification, as well as several MF such as extracellular matrix structural constituent, glycosaminoglycan binding, and extracellular matrix structural constituent conferring tensile strength (Figure 4C, Supplementary Table 2). KEGG analysis indicated that DEGs were particularly involved in focal adhesion, phagosome, protein digestion and absorption, complement and coagulation cascades, and ECM-receptor interaction pathways (Figure 4D and Supplementary Table 3). Two pathways, protein digestion and absorption ($P = 2.82E^{-08}$) and phagosome pathways ($P = 2.96E^{-07}$), which were highly related to the mRNAsi-related metabolic risk score model, are shown in Figures 4E, F.

Furthermore, GSEA showed that ascorbate and aldarate metabolism (NES = -1.89, $P = 0.002$), citrate cycle (TCA cycle) (NES = -2.28, $P = 0.002$), glyoxylate and dicarboxylate metabolism pathways (NES = -2.03, $P = 0.002$), propanoate metabolism (NES = -2.07, $P = 0.002$), arginine and proline metabolism (NES = -1.94, $P = 0.002$), pyruvate metabolism (NES = -2.18, $P = 0.002$), hallmark oxidative

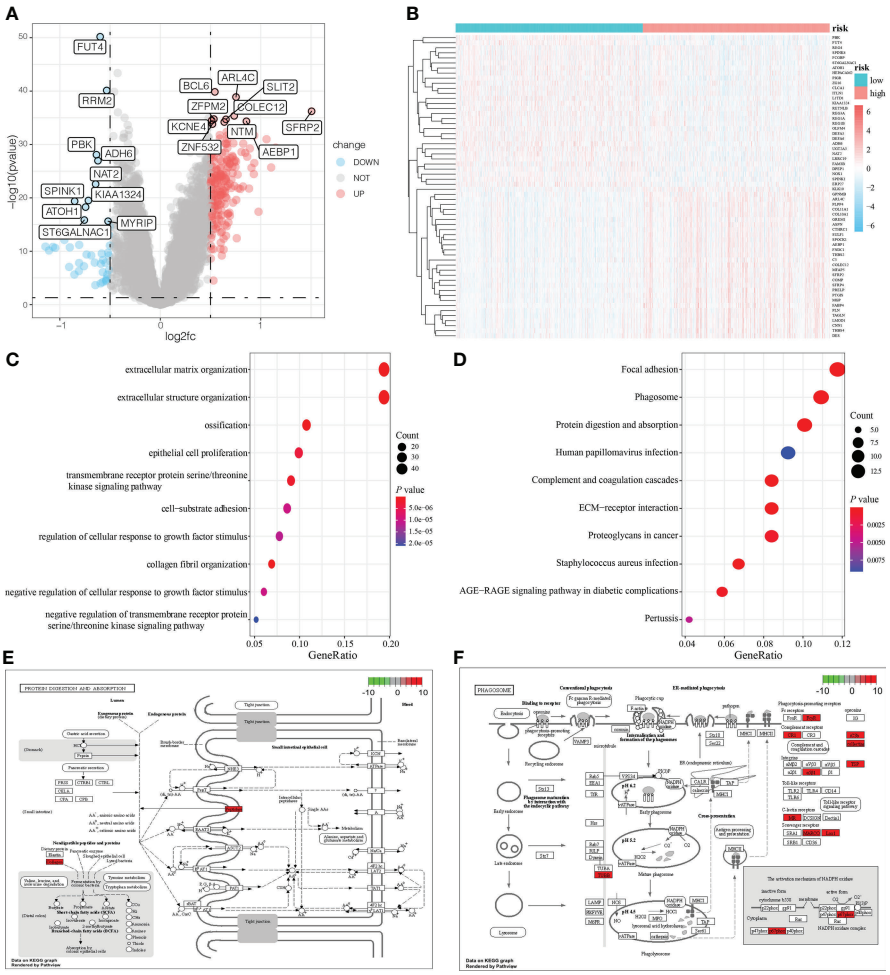


FIGURE 4
DEG analysis and functional enrichment analysis based on the mRNAsi-related metabolic risk score model. **(A, B)**. Volcano plot and heatmap showing the expression pattern of DEGs in two groups of patients with CRC. **(C, D)**. Biological processes (BP) and KEGG pathway analysis of DEGs in two groups of patients with CRC. **(E, F)**. The two pathways are closely related to the mRNAsi-related metabolic risk score model: protein digestion and absorption, and the phagosome pathway.

phosphorylation (NES = -2.86, $P < 0.001$), and hallmark fatty acid metabolism (NES = -1.93, $P < 0.001$), were abundant in low-risk patients, whereas hallmark hypoxia was significantly enriched in high-risk patients (NES = 2.04, $P < 0.001$), (Supplementary Figures 2A–I and Supplementary Table 4).

Construction of PPI network and related regulation network

We used the STRING database to establish the PPI network between DEGs, and imported the interaction between genes into Cytoscape software to obtain Figure 5A, in which the upregulated genes were represented in red and the downregulated genes were represented in blue.

The hub genes were analyzed using Cytoscape software (Figure 5B). GO semantic similarity analysis showed that the *CALD1* gene played an important role in the hub genes (Figure 5C). Subsequently, we conducted correlation analysis between hub genes and mRNAsi, and found a significant co-expression pattern between hub genes, whereas the relationship of each hub gene and mRNAsi was not consistent (Figure 5D). Finally, based on information about miRNA-mRNA interaction downloaded from the miRTarBase database; the hub genes obtained *via* the PPI network were used to construct the ceRNA network of miRNA-mRNA–lncRNA interaction (Figure 5E).

Immune contexture difference between high-risk and low-risk patients

We then evaluated the immune contexture heterogeneity between the high and low-risk score groups. As shown in Figure 6, the immune and stromal scores of the high-risk score group were significantly higher than those in the low-risk score group (both $P < 0.001$, Figures 6A, B). In addition, to evaluate the degree of immune cell infiltration in tumor tissue, we used the ssGSEA algorithm and obtained the relative enrichment scores of 28 subtypes of immune cells between the two groups, as shown in the heatmap in Figure 6C. The correlation analysis showed that the infiltration levels of most immune cells were positively correlated (Figure 6D). Further analysis revealed that infiltration of $CD4^+$T cells, $CD8^+$ T cells, B cells, dendritic cells, eosinophils, mast cells, macrophages, myeloid-derived suppressor cells (MDSCs), natural killer cells, regulatory T cells, and T helper cells was higher in the high-risk score group (Figure 6E). In addition, in this study the expression of HLA family members and several immunotherapy-related target genes, such as *CD274 (PD-L1)*, *CTLA-4*, and *LAG-3*, was

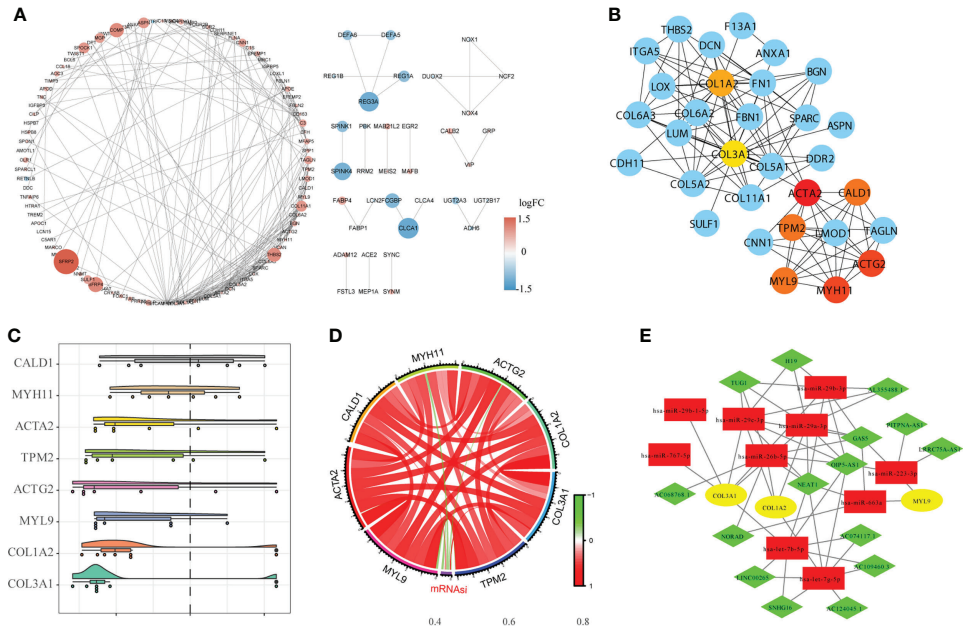


FIGURE 5
Construction of the protein-protein interaction network (PPI) and the ceRNA network. **(A)**. The results of PPI analysis were introduced into Cytoscape software for analysis, in which red represents upregulated genes, blue represents downregulated genes, and color depth and node size were positively correlated with log fold change (FC). **(B)**. The CytoHubba algorithm was used to identify and extract the top eight genes from the PPI network as the hub genes. **(C)**. GO semantic similarity analysis of importance of the eight hub genes. **(D)**. Circle diagram of the correlation between the hub gene and mRNAsi. **(E)**. Construction of the ceRNA interaction network based on hub genes.

elevated in the tumor environment of the high-risk group compared with the low-risk group (Figures 6F, G).

Sensitivity to chemotherapy and immunotherapy in high-risk and low-risk patients with CRC

To analyze the differences in the sensitivity of patients with CRC to different drugs and small-molecule drugs based on the risk score, we downloaded the CRC cell line gene mutation data and the half-maximal inhibitory concentration (IC50) values of several anticancer drugs from the GDSC database. In GDSC, IC50 values for patients with CRC were predicted based on the responses of cell lines to 138 chemotherapeutic agents and small-molecule anticancer agents. This suggested that patients in the high-risk score group were less susceptible to multiple chemotherapeutic and small-molecule anticancer drugs, including Metformin, PF.4708671, Sorafenib, Mitomycin, Methotrexate, and gemcitabine (Figure 7A, all $P < 0.05$).

Because of the important role of immune-checkpoint inhibitor (ICI) therapy in tumors, we examined the sensitivity of two groups of patients with CRC to ICI therapy using the TIDE algorithm, which models two mechanisms of immune-evasion: T-cell dysfunction and reduced T-cell infiltration, to predict the immunotherapy response. As shown in Figure 7B, although no significant differences in the scores for two immune markers CD8 and CD274 between the high-risk and low-risk score groups was observed, the TIDE score in the high-risk score group was lower than that in the low-risk score group, suggesting a better response to the ICI therapy in the high-risk score group than in the low-risk score group.

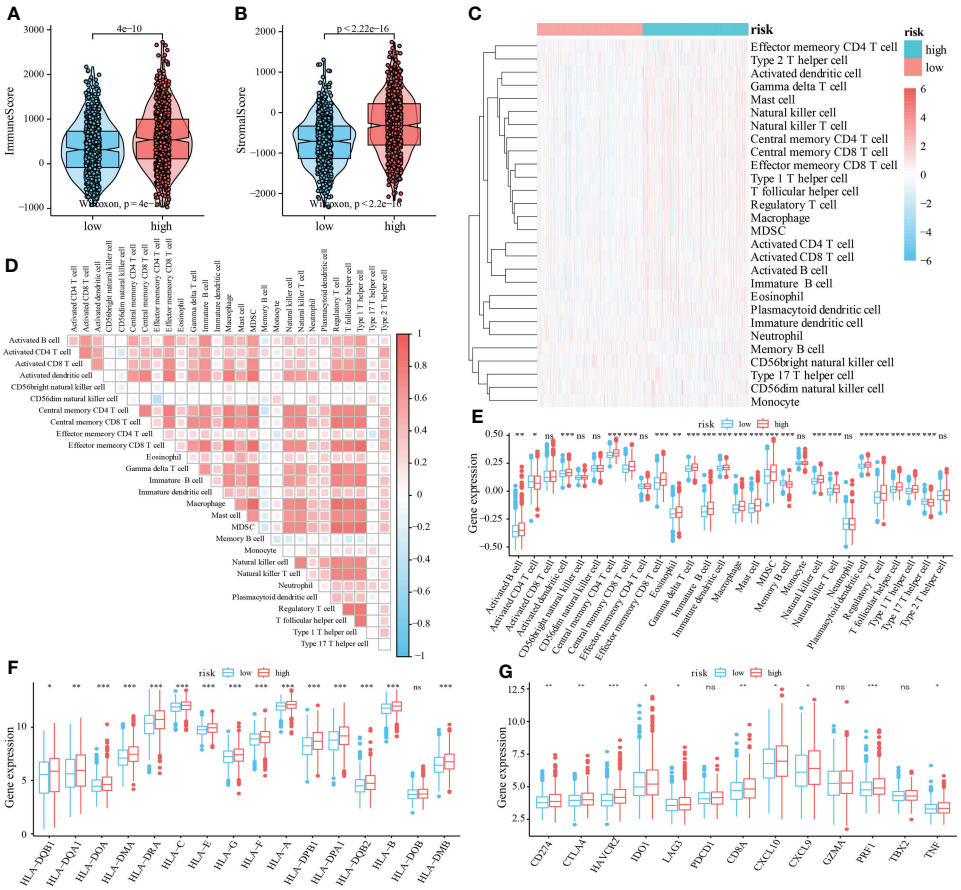


FIGURE 6
Relationship between mRNAsi-related metabolic risk score groups and infiltration of different immune cell subtypes. **(A, B)**. Differential analysis of immune scores and stromal scores between the high- and low-risk score group of patients with CRC. **(C)**. The heatmap showed the infiltration levels of 28 immune cell subtypes in CRC samples from TCGA and GEO datasets. **(D)**. Correlation heatmap showed the correlation between different levels of immune cell infiltration. **(E)**. Analysis of the difference of 28 levels of immune cell infiltration between two groups **(F)**. Multiple HLA family members, and **(G)** immunotherapy-related targets in high and low-risk score groups of patients with CRC. Differences were considered significant at $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, compared to the low-risk group. ns, not significant.

Analysis of genomic variation between high and low-risk score patients

Research has suggested that genomic variation affects tumor response to immunotherapy (38). Furthermore, we evaluated the differences in genomic variation in patients with CRC in the high and low-risk groups, including single nucleotide polymorphism (SNP), TMB, microsatellite instability (MSI), and CNV.

Difference in the level of the top SNP between the two groups was detected (Figure 8A). Furthermore, the TMB in the high-risk score group was higher than in the low-risk score group; however, no significant differences in MSI between the two groups were detected (Figures 8B, C). In addition, compared with the high-risk score group, the low-risk score group showed a significant increase in CNV, mainly characterized by deletion events (Figures 8D, E).

Construction and validation of clinical prediction nomogram based on mRNAsi-related metabolic risk score model

Next, we evaluated the association between mRNAsi-associated metabolic risk scores and clinicopathological characteristics in patients with CRC. The results showed no significant correlation between the risk scores and the age and gender of the patients (Figures 9A, B). However, high-risk scores were associated with lower mRNAsi and an advanced TNM state (Figures 9C, D).

In addition, univariate and multivariate Cox analyses showed that a high mRNAsi-related metabolic risk score was an independent predictor of prognosis in patients with CRC (Figure 9E and Supplementary Table 5). The mRNAsi-related metabolic risk score was then combined with different

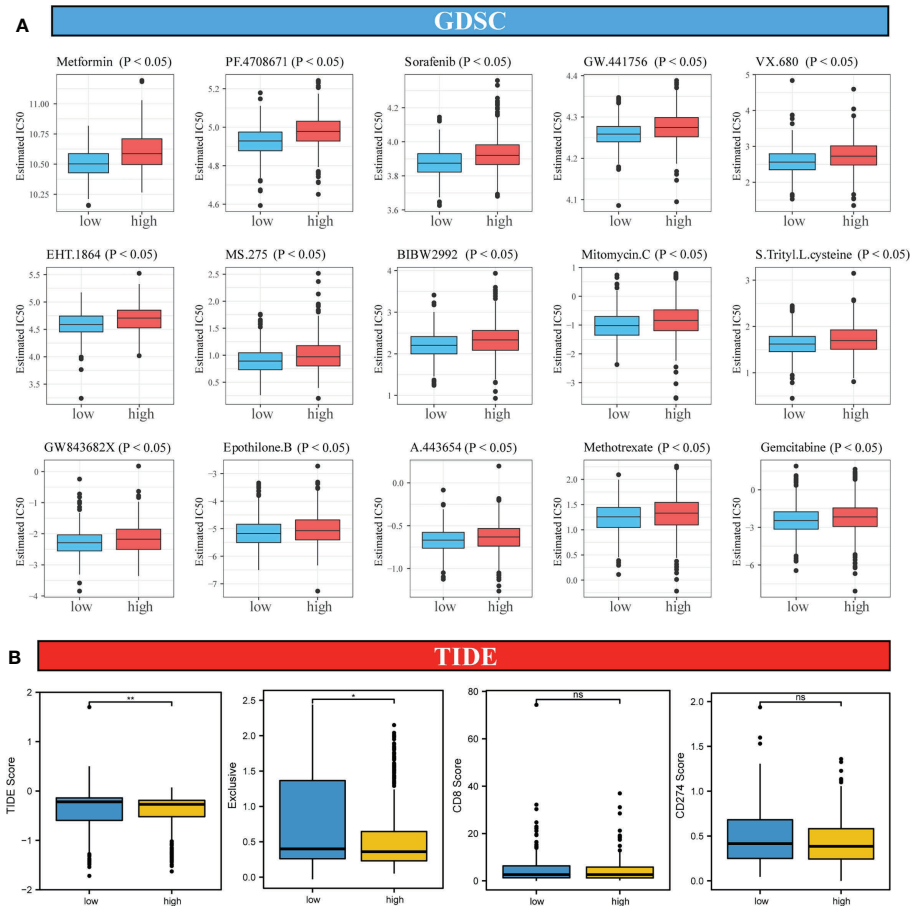


FIGURE 7
Analysis of sensitivity differences between high-risk and low-risk patients to different chemotherapeutic agents, small-molecule anticancer agents, and immunotherapy. **(A)**. Difference in sensitivity between high-risk and low-risk patients to 138 small-molecule anticancer agents and chemotherapeutic agents. **(B)**. Differences in TIDE score, immune exclusive, scores of immunotherapy targets, CD8 and CD274 in high-risk and low-risk score groups. Differences were considered significant at $^{*}P < 0.05$, $^{**}P < 0.01$, compared to the low-risk group. ns, not significant.

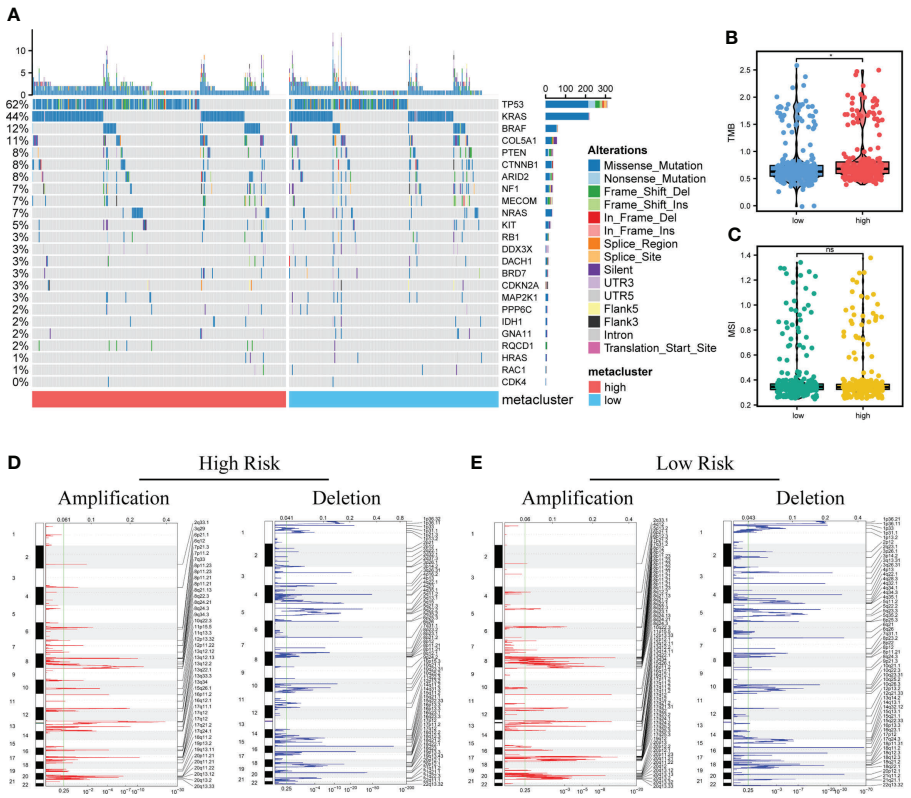


FIGURE 8
Analysis of genomic variation between high-risk and low-risk patients. **(A)**. Mutation profiles of common tumor-related genes in patients in high and low-risk score groups (Left: high-risk score group, right: low-risk score group). **(B, C)**. The difference of microsatellite instability (MSI) and tumor mutation burden (TMB) between two groups. **(D, E)**. Copy number variation in patients between two groups. Red indicates the amplified genes, and blue indicates the deleted genes. Differences were considered significant at $*P < 0.05$, compared with low-risk group. ns, not significant.

clinicopathological features to create a nomogram to predict OS in patients with CRC (Figure 9F). Then we preformed C-index to evaluate the differentiation of nomogram and found it has high discriminative ability (mean: 0.751 [range: 0.700–0.802]). In addition, the calibration curve showed that the 1-, 2-, and 3-year OS estimated by the nomogram matched with the actual OS values of the patients (Figure 9G).

High-risk score group correlated with malignant progression, worse prognosis, inferior adjuvant chemotherapy responsiveness of CRC

To further determine the clinical significance of the risk score model in CRC, we evaluated the correlation between the high and low-risk score groups and the clinicopathological characteristics of patients with CRC in the Zhongshan Hospital cohort. The mRNA expression of 21 mRNAsi-related metabolic genes was measured by qRT-PCR. The median expression level of risk score was used as the cutoff value. Patients were divided into high and low-risk score groups. The high-risk score group was positively correlated with right-sided colon, poorer differentiation, ucoid adenocarcinoma and signet-ring cell carcinoma, nerve invasion, surgical margin positivity, and higher TNM stage (all $P < 0.001$, Supplementary Table 6). These findings suggest the gene set defining the high-risk score group is potentially involved in tumor progression.

To investigate the association between the risk score model and long-term outcomes of patients with CRC, Kaplan–Meier analysis was performed. The high-risk score group predicted worse survival of patients with CRC in the Zhongshan Hospital cohort (OS: $P < 0.001$, log-rank = 13.102; DFS: $P < 0.001$, log-rank = 26.309; Figures 10A, B). These results indicate that the high-risk score was related to a worse outcome for patients with CRC.

In addition, we evaluated the interaction between the risk score model and therapeutic responsiveness to adjuvant chemotherapy (ACT) for TNM stage III-IV patients with CRC. In this study, ACT could improve patient survival in the

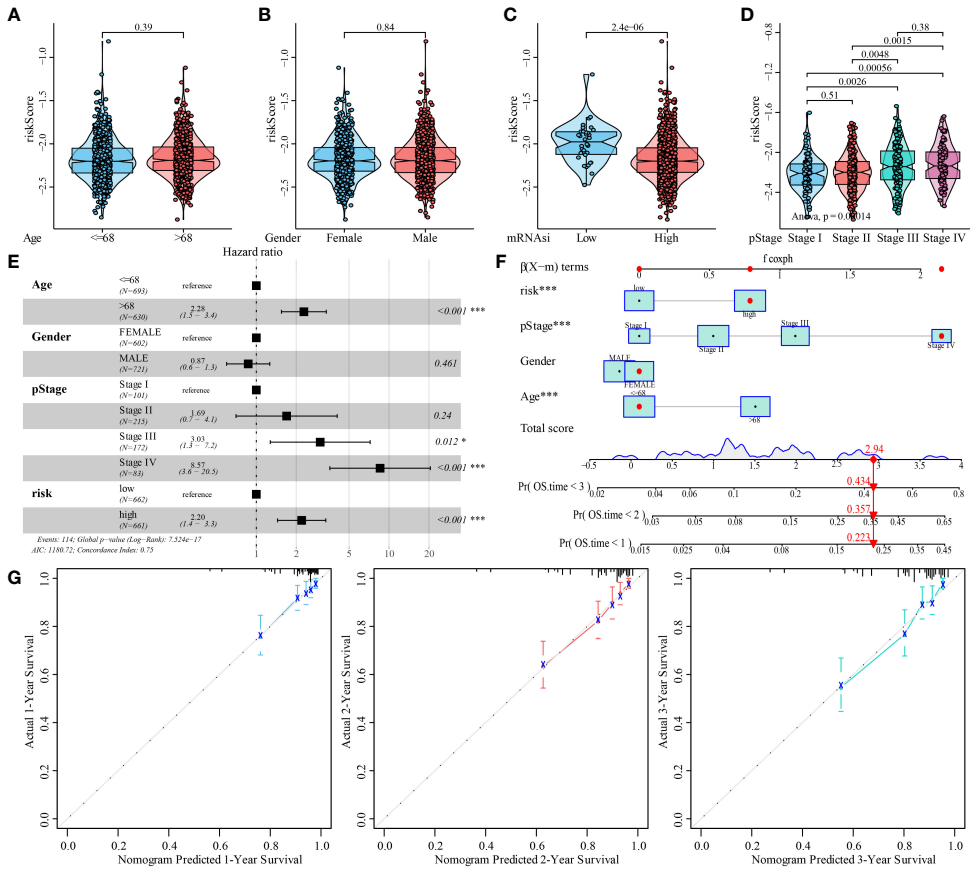


FIGURE 9
Analysis of the predictive ability of mRNAsi-related metabolic risk score model for the prognosis of patients with CRC. **(A–D)**. Analysis of correlation between mRNAsi-related metabolic risk scores and clinicopathological features of patients with CRC. **(E)**. Multivariate Cox regression analysis of HR and P values of risk score, combined with clinicopathological features. **(F)**. mRNAsi-related metabolic risk score combined with clinicopathological features to construct a clinical predictive model. **(G)**. The calibration curve of the nomogram showed that the risk score model had a good predictive ability for the overall survival rate of 1-, 2- and 3-year OS in patients. Differences were considered significant at $^{*}P < 0.05$, $^{***}P < 0.001$, compared to the reference.

low-risk score group (OS: $P = 0.043$, log rank =4.094, Figure 10C; DFS: $P = 0.005$, log rank = 7.860, Figure 10D) but had no significant benefit in the high-risk score group (OS: $P = 0.209$, log rank = 1.579, Figure 10E; DFS: $P = 0.413$, log rank = 0.670, Figure 10F). Therefore, these results suggest that the high-risk score group might have impaired therapeutic responsiveness to ACT in TNM stage III-IV CRC.

High-risk score group shaped immunoevasive contexture

To explore the underlying mechanism, we performed IHC staining of tumor-infiltrating immune cells in CRC tissues obtained from the Zhongshan Hospital cohort. The number of $CD8^{+}$T cells (P=0.0080), $CD19^{+}$B cells (P=0.0013), $Foxp3^{+}$Tregs (P<0.001), and CD11c dendritic cells (P=0.0028) was more abundant in the high-risk group (Figures 11A–H). But the ratio of $Foxp3^{+}$Treg cells to $CD8^{+}$T cells also increased markedly in the high-risk score group (P=0.029) (Figure 11I), suggesting a more immunosuppressive tumor microenvironment with increased Treg cell infiltration. We further investigated whether the high-risk score group could affect $CD8^{+}$T-cell function. The results indicated that $CD8^{+}$T cells in the high-risk score group showed an exhausted T-cell phenotype with increased expression of immunosuppressive checkpoints, programmed cell death protein 1 (PD-1) (P=0.0027) and programmed cell death-ligand 1 (PD-L1) (P=0.0013), and decreased expression of $CD8^{+}$T-cell effector molecules, granzyme B (GZMB)(P=0.0028) and perforin (PRF1) (P=0.0020), compared to the low-risk score group (Figures 11J-Q). Taken together, these data suggest that the high-risk score group may orchestrate an immunoevasive contexture and direct $CD8^{+}$T-cell dysfunction in CRC.

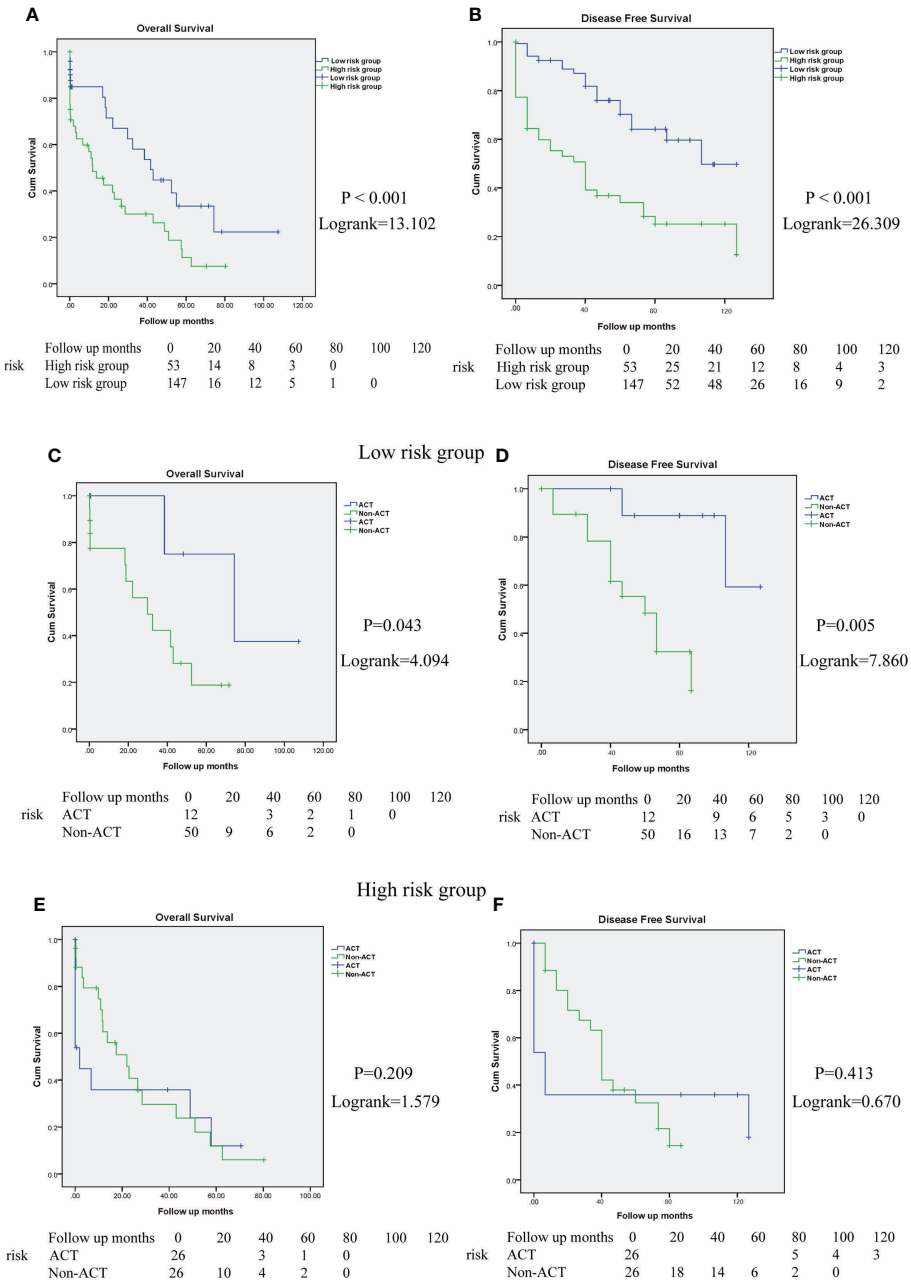


FIGURE 10
High-risk score group determines poor prognosis and impairs the ACT responsiveness of patients with CRC in Zhongshan cohort. **(A, B)**. Overall survival (OS) and disease-free survival (DFS) curves between high and low-risk score group in Zhongshan Hospital cohort. **(C, D)**. The OS and DFS curves for TNM stage III-IV patients with CRC in low-risk score group with or without ACT treatment. **(E, F)**. The OS and DFS curves for TNM stage III-IV patients with CRC in high-risk score group with or without ACT treatment.

Discussion

Tumor recurrence and drug resistance have always been obstacles to the treatment of CRC. Studies have shown that CSCs and metabolic reprogramming promote immunosuppression, immune escape, and therapeutic resistance by interacting with immune cells (39, 40). Therefore, by integrating differential expression analysis between high and low mRNAsi, metabolism-related genes, co-expression network analysis, and LASSO Cox regression analysis, 21 mRNAsi-related metabolic genes with the highest prognostic value were identified and used to construct the risk score model. After validation first in the

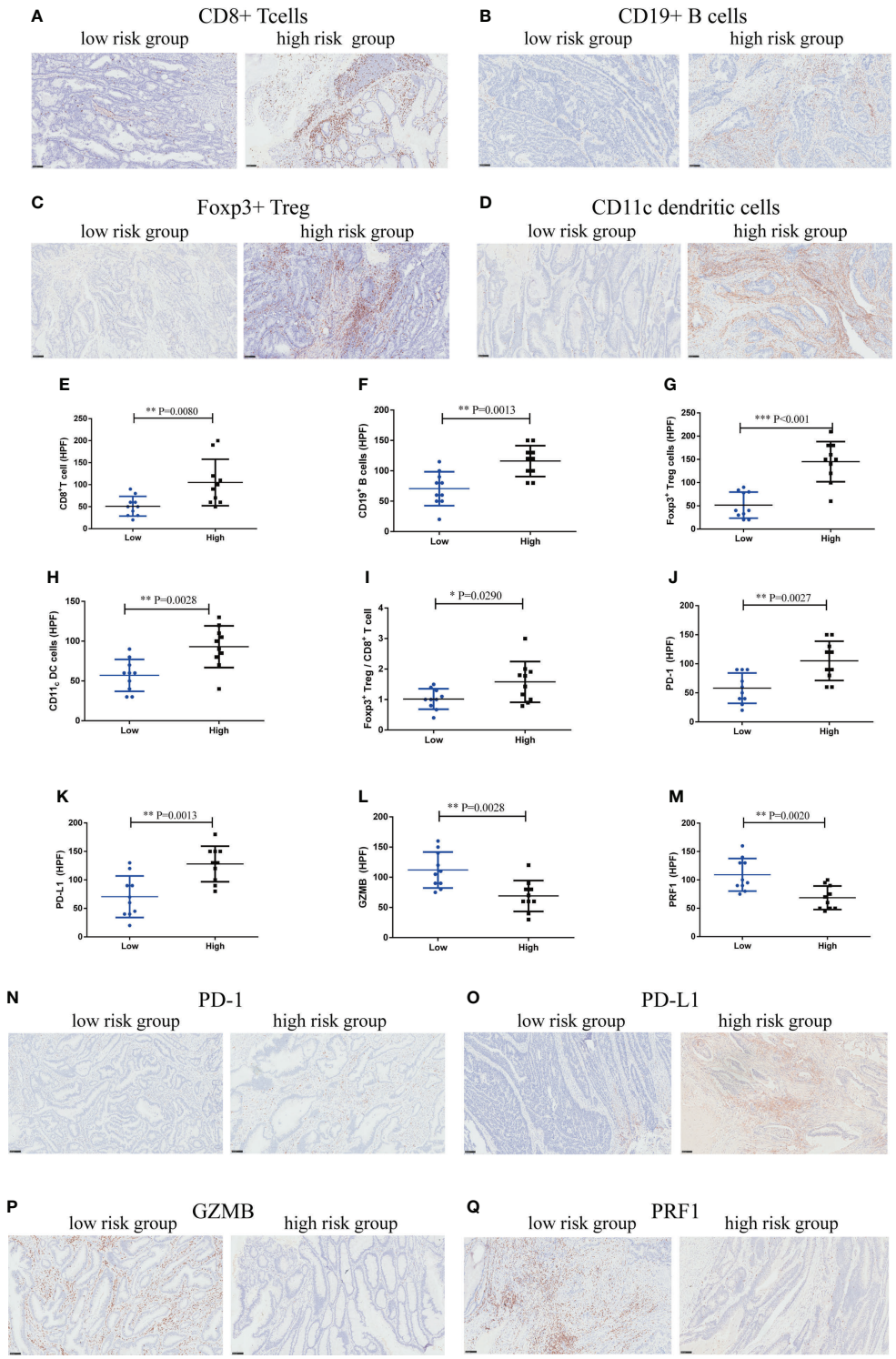


FIGURE 11
High-risk score group drives immunoevasive contexture and damages $CD8^+$ T-cell function in CRC in Zhongshan cohort. **(A–D)**. Representative immunohistochemical (IHC) staining of four significant tumor-infiltrating immune cell subtypes between high and low-risk score groups, including $CD8^+$ T cells, $CD19^+$ B cells, $Foxp3^+$ Tregs, $CD11_c$ dendritic cells. **(E–I)**. Comparison of $CD8^+$ T cells, $CD19^+$ B cells, $Foxp3^+$ Tregs, $CD11_c$ dendritic cells and the ratio of $Foxp3^+$ Tregs to $CD8^+$ T cells between two groups. **(J–M)**. Expression of immunosuppressive checkpoints (PD-1, PD-L1) and effector molecules (GZMB, PRF1) between two groups. **(N–Q)**. Representative IHC staining of immunosuppressive checkpoints (PD-1, PD-L1) and effector molecules (GZMB, PRF1) between two groups. n = 10 in each group, scale bar: 250um. Differences were considered significant at $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, compared to the low-risk group.

data retrieved from the TCGA and GEO databases, and then in the Zhongshan Hospital cohort, the CRC samples in the high-risk score group exhibited poor clinical outcome, increased immune-evasion, reduced sensitivity to chemotherapy, whereas potentially better response to immunotherapy, and higher genomic variation. This risk score model could be a tool to screen for patients with worse prognosis and inferior chemotherapy response, optimizing targeted treatment for CRC patients.

According to the GO functional enrichment results, most of these genes were clustered in functional groups related to the extracellular matrix (ECM). The ECM, as the main component of the tumor microenvironment, is considered to play a leading role in the progression of various cancers, including CRC, and promotes the invasion and metastasis of cancer (41–43). In addition, Ortensi et al. have demonstrated that the stemness characteristic of cancer tissue contributes to glioma invasiveness, which is closely related to the ECM (44). KEGG analysis also showed these genes were significantly abundant in inflammation, immunity, adhesion, invasion, and other processes. These results suggest that mRNAsi-related metabolic genes are involved in CRC metastasis and invasion.

Furthermore, GSEA revealed that hallmark hypoxia was significantly enriched in the high-risk score group. Our study and other studies have shown that hypoxia may promote glycolysis in CRC cells by activating the HIF-1α signaling pathway, thus promoting the proliferation and metastasis of CRC cells (45, 46). Hypoxia reduced the sensitivity of CRC to 5-fluorouracil chemotherapy (47). This suggests that mRNAsi-related metabolic genes participate in hypoxia-related pathways, leading to a poor prognosis and chemotherapy resistance in CRC.

Hub genes were obtained by PPI network construction and Cytoscape software analysis. GO semantic similarity analysis showed a key role for the *CALD1* gene. Li et al. found that CALD1 upregulated the expression of PD-L1 through the JAK/STAT signaling pathway and promoted malignant progression of bladder cancer (48). Several bioinformatics analyses and cellular studies have shown that CALD promoted the proliferation, metastasis, and invasion of CRC cells and is related to a reduction in OS (49). However, the exact mechanism of CALD involvement in CRC remains to be clarified.

Considering that the risk score model was derived from the stemness index and metabolism-related genes, which were significantly associated with antitumor immunity (13, 50), we further investigated the immune contexture heterogeneity between the high-risk and low-risk score groups. Tumor-infiltrating immune cell analysis showed that the high-risk score group had greater infiltration of CD8+ T cells, CD4+ T cells, B cells, Treg cells, dendritic cells, macrophages, MDSCs, neutrophils, regulatory T cells, and T helper cells, which is consistent with infiltrating immune cells from colorectal cancer in a highly inflammatory state (51). Meanwhile, Zhongshan cohort validated that $CD8^+$T cells, $CD19^+$B cells, $Foxp3^+$Tregs, and CD11c dendritic cells was more abundant in the high-risk group. But an increase in the ratio of $Foxp3^+$Treg cells to $CD8^+$T cells in high-risk group, which leads to an immunosuppressive microenvironment. Treg cells promote tumorigenesis and development by inhibiting adaptive anti-tumor immunity, which is the key mechanism of tumor immune escape (52). The ratio of Foxp3+Treg cells to CD8+T cells as is a better variable because they are more representative of the biological characteristics of infiltrating immune cells (53). The immune microenvironment of tumor is closely related to clinical outcome and drug resistance (54). Further immunostaining experiments confirmed $CD8^+$T-cell dysfunction with decreased levels of cytotoxic molecules (GZMB and PRF1) and increased the expression of immune checkpoints (PD-1 and PD L1) in the high-risk score group, resulting in a highly exhausted state and impaired immune function. The effect of tumor on immune cells can lead to T cell anergy or dysfunction, which promotes tumor escape and therapeutic drug resistance (55). Our results suggested that high-risk group induced an immunoevasive contexture and impaired antitumor immunity, which explains the poor clinical prognosis.

Immunotherapy is a novel cancer treatment approach. Although the effect is significant, only a fraction of the patients responds to the treatment (56, 57). In the TCGA and GEO databases, patients with CRC with high-risk scores had higher expression of immunotherapeutic molecules (such as PD-L1, CTLA4, HAVCR2 and LAG3), suggesting that CRC with high-risk scores may be more likely to be affected by the immune-checkpoint pathway, inhibit the antitumor immune response, and lead to deterioration of prognosis. Similarly, the TIDE algorithm predicted that patients with high-risk scores were more sensitive to ICI therapy. In the Zhongshan Hospital cohort, we confirmed that $CD8^+$T cells in high-risk score patients showed an exhausted T-cell phenotype with increased expression of the immunosuppressive checkpoint, PD-1 and PD-L1, compared with low-risk score patients. These findings also suggest that CRC patients with high-risk scores may clinically benefit from immunotherapy. The TMB score of patients with the high-risk score group is higher, suggesting that PD-1 blocking therapy has a certain curative effect on these patients (13). Low TMB is an important reason for patient resistance to immunotherapy (58). This provides a new approach for stratifying this subgroup of patients with CRC to identify those who may achieve a superior response to immunotherapy.

Because of the close relationship of the stemness index and chemotherapeutic drug resistance in cancer, we analyzed the predictive ability of the risk score to chemotherapeutic drug sensitivity and found that patients with high-risk scores were less susceptible to a variety of small molecular anticancer drugs and chemotherapeutic medicines in TCGA and GEO databases. In

the Zhongshan Hospital cohort, our findings suggest that the high-risk score group might have impaired therapeutic responsiveness to ACT in TNM stage III-IV CRC. Some studies have shown that in CRC, tumor stem cells lead to chemotherapy resistance by inhibiting antiapoptotic gene expression and reducing mitochondrial transcription initiation (59, 60). Due to the heterogeneity of CRC, the response of patients to chemotherapy is different, even at the same stage (61). This further suggests that patients with high-risk scores are potentially more suitable for immunotherapy than for traditional chemotherapy. However, the exact correlation between the risk score and the response to anticancer treatment needs to be further explored in a larger CRC cohort.

Finally, to improve clinical application, mRNAsi-related metabolic risk scores were combined with different clinicopathological characteristics to construct a prognostic nomogram and verify the predictive ability of the nomogram in TCGA and GEO datasets. In the Zhongshan Hospital cohort, a high-risk score was related to malignant progression and worse clinical outcomes in patients with CRC. This risk score model contains 21 important prognostic genes and has never been reported to identify the immunoevasive subgroup of patients with CRC in previous publications related to the CRC stemness index (62). Furthermore, this risk score model could help molecular typing and screening of differential subgroups to optimize personalized treatment and facilitate clinical translation.

Our study bears limitations. First, the mechanism of crosstalk between mRNAsi and metabolic reprogramming remains unclear, and further experimental studies are needed. Second, the clinical data of patients receiving immunotherapy in this study were limited, and the robust ability of the risk score model to predict immunotherapy responsiveness needs to be verified in a future larger immunotherapy cohort. Finally, although the new model and nomogram could accurately forecast the survival of patients with CRC in the TCGA and GEO databases and the Zhongshan Hospital cohort, more cell experiments, animal models, and clinical samples are needed to verify the value of this mRNAsi-related metabolic risk score model before developing immunotherapy strategies for subgroups of patients with CRC.

In this study, we proposed and validated a new risk score model according to 21 mRNAsi-related metabolic genes. The high-risk score group had a poorer clinical prognosis, inferior sensitivity to chemotherapy, a potentially better response to immunotherapy, and an immunoevasive environment, which sheds light on more accurate risk stratification and divides subgroups of patients with CRC for immunotherapy.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by This study was approved by the ethics committee of Zhongshan Hospital, Fudan University (B2022-068R2). The patients/participants provided their written informed consent to participate in this study.

Author contributions

MLW and CHM designed the study. MLW, TL, JZ, MMG, and WLZ, performed the study. MLW, WCG, CHS, QWL, YY, KN, and ZWZ analyzed the data. MLW and TL wrote the paper MLW, MLS and DZ revised the paper. All authors contributed to the article and approved the submitted version.

Funding

Our study was supported by the National Natural Science Foundation of China (No.82002538, 82072213); Shanghai Pujiang Talent Plan (No. 2020PJD013); Clinical Research Plan of SHDC (No. SHDC2020CR1005A); and National Key Research and Development Program of China (No. 2020YFC2008400).

Acknowledgments

We thank Rongkui Luo (Department of Pathology, Zhongshan hospital Fudan University) for providing technical consultation.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.950782/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Data collection and processing. Box plot of CRC samples in the GSE17536 **(A)** and GSE39582 **(B)** datasets. Box plots of TCGA, GSE17536, and GSE39582 datasets before **(C)** and after **(D)** merging and de-batching.

SUPPLEMENTARY FIGURE 2
GSEA of DEGs in high and low-risk score groups of patients with CRC. **(A–I)**. GSEA analysis of DEGs in high and low-risk score groups of patients with CRC.

SUPPLEMENTARY FIGURE 3
(A). The heatmap showing the expression of metabolism-related DEGs in patients with CRC. **(B)**. The heatmap shows the expression of 83 significantly differentially expressed mRNAsi-related metabolic genes in the CRC and normal tissues.

SUPPLEMENTARY TABLE 1
Association between risk score and clinical characteristics of patients with CRC.

SUPPLEMENTARY TABLE 2
GO analysis of DEGs in high and low-risk score groups of patients with CRC.

SUPPLEMENTARY TABLE 3
KEGG analysis of DEGs in high and low-risk score groups of patients with CRC.

SUPPLEMENTARY TABLE 4
GSEA of DEGs in high and low-risk score groups of patients with CRC.

SUPPLEMENTARY TABLE 5
Univariate and multivariate Cox analyses of OS prediction in patients with CRC based on mRNAsi-related metabolic risk score model.

SUPPLEMENTARY TABLE 6
Relationship between high and low-risk score groups and clinicopathological features of patients with CRC in Zhongshan Hospital cohort.

SUPPLEMENTARY TABLE 7
Primers used in reverse transcription PCR analysis.

SUPPLEMENTARY TABLE 8
Immunohistochemistry antibodies.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

2. Kyrochristos ID, Roukos DH. Comprehensive intra-individual genomic and transcriptional heterogeneity: Evidence-based colorectal cancer precision medicine. *Cancer Treat Rev* (2019) 80:101894. doi: 10.1016/j.ctrv.2019.101894

3. Zhai Z, Yu X, Yang B, Zhang Y, Zhang L, Li X, et al. Colorectal cancer heterogeneity and targeted therapy: Clinical implications, challenges and solutions for treatment resistance. *Semin Cell Dev Biol* (2017) 64:107–15. doi: 10.1016/j.semcdb.2016.08.033

4. Linnekamp JF, Wang X, Medema JP, Vermeulen L. Colorectal cancer heterogeneity and targeted therapy: a case for molecular disease subtypes. *Cancer Res* (2015) 75(2):245–49. doi: 10.1158/0008-5472.CAN-14-2240

5. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in colorectal cancer: A challenge for personalized medicine? *Int J Mol Sci* (2018) 19(12):3733. doi: 10.3390/ijms19123733

6. Yi Y, Hsieh IY, Huang X, Li J, Zhao W. Glioblastoma stem-like cells: Characteristics, microenvironment, and therapy. *Front Pharmacol* (2016) 7:477. doi: 10.3389/fphar.2016.00477

7. Bayik D, Lathia JD. Cancer stem cell-immune cell crosstalk in tumour progression. *Nat Rev Cancer* (2021) 21(8):526–36. doi: 10.1038/s41568-021-00366-w

8. Yamashina T, Baghdadi M, Yoneda A, Kinoshita I, Suzu S, Dosaka-Akita H, et al. Cancer stem-like cells derived from chemoresistant tumors have a unique capacity to prime tumorigenic myeloid cells. *Cancer Res* (2014) 74(10):2698–709. doi: 10.1158/0008-5472.CAN-13-2169

9. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* (2018) 173(2):338–54. doi: 10.1016/j.cell.2018.03.034

10. Yi L, Huang P, Zou X, Guo L, Gu Y, Wen C, et al. Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. *Pharmacol Res* (2020) 161:105144. doi: 10.1016/j.phrs.2020.105144

11. Chen X, Zhang D, Jiang F, Shen Y, Li X, Hu X, et al. Prognostic prediction using a stemness index-related signature in a cohort of gastric cancer. *Front Mol Biosci* (2020) 7:570702. doi: 10.3389/fmolb.2020.570702

12. Xu Q, Xu H, Chen S, Huang W. Immunological value of prognostic signature based on cancer stem cell characteristics in hepatocellular carcinoma. *Front Cell Dev Biol* (2021) 9:710207. doi: 10.3389/fcell.2021.710207

13. Wang Z, Wang Y, Yang T, Xing H, Wang Y, Gao L, et al. Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Brief Bioinform* (2021) 22(5):bbab032. doi: 10.1093/bib/bbab032

14. Li X, Wenes M, Romero P, Huang SC, Fendt SM, Ho PC. Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nat Rev Clin Oncol* (2019) 16(7):425–41. doi: 10.1038/s41571-019-0203-7

15. Klebanoff CA, Crompton JG, Leonardi AJ, Yamamoto TN, Chandran SS, Eil RL, et al. Inhibition of AKT signaling uncouples T cell differentiation from expansion for receptor-engineered adoptive immunotherapy. *JCI Insight* (2017) 2 (23):e95103. doi: 10.1172/jci.insight.95103

16. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* (2010) 138(3):958–68. doi: 10.1053/j.gastro.2009.11.005

17. Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL, et al. Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *Gastroenterology* (2012) 142(3):562–71. doi: 10.1053/j.gastro.2011.11.026

18. Williams CS, Bernard JK, Demory BM, Almohazey D, Washington MK, Smith JJ, et al. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. *Carcinogenesis* (2015) 36(7):710–18. doi: 10.1093/carcin/bgv049

19. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PloS Med* (2013) 10(5):e1001453. doi: 10.1371/journal.pmed.1001453

20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

21. Zeng L, Fan X, Wang X, Deng H, Zhang K, Zhang X, et al. Bioinformatics analysis based on multiple databases identifies hub genes associated with hepatocellular carcinoma. *Curr Genomics* (2019) 20(5):349–61. doi: 10.2174/1389202920666191011092410

22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A.* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

23. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *Omics* (2012) 16(5):284–87. doi: 10.1089/omi.2011.0118

24. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* (2019) 47(D1):D607–13. doi: 10.1093/nar/gky1131

25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303

26. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* (2014) 8 Suppl 4:S11. doi: 10.1186/1752-0509-8-S4-S11

27. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an r package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* (2010) 26(7):976–78. doi: 10.1093/bioinformatics/btq064

28. Wang Y, Zhao M, Zhang Y. Identification of fibronectin 1 (FN1) and complement component 3 (C3) as immune infiltration-related biomarkers for diabetic nephropathy using integrated bioinformatic analysis. *Bioengineered* (2021) 12(1):5386–401. doi: 10.1080/21655979.2021.1960766

29. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003

30. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460

31. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

32. Geeleher P, Cox N, Huang RS. pRRophetic: an r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* (2014) 9(9):e107468. doi: 10.1371/journal.pone.0107468

33. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41(Database issue): D955–61. doi: 10.1093/nar/gks1111

34. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-Scale public data reuse to model immunotherapy response and resistance. *Genome Med* (2020) 12 (1):21. doi: 10.1186/s13073-020-0721-z

35. Edge SB, Compton CC. The American joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* (2010) 17(6):1471–74. doi: 10.1245/s10434-010-0985-4

36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* (2001) 25 (4):402–08. doi: 10.1006/meth.2001.1262

37. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for r and s+ to analyze and compare ROC curves. *BMC Bioinf* (2011) 12:77. doi: 10.1186/1471-2105-12-77

38. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* (2015) 160(1-2):48–61. doi: 10.1016/j.cell.2014.12.033

39. Allavena P, Digifico E, Belgiovine C. Macrophages and cancer stem cells: a malevolent alliance. *Mol Med* (2021) 27(1):121. doi: 10.1186/s10020-021-00383-3

40. Bottcher M, Baur R, Stoll A, Mackensen A, Mougiakakos D. Linking immunoevasion and metabolic reprogramming in b-Cell-Derived lymphomas. *Front Oncol* (2020) 10:594782. doi: 10.3389/fonc.2020.594782

41. Kim MS, Ha SE, Wu M, Zogg H, Ronkon CF, Lee MY, et al. Extracellular matrix biomarkers in colorectal cancer. *Int J Mol Sci* (2021) 22(17):9185. doi: 10.3390/ijms22179185

42. Yuzhalin AE, Lim SY, Kutikhin AG, Gordon-Weeks AN. Dynamic matrisome: ECM remodeling factors licensing cancer progression and metastasis. *Biochim Biophys Acta Rev Cancer* (2018) 1870(2):207–28. doi: 10.1016/j.bbcan.2018.09.002

43. Crotti S, Piccoli M, Rizzolio F, Giordano A, Nitti D, Agostini M. Extracellular matrix and colorectal cancer: How surrounding microenvironment affects cancer cell behavior? *J Cell Physiol* (2017) 232(5):967–75. doi: 10.1002/jcp.25658

44. Ortensi B, Setti M, Osti D, Pelicci G. Cancer stem cell contribution to glioblastoma invasiveness. *Stem Cell Res Ther* (2013) 4(1):18. doi: 10.1186/scrt166

45. Weng ML, Chen WK, Chen XY, Lu H, Sun ZR, Yu Q, et al. Fasting inhibits aerobic glycolysis and proliferation in colorectal cancer *via* the Fdft1-mediated AKT/mTOR/HIF1alpha pathway suppression. *Nat Commun* (2020) 11(1):1869. doi: 10.1038/s41467-020-15795-8

46. Wu Z, Zuo M, Zeng L, Cui K, Liu B, Yan C, et al. OMA1 reprograms metabolism under hypoxia to promote colorectal cancer development. *EMBO Rep* (2021) 22(1):e50827. doi: 10.15252/embr.202050827

47. Nijhuis A, Thompson H, Adam J, Parker A, Gammon L, Lewis A, et al. Remodelling of microRNAs in colorectal cancer by hypoxia alters metabolism profiles and 5-fluorouracil resistance. *Hum Mol Genet* (2017) 26(8):1552–64. doi: 10.1093/hmg/ddx059

48. Li C, Yang F, Wang R, Li W, Maskey N, Zhang W, et al. CALD1 promotes the expression of PD-L1 in bladder cancer *via* the JAK/STAT signaling pathway. *Ann Transl Med* (2021) 9(18):1441. doi: 10.21037/atm-21-4192

49. Zheng H, Bai Y, Wang J, Chen S, Zhang J, Zhu J, et al. Weighted gene Co-expression network analysis identifies CALD1 as a biomarker related to M2 macrophages infiltration in stage III and IV mismatch repair-proficient colorectal carcinoma. *Front Mol Biosci* (2021) 8:649363. doi: 10.3389/fmolb.2021.649363

50. Zhang C, Chen T, Li Z, Liu A, Xu Y, Gao Y, et al. Depiction of tumor stemlike features and underlying relationships with hazard immune infiltrations based on large prostate cancer cohorts. *Brief Bioinform* (2021) 22(3). doi: 10.1093/bib/bbaa211

51. Vayrynen JP, Tuomisto A, Klintrup K, Makela J, Karttunen TJ, Makinen MJ. Detailed analysis of inflammatory cell infiltration in colorectal cancer. *Br J Cancer* (2013) 109(7):1839–47. doi: 10.1038/bjc.2013.508

52. Li C, Jiang P, Wei S, Xu X, Wang J. Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. *Mol Cancer* (2020) 19(1):116. doi: 10.1186/s12943-020-01234-1

53. Ino Y, Yamazaki-Itoh R, Shimada K, Iwasaki M, Kosuge T, Kanai Y, et al. Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. *Br J Cancer* (2013) 108(4):914–23. doi: 10.1038/bjc.2013.32

54. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* (2013) 14(10):1014–22. doi: 10.1038/ni.2703

55. Dersh D, Holly J, Yewdell JW. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat Rev Immunol* (2021) 21 (2):116–28. doi: 10.1038/s41577-020-0390-6

56. Billan S, Kaidar-Person O, Gil Z. Treatment after progression in the era of immunotherapy. *Lancet Oncol* (2020) 21(10):e463–76. doi: 10.1016/S1470-2045(20)30328-4

57. Schoenfeld AJ, Hellmann MD. Acquired resistance to immune checkpoint inhibitors. *Cancer Cell* (2020) 37(4):443–55. doi: 10.1016/j.ccell.2020.03.017

58. Touat M, Li YY, Boynton AN, Spurr LF, Iorgulescu JB, Bohrson CL, et al. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* (2020) 580(7804):517–23. doi: 10.1038/s41586-020-2209-9

59. Colak S, Medema JP. Human colonic fibroblasts regulate stemness and chemotherapy resistance of colon cancer stem cells. *Cell Cycle* (2016) 15(12):1531–37. doi: 10.4161/15384101.2014.973321

60. Colak S, Zimberlin CD, Fessler E, Hogdal L, Prasetyanti PR, Grandela CM, et al. Decreased mitochondrial priming determines chemoresistance of colon cancer stem cells. *Cell Death Diff* (2014) 21(7):1170–77. doi: 10.1038/cdd.2014.37

61. Kishore C, Bhadra P. Current advancements and future perspectives of immunotherapy in colorectal cancer research. *Eur J Pharmacol* (2021) 893:173819. doi: 10.1016/j.ejphar.2020.173819

62. Wei R, Quan J, Li S, Liu H, Guan X, Jiang Z, et al. Integrative analysis of biomarkers through machine learning identifies stemness features in colorectal cancer. *Front Cell Dev Biol* (2021) 9:724860. doi: 10.3389/fcell.2021.724860

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fimmu.2022.980911

Check for updates

OPEN ACCESS

EDITED BY
Jun Liu,
Yuebei People's Hospital, China

REVIEWED BY
Dan Wang,
Hubei University of Science and Technology, China
Sun Zhaoyang,
Fudan University, China

*CORRESPONDENCE
Daheng Yang
yangdaheng@njmu.edu.cn

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

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 29 June 2022
ACCEPTED 01 August 2022
PUBLISHED 23 August 2022

CITATION
Chen H, Wu M, Xia H, Du S, Zhou G, Long G, Zhu Y, Huang X and Yang D (2022) FLT3LG and IFITM3P6 consolidate T cell activity in the bone marrow microenvironment and are prognostic factors in acute myelocytic leukemia. *Front. Immunol.* 13:980911. doi: 10.3389/fimmu.2022.980911

FLT3LG and IFITM3P6 consolidate T cell activity in the bone marrow microenvironment and are prognostic factors in acute myelocytic leukemia

Haiyan Chen[1,2†], Meng Wu[2†], Hongping Xia[3†], Songjie Du[4†], Guoren Zhou[5†], Guangfeng Long[2], Yanan Zhu[6], Xu Huang[2] and Daheng Yang[2*]

[1]Institute for Cancer Research, School of Basic Medical Science of Xi'an Jiaotong University, Xi'an, China, [2]Department of Clinical Laboratory, Children's Hospital of Nanjing Medical University, Nanjing, China, [3]Department of Pathology, School of Basic Medical Sciences & Key Laboratory of Antibody Technique of National Health Commission, Nanjing Medical University, Nanjing, China, [4]Department of Medical Genetics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China, [5]Jiangsu Cancer Hospital & The Affifiliated Cancer Hospital of Nanjing Medical University & Jiangsu Institute of Cancer Research, Nanjing, China, [6]Translational Medicine Institute, Xi'an Jiaotong University Health Science Center, Xi'an, China

Acute myelocytic leukemia (AML) is a malignancy of the stem cell precursors of the myeloid lineage. CD4+ and CD8+ T cells play pivotal roles in influencing AML progression but are functionally suppressed in the bone marrow microenvironment. We aimed to find hub genes related to T cell exhaustion and suppression, thereby providing evidence for immunotherapy. In this study, gene transcriptome expression data from TCGA and TARGET databases were utilized to find key genes. Firstly, CIBERSORT immune cell infiltration algorithm and WGCNA method were used to identify CD4+ and CD8+ T cells-related genes. Univariate and multivariate cox regression analyses were then introduced to construct the overall survival prognosis model and included hub genes. The ESTIMATE and ssGSEA scoring methods were used to analyze the correlation between the hub genes and immune activity. Single-cell transcriptome analysis was applied to detect the immune cells expressing hub genes, hence, to detect exact mechanisms. Consequently, FLT3LG and IFITM3P6 were determined to be positively correlated with patients' overall

survival and microenvironment immune activity. Further study suggested FLT3-FLT3LG and IFITM3P6-miR-6748-3p-CBX7 signaling axes were involved in CD4+ and CD8+ T cells activation. This may be one of the mechanisms of T cells suppression in AML.

KEYWORDS

acute myelocytic leukemia, bone marrow microenvironment, dendritic cell activation, T cell activation, FLT3LG, IFITM3P6

Introduction

Acute myelocytic leukemia (AML) is a group of malignant clonal diseases that originate from the myeloid stem and progenitor cells and are highly heterogeneous in immunophenotype, cytogenetics, and molecular genetics (1). It's the most common acute leukemia in adults (~80% in this group), and the proportion of pediatric patients is about 20% (2). Similar to other malignant tumors, genetic variations consist of the main reason that leads to neoplastic changes and clonal proliferation. The incidence of AML elevated with age, from ~1.3 cases per 100 thousand population less than 65 years old, to 12.2 per 100 thousand population that over 65 years old (3). With a very variable prognosis and a high mortality rate, 5-year overall survival in AML cases is less than 50%, and 20% of elderly patients will survive 2 years after diagnosis (4, 5). In the past decades, treatment paradigms were still unchanged with survival curves remaining stagnant (6).

The current regimens include a high-intensity induction phase wherein cytotoxic chemotherapy, or other target-specific agents are administered based on disease profile as well as patient risk and comorbidities (6, 7). Due to the minimal residual disease in the bone marrow (BM) microenvironment, failure in treatment often occurred (8, 9). Therefore, a better understanding of the AML microenvironment is crucial for preventing tumor development and designing an effective treatment regimen. Studies have reported various disorders in T cell immunity, including increased T regulatory cells, decreased T helper cells, exhausted T cells, functional T cell suppression and abnormal activity of transcription factors in the presence of AML (10). Consequently, T cell-associated functions are changed to allow immunity evasion of tumor cells. The importance of T cell in antileukemia is confirmed by Lamble et al., they showed an association between T cells and clinical outcomes based on their research. Patients with higher BM T cell percentages (over 78.5% of total lymphocytes) presented favorable overall survival (11). In addition, an *in vitro* co-culture experiment of T cells and AML leukemic cells revealed that pooled CD4+ and CD8+ cells cytotoxic against blasts (32%, 30:1 E/T ratio) (12). The antitumor effect of CD8+ T cells is mainly attributed to it can recognize AML-derived mutated peptides and induce a cytotoxic effect on tumor cells. One important mechanism was revealed in a study that CD8+ T cells could disrupt AML progression caused by mutation of nucleophosmin 1 (NPM1) or FMS-like tyrosine kinase receptor 3 internal tandem duplication (FLT3-ITD) (13, 14). Other studies demonstrated that CD4+ T cells prompt apoptotic effect on AML cells and are mainly associated with INF-γ release (15).

It has been well documented cytotoxic effects of CD4+ and CD8+ T cells against AML cells, which consists of important mechanisms in influencing disease relapse and drug resistance in the tumor microenvironment (TME) (16, 17), blocking tumor progression, invasiveness, and metastasis, and even maintaining a cancer stem-like phenotype (18). But so far, the molecular and cellular mechanisms of T cell exhaustion and functional suppression remain unclear. A better understanding of these will give a deep insight into the balance of tumor immune surveillance and cancer cell evasion in TME, and help to develop novel immunotherapeutic approaches.

The rapid development of bioinformatics facilitates the exploration of potential mechanisms driving tumor progression. In this study, we sought to determine the relationship between hub genes and T cells in TME to reveal the landscape of AML progression.

Materials and methods

Data acquisition and processing

Gene transcriptome expression data of AML were obtained from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (https://ocg.cancer.gov/programs/target/data-matrix) databases. The blood gene expression data of healthy controls were obtained from the Genotype-Tissue Expression (GTEx) (https://gtexportal.org/home/datasets) database, including 755 cases. All expression data were raw count values. Before analysis, these values were converted into transcripts per million (TPM) values.

Batch correction of the merged data was performed by the normalizeBetweenArrays function of the *limma* package. Finally, the expression data were filtrated with a criterion by the average expression level of each gene across samples > 0.1, and the repeated genes were averaged.

Analysis of immune cell infiltration

Firstly, the immune cell infiltration landscape of TCGA and TARGET data was explored to find relationships between infiltrating immune cells, especially CD4+ and CD8+ T cells, and AML patients' overall survival (OS). A deconvolution algorithm CIBERSORT (https://cibersortx.stanford.edu/) calculates the composition and proportion of infiltrating immune cells (19). It has been widely used to reveal the immune cell subtypes infiltrating numerous cancers (20, 21). With matched R script v1.03 (last updated 07-10-2015) and feature genes expression matrix of 22 immune cell subsets (downloaded from CIBERSORT website) as background expression data, the infiltration features of these immune cells in each sample were determined. Samples with $P < 0.05$ were included for further analysis. The analysis was based on *e1071*, *parallel* and *preprocessCore* packages.

Screening of immune cell infiltration-related genes

Next, we introduced weighted gene co-expression network analysis (WGCNA), which is used to analyze gene expression patterns of multiple samples, to identify core modules and central genes that are associated with infiltrating CD4+ and CD8+ T cells (22). WGCNA method can cluster genes and form modules by similar gene expression patterns and analyze the relationship between modules and specific features. Candidate genes were screened by module membership (MM) > 0.8 and gene significance (GS) > 0.4, with both threshold of *P*-value < 0.05. Then these candidate genes were intersected between TCGA and TARGET data. Gene co-expression modules were identified using the *WGCNA* package.

Identification of hub genes

Based on those intersected genes, AML patients' overall survival prognosis model was constructed using TCGA data by univariate and multivariate cox regression methods. Univariate cox regression analysis for survival-related genes was determined by $P < 0.05$. Finally, genes included in the prognosis model were defined as hub genes. The analysis was based on *survival* and *survminer* packages.

Relationship between hub genes and immune activity in TME

To investigate the correlation of hub genes and immune activity in microenvironment, the single-sample gene set enrichment analysis (*ssGSEA*) algorithm, which standardizes the gene expression value of an AML sample by rank and calculates the enrichment fraction of 29 immune cell types (23) using the empirical cumulative distribution function (24), was implemented to group AML samples into different immune activity sets. Patients were assigned into high- and low-immune activity groups using the *GSVA* package.

Furtherly, the ESTIMATE method (25), an algorithm designed to calculate scores for reflecting the infiltration levels of immune cells and stromal cells within the TME on the foundation of their specific genes expression level, was introduced to validate the accuracy of immune activity grouping based on the TCGA data. Then the relationship between hub genes expression level and immune activity was determined.

Detection of hub genes at the single-cell level

Commonly, accurate positioning of gene expression in certain cells will contribute to a better understanding of their functional mechanism, and single cell RNA sequencing (scRNA) technology provides us a chance to realize the research.

Herein, the Tumor Immune Single-cell Hub (TISCH) (http://tisch.comp-genomics.org/home/), a database that integrated enormous scRNA data of various tumors, was used to obtain the accurate expression location of these genes. AML_GSE116256 (26) and PBMC_30K_10× datasets (a test data from 10× Genomics website) for AML and healthy control were obtained. Then, the immune subset of malignant cells in which genes are mainly expressed was explored to determine the precise mechanisms of action.

Functional enrichment analysis of hub genes

To find the function of these hub genes, gene set enrichment analysis (GSEA) was performed based on their expression level to complete gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. According to the screening criteria suggested by MsigDB website (https://www.gsea-msigdb.org/gsea/msigdb), enrichment terms were included with *NOM p-value* < 0.05 and *FDR q-value* < 0.25. GSEA was performed by gsea software (version 3.0) and visualization was based on the *clusterProfiler* package.

Pseudogene exploration

A pseudogene was included as a hub gene in this study. Pseudogenes are a special type of long non-coding RNAs (lncRNAs) that regulate different tumorigenic processes. To detect its function involved in competing for endogenous RNA (ceRNA) reaction in AML, we firstly retrieved its sequence information from the UCSC Genome Browser (http://genome.ucsc.edu/) and predicted its location in lncLocator (http://www.csbio.sjtu.edu.cn/bioinf/lncLocator/) and iLoc-LncRNA (http://lin-group.cn/server/iLoc-LncRNA/predictor.php) databases. Then its predicted mRNA sequence was put in miRDB (http://www.mirdb.org/) database to select its sponge-binding miRNAs.

LncRNAs are confirmed to be critical in influencing transcription factors involved in biological processes, therefore yielding or facilitating tumor development. We downloaded a gene set containing 318 tumor-related transcription factors (Supplementary Table 1) from the Cistrome Cancer (http://cistrome.org/CistromeCancer/) database, mapped them to TCGA expression data, and resulted in pseudogene-related transcription factors using the univariate Cox regression analysis method. Genes with a correlation coefficient > 0.5 and $P < 0.05$ were selected.

Additionally, the catRAPID omics database (http://service.tartaglialab.com/page/catrapid_group) was used to find potential genes interacting with the pseudogene. Moreover, pseudogene-related transcription factors and potentially related genes were intersected and put into the miRDB database to find potentially interacting miRNAs. Finally, pseudogene- and intersected transcription factors-related miRNAs were further intersected to determine those genes involved in the ceRNA mechanism.

Validation of genes in peripheral blood by qRT-PCR

Based on these findings, we collected ethylenediaminetetracetic acid disodium (EDTA-Na_2) anticoagulant peripheral whole blood of 24 AML patients and 19 healthy controls from the Children's Hospital of Nanjing Medical University and The First Affiliated Hospital of Xi'an Jiaotong University from November 2020 to February 2021. All patients were initially diagnosed with AML; their blood was obtained before treatment. For mRNA and lncRNA validation, total RNA was extracted using the RNAprep Pure Hi-Blood Kit (TianGen Biotechnology, China) and reverse transcribed using the PrimeScript RT reagent kit (Takara, Japan). Primer sequences, which were synthesized by Generay Biotechnology (Shanghai, China), are listed in Supplementary Table 2. For microRNA validation, the specific primers of real-time reverse transcription PCR (qRT-PCR) from Bulge-loop™ miRNA qRT-PCR Primer Sets (one RT primer and a pair of qPCR primers for each set) for hsa-miR-6748-3p quantification were designed by RiboBio (Guangzhou, China).

Quantification of all genes was based on a real-time fluorescent quantitative PCR assay. The SYBR Green dye was purchased from Takara (Japan). For mRNA and lncRNA, GAPDH was used as an endogenous control gene; U6 was used as an endogenous control for miRNA. The relative quantification for gene expression was defined as $2^{-\Delta\Delta CT}$ compared with control group. All PCR reactions were performed in triplicate.

Dual-luciferase reporter gene assay and ceRNA mechanism validation

This study was performed in human embryonal kidney 293T cell line, CD4+ T cell and CD8+ T cell. CD4+ and CD8+ T cells were extracted from EDTA-Na_2 anticoagulant peripheral whole blood (donated by Daheng Yang), using MojoSort™ Human CD4T/CD8T Cell Isolation Kit (BioLegend, USA). 293T cells and T cells were maintained in Dulbecco's Modified Eagle's medium (DMEM, Hyclone, USA) and 1640 medium (Hyclone, USA), respectively, with 10% fetal bovine serum (Gibco, USA), 100 U/ml penicillin and 100 U/ml streptomycin (Beyotime, China), and incubated at 5% CO_2 at 37°C.

PmirGLO-CBX7 vector plasmids (mutant and wild types) were constructed by Generay Biotechnology (Nanjing, China); micrON™ miR-6748-3p mimic, micrOFF™ miR-6748-3p inhibitor and their negative controls were synthesized by RiboBio; overexpression plasmid and shRNA plasmid of IFITM3P6 were constructed by Qingke Biotech (Beijing, China), target sequences were presented in Supplementary Table 2. All experiments were performed in 6-well plates (Corning, USA) in triplicates. Synthetic nucleic acids were transfected by lipofectamine 2000 (Invitrogen, USA) to cells. Reporter gene assay was performed using the Dual-Luciferase® Reporter Assay System Kit (Promega, USA) according to the manufacturer's instructions, plasmids were quantified by 4μg and miR-6748-3p mimic was quantified by 100nM per well. Upregulation and downregulation of miR-6748-3p were quantified by 100nM mimic and 150nM inhibitor per well, respectively. Overexpression and down expression of IFITM3P6 plasmids were quantified by 4μg, respectively. CBX7 expression levels were detected by qRT-PCR with GAPDH as endogenous control in triplicates.

RNA N^6-methyladenosine (m6A) and DNA methylation analysis

RNA m6A and DNA methylation are important biological processes affecting expression of genes. Their abnormalities will lead to different consequences in tumor progression. Our study

has included hub genes and their relevant genes, but aberrant expression mechanism is need to be elucidated. Methylation may be the main reason.

Herein, we downloaded DNA methylation data of 140 AML samples from the TCGA database using gdc-client tool (downloaded from https://gdc.cancer.gov/access-data/gdc-data-transfer-tool), and extracted methylation positions for each sample.

For RNA m6A methylation analysis, we compared the correlations of included genes with m6A-related genes by Pearson test. For DNA methylation analysis, correlations of gene expression with methylation and gene expression with methylation position were tested. All statistical significance was determined by Pearson coefficients > 0.3 or < -0.3, and $P < 0.05$.

Statistical analysis

All statistical analyses were performed using R software (version 4.0.3) and RStudio (version 1.2.1335). The Wilcox signed-rank test was used for extracting differentially expressed genes (DEGs) with a $|log_2 Fold\ Change\ (FC)| > 2$ and adjusted *P-value* < 0.01. PCR detection was analyzed by unpaired t-test using GraphPad Prism 8 software (version 8.0.1), with a $P < 0.05$.

Results

Data preprocessing and identification of DEGs

The flow chart of the study is shown in the Supplementary Figure 1. Clinical information from the TCGA and TARGET datasets is shown in Table 1. There were 151 samples from 151 patients in TCGA and 358 samples from 296 patients in the TARGET cohort. After data batch correction by the normalize BetweenArrays function of the *limma* package, principal component analysis (PCA) showed that the tumor samples were significantly different from the control samples (Supplementary Figures 2A, B). Heatmap of the top 1000 genes with the highest standard deviation (SD) values indicated that there were significant differences between groups (Supplementary Figures 2C, D). The results suggested that the batch effect of the corrected expression data could be ignored. The number of DEGs in TCGA and TARGET was 11084 and 8436, respectively. WGCNA analysis was based on these DEGs.

CD4+ and CD8+ T cells infiltration levels are associated with OS of AML patients

CIBERSORT deconvolution algorithm showed that the infiltration of 22 types of immune cell was significantly different between the tumor group and the control group in TCGA (tumor group, n = 121; control group, n = 375) and TARGET (tumor group, n = 224; control group, n = 751) datasets. The results revealed that the infiltration level of resting memory CD4+ T cells was significantly increased in tumors. However, the infiltration levels of CD8+ T cells and naïve CD4+ T cells, activated memory CD4+ T cells were markedly reduced in tumors, both in TCGA and TARGET samples (Figures 1A, B; Supplementary Figure 3).

In the TCGA dataset, CD8+ T cells and activated memory CD4+ T cells were significantly associated with the patient's survival. The OS rate of the high-level group was markedly higher than that of the low-level group ($P = 0.012$; $P = 0.012$). Similar results were observed for CD8+ T cells and naïve CD4+ T cells in the TARGET dataset ($P = 0.014$; $P< 0.001$). However, the infiltration of resting memory CD4+ T cells showed a contrary result ($P = 0.009$) (Figures 1C–G).

Identification of hub genes

The WGCNA method was used to screen out genes related to infiltrating CD4+ and CD8+ T cells. By combining the CIBERSORT results and the differential gene expression matrices, outlier samples were deleted through sample clustering to reduce sample derived deviation. A soft threshold power of 4 was chosen for TCGA data to construct weighted genes co-expression networks, and 23 color modules were finally obtained by merging similar modules. Similarly, a soft threshold power of 8 was selected for TARGET data, and 21 color modules were finally obtained (Supplementary Figure 4).

Interestingly, the correlation analysis between the module genes and immune cells showed some similar patterns in the two datasets. In the TCGA dataset, the green module was significantly correlated with CD8+ T cells and resting memory CD4 T+ cells (R = 0.7, P = 9.2e-106; R = 0.69, P = 1.3e-101, respectively) (Figures 2A–C). In the TARGET dataset, the yellow module was significantly correlated with CD8+ T cells and resting memory CD4+ T cells (R = 0.76, P = 5.3e-111; R = 0.44, P = 4.8e-29, respectively) (Figures 2D–F). In addition, infiltration of CD4+ and CD8+ T cells was markedly associated with the OS rate of AML patients. Consequently, 94 genes were obtained from the TCGA dataset and 89 genes from the TARGET dataset, according to the screening criteria. The

10.3389/fimmu.2022.980911

TABLE 1 Clinical characteristics of patients in TCGA and TARGET datasets.

Characteristics	Databases			
	TCGA		TARGET	
Age (year)	> 50	94	>12	122
	≤50	57	<=12	174
Gender	male	83	Male	159
	female	68	Female	137
Leukocyte ($\times10^9$/L)	>30	63	>30	182
	≤30	88	≤30	114
Blast cell (%)	>30	84	>30	270
	≤30	67	≤30	17
	NA	0	NA	9
FAB category	M0	15	M0	8
	M1	35	M1	37
	M2	38	M2	73
	M3	15	M3	0
	M4	29	M4	72
	M5	15	M5	54
	M6	2	M6	4
	M7	1	M7	9
	unknown	1	unknown	39
Survival status	alive	54	alive	158
	dead	97	dead	138

FAB, French-American-British.

intersection of the two datasets included 37 candidate genes (Figure 3A).

The univariate Cox regression method screened 6 OS-related genes from these candidate genes using TCGA data, including RP11-23J18.1, FLT3LG, LCK, RNF157, GOLGA7B and ZNF831. All of the 6 genes were found to be downregulated in the tumor samples and were significantly associated with the OS rate of AML patients (Figure 3B, Supplementary Figure 5). Multivariate cox regression analysis was used to build a survival prediction model, which finally included RP11-23J18.1 and FLT3LG (Figure 3C). The Kaplan-Meier curve showed that the OS rate in the lower-risk group was significantly higher than that in the higher-risk group (P = 4.251e-03), suggesting that the patients' overall survival time decreases with an increase of the risk score (Figure 3D, E).

RP11-23J18.1 and FLT3LG are positively correlated with immune activity

Next, the relationship between RP11-23J18.1 and FLT3LG and immune activity was explored. Firstly, we utilized the ESTIMATE algorithm to evaluate the TCGA expression matrix, obtained the immune cell score, stromal cell score, ESTIMATE score (the total score) and tumor purity score in the TME (Figure 4A), finding that immune cell score, stromal cell score and ESTIMATE score were positively correlated with immune activity (Figures 4B–D), while tumor purity was inversely correlated with immune activity (Figure 4E). That validated high- and low-immune activity grouping by ssGSEA method was accurate.

Further analysis showed that the expression levels of FLT3LG and RP11-23J18.1 genes were significantly higher in the high immunocompetence group compared with the low immunocompetence group in the TCGA ($P < 0.001$; $P < 0.001$; Figures 4F, G). These results suggest that FLT3LG and RP11-23J18.1 may play an important role in the activation process of CD4+ and CD8+ T cells. Moreover, GSEA analysis showed that FLT3LG and RP11-23J18.1 were mainly enriched in antigen receptor-mediated signaling transduction, cell killing, T cell migration and differentiation, interleukin regulation, etc. (Figure 5). These biological processes are closely related to T cells, which are involved in killing tumor cells.

FLT3LG-FLT3 axis involved in influencing dendritic cell activation

In TISCH database, we retrieved FLT3LG is mainly expressed in CD4+ and CD8+ T cells in healthy control (PBMC_30K_10×

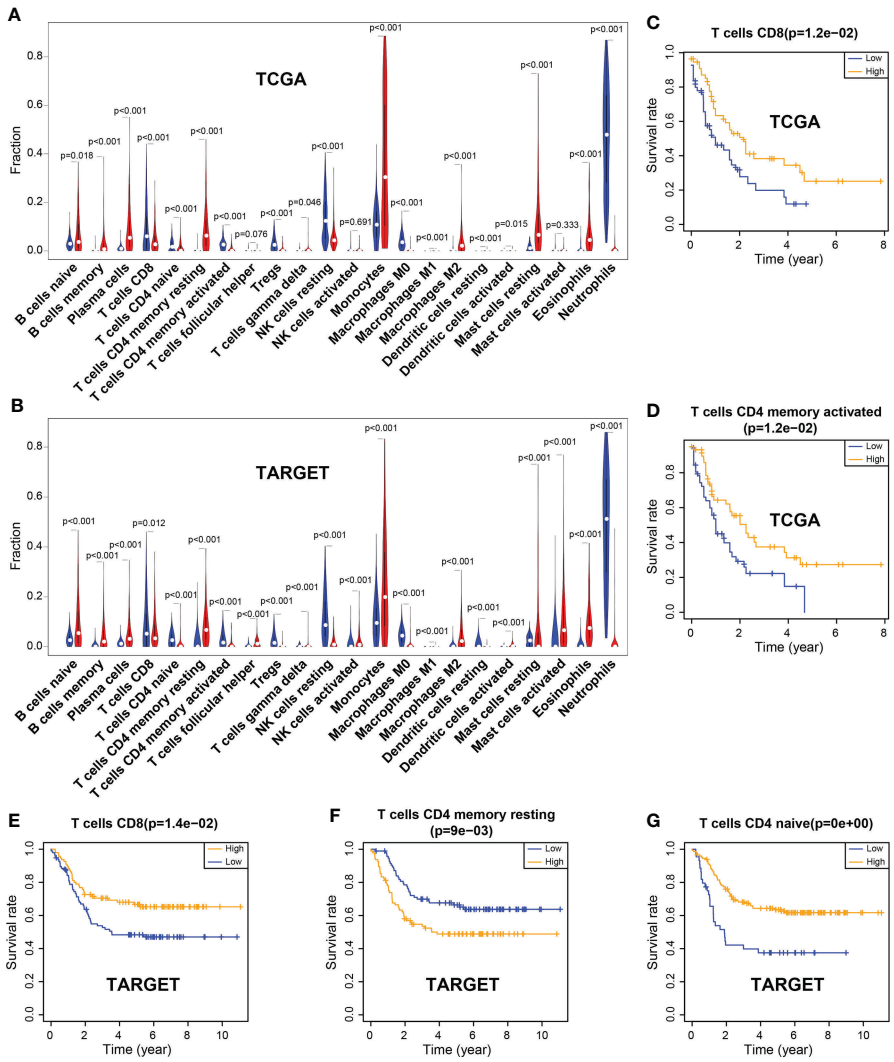


FIGURE 1
Differences of immune cell infiltration between the tumor and the normal group, and correlation between immune cells and the overall survival rate. **(A)** Violin plot of immune cell infiltration in TCGA. **(B)** Violin plot of immune cell infiltration in TARGET. **(C, D)** CD8 and activated memory CD4 T cells were associated with the overall survival (OS) rate, which was significantly higher in the high-level group than in the low-level group in TCGA. **(E–G)** CD8 T cells and naïve CD4 T cells were associated with the OS rate, which was markedly higher in the high-level group than in the low-level group in TARGET; however, resting memory CD4 T cells showed the opposite results.

dataset, Figures 6A, C). Its receptor, FLT3, is mainly expressed in dendritic cells (DCs) (Figure 6E). Pearson test determined FLT3LG expression was positively correlated with CD40 expression, which is the marker of dendritic cell activation (Figure 6G). In the AML_GSE116256 dataset, FLT3LG is mainly expressed in CD4+ T cells, but compared with the PBMC_30K_10× dataset, the expression level of FLT3LG is remarkably reduced (Figures 6B, D), and FLT3 is mainly expressed in leukemia cells and precursor cells (Figure 6F). This reminder us FLT3LG (ligand) may activate DCs by acting on FLT3 (receptor) under physiological conditions.

To prove it, we purchased primary dendritic cells (cat NO.: CP-R162) that were extracted from the bone marrow of rats by flow cytometry from Procell Life Science & Technology Corporation (Wuhan, China), cultured them in customized media (provided by the company). Added 1μg/ml, 10μg/ml, 50μg/ml, 100μg/ml and 200μg/ml of recombinant rat Flt3lg cytokine (PEPROTECH, USA) per well (6-well plate) to stimulate the activation of dendritic cells and performed qRT-PCR, found that expression of CD40 was higher than the negative control, and its expression presented positive dose-dependent (Figure 6H).

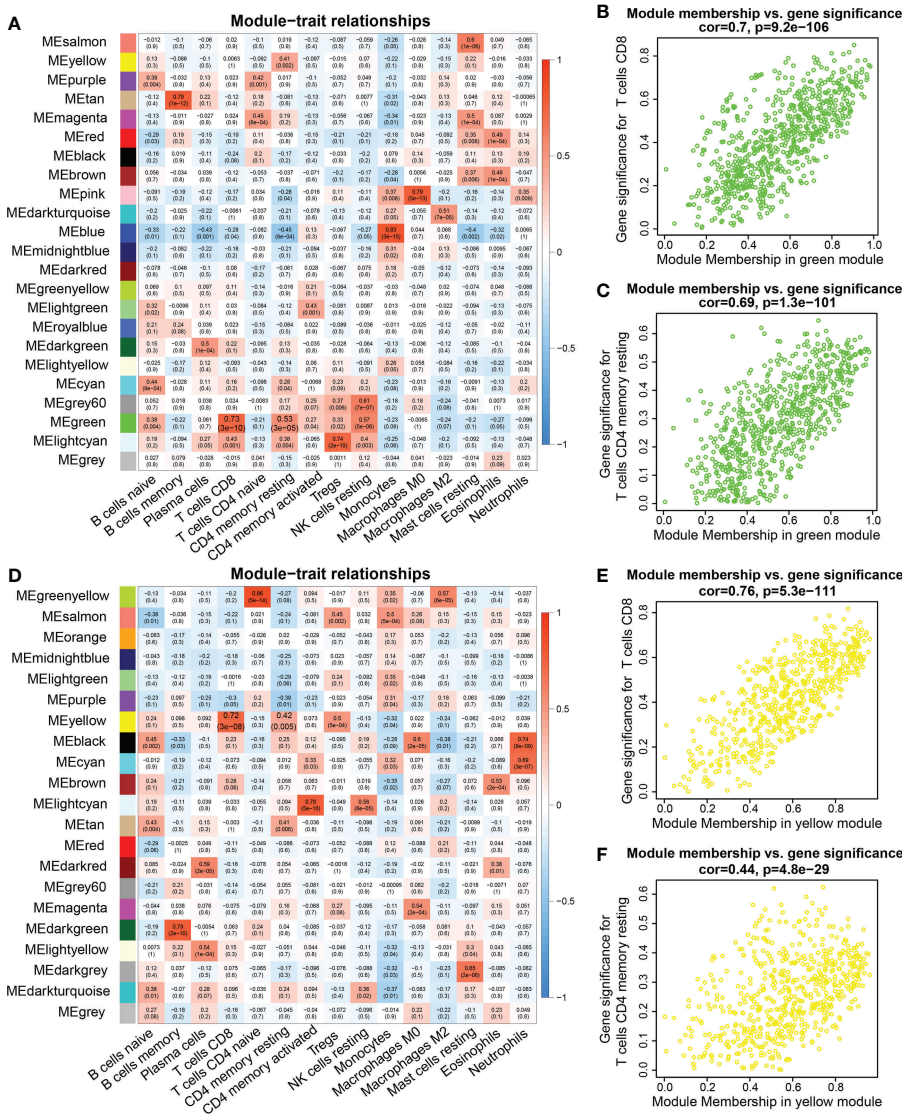


FIGURE 2
WGCNA analysis between genes and immune cell infiltration signatures. **(A, D)** Heatmap of the correlations between module eigengenes and immune cells infiltration. **(B)** Pearson correlation between gene significance (GS) in CD8 T cells and module membership (MM) in the green module in the TCGA database. **(C)** Pearson correlation between GS in memory CD4 T cells and MM in the green module in the TCGA database. **(E)** Correlation between GS in CD8 T cell and MM in the yellow module in the TARGET database. **(F)** Correlation between GS in resting memory CD4 T cells and MM in the yellow module in the TARGET database.

IFITM3P6 influences CBX7 expression by sponge binding miR-6748-3p

The pseudogene, RP11-23J18.1 (ensemble name ENSG00000258352), was put into the UCSC Genome Browser (http://genome.ucsc.edu/) and GeneCards (https://www.genecards.org/) databases. It was confirmed that it mostly matched with IFITM3P6 (IFITM3P pseudogene 6). The annotation for its gene version (ENSG00000258352.1_8) and transcript version (ENST00000553227.1_1) was validated level. Therefore, the predicted RNA sequences of IFITM3P6 provided by the UCSC database were obtained (Supplementary Table 3). In the NCBI database, the gene ID of IFITM3P6 was confirmed to be 643058 (*Homo sapiens*). Transcriptome sequencing analysis showed that IFITM3P6 had the highest mean expression in blood compared with other tissues and organs in its physiologic context in the GTEx (https://www.gtexportal.org/home/gene/ENSG00000258352) database (Supplementary Figure 6). In subsequent studies, IFITM3P6 will be used as the pseudogene instead of RP11-23J18.1.

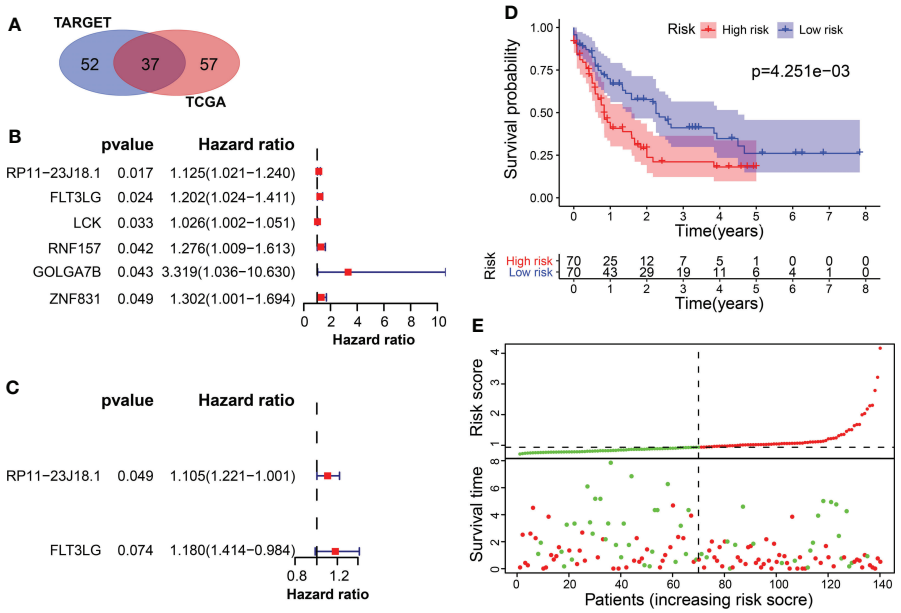


FIGURE 3
Screening of survival-related differentially expressed genes. **(A)** The intersection of candidate genes in TCGA and TARGET datasets. **(B)** Univariate cox regression analysis of 6 survival-related genes in TCGA. **(C)** Multivariate cox regression analysis included RP11-23J18.1 and FLT3LG as survival predicting gene models in TCGA. **(D, E)** Application of the survival predicting model in TCGA dataset.

In the lncLocator database, the predicted subcellular locations of IFITM3P6 were mainly in cytoplasm and cytosol with a score of 0.449968 and the nucleus with a score of 0.416618. In the iLoc-LncRNA database, the subcellular location was cytoplasm and cytosol, with a probability score of 0.956631. Based on these results, we hypothesized that IFITM3P6 is a lncRNA that exerts function in the cytoplasm. A total of 579 IFITM3P6-related genes were retrieved from the catRAPID omics database (Supplementary Table 4), which are potentially binding to IFITM3P6 directly or indirectly.

Next, 10 transcription factors were identified, which were significantly correlated with IFITM3P6. Then, 579 IFITM3P6-related genes were intersected with the 10 IFITM3P6-related transcription factors, including CBX7 as a core gene (Figure 7A).

In TCGA data, the expression of IFITM3P6 was positively correlated with CBX7 expression level by Pearson test (R = 0.57, P = 2.116e-14) (Figure 7B). scRNA analysis showed that CBX7 was mainly expressed in T cells, B cells, NK cells and plasma cells in a healthy control, while it presented a striking decrease in the AML data (Figures 7C, D).

Furthermore, 18 IFITM3P6-related miRNAs and 142 CBX7-related miRNAs were predicted in the miRDB database. miR-6748-3p was finally intersected from them (Figure 7E; Supplementary Table 5). Compared to healthy controls, IFITM3P6 and CBX7 were significantly decreased in AML blood and HL-60 cell (t = 9.665, P < 0.0001; t = 9.043, P < 0.0001, respectively). However, miR-6748-3p was significantly increased in AML blood samples (t = 4.197, P < 0.001) (Figure 7F).

The binding sites of miR-6748-3p to CBX7 and IFITM3P6 were predicted in the miRDB database (Figure 7G). Dual-luciferase reporter assay in 293T cells confirmed that CBX7 wild type and miR-6748-3p mimic co-transfected group significantly decreased luciferase activity compared with CBX7 wild type and miR-6748-3p negative control co-transfected group (P < 0.0001). The result was comparable with CBX7 mutant type and miR-6748-3p mimic co-transfected group (P < 0.0001) (Figure 7H), while the luciferase activity of the co-transfected group with miR-6748-3p mimic and CBX7 mutant type had no significance compared to control group, suggesting that miR-6748-3p could bind to the 3' UTR of CBX7, thereby interfering with its expression. Similar results were validated in CD4+ and CD8+ T cells (Supplementary Figure 7).

Compared with negative controls, mRNA of CBX7 presented significantly lower expression in miR-6748-3p mimic group, while presented higher expression in miR-6748-3p inhibitor group (Figures 7I, J). Furtherly, qRT-PCR revealed a higher expression of CBX7 compared with negative control when IFITM3P6 was upregulated, while its downregulation led contrary result (Figures 7K, L).

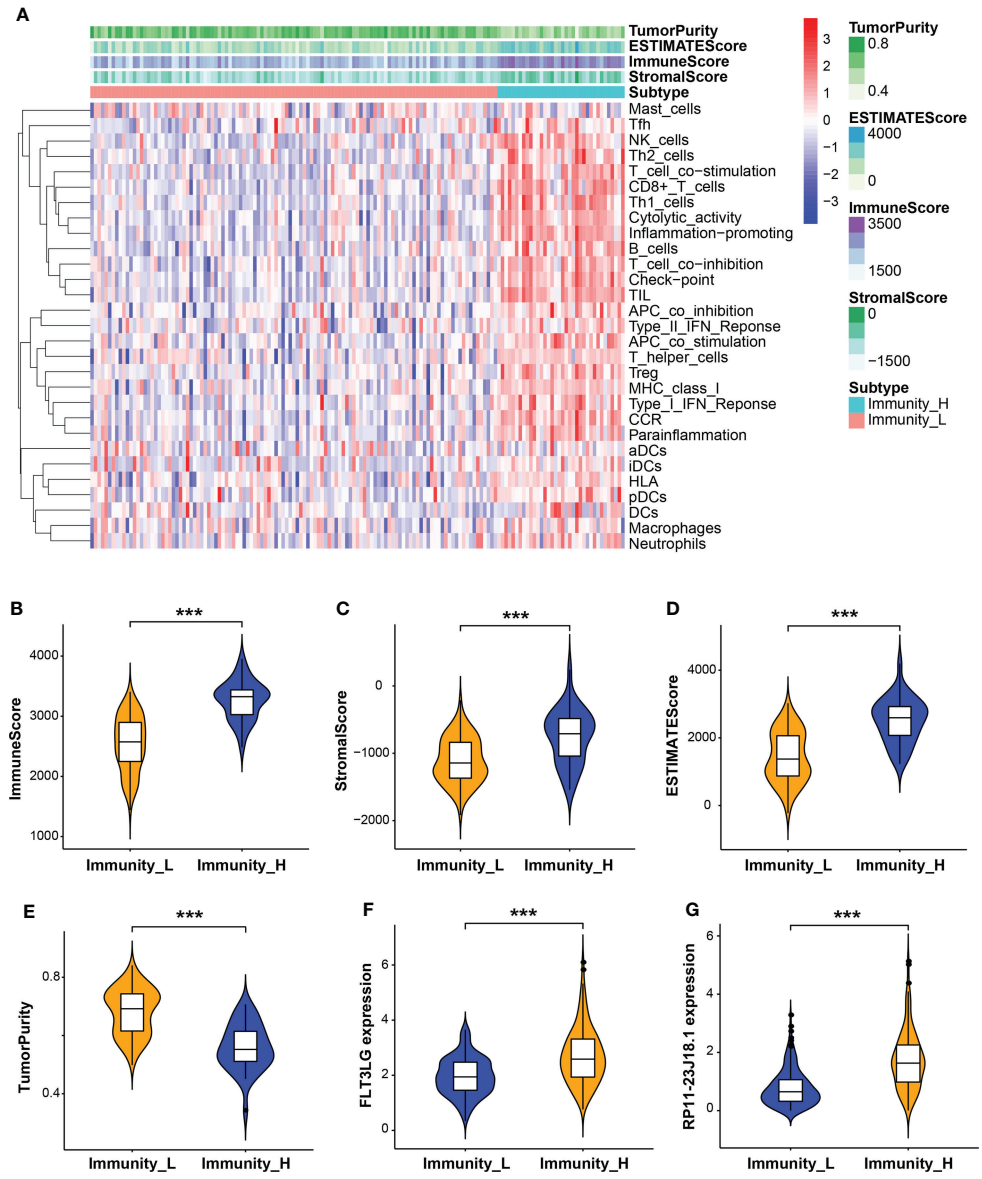


FIGURE 4
Evaluation of tumor microenvironment (TME). **(A)** Heatmap of immune cell, stromal cell and ESTIMATE scores and tumor purity in high and low immune activity groups. **(B–E)** Relationship between immune cell, stromal cell and ESTIMATE scores and tumor purity with immune activity in the TME. **(F, G)** Relationship between FLT3LG and RP11-23J18.1 expression and immune activity. ***$p < 0.001$.

RNA m6A and DNA methylation regulate gene expression

FLT3LG, IFITM3P6 and CBX7 were downregulated in AML samples, while FLT3 was upregulated. The mechanism underlying the cause of aberrant gene expression is worthy of exploring. We found that FLT3 expression is negatively correlated with its DNA methylation level (R = -0.338, P = 5.943e-04) (Figure 8A), in which the main methylation site is cg26472910 (Figure 8B). However, FLT3LG, IFITM3P6 and CBX7 expression presented non-significant correlations with their DNA methylation level.

Furthermore, the relationship between the expression of these genes and m6A-related gene expression was assessed, found that FLT3LG was negatively correlated with m6A writer and m6A reader genes, including RBMX (writer, R = -0.416, P = 1.068e-07, Figure 8C), ZC3H13 (writer, R = -0.311, P = 9.98e-05, Figure 8D), HNRNPC (reader, R = -0.418, P = 9.025e-08, Figure 8E) and YTHDF2 (reader, R = -0.368, P = 3.373e-06, Figure 8F), but positively correlated with m6A eraser gene,

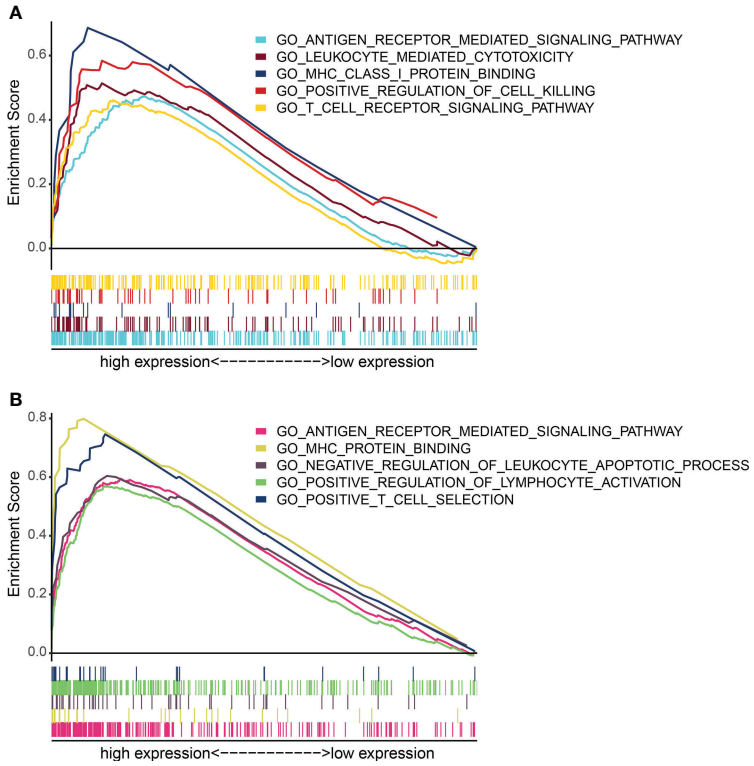


FIGURE 5
GO term enrichment analysis of FLT3LG and RP11_23J18.1 by GSEA method. **(A)** GO analysis of FLT3LG. **(B)** GO analysis of RP11_23J18.1.

ALKBH5 (R = 0.45, *P* = 6.78e-09, Figure 8G). IFITM3P6 was negatively correlated with m6A reader gene, EIF3 (R = -0.321, *P* = 5.748e-05, Figure 8H). CBX7 was negatively correlated with m6A writer and m6A reader genes, including METTL14 (writer, R = -0.383, *P* = 1.188e-06, Figure 8I), ZC3H13 (writer, R = -0.361, *P* = 5.157e-06, Figure 8J), HNRNPC (reader, R = -0.542, *P* = 6.941e-13, Figure 8K), YTHDF2 (reader, R = -0.406, *P* = 2.266e-07, Figure 8L) and EIF3 (reader, R = -0.582, *P* = 4.649e-15, Figure 8M).

FLT3LG, IFITM3P6 and CBX7 may exist coupling effects

Now that FLT3-FLT3LG and IFITM3P6-miR-6748-3p-CBX7 axes were determined in activating T cells, we were curious about whether they exist coupling effects. Interestingly, we found FLT3LG was positively correlated with IFITM3P6 (R = 0.41, *P* = 1.7e-08, Figure 9A) and CBX7 (R = 0.45, *P* = 5e-10, Figure 9B) in TCGA data. It's suggested that CBX7 may be the key node connecting FLT3LG and IFITM3P6.

Furtherly, KEGG enrichment analysis by GSEA method showed high expression of FLT3LG and IFITM3P6 were both enriched in NOTCH signaling pathway (Figures 9C, D). Meanwhile, CBX7 was determined to be positively correlated with NOTCH1 (R = 0.49, *P* = 4.9e-12, Figure 9E) and MAML2 (R = 0.53, *P* = 1.1e-13, Figure 9F).

NOTCH pathway has been well documented in activating CD4+ and CD8+ T cells. CBX7 was reported to involve in FasL suppression, the latter is an important factor suppressing T cell activation. This reminded us that NOTCH1, MAML2 and CBX7 in nucleus may form a suppressive complex to suppress FasL expression. In PPA-Pred2 database (https://www.iitm.ac.in/bioinfo/PPA_Pred/), a protein-protein affinity predictor, we retrieved their potential binding affinity by putting their protein sequences, and predicted value of Delta G (ΔG, binding free energy) between CBX7 and NOTCH1 is -12.58 kcal/mol, between CBX7 and MAML2 is -16.91 kcal/mol, between NOTCH1 and MAML2 is -30.07 kcal/mol. These provide evidence that CBX7, NOTCH1 and MAML2 exist binding potential to reduce FasL expression.

Ligand-receptor reaction is the indispensable condition for signaling pathway activation. Herein, we found Dll1 (delta like 1), the ligand of NOTCH receptor, in rat bone marrow-derived dendritic cells was significantly increased in Flt3lg-stimulating model, with a trend of dose-dependent (Figure 9G), revealing that activated dendritic cells, which is stimulated by FLT3LG-FLT3 axis, will further activate NOTCH pathway in T cells.

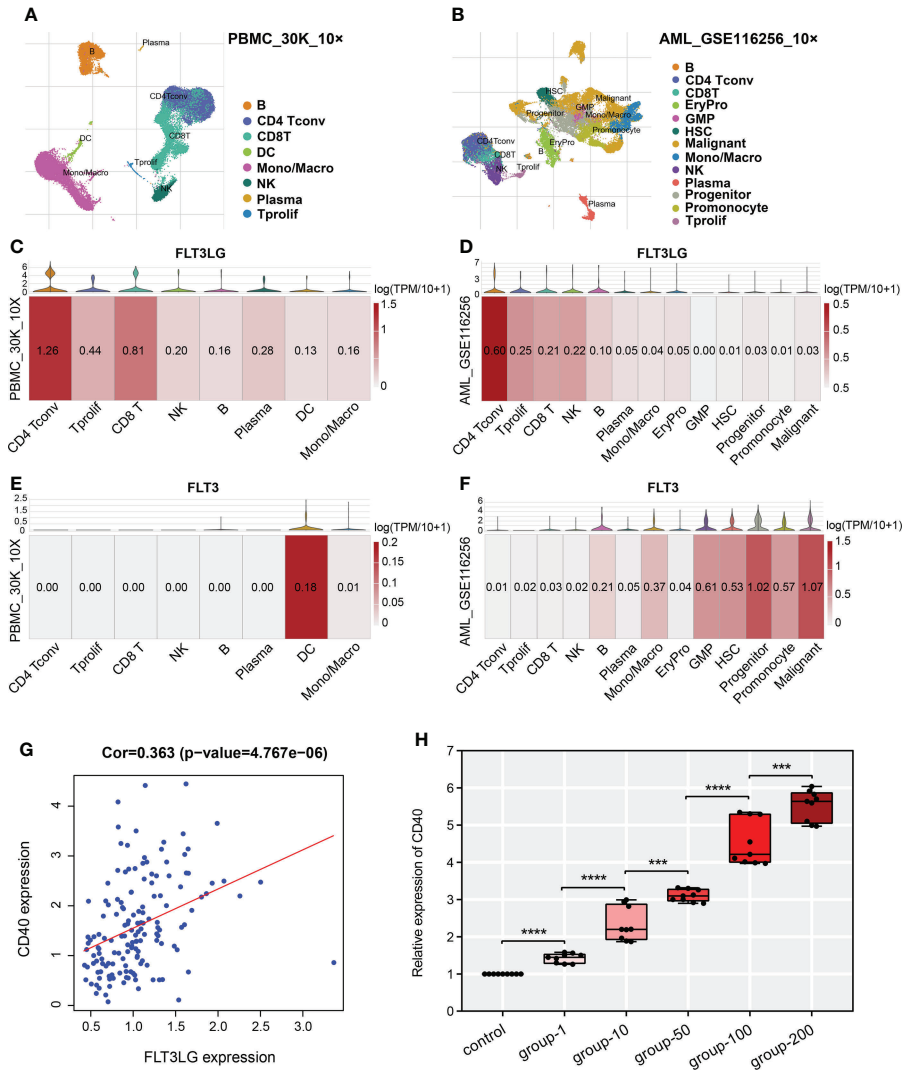


FIGURE 6
Single-cell transcriptome analysis revealed detailed gene location in various cell types. **(A, B)** UMAP cluster after cluster analysis of peripheral blood mononuclear cells (PBMC) and acute myelocytic leukemia (AML). **(C, E)** Violin plot and heatmap of FLT3LG and FLT3 expression in the PBMC_30K_10× dataset. **(D, F)** The transcription level of FLT3LG and FLT3 in cell types of the AML_GSE116256 dataset. **(G)** FLT3LG expression was positively correlated with CD40 expression in TCGA data. **(H)** CD40 expression was significantly elevated when Flt3lg cytokine activation in dendritic cells. ***$p < 0.001$; ****$p < 0.0001$.

Discussion

AML is a biologically and clinically heterogeneous disease, despite advances in supportive care and prognostic risk stratification that have been optimized for established treatments, overall long-term survival remains a challenge (3). Leukemia blasts exhibit impressive immunoediting capabilities under the selective immune pressure. The presence of T cells at the tumor site is the prerequisite for immune recognition and elimination of AML cells, also for any therapy leveraging on this condition. In a mouse model of AML, deletional CD4+ and CD8 + T cell tolerance induction is attributed to leukemia antigen presentation by immature antigen-presenting cells (DCs) or splenic CD8α+ dendritic cells (27, 28). Moreover, a recent study demonstrated loss of plasmacytoid DC differentiation was associated with persistence of the residual disease after AML treatment and unfavorable outcomes (29).

Our present research and those previously reported studies have confirmed that FMS-related tyrosine kinase 3 (FLT3) is mainly physiologically expressed in DCs (30), but pathologically expressed in malignant and progenitor cells. FLT3 ligand (FLT3LG) is predominantly produced by lymphocytes, especially T cells (31). Admittedly, FLT3LG has been shown to bind to FLT3 (receptor) on DCs to stimulate their differentiation

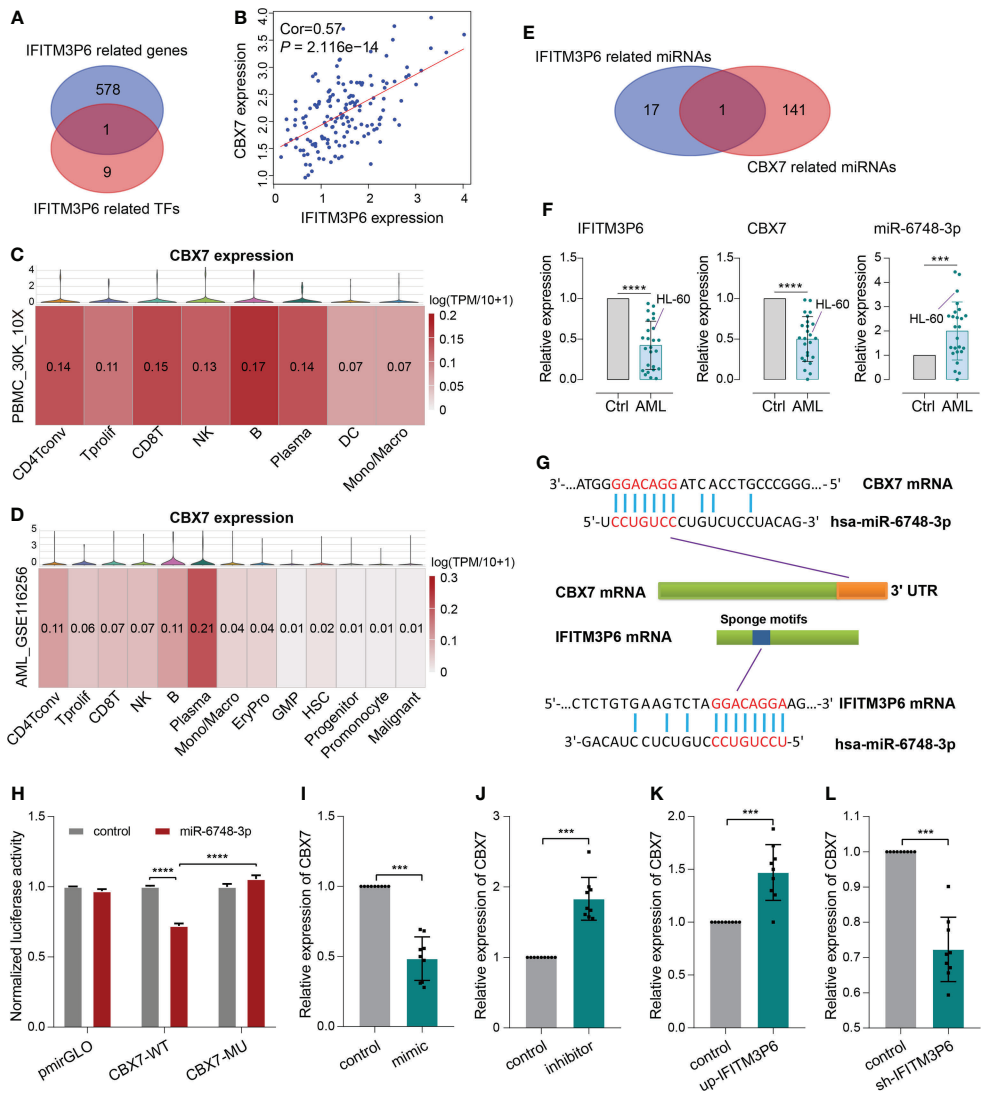


FIGURE 7
ceRNA mechanism detection of IFITM3P6 in T cell activation. **(A, B)** CBX7 was determined as a hub transcription factor that is computationally correlated with IFITM3P6. **(C, D)** Location of CBX7 in cell subtypes in healthy and AML single-cell transcriptome sequencing datasets. **(E)** Venn diagram of IFITM3P6- and CBX7-related miRNAs. **(F)** Relative expression of IFITM3P6, CBX7 and miR-6748-3p compared between the AML group and healthy control group. **(G)** Potential competing endogenous RNA (ceRNA) mechanism of IFITM3P6 and CBX7. **(H)** Dual-luciferase reporter assay revealed that miR-6748-3p could bind to the 3′ UTR of CBX7. **(I, J)** MiR-6748-3p expression was negatively correlated with CBX7 expression. **(K, L)** IFITM3P6 expression was positively correlated with CBX7 expression. ***$p < 0.001$; ****$p < 0.0001$.

and expansion, facilitating tumor antigen cross-presentation and anticancer immune responses (32). However, in a malignant context, FLT3 mutations induce receptor dimerization, resulting in constitutive activation of PI3K-AKT, RAS-MEK-MAPK and STAT-5 signaling pathways (33).

We found that low DNA methylation occurred in AML, induced upregulation of FLT3, that's a reason for eliciting pro-tumor signaling pathway activation. Targeting the methylation site (cg26472910) may lead to the elimination of FLT3 upregulation. FLT3LG expression was positively correlated with the N^6-methyladenosine (m6A) eraser gene (ALKBH5) but negatively correlated with the writer (RBMX, ZC3H13) and reader (HNRNPC, YTHDF2) genes, and induced its low expression. Considering FLT3LG is positively correlated with immune activity in the TME, this might be one of the reasons for the attenuated power of DCs in tumor antigen presentation. Regarding DCs function, it's been well documented as the strongest presenting cell in activating T cells. Hence, FLT3LG-FLT3 axis is a key way to promote DC-based cross-priming of antileukemia T cells.

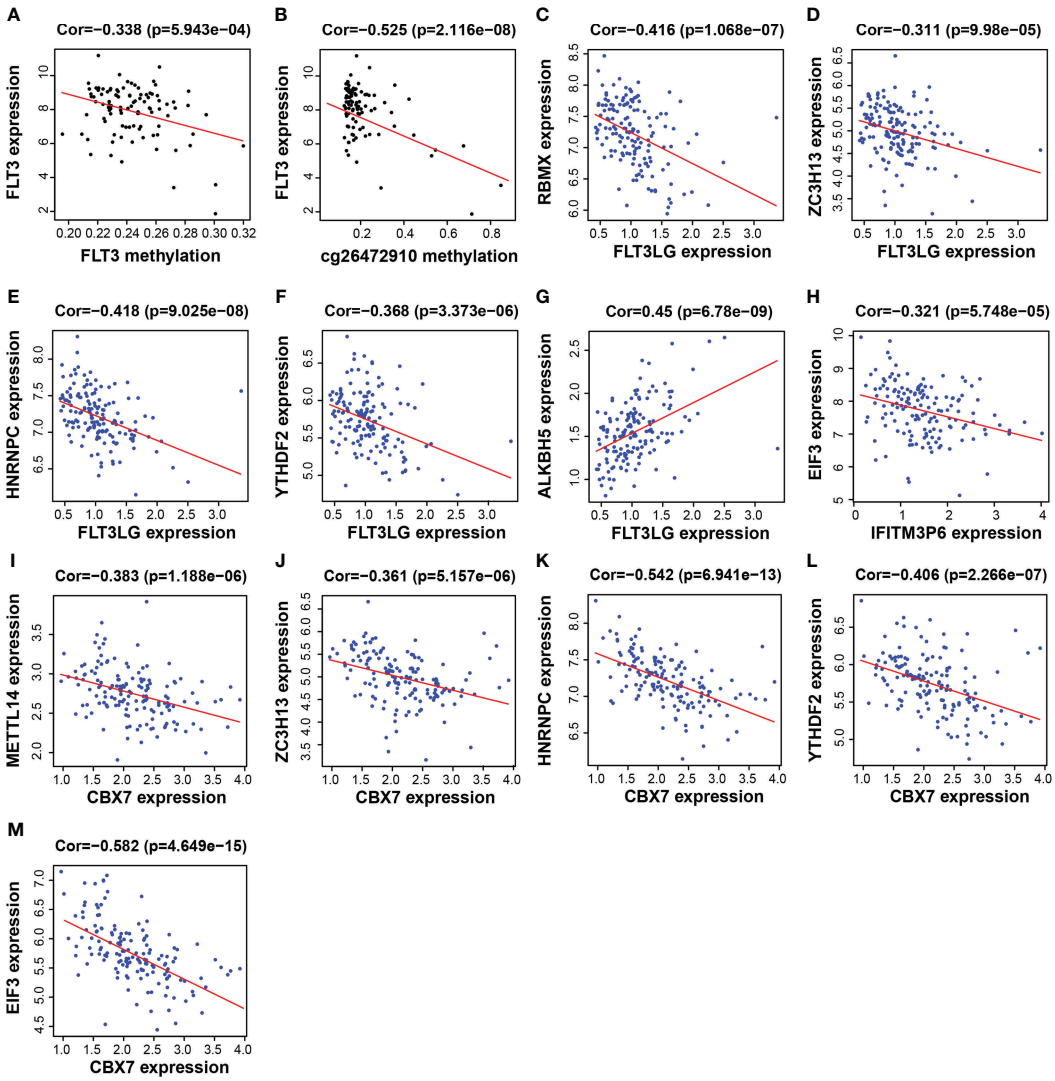


FIGURE 8
Functional mechanism analysis for FLT3LG, IFITM3P6 and CBX7. **(A, B)** Relationship between FLT3 DNA methylation and FLT3 expression and the main methylation site. **(C–G)** Relationship between FLT3LG expression and N^6-methyladenosine (m6A) methylation-related gene expression. **(H)** Relationship between IFITM3P6 expression and m6A methylation-related gene expression. **(I–M)** Relationship between CBX7 expression and m6A methylation-related gene expression.

IFITM3P6 was identified as a special type of lncRNA that regulates AML processes and positively correlated with immune activation in this study. It was predicted to be mainly in the cytoplasm and hypothesized that it may exert functions *via* the ceRNA mechanism. Based on these findings and analysis of multiple databases, we found that the IFITM3P6-miR-6748-3p-CBX7 axis may play an important role in regulating the activity of T cells.

Chromobox homolog 7 (CBX7) is a polycomb protein involved in the formation of polycomb repressive complex 1. Low expression of the CBX7 gene is associated with poor prognosis in most cancers (34). It has been reported that CBX7 represses FasL expression in CD4+ T cells, consequently preventing CD4+ T cell apoptosis (35). Here, we also found that CBX7 expression was negatively correlated with caspase 3 expression (data not presented), suggesting that the IFITM3P6-miR-6748-3p-CBX7 axis modulates T cells activity and apoptosis.

M6A is methylation that occurs in the N^6-position of adenosine, which is the most prevalent internal modification in eukaryotic mRNA. Accumulating evidence suggests that m6A modulates gene expression (36). We found that EIF3 (m6A reader) was reported to be upregulated in AML (37) and was negatively correlated with IFITM3P6, suggesting that m6A

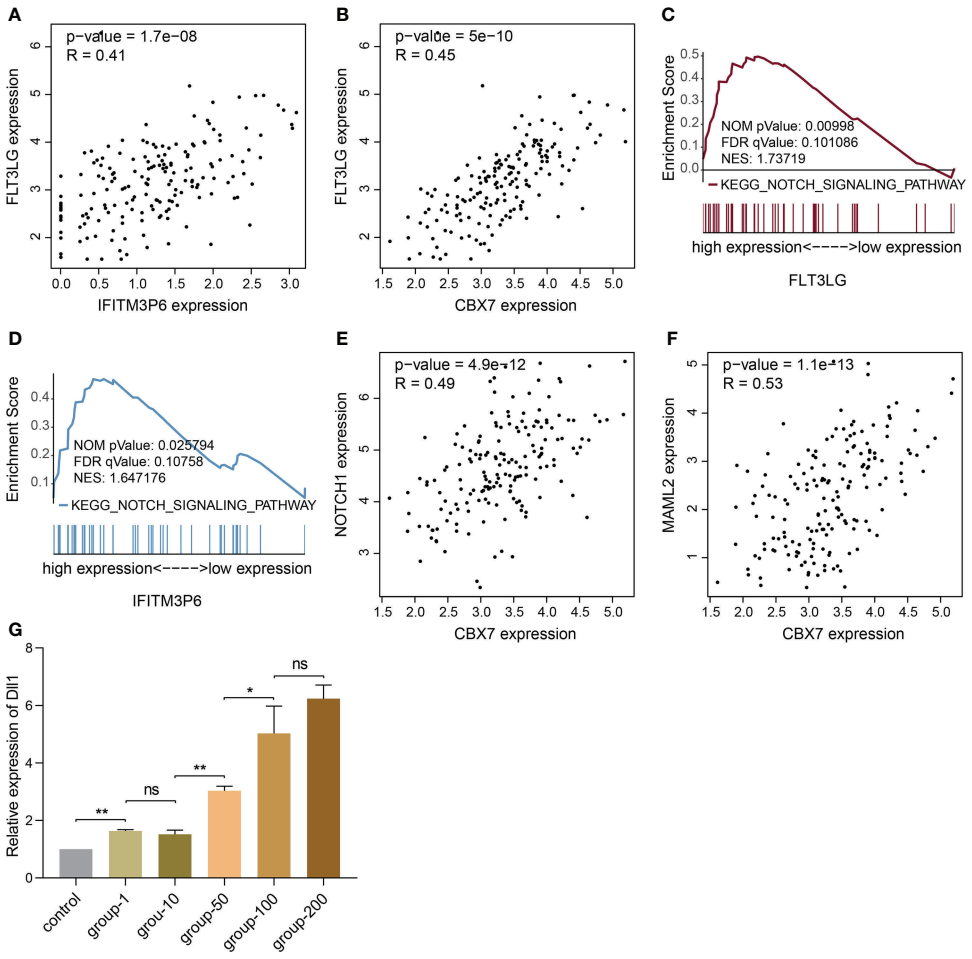


FIGURE 9
The potential relationship of FLT3LG and IFITM3P6 in activating T cells. **(A)** FLT3LG is significantly correlated with IFITM3P6 in TCGA data. **(B)** FLT3LG is positively correlated with CBX7. **(C, D)** FLT3LG and IFITM3P6 are enriched in NOTCH signaling pathway. **(E, F)** CBX7 is positively correlated with NOTCH1 and MAML2. **(G)** Dll1 is increased in rat bone marrow-derived dendritic cells along with elevated concentration (μg/ml) of Flt3lg stimulation. *$p < 0.05$; **$p < 0.01$; ns, no significance.

modification could reduce the expression of IFITM3P6. m6A writer and reader genes, including METTL14 (writer), ZC3H13 (writer), HNRNPC (reader) and EIF3 (reader) were also negatively correlated with CBX7 expression. In addition, they are upregulated in AML (38, 39). METTL14 and ZC3H13 are accessory subunits that form a stable complex with METTL3 and play key roles in substrate recognition (40–42). HNRNPC is mainly involved in mediating mRNA splicing (43). Consequently, m6A modification in combination with IFITM3P6-miR-6748-3p-CBX7 regulatory axis downregulates CBX7 in AML, thereby prompting T cells apoptosis.

In a summary, the downregulation of FLT3LG induced by m6A modification and the ectopic expression of FLT3 in malignant cells and progenitor cells ultimately decrease the ability of DCs to present antigen to T cells. What's more, m6A modification induces the downregulation of IFITM3P6 to regulate the expression of CBX7 through the ceRNA mechanism, thereby inducing T cell apoptosis in AML. Additionally, we found that FLT3LG and IFITM3P6 were both enriched in NOTCH signaling pathway. Evidence showed they may a exert synergy effect by promoting the formation of NOTCH1-MAML2-CBX7 suppressive transcription system, combining with the canonical pathway molecule like RBP-Jκ, to block FasL expression.

The role of IFITM3P6-miR-6748-3p-CBX7 and FLT3LG-FLT3 axes in AML was revealed for the first time, but the experimental verification was not solid. Whether these key genes could exert function on AML cells directly, such as the landscape of exosome of IFITM3P6 and/or miR-6748-3p secreted by T cells to interact with tumor cells, these questions remain to be explored. What's more, the FLT3 gene located near the membrane region mutation has been found in about 30% of

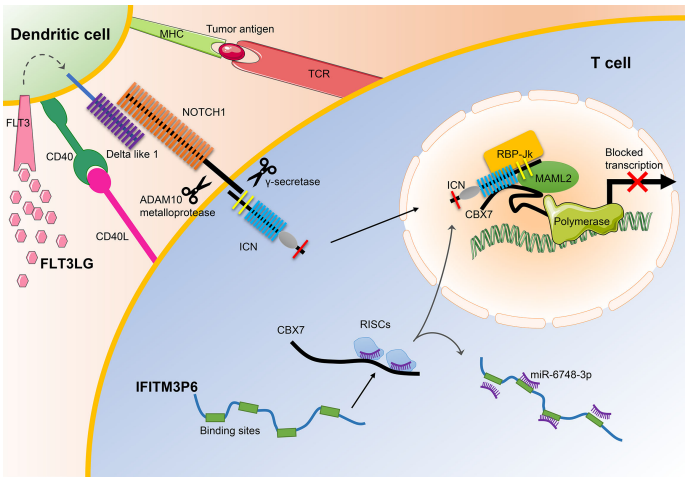


FIGURE 10
The main mechanisms of FLT3LG and IFITM3P6 in T cell activation. On the one hand, increased DLL1 by stimulation of FLT3LG to FLT3 in dendritic cells, transduct signals to NOTCH1 receptor expressed by T cells. Ligand/receptor binding triggers sequential proteolytic cleavage of NOTCH receptor, first by the ADAM10 metalloprotease and by the γ-secretase complex. These cleavages release intracellular NOTCH (ICN) into the nucleus. On the other hand, IFITM3P6 as a kind of ceRNA inhibit targeted degradation of CBX7 by miR-6748-3p. They finally promote the formation of NOTCH1-MAML2-CBX7 complex system, and suppress expression of FasL.

patients with AML. It is an internal tandem duplication (ITD) mutation composed of exon 11, intron 11 and exon 12, which is called the FLT3/ITD gene mutation. The impact of ITD on FLT3 was not evaluated in the current study. FLT3-TKD (tyrosine kinase domain) is another form of mutation. These two mutations are associated with recurrence and drug resistance.

In general, our results demonstrate that IFITM3P6 and FLT3LG might serve as prognostic markers in AML and may be used as potential therapeutic targets for the treatment of leukemia in the future (Figure 10).

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 Ethics committee of Children's Hospital of Nanjing Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

HC: Conceptualization, methodology, software, writing-original draft, supervision, writing-review & editing. MW: Methodology, validation, investigation, writing-original draft, funding acquisition. HX: Writing-review & editing, visualization, funding acquisition. SD: Methodology, investigation, validation. GL: Formal analysis. GZ: Data curation, software, writing-review & editing, funding acquisition. YZ: Methodology, software. XH: Data curation. DY: Conceptualization, methodology, software, data curation, supervision, writing-original draft, writing-review & editing, project administration, funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Academic Technology Program of Nanjing Medical University (No. NMUB2020096) and Nanjing Medical Key Science and Technology Development Project Fund (No. YKK19105), The Recruitment Program of Overseas High-Level Young Talents, "Innovative and Entrepreneurial Team" (No. (2018)2015), Science and Technology Grant of Jiangsu Province (BE2019758), Beijing Xisike Clinical Oncology Research Foundation, Shaanxi

Province Hundred Talents and 'Young Top Talent' of Xi'an Jiaotong University.

Acknowledgments

We thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.980911/full#supplementary-material

References

1. Yang X, Wang J. Precision therapy for acute myeloid leukemia. *J Hematol Oncol* (2018) 11(1):3. doi: 10.1186/s13045-017-0543-7

2. Horibe K, Saito AM, Takimoto T, Tsuchida M, Manabe A, Shima M, et al. Incidence and survival rates of hematological malignancies in Japanese children and adolescents (2006-2010): Based on registry data from the Japanese society of pediatric hematology. *Int J Hematol* (2013) 98(1):74–88. doi: 10.1007/s12185-013-1364-2

3. De Kouchkovsky I, Abdul-Hay M. 'Acute myeloid leukemia: A comprehensive review and 2016 update'. *Blood Cancer J* (2016) 6(7):e441. doi: 10.1038/bcj.2016.50

4. Gregory TK, Wald D, Chen Y, Vermaat JM, Xiong Y, Tse W. Molecular prognostic markers for adult acute myeloid leukemia with normal cytogenetics. *J Hematol Oncol* (2009) 2:23. doi: 10.1186/1756-8722-2-23

5. Riva L, Luzi L, Pelicci PG. Genomics of acute myeloid leukemia: The next generation. *Front Oncol* (2012) 2:40. doi: 10.3389/fonc.2012.00040

6. Pelcovits A, Niroula R. Acute myeloid leukemia: A review. *R I Med J (2013)* (2020) 103(3):38–40.

7. Swaminathan M, Wang ES. Novel therapies for aml: A round-up for clinicians. *Expert Rev Clin Pharmacol* (2020) 13(12):1389–400. doi: 10.1080/17512433.2020.1850255

8. Zeng Z, Shi YX, Samudio IJ, Wang RY, Ling X, Frolova O, et al. Targeting the leukemia microenvironment by Cxcr4 inhibition overcomes resistance to kinase inhibitors and chemotherapy in aml. *Blood* (2009) 113(24):6215–24. doi: 10.1182/blood-2008-05-158311

9. Buckley SA, Kirtane K, Walter RB, Lee SJ, Lyman GH. Patient-reported outcomes in acute myeloid leukemia: Where are we now? *Blood Rev* (2018) 32(1):81–7. doi: 10.1016/j.blre.2017.08.010

10. Hao F, Sholy C, Wang C, Cao M, Kang X. The role of T cell immunotherapy in acute myeloid leukemia. *Cells* (2021) 10(12):3376. doi: 10.3390/cells10123376

11. Lamble AJ, Lind EF. Targeting the immune microenvironment in acute myeloid leukemia: A focus on T cell immunity. *Front Oncol* (2018) 8:213. doi: 10.3389/fonc.2018.00213

12. Barbui AM, Borleri G, Conti E, Ciocca A, Salvi A, Mico C, et al. Clinical grade expansion of Cd45ra, Cd45ro, and Cd62l-positive T-cell lines from hla-compatible donors: High cytotoxic potential against aml and all cells. *Exp Hematol* (2006) 34(4):475–85. doi: 10.1016/j.exphem.2005.12.012

13. Graf C, Heidel F, Tenzer S, Radsak MP, Solem FK, Britten CM, et al. A neoepitope generated by an Flt3 internal tandem duplication (Flt3-itd) is recognized by leukemia-reactive autologous Cd8+ T cells. *Blood* (2007) 109(7):2985–8. doi: 10.1182/blood-2006-07-032839

14. Greiner J, Ono Y, Hofmann S, Schmitt A, Mehring E, Gotz M, et al. Mutated regions of nucleophosmin 1 elicit both Cd4(+) and Cd8(+) T-cell responses in patients with acute myeloid leukemia. *Blood* (2012) 120(6):1282–9. doi: 10.1182/blood-2011-11-394395

15. Zhong RK, Lane TA, Ball ED. Generation of T-cell lines to autologous acute myeloid leukemia cells by competitive limiting dilution culture of acute myeloid leukemia mononuclear cells. *Exp Hematol* (2008) 36(4):486–94. doi: 10.1016/j.exphem.2007.11.012

16. Zhang H, Deng T, Liu R, Ning T, Yang H, Liu D, et al. Caf secreted mir-522 suppresses ferroptosis and promotes acquired chemo-resistance in gastric cancer. *Mol Cancer* (2020) 19(1):43. doi: 10.1186/s12943-020-01168-8

17. Hernandez-Camarero P, Lopez-Ruiz E, Marchal JA, Peran M. Cancer: A mirrored room between tumor bulk and tumor microenvironment. *J Exp Clin Cancer Res* (2021) 40(1):217. doi: 10.1186/s13046-021-02022-5

18. Liu L, Zhang Z, Zhou L, Hu L, Yin C, Qing D, et al. Cancer associated fibroblasts-derived exosomes contribute to radioresistance through promoting colorectal cancer stem cells phenotype. *Exp Cell Res* (2020) 391(2):111956. doi: 10.1016/j.yexcr.2020.111956

19. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337

20. Zhao SG, Lehrer J, Chang SL, Das R, Erho N, Liu Y, et al. The immune landscape of prostate cancer and nomination of pd-L2 as a potential therapeutic target. *J Natl Cancer Inst* (2019) 111(3):301–10. doi: 10.1093/jnci/djy141

21. Zhang S, Zhang E, Long J, Hu Z, Peng J, Liu L, et al. Immune infiltration in renal cell carcinoma. *Cancer Sci* (2019) 110(5):1564–72. doi: 10.1111/cas.13996

22. Langfelder P, Horvath S. Wgcna: An r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

23. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carre C, et al. Rna-seq signatures normalized by mrna abundance allow absolute deconvolution of human immune cell types. *Cell Rep* (2019) 26(6):1627–40 e7. doi: 10.1016/j.celrep.2019.01.041

24. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic rna interference reveals that oncogenic kras-driven cancers require Tbk1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460

25. Zhao K, Yang H, Kang H, Wu A. Identification of key genes in thyroid cancer microenvironment. *Med Sci Monit* (2019) 25:9602–8. doi: 10.12659/MSM.918519

26. Sun D, Wang J, Han Y, Dong X, Ge J, Zheng R, et al. Tisch: A comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res* (2021) 49(D1):D1420–D30. doi: 10.1093/nar/gkaa1020

27. Zhang L, Chen X, Liu X, Kline DE, Teague RM, Gajewski TF, et al. Cd40 ligation reverses T cell tolerance in acute myeloid leukemia. *J Clin Invest* (2013) 123(5):1999–2010. doi: 10.1172/JCI63980

28. Kline DE, MacNabb BW, Chen X, Chan WC, Fosco D, Kline J. Cd8alpha(+) dendritic cells dictate leukemia-specific Cd8(+) T cell fates. *J Immunol* (2018) 201(12):3759–69. doi: 10.4049/jimmunol.1801184

29. Xiao W, Goldberg AD, Famulare CA, Devlin SM, Nguyen NT, Sim S, et al. Loss of plasmacytoid dendritic cell differentiation is highly predictive for post-induction measurable residual disease and inferior outcomes in acute myeloid leukemia. *Haematologica* (2019) 104(7):1378–87. doi: 10.3324/haematol.2018.203018

30. Abrahamsson H, Jensen BV, Berven LL, Nielsen DL, Saltyte Benth J, Johansen JS, et al. Antitumour immunity invoked by hepatic arterial infusion of first-line oxaliplatin predicts durable colorectal cancer control after liver metastasis ablation: 8-12 years of follow-up. *Int J Cancer* (2020) 146(7):2019–26. doi: 10.1002/ijc.32847

31. Barry KC, Hsu J, Broz ML, Cueto FJ, Binnewies M, Combes AJ, et al. A natural killer-dendritic cell axis defines checkpoint therapy-responsive tumor microenvironments. *Nat Med* (2018) 24(8):1178–91. doi: 10.1038/s41591-018-0085-8

32. Pol JG, Le Naour J, Kroemer G. Flt3lg - a biomarker reflecting clinical responses to the immunogenic cell death inducer oxaliplatin. *Oncoimmunology* (2020) 9(1):1755214. doi: 10.1080/2162402x.2020.1755214

33. Zhou J, Bi C, Janakakumara JV, Liu SC, Chng WJ, Tay KG, et al. Enhanced activation of stat pathways and overexpression of survivin confer resistance to Flt3 inhibitors and could be therapeutic targets in aml. *Blood* (2009) 113(17):4052–62. doi: 10.1182/blood-2008-05-156422

34. Pallante P, Forzati F, Federico A, Arra C, Fusco A. Polycomb protein family member Cbx7 plays a critical role in cancer progression. *Am J Cancer Res* (2015) 5 (5):1594–601.

35. Li J, Li Y, Cao Y, Yuan M, Gao Z, Guo X, et al. Polycomb chromobox (Cbx) 7 modulates activation-induced Cd4+ T cell apoptosis. *Arch Biochem Biophys* (2014) 564:184–8. doi: 10.1016/j.abb.2014.10.004

36. He L, Li H, Wu A, Peng Y, Shu G, Yin G. Functions of N6-methyladenosine and its role in cancer. *Mol Cancer* (2019) 18(1):176. doi: 10.1186/s12943-019-1109-9

37. Weng H, Huang H, Wu H, Qin X, Zhao BS, Dong L, et al. Mettl14 inhibits hematopoietic Stem/Progenitor differentiation and promotes leukemogenesis *Via* mrna M(6)a modification. *Cell Stem Cell* (2018) 22(2):191–205.e9. doi: 10.1016/j.stem.2017.11.016

38. Ianniello Z, Fatica A. N6-methyladenosine role in acute myeloid leukaemia. *Int J Mol Sci* (2018) 19(8):2345. doi: 10.3390/ijms19082345

39. Zheng X, Gong Y. Functions of rna N(6)-methyladenosine modification in acute myeloid leukemia. *biomark Res* (2021) 9(1):36. doi: 10.1186/s40364-021-00293-w

40. Wang P, Doxtader KA, Nam Y. Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol Cell* (2016) 63(2):306–17. doi: 10.1016/j.molcel.2016.05.041

41. Knuckles P, Lence T, Haussmann IU, Jacob D, Kreim N, Carl SH, et al. Zc3h13/Flacc is required for adenosine methylation by bridging the mrna-binding factor Rbm15/Spenito to the M(6)a machinery component Wtap/Fl(2)D. *Genes Dev* (2018) 32(5-6):415–29. doi: 10.1101/gad.309146.117

42. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A Mettl3-Mettl14 complex mediates mammalian nuclear rna N6-adenosine methylation. *Nat Chem Biol* (2014) 10(2):93–5. doi: 10.1038/nchembio.1432

43. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. N(6)-Methyladenosine-Dependent rna structural switches regulate rna-protein interactions. *Nature* (2015) 518(7540):560–4. doi: 10.1038/nature14234

COPYRIGHT

© 2022 Chen, Wu, Xia, Du, Zhou, Long, Zhu, Huang 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 | Frontiers in Immunology

TYPE Original Research
PUBLISHED 26 August 2022
DOI 10.3389/fimmu.2022.977338

Check for updates

OPEN ACCESS

EDITED BY
Hao Zhang,
Xiangya Hospital Central South University, China

REVIEWED BY
Xuehui Hong,
Xiamen University, China
Maoni Guo,
University of Macau, Macao SAR, China

*CORRESPONDENCE
Yihua Sun
600611@hrbmu.edu.cn
Shuyun Zhang
13214501198@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 24 June 2022
ACCEPTED 09 August 2022
PUBLISHED 26 August 2022

CITATION
Xia Y, Zhang R, Wang M, Li J, Dong J, He K, Guo T, Ju X, Ru J, Zhang S and Sun Y (2022) Development and validation of a necroptosis-related gene prognostic score to predict prognosis and efficiency of immunotherapy in gastric cancer. *Front. Immunol.* 13:977338. doi: 10.3389/fimmu.2022.977338

Development and validation of a necroptosis-related gene prognostic score to predict prognosis and efficiency of immunotherapy in gastric cancer

Yan Xia[1†], Rongzheng Zhang[2†], Mingzhu Wang[1], Jiaqi Li[2], Jianming Dong[2], Kaitong He[1], Ting Guo[1], Xiaomei Ju[2], Jiaqiu Ru[2], Shuyun Zhang[2*] and Yihua Sun[1*]

[1]Department of Clinical Laboratory, Harbin Medical University Cancer Hospital, Harbin, China, [2]Scientific Research Center, the Second Affiliated Hospital of Harbin Medical University, Harbin, China

Necroptosis is a novel type of regulated cell death that is intimately associated with a variety of tumors. However, how necroptosis affects the identification of gastric cancer (GC) remains unclear. Here we seek to find new potential necroptosis-related biomarkers to predict GC prognosis and immunotherapy effect. We used Cox analysis to obtain shared prognostic markers related to necroptosis from five datasets (TCGA and four GEO datasets). Then, a necroptosis-related gene prognostic score (NRGPS) system was constructed using LASSO Cox regression, NRGPS consisting of three necroptosis-related mRNAs (*AXL*, *RAI14*, and *NOX4*) was identified, 31 pairs of GC and adjacent normal tissues from the Second Hospital of Harbin Medical University were collected and Real-Time Quantitative PCR (RT-qPCR) was used to detect the relative expression levels of the three necroptosis-related mRNAs, and external validation was performed on four GEO datasets (GSE84437, GSE26901, GSE62254 and GSE15459). In this study, Overall survival (OS) in the high-NRGPS group was significantly lower than in the low-NRGPS group. Cox regression analyses showed that NRGPS was an independent prognostic variable. Tumor-mutation-burden (TMB), tumor microenvironment (TME), microsatellite instability (MSI), and Tumor Immune Dysfunction and Exclusion (TIDE) scoring were used as predictors of the immunotherapy response. A cancer-friendly immune microenvironment, a high TIDE score, a low TMB, and a low MSI were all characteristics of the high-NRGPS group, and they all consistently showed that the issues seen there are related to immune escape in GC. The combination of three candidate genes may be an effective method for diagnostic assessment of GC prognosis and immunotherapy efficacy.

KEYWORDS

necroptosis gene, gastric cancer, bioinformatics analysis, prognosis, tumor microenvironment

Introduction

Over one million new cases of gastric cancer (GC) and close to 80,000 GC-related fatalities were reported in 2020, making it one of the most prevalent cancers in the world (1). Immunotherapy has improved the treatment strategy for several types of malignancies, including GC, as ICB has the potential to induce durable immune responses in different types of cancers (2, 3). However, only one-third of individuals with the majority of malignancies benefit with checkpoint inhibitors (4). This resistance is usually through a well-established mechanism of immune evasion of the cancer cells, leading to the spread of the disease (5). GC prognosis is currently based on the Lauren and World Health Organization classification as well as the Tumor-Node-Metastasis (TNM) staging system (6). However, patients in the same stage of cancer may demonstrate significantly different prognoses. Although many patients have similar clinical characteristics and associated treatment options, outcomes may differ markedly. Thus, to improve the prognosis of GC patients, it is necessary to investigate trustworthy biomarkers that can reliably predict prognosis and identify targets for prospective treatment.

Necroptosis is a new type of programmed necrotic cell death, which is regulated differently from apoptosis. The necroptosis signaling pathway includes activated receptor-interacting protein kinases (RIPKs) as well as mixed lineage kinase domain-like pseudokinases (MLKL) (7). Necroptosis induces strong cross-priming of anti-tumor $CD8^+$ T cells by releasing damage-associated molecular patterns, thereby suppressing tumor progression (8). A recent study showed that nano-vaccines that mimic necrotic cancer cells could enhance immunity in mice through the proliferation of $NKG2D^+$ natural killer cells and $CD8^+$ T cells, ultimately enhancing the anti-tumor effects (9). Another study showed that the activation of RIPK3 results in a reliable derepression of tripartite motif-containing 28 in cancer cells, thereby inducing increased production of immunostimulatory cytokines in the tumor microenvironment (TME), thus promoting potent cytotoxic antitumor immunity (10).

Recent basic experiments and bioinformatics analyses have revealed that MLKL mRNA expression levels are dramatically lower in GC than normal tissue and that GC patients with low MLKL expression have a poorer prognosis compared to normal tissues (11). Another study found that astaxanthin caused AGS death in GC cell lines by inducing increased NADPH oxidase activity, ROS production, and phosphorylation of RIP1/RIP3/MLKL (12). This indicates that necroptosis is related to the prognosis and treatment of GC. Necroptosis is a new target for cancer therapy, considering its critical role in cancer biology (13). The quality and quantity of innate necroptosis-centered immune cell rely on the inflammatory background, tissue type, and other individual circumstances (7). However, only a small number of research have examined the connection between necroptosis-related genes (NRGs) and the prognosis and therapy of GC (14, 15).

In this study, The Cancer Genome Atlas (TCGA) was searched to develop prognostic NRGs markers for GC. We identified three mRNA markers with reliable prognostic expression and experimentally validated NRGPS using RT-qPCR assay for GC and adjacent normal tissues to detect their relative mRNA level, as well as independent external validation using the GSE84437, GSE26901, GSE62254 and GSE15459 datasets. It was finally demonstrated that NRGPS could be used as an independent prognostic indicator for GC. Our aim is to fully investigate the expression profile of necroptosis and the predicted effect of immunotherapy treatment, thus identifying potential targets for the treatment of GC. We hope to our study enable the improved stratification of GC patients, thus facilitating personalized treatment decisions.

Materials and methods

Clinical specimen collection

Thirty-one pairs of GC and healthy tissue samples from cancer were gathered and frozen in liquid nitrogen at the Second Hospital of the Harbin Medical University.All specimens were judged by experienced pathologists and informed consent and permission was obtained from the Medical Ethics Committee of the Second Affiliated Hospital of Harbin Medical University. Detailed clinicopathological data of the patients are shown in Supplementary Table S1.

Acquisition of Information of patients with GC

The flow chart of this study is shown in Figure 1. The RNA-Seq data, clinical information, and somatic mutations of GC patients (375 tumor samples, 32 normal samples) were obtained from TCGA (https://portal.gdc.cancer.gov/). The external validation cohorts GSE84437 (n = 433), GSE26901 (n = 109), GSE62254 (n = 300) and GSE15459 (n = 192) were obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/). In addition, GSE29272 (134 pairs) and GSE63089 (45 pairs) with adjacent normal tissues as controls, and the melanoma cohort (GSE78220) with anti-PD-1 checkpoint inhibition therapy were similarly derived from GEO. Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/) contains data from TCGA and Genotype-Tissue Expression (GTEx). 583 NRGs sets were derived from GeneCards (https://www.genecards.org/). After 12 non-coding RNAs were removed, 571 necroptosis-related mRNAs were obtained.

Identication of prognosis markers

First, we extracted necroptosis-related gene expression data from TCGA and GEO cohorts. Then, the non-technical batch effect between TCGA and GEO data was corrected before analysis by using the "sva" R package. In addition, we also intersect the gene symbols of TCGA cohort with GEO cohorts to ensure that the genes obtained from subsequent analysis are shared by the five data sets. Next, univariable Cox regression analysis was performed for each cohort to further screen for potential prognostic markers. The venn diagrams were drawn using the "VennDiagram" R package to show common prognostic markers in all five cohorts.

Construction and validation of the necroptosis-related gene prognostic score system

The Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression method was used to create the NRGPS system, and tenfold cross validation was used to estimate the ideal coefficients based on the partial likelihood deviance. To determine the patient risk scores, regression coefficients of the genes and matching mRNA expression levels were employed.The formula for calculating NRGPS is as follows:

$$NRGPS = \sum_{i=1}^{n} \text{coef}(\text{gene}i) * \text{expr}(\text{gene}i)$$

The median NRGPS was used to separate the high-NRGPS group from the low-NRGPS group. Using the Kaplan-Meier curve, the survival status of both the high-NRGPS and low-NRGPS groups was examined. The prognostic significance of the created marker for predicting survival was evaluated using a time-dependent receiver operating characteristic (ROC) curve analysis. The independent prognostic significance of NRGPS was assessed using univariable and multivariable Cox regression analysis.At the same time, we performed the same analysis on four GEO cohorts to verify. The above analysis was performed using the "survminer", "survival", and "timeROC" R packages.

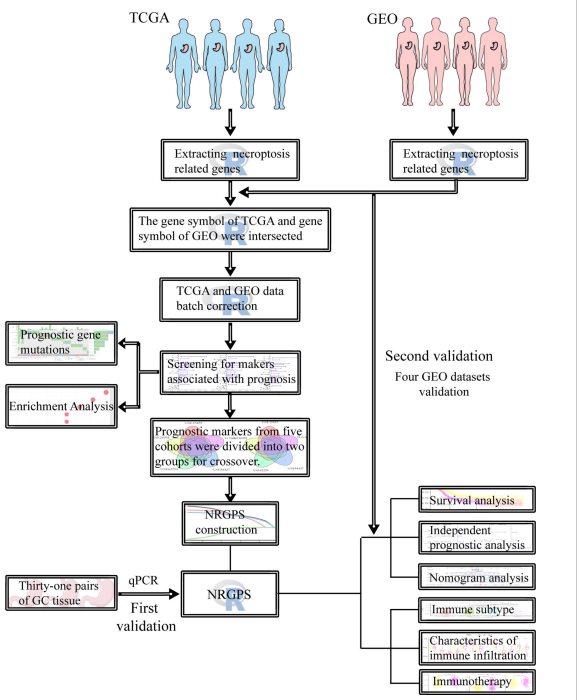


FIGURE 1
Flowchart of this study.

Detection of NRGs mRNA expression levels in GC tissues by RT-qPCR

Total RNA was extracted from GC and Normal using Trizol reagent (SM129-02, Sevenbio, China); cDNA was synthesized using a reverse transcription kit (1119ES60, Yeasen, China); SYBR Green Master Mix kit (11184ES03, Yeasen, China) and RT-qPCR instrument (SLAN-96p Shanghai hongshi, China) were used for RT-qPCR, glyceraldehyde-3phosphate dehydrogenase (GAPDH) was used as an internal control, and three NRGs quantification was based on the $2^{-\Delta\Delta Ct}$ method, and the primer sequences are shown in Supplementary Table 2.

Function enrichment analysis

The "clusterProfiler", "enrichplot", and "ggplot2" R packages were used to conduct the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway studies (16, 17). Then, we performed a Gene Set Enrichment Analysis (GSEA) to investigate potential biological courses.c2.cp.kegg.v7.4.symbols.gmt was chosen as the reference file, and meaningful biological processes and pathways were enriched to FDR< 0.05. The protein-protein interactions (PPI) of prognostic markers were shown using a string database (https://www.string-db.org/).

Development of nomogram based on NRGPS and clinical features

Nomogram was created to predict the survival probability of gastric cancer patients at 3 and 5 years. The ROC curve of nomogram is drawn to evaluate its accuracy. The "rms","regplot", "survival", "survminer" and "timeROC" R packages were used for the above analysis.

Correlation analysis to identify an association between NRGPS and immune typing

Based on immune typing files (Subtype_Immune_Model_Based.txt), GC patients were classified into six subtypes (18). Then observe whether there are differences in NRGPS between patients with different immune subtypes.

Immune cell infiltration analysis

First, on the TCGA cohort, it was done using single sample gene set enrichment analysis (ssGSEA).The relative abundance of 16 immune cell infiltrates and the activity of 13 immunological-related pathways in the TME were represented by enrichment scores.Next, the correlation between three candidate genes and NRGPS and immune cells and pathways was analyzed. The relationship between immune cell infiltration and GC patients' survival time was further analyzed. Immune cell infiltration and NRGPS were combined to perform a survival analysis of the GC patients. The absolute mode of cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT-ABS) technique was utilized to confirm the immune cells' invasion of TME. Finally, to confirm the relationship between the three candidate genes and macrophages, estimating the proportion of Immune and cancer cells (EPIC), tumor immune estimation resource (TIMER), comprehensive bioinformatic deconvolution (xCELL), and microenvironment cell populations- counter (MCPCOUNTER) techniques were employed. The data of CIBERSORT-ABS in TCGA cohort comes from TIMER (http://timer.cistrome.org/). The EPIC, TIMER, xCELL and MCPCOUNTER algorithms are implemented based on the TIMER website.

Prediction of immunotherapy effect

Based on the somatic mutation data downloaded from TCGA, the number of mutations in each gene in the sample was first counted to analyze the mutation status in the high- and low-NRGPS groups. Then, the "limma", "ggplot2", "ggpubr", and "ggExtra" R packages were used for the following: analyze the difference in the TMB of the high- and low-NRGPS groups; identify a correlation between NRGPS and TMB; Then, The link between TMB and prognosis was assessed using the TMB value and the accompanying survival data. Finally, NRGPS and TMB were combined to perform a survival analysis of the GC samples.

We plotted the percentage histogram of microsatellite instability (MSI) states of GC patients in the high- and low-NRGPS groups. Then, the boxplot is drawn to show the NRGPS differences of different MSI groups. The MSI status of GC patients in the TCGA cohort was downloaded from The Cancer Immunome Atlas (TCIA; https://tcia.at/home) database.

We further analyzed the correlation between the NRGPS and immune checkpoints and plotted the correlation plot using the "corrplot" R package. Then, the expression of immune checkpoints between high- and low-NRGPS groups was compared.

Immune escape and treatment were evaluated by subjecting the high- and low-NRGPS groups to tumor immune dysfunction and exclusion (TIDE; http://tide.dfci.harvard.edu/). The TIDE score, dysfunction score, and immune exclusion score were obtained from TIDE.

The GSE78220 cohort was used to verify the effect of NRGPS in predicting immunotherapy. We divided the GSE78220 cohort into high- and low-NRGPS groups according to the median NRGPS, and plotted the survival curve and ROC curve. Then, the boxplot was drawn to compare the NRGPS of patients who responded and did not respond to the treatment. In addition, we drew a histogram and compared whether there was a difference between the proportion of patients who responded and did not respond to the treatment in the high- and low-NRGPS groups.

Statistical analysis

Statistical analysis of this study was conducted by R (Version 4.1.2) and SPSS software (version 25.0). Including Cox regression analysis, Lasso analysis, Kaplan-Meier survival analysis, ROC curve analysis, independent prognostic analysis, functional analysis, nomogram analysis, immune cell infiltration analysis, correlation analysis, TMB analysis and TIDE analysis. To compare the differences between the two groups of data, we used the Wilcoxon test. The Spearman method was used for correlation analysis. RT-qPCR data did not conform to normal distribution, and non-parametric Wilcoxon's matched-pairs test was conducted. The paired samples of GSE29272 and GSE63089 also used non-parametric Wilcoxon's matched-pairs test. Fisher's exact test was used to determine the proportion of patients who responded to treatment in the high- and low-NRGPS groups of the GSE29272 cohort. The R packages and statistical methods used by GEO validation cohorts are consistent with TCGA. $p< 0.05$ was considered statistically significant.

Results

Identification of prognosis markers and functional enrichment analysis

To explore the prognostic value of the NRGs, we conducted a univariable Cox regression analysis for GC patients in TCGA cohort and GEO cohorts. The results showed that 35 NRGs were significantly associated with GC prognosis in TCGA cohort

(Figure 2A). NRGs of the four GEO cohorts significantly associated with prognosis are presented in Supplementary Table 3. To learn more about the function of prognostic markers in GC, we conducted GO and KEGG pathway enrichment analyses. The GO analysis revealed that prognostic indicators were enriched in necroptosis development and maintenance,such as "extrinsic apoptotic signaling pathway", "regulation of mRNA stability", and "mRNA stabilization" (Figure 2B). The most abundant pathways in the KEGG analysis were related to "MAPK signaling pathway", "apoptosis" and "pathogenic escherichia coli infection" (Figure 2C). Together, these findings are associated with programmed cell death and tumor progression. The PPI network of 21 nodes and 24 edges showed a complex relationship between these prognostic markers in GC (Figure 2D).

Development and validation of the NRGPS system

To better understand the mechanism by which these 35 NRGs affected the prognosis of the GC patients, we further analyzed their somatic mutation status in GC samples. The result indicated that 135 of 433 (31.18%) GC samples demonstrated genetic mutations. The missense mutation was the most common type of variation (Figure 3A).

In order to ensure the accuracy of prognostic markers, we divided prognostic markers into risk factor (HR > 1) and protective factor (HR< 1) groups according to HR values obtained by univariable Cox regression analysis, and then divided prognostic markers from five cohorts into two groups for intersection. Finally, in the risk factor group, there were five intersection genes (*TUBB6*, *LGALS1*, *AXL*, *RAI14*, and *NOX4*) in five cohorts (Figure 3B).

Lasso Cox regression was performed on the five risk factor genes to obtain the NRGs markers (Figures 3C, D). Then, an NRGPS was constructed as follows: score = 0.162 × expression quantity of *AXL* + 0.152 × expression quantity of *NOX4* + 0.153 × expression quantity of *RAI14*), with the median NRGPS of TCGA cohort as the critical value. In addition, scatter plots indicated that GC patients in the high-NRGPS group had a higher proportion of death and shorter survival time than those in the low-NRGPS group (Figures 3E, F). The Kaplan–Meier curve indicated that the overall survival (OS) of the high-NRGPS group was shorter than that of the low-NRGPS group (Figure 3G). The area under the curve (AUC) in the time-dependent ROC analysis was 0.622 at 3 years and 0.715 at 5 years (Figure 3H), indicating great specificity and sensitivity of the NRGPS in predicting the OS. To assess the value of NRGPS as a standalone prognostic marker, we performed univariable and multivariable Cox analyses to identify its correlation with age (≥ 60 years vs< 60 years), gender (male vs female), Stage (StageIII-IV vs StageI-II) in the GC patients in TCGA cohorts. Univariable Cox regression analysis showed that the NRGPS was associated with GC prognosis in TCGA cohort (Figure 4A). Multivariable Cox regression confirmed that the NRGPS was an independent predictor of survival after correcting for other clinical confounding factors in TCGA cohort (Figure 4B). In addition, we also performed a stratified analysis to assess whether NRGPS is applicable to patients with different clinical characteristics (Figure S1A).

Validation of 3 NRGs in GC tissue specimens

RT-qPCR was used to detect the relative mRNA level of 3 NRGs in 31 pairs of GC and adjacent normal tissues, Results from RT-qPCR revealed that the expression levels of *AXL*, *RAI14*, and *NOX4* were higher in GC tissues than in nearby normal tissues next to cancer (Figure 4C), further demonstrating from the perspective of basic experiments that the reliability of NRGPS for determining GC prognosis. In addition, we also added three data sets and obtained the same trend results (Figures 4D–F).

GEO external verification of NRGPS

To evaluate the stability of NRGPS—constructed from TCGA cohort—as a prognostic biomarker for GC, four independent GEO data sets were used for validation. Then, according to the median NRGPS of TCGA cohort, 240 GC patients were included in the high-NRGPS group and 191 GC patients were included in the low-NRGPS group in GSE84437 cohort; 51 GC patients were included in high-NRGPS group and 58 GC patients were included in the low-NRGPS group in GSE26901 cohort; 140 GC patients were included in the high-NRGPS group and 160 GC patients were included in the low-NRGPS group in GSE62254 cohort; 114 GC patients were included in the high-NRGPS group and 77 GC patients were included in the low-NRGPS group in GSE15459 cohort. The TCGA cohort's NRGPS distribution, survival status, and survival time were comparable in the GEO cohorts (Figures 5A–D). The results proved that the NRGPS has excellent stability.

In addition, we also analyzed Kaplan–Meier curve and the AUC in the time-dependent ROC analysis of four GEO cohorts to further verify NRGPS. As shown in Figures 5E–H, the OS of the high-NRGPS group was lower than of the low-NRGPS group.The AUC in the time-dependent ROC curve of the four cohorts showed that NRGPS had excellent predictive efficacy (Figures 5I–L). Finally, we performed univariable and multivariable Cox analyses on four GEO validation cohorts to further determine the independent predictors that NRGPS can be used (Figures 6A–H).

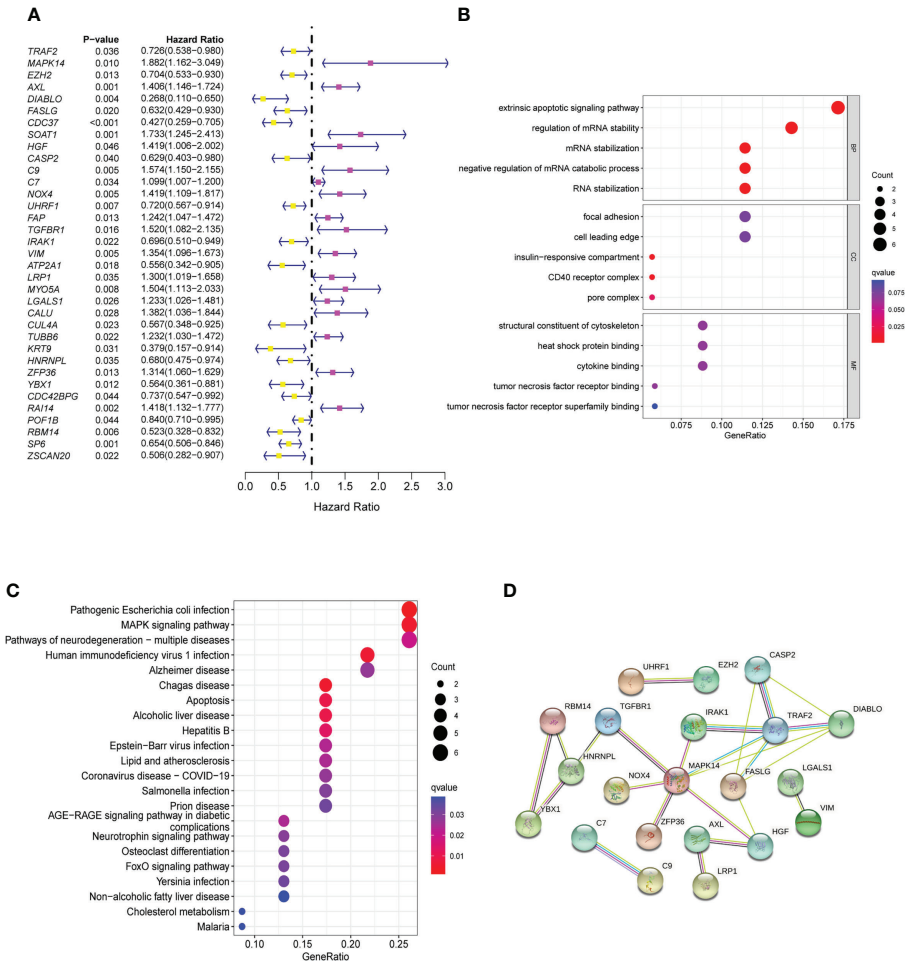


FIGURE 2
Identification and functional enrichment analysis of prognostic markers. **(A)** The forestmap of 35 prognostic markers of GC patients in the TCGA cohort. **(B)** The GO enrichment analysis of the prognostic markers. **(C)** The KEGG pathway analysis of the prognostic markers. **(D)** The PPI network was constructed through 35 prognostic markers. The interaction score was set to 0.4. GC, Gastric cancer; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction.

Nomogram and GSEA analysis of NRGPS

To help in determining each patient's specific prognosis for GC, we developed a nomogram based on NRGPS and clinical characteristics. In order to accurately evaluate the prognosis of each patient with GC, we constructed a 3-year and 5-year prognostic nomogram model combined with age, gender, stage and NRGPS (Figure 7A). For example, when a 60-year-old male patient is in stage III and NRGPS of 2.7, he will get 178 total points, which means that his probability of survival in less than 3 years and less than 5 years is 0.477 and 0.6, respectively. Next, we demonstrate the nomogram model's predictive power using a time-dependent ROC analysis. The AUC is 0.710 in 3 years and 0.708 in 5 years (Figure 7B). At the same time, four GEO cohorts were used to verify the Nomogram model (Figures S2, S3). For GC patients in the GSE15459 cohort, a Nomogram model was used to predict the prognosis, and the AUC reached 0.872 in 3 years, 0.891 in 5 years (Figure S3B). These results suggest that the nomogram has accurate and stable prognostic ability.

Based on the KEGG gene sets in the high- and low-NRGPS groups, GSEA was carried out to investigate variations in biological characteristics between the two groups.The enriched pathways in the high-NRGPS group were "Neuroactive ligand-receptor interaction", "Calcium signaling pathway", "Vascular smooth muscle contraction", and "Focal adhesion" (Figure 7C). The enriched pathways in the low-NRGPS group were "DNA replication" and "cell cycle" (Figure 7D). These findings demonstrated that NRGPS can reliably detect tumor progression and that the high-NRGPS group was related to tumor development and metastasis.

10.3389/fimmu.2022.977338

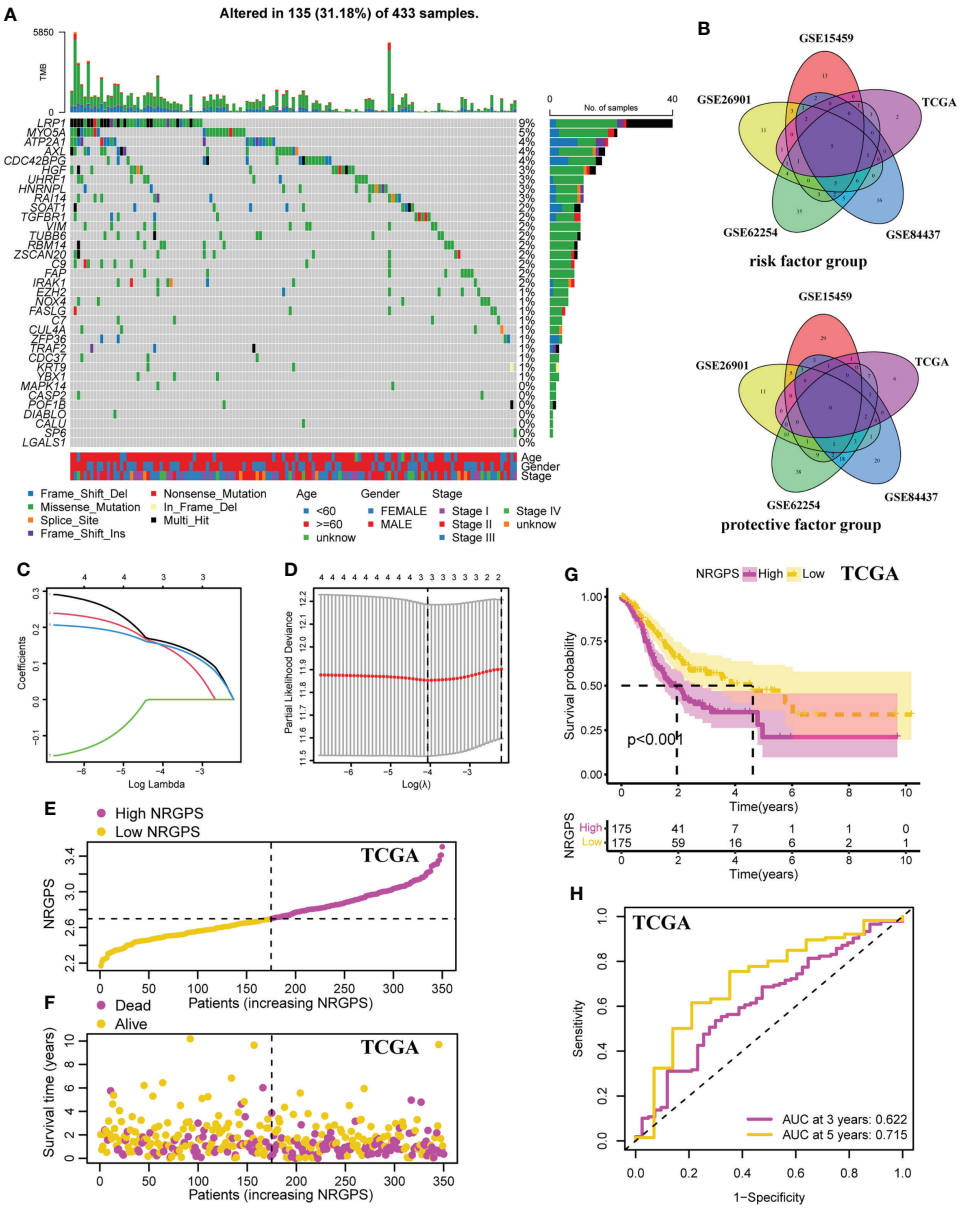


FIGURE 3
Development of the NRGPS System. **(A)** The waterfall plot of the prognostic associated genes mutations. **(B)** The venn plots of intersection of prognostic markers in five cohorts were shown. **(C)** LASSO coefficient profiles of prognostic markers. **(D)** The tuning parameters were cross-validated in the LASSO model. **(E)** Distribution of NRGPS in TCGA cohort. **(F)** Survival status in the high-NRGPS and low-NRGPS groups of TCGA cohort. **(G)** Kaplan-meier survival analysis in TCGA cohort. **(H)** The ROC curve of TCGA cohort verifies the prediction ability of this prediction model. LASSO, Least Absolute Shrinkage and Selection Operator. NRGPS, necroptosis-related gene prognostic score.

Analysis of TME

By analyzing human tumors in TCGA, six kinds of immune infiltration were identified, namely wound healing (C1), INF-γ dominant (C2), inflammatory (C3), lymphocyte depleted (C4), immunologically quiet (C5), and TGF-β dominant (C6) (18), corresponding to tumor promotion and tumor inhibition, respectively. No GC patient belonged to the C5 subtype, and only four GC samples belonged to the C6 subtype; thus, C5 and C6 were not included in this study. According to our analysis of the associations between the GC sample subtypes and NRGPS, C4 was related with low-NRGPS groups, whereas C3 was connected with high-NRGPS groups (Figure 8A).

The activation of 13 immune-related pathways and the infiltrating status of 16 immune cells in the TCGA cohort were investigated using ssGSEA. As shown in Figure 8B, the

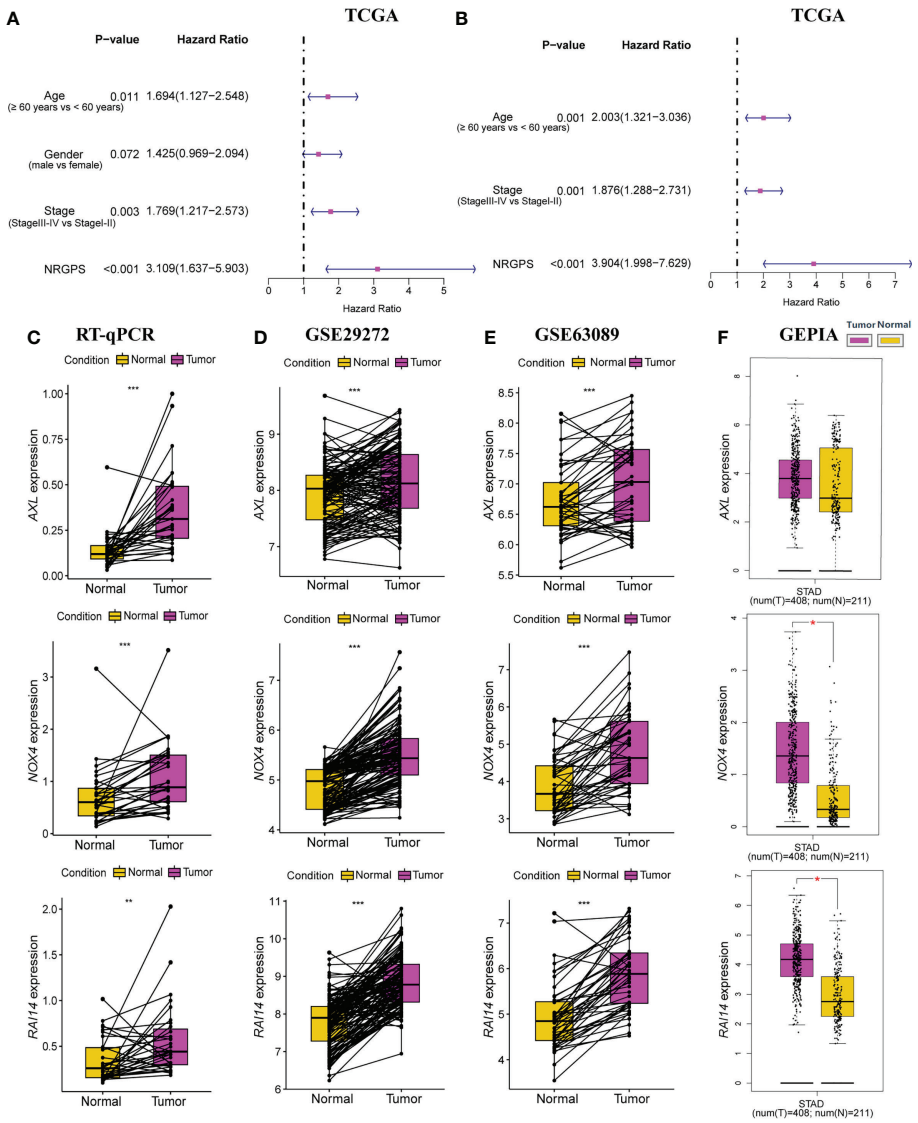


FIGURE 4
Independent prognostic analysis of prognostic model in TCGA cohort was performed. **(A)** The forestplot of univariable Cox regression analysis of NRGPS and clinical characteristics in TCGA cohort. **(B)** The forestplot of multivariable Cox regression analysis of NRGPS and clinical characteristics in TCGA cohort. **(C)** RT-qPCR was used to detect the relative mRNA levels of three NRGs in GC and adjacent normal tissues. **(D)** Relative mRNA levels of three NRGs in GC and adjacent normal tissues in GSE29272 dataset. **(E)** Relative mRNA levels of three NRGs in GC and adjacent normal tissues in GSE63089 dataset. **(F)** Relative mRNA levels of three NRGs in GC and adjacent normal tissues in GEPIA dataset. NRGPS, necroptosis-related gene prognostic score; NRG, necroptosis-related genes; GEPIA, Gene Expression Profiling Interactive Analysis. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

three candidate genes and NRGPS are closely related to immune cells and pathways. Next, we discovered that the level of immune cell infiltration in the high-NRGPS group in TCGA cohort was higher, especially with respect to B-cells, CD8 T cells, dendritic cells, immature dendritic cells, Macrophages cells, mast cells, neutrophils, plasmacytoid dendritic cells, helper T cells, tumor-infiltrating lymphocytes, and regulatory T cells (than in the low-NRGPS group) (Figure 8C). Moreover, the activity of nine immune pathways, i.e., antigen-presenting cell co-inhibition, antigen-presenting cell co-stimulation, C–C chemokine receptor, Check-point, para-inflammation, T cell co-inhibition, T cell co-stimulation, type I interferon response, and type II interferon response in the high-NRGPS group was higher than that in the low-NRGPS group in TCGA cohort (Figure 8D). The relatively high level of immune cell infiltration and pathway activation in the high-NRGPS group was also consistent with the C3 subtype of GC tissues, with the highest number of "inflammatory" features, as shown in our previous results. In

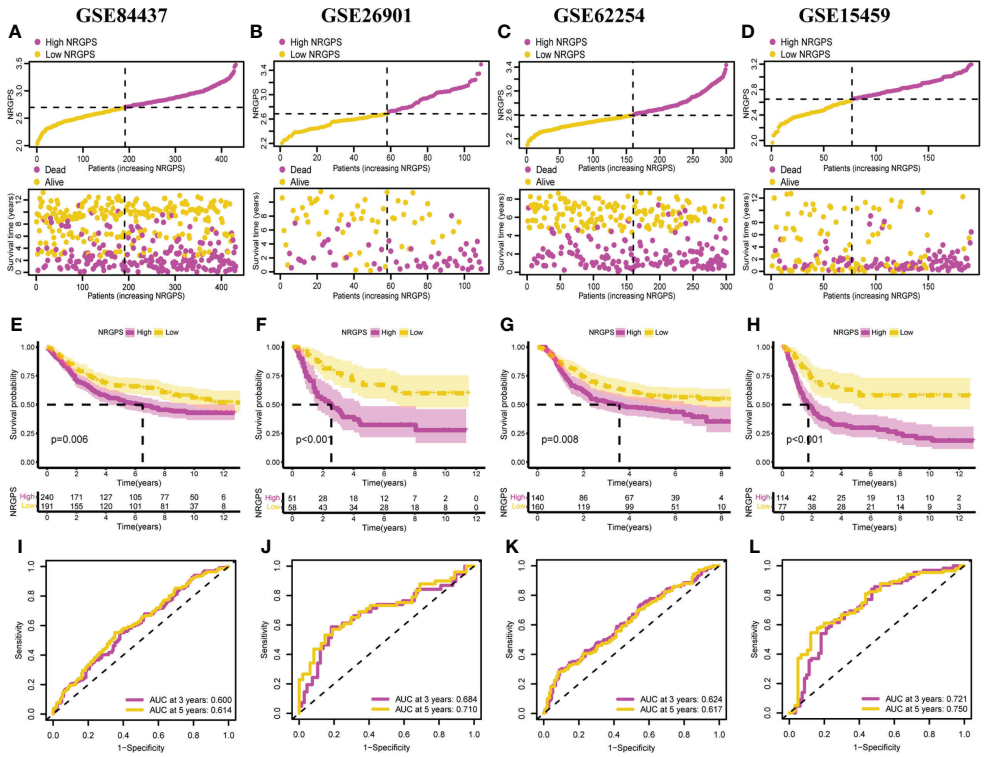


FIGURE 5
Validation of the NRGPS System. Distribution of NRGPS and survival status in the high-NRGPS and low-NRGPS groups in **(A)** GSE84437, **(B)** GSE62901, **(C)** GSE62254, **(D)** GSE15459. Kaplan-Meier survival analysis in **(E)** GSE84437, **(F)** GSE62901, **(G)** GSE62254, **(H)** GSE15459. The ROC analysis in **(I)** GSE84437, **(J)** GSE62901, **(K)** GSE62254, and **(L)** GSE15459 verifies the prediction ability of this prediction model. NRGPS, necroptosis-related gene prognostic score; GC, Gastric cancer; ROC, receiver operating characteristic.

addition, We analyzed the immune cell between the high- and low-NRGPS groups in TCGA cohort using the CIBERSORT-ABS algorithm. The levels of CD8 T cells,resting memory CD4 T cells, activated NK cells, Monocytes, M1 Macrophages, M2 Macrophages, and activated Mast cells were higher in high-NRGPS group than in the low-NRGPS group (Figure 8E). Again, this supports the results of the above analysis. On the one hand, this result suggests that inflammatory cells are associated with the prognosis of GC patients. On the other hand, it provides a basis for the possible regulatory role of NRGPS in the TME, which may affect the prognosis of GC patients.

Since differences in the degree of macrophage infiltration were observed in the high- and low-NRGPS groups, the OS of GC patients with different macrophage infiltration was analyzed. It can be found from the results that the OS of GC patients with high macrophage infiltration is relatively low (Figure 9A). Next, we combined macrophages and NRGPS to analyze the survival of patients. High-macrophages-high-NRGPS patients had the lowest OS (Figure 9B). The results confirmed our speculations. Finally, we confirmed the close relationship between the three candidate genes and macrophages through the algorithms of EPIC, TIMER, xCELL and MCPCOUNTER (Figure 9C).

The effect of immunotherapy in the different NRGPS groups

To investigate the connection between somatic mutation and NRGPS in more detail, analysis was done on the variations in somatic mutation distribution between the TCGA cohort's high-NRGPS and low-NRGPS groups. Somatic mutations were found in 88.89% of the GC patients in the high-NRGPS group and 92.4% of the GC patients in the low-NRGPS group (Figures 10A, B). Missense mutations, frameshift deletions, as well as nonsense mutations were the most frequent mutations in GC patients, a discovery that is in line with the findings of earlier research (19).

Additionally, the TMB of the two groups was calculated and analyzed, showing that the TMB level was significantly higher in the low-NRGPS group than in the high-NRGPS group (Figure 10C). The TMB was negatively correlated with NRGPS (r = -0.3, p< 0.001) (Figure 10D). The OS of the high-TMB group was higher than that of the low-TMB group (Figure 10E). To

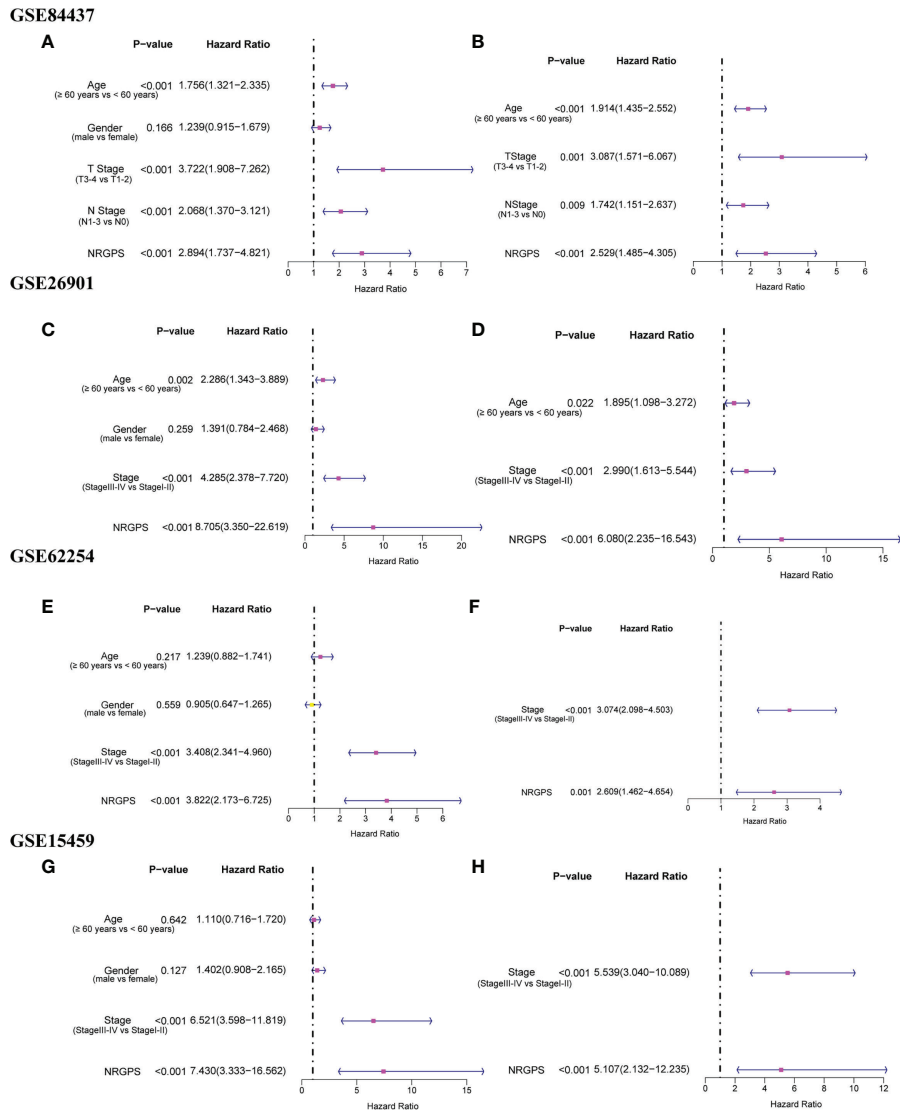


FIGURE 6
Independent prognostic analysis of prognostic model in GEO validation cohorts was performed. The forestplot of univariable Cox regression analysis of NRGPS and clinical characteristics in **(A)** GSE84437, **(C)** GSE62901, **(E)** GSE62254, **(G)** GSE15459. The forestplot of multivariable Cox regression analysis of NRGPS and clinical characteristics in **(B)** GSE84437, **(D)** GSE62901, **(F)** GSE62254, **(H)** GSE15459. NRGPS,necroptosis-related gene prognostic score.

further prove that NRGPS and TMB could be used together for the prognostic assessment of GC, we combined the high- and low-NRGPS groups with the high- and low-TMB groups (Figure 10F), and the results showed that patients with high-TMB-low-NRGPS had the longest survival time, while patients with low-TMB-high-NRGPS had the shortest survival time. This indicated that the NRGPS combined with TMB level has excellent predictive power for the prognosis of GC patients.

Then, we analyzed the MSI of GC patients in the TCGA cohort. We discovered that the high-NRGPS group contained more MSS patients and fewer MSI-H patients (Figure 11A). This also predicted a poor prognosis for patients in the high-NRGPS group. Furthermore, patients with MSS had a higher NRGPS than those with MSI-H, demonstrating once more that NRGPS can accurately predict the prognosis of GC patients (Figure 11B).

In addition, we performed a correlation analysis between NRGPS and immune checkpoints. NRGPS is positively correlated with *PDCD1LG2* and *HAVCR2*, and the expression of *PDCD1LG2* and *HAVCR2* is significantly higher in GC patients in the high-NRGPS group (Figure S4A).

Because the TIDE score and immune escape potential are positively associated, tumor patients with higher TIDE scores are

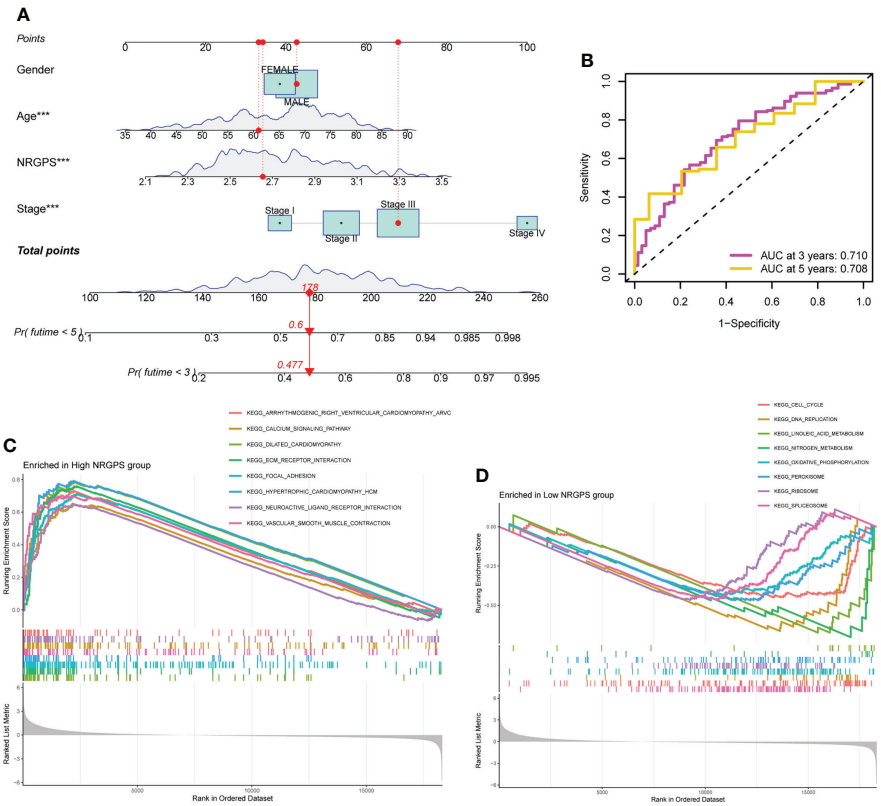


FIGURE 7
Nomograph Model and GSEA Analysis. **(A)** Nomogram of NRGPS and clinical factors predicting survival probability of GC patients. **(B)** The ROC curve verifies the predictive ability of the nomogram. **(C)** Enrichment pathways in high-NRGPS group. **(D)** Enrichment pathways in low-NRGPS group. GSEA, Gene Set Enrichment Analysis; NRGPS, necroptosis-related gene prognostic score; GC, Gastric cancer; ROC, receiver operating characteristic. ***$p < 0.001$.

less likely to benefit from immunotherapy. Our results showed that compared with that in the high-NRGPS group, the low-NRGPS group had a lower TIDE score, and NRGPS was positively correlated with TIDE ($r = 0.4$, $p< 0.001$), implying that the low-NRGPS group GC patients might benefit more from the immunotherapy than the high-NRGPS group GC patients (Figure 11C). Additionally, we discovered that the T cell dysfunction and T cell exclusion scores in the high-NRGPS group were higher than those in the low-NRGPS group, and that NRGPS was also positively connected with these scores (Figures 11D, E). This indicates that NRGPS has a high degree of confidence in determining the effectiveness of immunotherapy.

Finally, the GSE78220 cohort verified the predictive ability of NRGPS for immunotherapy. According to the median NRGPS of the GSE78220, patients were divided into high-NRGPS group and low-NRGPS group. With the increase in NRGPS, the survival rate of patients also decreased gradually (Figure 11F). There was a significant difference in survival time between high- and low-NRGPS groups, and NRGPS also had good prediction ability (Figures 11G, H). Then, the NRGPS of the non-response group was significantly higher than that of the response group (Figure 11I). In addition, there are also differences in the proportion of patients with or without response to immunotherapy in the high- and low-NRGPS group, and the proportion of patients with non-response in the high-NRGPS group is larger (Figure 11J).

Discussion

Necroptosis is closely related to tumorigenesis and immunity and is a target for tumor therapy. In the tumor microenvironment, necroptosis is mostly considered as pro-inflammatory cell death (20), but its underlying inflammation also has the potential to promote tumor development and metastasis through genomic instability, cell proliferation and angiogenesis proliferation and angiogenesis to promote tumor development and metastasis (21). Necroptosis plays different roles in different cancers, with necroptosis-associated protein RIPK3 being expressed at low levels in leukemia, colorectal

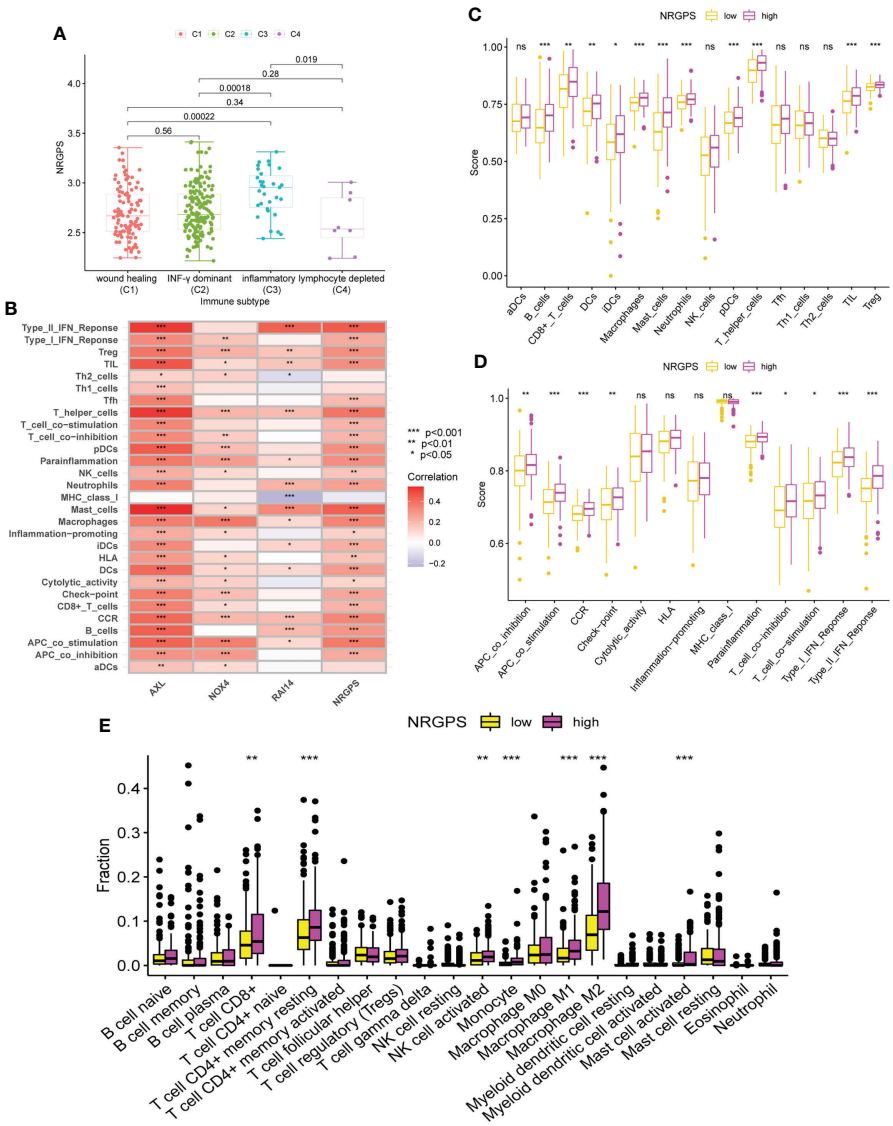


FIGURE 8
Analysis of immune subtypes and immune cell infiltration in the TCGA cohort. **(A)** The boxplot of differences in NRGPS between immune subtype groups. **(B)** Correlation analysis of NRGPS and 3 candidate genes with immune cells and signaling pathways (ssGSEA algorithm). **(C)** The boxplot of 16 immune cell differences in the low- and high-NRGPS groups (ssGSEA algorithm). **(D)** The boxplot of 13 immune signaling pathway differences in the low- and high-NRGPS groups (ssGSEA algorithm). **(E)** The boxplot of 22 immune cell infiltration differences between high-NRGPS and low-NRGPS groups in TCGA cohort (CIBERSORT-ABS algorithm). NRGPS, necroptosis-related gene prognostic score; ssGSEA, single sample gene set enrichment analysis. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns, not significant.

cancer, breast cancer and melanoma relative to adjacent normal tissues, and at increased levels in lung cancer and pancreatic cancer (8). Recent research showed that necroptosis was associated with a critical role in GC development (22). However, it remains unclear whether necroptosis can predict prognosis and immunotherapy efficacy for GC patients. Typically, either TNM stage systems or serum markers (CEA, CA19-9, and CA125) are employed to monitor progression and to predict prognosis in GC patients. Nevertheless, these methods are not unsatisfactory, having low accuracy and high non-specificity, especially for GC patients with highly heterogeneous (23). In this study, a new NRGPS system for GC was established and four independent GEO external validation were performed. The ability of NRGPS to predict the effect of immunotherapy was also confirmed in the GSE78220 cohort. The results demonstrated that the NRGPS system could accurately forecast the prognosis and immunotherapy sensitivity of GC patients.

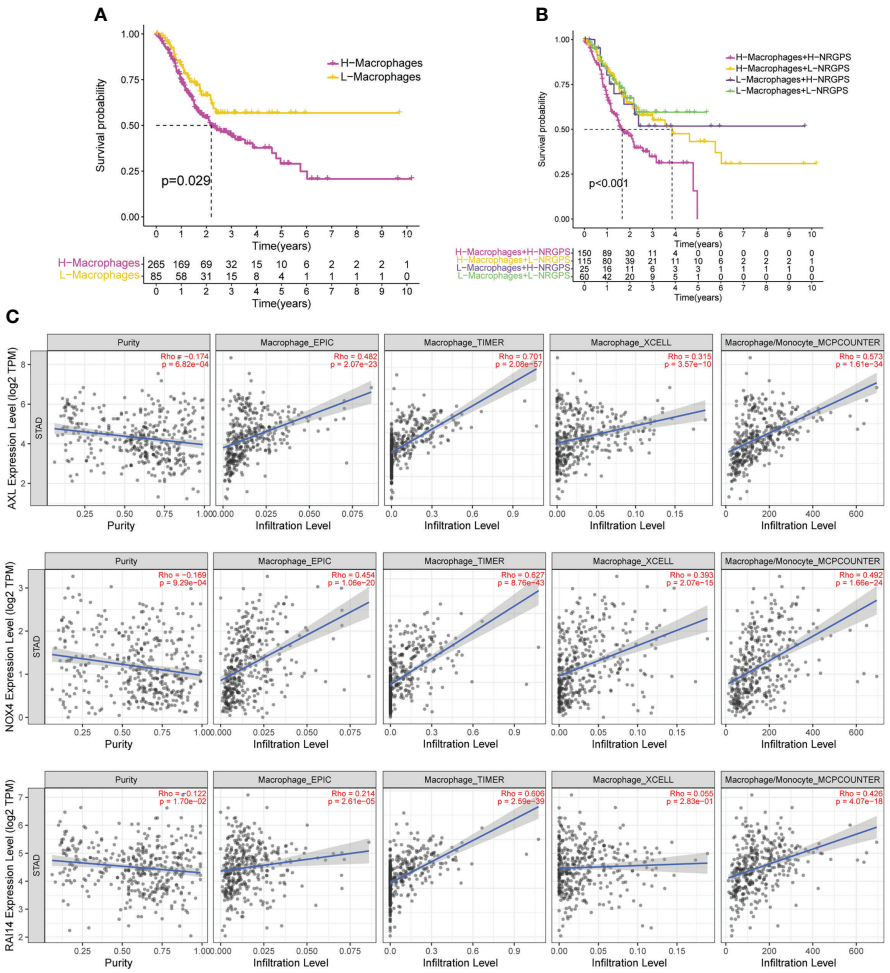


FIGURE 9
NRGPS and macrophages in TME were analyzed. **(A)** Difference in survival time between high- and low-macrophage groups (ssGSEA algorithm). **(B)** Survival analysis of macrophage combined NRGPS (ssGSEA algorithm). **(C)** EPIC, TIMER, XCELL and MCPCOUNTER algorithms were used to analyze the correlation between the three candidate genes and macrophages. NRGPS, necroptosis-related gene prognostic score; TME, tumor microenvironment.

In order to understand how NRGs promote GC progression, we first obtained 35 NRGs significantly associated with GC prognosis by univariable Cox regression analysis in the TCGA cohort. Then,it was revealed by GO and KEGG pathway analysis that these NRGs were mostly implicated in several pathways linked to tumor growth.In the risk factor group, there were 5 cross-genes in TCGA and 4 independent GEO datasets. The NRGPS system for gastric cancer patients consisting of three NRGs was further constructed using lassoCox regression analysis, and this system had an excellent performance in discriminating the high-NRGPS group with a poorer prognosis. The results demonstrated that the high-NRGPS group was significantly associated with shorter survival time in TCGA and four GEO cohorts. Moreover, NRGPS could independently predict the OS for GC patients. We used the GSEA analysis to investigate the gene sets enriched in the two NRGPS groups in order to more thoroughly examine the function of these three marker genes in GC, and the results showed that the gene sets of the high-NRGPS group were mainly enriched in the “Neuroactive ligand-receptor interaction”, “Calcium signaling pathway”, “Vascular smooth muscle contraction”, and “Focal adhesion”, which are associated with tumor development (24–27).

The NRGPS was made up of three NRGs, such as *AXL*, *RAI14*, and *NOX4*. Upregulated of these genes in the GC tumor tissues was link to a poor prognosis. The tumor microenvironment’s immunosuppression and the survival, proliferation, migration, invasion, and metastasis of tumor cells are all significantly influenced by the oncogenic receptor tyrosine kinase *AXL* (28–30). *AXL* increases the expression of *ZEB1* in GC cells, promoting EMT, invasion, and proliferation (31). *AXL* inhibitors control the

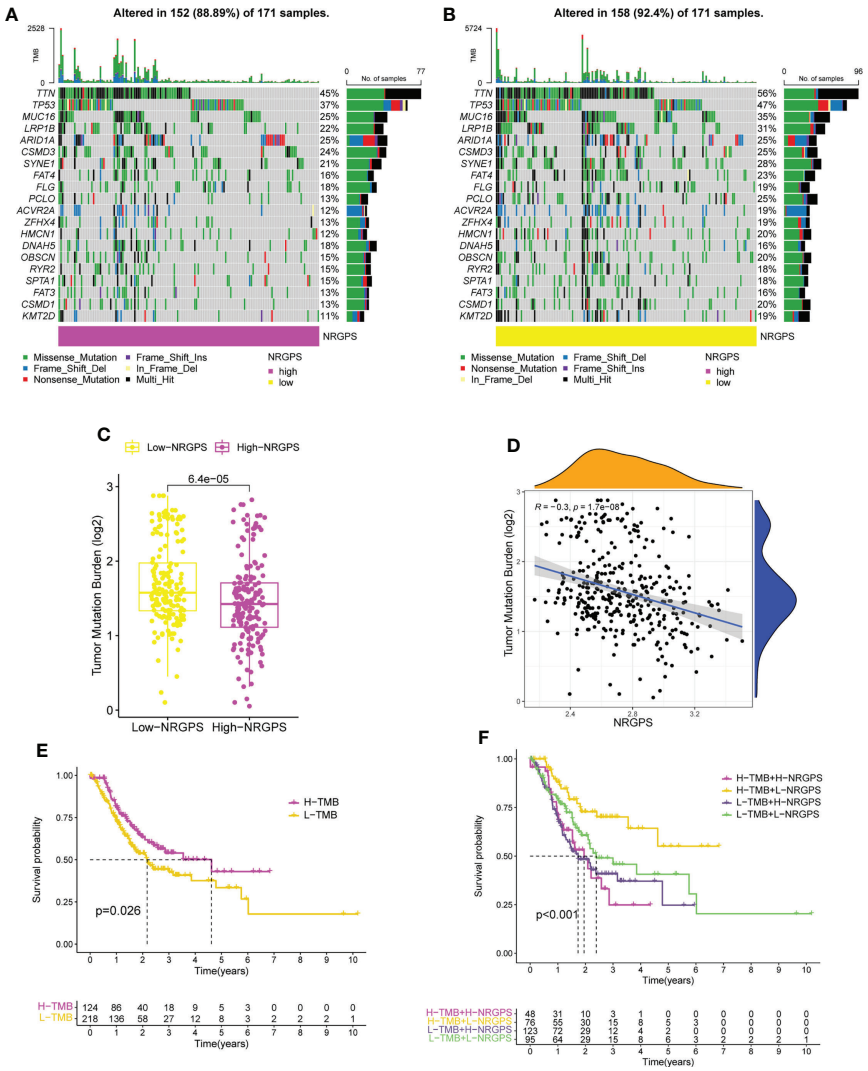


FIGURE 10
TMB analysis and survival analysis of patients in TCGA cohort. **(A)** The waterfall plot of somatic mutation in high-NRGPS group. **(B)** The waterfall plot of somatic mutation in low-NRGPS group. **(C)** The boxplot of TMB differences between low-NRGPS and high-NRGPS groups. **(D)** Correlation analysis between TMB and NRGPS. **(E)** Difference in survival time between high- and low-TMB groups. **(F)** Survival analysis of TMB combined NRGPS. TMB, Tumor mutation burden; NRGPS, necroptosis-related gene prognostic score.

polarization of macrophages to boost tumor immunity (32). Targeting of *AXL* receptors is particularly well suited to enhance the efficacy of immune checkpoint inhibitors (33). Retinoic acid induced 14 (*RAI14*) was originally identified in all-trans retinoic acid-induced human retinal pigment epithelial cells (34), *RAI14* may be connected to the growth and invasion of cancer cells in several malignancies, according to current investigations (35). High expression of *RAI14* may enhance the translocation of esophageal tumor cells (36). *RAI14* is significantly expressed in GC, and its knockdown slows the development of GC (37).

The nicotinamide adenine dinucleotide phosphate (NADPH) oxidase 4 (*NOX4*) is one of the most important NADPH isoforms in endothelial cells, and it has been reported that the receptor-interacting protein 1 is involved in tumor necrosis factor-α-induced reactive oxygen species generation and necroptosis through interaction with NADPH oxidase (38). Through the recruitment of M2 TAM through the generation of different cytokines in response to ROS/PI3K signaling, tumor *NOX4* stimulates the proliferation of non-small cell lung cancer cells (39). In GC patients, the high expression of *NOX4* results in a bad prognosis (40).

We next further demonstrated that NRGPS is associated with multiple immune cell recruitment, so we further performed TME analysis.

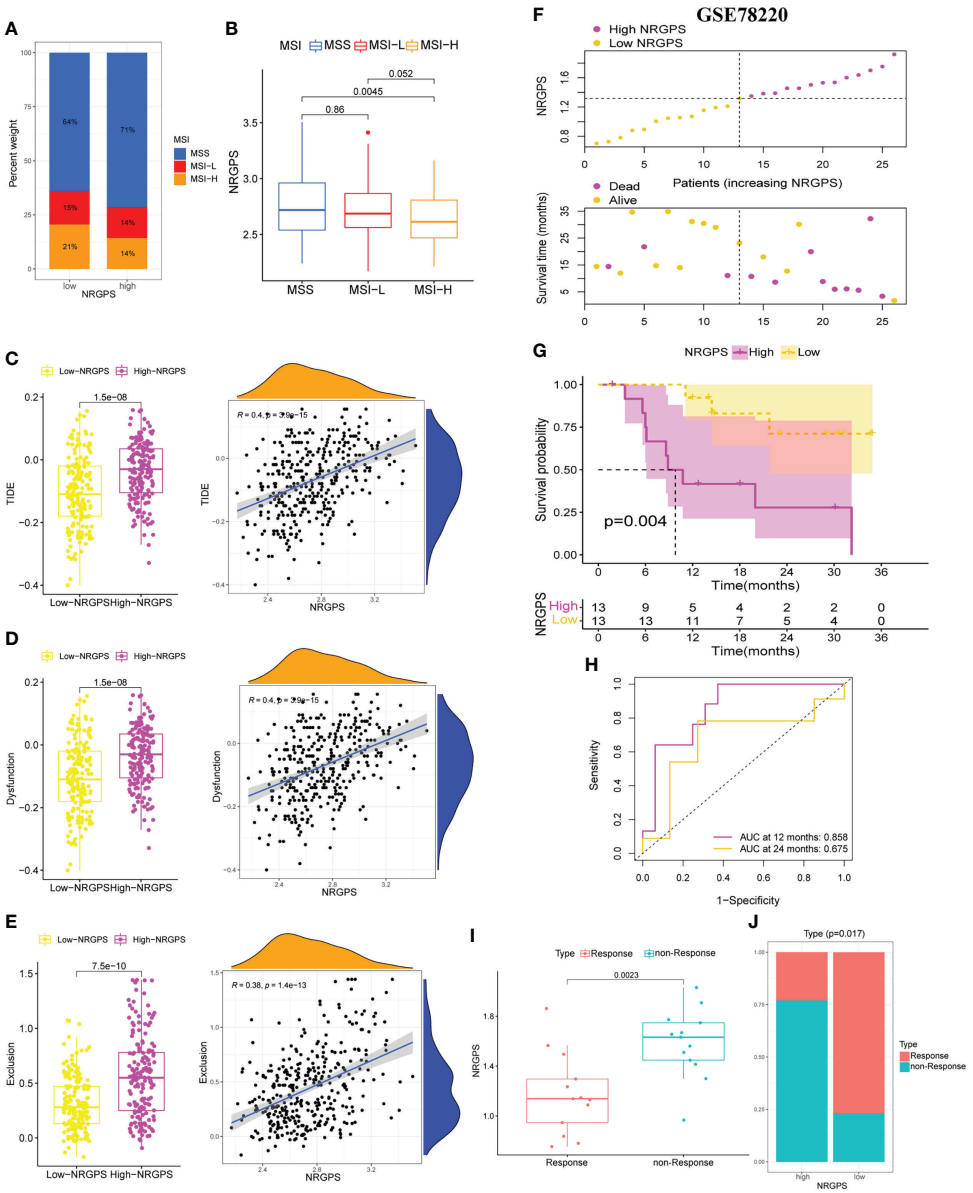


FIGURE 11
Differences in MSI statuse, TIDE score between high-NRGPS and low-NRGPS groups in the TCGA cohort. **(A)** Distribution of patients with different MSI statuses in high- and low-NRGPS groups. **(B)** The boxplot of NRGPS differences between groups with different MSI. **(C)** The boxplots of differences between TIDE in low- and high-NRGPS groups and correlation analysis of TIDE with NRGPS. **(D)** The boxplots of differences between Immune Dysfunction scores in low- and high-NRGPS groups and correlation analysis of Immune Dysfunction scores with NRGPS. **(E)** The boxplots of differences between Immune Exclusion scores in low- and high-NRGPS groups and correlation analysis of Immune Exclusion scores with NRGPS. **(F)** The distribution of NRGPS and the survival status of the high- and low-NRGPS groups. **(G)** Kaplan-meier survival analysis in GSE78220 cohort. **(H)** The ROC analysis of GSE78220 cohort. **(I)** The boxplot of NRGPS differences between immunotherapy responde and non-responde groups. **(J)** Distribution of patients who responded and did not respond to immunotherapy in the high- and low-NRGPS groups. MSI, Microsatellite Instability; TIDE, Tumor Immune Dysfunction, and Exclusion; NRGPS, necroptosis-related gene prognostic score, ROC, receiver operating characteristic.

The imbalance in the ratio of immune cell components is highly correlated with poor prognosis in cancer patients (41). Moreover, the immune cells in the TME can be used in the prognostic assessment of a variety of tumors (42). Recent studies have identified that necroptosis can regulate the TME (43). On this basis, we evaluated the GC TME based on the NRGPS. In this study, the degree of immune cell infiltration varied significantly between the groups with high and low NRGPS.In

patients with high-NRGPS, the majority of immune cells were heavily invaded.

One of the most crucial elements of the tumor immunosuppressive microenvironment are tumor-associated macrophages (TAM),which is influenced by the surrounding TME, and macrophages show a continuous activation state (44, 45). In our results, TAM infiltration was relatively high in the high-NRGPS group and was abundant in immune cells recruited by NRGPS, *AXL*, *RAI14*, and *NOX4*, and in the combined TAM and NRGPS analysis results, we found that GC patients in the high-TAM and high-NRGPS groups had the shortest survival times.M1-like macrophages and M2-like macrophages are the two main classifications of activated macrophages. The inflammatory response is intimately correlated with both M1 and M2 macrophages, with M1 macrophages primarily generating pro-inflammatory cytokines to take part in the pro-inflammatory response., stimulating the Th1 response of T cells (IFN-γ) and further enhancing the M1 macrophage response, and M2 macrophages mainly participating in the anti-inflammatory response (46, 47).

It is well known that M1 macrophages promote the attack on tumor cells, while M2 macrophages have been consistently associated with cancer metastasis and poor prognosis (48), but it has also been suggested that CD68+ HLA-DR+ M1-type macrophages in the tumor microenvironment can promote tumor migration through the NF-KB signaling pathway, which in turn promotes tumor progression (49). Moreover, recent studies have shown that M1-type macrophages polarized by exosomes promote the malignant migration of oral squamous cell carcinoma (50). It is clear that the complexity of the tumor microenvironment is not only in the differences between individuals with different diseases, but also in the different parts of the same tumor tissues of patients, which may have different effects depending on the microenvironment in which the tumor is located (51–53).

Thus, It is obvious that the biology of M1-type macrophages in relation to cancer is complex and fascinating.Additionally, we discovered that patients with GC in the high-NRGPS group with pro-tumor immune cell infiltration had a much shorter survival time than those in the low-NRGPS group. This suggests that in the high-NRGPS group, these immunosuppressive cells were manipulated to protect them from the body's immune response, which is also known as "sabotage" (54).

Ultimately, the tuRecent studies have demonstrated the significant association of TMB and TME in GC patients regarding tumorigenesis, tumor progression, and drug resistance (55). Nevertheless, the exact mechanisms by which necroptosis plays a role during tumor immunotherapy remain largely unknowable. We further analyzed the TMB and TIDE in the different subgroups to assess the immunotherapy response.

The TMB is another index to evaluate patient reaction to immunotherapy independent of the programmed cell death-ligand 1 expression level (56, 57). A comprehensive analysis of 27 cancer types showed a positive correlation of benefit between TMB and immunotherapy (58). According to a recent Panel Sequencing study, patients with elevated TMB had longer OS., and tests for treatment response to immunotherapy were performed, suggesting that TMB could be used as a predictive biomarker for advanced GC patients treated with immunotherapy (59). In the field of cancer biomarkers, although thousands of expression signatures have been nominated to be used as biomarkers, few have found reliable clinical use, therefore the expression signature of the marker must be consistent with reproducible genetic marker alterations (60).

Consistently, in our study, the TMB was lower in the high-NRGPS group than in the low-NRGPS group. The NRGPS was negatively correlated with the TMB, and the combination of NRGPS with the TMB interestingly revealed that GC patients in the high-TMB and low-NRGPS groups had the longest survival time, and GC patients in the low-TMB and high-NRGPS groups had the shortest survival time, demonstrating the reliability of NRGPS as a prognostic evaluation index for the immunotherapy of GC patients. In China, about 95% of GC patients have the characteristics of microsatellite stability (MSS) (61). Patients with GC in the MSI-H group had a higher survival rate and significantly more benefit from immunotherapy in contrast to the MSS/MSI-L group (62, 63). Consistently, according to our research, the high-NRGPS group of GC patients had a larger percentage of MSS and a lower percentage of MSI-H.

Programmed cell death-ligand 1 level and TMB are less accurate than the TIDE at predicting survival outcomes in cancer patients receiving immunotherapy medications (64, 65). Recent research has demonstrated its usefulness in predicting or evaluating the impact of immunotherapy (66–68), and it has also been affirmed in GC studies (69). Because anti-tumor immune escape is more likely in patients with higher TIDE scores, immunotherapy response rates are lower in these patients (64). Consistent with the TMB result, the GC patients with more immune dysfunction in the high-NRGPS group were more likely to resort to immunotherapy and receive immunotherapy less effectively than those in the low-NRGPS group.mor escapes immune surveillance and flees the immune system.

Our findings show that the high-NRGPS group was characterized by a pro-cancer immune microenvironment, high TIDE score, and low TMB, low MSI, which demonstrated that the high-NRGPS group was correlated with immune escape in GC, and therefore NRGPS could be used as a new biomarker to accurately predict the prognosis and immunotherapy efficacy in GC patients. Some limitations must be highlighted. First, A large number of clinical specimens and relevant information should be used to further verify the prognostic characteristics of NRGs. Second, the mechanisms by which necroptosis shapes the TME features in GC are unclear and would be experimentally investigated in the future.

Conclusion

In conclusion, our study defined a new prognostic signal composed of three necroptosis-related genes and validated it using RT-qPCR methodsas, well as independent external validation using the GSE84437, GSE26901, GSE62254 and GSE15459 datasets. This signal was verified in GC prognosis and immunotherapy effects and was proven to be independently related with OS in the TCGA cohort and the GEO validation cohorts.It is a prognostic classifier that can be applied to clinical decision-making for therapy and personalised prognosis.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Second Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YX, RZZ performed the experiments. YX, RZZ, and JQL wrote and revised the manuscript. XMJ, JMD, and MZW helped or performed experiments and analyses. JQR, TG and KTH downloaded and collated the data. YHS and SYZ was responsible for supervising the study. All authors read and gave final approval ofthe manuscript.

Funding

The present study was supported by the National Natural Science Foundation of China (No. 81772253); the Natural Science Foundation of Heilongjiang Province (No.LH2021H082); Scientific research project of Heilongjiang Provincial Health Commission (No. 2017-085).

Acknowledgments

Figure 1 was produced with the assistance of Servier Medical Art (https://smart.servier.com).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.977338/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

2. Zhao Q, Cao L, Guan L, Bie L, Wang S, Xie B, et al. Immunotherapy for gastric cancer: dilemmas and prospect. *Brief Funct Genomics* (2019) 18(2):107–12. doi: 10.1093/bfgp/ely019

3. Zhang J, Zhang Y, Qu B, Yang H, Hu S, Dong X. If small molecules immunotherapy comes, can the prime be far behind? *Eur J Med Chem* (2021) 218:113356. doi: 10.1016/j.ejmech.2021.113356

4. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet* (2020) 396(10251):635–48. doi: 10.1016/S0140-6736(20)31288-5

5. Sprooten J, De Wijngaert P, Vanmeerbeerk I, Martin S, Vangheluwe P, Schlenner S, et al. Necroptosis in immuno-oncology and cancer immunotherapy. *Cells* (2020) 9(8):1823. doi: 10.3390/cells9081823

6. Díaz Del Arco C, Ortega Medina L, Estrada Muñoz L, García Gómez de Las Heras S, Fernández Aceñero MJ. Is there still a place for conventional histopathology in the age of molecular medicine? laurén classification, inflammatory infiltration and other current topics in gastric cancer diagnosis and prognosis. *Histol Histopathol* (2021) 36(6):587–613. doi: 10.14670/HH-18-309

7. Li X, Dong G, Xiong H, Diao H. A narrative review of the role of necroptosis in liver disease: a double-edged sword. *Ann Transl Med* (2021) 9(5):422. doi: 10.21037/atm-20-5162

8. Gong Y, Fan Z, Luo G, Yang C, Huang Q, Fan K, et al. The role of necroptosis in cancer biology and therapy. *Mol Cancer* (2019) 18(1):100. doi: 10.1186/s12943-019-1029-8

9. Tang R, Xu J, Zhang B, Liu J, Liang C, Hua J, et al. Ferroptosis, necroptosis, and pyroptosis in anticancer immunity. *J Hematol Oncol* (2020) 13(1):110. doi: 10.1186/s13045-020-00946-7

10. Park HH, Kim HR, Park SY, Hwang SM, Hong SM, Park S, et al. RIPK3 activation induces TRIM28 derepression in cancer cells and enhances the anti-tumor microenvironment. *Mol Cancer* (2021) 20(1):107. doi: 10.1186/s12943-021-01399-3

11. Sun W, Yu W, Shen L, Huang T. MLKL is a potential prognostic marker in gastric cancer. *Oncol Lett* (2019) 18(4):3830–6. doi: 10.3892/ol.2019.10687

12. Kim S, Lee H, Lim JW, Kim H. Astaxanthin induces NADPH oxidase activation and receptor–interacting protein kinase 1–mediated necroptosis in gastric cancer AGS cells. *Mol Med Rep* (2021) 24(6):837. doi: 10.3892/mmr.2021.12477

13. Choi ME, Price DR, Ryter SW, Choi AMK. Necroptosis: a crucial pathogenic mediator of human disease. *JCI Insight* (2019) 4(15):e128834. doi: 10.1172/jci.insight.128834

14. Zhao Z, Liu H, Zhou X, Fang D, Ou X, Ye J, et al. Necroptosis-related lncRNAs: Predicting prognosis and the distinction between the cold and hot tumors in gastric cancer. *J Oncol* (2021) 2021:6718443. doi: 10.1155/2021/6718443

15. Wang N, Liu D. Identification and validation a necroptosis–related prognostic signature and associated regulatory axis in stomach adenocarcinoma. *Onco Targets Ther* (2021) 14:5373–83. doi: 10.2147/OTT.S342613

16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

17. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2 (3):100141. doi: 10.1016/j.xinn.2021.100141

18. Tamborero D, Rubio-Perez C, Muiños F, Sabarinathan R, Piulats JM, Muntasell A, et al. A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin Cancer Res* (2018) 24(15):3717–28. doi: 10.1158/1078-0432.CCR-17-3509

19. Usui G, Matsusaka K, Mano Y, Urabe M, Funata S, Fukayama M, et al. DNA Methylation and genetic aberrations in gastric cancer. *Digestion* (2021) 102(1):25–32. doi: 10.1159/000511243

20. Negroni A, Colantoni E, Cucchiara S, Stronati L. Necroptosis in intestinal inflammation and cancer: New concepts and therapeutic perspectives. *Biomolecules* (2020) 10(10):1431. doi: 10.3390/biom10101431

21. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* (2010) 140(6):883–99. doi: 10.1016/j.cell.2010.01.025

22. Guo D, Zhang W, Yang H, Bi J, Xie Y, Cheng B, et al. Celastrol induces necroptosis and ameliorates inflammation *via* targeting biglycan in human gastric carcinoma. *Int J Mol Sci* (2019) 20(22):5716. doi: 10.3390/ijms20225716

23. Shao Y, Jia H, Li S, Huang L, Aikemu B, Yang G, et al. Comprehensive analysis of ferroptosis-related markers for the clinical and biological value in gastric cancer. *Oxid Med Cell Longev* (2021) 2021:7007933. doi: 10.1155/2021/7007933

24. Sousa B, Pereira J, Paredes J. The crosstalk between cell adhesion and cancer metabolism. *Int J Mol Sci* (2019) 20(8):1933. doi: 10.3390/ijms20081933

25. Sulzmaier FJ, Jean C, Schlaepfer DD. FAK in cancer: Mechanistic findings and clinical applications. *Nat Rev Cancer* (2014) 14(9):598–610. doi: 10.1038/nrc3792

26. Monteith GR, Prevarskaya N, Roberts-Thomson SJ. The calcium-cancer signalling nexus. *Nat Rev Cancer* (2017) 17(6):367–80. doi: 10.1038/nrc.2017.18

27. Li J, Rong MH, Dang YW, He RQ, Lin P, Yang H, et al. Differentially expressed gene profile and relevant pathways of the traditional Chinese medicine cinobufotalin on MCF 7 breast cancer cells. *Mol Med Rep* (2019) 19(5):4256–70. doi: 10.3892/mmr.2019.10062

28. Ludwig KF, Du W, Sorrelle NB, Wnuk-Lipinska K, Topalovski M, Toombs JE, et al. Small-molecule inhibition of axl targets tumor immune suppression and enhances chemotherapy in pancreatic cancer. *Cancer Res* (2018) 78(1):246–55. doi: 10.1158/0008-5472.CAN-17-1973

29. Skinner HD, Giri U, Yang LP, Kumar M, Liu Y, Story MD, et al. Integrative analysis identifies a novel AXL-PI3 kinase-PD-L1 signaling axis associated with radiation resistance in head and neck cancer. *Clin Cancer Res* (2017) 23(11):2713–22. doi: 10.1158/1078-0432.CCR-16-2586

30. Zhang G, Wu LW, Mender I, Barzily-Rokni M, Hammond MR, Ope O, et al. Induction of telomere dysfunction prolongs disease control of therapy-resistant melanoma. *Clin Cancer Res* (2018) 24(19):4771–84. doi: 10.1158/1078-0432.CCR-17-2773

31. He L, Lei Y, Hou J, Wu J, Lv G. Implications of the receptor tyrosine kinase axl in gastric cancer progression. *Onco Targets Ther* (2020) 13:5901–11. doi: 10.2147/OTT.S257606

32. Goyette MA, Elkholi IE, Apcher C, Kuasne H, Rothlin CV, Muller WJ, et al. Targeting Axl favors an antitumorigenic microenvironment that enhances immunotherapy responses by decreasing Hif-1α levels. *Proc Natl Acad Sci USA* (2021) 118(29):e2023868118. doi: 10.1073/pnas.2023868118 Goyette MA, Elkholi IE, Apcher C, Kuasne H, Rothlin CV, Muller WJ. Targeting Axl favors an antitumorigenic microenvironment that enhances immunotherapy responses by decreasing Hif-1α levels. *Proc Natl Acad Sci USA* (2021) 118(29):e2023868118. doi: 10.1073/pnas.2023868118

33. Engelsen AST, Lotsberg ML, Abou Khouzam R, Thiery JP, Lorens B, Chouaib S, et al. Dissecting the role of AXL in cancer immune escape and resistance to immune checkpoint inhibition. *Front Immunol* (2022) 13:869676. doi: 10.3389/fimmu.2022.869676

34. Kutty RK, Kutty G, Samuel W, Duncan T, Bridges CC, El-Sherbeeny A, et al. Molecular characterization and developmental expression of NORPEG, a novel gene induced by retinoic acid. *J Biol Chem* (2001) 276(4):2831–40. doi: 10.1074/jbc.M007421200

35. Yuan W, Zheng Y, Huo R, Lu L, Huang XY, Yin LL, et al. Expression of a novel alternative transcript of the novel retinal pigment epithelial cell gene NORPEG in human testes. *Asian J Androl* (2005) 7(3):277–88. doi: 10.1111/j.1745-7262.2005.00040.x

36. Wang J, Cai Y, Luo J, Sun Z, Yu J, Yan F, et al. RAI14 silencing suppresses progression of esophageal cancer *via* the STAT3 pathway. *Aging (Albany NY)* (2020) 12(18):18084–98. doi: 10.18632/aging.103613

37. Chen C, Maimaiti A, Zhang X, Qu H, Sun Q, He Q, et al. Knockdown of RAI14 suppresses the progression of gastric cancer. *Onco Targets Ther* (2018) 11:6693–703. doi: 10.2147/OTT.S175502

38. Ago T, Kitazono T, Ooboshi H, Iyama T, Han YH, Takada J, et al. Nox4 as the major catalytic component of an endothelial NAD(P)H oxidase. *Circulation* (2004) 109(2):227–33. doi: 10.1161/01.CIR.0000105680.92873.70

39. Zhang J, Li H, Wu Q, Chen Y, Deng Y, Yang Z, et al. Tumoral NOX4 recruits M2 tumor-associated macrophages *via* ROS/PI3K signaling-dependent various cytokine production to promote NSCLC growth. *Redox Biol* (2019) 22:101116. doi: 10.1016/j.redox.2019.101116

40. You X, Ma M, Hou G, Hu Y, Shi X. Gene expression and prognosis of NOX family members in gastric cancer. *Onco Targets Ther* (2018) 11:3065–74. doi: 10.2147/OTT.S161287

41. Li Y, He X, Fan L, Zhang X, Xu Y, Xu X. Identification of a novel immune prognostic model in gastric cancer. *Clin Transl Oncol* (2021) 23(4):846–55. doi: 10.1007/s12094-020-02478-5

42. Giraldo NA, Sanchez-Salas R, Peske JD, Vano Y, Becht E, Petitprez F, et al. The clinical role of the TME in solid cancer. *Br J Cancer* (2019) 120(1):45–53. doi: 10.1038/s41416-018-0327-z

43. Meng MB, Wang HH, Cui YL, Wu ZQ, Shi YY, Zaorsky NG, et al. Necroptosis in tumorigenesis, activation of anti-tumor immunity, and cancer therapy. *Oncotarget* (2016) 7(35):57391–413. doi: 10.18632/oncotarget.10548

44. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol* (2008) 8(12):958–69. doi: 10.1038/nri2448

45. Coussens LM, Zitvogel L, Palucka AK. Neutralizing tumor-promoting chronic inflammation: A magic bullet? *Science* (2013) 339(6117):286–91. doi: 10.1126/science.1232227

46. Kim J, Bae JS. Tumor-associated macrophages and neutrophils in tumor microenvironment. *Mediat Inflamm* (2016) 2016:6058147. doi: 10.1155/2016/6058147

47. Qiu SQ, Waaijer SJH, Zwager MC, de Vries EGE, van der Vegt B, Schröder CP. Tumor-associated macrophages in breast cancer: Innocent bystander or important player? *Cancer Treat Rev* (2018) 70:178–89. doi: 10.1016/j.ctrv.2018.08.010

48. Gabrilovich DI, Ostrand-Rosenberg S, Bronte V. Coordinated regulation of myeloid cells by tumours. *Nat Rev Immunol* (2012) 12(4):253–68. doi: 10.1038/nri3175

49. Wang H, Wang X, Li X, Fan Y, Li G, Guo C, et al. CD68(+)HLA-DR(+) M1-like macrophages promote motility of HCC cells *via* NF-κB/FAK pathway. *Cancer Lett* (2014) 345(1):91–9. doi: 10.1016/j.canlet.2013.11.013

50. Xiao M, Zhang J, Chen W, Chen W. M1-like tumor-associated macrophages activated by exosome-transferred THBS1 promote malignant migration in oral squamous cell carcinoma. *J Exp Clin Cancer Res* (2018) 37(1):143. doi: 10.1186/s13046-018-0815-2

51. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* (2011) 469(7330):356–61. doi: 10.1038/nature09650

52. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* (2013) 501 (7467):338–45. doi: 10.1038/nature12625

53. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* (2013) 152(4):714–26. doi: 10.1016/j.cell.2013.01.019

54. Poschke I, Mougiakakos D, Kiessling R. Camouflage and sabotage: tumor escape from the immune system. *Cancer Immunol Immunother* (2011) 60(8):1161–71. doi: 10.1007/s00262-011-1012-8

55. Kumagai S, Togashi Y, Sakai C, Kawazoe A, Kawazu M, Ueno T, et al. An oncogenic alteration creates a microenvironment that promotes tumor progression by conferring a metabolic advantage to regulatory T cells. *Immunity* (2020) 53 (1):187–203.e8. doi: 10.1016/j.immuni.2020.06.016

56. Hodges TR, Ott M, Xiu J, Gatalica Z, Swensen J, Zhou S, et al. Mutational burden, immune checkpoint expression, and mismatch repair in glioma: implications for immune checkpoint immunotherapy. *Neuro Oncol* (2017) 19 (8):1047–57. doi: 10.1093/neuonc/nox026

57. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-Small-Cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol* (2018) 36(7):633–41. doi: 10.1200/JCO.2017.75.3384

58. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med* (2017) 377(25):2500–1. doi: 10.1056/NEJMc1713444

59. Kim J, Kim B, Kang SY, Heo YJ, Park SH, Kim ST, et al. Tumor mutational burden determined by panel sequencing predicts survival after immunotherapy in patients with advanced gastric cancer. *Front Oncol* (2020) 10:314. doi: 10.3389/fonc.2020.00314

60. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495

61. Li Z, Jia Y, Zhu H, Xing X, Pang F, Shan F, et al. Tumor mutation burden is correlated with response and prognosis in MSS gastric cancer patients undergoing neoadjuvant chemotherapy. *Gastric Cancer* (2021) 24(6):1342–54. doi: 10.1007/s10120-021-01207-3

62. Zhang T, Yu S, Zhao S. Development and verification of a microsatellite instability-related risk signature for predicting survival and therapy effectiveness in gastric cancer. *J Gastrointest Oncol* (2022) 13(1):84–101. doi: 10.21037/jgo-21-808

63. Cao L, Li T, Ba Y, Chen E, Yang J, Zhang H. Exploring immune-related prognostic signatures in the tumor microenvironment of colon cancer. *Front Genet* (2022) 13:801484. doi: 10.3389/fgene.2022.801484

64. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

65. Keenan TE, Burke KP, Van Allen EM. Genomic correlates of response to immune checkpoint blockade. *Nat Med* (2019) 25(3):389–402. doi: 10.1038/s41591-019-0382-x

66. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat Med* (2019) 25(12):1916–27. doi: 10.1038/s41591-019-0654-5

67. Bretz AC, Parnitzke U, Kronthaler K, Dreker T, Bartz R, Hermann F, et al. Domatinostat favors the immunotherapy response by modulating the tumor immune microenvironment (TIME). *J Immunother Cancer* (2019) 7(1):294. doi: 10.1186/s40425-019-0745-3

68. George AP, Kuzel TM, Zhang Y, Zhang B. The discovery of biomarkers in cancer immunotherapy. *Comput Struct Biotechnol J* (2019) 17:484–97. doi: 10.1016/j.csbj.2019.03.015

69. Ren Q, Zhu P, Zhang H, Ye T, Liu D, Gong Z, et al. Identification and validation of stromal-tumor microenvironment-based subtypes tightly associated with PD-1/PD-L1 immunotherapy and outcomes in patients with gastric cancer. *Cancer Cell Int* (2020) 20:92. doi: 10.1186/s12935-020-01173-3

COPYRIGHT

© 2022 Xia, Zhang, Wang, Li, Dong, He, Guo, Ju, Ru, Zhang and Sun. 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 Immunology

TYPE Original Research
PUBLISHED 30 August 2022
DOI 10.3389/fimmu.2022.998236

Check for updates

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Xu Qi-Tong,
Nanjing Medical University, China
Zewei Tu,
The Second Affiliated Hospital of Nanchang University, China

*CORRESPONDENCE
Peipei Gong
ntnsgpp@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 19 July 2022
ACCEPTED 03 August 2022
PUBLISHED 30 August 2022

CITATION
Wang W, Lu Z, Wang M, Liu Z, Wu B, Yang C, Huan H and Gong P (2022) The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma.
Front. Immunol. 13:998236.
doi: 10.3389/fimmu.2022.998236

COPYRIGHT
© 2022 Wang, Lu, Wang, Liu, Wu, Yang, Huan and Gong. 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.

The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma

Weichen Wang†, Zhichao Lu†, Maoyu Wang†, Zongheng Liu, Bing Wu, Chengkai Yang, He Huan and Peipei Gong*

Department of Neurosurgery, Affiliated Hospital of Nantong University, Medical School of Nantong University, Nantong, China

Background: Copper ions are essential for cellular physiology. Cuproptosis is a novel method of copper-dependent cell death, and the cuproptosis-based signature for glioma remains less studied.

Methods: Several glioma datasets with clinicopathological information were collected from TCGA, GEO and CGGA. Robust Multichip Average (RMA) algorithm was used for background correction and normalization, cuproptosis-related genes (CRGs) were then collected. The TCGA-glioma cohort was clustered using ConsensusClusterPlus. Univariate Cox regression analysis and the Random Survival Forest model were performed on the differentially expressed genes to identify prognostic genes. The cuproptosis-signature was constructed by calculating CuproptosisScore using Multivariate Cox regression analysis. Differences in terms of genomic mutation, tumor microenvironment, and enrichment pathways were evaluated between high- or low-CuproptosisScore. Furthermore, drug response prediction was carried out utilizing pRRophetic.

Results: Two subclusters based on CRGs were identified. Patients in cluster2 had better clinical outcomes. The cuproptosis-signature was constructed based on CuproptosisScore. Patients with higher CuproptosisScore had higher WHO grades and worse prognosis, while patients with lower grades were more likely to develop IDH mutations or MGMT methylation. Univariate and Multivariate Cox regression analysis demonstrated CuproptosisScore was an independent prognostic factor. The accuracy of the signature in prognostic prediction was further confirmed in 11 external validation datasets. In groups with high-CuproptosisScore, PIK3CA, MUC16, NF1, TTN, TP53, PTEN, and EGFR showed high mutation frequency. IDH1, TP53, ATRX, CIC, and FUBP1 demonstrated high mutation frequency in low-CuproptosisScore group. The level of immune infiltration increased as CuproptosisScore increased. SubMap analysis revealed patients with high-CuproptosisScore may respond to anti-PD-1 therapy. The IC50 values of Bexarotene, Bicalutamide, Bortezomib, and

Cytarabine were lower in the high-CuproptosisScore group than those in the low-CuproptosisScore group. Finally, the importance of IGFBP2 in TCGA-glioma cohort was confirmed.

Conclusion: The current study revealed the novel cuproptosis-based signature might help predict the prognosis, biological features, and appropriate treatment for patients with glioma.

KEYWORDS

glioma, cuproptosis, signature, clusters, bioinformatics

Introduction

Glioma is the most common primary brain tumor, accounting for about 40% of all brain tumors (1), among which glioblastoma (GBM) is the most malignant brain tumor. According to the classification of World Health Organization (WHO), gliomas are classified into four different histopathological grades: Grade I, II, III and IV, of which WHO II and III are considered low-grade gliomas (LGG). Glioblastoma (GBM, WHO IV) characterized by new angiogenesis is the most aggressive molecular subtype of glioma (2, 3). The median survival of LGG can be achieved from five to ten years through the administration of surgery, radiotherapy, and chemotherapy combination treatments, whereas the median survival of GBM is normally around one or two years (4, 5). The prognosis of glioma patients is divergent, which may be related to different tumor grades, mutation of isocitrate dehydrogenase (IDH) (6), amplification of epidermal growth factor receptor (EGFR) (7) and other factors. The current glioma prognostic evaluation model is mainly based on clinical factors, which has limited predictive ability (8–11). Therefore, a better prognostic evaluation system is needed.

Gene-regulated cell death known as "programmed cell death" is crucial for tissue homeostasis and growth, it also takes part in several pathological processes (12). At present, various types of cell death, such as necroptosis, ferroptosis, and pyroptosis, have been found to belong to necrotic programmed cell death (12). Researchers have found that cell death is closely related to tumorigenesis and prognosis. In the process of tumor development, cell death often occurs in the intratumoral area due to metabolic stress, such as hypoxia or glucose deprivation (13). Consequently, triggering programmed cell death could be a potential strategy for novel tumor therapy. Currently, knowledge of programmed cell death in cancer is continuously updated as more types of programmed cell death are discovered and recognized. Biology has long recognized copper as a vital component in all living things, from bacteria and fungi to mammals and plants, where it is a must for survival (14, 15). In humans, it binds to enzymes that assist in blood clotting, hormone maturation, and cell processing of energy, however, excessive copper can cause cell death (14). Cuproptosis as a new type of cell death is modulated and regulated by copper in cells. Copper ion directly binds to the lipoacylated components in the tricarboxylic acid cycle, leading to abnormal aggregation of lipoacylated proteins and loss of iron-sulfur cluster proteins, which leads to protein toxic stress and ultimately leads to cell death (15). Cuproptosis in glioma, however, has not yet been studied in depth.

The establishment of a glioma prognosis prediction model based on transcriptome data combined with clinical data can improve the prediction ability to a certain extent, which has highly significant clinical significance. In this study, glioma gene expression data and clinical data were collected from open databases, combined with reported cuproptosis-related genes (CRGs), was used to establish and verify cuproptosis-related clusters and signature. Subsequently, the prognosis, immune status, and treatment response of patients was also explored based on the cuproptosis-related clusters and signature.

Materials and methods

Collection and preprocessing of data for glioma

Transcripts and clinical data of glioma samples, including survival status, IDH status, grade, gender, and age, were collected from TCGA database based on UCSC Xena platform (16, 17). A total of 656 glioma patients with corresponding data were enrolled. Meanwhile, the gene expression profile of the control (non-tumoral samples) were also obtained from Genome Tissue Expression (GTEx) project (https://www.gtexportal.org) (18, 19). In addition, 11 glioma-cohorts (CGGA311, CGGA668, GSE108474, GSE13041, GSE16011, GSE43289, GSE43378, GSE4412, GSE4412, GSE68838, and GSE83300) were collected from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) or Chinese Glioma

Genome Atlas (CGGA, http://www.cgga.org.cn/). Robust Multichip Average (RMA) algorithm was used for background correction and normalization (20). Data in the form of fragments per kilobase million (FPKM) was transformed into transcripts per kilobase million (TPM). The list of cuproptosis-related genes (CRGs) refers to the previous literature (14). In the end, ten CRGs were included in our study: FDX1, LIAS, LIPT1, DLD, DLAT, PDHA1, POHB, MTF1, GLS, and CDKN2A.

Establishment of cuproptosis-clusters and cuproptosis-signature

Based on the collected ten CRGs, the TCGA-glioma cohort was clustered using ConsensusClusterPlus package (21). Next, principal component analysis (PCA) (22) was further carried out to assess patterns associated with cuproptosis. Limma package was used to identify the differentially expressed genes (DEGs) in cuproptosis-clusters (logFC>1, P<0.05) (23). Subsequently, Univariate Cox regression analysis was performed to identify prognostic DEGs preliminarily (24). Subsequently, more valuable prognostic genes were screened out based on Random Survival Forest model (variable importance>0.25) (25, 26). To construct a cuproptosis-signature, Multivariate Cox regression analysis was used to estimate and weight the regression coefficients of the prognostic genes, and the CuproptosisScore for each glioma sample was calculated. According to the best optimal cutoff, the patients were divided into high- or low-CuproptosisScore subgroups. The association between overall survival (OS) and CuproptosisScore was analyzed using Kaplan-Meier curves. ROC curves were further utilized to validate the efficiency and accuracy of CuproptosisScore in predicting outcomes at one-, two-, and three-year. In addition, Univariate or Multivariate Cox regression analyses of CuproptosisScore and several clinical factors were performed to verify the independence of CuproptosisScore in predicting prognosis.

Genomic mutation analysis for cuproptosis-signature

The data of somatic mutations (16, 27) or copy number variation (CNV) (16) was acquired from TCGA. Genomic Identification of Significant Targets in Cancer (GISTIC) (28) algorithm was used to assess genomic characterization and CNV landscape.

Analysis of immune infiltration

Immune cell abundance (immune score), stromal cell infiltrating level (stromal score), and tumor purity (ESTIMATE score) were estimated *via* ESTIMATE (The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression) algorithm (29). Using Tumor Immune Estimation Resource 2.0 (TIMER 2.0, http://timer.cistrome.org/) (30), a comprehensive analysis of immune infiltration in the tumor microenvironment of glioma was carried out. MCPcounter algorithm was used to estimate the relative proportions of ten immune cells in glioma (31). The infiltration of 28 immune cells was indicated by enrichment scores, which were calculated by single sample gene set enrichment analysis (ssGSEA) using Gene Set Variation Analysis (GSVA) R package (32, 33). Immunomodulators associated with seven different immune processes (Antigen presentation, Cell adhesion, Co-inhibitor, Co-stimulator, Ligand, Receptor and Other) were obtained from the previous literature (34). The response of glioma to anti-PD1 and anti-CTLA4 therapy was evaluated by Submap algorithm (35–37).

Enrichment pathway analysis

All gene sets from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were downloaded from the MSigDB database (38). Gene Set Enrichment Analysis (GSEA) (39) and Gene Set Variation Analysis (GSVA) (33) were carried out according to clusterProfiler and GSVA packages, respectively.

Drug response prediction

Pharmacogenomic data from Genomics of Drug Sensitivity in Cancer (GDSC) (40) database was used to predict drug sensitivity in the enrolled glioma cases. The half maximal inhibitory concentration (IC50) value was calculated by pRRophetic package to reflect the drug response (41).

Immunohistochemistry (IHC) staining

The tissue sections through deparaffinization and dehydration were incubated with polyclonal rabbit anti-human IGFBP2 antibodies (1:50, Proteintech, 11065-3-AP) overnight at 4°C after epitope retrieval, H2O2 treatment, and non-specific antigens blocking. Next, sections were incubated with secondary antibodies (1:1000, Proteintech, SA00001-2) for two hours at room temperature, and then the signal was detected by an enhanced DAB staining kit (Proteintech, China).

Western blot

Tumor tissues as well as normal tissues were lysed in RIPA buffer (Solarbio, Beijing, China), protease and phosphatase

inhibitors were added, and then denatured at 100°C for 15 min. The protein samples were then separated by 10% SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membranes. Next, PVDF membranes were blocked with 5% skim milk powder solution for 1 hour, and incubated with primary antibodies, including anti-IGFBP2 antibody (1:1000, Proteintech, 11065-3-AP), anti-GAPDH antibody (1:5000, Abcam, ab9485) overnight, followed by secondary antibodies (1:2000, Proteintech, SA00001-2) for 2 hours at room temperature, observed with the ECL kit chemiluminescence reagent (Billerica Millipore, MA, USA). Protein band signals were detected by the Chemidoc detection system (Bio-Rad, Hercules, CA, USA) and quantified by ImageJ software (National Institutes of Health, USA).

Statistical analysis

The Wilcoxon test was used to compare non-normally distributed data. The T-test was used to compare normally distributed variables between two groups. The R package survminer was used to estimate OS between two groups using Kaplan-Meier survival plots. Cox regression of survival analysis was also performed by survival. Time-dependent receiver operating characteristics (ROC) curves were plotted using R package timeROC. All heatmaps were performed through pheatmap package. The data were visualized by ggplot2 (V4.1.2). P<0.05 was considered statistically significant.

Results

Characteristics of cuproptosis-clusters for TCGA-glioma

The clinical information of patients from TCGA was listed in the Supplemental Table S1. The correlations among the ten CRGs were mostly positive, and the most strongly associated variables were DLD and DLAT (Figure 1A). Based on ConsensusClusterPlus, the optimal number of clusters was determined, k=2 (Figure 1B). Furthermore, PCA analysis was further used to validate that patients in the two subclusters clustered separately, which confirmed the reliability of the clustering results (Figure 1C). Patients with glioma in cluster2 had significantly better clinical outcomes than those in cluster1 (Figure 1D). Compared with the other subcluster, the expressions of FDX1, DLD, DLAT, PDHB, GLS were relatively high in cluster1, while the expression of CDKN2A was relatively high in cluster2, which indicates that these CRGs may be genetic markers for identifying different clusters (Figure 1E). Interestingly, we discovered that cluster2 had a greater percentage of IDH mutation status, which may be one of the factors contributing to a better prognosis of this subcluster (Figure 1E).

Establishment of cuproptosis-signature

In the two subclusters, a total of 27 differentially expressed genes (DEGs) were identified (logFC>1, P<0.05), and the volcano map accurately reflected the gene expression differences between the two subclusters (Figure 2A). After Univariate Cox regression analysis, 16 potential pro-oncogenes (HR>1; CYTOR, EMP3, OCIAD2, PLA2G5, FABP5P7, IGFBP2, TSTD1, TIMP1, RBP1, METTL7B, POSTN, CHI3L1, H19, CXCL14, LTF and ENC1) and 9 potential suppressor genes were identified (HR<1; CAMK2A, LINC01088, CDKN2B, LINC00689, TPTEP1, C5orf38, KLRC2, VIPR2, and SMOC1) (Figure 2B). Figure 2C showed the distribution of error rates in Random Survival Forest model, after which the relative importance of seven genes (H19, CYTOR, IGFBP2, EMP3, KLRC2, C5orf38, and CHI3L1) was established (variable importance>0.25, Figure 2D). Multivariate Cox regression analysis was used to develop the cuproptosis-signature, and the CuproptosisScore for each glioma sample was calculated according to the following formula: $0.0621*Expr_{H19}+0.0196*Expr_{CYTOR}+0.2739*Expr_{IGFBP2}+0.0183*Expr_{KLRC2}+0.0036*Expr_{C5orf38}+0.1406*Expr_{CHI3L1}$. Heatmap displayed the distribution of six genes in cuproptosis-signature, CuproptosisScore and the clinical characteristics (Figure 2E). It was clear that a higher CuproptosisScore was linked to higher expressions of H19, CHI3L1, CYTOR, and IGFBP2 and, in contrast, was associated with lower expressions of KLRC2 and C5orf38 (Figure 2E). In the meantime, IDH mutation status was more likely to be present in glioma patients with lower Cuprop tosisScore. (Figure 2E).

Prognostic potential of cuproptosis-signature

Next, we analyzed the CuproptosisScore of TCGA patients among WHO grades, mutation status and MGMT methylation status (Figure 3A). Patients with higher CuproptosisScore had higher WHO grades, while patients with lower grades were more likely to develop IDH mutations or MGMT methylation (Figure 3A), all of which may explain the significantly better clinical outcomes of patients with lower CuproptosisScore (P<0.001, Figure 3B). In addition, Univariate and Multivariate Cox regression analysis of CuproptosisScore and clinicopathologic features demonstrated that both CuproptosisScore and Grade were independent prognostic factors for patients with glioma (Figure 3C). The survival ROC curves predicted by the cuproptosis-signature showed that the AUCs were all greater than 0.8, indicating the effectiveness of the cuproptosis-signature in predicting prognosis for glioma at the 1-year (AUC=0.898), 2-year (AUC=0.922), 3-year (AUC=0.918), 4-year (AUC=0.867), and 5-year (AUC=0.828) time points (Figure 3D). Furthermore, we conducted Univariate Cox regression analysis on the OS

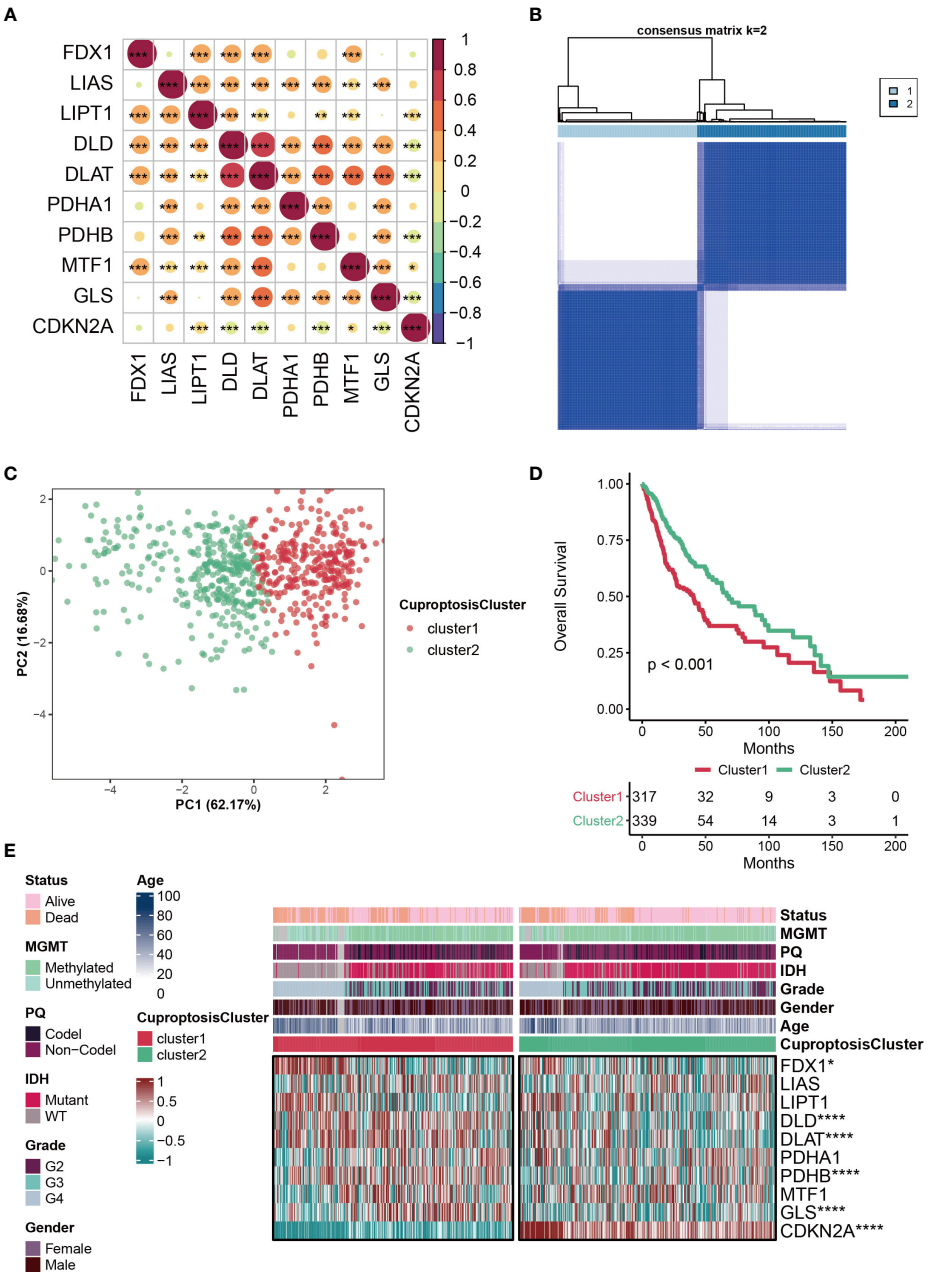


FIGURE 1
Characteristics of CuproptosisCluster in TCGA-glioma cohort. **(A)** The correlations among the ten cuproptosis-related genes (CRGs). The color represents the correlation coefficient. **(B)** Cluster diagram for subtype analysis of glioma samples. The intragroup correlations were the highest and the inter-group correlations were low when k=2. **(C)** PCA analysis for the two subclusters. **(D)** Kaplan-Meier survival curve showing survival probability of cluster1 and cluster2. **(E)** Heatmap showing the expression levels of the ten cuproptosis-related genes (CRGs) in different clinical features and clusters. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

(overall survival) of glioma patients based on the external validation data sets, and the results showed that HR was greater than 1 in all of the 11 data sets, which further validated the accuracy of our constructed cuproptosis-signature in prognostic prediction (Figure 3E).

Genomic mutation analysis for cuproptosis-signature

GISTIC algorithm was used to assess the genomic characterization landscape between high- CuproptosisScore

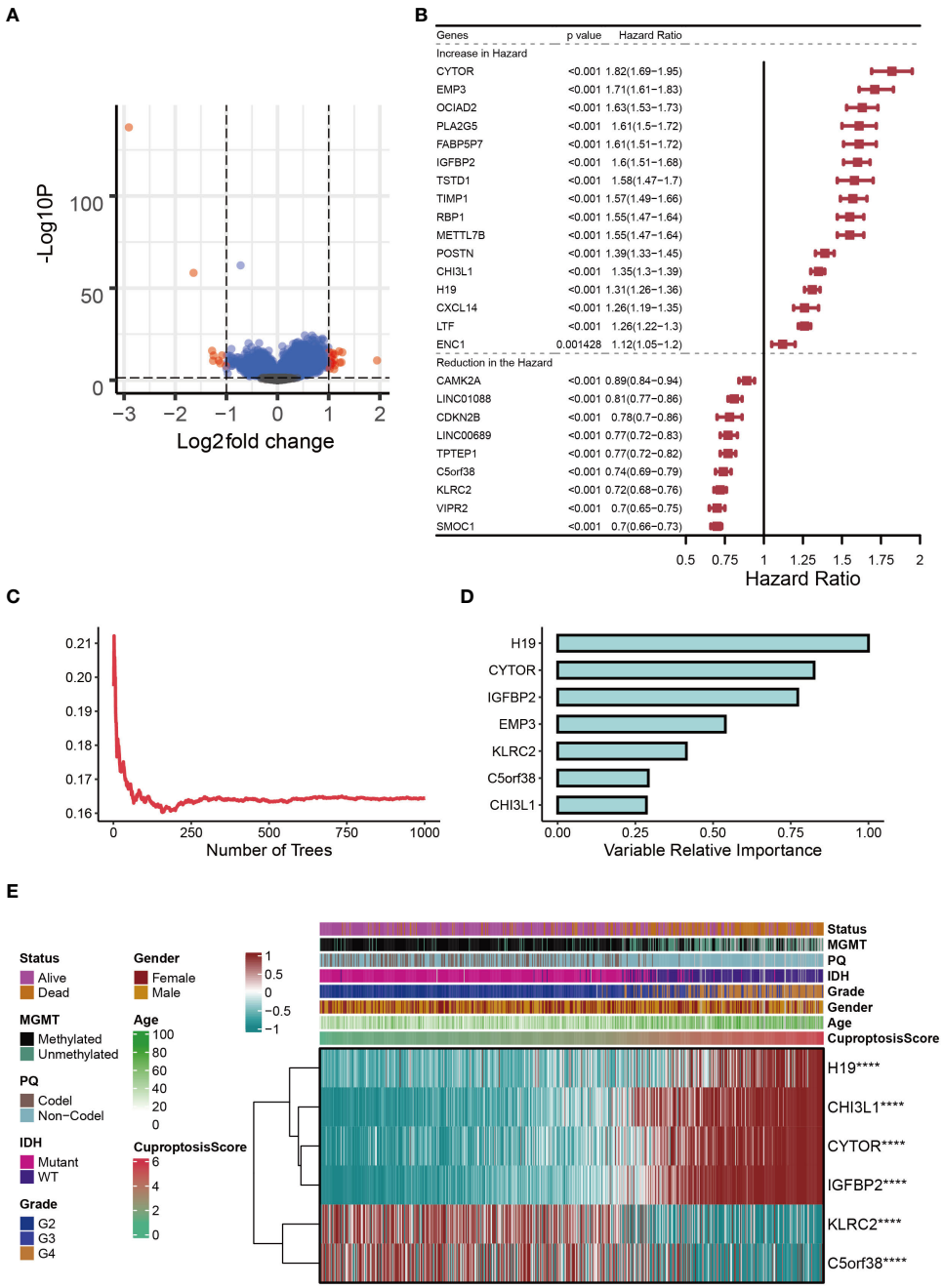


FIGURE 2
Establishment of cuproptosis-signature. **(A)** The volcano map reflects the differentially expressed genes identified (logFC > 1, P < 0.05). **(B)** The forest figure for Univariate Cox regression analysis of the differentially expressed genes. **(C)** The distribution of error rates in Random Survival Forest model. **(D)** The variable relative importance of the seven genes. **(E)** Heat map showing the relationship between six genes in the cuproptosis-signature and CuproptosisScore distribution and its clinical characteristics. ****P < 0.0001.

and low-CuproptosisScore subgroups, which was shown in Figure 4A. In patients with high-CuproptosisScore, PIK3CA, MUC16, NF1, TTN, TP53, PTEN, and EGFR had high mutation frequency (over 10%, Figure 4B), while in those with low-CuproptosisScore, IDH1, TP53, ATRX, CIC, and FUBP1 had high mutation frequency (over 10%, Figure 4C). TP53 had high mutation rates in both groups (26% and 51%, respectively). In agreement with the findings above, the mutation rate of IDH1 was particularly high in the low-CuproptosisScore group, reaching as high as 89% (Figure 4C).

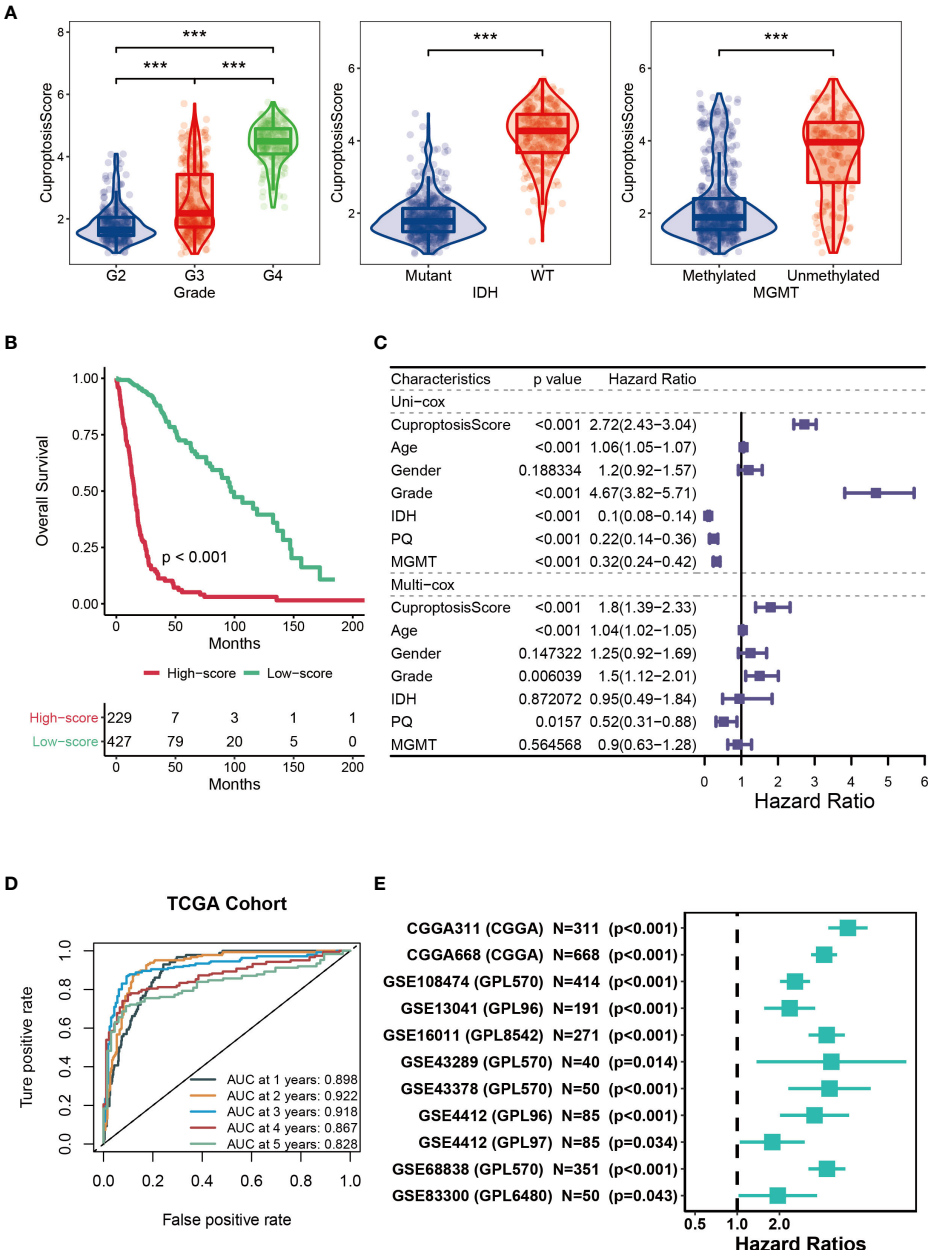


FIGURE 3
Prognostic potential of cuproptosis-signature. **(A)** The violin figures for comparing the CuproptosisScore of TCGA patients among WHO grades, mutation status and MGMT methylation status. **(B)** Kaplan-Meier survival curve showing survival probability of high-CuproptosisScore or low-CuproptosisScore subgroups. **(C)** The forest figure for Univariate or Multivariate Cox regression analysis of CuproptosisScore and clinicopathologic features. **(D)** The 1-year, 2-year, 3-year, 4-year, and 5-year survival ROC curves are predicted by the cuproptosis-signature. **(E)** Univariate Cox regression analysis of the cuproptosis-signature in 11 external validation data sets. ***P < 0.001.

Immune status for cuproptosis-signature

Based on ESTIMATE, MCPcounter, ssGSEA, and TIMER algorithms mentioned in the Methods section, the heatmap showed the abundance of infiltrating immune cell populations at different CuproptosisScores (Figure 5A). In general, the level of immune infiltration increased as the CuproptosisScore increased (Figure 5A). However, it was observed that patients with lower CuproptosisScores had more tumor purity (Figure 5A). In addition, our results showed that glioma

10.3389/fimmu.2022.998236

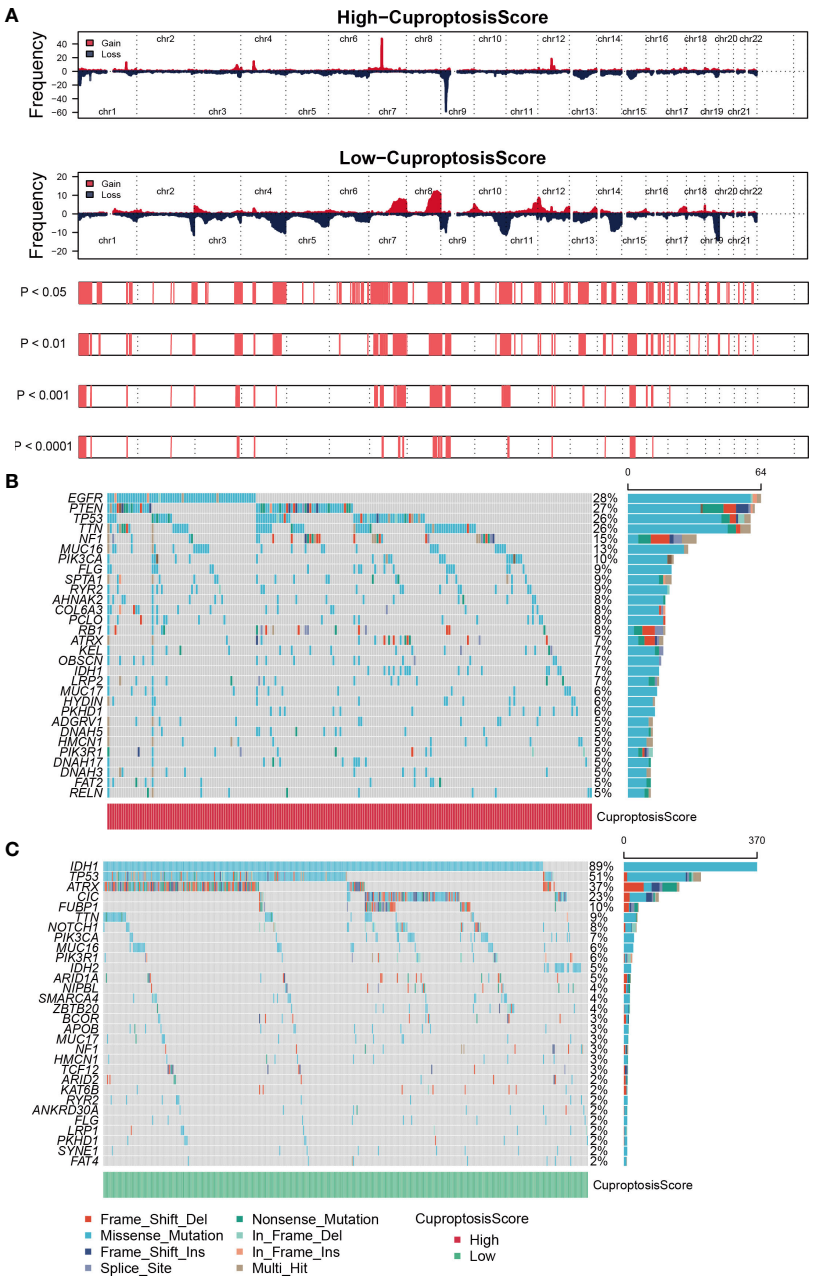


FIGURE 4
Genomic mutation analysis for cuproptosis-signature. **(A)** Genomic characterization landscape of high-CuproptosisScore or low-CuproptosisScore subgroups. **(B)** Gene mutation frequency in high-CuproptosisScore. **(C)** Gene mutation frequency in low-CuproptosisScore.

patients with high CuproptosisScores also had higher levels of TMB (Figure 5B), GEP (Figure 5C), and CYT (Figure 5D). GSVA analysis also suggested that patients with high CuproptosisScores were enriched in immune-related pathways, such as negative regulation of macrophage apoptotic process, macrophage fusion, B cell receptor signaling pathway, T cell receptor signaling pathway, and primary immunodeficiency (Figure 5E).

Immunotherapy and chemotherapy of cuproptosis-signature

Immunomodulators (IMs) are closely related to the immunotherapy of malignant tumors, as well, agonists and antagonists for immunomodulators are also being studied (42). The expression of IM-related genes varied across high-CuproptosisScore or low-CuproptosisScore subgroups

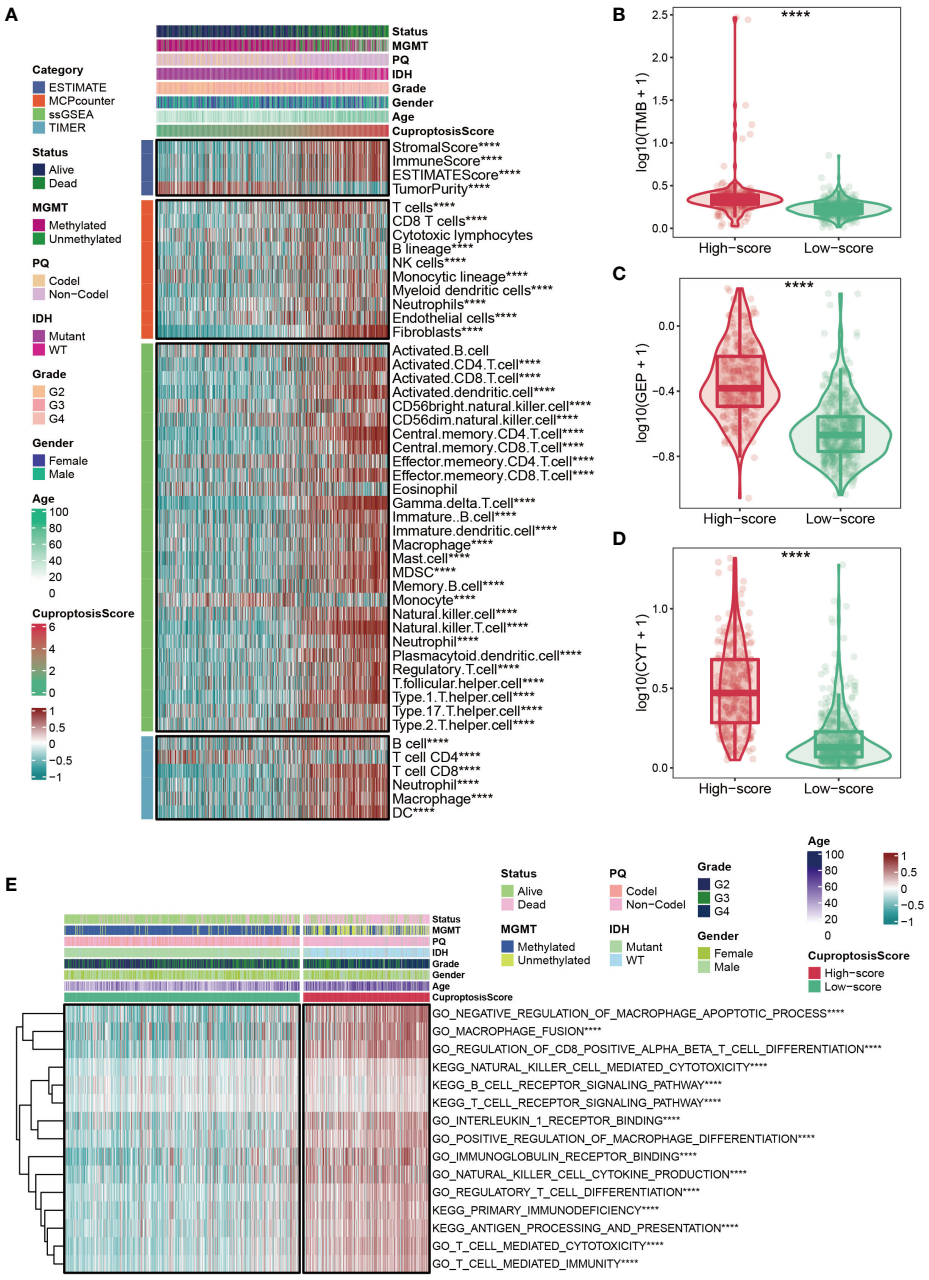


FIGURE 5
Immune status for cuproptosis-signature. **(A)** The heatmap shows the abundance of infiltrating immune cell populations at different CuproptosisScores. **(B–D)** Glioma patients with high CuproptosisScores had higher levels of TMB **(B)**, GEP **(C)**, and CYT **(D)**. **(E)** The heatmap shows CuproptosisScores, clinical features, and immune-related pathways based on GSVA analysis. ****P < 0.0001.

(Figure 6A). As the heatmap demonstrated that the level of TNFSF9, IL13, and TIGIT, showed no difference between the two groups, VTCN1, TNF, CX3CL1, IL12A, HMGB1, EDNRB, and TLR4 were highly expressed in the low-CuproptosisScore group, and the remaining genes were highly expressed in the high-CuproptosisScore group (Figure 6A). In addition, SubMap analysis revealed patients with high-CuproptosisScore may respond to anti-PD-1 therapy (Figure 6B). This may be due to the high expression of IMs in this group of patients. We also investigated the IC50 values of four chemotherapeutics (Bexarotene, Bicalutamide, Bortezomib, and Cytarabine) between the high- and low-CuproptosisScore groups. Results showed that IC50 values of patients in the high- CuproptosisScore group were lower than those in the low-CuproptosisScore group, suggesting

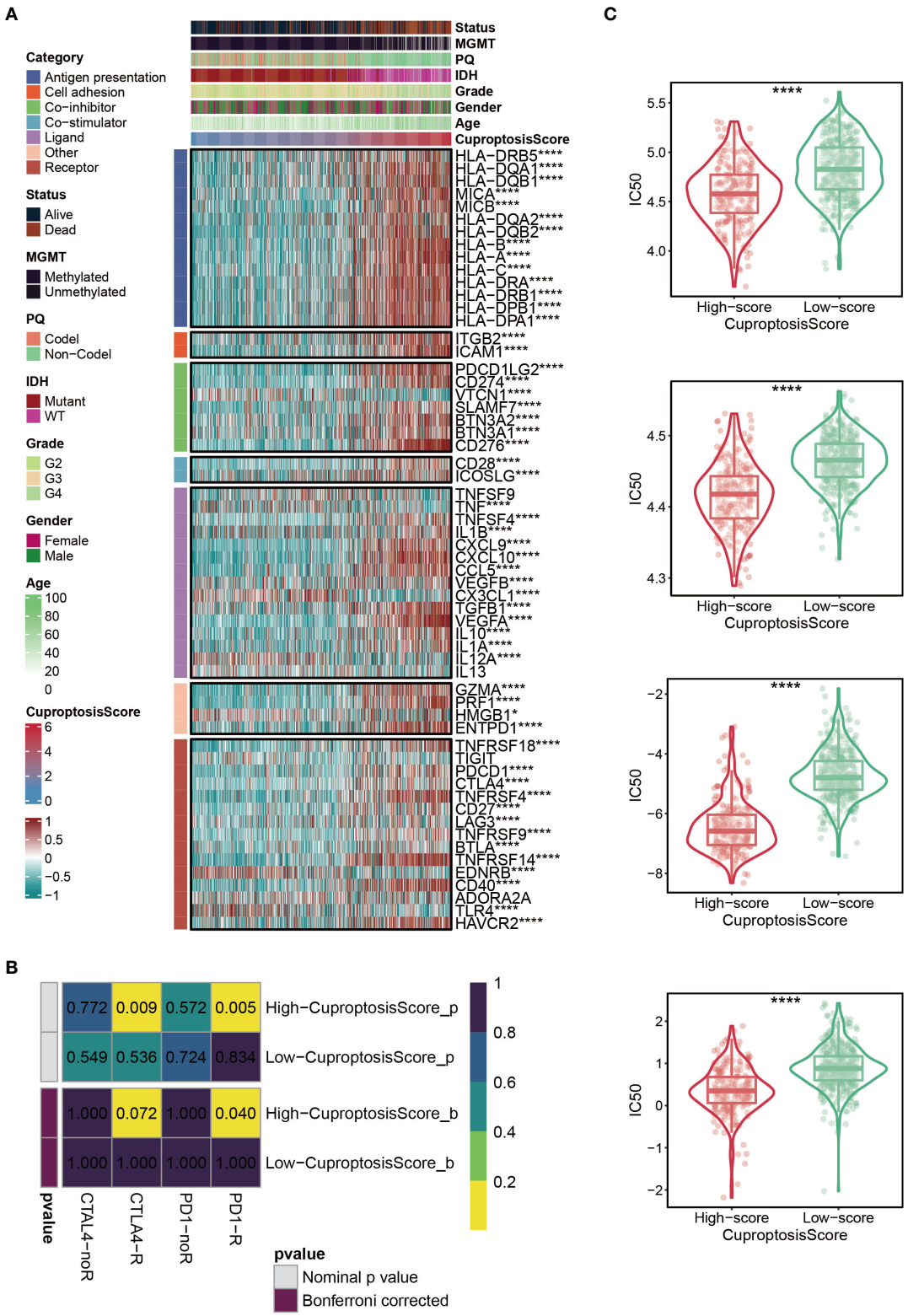


FIGURE 6
Immunotherapy and chemotherapy of cuproptosis-signature. **(A)** Correlation of CuproptosisScore with seven immunomodulators in gliomas. **(B)** SubMap analysis for cuproptosis-signature in gliomas. **(C)** Box plots of estimated IC50 for several chemotherapeutic agents in the high- or low-CuproptosisScore groups. *P < 0.05; **P < 0.01; ***P < 0.001, ****P < 0.0001.

that patients in the high-CuproptosisScore group were more likely to benefit from these four drugs (Figure 6C).

The importance of IGFBP2 in TCGA-glioma cohort

The aforementioned results showed that the cuproptosis-signature we created has substantial clinical significance. Next, we randomly selected one gene from this signature, IGFBP2, and explored its important value in gliomas. As is evident from WB and IHC results that IGFBP2 was significantly higher in the six tumor tissues than in the paired adjacent tumor tissues (Figures 7A–C). The patients with IGFBP2 expression values were listed in the Supplemental Table S2. The expression level of IGFBP2 was further compared between the glioma sample and the healthy control sample, it was found that IGFBP2 was significantly overexpressed in the cancer tissue (Figure 7D). In terms of the survival curve, glioma patients with low expression of IGFBP2 had better survival, indicating that IGFBP2 may be a promoter of the malignant progression of glioma. The 1-year (AUC=0.877), 2-year (AUC=0.92), 3-year (AUC=0.91), 4-year (AUC=0.858), and 5-year (AUC=0.822) survival ROC curves predicted by IGFBP2 revealed that the AUCs were all higher than 0.8, indicating the efficiency of IGFBP2 in predicting prognosis for glioma (Figures 7E, F). Therefore, IGFBP2 is highly likely to be the oncogenic gene of glioma. The heatmap (Figure 7G) showed that the expression value of IGFBP2 was positively correlated with the expression value of eight immune checkpoints (LAG3, CD274, PDCD1LG2, TNFRSF9, PDCD1, CTLA4, CD247, and TNFRSF4). Moreover, we carried out GSEA analysis to explore cancer and immune-related signaling pathways positively modulated by IGFBP2. We found six signaling pathways (Figure 7H): immune response, T cell receptor signaling pathway, regulation of immune response, pathways in cancer, p53 signaling pathway, and TF signaling pathway.

Discussion

Gliomas, especially glioblastoma (GBM), are the most destructive brain tumors within the human nervous system (43). Despite improvements in glioma diagnosis and treatment in recent years, gliomas are still difficult to treat with surgery alone due to their invasive and quickly proliferating nature. Patients with postoperative recurrence have a poor prognosis, with the median survival time only being extended by a few months (44, 45). One of the crucial characteristics of tumor cells is their resistance to cell death. Unrestricted proliferation is typical for tumor cells, and they overcome growth inhibition by resisting death and avoiding being killed by immune cells. However, due to metabolic stress, such as hypoxia and glucose deprivation, necrotizing cell death often occurs in the interior of solid tumors, which affects the occurrence and development of tumors by reshaping the tumor microenvironment. With the discovery of ever-more programmed death modes and the elucidation of associated molecular mechanisms, our understanding of the role of cell death in tumor is constantly updated. Since multiple forms of cell death occur simultaneously in tumors, an in-depth study of cell death on the occurrence and development of tumors can help us better understand their pathogenesis and pave the way for the creation of effective anti-tumor medications. For example, abnormalities of apoptosis pathways play critical roles in tumorigenesis, and tumor cell avoidance of apoptosis has long been thought to lead to primary or acquired therapeutic resistance (46). Necroptosis has both pro-tumor and anti-tumor effects in different types of cancer (13). Inducing necroptosis of tumor cells is an important way to overcome chemotherapy resistance of tumor cells. Finding a novel way to precisely regulate necroptosis might be an essential research target in the field of tumor therapy in the future (47). Numerous pieces of evidence suggest that pyrotopia plays an important role in tumor progression, and inducing pyrotopia has become one of the focuses of cancer immunotherapy (48–51). Ferroptosis is a type of cell death induced by oxidative stress. Cancer cells metabolize more efficiently than normal cells, with a higher ROS load (52) and require large amounts of iron, thus they are more sensitive to ferroptosis than normal cells (53). However, cancer cells also employ additional genetic or epigenetic mechanisms to combat elevated ROS levels, thereby reducing their sensitivity to ferroptosis (54). Therefore, ferroptosis is closely related to the occurrence and development of tumors.

Copper ions can be combined with a variety of proteins or enzymes, as cofactors or structural components, involved in the regulation of energy metabolism, mitochondrial respiration, antioxidant, and other physiological processes (55, 56). The content of copper ions maintains a dynamic balance, which can lead to oxidative stress (55) and abnormal autophagy (56), and thus induce a variety of copper or copper ion-related diseases. Tsvetkov et al. proposed for the first time that a new method of cell death with copper dependence, which was called cuprotosis (14). Several studies had shown that copper metabolism was associated with tumorigenesis, and cancer cells have a higher demand for copper than normal cells (57–60). Wang et al. found that blocking Cu^{2+} transport can cause oxidative stress and decrease cellular ATP levels, which in turn activates AMP-activated protein kinase (AMPK), leading to reduced adipogenesis and inhibiting tumor cell proliferation (61). Studies have confirmed that copper is closely related to the expression level of hypoxia-inducible factor 1α (HIF-1α) (62). The use of copper chelating agent tetrathiomolybdate can significantly reduce the content of Cu^{2+} *in vivo*, and dramatically reduce tumor angiogenesis, restrain tumor growth, and reduce the invasion of breast cancer cells (63). In conclusion, cuprotosis

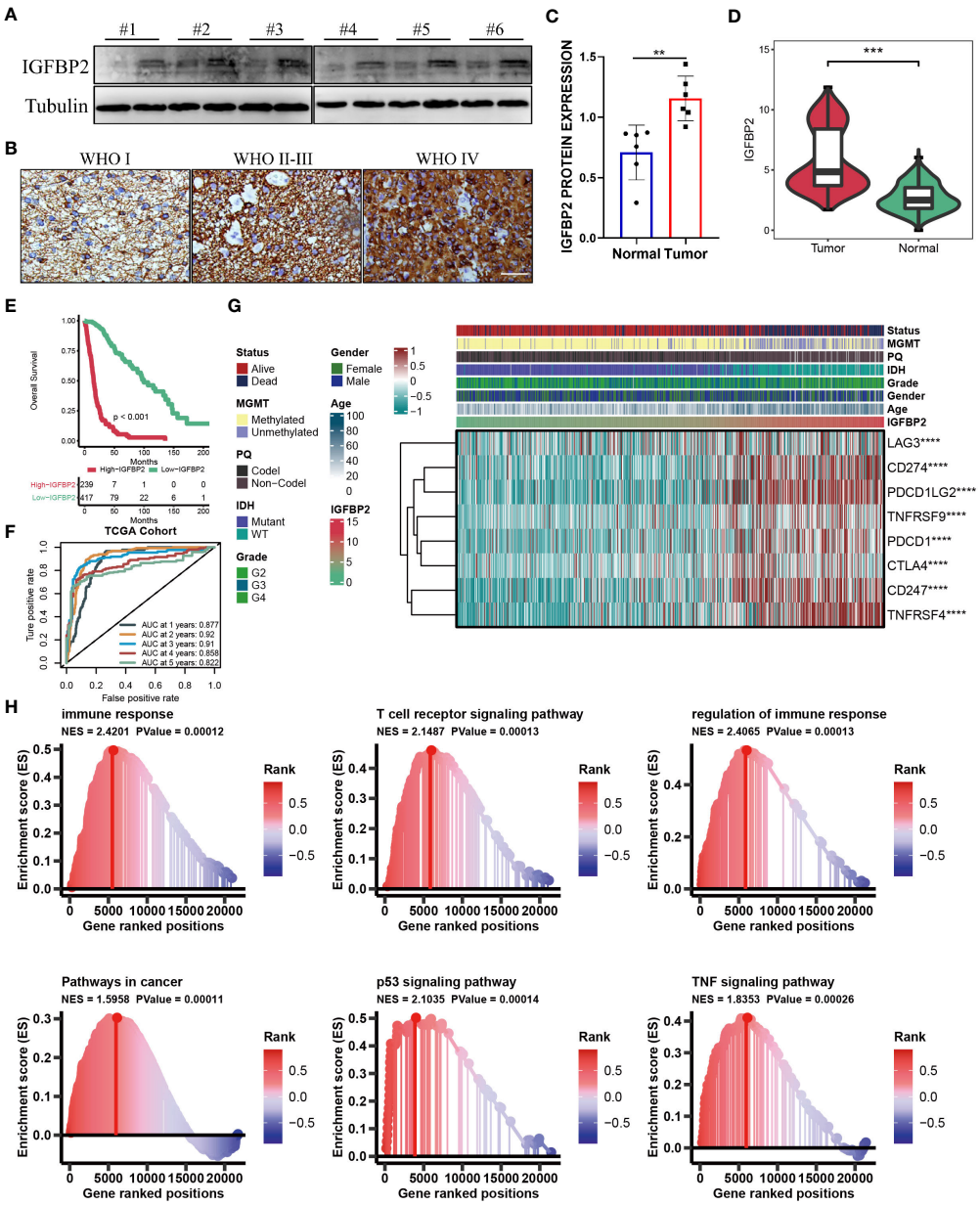


FIGURE 7
(A) WB for IGFBP2 in 3 pairs patients from Nantong cohort. **(B)** Represented IHC for IGFBP2 in three parents with different WHO stage from Nantong cohort. **(C)** Boxplot of IHC for IGFBP2 in six pairs parents from Nantong cohort. **(D)** The expression level of IGFBP2 in glioma sample and the control normal sample. **(E)** Kaplan-Meier survival curve showing survival probability of high- or low-expression IGFBP2. **(F)** The 1-year, 2-year, 3-year, 4-year, and 5-year survival ROC curves are predicted by the expression of IGFBP2. **(G)** The heat map shows the correlation between IGFBP2 and eight immune checkpoints in TCGA. **(H)** GSEA maps of cancer and immune-related signaling pathways positively modulated by IGFBP2. **P < 0.01; ***P < 0.001; ****P < 0.0001.

is a novel kind of cellular regulatory death that impacts copper metabolism. The identification of cuprotosis molecular pathways has implications for the mechanism of cuprotosis, cancer drug discovery, and a deeper understanding of copper metabolic diseases. In this study, we included ten cuproptosis-related genes (CRGs): FDX1, LIAS, LIPT1, DLD, DLAT, PDHA1, POHB, MTF1, GLS, and CDKN2A. The correlations among the ten CRGs were primarily positive in TCGA-glioma cohort. In addition, a prognostic signature based on CuproptosisScore was established to explore its prognostic and

clinical value in glioma. The current research provides a reference for exploring the mechanism of cuprotosis in the development of glioma.

As for the ten CRGs, the phosphorylation and dephosphorylation of PDHA1 (Pyruvate Dehydrogenase E1 alpha Subunit) are key modulators of deactivation and activation of PDC(Pyruvate Dehydrogenase Complex) (64). It was reported that the increasing level of PDHA1 was observed in the higher grade of glioma and PDHA1 could regulate the migration of glioma cells (65). LINC00665 promoted MTF1 degradation, and MTF1 bound to the promoter region of GTSE1 and transcription promoted GTSE1 expression, which proved that LINC00665/MTF1/GTSE1 axis played an important role in regulating the biological behavior of glioma cells (66). GLS are oncogenic genes of glioma (67, 68) and Qiangzhen Huang et al. found that GLS could regulate the effect of SNAP25 in glioma (66, 69). CDKN2A homozygous deletion was reported to serve as an adverse prognostic factor for IDH-mutant gliomas (70–72). However, to our knowledge, the role of the remaining CRGs in glioma has not been reported in the literature.

The complexity of gliomas is mainly reflected by their molecular heterogeneity. Molecular subtypes can well predict the occurrence and development of glioma polymorphism, which can assist us in developing better treatments (73). Mesenchymal subtypes are particularly malignant compared to other subtypes (neurogenic, canonical, and preneurotic) according to TCGA classification, with recurrent GBM always fatal and often presenting as a mesenchymal phenotype (74–76). In addition, mesenchymal subtypes of gliomas expressed higher levels of angiogenic markers in addition to higher levels of necrosis (74, 77). It has been reported that the transition from the former neural subtype to the mesenchymal subtype is closely associated with treatment resistance and poor prognosis (78). Currently, no fully verified and feasible classification system has been applied to clinical practice, and the glioma classification system needs to be continuously explored and improved. In this study, we determined the optimal number (k=2) of clusters based on R package (Figure 1B). Furthermore, the reliability of the clustering results was confirmed by PCA analysis (Figure 1C). Patients with glioma in cluster2 had significantly better clinical outcomes than those in cluster1 (Figure 1D). Interestingly, it was observed that cluster2 had a higher proportion of IDH mutation status, which may be one of the reasons for the better prognosis of this subcluster (Figure 1E). In the two subclusters, a total of 27 differentially expressed DEGs were identified (Figure 2A). When combined with the aforementioned findings, our study sheds light on the need for a new glioma typing system.

Although the WHO classification system has been used for many years to predict the prognosis of patients with glioma, it is occasionally inaccurate due to the heterogeneity of the tumor. In addition to identifying potential biomarkers, new advances in bioinformatics and genome sequencing can help predict cancer patient outcomes and treatment strategies (79, 80). Studies have shown that the prognostic value of a single biomarker is limited, and it is better to integrate multiple biomarkers into a single model (81). For example, three IncRNAs can predict the prognosis of colorectal cancer (CRC) based on a network of metastasis-related competing endogenous RNAs (ceRNA) (82). By extracting TCGA-related data, four IncRNA signals can effectively predict the survival time of lung adenocarcinoma (LUAD) (83). Importantly, recent studies have confirmed the predictive power of some IncRNA prognostic signatures in gliomas, such as immune-associated IncRNAs and autophagy-associated IncRNAs, which have strong prognostic potential for glioma patients (44, 45). In this study, we also used bioinformatics methods to identify a signature containing multiple genes based on CuproptosisScore. Patients with higher CuproptosisScore had higher WHO grades, while patients with lower grades were more likely to develop IDH mutations or MGMT methylation (Figure 3A) and patients with lower CuproptosisScore had the significantly better clinical outcomes (Figure 3B). In addition, Univariate and Multivariate Cox regression analysis of the signature demonstrated that the cuproptosis-signature was an independent prognostic factor for patients with glioma (Figure 3C). The survival ROC curves indicated the efficiency of the cuproptosis-signature in predicting prognosis for glioma (Figure 3D). Furthermore, we conducted Univariate Cox regression analysis on the OS of glioma patients based on the external validation data sets, and the results demonstrated the accuracy of our constructed cuproptosis-signature in prognostic prediction. In conclusion, the signature we have identified has excellent prognostic power (Figure 3E).

We randomly selected one gene from this signature, IGFBP2, and explored its important value in gliomas. In glioma, IGFBP2 is often involved in the activation of PTEN, AKT and other related pathways, leading to enhanced invasiveness and malignancy (84, 85). Studies have shown that overexpression of IGFBP2 can increase the malignant degree of glioma and up-regulate the expression of invasion protein MMP2, thereby enhancing the invasion ability of glioma cells (86). Previous studies have confirmed that the expression levels of IGFBP2 transcripts and proteins are positively correlated with the malignant degree of glioma, suggesting that IGFBP2 plays an important role in malignant transformation, tumor necrosis and metastasis of glioma (87, 88). In our study, we also found that IGFBP2 is highly likely to be the oncogenic gene of glioma.

However, our study still has several shortcomings. Due to the lack of data of clinical samples collected by us, the prognostic cuproptosis-signature constructed in this study was based on the public database, which may have a certain bias in the source of samples. The hypothesis obtained in this study was not verified by experimental results, and the next step is to be confirmed by various *in vivo* and *in vitro* experiments and larger multicenter studies. In conclusion, this prognostic cuproptosis-signature still

needs to be further tested, evaluated and applied in a wide range of clinical settings.

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

WW, ZCL, and MW carried out experiments and analysis. WW, PG, ZCL, MW, ZHL, BW, HH, and CY wrote the manuscript. WW and PG conceived the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Medical Scientific Research Project of Jiangsu Provincial Health Commission (H2019058).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.998236/full#supplementary-material

References

1. Gusyatiner O, Hegi ME. Glioma epigenetics: From subclassification to novel treatment options. *Semin Cancer Biol* (2018) 51:50–8. doi: 10.1016/j.semcancer.2017.11.010

2. Samioti SE, Benos LT, Sarris IE. Effect of fractal-shaped outer boundary of glioblastoma multiforme on drug delivery. *Comput Methods Programs BioMed* (2019) 178:191–9. doi: 10.1016/j.cmpb.2019.06.031

3. Polivka J Jr., Polivka J, Holubec L, Kubikova T, Priban V, Hes O, et al. Advances in experimental targeted therapy and immunotherapy for patients with glioblastoma multiforme. *Anticancer Res* (2017) 37:21–33. doi: 10.21873/anticanres.11285

4. Lapointe S, Perry A, Butowski NA. Primary brain tumours in adults. *Lancet* (2018) 10145:432–46. doi: 10.1016/s0140-6736(18)30990-5

5. Garcia CR, Slone SA, Dolecek TA, Huang B, Neltner JH, Villano JL. Primary central nervous system tumor treatment and survival in the united states, 2004-2015. *J Neurooncol* (2019) 144:179–91. doi: 10.1007/s11060-019-03218-8

6. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med* (2015) 372:2499–508. doi: 10.1056/NEJMoa1407279

7. Eskilsson E, Røsland GV, Solecki G, Wang Q, Harter PN, Graziani G, et al. EGFR heterogeneity and implications for therapeutic intervention in glioblastoma. *Neuro Oncol* (2018) 20:743–52. doi: 10.1093/neuonc/nox191

8. Zhao YY, Chen SH, Hao Z, Zhu HX, Xing ZL, Li MH. A nomogram for predicting individual prognosis of patients with low-grade glioma. *World Neurosurg* (2019) 130:e605–12. doi: 10.1016/j.wneu.2019.06.169

9. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta A. & dutta, a. a prognostic signature for lower grade gliomas based on expression of long non-coding RNAs. *Mol Neurobiol* (2019) 56:4786–98. doi: 10.1007/s12035-018-1416-y

10. Razzak M. Genetics: new molecular classification of gastric adenocarcinoma proposed by the cancer genome atlas. *Nat Rev Clin Oncol* (2014) 11:499. doi: 10.1038/nrclinonc.2014.138

11. Deng S, Li Y, Yi G, Lei B, Guo M, Xiang W, et al. Overexpression of COX7A2 is associated with a good prognosis in patients with glioma. *J Neurooncol* (2018) 136:41–50. doi: 10.1007/s11060-017-2637-z

12. Vanden Berghe T, Linkermann A, Jouan-Lanhouet S, Walczak H, Vandenabeele P. Regulated necrosis: the expanding network of non-apoptotic cell death pathways. *Nat Rev Mol Cell Biol* (2014) 15:135–47. doi: 10.1038/nrm3737

13. Qin X, Ma D, Tan YX, Wang HY, Cai Z. The role of necroptosis in cancer: A double-edged sword? *Biochim Biophys Acta Rev Cancer* (2019) 1871:259–66. doi: 10.1016/j.bbcan.2019.01.006

14. Tsvetkov P, Coy S, Petrova B, Dreishpoon M, Verma A, Abdusamad M, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. *Science* (2022) 375:1254–61. doi: 10.1126/science.abf0529

15. Kim BE, Nevitt T, Thiele DJ. Mechanisms for copper acquisition, distribution and regulation. *Nat Chem Biol* (2008) 4:176–85. doi: 10.1038/nchembio.72

16. Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome atlas (TCGA). *Methods Mol Biol* (2016) 1418:111–41. doi: 10.1007/978-1-4939-3578-9_6

17. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via Xena Platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8

18. Human genomics. the genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* (2015) 348:648–60. doi: 10.1126/science.1262110

19. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* (2015) 21:938–45. doi: 10.1038/nm.3909

20. Katz S, Irizarry RA, Lin X, Tripputi M, Porter MW. A summarization approach for affymetrix GeneChip data using a reference training set from a large, biologically diverse database. *BMC Bioinf* (2006) 7:464. doi: 10.1186/1471-2105-7-464

21. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *Omics* (2010) 14:109–13. doi: 10.1089/omi.2009.0083

22. Ringnér M. What is principal component analysis? *Nat Biotechnol* (2008) 26:303–4. doi: 10.1038/nbt0308-303

23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007

24. van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basic concepts and methods of cox regression. *Kidney Int* (2008) 74:705–9. doi: 10.1038/ki.2008.294

25. Taylor JM. Random survival forests. *J Thorac Oncol* (2011) 6:1974–5. doi: 10.1097/JTO.0b013e318233d835

26. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med* (2019) 38:558–82. doi: 10.1002/sim.7803

27. Ng PK, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, et al. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* (2018) 33:450–62.e410. doi: 10.1016/j.ccell.2018.01.021

28. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12:R41. doi: 10.1186/gb-2011-12-4-r41

29. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

30. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* (2017) 77:e108–10. doi: 10.1158/0008-5472.Can-17-0307

31. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17:218. doi: 10.1186/s13059-016-1070-5

32. Xiao B, Liu L, Li A, Xiang C, Wang P, Li H, et al. Identification and verification of immune-related gene prognostic signature based on ssGSEA for osteosarcoma. *Front Oncol* (2020) 10:607622. doi: 10.3389/fonc.2020.607622

33. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

34. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. *Immunity* (2019) 51:411–2. doi: 10.1016/j.immuni.2019.08.004

35. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PloS One* (2007) 2:e1195. doi: 10.1371/journal.pone.0001195

36. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* (2016) 351:1463–9. doi: 10.1126/science.aaf1490

37. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia* (2019) 21:591–601. doi: 10.1016/j.neo.2019.04.003

38. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004

39. Subramanian A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

40. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41:D955–961. doi: 10.1093/nar/gks1111

41. Geeleher P, Cox N, Huang RS. pRRophetic: an r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468

42. Tang J, Shalabi A, Hubbard-Lucey VM. Comprehensive analysis of the clinical immuno-oncology landscape. *Ann Oncol* (2018) 29:84–91. doi: 10.1093/annonc/mdx755

43. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* (2015) 16:e173–180. doi: 10.1016/s1470-2045(14)71116-7

44. Luan F, Chen W, Chen M, Yan J, Chen H, Yu H, et al. An autophagy-related long non-coding RNA signature for glioma. *FEBS Open Bio* (2019) 9:653–67. doi: 10.1002/2211-5463.12601

45. Xia P, Li Q, Wu G, Huang Y. An immune-related lncRNA signature to predict survival in glioma patients. *Cell Mol Neurobiol* (2021) 41:365–75. doi: 10.1007/s10571-020-00857-8

46. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013

47. Chefetz I, Grimley E, Yang K, Hong L, Vinogradova EV, Suciu R, et al. A pan-ALDH1A inhibitor induces necroptosis in ovarian cancer stem-like cells. *Cell Rep* (2019) 26:3061–75.e3066. doi: 10.1016/j.celrep.2019.02.032

48. Wang Q, Wang Y, Ding J, Wang C, Zhou X, Gao W, et al. A bioorthogonal system reveals antitumour immune function of pyroptosis. *Nature* (2020) 579:421–6. doi: 10.1038/s41586-020-2079-1

49. Kaplanov I, Carmi Y, Kornetsky R, Shemesh A, Shurin GV, Shurin MR, et al. Blocking IL-1β reverses the immunosuppression in mouse breast cancer and synergizes with anti-PD-1 for tumor abrogation. *Proc Natl Acad Sci U.S.A.* (2019) 116:1361–9. doi: 10.1073/pnas.1812266115

50. Kaplanski G. Interleukin-18: Biological properties and role in disease pathogenesis. *Immunol Rev* (2018) 281:138–53. doi: 10.1111/imr.12616

51. Wang Y, Gao W, Shi X, Ding J, Liu W, He H, et al. Chemotherapy drugs induce pyroptosis through caspase-3 cleavage of a gasdermin. *Nature* (2017) 547:99–103. doi: 10.1038/nature22393

52. Tsoi J, Robert L, Paraiso K, Galvan C, Sheu KM, Lay J, et al. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell* (2018) 33:890–904.e895. doi: 10.1016/j.ccell.2018.03.017

53. Carlisle AE, Lee N, Matthew-Onabanjo AN, Spears ME, Park SJ, Youkana D, et al. Selenium detoxification is required for cancer-cell survival. *Nat Metab* (2020) 2:603–11. doi: 10.1038/s42255-020-0224-7

54. Rojo de la Vega M, Chapman E, Zhang DD. NRF2 and the hallmarks of cancer. *Cancer Cell* (2018) 34:21–43. doi: 10.1016/j.ccell.2018.03.022

55. Guo H, Wang Y, Cui H, Ouyang Y, Yang T, Liu C, et al. Copper induces spleen damage through modulation of oxidative stress, apoptosis, DNA damage, and inflammation. *Biol Trace Elem Res* (2022) 200:669–77. doi: 10.1007/s12011-021-02672-8

56. Jian Z, Guo H, Liu H, Cui H, Fang J, Zuo Z, et al. Oxidative stress, apoptosis and inflammatory responses involved in copper-induced pulmonary toxicity in mice. *Aging (Albany NY)* (2020) 12:16867–86. doi: 10.18632/aging.103585

57. Atakul T, Altinkaya SO, Abas BI, Yenisey C. Serum copper and zinc levels in patients with endometrial cancer. *Biol Trace Elem Res* (2020) 195:46–54. doi: 10.1007/s12011-019-01844-x

58. Feng Y, Zeng JW, Ma Q, Zhang S, Tang J, Feng JF, et al. Serum copper and zinc levels in breast cancer: A meta-analysis. *J Trace Elem Med Biol* (2020) 62:126629. doi: 10.1016/j.jtemb.2020.126629

59. Pavithra V, Sathisha TG, Kasturi K, Mallika DS, Amos SJ, Ragunatha S, et al. Serum levels of metal ions in female patients with breast cancer. *J Clin Diagn Res* (2015) 9:BC25–c27. doi: 10.7860/jcdr/2015/11627.5476

60. Zhang M, Shi M, Zhao Y. Association between serum copper levels and cervical cancer risk: A meta-analysis. *Biosci Rep* (2018) 38:4. doi: 10.1042/bsr20180161

61. Wang J, Luo C, Shan C, You Q, Lu J, Elf S, et al. Inhibition of human copper trafficking by a small molecule significantly attenuates cancer cell proliferation. *Nat Chem* (2015) 7:968–79. doi: 10.1038/nchem.2381

62. Zimna A, Kurpisz M. Hypoxia-inducible factor-1 in physiological and pathophysiological angiogenesis: Applications and therapies. *BioMed Res Int* (2015) 2015:549412. doi: 10.1155/2015/549412

63. MacDonald G, Nalvarte I, Smirnova T, Vecchi M, Aceto N, Dolemeyer A, et al. Memo is a copper-dependent redox protein with an essential role in migration and metastasis. *Sci Signal* (2014) 7:ra56. doi: 10.1126/scisignal.2004870

64. Holness MJ, Sugden MC. Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem Soc Trans* (2003) 31:1143–51. doi: 10.1042/bst0311143

65. Yang Z, Wang Y, Zhang L, Zhao C, Wang D. Phosphorylated form of pyruvate dehydrogenase α1 mediates tumor necrosis factor α-induced glioma cell migration. *Oncol Lett* (2021) 21:176. doi: 10.3892/ol.2021.12437

66. Ruan X, Zheng J, Liu X, Liu Y, Liu L, Ma J, et al. lncRNA LINC00665 stabilized by TAF15 impeded the malignant biological behaviors of glioma cells *via* STAU1-mediated mRNA degradation. *Mol Ther Nucleic Acids* (2020) 20:823–40. doi: 10.1016/j.omtn.2020.05.003

67. Martín-Rufián M, Nascimento-Gomes R, Higuero A, Crisma AR, Campos-Sandoval JA, Gómez-García MC, et al. Both GLS silencing and GLS2 overexpression synergize with oxidative stress against proliferation of glioma cells. *J Mol Med (Berl)* (2014) 92:277–90. doi: 10.1007/s00109-013-1105-2

68. Szeliga M, Bogacińska-Karaś M, Różycka A, Hilgier W, Marquez J, Albrecht J. Silencing of GLS and overexpression of GLS2 genes cooperate in decreasing the proliferation and viability of glioblastoma cells. *Tumour Biol* (2014) 35:1855–62. doi: 10.1007/s13277-013-1247-4

69. Huang Q, Lian C, Dong Y, Zeng H, Liu B, Xu N, et al. SNAP25 inhibits glioma progression by regulating synapse plasticity via GLS-Mediated. *Glutaminolysis Front Oncol* (2021) 11:698835. doi: 10.3389/fonc.2021.698835

70. Appay R, Dehais C, Maurage CA, Alentorn A, Carpentier C, Colin C, et al. CDKN2A homozygous deletion is a strong adverse prognosis factor in diffuse malignant IDH-mutant gliomas. *Neuro Oncol* (2019) 21:1519–28. doi: 10.1093/neuonc/noz124

71. Pessôa IA, Amorim CK, Ferreira WAS, Sagica F, Brito JR, Othman M, et al. Detection and correlation of single and concomitant TP53, PTEN, and CDKN2A alterations in gliomas. *Int J Mol Sci* (2019) 20:11. doi: 10.3390/ijms20112658

72. Lu VM, O'Connor KP, Shah AH, Eichberg DG, Luther EM, Komotar RJ, et al. The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: A systematic review of the contemporary literature. *J Neurooncol* (2020) 148:221–9. doi: 10.1007/s11060-020-03528-2

73. Wirsching HG, Galanis E, Weller M. Glioblastoma. *Handb Clin Neurol* (2016) 134:381–97. doi: 10.1016/b978-0-12-802997-8.00023-2

74. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* (2006) 9:157–73. doi: 10.1016/j.ccr.2006.02.019

75. Kwon SM, Kang SH, Park CK, Jung S, Park ES, Lee JS, et al. Recurrent glioblastomas reveal molecular subtypes associated with mechanistic implications of drug-resistance. *PloS One* (2015) 10:e0140528. doi: 10.1371/journal.pone.0140528

76. Wood MD, Reis GF, Reuss DE, Phillips JJ. Protein analysis of glioblastoma primary and posttreatment pairs suggests a mesenchymal shift at recurrence. *J Neuropathol Exp Neurol* (2016) 75:925–35. doi: 10.1093/jnen/nlw068

77. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* (2010) 17:98–110. doi: 10.1016/j.ccr.2009.12.020

78. Yoon SJ, Shim JK, Chang JH, Moon JH, Roh TH, Sung KS, et al. Tumor mesenchymal stem-like cell as a prognostic marker in primary glioblastoma. *Stem Cells Int* (2016) 2016:6756983. doi: 10.1155/2016/6756983

79. Desany B, Zhang Z. Bioinformatics and cancer target discovery. *Drug Discovery Today* (2004) 9:795–802. doi: 10.1016/s1359-6446(04)03224-6

80. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* (2009) 37:1–13. doi: 10.1093/nar/gkn923

81. Tian X, Zhu X, Yan T, Yu C, Shen C, Hong J, et al. Differentially expressed lncRNAs in gastric cancer patients: A potential biomarker for gastric cancer prognosis. *J Cancer* (2017) 8:2575–86. doi: 10.7150/jca.19980

82. Liu S, Cao Q, An G, Yan B, Lei L. Identification of the 3-lncRNA signature as a prognostic biomarker for colorectal cancer. *Int J Mol Sci* (2020) 21:24. doi: 10.3390/ijms21249359

83. Sui J, Yang S, Liu T, Wu W, Xu S, Yin L, et al. Molecular characterization of lung adenocarcinoma: A potential four-long noncoding RNA prognostic signature. *J Cell Biochem* (2019) 120:705–14. doi: 10.1002/jcb.27428

84. Moore LM, Holmes KM, Smith SM, Wu Y, Tchougounova E, Uhrbom L, et al. IGFBP2 is a candidate biomarker for Ink4a-arf status and a therapeutic target for high-grade gliomas. *Proc Natl Acad Sci USA* (2009) 106:16675–9. doi: 10.1073/pnas.0900807106

85. Azar WJ, Zivkovic S, Werther GA, Russo VC. IGFBP-2 nuclear translocation is mediated by a functional NLS sequence and is essential for its pro-tumorigenic actions in cancer cells. *Oncogene* (2014) 33:578–88. doi: 10.1038/onc.2012.630

86. Zhou YH, Hess KR, Liu L, Linskey ME, Yung WK. Modeling prognosis for patients with malignant astrocytic gliomas: quantifying the expression of multiple genetic markers and clinical variables. *Neuro Oncol* (2005) 7:485–94. doi: 10.1215/s1152851704000730

87. Zheng S, Houseman EA, Morrison Z, Wrensch MR, Patoka JS, Ramos C, et al. DNA Hypermethylation profiles associated with glioma subtypes and EZH2 and IGFBP2 mRNA expression. *Neuro Oncol* (2011) 13:280–9. doi: 10.1093/neuonc/noq190

88. Elmlinger MW, Deininger MH, Schuett BS, Meyermann R, Duffner F, Grote EH, et al. *In vivo* expression of insulin-like growth factor-binding protein-2 in human gliomas increases with the tumor grade. *Endocrinology* (2001) 142:1652–8. doi: 10.1210/endo.142.4.8084

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fimmu.2022.950365

Check for updates

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Guo Wen Zhi,
Zhengzhou University, China
Qianqian Song,
Wake Forest School of Medicine,
United States

*CORRESPONDENCE
Yuanji Xu
xuyuanji@fjmu.edu.cn
Yingyong Hou
hou.yingyong@zs-hospital.sh.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 22 May 2022
ACCEPTED 22 August 2022
PUBLISHED 08 September 2022

CITATION
Wang H, Zhang Y, Chen L, Liu YF, Xu C, Jiang D, Song Q, Wang H, Wang L, Lin Y, Chen Y, Chen J, Xu Y and Hou Y (2022) Identification of clinical prognostic features of esophageal cancer based on m6A regulators.
Front. Immunol. 13:950365.
doi: 10.3389/fimmu.2022.950365

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

Identification of clinical prognostic features of esophageal cancer based on m6A regulators

Huimei Wang[1†], Yiping Zhang[2†], Lin Chen[2†], Yufeng Liu[1†], Chen Xu[1], Dongxian Jiang[1], Qi Song[1], Haixing Wang[1], Liyan Wang[2], Yu Lin[3], Yuanmei Chen[4], Junqiang Chen[3], Yuanji Xu[3*] and Yingyong Hou[1*]

[1]Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai, China, [2]Clinical Oncology School of Fujian Medical University, Fuzhou, China, [3]Department of Radiation Oncology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China, [4]Department of Thoracic Surgery, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China

Background: Esophageal cancer (ESCA) is a common malignancy with high morbidity and mortality. n6-methyladenosine (m6A) regulators have been widely recognized as one of the major causes of cancer development and progression. However, for ESCA, the role of regulators is unclear. The aim of this study was to investigate the role of m6A RNA methylation regulators in the immune regulation and prognosis of ESCA.

Methods: RNA-seq data were downloaded using the Cancer Genome Atlas (TCGA) database, and the expression differences of m6A RNA methylation regulators in ESCA were analyzed. Further m6A methylation regulator markers were constructed, and prognostic and predictive values were assessed using survival analysis and nomograms. Patients were divided into low-risk and high-risk groups. The signature was evaluated in terms of survival, single nucleotide polymorphism (SNP), copy number variation (CNV), tumor mutation burden (TMB), and functional enrichment analysis (TMB). The m6A expression of key genes in clinical specimens was validated using quantitative reverse transcription polymerase chain reaction (qRT-PCR).

Results: In ESCA tissues, most of the 23 regulators were significantly differentially expressed. LASSO regression analysis included 7 m6A-related factors (FMR1, RBMX, IGFBP1, IGFBP2, ALKBH5, RBM15B, METTL14). In addition, this study also identified that the risk model is associated with biological functions, including base metabolism, DNA repair, and mismatch repair. In this study, a nomogram was created to predict the prognosis of ESCA patients. Bioinformatics analysis of human ESCA and normal tissues was performed using qRT-PCR. Finally. Seven genetic features were found to be

associated with m6A in ESCA patients. The results of this study suggest that three different clusters of m6A modifications are involved in the immune microenvironment of ESCA, providing important clues for clinical diagnosis and treatment.

KEYWORDS

esophageal cancer, m6A RNA methylation regulators, prognosis, signature, the cancer genome atlas, nomogram

Introduction

Esophageal cancer is a common malignancy, and in 2020, ESCA had the 7th highest incidence (604,000 new cases) and 6th highest incidence (544,000 deaths) of all cancers (1). ESCA has a high mortality rate due to the lack of effective diagnostic and treatment strategies.

N6-methyladenosine (m6A) is the most common epigenetic RNA modification that plays an important role in the regulation of malignancies (2). Despite the potential of m6A for the diagnosis and treatment of ESCA (3, 4), its potential targets and mechanisms remain unclear.

m6A refers to the methylation reaction at the sixth position of adenosine. m6A methylation abnormalities play an important role in many diseases, especially in tumors (5). Methyltransferases (METTL3, METTL14, METTL16, WTAP, KIAA1429, ZC3H13, RBM15), demethylases (FTO, ALKBH5), binding proteins (HNRNPC, HNRNPA2B1, YTHDF1, YTHDF2, YTHDC2, YTHDC2, YTHDC1). m6A methylation regulators have important effects on ESCA progression, proliferation, migration, and invasion. hNRNPA2B1 affects the prognosis of ESCA by regulating the miR-17-92 cluster as an oncogenic factor (6). In addition, HNRNPA2B1 promotes ESCA progression through upregulation of fatty acid synthase ACLY, ACC1 (7). ALKBH5 exerts tumor suppressive effects by inhibiting miR-194-2 biogenesis through demethylation of pri-miR-194-2, thereby inhibiting RAl1 (8). Another study showed that FTO is involved in oncogenesis of ESCA through upregulation of MMP13 (9). In addition, METTL3 may also promote ESCA proliferation and invasion by regulating multiple pathways, such as AKT (10), Notch (11), COL12A1/MAPK (12) and other signaling pathways.

In addition, Xu et al. (13) demonstrated that eight regulators (KIAA1429, HNRNPC, RBM15, METTL3, WTAP, YTHDF1, YTHDC1, YTHDF2) were significantly upregulated in ESCA tissues. The above results suggest that the prognostic features of two genes, ALKBH5 and HNRNPC, have a predictive effect on prognosis. In another study (7), HNRNPA2B1, ALKBH5 was the prognostic signature consisting of HNRNPA2B1 and ALKBH5 (7). In addition, a recent study found that m6A methylation regulators may be important mediators of PD-L1 expression and immune cell infiltration, which may strongly influence the tumor microenvironment of esophageal squamous cell carcinoma (ESCC) (14). Furthermore, Saiyan et al. (15) suggested that Flap endonuclease 1 Facilitated hepatocellular carcinoma progression by enhancing USP7/MDM2-mediated P53 inactivation. However, previous bioinformatics studies were relatively simple: no multi-omics integration analysis or assessment of tumor mutation burden (TMB) was used, or focusing on exploration of the single gene, or without experimental validation. The aim of this study was to investigate the molecular targets and therapeutic mechanisms of ESCA.

Methods

Data acquisition and processing

RNA transcriptome data (FPKM) format was obtained from the TCGA public database (https://portal.gdc.cancer.gov/), along with copy number variation (CNV), somatic mutations, corresponding clinical data, TNM classification, survival information, and prognostic data for ESCA patients. To ensure data consistency, FPKM was transformed into transcripts per kilobase (TPM) values. Patients with no clinical information were excluded from this study. Finally, 161 ESCA samples and 11 adjacent tissues were included in the analysis. For CNV analysis, the "Circos" R package was used, and CNV genes were mapped on 23 chromosome pairs. Somatic mutation data were obtained from the TCGA database, and somatic mutation data were visualized using the "maftools" software (16). In addition, a TMB examination was performed for each patient.

Construction and prediction of predictive features

First, the differences in m6A-related regulators expression between ESCA samples and normal tissues, and the relationship

between regulator expression and prognosis of ESCA patients were analyzed. To determine the prognostic value of m6A-associated regulators, TCGA-ESCARNA-seq candidate risk regulators were selected in combination with LASSO regression analysis to reduce the dimensionality and select representative indicators. Subsequently, the selected genes were subjected to dimensionality reduction analysis to determine whether the selected genes had independent prognostic value. Finally, the minimum criterion was used to determine the corresponding specification coefficients. The regression coefficients were estimated based on the LASSO regression model, and the results were calculated as follows.

$$\text{riskScore} = \sum_i \text{Coefficient}(\text{hub gene}_i) \times \text{mRNA Expression}(\text{hub gene}_i)$$

ESCA patients were divided into low-risk and high-risk subgroups according to median risk score.

Genomic and functional analysis

Gene ontology (GO) functional analysis is a common method for large-scale functional enrichment studies, including biological process (BP), molecular function (MF), and cellular component (CC). GO analysis was performed using the clusterProfiler (17) R package, based on differential gene expression analysis between high-risk and low-risk groups. False discovery rate $q<0.05$ was considered statistically significant.

Gene set enrichment analysis (GSEA) is a computational method to analyze whether statistical differences exist in a particular gene set. In this study, the GSEA method was used to analyze TCGA-ESCA RNA-seq data to explore the differences in BP between different sets." The h.all.v7.2.symbols.gm t set was downloaded from the MSigDB (18) database and used for GSEA. $p<0.05$, statistically significant.

Quantitative validation of pivotal genes using reverse transcription polymerase chain reaction

The expression of IGFBP2, ALKBH5, FMR1, RBMX in 15 pairs of ESCA and adjacent esophageal tissues was detected by quantitative reverse transcription polymerase chain reaction (qRT-PCR). IGFBP2, ALKBH5, FMR1, RBMX, were acquired from BioSune (Shanghai, China). qRT-PCR analysis was performed using the Hieff® qPCR ® qRT-PCR system (Applied Biosciences, USA, USA) was used for qRT-PCR analysis. Reactions were performed at 95°C for 10 min,95°C, 40 cycles for 15 s, and 60°C for 1 min. The relative levels of gene expression were calculated by the 2-Δ Ct method using GAPDH as a reference gene. The study was conducted according to the Declaration of Helsinki of the World Medical Association. The School of Clinical Oncology of Fujian Medical University approved the use of human tissues.

Consistent clustering study of m6A-related genes

To assess genetic identity, molecules significantly associated in the m6A risk model were analyzed using the Spearman method. p values were adjusted with the Benjamin-Hochberg test. When the absolute correlation coefficient was greater than 0.3 and $P<0.01$, it was significantly associated with genetic correlation. Tumor samples were clustered into distinct GeneClusters using the Kaplan-Meier method, which partitions around significantly correlated molecular expression and Euclidean measurement distances. Specifically, clustering analysis was performed using the ConsensusClusterPlus (19) R package, with 1000 cycles of calculation to ensure stability and reliability of the classification. To study the signature genomes, the Boruta algorithm was used to perform a dimensionality reduction analysis of significantly related genomes. Then, two classes, signature gene A and signature B, were clustered according to signature gene expression changes and visualized using the ComplexHeatmap package in the R software.

Copy number variation analysis

For CNV analysis, masked copy number segment datasets for different risk groups were downloaded from TCGA-ESCA. The data were examined using GISTIC 2.0 (20). GenePattern 5 was used for the above analysis.

Construct and validate the prediction nomogram

To improve the value of the signature in clinical practice, clinical factors (T, N, M, TNM) and m6A risk score were used as prognostic nomograms to evaluate the probability of OS occurrence in ESCA patients at 1, 2, and 3 years. To quantify the discriminatory performance of the nomogram, the concordance index (C-index) of T, N, M, m6A risk score, TNM, TNM+m6A risk score were compared. Calibration curves, time-dependent receiver operating characteristic (ROC) curves, and decision curve analysis (DCA) were used to examine the TNM and TNM+m6A risk scores.

Results

Genetic variation of m6A-related genes

To analyze the overall expression of m6A-related genes in ESCA patients, this paper first analyzed mutations and gene expression levels, including single nucleotide polymorphisms (SNPs), CNV, and gene expression. Among the 172 samples, 23 samples showed SNPs of m6A-related regulators, mainly missense mutations (Figure 1A). Subsequently, we summarized the incidence of CNV for the 23 m6A-associated regulators in the ESCA samples. Figure 1B shows the altered CNV chromosomal location. CNV alterations were widespread in m6A-related genes, with reduced copy number in most patients (Figure 1C). In addition, the mRNA expression levels of m6A-related genes were analyzed between ESCA samples in this study, and the results showed that all genes were differentially expressed except METTL14, ZC3H13, RBM15, YTHDC2, and IGFBP1 (Figure 1D). Spearman correlation analysis was applied to correlate the 23 m6A RNA methylation regulators. From Figure 1E, the transcriptome associations were explored, and we suggested that there are close correlations among writers, erasers and readers. The correlation between RBMX and HNRNPC, HNRNPA2B1 was the highest (P<0.01), while the correlation between YTHDF3, VIRMA was the highest (P<0.01) (P<0.01).

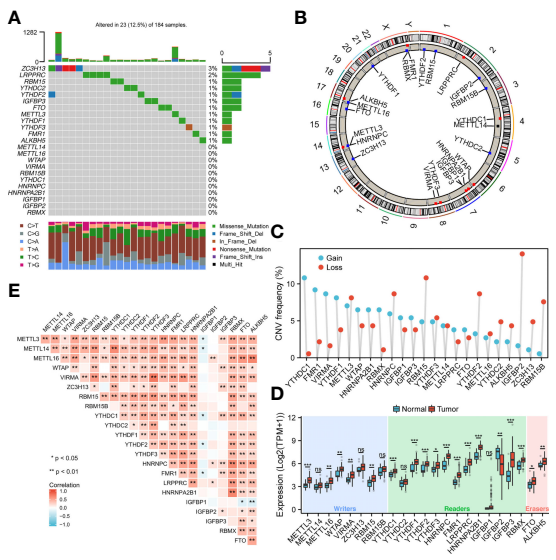


FIGURE 1
Genetic variants of m6A-related genes. **(A)** SNP of m6A related genes in 23 samples. **(B)** Location of CNV alterations on chromosomes. **(C)** Frequency of CNV in m6A related genes. Blue represents amplification, orange represents deletion. **(D)** Genetic variants of m6A-related genes. **(E)** Diagram showing the relationship between different m6A-related genes through Pearson correlation analysis. Red and blue represent a positive and negative correlation, respectively. *p < 0.05, **p < 0.01, ***p < 0.001; ns, not significant.

Construction and validation of a prognostic risk model based on seven m6A methylation regulators

Next, we analyzed the role of m6A-related regulators in ESCA patients. The m6A regulatory network shown in Figure 2A reveals the interactions with m6A-related genes,

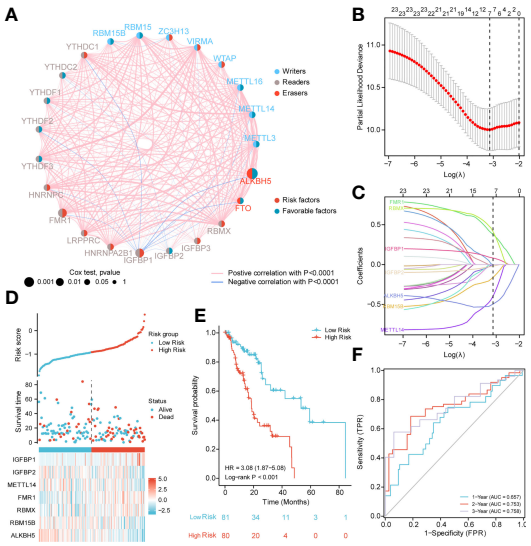


FIGURE 2
PPI network and prognostic signatures construction and prediction. **(A)** Interaction among m6A-related regulators. Circle size indicates the effect of each gene on survival, the larger the size, the greater the effect; on the right half of the circle, red represents risk prognostic factors and blue represents favorable factors; on the left side of the circle, red represents recognition proteins (readers), blue represents methyltransferases (writers), and brown demethyltransferases (erasers); lines that connect genes exhibit genetic interactions, red and blue represent positive and negative associations, respectively. **(B)** Partial likelihood deviance of different numbers of variables. One-thousand-fold cross-validation was applied for tuning penalty parameter selection. **(C)** LASSO analysis identified seven m6A-related genes in the 23 m6A-related regulators cohort. Each curve corresponds to one gene (cyan, FMR1; green, RBMX; pink, IGFBP1; brown IGFBP2; blue, ALKBH5; orange, RBM15B; purple, METTL14). **(D)** Risk score, distribution of patient survival status between the low- and high–risk groups, and expression heat maps of seven m6A-related regulators. **(E)** Kaplan–Meier curves indicated that there is a strong relationship between high and low m6Ascore and the overall survival rate. **(F)** ROC curve was applied to assess the predictive efficiency of the prognostic risk signature.

nodes, and their role in ESCA prognosis. It was found that not only the m6A regulators share the same functional class, but also the expression of functional class m6A regulators was significantly correlated.

In addition, to quantitatively evaluate the effect of m6A-related regulators on the prognosis of each ESCA patient, we constructed a risk model of m6A-related regulator expression. First, based on the results of LASSO regression analysis, the minimum-minimum criterion and optimal-minimum criterion M6A-related genes, including FMR1, RBMX, IGFBP1, IGFBP2, ALKBH5, RBM15B, METTL14, were used (Figure 2B–C). Meanwhile, the penalty coefficients of the characteristic regulators were calculated using LASSO analysis, and the risk index was established by multiplying the gene expression by the corresponding coefficients. The risk score of each sample was then calculated based on the median of m6A scores and divided into two groups of low risk and high risk. The risk score distribution, survival status and characteristic gene expression patterns are shown in Figure 2D. Kaplan-Meier survival analysis showed that OS was significantly lower in the high-risk group ESCA patients than in the low-risk group (log-rank p<0.001, Figure 2E).

The sensitivity and predictive specificity of the risk scores were investigated using ROC curve analysis. The AUC values were 0.657, 0.753, 0.758, and 0.758, respectively (Figure 2F). The AUC values showed that the risk scores significantly predicted the prognosis of patients with ESCA.

GO enrichment and genome enrichment analysis

In this study, GO analysis was used to explore the biological functions of the low-risk and high-risk groups. The results showed the processes of organ identity maintenance, negative regulation of synaptic vesicle extravasation, tubular boys, meiotic telomere and nuclear envelope attachment, meiotic telomere aggregation, telomere localization, neuronal action potential regulation, purine nucleotide metabolism in the high-risk group animals. decoder complex, tetrahydrobiopterin biosynthesis process, reduced food intake, brush border assembly, galactolipid biosynthesis, glycosylceramide biosynthesis, transmembrane transport of pyrimidine compounds, and purine nucleobase transport (Figure 3A). Next, we performed the GSEA analysis shown in Figure 3B. The high-risk group was associated with antigen processing mechanisms, EMT-1, and mismatch repair. The low-risk group was correlated with CD8-T, immune checkpoint, EMT2, pan-F TBRS, angiogenesis, Fanconi anemia, DNA damage repair, WNT target, and DNA damage response.

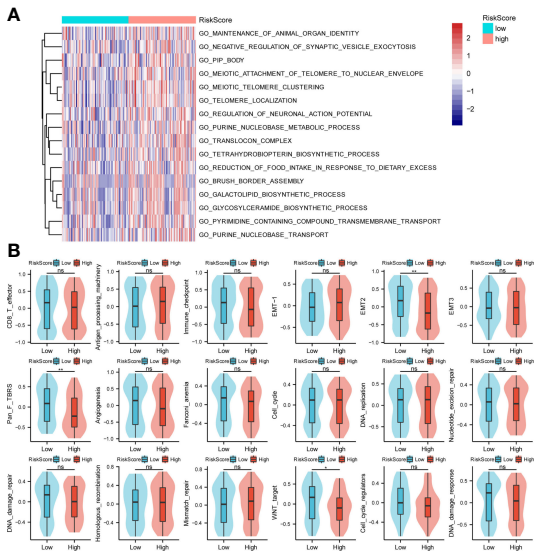


FIGURE 3
GO enrichments and gene set variation analysis. **(A, B)** Bioinformatics analysis of low- and high-risk groups. *p < 0.05, **p < 0.01, ns, not significant.

Validation of *IGFBP2 ALKBH5, FMR1*, and *RBMX* by qRT-PCR

IGFBP2, ALKBH5, FMR1, and RBMX were selected as the study subjects and validated by qRT-PCR method (Table 1). RESULTS: The expression level of IGFBP2 in normal tissues was significantly higher than that in ESCA tissues (Figure 4).

Constructing genetic traits based on the m6A risk model

To better understand the biology of phenotypes associated with the m6A risk model, genes significantly associated with m6A risk scores were analyzed using the Spearman method (Cor|>0.3 & p.adjusted<0.01). A total of 741 associated genes were identified. Subsequently, based on the expression of these genes, the unsupervised clustering method was used to classify ESCA patients into three subtypes, named GeneClusters A, B, and C. The dimensionality of the associated gene clusters was reduced using the Boruta algorithm to obtain the signature gene clusters. Based on the trend of signature gene expression, the signature genes were classified into two groups, A and B. The relationship between GeneCluster groups, m6A signature group, m6A risk score and clinical prognosis was further analyzed (Figure 5A). Meanwhile, the results of survival analysis showed significant differences in the prognosis of patients with the three

Table 1 Primers of IGFBP2, ALKBH5, FMR1, RBMX, and GAPDH.

Primer	Forward (5′ to 3′)	Reverse (5′ to 3′)
IGFBP2	TGCAGACAATGGCGATGACC	GGTGCTGCTCAGTGACCTTC
ALKBH5	GACAAGGAAGAGAACCGGCG	GCATCTTCACCTTTCGGGCA
FMR1	GCCAAAGAGGCGGCACATAA	CGCAGACTCCGAAAGTGCAT
RBMX	CCCAGCAGACGCTAAGGATG	CTACGAGAGGGCAGCGGTTC
GAPDH	GGAGCGAGATCCCTCCAAAAT	GGCTGTTGTCATACTTCTCATGG

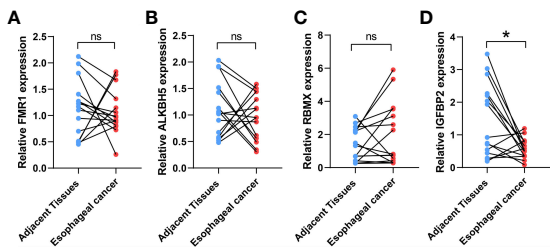


FIGURE 4
Validation of the expression levels of target *IGFBP2*. The *IGFBP2* expression in ESCA (n = 15), and adjacent normal tissues (n = 15) was evaluated by qRT-PCR including FMR1 **(A)**, AKBH5 **(B)**, RBMX **(C)**, IGFBP2 **(D)**; the results were analyzed using paired sample t test. Results are expressed as mean ± standard deviation (SD). *p< 0.05, *p< 0.05, ns, not significant.

gene subtypes, with the GeneCluster C group having the worst prognosis (log-rank p=0.003, Figure 5B).

Effect of genetic variant risk score

To better understand the effect of high and low m6A scoring methods on the level of genetic variants, this study analyzed

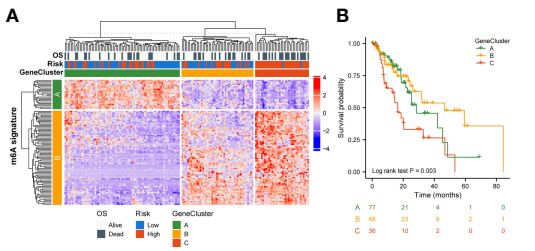


FIGURE 5
Genetic signature of risk grouping. **(A)** Relationship between GeneCluster groups, m6A signature groups, m6A risk scores, and clinical prognosis. **(B)** Kaplan–Meier curves indicate that there is a strong relationship between GeneClusters A, B, and C and the overall survival rate.

single nucleotide mutations in driver genes during common tumorigenesis, with differences in SNP levels between groups (Figure 6A). The overall level analysis showed that TMB (P=0.051, Figure 6B) was slightly correlated between the low-risk and high-risk groups, and TMB was significantly lower in GeneCluster A than in groups B and C (Figure 6C). In addition, the study of CNV variation frequency showed that the variation of gene CNV in the high-risk group was mainly focused on gene amplification (Figure 6D), while the low-risk group deleted gene copy number relatively (Figure 6D).

Construction of a clinical prediction model based on the m6A risk score

Next, to quantify OS prediction, we combined risk scores with independent clinical characteristics (T, N, M, and TNM) to construct a nomogram (Figure 7A). To verify the different

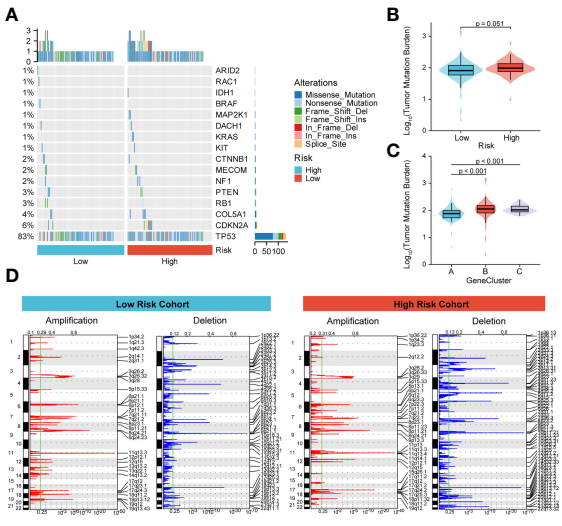


FIGURE 6
Molecular profiling of high and low m6A score groups. **(A)** Distribution of driving genes during common tumorigenesis between high and low m6A score samples; **(B, C)** Tumor mutation burden distribution in the different m6A score samples and GeneCluster groups; **(D)** Distribution of copy number amplifications and deletions in high and low m6A score samples.

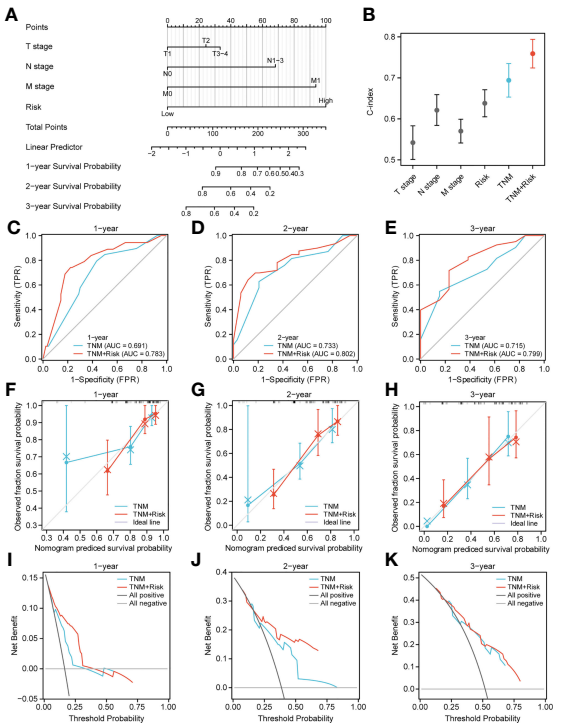


FIGURE 7
Nomogram analysis. **(A)** Nomogram composed of stage, T, N, M, and risk score for the prediction of 1-, 2-, and 3-years OS probability. **(B)** C-index analyses of T stage, N stage, M stage, risk score, TNM, TNM +risk score. **(C–E)** ROC curve for the nomogram based on TNM prognosis model and TNM +risk score model 1-, 3-, and 5-y survival. **(F–H)** Calibration plot of the TNM prognosis model and TNM +risk score model for 1-year **(F)**, 2-year **(G)**, and 3-year **(H)** OS. **(I–K)** DCA curve of TNM prognosis model and TNM +risk score for 1, 2, and 3 years.

predictive effects of m6A risk scores on the combined T, N, and M stages, this study compared the T, N, M,m6A risk scores using the scales of the TNM risk scale; and a TNM+m6A risk score model was developed (Figure 7B). The results showed that the TNM+m6A prediction model had a better predictive effect. To verify the prognosis of TNM and the predictive value of the TNM+m6A risk score model, time-dependent ROC curves were established (Figures 7C–E). the AUC of the TNM model for 1, 2, and 3-year OS were 0.691,0.733, and 0.715, respectively; the AUC of the TNM+m6A risk score model for OS 1,2, and 3-year OS were 0.783, These results suggest that the TNM+ risk score model has a higher predictive value than the TNM model. Also, calibration curves were generated to test the correctness of the models. With the calibration curves, we found that the survival curves predicted by both models at 1, 2, and 3 years were very close to the observed survival curves, indicating that the nomograms were highly predictive (Figures 7F–H). In addition, DCA showed that the TNM+m6A risk score model had a broader clinical benefit than the TNM prognostic model, but the benefit of 3-year OS was similar (Figures 7I–K).

Discussion

ESCA is a lethal malignancy. Despite advances in surgery, radiation therapy, chemotherapy and immunotherapy, the 5-year survival rate of ESCA remains low due to late detection and lack of precise treatment (21). Therefore, it is important to gain insight into the mechanisms of oncogenicity of ESCA. ESCA is closely associated with lifestyle environmental factors that can alter genomic inheritance and epigenetics (22). m6A modification is a novel regulatory mechanism of eukaryotic gene expression that controls gene expression through reversible epigenetic modifications (23).

In the present study, CNV alterations in the m6A regulator were prevalent in ESCA patients and mostly concentrated in copy number deletions. However, SNP was low in the m6A-regulator. We further demonstrated general differences and positive correlations in the expression of the 23 m6A-gene regulators in ESCA. Next, prognostic scores (high vs. low risk) were established based on the expression levels of FMR1, RBMX, IGFBP1, IGFBP2, ALKBH5, RBM15B, and METTL14 genes. These patterns could well predict the survival of ESCA patients. Initial analysis of biological functions in the high-risk versus low-risk groups was performed by GO functional annotation and genomic variation analysis. Based on this, the TCGA-ESCA cohort was clustered according to the high-risk score moderator. There were significant survival status differences among the three GeneClusters. In addition, there were some differences between the high-risk and low-risk groups according to prognostic characteristics, with CNV alterations mainly focused on gene amplification and the opposite in the low-risk group. Finally, a nomogram was constructed combining risk scores and independent clinical characteristics (T, N, M and TNM). The results showed that the TNM+m6A model was the best predictor. The calibration curves showed a significant agreement between prediction and actual survival probability.

First, 17 m6A regulators (METTL3, METTL16, WTAP, VIRMA, RBM15, YTHDC1, YTHDF2, YTHDF3, HNRNPC, FMR1, LRPPRC, HNRNPA2B1, IGFBP2, IGFBP3, RBMX, FTO, ALKBH5 gene expression differences. The expression of IGFBP2 in normal tissues was significantly higher than that in ESCA tissues as confirmed by qRT-PCR; IGFBP2 is involved in various oncogenic processes, such as epithelial-to-mesenchymal

transition, cell migration, invasion, angiogenesis, stemness, transcriptional activation, epigenetic programming, etc. A recent study showed that rs1470579 CC genotype IGFBP2 is protective against adenocarcinoma of the esophagogastric junction (24). Based on previous findings, we suggest that IGFBP2 may be a key regulator affecting the prognosis of patients with ESCA. reduced IGFBP2 gene expression may be associated with poor prognosis. Genomic mutation analysis showed that SNPs and CNVs of gene regulatory genes were associated with ESCA.

LASSO regression analysis showed that the prognostic features of ESCA patients-FMR1, RBMX, IGFBP1, IGFBP2, ALKBH5, RBM15B, METTL14-predicted the OS of ESCA patients. ALKBH5 promotes ESCC proliferation and its mechanism of action is cell cycle regulation (25). IGFBP1 has a regulatory role in cell proliferation and invasion under the regulation of miRNAs. miRNAs regulate oncogenes with regulatory cell proliferation and invasive effects. For example, miR-454-3P can act on IGFBP1 through ERK and AKT signaling, thereby inhibiting its proliferation, invasion, and apoptosis (26). However, the mechanism of action of FMR1, IGFBP1, RBM15B, and METTL14 is still unclear. We will explore their relationship with the development of ESCA in a future study. The predictive potential of these seven factors combined is much greater than that of individual factors. Although the ROC curve did not show strong predictive power in 4-5 years, the number of patients in year 4-5 was too small, which may lead to an unstable ROC curve.

In addition, this study also analyzed the GO enrichment and GSVA of ESCA patients based on risk scores using TCGA information. The high-risk group may be involved in telomere localization, base metabolism, base translocation, and mismatch repair. This may be the reason for the poor prognosis of ESCA due to risk modifiers.

On this basis, the TCGA-ESCA cohort was clustered according to the high-risk score moderator. The results of the survival analysis showed significant differences in prognosis among the three gene cluster groups, with the GeneCluster C group having the worst prognosis. Most samples had high risk scores in GeneCluster C. These results suggest that the regulators involved in the m6A risk model are significantly correlated with ESCA characteristics.

We then combined risk scores and independent clinical characteristics (T, N, M, TNM) to construct a nomogram that allowed the prediction to be quantified. The results showed that the TNM combined with the risk score model was able to predict the prognosis of ESCA patients. This improves the value of this prognostic feature for clinical application.

The study has several limitations. First, the lack of complete clinical data from TCGA may affect the results; therefore, the statistical power may not be high. Further improvement of sample size, sequencing data and clinical information is essential. Second, the mechanism of IGFBP2's effect on the prognosis of ESCA has not been explored in depth. Therefore, this study will focus on the mechanism of IGFBP2's action as a tumor suppressor. Finally, this study was based on bioinformatics and qRT-PCR techniques to analyze the results of ESCA patient tissues.

Conclusions

The above results suggest that abnormal expression of 17 m6A- RNA methylation regulators is associated with survival outcome in ESCA patients. The risk score was then combined with TNM to quantify OS prediction. This study highlights the important role of RNA modifications in the formation of ESCA and also provides potential biomarkers for the selection of therapeutic approaches.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Oncology Ethics Committee of Fujian Province, China (No. K202-027-01). The patients/participants provided their written informed consent to participate in this study.

Author contributions

YX and YH proposed and managed the study. HMW and YZ wrote the manuscript. LC and YFL collected the material and CX, DJ, QS, and HXW performed the bioinformatic analysis. LW and JC provided the samples. Data analysis consisted of YL and YC. All authors contributed to the article and approved the submitted version.

Funding

Shanghai Municipal Health Commission (#20214Y0275), National Natural Science Foundation of China (8170 2372), Shanghai Science and Technology Commission (19441904000), Shanghai Key Clinical Specialties (#shslczd zk01302), Shanghai Science and Technology Development Fund (#19MC1911000), National Natural Science Foundation of China (U21A20377), Natural Science Foundation of Fujian Province (#2020J011123).

Acknowledgments

Thanks to editage for providing thesis revision support.

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

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21609

2. Zhou Z, Lv J, Yu H, Han J, Yang X, Feng D, et al. Mechanism of RNA modification N6-methyladenosine in human cancer. *Mol Cancer* (2020) 19:104. doi: 10.1186/s12943-020-01216-3

3. Zhang X, Lu N, Wang L, Wang Y, Li M, Zhou Y, et al. Recent advances of mA methylation modification in esophageal squamous cell carcinoma. *Cancer Cell Int* (2021) 21:421. doi: 10.1186/s12935-021-02132-2

4. Guo W, Tan F, Huai Q, Wang Z, Shao F, Zhang G, et al. Comprehensive analysis of PD-L1 expression, immune infiltrates, and m6A RNA methylation regulators in esophageal squamous cell carcinoma. *Front Immunol* (2021) 12:669750. doi: 10.3389/fimmu.2021.669750

5. He L, Li H, Wu A, Peng Y, Shu G, Yin G. Functions of N6-methyladenosine and its role in cancer. *Mol Cancer.* (2019) 18:176. doi: 10.1186/s12943-019-1109-9

6. Li K, Chen J, Lou X, Li Y, Qian B, Xu D, et al. HNRNPA2B1 affects the prognosis of esophageal cancer by regulating the miR-17-92 cluster. *Front Cell Dev Biol* (2021) 9:658642. doi: 10.3389/fcell.2021.658642

7. Guo H, Wang B, Xu K, Nie L, Fu Y, Wang Z, et al. mA reader HNRNPA2B1 promotes esophageal cancer progression *via* up-regulation of ACLY and ACC1. *Front Oncol* (2020) 10:553045. doi: 10.3389/fonc.2020.553045

8. Chen P, Li S, Zhang K, Zhao R, Cui J, Zhou W, et al. N-methyladenosine demethylase ALKBH5 suppresses malignancy of esophageal cancer by regulating microRNA biogenesis and RAI1 expression. *Oncogene* (2021) 40:5600–12. doi: 10.1038/s41388-021-01966-4

9. Liu S, Huang M, Chen Z, Chen J, Chao Q, Yin X, et al. FTO promotes cell proliferation and migration in esophageal squamous cell carcinoma through up-regulation of MMP13. *Exp Cell Res* (2020) 389:111894. doi: 10.1016/j.yexcr.2020.111894

10. Hou H, Zhao H, Yu X, Cong P, Zhou Y, Jiang Y, et al. METTL3 promotes the proliferation and invasion of esophageal cancer cells partly through AKT signaling pathway. *Pathol Res Pract* (2020) 216:153087. doi: 10.1016/j.prp.2020.153087

11. Han H, Yang C, Zhang S, Cheng M, Guo S, Zhu Y, et al. METTL3-mediated mA mRNA modification promotes esophageal cancer initiation and progression *via* notch signaling pathway. *Mol Ther Nucleic Acids* (2021) 26:333–46. doi: 10.1016/j.omtn.2021.07.007

12. Li J, Li Z, Xu Y, Huang C, Shan B. METTL3 facilitates tumor progression by COL12A1/MAPK signaling pathway in esophageal squamous cell carcinoma. *J Cancer* (2022) 13:1972–84. doi: 10.7150/jca.66830

13. Xu LC, Pan JX, Pan HD. Construction and validation of an m6A RNA methylation regulators-based prognostic signature for esophageal cancer. *Cancer Manag Res* (2020) 12:5385–94. doi: 10.2147/CMAR.S254870

14. Guo W, Tan F, Huai Q, Wang Z, Shao F, Zhang G, et al. Comprehensive analysis of PD-L1 expression, immune infiltrates, and m6A RNA methylation regulators in esophageal squamous cell carcinoma. *Front Immunol* (2021) 12:669750. doi: 10.3389/fimmu.2021.669750

15. Bian S, Ni W, Zhu M, Zhang X, Qiang Y, Zhang J, et al. Flap endonuclease 1 facilitated hepatocellular carcinoma progression by enhancing USP7/MDM2-mediated P53 inactivation. *Int J Biol Sci* (2022) 18(3):1022–38. doi: 10.7150/ijbs.68179

16. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28:1747–56. doi: 10.1101/gr.239244.118

17. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

18. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004

19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170

20. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci U.S.A.* (2007) 104:20007–12. doi: 10.1073/pnas.0710052104

21. Deng W, Zhang W, Yang J, Ni W, Yu S, Li C, et al. Nomogram to predict overall survival for thoracic esophageal squamous cell carcinoma patients after radical esophagectomy. *Ann Surg Oncol* (2019) 26:2890–8. doi: 10.1245/s10434-019-07393-w

22. Liu K, Ti Z, Wang J, Chen Y, Zhang R, Lan X, et al. Etiology, cancer stem cells and potential diagnostic biomarkers for esophageal cancer. *Cancer Lett* (2019) 458:21–8. doi: 10.1016/j.canlet.2019.05.018

23. Huang H, Weng H, Chen J. mA modification in coding and non-coding RNAs: Roles and therapeutic implications in cancer. *Cancer Cell* (2020) 37:270–88. doi: 10.1016/j.ccell.2020.02.004

24. Miyake H, Hara I, Yamanaka K, Muramaki M, Gleave M, Eto H. Introduction of insulin-like growth factor binding protein-2 gene into human bladder cancer cells enhances their metastatic potential. *Oncol Rep* (2005) 13:341–5. doi: 10.1158/1078-0432.CCR-08-0408

25. Nagaki Y, Motoyama S, Yamaguchi T, Hoshizaki M, Sato T, Sato T, et al. m6A demethylase ALKBH5 promotes proliferation of esophageal squamous cell carcinoma associated with poor prognosis. *Genes Cells* (2020) 25:547–61. doi: 10.1111/gtc.12792

26. Yan A, Wang C, Zheng L, Zhou J, Zhang Y. MicroRNA-454-3p inhibits cell proliferation and invasion in esophageal cancer by targeting insulin-like growth factor 2 mRNA-binding protein 1. *Oncol Lett* (2020) 20:359. doi: 10.3892/ol.2020.12223

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fimmu.2022.983636

Check for updates

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Meng Meng,
Capital Medical University, China
Yancheng Cui,
Peking University People's Hospital, China

*CORRESPONDENCE
Shiqian Lan
120913676@qq.com
Kaihong Chen
chenkaihong1964@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 01 July 2022
ACCEPTED 02 August 2022
PUBLISHED 08 September 2022

CITATION
Zhong C, Xie T, Chen L, Zhong X, Li X, Cai X, Chen K and Lan S (2022) Immune depletion of the methylated phenotype of colon cancer is closely related to resistance to immune checkpoint inhibitors.
Front. Immunol. 13:983636.
doi: 10.3389/fimmu.2022.983636

COPYRIGHT
© 2022 Zhong, Xie, Chen, Zhong, Li, Cai, Chen and Lan. 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.

Immune depletion of the methylated phenotype of colon cancer is closely related to resistance to immune checkpoint inhibitors

Chengqian Zhong[1†], Tingjiang Xie[2†], Long Chen[2†], Xuejing Zhong[3], Xinjing Li[4], Xiumei Cai[1], Kaihong Chen[5*] and Shiqian Lan[1*]

[1]Department of Digestive Endoscopy center, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [2]Department of Gastrointestinal Surgery, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [3]Department of Science and Education, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [4]Department of Pathology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China, [5]Department of Cardiology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, China

Background: Molecular typing based on single omics data has its limitations and requires effective integration of multiple omics data for tumor typing of colorectal cancer (CRC).

Methods: Transcriptome expression, DNA methylation, somatic mutation, clinicopathological information, and copy number variation were retrieved from TCGA, UCSC Xena, cBioPortal, FireBrowse, or GEO. After pre-processing and calculating the clustering prediction index (CPI) with gap statistics, integrative clustering analysis was conducted *via* MOVICS. The tumor microenvironment (TME) was deconvolved using several algorithms such as GSVA, MCPcounter, ESTIMATE, and PCA. The metabolism-relevant pathways were extracted through ssGSEA. Differential analysis was based on limma and enrichment analysis was carried out by Enrichr. DNA methylation and transcriptome expression were integrated *via* ELMER. Finally, nearest template or hemotherapeutic sensitivity prediction was conducted using NTP or pRRophetic.

Results: Three molecular subtypes (CS1, CS2, and CS3) were recognized by integrating transcriptome, DNA methylation, and driver mutations. CRC patients in CS3 had the most favorable prognosis. A total of 90 differentially mutated genes among the three CSs were obtained, and CS3 displayed the highest tumor mutation burden (TMB), while significant instability across the entire chromosome was observed in the CS2 group. A total of 30 upregulated mRNAs served as classifiers were identified and the similar diversity in clinical outcomes of CS3 was validated in four external datasets. The heterogeneity in the TME and metabolism-related pathways were also observed in the three

CSs. Furthermore, we found CS2 tended to loss methylations while CS3 tended to gain methylations. Univariate and multivariate Cox regression revealed that the subtypes were independent prognostic factors. For the drug sensitivity analysis, we found patients in CS2 were more sensitive to ABT.263, NSC.87877, BIRB.0796, and PAC.1. By Integrating with the DNA mutation and RNA expression in CS3, we identified that SOX9, a specific marker of CS3, was higher in the tumor than tumor adjacent by IHC in the in-house cohort and public cohort.

Conclusion: The molecular subtypes based on integrated multi-omics uncovered new insights into the prognosis, mechanisms, and clinical therapeutic targets for CRC.

KEYWORDS

colorectal cancer, molecular subtype, machine learning, methylation, immunotherapy

Introduction

Colorectal cancer (CRC) is the third most common malignant tumor in the world and the fourth major cause of cancer death (1). The diagnosis of CRC is often in the middle and late stages with poor prognosis, and distant metastasis is the main cause of death in colorectal cancer patients. With the continuous improvement in medical level, comprehensive treatment measures such as surgery, radiotherapy, and chemotherapy, targeted therapy, and immunotherapy have improved the overall survival (OS) of patients with CRC, but their overall efficacy is still poor, and the 5-year survival rate of patients with metastatic CRC is only about 14% (1). Therefore, how to effectively evaluate the prognosis of different CRC patients is an urgent problem to be solved.

At present, the most widely used prognostic staging system for CRC is the TNM (Tumor, Node, and Metastasis) staging system, which is easy to observe from clinical information and is the benchmark for the establishment of clinical treatment plans for patients. However, the TNM staging system mainly relies on expert opinions, and the features used are relatively single. The abnormal phenomenon of the TNM staging system in CRC (the prognosis of patients at stage IIB/C is significantly worse than that of patients at stage III A) results in its limited ability of personalized and accurate clinical decision (2). At the same time, as a population-based system, the TNM staging system has been questioned about its application to individual patients (3). The latest eighth edition of the TNM staging system included biomarkers as new prognostic factors in some cancer staging (3). Therefore, it is necessary to introduce new prognostic factors to the existing TNM staging system in order to more accurately assess the prognosis of patients and formulate treatment plans.

Cancer is a complex disease with high heterogeneity, even patients with the same histopathological classification will have different gene mutations (4). Hence, personalized prevention, diagnosis, and treatment should be done according to the clinical and omics characteristics of different patients (5). For CRC, microsatellite instability (MSI), DNA mismatch repair (MMR), and the results of molecular tests such as RAS mutation and BRAF VE6000 are used to determine the prognosis (3, 6). It is possible to combine clinical and omics information for more personalized prognostic analysis of cancer, but it is difficult for a single omics data to fully account for all factors in a complex disease such as cancer, making it difficult for researchers to derive data from millions of single-nucleotide variations (SNV) to find the key gene that actually causes the disease (7). In recent years, more and more researchers have carried out integrated analysis of various omics data and obtained certain results (8, 9). However, most prognostic studies of CRC are limited to one set of omics, such as gene expression (10) or DNA methylation (11), and few studies that consider multiple omics data have failed to effectively combine multiple omics data with clinical data (12).

Abbreviations: CRC, Colorectal Cancer; TME, Tumor Microenvironment; TMB, Tumor Mutation Burden; OS, Overall Survival; MSI, Microsatellite Instability; MMR, Mismatch Repair; SNV, Single Nucleotide Variation; TCGA, The Cancer Genome Atlas; COAD, Colon Adenocarcinoma; READ, Rectum Adenocarcinoma; FPKM, Fragments Per Kilobase Million; CNV, Copy Number Variation; RMA, Robust Multichip Average; TPM, Transcripts Per Kilobase Million; PCA, Principal Component Analysis; CPI, Clustering Prediction Index; GSVA, Gene Set Variation Analysis; MeTIL, Methylation of Tumour-Infiltrating Lymphocyte; GSEA, Gene Set Enrichment Analysis; ssGSEA, single-sample GSEA; MSigDB, Molecular Signatures Database; DMP, Differentially Methylated Probe; DPI, Data Processing Inequality; CS, Cancer Subtype; PFS, Progression-Free Survival; CIRC, Co-ordinate Immune Response Cluster.

Therefore, how to integrate clinical data and omics data and apply them to the prognosis of CRC is of great significance.

The Cancer Genome Atlas (TCGA) is a platform that integrates clinical data, survival information, and multiple omics data for 33 cancers. Through the integration and analysis of multiple omics, cancer subtype classification, biomarker discovery, and survival prognosis analysis can be carried out (13–15). Herein, using data from TCGA and other public databases, we developed a classifier based on multi-omics integration for the prognosis prediction of CRC for the first time. We evaluated the differences in genomic heterogeneity, transcriptome biomarkers, TME landscape, metabolism-related pathways, epigenetic regulation, and chemotherapeutic drug sensitivity among the molecular subtypes of CRC. Multivariate Cox regression analysis confirmed the independent prognostic value of our subtype system. In summary, the molecular subtypes based on integrated multi-omics uncovered new insights into the prognosis, mechanisms, and clinical therapeutic targets for patients with CRC.

Materials and methods

Study population

Molecular data of patients diagnosed with CRC were retrieved from TCGA (13). Transcriptome expression profiles of the TCGA-COAD (colon adenocarcinoma) and TCGA-READ (rectum adenocarcinoma) projects quantified by the number of fragments per kilobase million (FPKM) were downloaded from the UCSC Xena (https://xenabrowser.net/), including 616 fresh-frozen samples with primary malignancy and 51 adjacent normal samples. The DNA methylation profile quantified by Illumina HumanMethylation 450K-array platform was downloaded from the UCSC Xena (https://xenabrowser.net/) under the projects of TCGA-COAD and TCGA-READ, respectively, including a total of 387 primary colorectal tumour samples and 45 adjacent normal samples. Somatic mutation data, patients' clinicopathological information, and survival data were retrieved from cBioPortal (http://www.cbioportal.org/datasets) (16). Copy number variation (CNV) data was collected from FireBrowse (http://firebrowse.org/) (17). For the purpose of multi-omics integrative clustering, 306 primary colorectal tumour samples with available transcriptome expression, DNA methylation, and somatic mutation profiles were identified for this study. Another four independent cohorts downloaded from GEO, including GSE14333 (18), GSE17538 (19), GSE38832 (20), and GSE39582 (21), comprised of a total of 1,159 CRCs with gene expression matrix and corresponding clinicopathological information. Of these external validation cohorts, gene expression matrices were profiled by Affymetrix Human Genome U133 Plus 2.0 Array. The Robust Multichip Average (RMA) algorithm was used for background correction and normalization (22).

Data pre-processing for gene expression and DNA methylation profiles

For the FPKM data of high-throughput sequencing from TCGA, Ensembl IDs for mRNAs were transformed to gene symbols by GENCODE 22. The FPKM values were transferred into transcripts per kilobase million (TPM) values, which showed more similarity to those derived from microarray and more comparable between samples (23). For microarray data retrieved from GEO database, we performed RMA normalization and processing using default settings for background correction and normalization by R package affy (24). Affymetrix probe ID was annotated with gene symbols according to the GPL570 platform. For multiple probes that mapped to one gene, mean value of expression was considered. The potential cross-dataset batch effect was removed under an empirical Bayes framework, namely, ComBat, by the R package sva (25), and the batch effect was further investigated using principal component analysis (PCA) for transcriptome profiles. For DNA methylation, we performed logit transforms β-values before ComBat adjustment and then computed the reverse logit transformation following the ComBat adjustment (26). Subsequently, we used R package ChAMP to comprehensively filter the methylation matrix. To be specific, probes with detection P value > 0.01, probes with <3 beads in at least 5% of samples per probe, all non-CpG probes, all SNP-related probes, all multi-hit probes, and probes located on sex chromosomes were removed in the first place (26, 27).

Integrative clustering based on multi-omics profiles

To perform integrative clustering analysis, we processed the TCGA multi-omics data sets to form three data matrices with columns corresponding to the common samples (n = 306) and rows corresponding to the omics features. The transcriptome expression profile was first $\log_2$ transformed. For the methylation data, we extracted probes located in promoter CpG islands, and for genes having more than one probe mapping to its promoter, the median β value was considered to identify 10,263 methylated genes. For the mutation matrix, a gene was considered mutated (entry of 1) if it contained at least one type of the following nonsynonymous variations: frameshift deletion/insertion, in-frame deletion/insertion, missense/nonsense/nonstop mutation, splice site or translation start site mutation; otherwise, 0 was used to designate wild-type status. To better fit the model and accelerate the clustering efficiency, features with flat values were removed. Specifically, we used the top 1,500 most variable mRNAs, and methylation genes according to the median absolute deviation. Additionally, 20 genes that were previously identified as driver mutations for

colorectal carcinoma were selected for cancer subtyping (28). To find an optimal clustering number, we calculated the clustering prediction index (CPI) and gap statistics using R package MOVICS (29). Consequently, integrative clustering of the TCGA cohort was conducted by R package MOVICS using a Bayesian latent variable model (29, 30).

Deconvolution of tumour microenvironment

To estimate the cell abundance of TME, we retrieved from the previous study a compendium of microenvironment genes related to specific microenvironment cell subsets, which consisted of 364 genes representing 24 microenvironment cell types (31, 32). We then used gene set variation analysis (GSVA) on these gene sets to generate enrichment scores for each cell using the R package GSVA (33). Additionally, quantification of the absolute abundance of eight immune and two stromal cell populations in heterogeneous tissues from transcriptomic data was conducted by the R package MCPcounter (34). The presence of infiltrating immune/stromal cells in the tumour tissue was estimated by the R package ESTIMATE (35). Additionally, the individual DNA methylation of tumour-infiltrating lymphocyte (MeTIL) score in the TCGA cohort was calculated using PCA according to the protocols described in the literature (36).

Single sample enrichment for metabolism-relevant pathways

The 115 metabolism-relevant gene signatures were achieved from previously published study (37), and were quantified by using single-sample GSEA (ssGSEA) approach through R package GSVA (38). Specifically, we extracted three main categories of these metabolism-relevant pathways, including carbohydrate metabolism, amino acid metabolism, and lipid metabolism.

Differential analysis and functional enrichment

Differential expression analyses were conducted using the R package "*limma*" (39). Gene set enrichment analysis (GSEA) was performed based on pre-ranked gene list according to the descending ordered $\log_2$FoldChange value derived from differential expression analysis; we then leveraged R package clusterProfiler to determine functional enrichment based on Hallmark gene set background that was retrieved from Molecular Signatures Database (MSigDB) (40, 41). The differentially methylated probes (DMPs) were obtained by R package ChAMP (26). Specifically, we considered probe to have significantly gained methylation if its corresponding mean β-value was greater than 0.3 in the specific subtype but less than 0.2 in the reference subtype with $P<0.05$ and FDR<0.05; vice versa for probes that significantly lost methylation. Gene-list based enrichment analysis was conducted by an integrative and collaborative website tool (Enrichr; https://maayanlab.cloud/Enrichr/) (42).

Cancer subtype characterization and visualization

As previously developed R package MOVICS provides powerful functions to comprehensively characterize cancer subtypes and create feature rich customizable visualizations with minimal effort, we therefore characterized the identified colorectal subtypes from multiple aspects, including survival rate, mutational frequency, fraction of copy number-altered genome (FGA), and clinical characteristics. All parameters were set to default values (29).

Integrative analysis of DNA methylation and transcriptome expression

We used R package ELMER to investigate the crosstalk between DNA methylation and transcriptome expression under an integrative analytic pipeline (43). For probes that are located in promoters, we identified putative genes that were significantly downregulated due to the hypermethylation of promoter probes. Next, the closest 20 upstream and downstream genes were collected for each probe, and for each candidate probe-gene pair, the Mann-Whitney U test was harnessed to test the null hypothesis that overall gene expression in the specific group was less than or equal to that in the reference group. For probes that are located in enhancers (distal probes that are at least 2Kb away from transcription start site on human chromosomes), hypomethylated enhancer mode with overexpressed gene expression pattern was investigated accordingly.

Nearest template prediction

Gene-expression signature-based classification was conducted using NTP algorithm, which provided a convenient model-free approach to make category prediction at single-sample level using only a list of signature genes and a test dataset, which was flexible and beneficial in external cohort application (44, 45).

Analysis of regulons

Transcriptional regulatory networks (regulons) were constructed for 71 candidate regulators associated with cancerous chromatin remodelling (46). As described in the previous study (31), potential associations between a regulator and all possible target genes were revealed by mutual information and Spearman's correlation, and associations were dropped *via* permutation analysis if the corresponding FDR was greater than 0.00001. Unstable associations were also eliminated through bootstrapping (1,000 re-samplings, consensus bootstrap>95%), and the weakest associations were removed by data processing inequality (DPI) filtering embedded in the R package RTN (47). Regulon activity scores for all samples were calculated by two-tailed GSEA.

Therapeutic response analysis

We employed R package pRRophetic to predict the chemotherapeutic sensitivity for each colorectal sample using the parameters by default (48, 49). For immunotherapy, we retrieved a published data set consisting of 47 patients with melanoma who responded to anti-CTLA4 or anti-PD1 blockades (50), and then harnessed subclass mapping to predict the clinical response to immune checkpoint blockade (51).

Immunohistochemical staining

The 50 pairs of CRC tumor and adjacent normal tissue Microarray (D216Re01) were purchased from Xi'an bioaitech Co., Ltd (Xi'an, China). Immunohistochemical staining was performed on normal and the paired tumor tissue slides. The slides were incubated with rabbit polyclonalanti-SOX9 (EPR14335, 1:2000); antibodies at 4℃ overnight. SOX9 expression was evaluated by using a system considering the staining intensity (0 means negative 1 means weak; 2 means moderate; and 3 means strong) and the percentage of positively stained cells (<5%=05% to <25%=1, 25% to 50%=2, >50 to <75%=3, >75%=4). The final score was calculated by multiplying the extent score by the intensity score.

Statistical analyses

All statistical analyses were performed by R (Version 4.0.2). We used Fisher's exact test for categorical data, Kruskal–Wallis one-way analysis of variance for continuous data, a log-rank test for Kaplan-Meier curve, and Cox regression for hazard ratio. For all comparisons, a two-sided $P < 0.05$ was considered statistically significant.

Results

Multi-omics integrative molecular subtype of colorectal cancer

We combined expression profiles of TCGA-COAD and TCGA-READ, and further removed the potential batch effect (Figure 1A). We determined the optimal cluster number of three taking into account two clustering statistics and previous molecular classifications (Figure 1B). Subsequently, integrative clustering identified three robust cancer subtypes (CSs), which were characterized by distinct molecular patterns across transcriptome mRNA expression, DNA methylation and colorectal cancerous driver mutations (Figure 1C). Of note, these classifications were not associated with major clinical features (all $P > 0.05$; Supplementary Table S1); our classification system was tightly associated with overall survival rate (OS; $P = 0.001$; Figure 1D) and progression-free survival rate (PFS; $P = 0.009$);. Generally, CS3 showed the most favourable prognosis among three clusters.

Genomic heterogeneity of colorectal cancer subtype

To investigate the genomic heterogeneity of these molecular subtypes further, we investigate the differentially mutated genes among our classifications, leading to a total of 90 genes (FDR < 0.05 and mutational frequency > 10%; Figure 2A). Among these 90 genes, 11 genes were previously identified as driver mutations in colorectal cancer, including *PIK3CA*, *APC*, *BRAF*, *KRAS*, *TP53*, *FBXW7*, *AMER1*, *TCF7L2*, *SOX9*, *ARID1A*, and *SMAD4* (Supplementary Table S2). Additionally, we found that CS3 showed a significantly higher tumour mutation burden (TMB, $P = 0.002$; Figure 2B) than the other two subtypes. We then investigated chromosomal instability by calculating the FGA scores and found that CS2 had significant instability across the entire chromosome as compared to the other two subtypes with significantly higher copy number loss or gain ($P < 0.001$; Figure 2C). We showed three types distinguishing composite copy number profiles: gistic score (Figure 2D), and percentage/frequency.

Identification of transcriptome biomarkers for colorectal cancer subtype

Given that transcriptome-level data were the most commonly used molecular profiles in cancer research, we identified 30 mRNAs with uniquely and significantly upregulated expression as classifiers for each subtype in the TCGA cohort, and a 90-gene signature was generated (Figure 3F; Supplementary Table S3). To

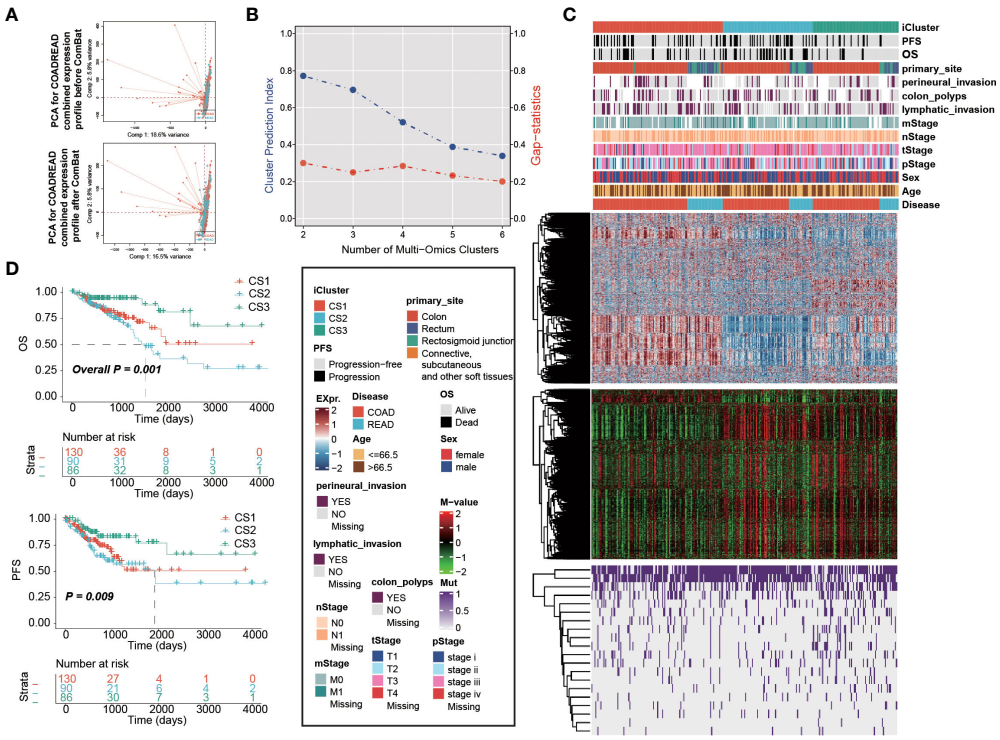


FIGURE 1
Multi-omics integrative molecular subtype of colorectal cancer. Principal component analysis to investigate the potential batch effect between TCGA-COAD and TCGA-READ. **(A)** before and after Combat. **(B)** Identification of optimal clustering number by calculating CPI and Gaps-statistics. **(C)** Comprehensive heatmap showing the molecular landscape of three cancer subtypes of colorectal carcinoma using integrative clustering. Kaplan-Meier curves of **(D)** OS and PFS with log-rank test for 306 patients with colorectal cancer according to the current molecular classification.

test the reproducibility of our identified colorectal molecular subtypes, we combined four external datasets as GEO cohort of which expression profiles were measured by microarray platform; batch effect across different datasets were removed (Figures 3A, B). We then predict the identified molecular subtypes in the GEO cohort (n = 1,159) using NTP algorithm, which classified each sample in the GEO cohort as one of the identified CS (Figure 3C). Of note, a total of 961 cases of GEO cohort were predicted with confidence (FDR < 0.05) and those cases were used for the downstream analyses. Likewise, CS3 presented with the most favourable clinical outcome out of the three subtypes ($P = 0.008$; Figures 3D, E).

Delineation of metabolism-related pathways in colorectal cancer subtype

Oncogenic heatmap with cancer associated mutations in tcga coadread (Figure 4A). Boxplot for oncogenetic pathways in iclusters of tcga coadread(Figure 4B). Similarly, GSEA is run for each subtype based on its corresponding DEA result to identify subtype-specific functional pathways (Figures 4C, D). Since Metabolic pathways regulate colorectal cancer initiation and progression, we further explored whether distinct subtypes had different metabolic characteristics in both TCGA and GEO cohort (Figure 4E). Of note, we found global dysfunction of metabolism-related pathways among three molecular subtypes, and generally CS3 showed relatively higher enrichment level of carbohydrate, amino acid, and lipid metabolism-relevant pathways, which may suggest that these colorectal cancers preserved the default metabolic program of normal colon and rectum, leading to a generally good clinical outcome.

Tumour microenvironment landscape of colorectal cancer subtype

Since cancer immunity plays a critical role in tumour progression, we suspected that the tumour microenvironment may vary a lot among these molecular subtypes. Since cancer immunity plays a critical role in tumour progression, we suspected that the tumour microenvironment may vary a lot among these molecular subtypes. Therefore, we investigated the specific immune cell infiltration status of samples in the TCGA

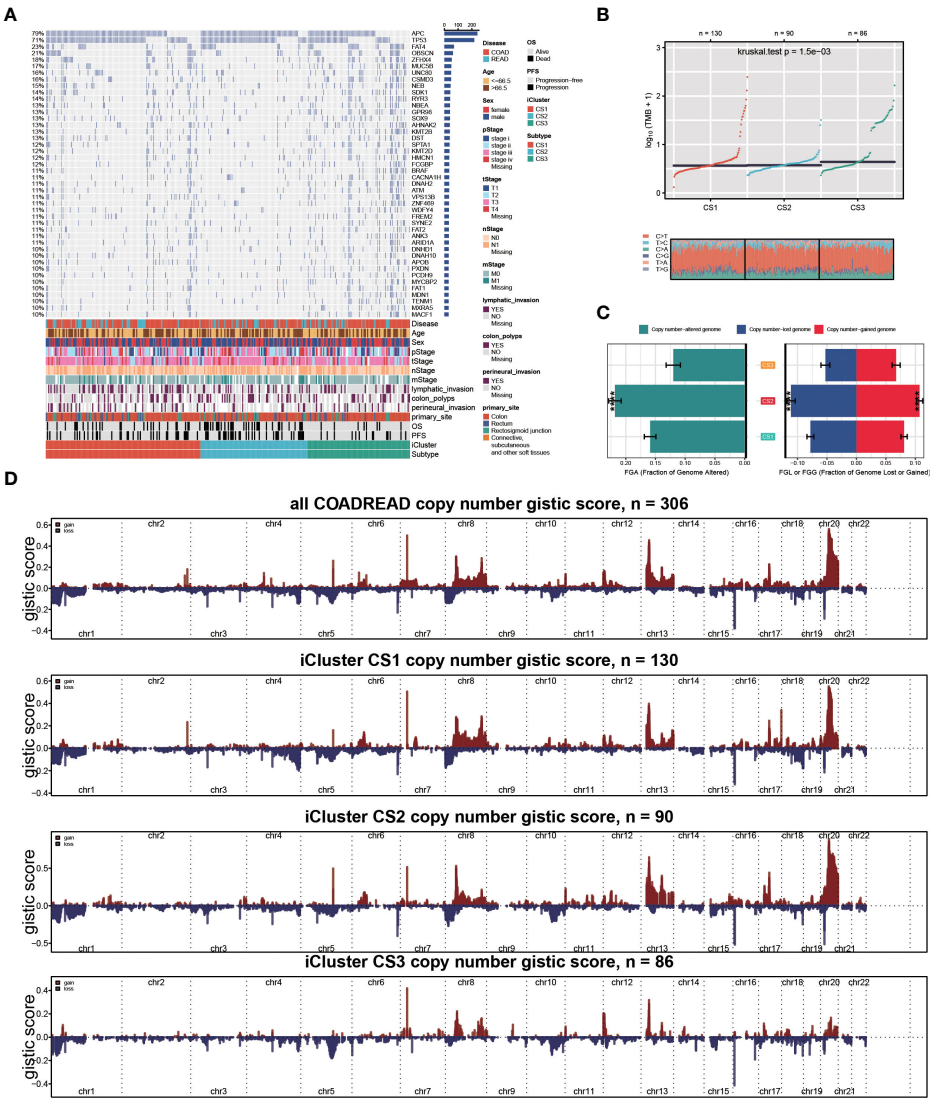


FIGURE 2
Genomic heterogeneity of colorectal cancer subtype. **(A)** OncoPrint showing the distribution of genes that were differentially mutated between three cancer subtypes. **(B)** Distribution of TMB and TiTv (transition to transversion) between two epigenetic phenotypes. **(C)** Barplot showing the distribution of FGA and fraction genome gain/loss (FGA/FGG). Bar charts are presented as the mean ± standard error of the mean. **(D)**three types distinguishing composite copy number. ****p<0.0001.

cohort. To be specific, we quantified the infiltration levels of several microenvironment cell types using different approach, and surveyed the colorectal samples for the expression of genes representing immune checkpoint targets. The analysis of gene expression signatures suggested that CS1 was highly immune-infiltrated, CS3 showed relatively higher immunocyte infiltration, while CS2 was generally immune-depleted (Figure 5A). This finding may converge to the poor overall survival of CS2 versus other molecular subtypes. Compared to the other subtypes, CS1 had relatively higher expression of several genes that represent potential targets for immunotherapy, including CD274 (PDL1), PDCD1 (PD1), CD247 (CD3), PDCD1LG2 (PDL2), CTLA4 (CD152), TNFRSF9 (CD137), TNFRSF4 (CD134) and TLR9 (Sup_S2). Interestingly, CS1 enriched for B cell, CD8 T cells but may lack CD4 memory activated cells (Sup_S2); previous study showed the ratio of CD4/CD8 may play prognostic role in several cancer subtypes (52, 53). Additionally, we found that interferon-γ pathway was significantly activated in CS1 (FDR < 0.001; Figure 5B), which made us hypothesized that CS1 may be beneficial from immune checkpoint inhibitors. In this manner, we performed subclass mapping of TCGA cohort and revealed that only the CS1 showed high transcriptome-level similarity to a group of patients with melanoma who responded to anti-

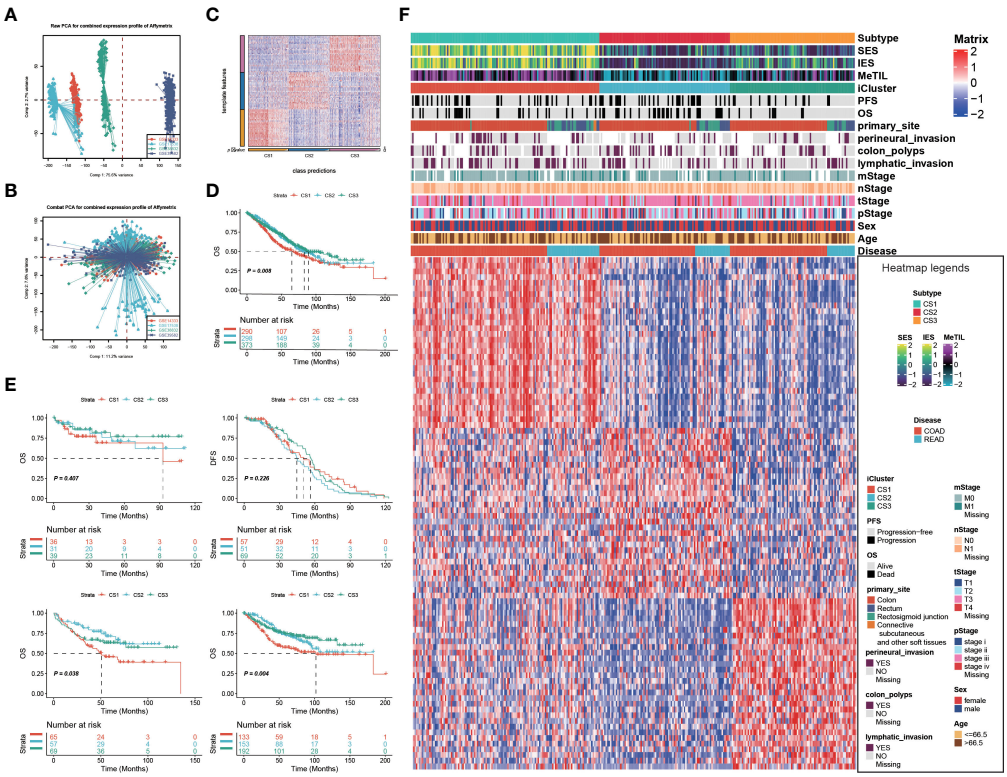


FIGURE 3
Identification of transcriptome biomarkers for colorectal cancer subtype. Principal component analysis to investigate the potential batch effect among four GEO datasets **(A)** before and **(B)** after Combat. **(C)** Heatmap showing the transcriptome expression pattern of the 120-gene signature in nearest template predicted cancer subtype of GEO cohort. **(D)** Kaplan-Meier curves of OS with log-rank test for 961 patients with colorectal cancer according to the eligible predicted classification. **(E)** KM of os using ntp in GEO **(F)** Heatmap showing the transcriptome expression pattern of the 90-gene signature (30 uniquely significantly upregulated genes in each cancer subtype) in TCGA cohort.

CTLA4 or anti-PD1 blockades ($P < 0.05$, adjusted $P \leq 0.25$; Figure 5C), which indicated that the current classification may be useful to identify ideal candidates of patients with colorectal cancer for immunotherapy. The tumour microenvironment landscape was generally validated in GEO cohort. Consistently, CS1 in GEO cohort significantly activated interferon-γ pathway, and showed higher likelihood of responding to immune checkpoint inhibitors.

Epigenetic regulation in colorectal cancer subtype

Given the different transcription profiles among the three CRC subtypes, we then asked if this could mirror the epigenetic aspect. To this end, we identified differentially methylated probes for each subtype, and we found CS2 tended to loss methylations (n = 240) as compared to other subtypes (Supplementary Table S4). Notably, these probes losing DNA methylation were significantly enriched in enhancers compared to the 450K array background (P<0.001; Figure 6A). As to CS3, we found this subtype tended to gain methylations (n=249) compared to other subtypes (Supplementary Table S5), and those probes gaining methylation significantly enriched in promoter CpG islands (P<0.001; Figure 6B). To further investigate the crosstalk between epigenetic DNA methylation and transcriptome expression, we performed integrative analysis combining both gene expression and DNA methylation profiles using ELMER pipeline. First, for CS2, we identified distal probes that are 2Kb away from the transcription start site of the human chromosome, and performed differential methylation analysis at probe level to identify probes with difference of β-value greater than 0.1 (FDR<0.05) in CS2 compared to other subtypes, ending up with a total of 3,683 distal probes/enhancers (Supplementary Table S6). Next, ELMER searched for the nearby 20 genes corresponding to these probes, and further predicted enhancer-gene linkages using associations between DNA methylation at enhancers and expression of 20 nearby genes of the CpG sites; such analysis identified a total of 2,533 pairs corresponding to 1,003 genes (Figure 6C; Supplementary Table S7). To understand the biologic relevance of these genes that were epigenetically activated, we harnessed Enrichr and found

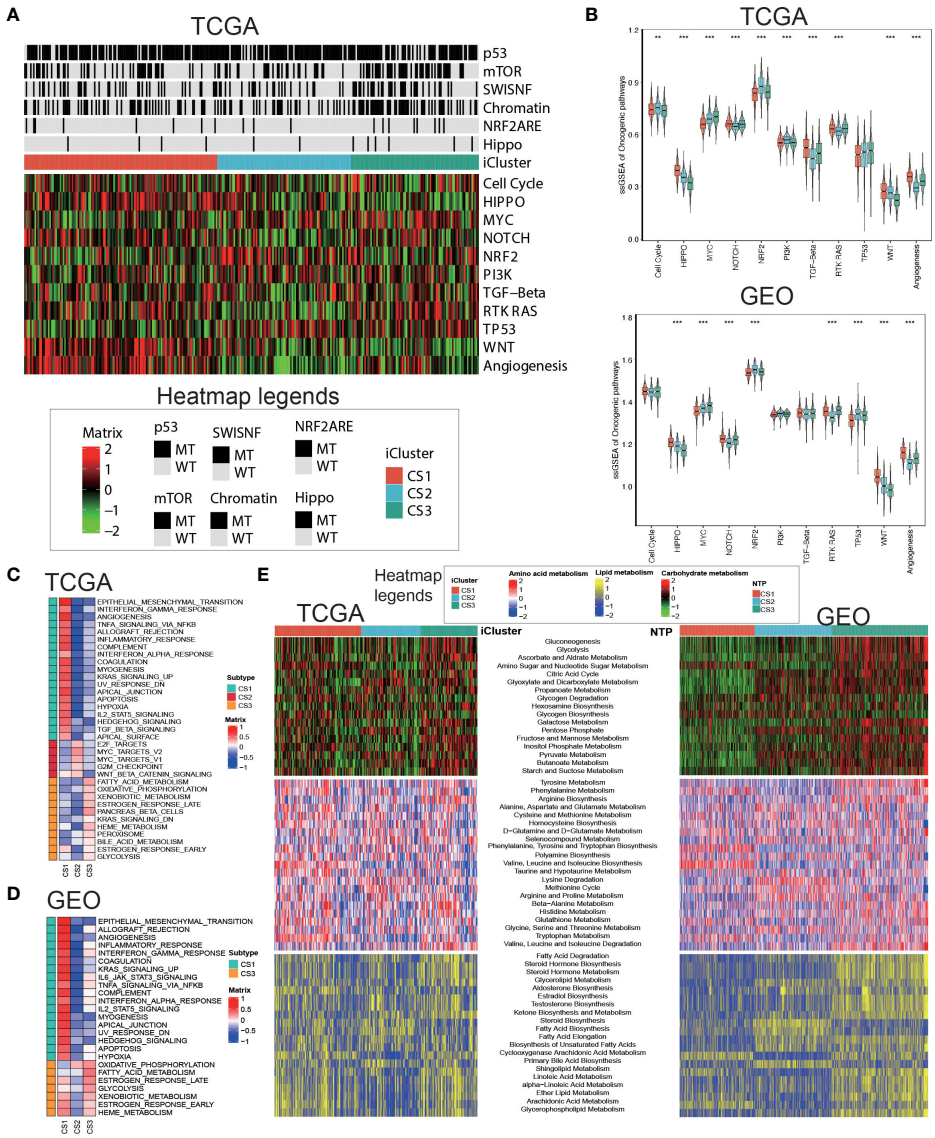


FIGURE 4
Delineation of metabolism-related pathways in colorectal cancer subtype. **(A)** Oncogenic heatmap with cancer associated mutations in tcga coadread. **(B)** Boxplot for oncogenetic pathways in iclusters of tcga coadread. **(C)** Upregulated hallmark pathway heatmap in tcga_using_upregulated_pathways. **(D)** Upregulated hallmark pathway heatmap in geo_using_upregulated_pathways. **(E)** Heatmap showing transcriptome enrichment score of three metabolic categories in TCGA and GEO cohorts. **p < 0.01; ***p < 0.001.

that these genes were significantly enriched in MYC Hallmark pathways (P =0.006, FDR=0.24; Supplementary Table S8). Previous study demonstrated that MYC oncogene was associated the suppression in tumour immunity (54), which suggest that the downregulation of MYC oncogenic pathway may contribute shaping the immune-depleted tumour microenvironment of CS2. Using the similar strategy, we investigated CS3, but we searched for promoter-gene pairs that showed epigenetically silencing mode. In this manner, ELMER identified a total of 1,063 promoters that gained methylation in CS3 versus other subtypes (Supplementary Table S9), and a total of 3,212 promoter-gene pairs were identified to be epigenetically silenced in CS3 (Figure 6D; Supplementary Table S10). Enrichr revealed that these genes are significantly enriched in epithelia-mesenchymal transition (EMT) hallmark pathway (P<0.001, FDR<0.001; Supplementary Table S11). Down-regulation of EMT may decrease tumour-initiating and metastatic potential of cancer cells (55), which lead to good prognosis of CS3. In addition, activity profiles of regulons associated with chromatin remodelling highlighted additional potential regulatory differences among three colorectal cancer subtypes, indicating that epigenetically driven transcriptional networks might be

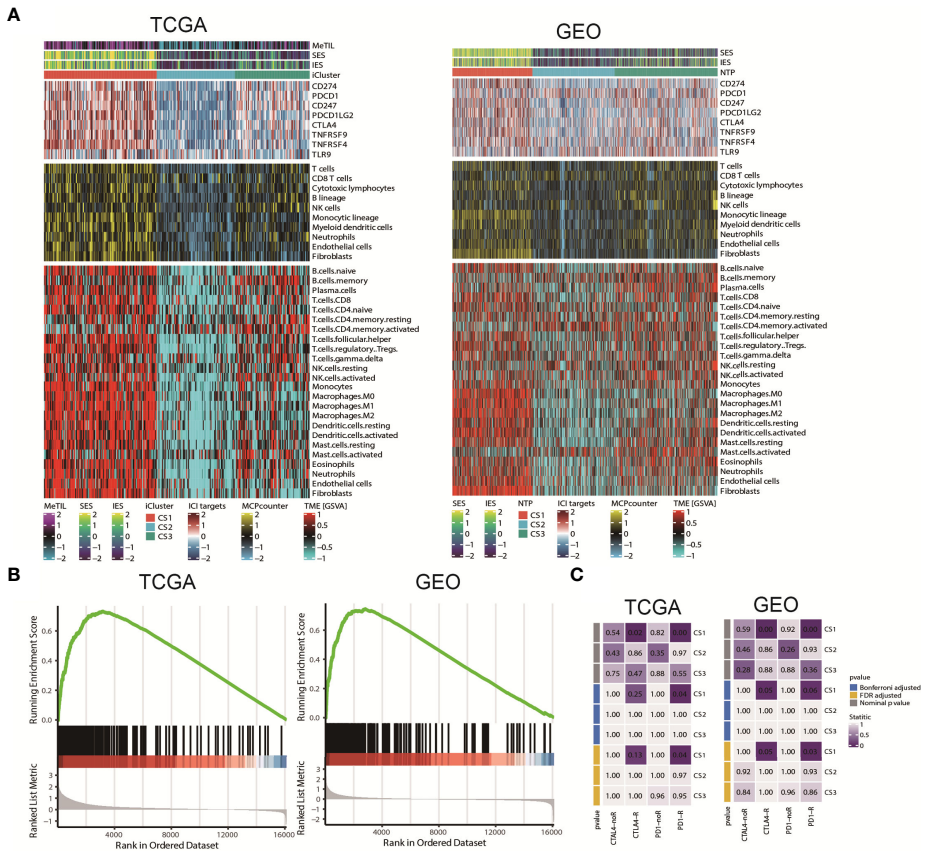


FIGURE 5
Tumour microenvironment landscape of colorectal cancer subtype. **(A)** Heatmap showing the immune profile in the TCGA and GEO cohort, with the top panel showing the expression of genes involved in immune checkpoint targets, the middle panel showing the enrichment level of 10 microenvironment cell types using MCPcounter approach, and the bottom panel showing the 24 microenvironment cells using GSVA approach; DNA methylation of tumour-infiltrating lymphocytes (MeTILs) were annotated at the top of the heatmap. The immune enrichment score and stromal enrichment score were annotated at the top of the heatmap. **(B)** GSEA plot showing activation of interferon-γ hallmark pathway. **(C)** Subclass analysis manifested that CS1 subtypes could be more sensitive to the immune checkpoint inhibitors.

important differentiators of these molecular subtypes (Figures 6E, F).

Independent prognostic value of colorectal cancer subtype

We then surveyed that whether the current classification was an independent prognostic factor in colorectal cancers from TCGA cohort. As the generally favourable prognosis of CS3, we therefore considered the CS3 as the non-aggressive subtype while patients belonged to CS1 or CS2 were aggressive in clinical setting. In this manner, univariate Cox regression model was first conducted to filter out prognostic clinical characterizations concerning OS and PFS; multivariate Cox regression was subsequently performed based on those prognosis-relevant features. Using such strategy, we found that the current classification remained the independent prognostic factor after adjusting clinical prognostic features with respect to OS ($P = 0.026$) and PFS ($P = 0.032$) (Figures 7A, B).

Potential therapeutic strategy for colorectal cancer subtype

Considering the significantly poor clinical outcome of CS2 in colorectal cancer, we decided to infer potential anticancer agents that may show clinical efficiency for patients belonging to CS2 through an *in-sillico* drug screening approach. To this end, we constructed ridge regression model between cell lines and corresponding drug sensitivity and applied the predictive model to each of the colorectal cases in both TCGA and GEO cohorts (Supplementary Tables 12, 13). A total of four drugs were discovered to be potentially effective in treating patients with CS2 phenotype as compared to other cases, including ABT.263, NSC.87877, BIRB.0796, and PAC.1 (all, $P < 0.01$; Figures 7C, D).

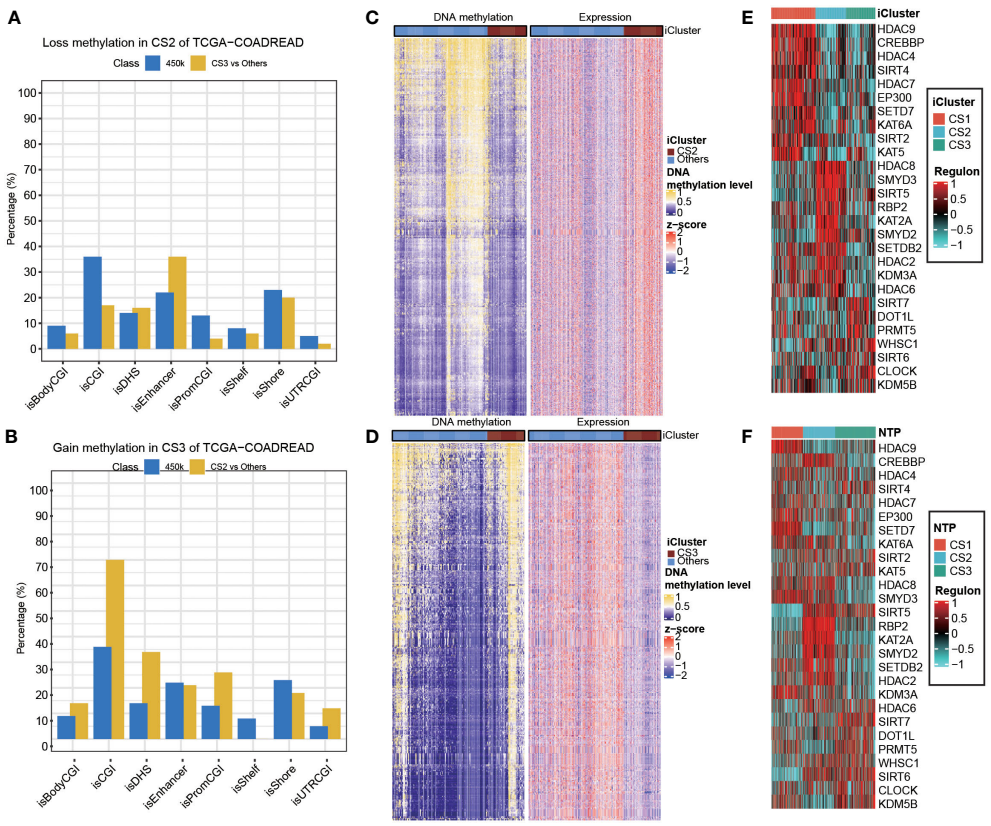


FIGURE 6
Epigenetic regulation in colorectal cancer subtype. Heatmap showing activity of regulon relevant to potential regulators associated with chromatin remodelling in both **(A)** TCGA and **(B)** GEO cohorts. Heatmap showing the association between DNA methylation and gene expression, presenting with **(C)** an epigenetic activation pattern in CS2 and **(D)** an epigenetically silencing pattern in CS3 of TCGA cohort. Barplots showing the region-specific distribution of DMPs comparing to the Illumina 450karray background for the **(E)** CS2 and **(F)** CS3 molecular classification in TCGA cohort.

Identified a biomarker for multi-omics molecular subtype

To apply the molecular subtype better in the clinic, we identified a biomarker for our molecular subtype which most based on the DNA mutation and RNA expression among different subtypes. By using the Chi-square test for DNA mutation and fold change with adjust FDR value for RNA expression, then, we detected SOX9 was a significant gene in the CS3 subtype. Through the IHC experiment, we found SOX9 was higher in the 50 tumor tissue than the 50 tumor adjacent tissue. IHC showed the represent sample in adjacent and tumor samples (Figures 8A, B). TCGA-COAD public cohort also confirmed that SOX9 was higher in tumor tissue than that in adjacent tissue (Figures 8C, D). SOX9 mainly located in the nucleoplasm of cell in A-431, U-2 OS and U-251 MG multi cell lines by immunofluorescence with HPA001758 antibody in the Human Protein Atlas (HPA) (Figure 8E).

Discussion

The high incidence and mortality of CRC have brought a huge burden on patients. How to effectively judge the prognosis of CRC patients and correctly evaluate the severity of the disease of CRC patients are the main objectives of the study on the prognosis of CRC. The prognosis of patients based on traditional tumor typing is often very different. Molecular typing of tumors can better reflect the differences in internal molecular characteristics of tumors, which is the basis for the realization of precision medicine. Accurate identification of patients' molecular subtypes will help to accurately predict patient prognosis and develop personalized treatment plans.

Currently, tumor molecular subtype studies are mainly based on single omics data, such as transcriptomics, proteomics, genomics, etc (56–61). Bhattacharjee et al. divided lung adenocarcinoma into 4 subtypes by analyzing gene expression profile data from lung adenocarcinoma patients, and found that abnormal expression profile can be used to

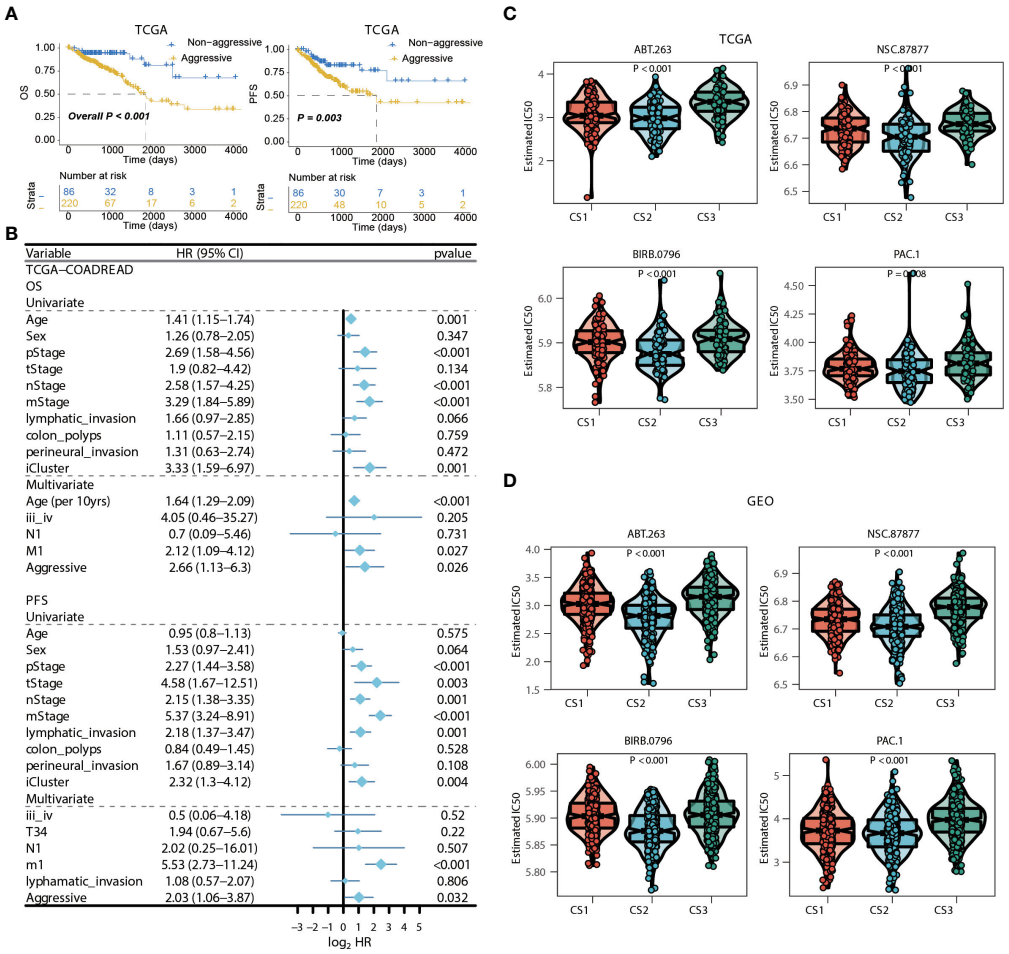


FIGURE 7
Identification of transcriptome biomarkers for colorectal cancer subtype. **(A)** KM of os using movics agreesiveness in coadread of tcga. **(B)** Forest plot showing the hazard ratio (95% CI) in univariate and multivariate Cox regressions with the corresponding *P* values. **(C, D)** Boxplot showing the distribution of estimated IC50 among three cancer subtypes based on GDSC database, **(C)** TCGA, **(D)** GEO.

distinguish primary and metastatic adenocarcinoma of lung (58). Based on genomic CNV data, Shibata et al. divided lung adenocarcinoma into three subtypes by unsupervised clustering analysis, and found that patients with EGFR mutations had shorter disease-free survival times (60). As for CRC, Roepman et al. conducted unsupervised classification of genome-wide data of CRC patients based on EMT, microsatellite instability caused by mismatch repair gene defects, and high mutation frequency associated with cell proliferation (62). Meanwhile, Lai et al. proposed the co-ordinate immune response cluster (CIRC), and identified four patient groups by this method (63). Zhang et al. identified two molecular subtypes, C1 and C2, based on cell cycle-related genes. PIK3CA, RYR2 and FBXW7 mutations were more frequent in C1, and the clinical characteristics and prognosis of patients were relatively poor (64). In addition to the above genotyping based on gene mutations and cytogenetic changes in the genome (10, 21, 65–68), CRC was also classified based on differences in gene expression profiles and proteomic (69–72)biomarkers. Therefore, molecular typing based on omics data can effectively identify clinically relevant tumor subtypes, which plays a very important role in judging patient prognosis and guiding clinical treatment.

Nevertheless, any single omics data can only reflect the intrinsic molecular characteristics of tumors from a single perspective, and the contribution of single-omics analysis to tumor typing is one-sided. Therefore, the integration of multi-omics information can simultaneously capture the heterogeneity of tumors in different omics and integrate the information from multiple perspectives to identify more accurate tumor molecular typing. As the high heterogeneity of tumors is determined by multiple omics, such as genome, epigenome, transcriptome, and proteome, the analysis of data from different omics sources is expected to better reveal the mechanism of tumor genesis and development. For the first time, Matan Hofree et al. integrated genomic mutations and protein interaction networks for molecular typing of tumors to identify subtypes significantly

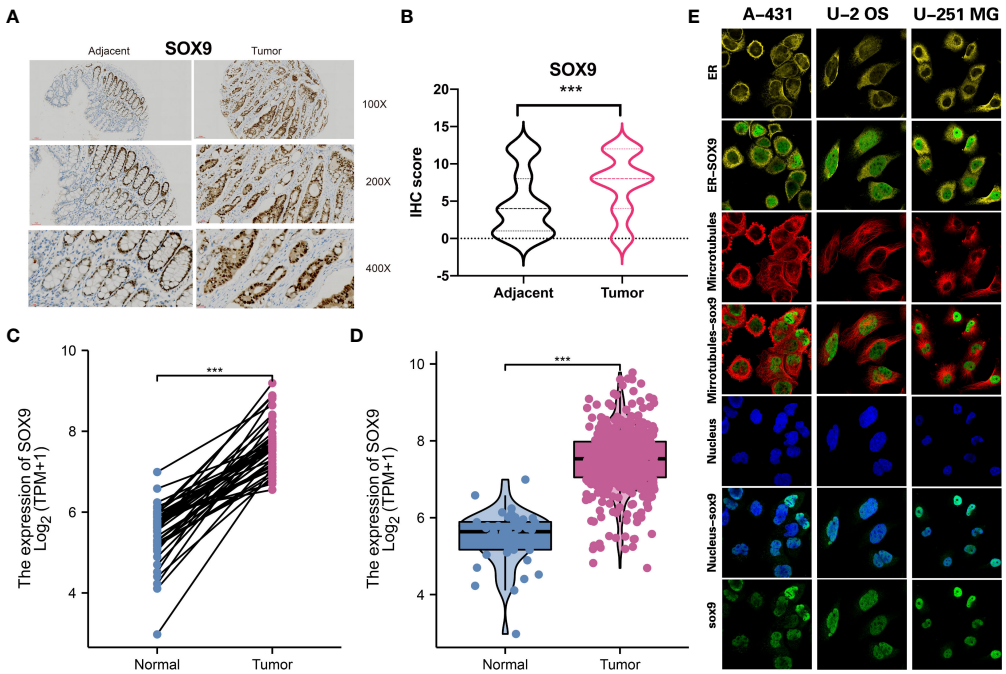


FIGURE 8
Molecular subtype biomarkers validated by wet experiment. **(A)** SOX9 protein expression of the represent sample in adjacent and tumor samples by IHC. **(B)** Pair-test for SOX9 protein expression between 50 tumor tissue and 50 tumor adjacent tissue by IHC. **(C)** and **(D)** SOX9 gene expression between tumor tissue and adjacent tissue in TCGA-COAD public cohort. **(E)** The location of SOX9 in A-431, U-2 OS and U-251 MG in the Human Protein Atlas(HPA). ***p<0.001.

associated with clinical features (73). Ronglai Shen et al. integrated genomic mutations, CNV and transcriptome expression profiles to obtain tumor classification based on iCluster (74). Herein, using transcriptome, DNA methylation, and driver mutations of CRC, we developed a classifier based on multi-omics integration for the prognosis prediction of CRC for the first time. At present, many studies have proved that CRC is the result of accumulation of multiple gene mutations and epigenetic modifications, and DNA hypermethylation or hypomethylation can be used as epigenetic biomarkers to predict the occurrence and prognostic effects of CRC (75–77). Driver mutations in the genome can be viewed as responsible for molecular changes associated with CRC progression, so targeting such genes for the elimination of multiple CRC gene dependencies could significantly improve efficacy (78). In conclusion, CRC can be comprehensively understood from multiple omics based on transcriptome, DNA methylation, and driver mutation levels to predict prognosis and guide clinical medication.

In the medical field, prognostic models need to undergo extensive and rigorous validation before they can be used in practice, and they also need to be continuously evaluated by feedback. At present, due to the different data standards and coding systems used by different sources, the output platforms and schemes of omics data also have certain heterogeneity. Therefore, the current integrated prognostic models are often internally verified by resampling or cross-validation. The few externally validated integrated prognostic models often involve only one type of omics data and have been externally validated in only a few open data sets, making it difficult for the current integrated prognostic models to be applied in clinical practice. In order to verify the reproducibility of the colorectal molecular subtypes we identified, we combined four external datasets from GEO cohort. We removed batch effects across different datasets and predicted the identified molecular subtypes in the GEO cohort using NTP algorithm. CS3 presented with the most favourable clinical outcome out of the three subtypes, indicating the accuracy of the subtype system.

Beyond that, there are several new findings and notable advantages to our study. TME and tumor cells interact and co-evolve to drive tumor growth and progression, and also play an important role in regulating tumor sensitivity to treatment (79). The results showed that CS1 was highly immune-infiltrated, CS3 showed relatively higher immunocyte infiltration, while CS2 was generally immune-depleted, explaining the difference in prognosis. Immunotherapy is an important treatment for CRC. We compared the responses of the three subtypes to immune checkpoint inhibitors. Abnormal metabolism is

closely related to the occurrence, development, recurrence, metastasis, and prognosis of CRC. We found that the enrichment level of carbohydrate, amino acid, and lipid metabolism-relevant pathways in CS3 was higher. Our results showed that CS2 tended to loss methylations while CS3 tended to gain methylations. ABT.263 is a small molecule Bcl-2 inhibitor that can induce cell apoptosis (80). BIRB.0796 is one of the most potent compounds of (81) p38 inhibitors. PAC.1 (Caspase activator) is an effective procaspase-3 activator, which acts on primary cancer cells and induces apoptosis (82). Our findings showed that ABT.263, NSC.87877, BIRB.0796, and PAC.1 were discovered to be potentially effective in treating patients with CS2 phenotype.

Nonetheless, some limitations of the current study should not be ignored. Hence, the cases of CRC patients were relatively small; more cases are needed to confirm our conclusions. The molecular subtypes of CRC were based on retrospective cohorts. Therefore, prospective studies are needed in the future. Even though we developed molecular subtypes based on integrated multi-omics, the metabolomics and proteomics data were missing because the relevant omics information was not available in the TCGA database. With the development of information technology and genetic testing technology, more and more clinical data in the form of accessible electronic medical records and shared omics data are available. The rapid development of artificial intelligence technology can further mine the correlation and interaction between different scales of data and more effectively use different scales of data for information complementarity to achieve a more accurate prediction model. Therefore, it is of great significance to further improve and optimize the multi-omics analysis based on this study, realize the multi-center collaborative multi-omics integrated analysis, and apply it to the prognostic analysis of CRC.

Conclusion

Taken together, we carried out multi-omics analysis of transcriptome mRNA expression, DNA methylation, and colorectal cancerous driver mutations. Three molecular subtypes were constructed and clinical significances, such as prognosis, mechanisms, and clinical therapeutic targets were observed among them. Besides, the subtypes were independent prognostic factors.

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, TX, LC, XZ, XL, KC, and XC designed the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Sponsored by Longyan City Science and Technology Plan Project (Grant number: 2020LYF17029).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.983636/full#supplementary-material

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551

2. Li J, Guo BC, Sun LR, Wang JW, Fu XH, Zhang SZ, et al. TNM staging of colorectal cancer should be reconsidered by T stage weighting. *World J Gastroenterol* (2014) 20:5104–12. doi: 10.3748/wjg.v20.i17.5104

3. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388

4. Roswall P, Bocci M, Bartoschek M, Li H, Kristiansen G, Jansson S, et al. Microenvironmental control of breast cancer subtype elicited through paracrine platelet-derived growth factor-CC signaling. *Nat Med* (2018) 24:463–73. doi: 10.1038/nm.4494

5. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LJ, Kinzler KW. Cancer genome landscapes. *SCIENCE* (2013) 339:1546–58. doi: 10.1126/science.1235122

6. Weiser MR. AJCC 8th edition: Colorectal cancer. *Ann Surg Oncol* (2018) 25:1454–5. doi: 10.1245/s10434-018-6462-1

7. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. *NATURE* (2010) 464:993–8. doi: 10.1038/nature08987

8. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* (2017) 18:83. doi: 10.1186/s13059-017-1215-1

9. Ideker T, Dutkowski J, Hood L. Boosting signal-to-noise in complex biology: prior knowledge is power. *CELL* (2011) 144:860–3. doi: 10.1016/j.cell.2011.03.007

10. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* (2015) 21:1350–6. doi: 10.1038/nm.3967

11. Kang KJ, Min BH, Ryu KJ, Kim KM, Chang DK, Kim JJ, et al. The role of the CpG island methylator phenotype on survival outcome in colon cancer. *GUT LIVER* (2015) 9:202–7. doi: 10.5009/gnl13352

12. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* (2014) 11:333–7. doi: 10.1038/nmeth.2810

13. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al . Comprehensive molecular characterization of human colon and rectal cancer. *NATURE* (2012) 487:330–7. doi: 10.1038/nature11252

14. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Veizer JK, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *NATURE* (2012) 490:61–70. doi: 10.1038/nature11412

15. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *CELL* (2014) 158:929–44. doi: 10.1016/j.cell.2014.06.049

16. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6:l1. doi: 10.1126/scisignal.2004088

17. Chabanais J, Labrousse F, Chaunavel A, Germot A, Maftah A. POFUT1 as a promising novel biomarker of colorectal cancer. *Cancers (Basel)* (2018) 10(11):411. doi: 10.dio: 10.3390/cancers10110411

18. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage b and c colorectal cancer. *Clin Cancer Res* (2009) 15:7642–51. doi: 10.1158/1078-0432.CCR-09-1431

19. Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL, et al. Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *GASTROENTEROLOGY* (2012) 142:562–71. doi: 10.1053/j.gastro.2011.11.026

20. Tripathi MK, Deane NG, Zhu J, An H, Mima S, Wang X, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res* (2014) 74:6947–57. doi: 10.1158/0008-5472.CAN-14-1592

21. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PloS Med* (2013) 10:e1001453. doi: 10.1371/journal.pmed.1001453

22. Katz S, Irizarry RA, Lin X, Tripputi M, Porter MW. A summarization approach for affymetrix GeneChip data using a reference training set from a large, biologically diverse database. *BMC Bioinf* (2006) 7:464. doi: 10.1186/1471-2105-7-464

23. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* (2012) 131:281–5. doi: 10.1007/s12064-012-0162-3

24. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of affymetrix GeneChip data at the probe level. *BIOINFORMATICS* (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405

25. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *BIOINFORMATICS* (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

26. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for illumina BeadChips. *BIOINFORMATICS* (2017) 33:3982–4. doi: 10.1093/bioinformatics/btx513

27. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of infinium DNA methylation BeadChip probes. *Nucleic Acids Res* (2017) 45:e22. doi: 10.1093/nar/gkw967

28. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *CELL* (2018) 173:371–85. doi: 10.1016/j.cell.2018.02.060

29. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: an r package for multi-omics integration and visualization in cancer subtyping. *BIOINFORMATICS* (2020) 14:btaa1018. doi: 10.1093/bioinformatics/btaa1018

30. Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *BIOSTATISTICS* (2018) 19:71–86. doi: 10.1093/biostatistics/kxx017

31. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *LANCET* (2021) 396:1979–93. doi: 10.1016/S0140-6736(20)32466-1

32. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. *Clin Cancer Res* (2019) 25:5002–14. doi: 10.1158/1078-0432.CCR-18-3524

33. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

34. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17:218. doi: 10.1186/s13059-016-1070-5

35. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

36. Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA Methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest* (2017) 127:3090–102. doi: 10.1172/JCI91095

37. Rosario SR, Long MD, Affronti HC, Rowsam AM, Eng KH, Smiraglia DJ. Pan-cancer analysis of transcriptional metabolic dysregulation using the cancer genome atlas. *Nat Commun* (2018) 9:5330. doi: 10.1038/s41467-018-07232-8

38. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *NATURE* (2009) 462:108–12. doi: 10.1038/nature08460

39. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007

40. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

41. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004

42. Xie Z, Bailey A, Kuleshov MV, Clarke D, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. *Curr Protoc* (2021) 1:e90. doi: 10.1002/cpz1.90

43. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *BIOINFORMATICS* (2019) 35:1974–7. doi: 10.1093/bioinformatics/bty902

44. Hoshida Y. Nearest template prediction: A single-sample-based flexible class prediction with confidence assessment. *PloS One* (2010) 5:e15543. doi: 10.1371/journal.pone.0015543

45. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an r package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci REP-UK* (2017) 7:16618. doi: 10.1038/s41598-017-16747-x

46. Audia JE, Campbell RM. Histone modifications and cancer. *Cold Spring Harb Perspect Biol* (2016) 8:a19521. doi: 10.1101/cshperspect.a019521

47. Castro MA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat Genet* (2016) 48:12–21. doi: 10.1038/ng.3458

48. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47

49. Geeleher P, Cox N, Huang RS. pRRophetic: an r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468

50. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *SCIENCE* (2016) 351:1463–9. doi: 10.1126/science.aaf1490

51. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia (New York N.Y.)* (2019) 21:591–601. doi: 10.1016/j.neo.2019.04.003

52. Matkowski R, Gisterek I, Halon A, Lacko A, Szewczyk K, Staszek U, et al. The prognostic role of tumor-infiltrating CD4 and CD8 T lymphocytes in breast cancer. *Anticancer Res* (2009) 29:2445–51.

53. Shah W, Yan X, Jing L, Zhou Y, Chen H, Wang Y. A reversed CD4/CD8 ratio of tumor-infiltrating lymphocytes and a high percentage of CD4(+)FOXP3(+) regulatory T cells are significantly associated with clinical outcome in squamous cell carcinoma of the cervix. *Cell Mol Immunol* (2011) 8:59–66. doi: 10.1038/cmi.2010.56

54. Wu X, Nelson M, Basu M, Srinivasan P, Lazarski C, Zhang P, et al. MYC oncogene is associated with suppression of tumor immunity and targeting myc induces tumor cell immunogenicity for therapeutic whole cell vaccination. *J Immunother Cancer* (2021) 9(3):e001388. doi: 10.1136/jitc-2020-001388

55. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol* (2019) 20:69–84. doi: 10.1038/s41580-018-0080-4

56. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *P Natl Acad Sci USA* (2003) 100:8418–23. doi: 10.1073/pnas.0932692100

57. Sun Y, Zhang W, Chen D, Lv Y, Zheng J, Lilljebjorn H, et al. A glioma classification scheme based on coexpression modules of EGFR and PDGFRA. *Proc Natl Acad Sci U.S.A.* (2014) 111:3538–43. doi: 10.1073/pnas.1313814111

58. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U.S.A.* (2001) 98:13790–5. doi: 10.1073/pnas.191502998

59. Pena-Llopis S, Vega-Rubin-de-Celis S, Liao A, Leng N, Pavia-Jimenez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* (2012) 44:751–9. doi: 10.1038/ng.2323

60. Shibata T, Uryu S, Kokubu A, Hosoda F, Ohki M, Sakiyama T, et al. Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: its association with clinicopathologic features. *Clin Cancer Res* (2005) 11:6177–85. doi: 10.1158/1078-0432.CCR-05-0293

61. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *NATURE* (2000) 406:747–52. doi: 10.1038/35021093

62. Roepman P, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int J Cancer* (2014) 134:552–62. doi: 10.1002/ijc.28387

63. Lal N, Beggs AD, Willcox BE, Middleton GW. An immunogenomic stratification of colorectal cancer: Implications for development of targeted immunotherapy. *ONCOIMMUNOLOGY* (2015) 4:e976052. doi: 10.4161/2162402X.2014.976052

64. Zhang Z, Ji M, Li J, Wu Q, Huang Y, He G, et al. Molecular classification based on prognostic and cell cycle-associated genes in patients with colon cancer. *Front Oncol* (2021) 11:636591. doi: 10.3389/fonc.2021.636591

65. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* (2013) 231:63–76. doi: 10.1002/path.4212

66. Schlicker A, Beran G, Chresta CM, McWalter G, Pritchard A, Weston S, et al. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med Genomics* (2012) 5:66. doi: 10.1186/1755-8794-5-66

67. Zhang J, Wang N, Wu J, Gao X, Zhao H, Liu Z, et al. 5-methylcytosine related LncRNAs reveal immune characteristics, predict prognosis and oncology treatment outcome in lower-grade gliomas. *Front Immunol* (2022) 13:844778. doi: 10.3389/fimmu.2022.844778

68. Perez-Villamil B, Romera-Lopez A, Hernandez-Prieto S, Lopez-Campos G, Calles A, Lopez-Asenjo JA, et al. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. *BMC Cancer* (2012) 12:260. doi: 10.1186/1471-2407-12-260

69. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. *NATURE* (2014) 513:382–7. doi: 10.1038/nature13438

70. Wang J, Mouradov D, Wang X, Jorissen RN, Chambers MC, Zimmerman LJ, et al. Colorectal cancer cell line proteomes are representative of primary tumors and predict drug sensitivity. *GASTROENTEROLOGY* (2017) 153:1082–95. doi: 10.1053/j.gastro.2017.06.008

71. Mischak H, Ioannidis JP, Argiles A, Attwood TK, Bongcam-Rudloff E, Broenstrup M, et al. Implementation of proteomic biomarkers: making it work. *Eur J Clin Invest* (2012) 42:1027–36. doi: 10.1111/j.1365-2362.2012.02674.x

72. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat Rev Clin Oncol* (2020) 17:11–32. doi: 10.1038/s41571-019-0241-1

73. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat Methods* (2013) 10:1108–15. doi: 10.1038/nmeth.2651

74. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinf (Oxford England)* (2009) 25:2906–12. doi: 10.1093/bioinformatics/btp543

75. Akimoto N, Zhao M, Ugai T, Zhong R, Lau MC, Fujiyoshi K, et al. Tumor long interspersed nucleotide element-1 (LINE-1) hypomethylation in relation to age of colorectal cancer diagnosis and prognosis. *Cancers (Basel)* (2021) 13(9):2016. doi: 10.3390/cancers13092016

76. Hagg S, Jylhava J. Should we invest in biological age predictors to treat colorectal cancer in older adults? *Eur J Surg Oncol* (2020) 46:316–20. doi: 10.1016/j.ejso.2019.11.003

77. Liu J, Liu Z, Zhang X, Yan Y, Shao S, Yao D, et al. Aberrant methylation and microRNA-target regulation are associated with downregulated NEURL1B: a diagnostic and prognostic target in colon cancer. *Cancer Cell Int* (2020) 20:342. doi: 10.1186/s12935-020-01379-5

78. Raskov H, Soby JH, Troelsen J, Bojesen RD, Gogenur I. Driver gene mutations and epigenetics in colorectal cancer. *Ann Surg* (2020) 271:75–85. doi: 10.1097/SLA.0000000000003393

79. Jin K, Ren C, Liu Y, Lan H, Wang Z. An update on colorectal cancer microenvironment, epigenetic and immunotherapy. *Int IMMUNOPHARMACOL* (2020) 89:107041. doi: 10.1016/j.intimp.2020.107041

80. Chang J, Wang Y, Shao L, Laberge RM, Demaria M, Campisi J, et al. Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice. *Nat Med* (2016) 22:78–83. doi: 10.1038/nm.4010

81. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, et al. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat Struct Biol* (2002) 9:268–72. doi: 10.1038/nsb770

82. Putt KS, Chen GW, Pearson JM, Sandhorst JS, Hoagland MS, Kwon JT, et al. Small-molecule activation of procaspase-3 to caspase-3 as a personalized anticancer strategy. *Nat Chem Biol* (2006) 2:543–50. doi: 10.1038/nchembio814

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fimmu.2022.1010434

Check for updates

OPEN ACCESS

EDITED BY
Hao Zhang,
Xiangya Hospital Central South University, China

REVIEWED BY
Hongze Chen,
Tulane University, United States
Zhen Ning,
Dalian Medical University, China

*CORRESPONDENCE
Xiaoming Zou
zou4930@163.com
Jiacheng Li
ljc911201@163.com

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 03 August 2022
ACCEPTED 24 August 2022
PUBLISHED 12 September 2022

CITATION
Yang S, Zou X, Li J, Yang H, Zhang A, Zhu Y, Zhu L and Zhang L (2022) Immunoregulation and clinical significance of neutrophils/NETs-ANGPT2 in tumor microenvironment of gastric cancer.
Front. Immunol. 13:1010434.
doi: 10.3389/fimmu.2022.1010434

COPYRIGHT
© 2022 Yang, Zou, Li, Yang, Zhang, Zhu, Zhu 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.

Immunoregulation and clinical significance of neutrophils/NETs-ANGPT2 in tumor microenvironment of gastric cancer

Shifeng Yang[1,2], Xiaoming Zou[1*], Jiacheng Li[3*], Hao Yang[1], Ange Zhang[1,2], Yanli Zhu[4], Lei Zhu[1] and Lisha Zhang[1]

[1]Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]The Key Laboratory of Myocardial Ischemia, Ministry of Education, Harbin, China, [3]Department of General Surgery, The First Affiliated Hospital of Jiamusi University, Jiamusi, China, [4]Key laboratory of Microecology-immune Regulatory Network and Related Diseases School of Basic Medicine, Jiamusi University, Jiamusi, China

Although significant progress has been made in the study of gastric cancer (GC), clinicians lack reliable protein markers for accurate diagnosis and tumor stratification. Neutrophil extracellular traps (NETs) are networks of extracellular fibers composed of DNA from neutrophils. We have previously reported that abundant NETs are deposited in GC, damaging human umbilical vein endothelial cells (HUVECs) and triggering the release of tissue factors, leading to a hypercoagulable state in GC. However, the specific effects of NETs on HUVECs are unclear. We aimed to explore the functional changes caused by NETs on HUVECs, providing evidence that NETs may fuel GC progression. Through quantitative proteomics, we identified 6182 differentially expressed proteins in NET-stimulated HUVECs by TMT. The reliability of the TMT technique was confirmed by parallel reaction monitoring (PRM) analysis of 17 differentially expressed proteins. Through bioinformatics analysis, we found that NETs upregulate ANGPT2 in HUVECs. We comprehensively analyzed the prognosis, biological function, immune response, and therapeutic value of ANGPT2 in GC. We found that overexpression of ANGPT2 in GC is associated with poor prognosis and potentially regulates multiple biological functions. At the same time, ANGPT2 also predicted immunotherapeutic and chemotherapeutic responses in GC. In conclusion, NETs promoted ANGPT2 overexpression in the GC microenvironment. In the future, the neutrophil/NETs-ANGPT2 axis may provide a new target for the treatment of GC.

KEYWORDS

gastric cancer, human umbilical vein endothelial cells, neutrophil extracellular traps (NETs), ANGPT2, tumor microenvironment

Introduction

Gastric cancer (GC), one of the most common digestive tract malignant tumors, severely threatens human health worldwide, with high morbidity and mortality rates (1, 2). GC has strong heterogeneity and rapid progression, with the liver being the most common organ involved in hematogenous metastasis (3). The five-year survival rate of patients with GC after liver metastasis is less than 20% (4). New tumor markers, treatment targets, and therapeutic strategies are urgently required to improve the prognosis and the postoperative survival time of patients with GC.

The importance of the tumor microenvironment (TME) in tumorigenesis is being extensively explored (5, 6). GC is a dynamic entity. There is continuous communication between GC cells and their surrounding environment, which fuels and maintains phenotypic and heterogeneous factors of cancer cells (7, 8). The complex TME of GC includes fibroblasts, immune cells, adipocytes, vascular endothelial cells, and extracellular matrix (9). Tumor-associated neutrophils (TANs) are important matrix partners in carcinogenesis. Neutrophils can increase gastric cancer cell migration and invasion abilities (10–12). Neutrophil extracellular traps (NETs) are defensive mechanisms deployed by neutrophils (13, 14). Current studies suggest that NETs may exert double-edge effects on tumors, possibly due to the physicochemical properties of NETs (15, 16). NETs are DNA-based network structures rich in a variety of active proteins (17, 18). NETs can capture circulating tumor cells through DNA reticular structures. Cools-Lartigue et al. have shown that NETs can prevent circulating tumor cells from entering the vascular system of the lung and liver, suggesting that NETs are closely related to tumor progression (19).

Although endothelial cells are relatively stable cells in the human body, they can re-enter the cell cycle under certain triggering conditions (20). Some studies have shown that vascular occlusion and blood hypercoagulability adversely affect the hemodynamic stability of patients, possibly increasing the number of circulating tumor cells captured by NETs in tumor patients (21–23). However, NETs can damage endothelial cells, and captured tumor cells can spread, forming new metastatic foci after adhering to activated endothelial cells (15). Therefore, clarifying the mechanism between NETs and endothelial cells is pivotal to explaining tumor occurrence and development. In this study, we found that NETs act on HUVECs, triggering the release of coagulation-related factors and promoting tube formation. Therefore, the interaction between NETs and HUVECs is worth exploring.

In our study, we used proteomic technology to characterize differentially expressed proteins in NET-stimulated HUVECs. Parallel reaction monitoring (PRM) analysis was used to validate 17 selected proteins. Finally, ANGPT2 was chosen as the target protein. ANGPT2 belongs to the angiopoietin (Ang) family (24). ANGPT2 expression plays a vital role in vascular remodeling and inflammation (25). As an early autocrine initiator of angiogenesis, ANGPT2 first destabilizes resting blood vessels, thus allowing VEGF to drive the proliferation and chemotactic migration of vascular buds. ANGPT2 also activates TIE2-expressing monocytes/macrophages (TEM), which promote angiogenesis, tumor formation, metastasis, and immunosuppression (26). ANGPT2 regulates vascular remodeling and tumor progression under many pathological conditions through different effects on TIE2 signaling. Targeting the ANG-TIE2 axis can significantly improve the effect of tumor immunotherapy (27). Therefore, it also proves that ANGPT2 participates in tumor immune response and targeting neutrophil/NETs-ANGPT2 may be a new target of anti-tumor therapy in the future.

Materials and methods

Patients and tissue samples

This study included 20 healthy participants and 60 patients with newly diagnosed primary gastric cancer by pathological examination who were admitted to the Second Affiliated Hospital of Harbin Medical University from October 2019 to April 2021. We based the evaluation (TNM), staging, and histological classification of gastric cancer on the 8th edition of the American Joint Commission on Cancer (34). The inclusion criterium was patients aged 18–65 years without any endocrine, cardiovascular, hematological, and infectious diseases. The exclusion criteria were as follows: patients who are not pregnant, without coexisting cancers, and those who did not receive any antineoplastic treatment before surgery. Samples of cancerous and para-cancerous tissues of patients with gastric cancer were collected. All patients provided informed consent, and this study was approved by the Internal Audit and Ethics Committee of the Second Affiliated Hospital of Harbin Medical University. Ethical review approval document No : KY2021-075.

Isolation of neutrophils

A neutrophil isolation kit (TBD sciences, Tianjin, China) was used to isolate neutrophils from blood samples of patients with GC or healthy controls. Venous blood samples were collected using a 5 ml catheter containing 3.2% sodium citrate. Then, a 5 mL neutrophil separation solution was mixed with anticoagulants and whole blood. The neutrophil layer was added to the erythrocyte separation solution and centrifuged repeatedly at 450 ×g for 5 min at 24°C until red blood cells disappeared. Neutrophils were resuspended in 1 mL PBS and counted.

Generation, isolation, and preparation of NETs

Purified neutrophils (1×10^6/well) were seeded in a 6-well plate, stimulated with 100 nM PMA, and cultured at 37°C, 5% CO_2, for 4 h. Then, the medium was gently removed, leaving NETs and neutrophils attached to the bottom. Pre-cooled PBS without calcium and magnesium was added to elute NETs and neutrophils. Liquid samples were collected in 6-well plates, centrifuged at 450 ×g, 4°C, for 10 min, and the supernatant was collected. The collected supernatant was centrifuged at 4°C 15000 ×g for 15 min. The concentration of NETs was determined using a micro-DNA instrument, and samples were stored at -20°C.

Cell culture

Primary human umbilical vein endothelial cells (HUVECs) were purchased from PROCELL (Wuhan, China). All cell lines were verified by a short tandem repeat (STR) map and negative detection of mycoplasma contamination. All cell lines were cultured at 37°C in a 5% CO_2 humidified laboratory and cultured in ECM (Scien Cell, USA) medium containing 10% fetal bovine serum (FBS) for 5 days; then, the medium was replaced with FBS-free medium.

Stimulation of HUVECs

HUVECs were cultured in a medium containing NETs (NETs were derived from neutrophils from patients with GC, as described above) at 37°C, 5% CO_2, for 24 h. To conduct the inhibition test, isolated NETs were pretreated with DNase-1 (100 U/mL) at 37°C for 1 h, and then co-cultured with HUVECs. At a specific time, point, HUVECs were collected by centrifugation for follow-up experiments, including western blot, after protein extraction and measurement of target protein expression.

Tube forming experiment

The Corning matrix glue was diluted with matrix glue at a 1:1 ratio in a 24-well plate, with 200 µl per well. The mix was incubated at 37°C for 30 min to obtain a gelatinous base glue. HUVECs were digested, and a single-cell suspension was prepared. The conditioned medium was replaced by FBS-free medium. Next, 20 000 cells were seeded in each well. The process of tubule formation was observed under a microscope at 37°C for 2–4 h.

Immunofluorescence

HUVECs were stimulated with 0.4 µg/mL NETs prepared hours in advance, and then an immunofluorescence experiment was carried out. The medium in the 24-well plate was removed, cells were fixed in 1% PFA for 15 min, and washed twice with PBS. Samples were then incubated in 3% BSA for 30 min, washed twice with PBS, and incubated overnight at 4°C with primary antibodies: diluted anti-VWF (Abcam), anti-ANGPT2 (Affinity) (13% BSA 500). Specimens were washed twice with PBS and incubated with secondary antibodies: diluted AF-488 or AF-594 antibody for 30 min. For nuclear staining, cells were incubated with DAPI for 10 min and washed with PBS twice. Sharp tweezers were used to remove adherent cells from each well and deposit them with the cellular side upside down on a slide coated with anti-quenched glycerol. Specimens were observed and imaged under a confocal microscope. All results were expressed as mean ± standard deviation (SD). SPSS or GraphPad Prism 8.0 were used to analyze all data. Multiple groups were compared and analyzed by one-way. ANOVA. Two pairs variables were analyzed by paired t-test. Statistical significance was determined when $P< 0.05$.

TMT-MS/MS labeled

The trypsin-hydrolyzed peptides were desalted with StrataXC18 (Phenomenex) and then freeze-dried in a vacuum. Peptides were dissolved with 0.5 M TEAB and labeled according to the operating instructions of the TMT kit. Briefly, after the labeled reagent is thawed, it is dissolved with acetonitrile, mixed with the peptide, and incubated at room temperature for 2 h; the labeled peptide is mixed, desalted, and vacuum freeze-dried.

Quantitative analysis

The results of database search of MS data show the signal intensity value (Intensity) of each peptide in different samples. According to this information, the relative quantity of protein is calculated by the following steps: First of all, the signal intensity values (I) of peptides in different samples were centralized, and the relative quantitative values (R) of peptides in different samples were obtained. The formula is as follows: where i represents the sample and j represents the peptide:

$$Rij = Iij/Mean(Ij)$$

The relative quantitative value of protein was calculated. The relative quantitative value of protein was expressed by the median of the relative quantitative value of specific peptide corresponding to protein. The formula is as follows: where k

represents protein. J represents the specific peptide to which the protein belongs:

$$Rik = Median\,(NRij, j \in k)$$

Differential protein screening.

Protein difference analysis first selects the sample pairs that need to be compared, and takes the ratio of the mean value of all biological repetitive quantitative values of each protein in the comparison sample pair as the difference multiple (Fold Change, FC). For example, the multiple of protein difference between sample group A and sample group B is calculated. The calculation formula is as follows: R represents relative quantity of protein, i represents sample, k represents protein:

$$FC_{A/B,k} = Mean(R_{ik}, i \in A)/Mean(R_{ik}, i \in B)$$

According to the above difference analysis, when P value ≤0.05, the change threshold of differential expression was more than 1.3 as the threshold of significant up-regulation, and less than 1.3 as the threshold of significant down-regulation.

Establishment of animal model

Animal experiments were carried out at the Animal Experimental Center of the Key Laboratory of Myocardial Ischemia of the Second Affiliated Hospital of Harbin Medical University in strict accordance with the scheme approved by the Animal Care and Use Committee. Ten BALB/c nude mice aged 5–7 weeks were purchased from Weitong Lihua Experimental Animal Technology Co. Ltd and kept in a 22°C aseptic animal house. Food and autoclaved water were provided. The mice were randomly divided into the control (n = 5) and DNAse-1 treatment (n = 5) groups. All mice were anesthetized with 2% isoflurane mixture; the axillary skin of the mice was disinfected with sterilized cotton balls; 1 × 106 HGC-27 gastric cancer cells were subcutaneously injected into the armpit; and cotton swabs were applied to stop the bleeding. The mice in the treatment group were intraperitoneally injected with deoxyribonuclease (DNase-I, 50μg/mouse, Roche) every 12 h until they were euthanized. The tumor volume in each mouse was measured once every 3 days and calculated using the following formula: 0.5 × length × width2. Euthanasia was performed on all mice after 18 days. The tumor tissues were harvested, weighed, soaked in 4% paraformaldehyde for 24 h, embedded in paraffin, and divided into 4-μm paraffin sections for follow-up immunohistochemical staining.

Screening of ANGPT2 genes

NET+ vs. Control difference analysis was performed with a cutoff of P-value< 0.05. 123differentially upregulated proteins were identified. We then used machine learning algorithms, svm-rfe and random Forest-rfe, to reduce dimensionality, yielding 29 proteins. Then we selected the 17 most valuable proteins according to the literature.

STAD dataset and preprocessing

STAD-related data and corresponding clinical information were downloaded and collected from TCGA (https://xenabrowser.net/). The transcriptional spectrum of 414 patients with STAD cancer was obtained from TCGA, of which 388 patients with STAD had complete OS information. 36 paracancerous STAD samples were obtained from TCGA. The transcriptional data from 174 normal gastric samples were collected from the genotype-tissue expression project (GTEx; https://www.gtexportal.org). The fragment (FPKM) value per kilobase was converted to the transcript (TPM) value per kilobase.

Carcinogenic characteristics of ANGPT2

The difference of ANGPT2 between GC samples and non-cancer samples was analyzed in TCGA and GTEx. We estimated the ANGPT2 of 388 patients in the TCGA-STAD dataset, then sorted patients into high ANGPT2 and low ANGPT2 groups according to the P-value of the best cutoff. Kaplan–Meier curves were used to analyze the association between OS and ANGPT2. Univariate and multivariate Cox regression analysis was performed for ANGPT2.

Genome change

Somatic mutation data and somatic copy number variation (CNV) were collected from TCGA. Genomic identification (GISTIC) analysis of important targets in cancer is used to evaluate genomic characteristics. GISTIC2.0 analysis (https://gatk.broadinstitute.org) was used to evaluate the increase or loss of copy number of CNV landscape and amplification or deletion peaks.

The immunological characteristics of TME

We used the Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression (ESTIMATE) algorithm to estimate the abundance of immune cells and the level of stromal cell infiltration. The level of immune infiltrating cells in STAD was analyzed comprehensively using Tumor

Immune Estimation Resource2.0 (TIMER2.0; http://timer.cistrome.org/) network server. The relative proportion of ten kinds of immune cells in the tumor was estimated by the MCPcounter algorithm. The infiltration level of 28 kinds of immune cells was expressed as the enrichment fraction based on the corresponding characteristics using a single sample genome enrichment analysis (ssGSEA) using the GSVA R package. The response of STAD to immunotherapy was evaluated by a submap algorithm.

Functional analysis

Gene set variation analysis (GSVA) was carried out with the GSVA R package.

Prediction of drug response

Pharmacogenomics data from cancer drug sensitivity genomics (GDSC, https://www.cancerrxgene.org/) are used to predict the drug sensitivity of included cases. Drug sensitivity was calculated with the oncoPredict R package.

Statistical analysis

Wilcoxon test was used to compare non-normally distributed data. The t-test was used to compare normally distributed variables between two groups. Using the R package survminer, the Kaplan–Meier survival graph was used to estimate the OS between two groups. The Cox regression of survival analysis was conducted with the R package survival. All heat maps were drawn with the R pheatmap package. The data were visualized with R ggplot2 package. $P< 0.05$ was considered statistically significant.

Results

3.1 Gastric cancer tissue shows multiple blood vessels with abundant NETs deposition around blood vessels

Immunofluorescence staining showed that GC tissue (n = 30) had more NET deposition (Figures 1A, B) than adjacent tissue (n=30) and staining of HE showed microvessels in GC; the worse the staging, the more abundant the microvessels (Figures 1A, C). CD31 immunofluorescence staining was used to characterize the blood vessels of GC, several NETs deposits were found around the vessels (Figures 1C, D).

NETs enhance the angiogenesis ability of HUVECs and the expression of damage factor vWF

After stimulating HUVECs with 0.4 μg/ml NETs, more tubules (Figure 1E) were formed compared with the PBS group. Immunofluorescence staining showed that VWF expression and the intercellular space increased after HUVECs were stimulated with NETs (Figure 1F). DNAse-1 could attenuate the above phenomena (Figures 1E, F). Figure 1G shows the result of quantitative analysis of the mean fluorescence intensity of VWF, which is statistically significant. Figure 1H shows scanning electron microscope images of NETs and cell-free NETs.

Identification of differentially expressed proteins in NET-stimulated HUVECs by TMT-MS/MS and PRM

For TMT-MS/MS analysis of NET-stimulated and unstimulated HUVECs, a total of 363519 secondary mass spectra were collected, including 83412 effective mass spectra, with a utilization rate of 22.9%. Among quantifiable proteins, 123 upregulated and 73 downregulated proteins were observed in the NET-stimulated group but not in the unstimulated group (Figure 2A). Then the dimension of 123 upregulated differential proteins was reduced with machine learning, reduce the number of proteins to 29 (Figure 2B). By reviewing the literature related to tumor progression and angiogenesis, 17 proteins were selected from the above 29 proteins for PRM analysis, which, namely ACE, ANGPT2, CCN1, CD34, GDF15, HTRA1, HTRA3, IFI16, IGFBP7, LAMA4, LAMC1, MRPS15, MYO1B, PLVAP, RPL34, RPL4, and RPL6 (Figures 2C, D). The results showed that the expression of these proteins was significantly upregulated in the NETs-stimulated group.

ANGPT2 is associated with poor prognosis of patients with gastric cancer and participates in carcinogenesis

The TCGA database analysis indicated that ANGPT2 expression in GC tissues (414) was significantly higher than in para-cancerous (36) and normal tissues in GTEx (174) (Figure 3A). The Kaplan–Meier curve results showed overall svrvival (OS) that high ANGPT2 expression was associated with worse survival (Figure 3B). Univariate and multivariate analysis based on TCGA dataset revealed that ANGPT2 can work as an independent risk factor in GC. (Figure 3C). GSVA analysis highlighted that ANGPT2 was significantly associated with

10.3389/fimmu.2022.1010434

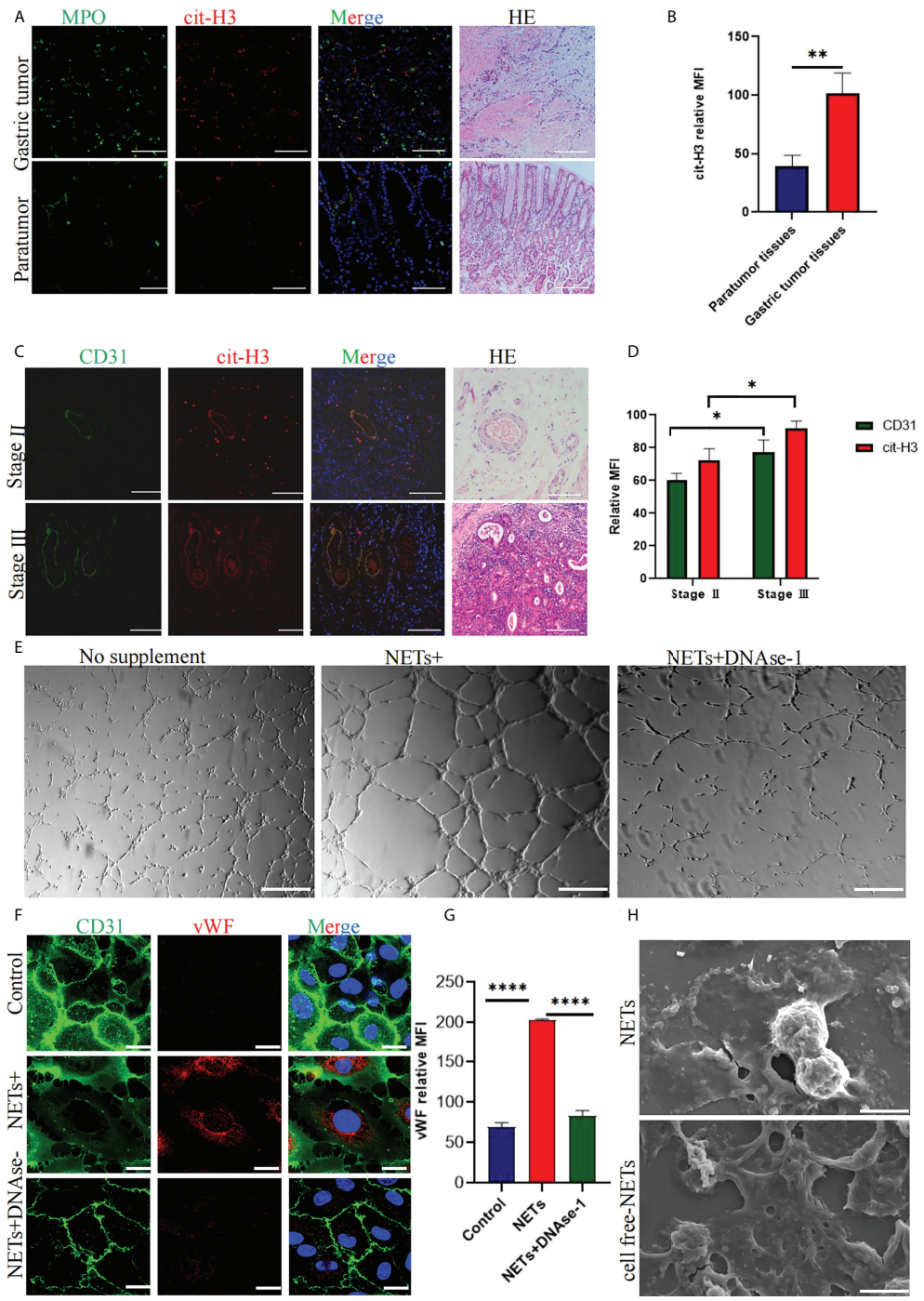


FIGURE 1
Expression of NETs in gastric cancer (GC) and paracancerous tissues and quantification of related factors in NET-stimulated HUVECs. **(A)** NETs deposition in GC and paracancerous tissues was analyzed by immunohistochemical staining. HE staining was used to analyze the distribution of microvessels in GC and paracancerous tissues. **(B)** Quantitative analysis of cit-H3 in GC and paracancerous tissues. **(C)** The localization of CD31 and NETs marker cit-H3 in GC. HE staining was used to analyze the distribution of microvessels in different stages of GC. **(D)** Mean fluorescence intensity of CD31 and cit-H3 in tumor tissues at different stages. **(E)** Tubule formation of HUVECs stimulated with PBS, NETs, and NETs+DNAse-1 was analyzed by a tube forming experiment. **(F, G)** The vWF expression on HUVECs was detected by confocal microscopy and analyzed with Image J software [the expression is indicated as the mean fluorescence intensity (MFI). **(H)** NETs and cell-free NETs imaged by scanning electron microscope. 20× objective. Scale bars indicate 20 mm. Arrows indicated ETs. All values are mean ± SD. ****P < 0.0001, **P < 0.01, *P < 0.05 by one -way ANOVA or t-test.

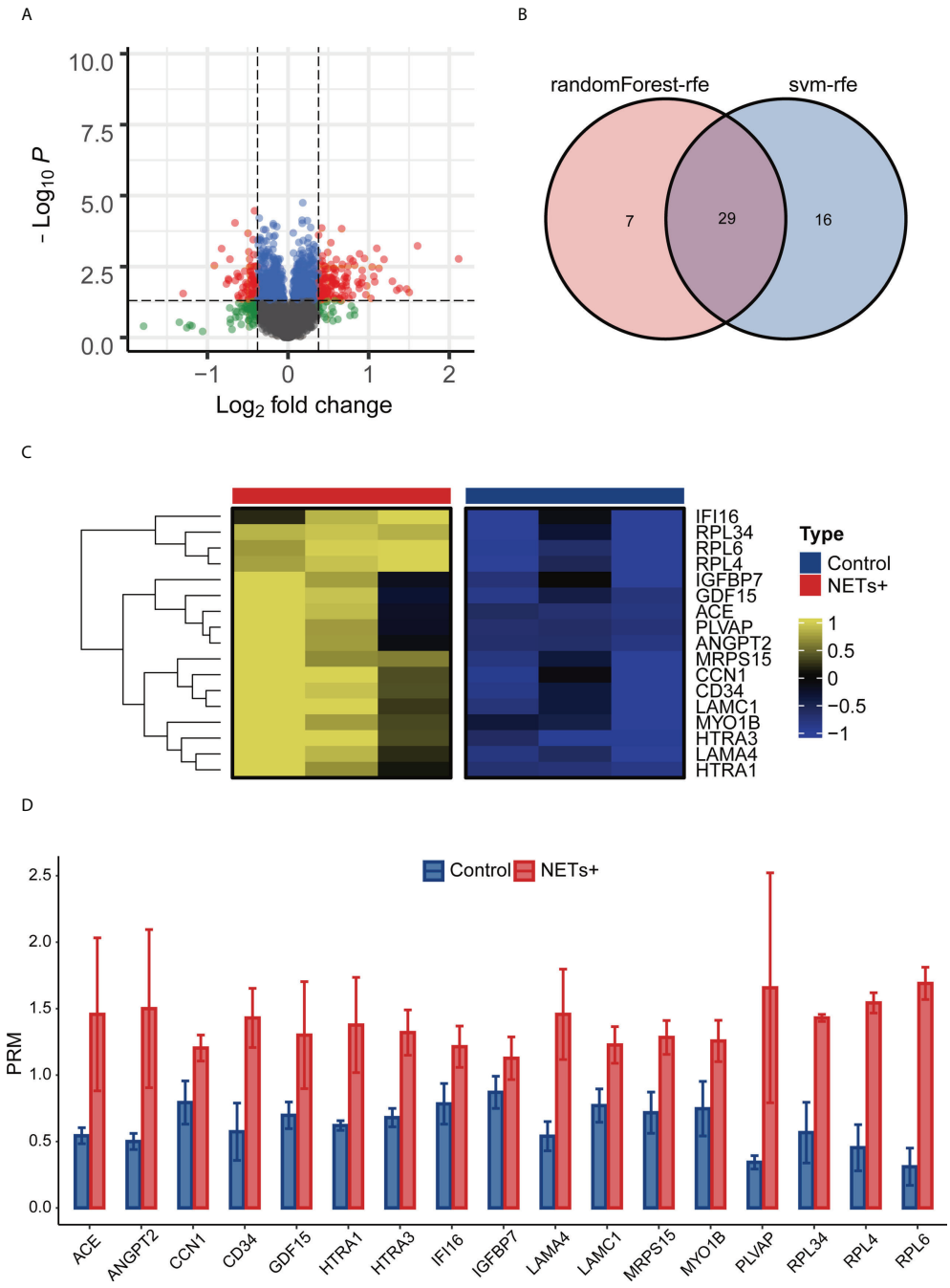


FIGURE 2
Screening ANGPT 2 based on TMT-MS/MS data and PRM analysis. **(A)** Volcanic map of the results of NETs + vs. Control difference analysis. **(B)** Svm-rfe and randomForest-rfe reduce dimensionality based on up-regulating differential proteins. **(C)** PRM analysis heat map of 17 differential proteins. **(D)** Analysis of bar plot by PRM of 17 differential proteins.

many important tumor-related pathways, such as the p53 signaling pathway, regulation of DNA damage response signal transduction, p53 class mediator, JAK-STAT signaling pathway, and cell migration involved in angiogenesis. We also observed that ANGPT2 participated in immune response-related processes, such as the T cell receptor signaling pathway and leukocyte trans endothelial migration (Figure 3D). Therefore, the high ANGPT2 expression can activate a variety of tumor-related pathways and immune responses, suggesting that ANGPT2 plays a significant role in the occurrence and development of GC.

ANGPT2 expression and the diversity of genomic changes

CNV in the human genome can affect gene expression by altering gene expression, destroying regulation or coding regions, or changing genome structure, ultimately affecting the normal function of genes. CNVs have been directly or indirectly related to various diseases. Cancer genomes are usually characterized by somatic CNV, often accompanied by amplification of proto-oncogenes or deletion of tumor suppressor genes. CNV and somatic mutation analysis were conducted in the TCGA GC data set. The results of the global CNV map showed that the amplification of the high ANGPT2 group was concentrated in Chr17, especially 17q12, and the deletion was concentrated in Chr4 and Chr16, especially loci 4q35.2 and 16q23.1. In the low ANGPT2 group, the increase of Chr17, the loss of Chr9, and the loss of Chr16 were identified. The amplification was mainly found in 17q12, and the deletion was mainly concentrated in 9p21.3 and 16q23.1 (Figure 4A, Supplementary Figure S1). The global view of mutation distribution shows that actin (TTN) and cell tumor antigen p53 (TP53) mutations are the most abundant in both the high ANGPT2 and low ANGPT2 groups, 51% and 44%, respectively (Figure 4B). The next three most common mutations in the high ANGPT2 group were MUC16 (31%), ARIDEA (27%), and LRP1B (26%). In the low ANGPT2 group, LRP1B (32%), MUC16 (30%), and SYNE1 (30%) (Figure 4B).

The mutation frequency ratio between the high ANGPT2 and low ANGPT2 groups was assessed by Fisher's exact test and sorted by increasing P-values. The sudden change load of the high ANGPT2 group was lower than that of the low ANGPT2 group (Supplementary Figure S2A). At the same time, we analyzed the coexistence or mutually exclusive mutation (Supplementary Figures S2B, C) of the 25 most frequent mutations. In the high ANGPT2 group, TTN and CSMD1 mutations frequently coexisted. ARID1A and PIK3CA, SYNE1 and SPTA1, FAT4 and DMD were also closely related mutation sites. In the low ANGPT2 group, the common co-mutations were FLG, SACS, OBSCN, and PCLO. Moreover, we found that the changes of some gene pairs were mutually exclusive, such as TP53 and PIK3CA, TP53 and ARID1A in the high ANGPT2 group; TP53 and ARID1A in the low ANGPT2 group (Supplementary Figures S2B, C).

Immune correlation of ANGPT2 in TCGA cohort

ESTIMATE analysis showed that the ANGPT2 immune and stromal scores were significantly increased (Figure 5A) in the high expression group, implying worse patient prognoses. We further analyzed the correlation between ANGPT2 expression and neutrophil infiltration. We found a positive correlation based on ESTIMATE, MCP counter, ssGSEA, and TIMER algorithms (Figure 5B), consistent with our results. Neutrophil infiltration correlates with NET deposition. At the same time, we also analyzed the correlation between ANGPT2 and other immune cell infiltration, and found that the high expression of ANGPT2 was significantly correlated with Macrophages M0 、NK cells resting and Mast cells activated (Supplementary Figures S3A, B). So we have reason to think that ANGPT2 is involved in the regulation of immune microenvironment for gastric cancer.

Potential targets of immunotherapy and chemotherapy in patients with gastric cancer with high ANGPT2 expression

Since immune checkpoint inhibitors are the basis of immunotherapy, immune checkpoint expression is of great significance in guiding clinical practice. The submap result showed that the response to immunotherapy differed between the high ANGPT2 and the low ANGPT2 groups. The result indicated that GC patients with high expression of ANGPT2 are more likely to respond with immunotherapy (Figure 6A), which is a significant reference in accurately choosing immunotherapy.

We analyzed the correlation between ANGPT2 and multiple immune checkpoints. The results showed that ANGPT2 was closely related to the level of multiple immune checkpoints in GC, PD-L1, PD-L2, and CTLA-4 (Figure 5C) are the most significant. These results suggest that inhibition of ANGPT2JI may be beneficial to immunotherapy.

We used the GDSC database to evaluate chemotherapeutic drug responses in high ANGPT2 and low ANGPT2 groups and measured accuracy by 10x cross-validation. The results showed the sensitivity of the high expression group to the following chemotherapeutic drugs, Rapamycin, LCL161, CZC24832, OTX015, AZD5153, GDC0810, UMI-77, BPD-00008900, Ulixertinib, Uprosertib, GSK591, and AT13148 was lower relative to the low expression group (Figure 6B). Therefore, patients with low ANGPT2 expression might experience better chemotherapeutic outcomes, and the efficacy of these chemotherapeutic drugs is worthy of further exploration.

ANGPT2 is highly expressed in gastric cancer, and NETs promote ANGPT2 expression in HUVECs

To further verify ANGPT2 expression in GC specimens, we found that ANGPT2 was highly expressed in cancer tissues from patients with GC than in para-cancerous tissues. ANGPT2 was more overexpressed in stage II/III than in stage I patients (Figures 7A, C). Compared with the control group, intraperitoneal injection of DNAse-1 reduced the growth

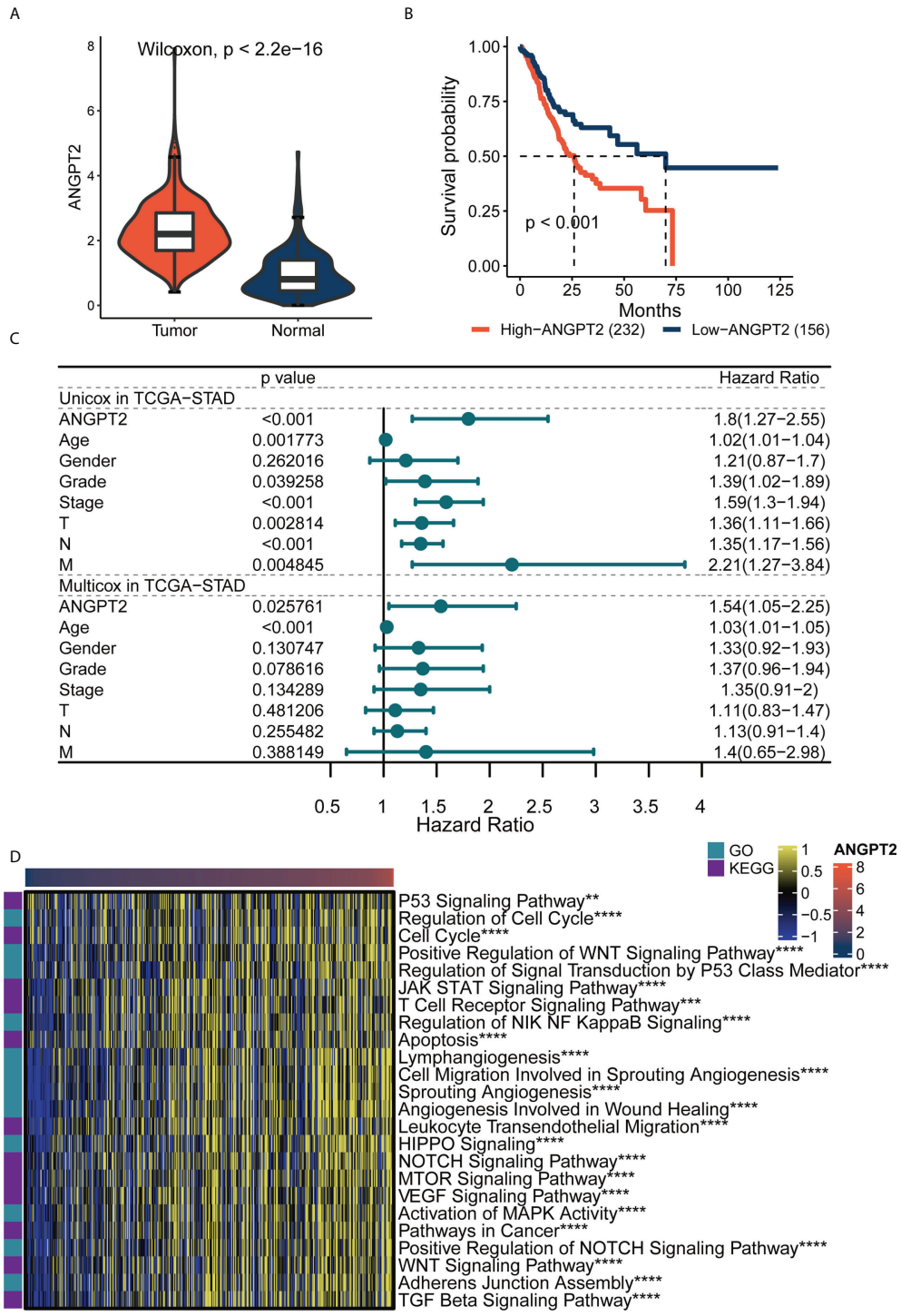


FIGURE 3
The importance of ANGPT2 gene in TCGA-STAD. **(A)** ANGPT2 expression level between STAD samples from TCGA cancer database and TCGA paracancerous plus GTEx normal samples. **(B)** Kaplan–Meier curves of high and low ANGPT2 groups in TCGA data. **(C)** Forest map of univariate and multivariate cox regression based on TCGA dataset and clinical variables. **(D)** GSVA thermomaps show functional pathways (GO and KEGG cancer-related pathways) significantly related to ANGPT2 in TCGA. **P < 0.01, ***P < 0.001 and ****P < 0.0001.

10.3389/fimmu.2022.1010434

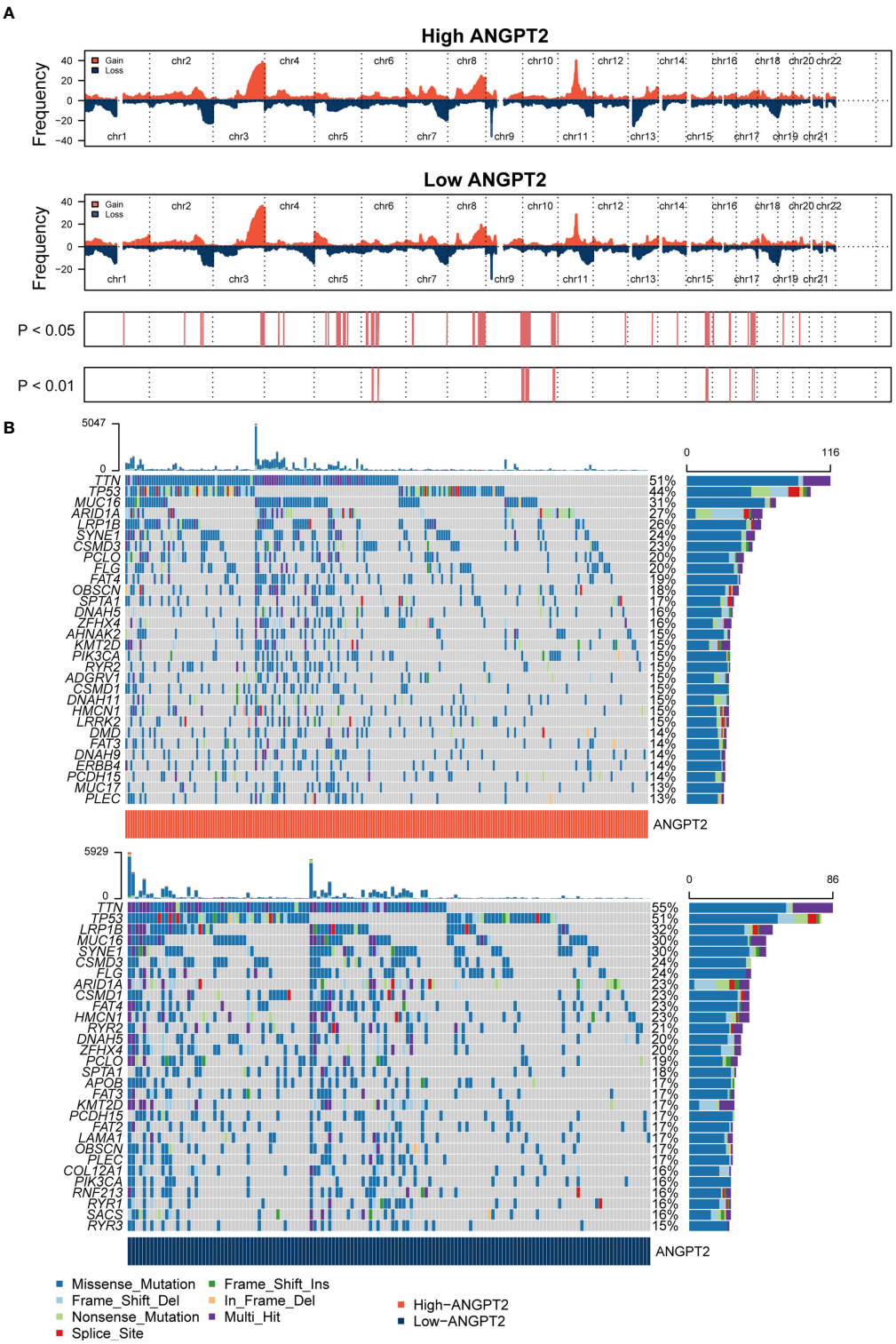


FIGURE 4
Genomic changes related to ANGPT2 in STAD samples **(A)** Changes in the copy number of ANGPT2 between high and low STAD groups. **(B)** Somatic mutation waterfall map in STAD of high and low ANGPT2 groups.

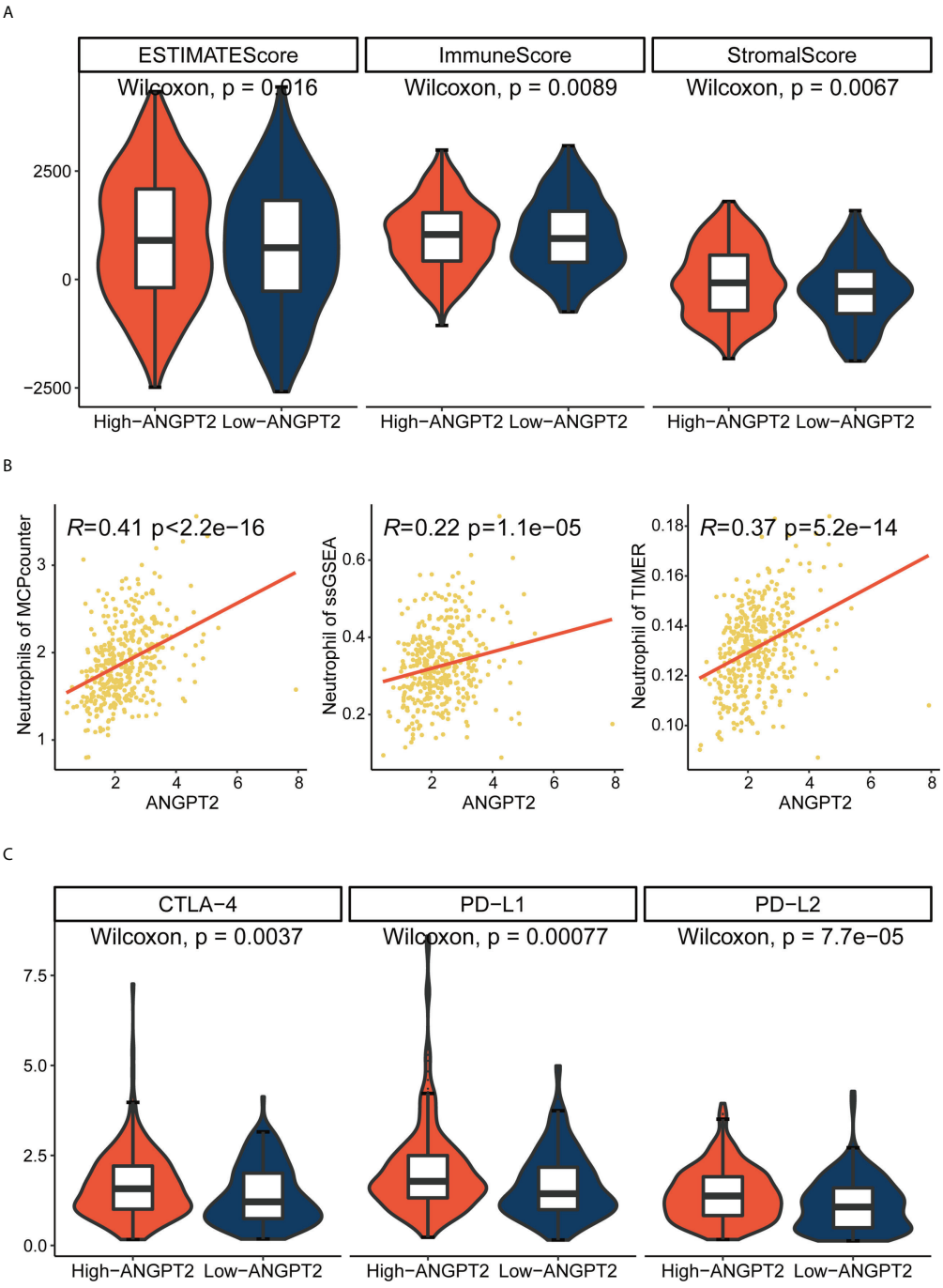


FIGURE 5
Relationship between ANGPT2 and immune characteristics in TCGA cohort. **(A)** Changes of ANGPT2 and ESTIMATE score, immune score, and stromal score. **(B)** Correlation between ANGPT2 and neutrophil cell infiltration level in three immune infiltration estimation algorithms. **(C)** The relationship between ANGPT2 and three classical immune checkpoints.

volume of the subcutaneous gastric cancer tumors (Figures 7E, F). Staining of GC tissue derived from a subcutaneous tumor model in nude mice showed that ANGPT2 expression in DNAse-1-treated mice was significantly lower than in the control group (Figures 7B, D). After stimulating HUVECs with NETs, we found that ANGPT2 expression increased significantly, as evidenced by Western blot, and began to change significantly when NET concentration was 0.4 μg/ml.

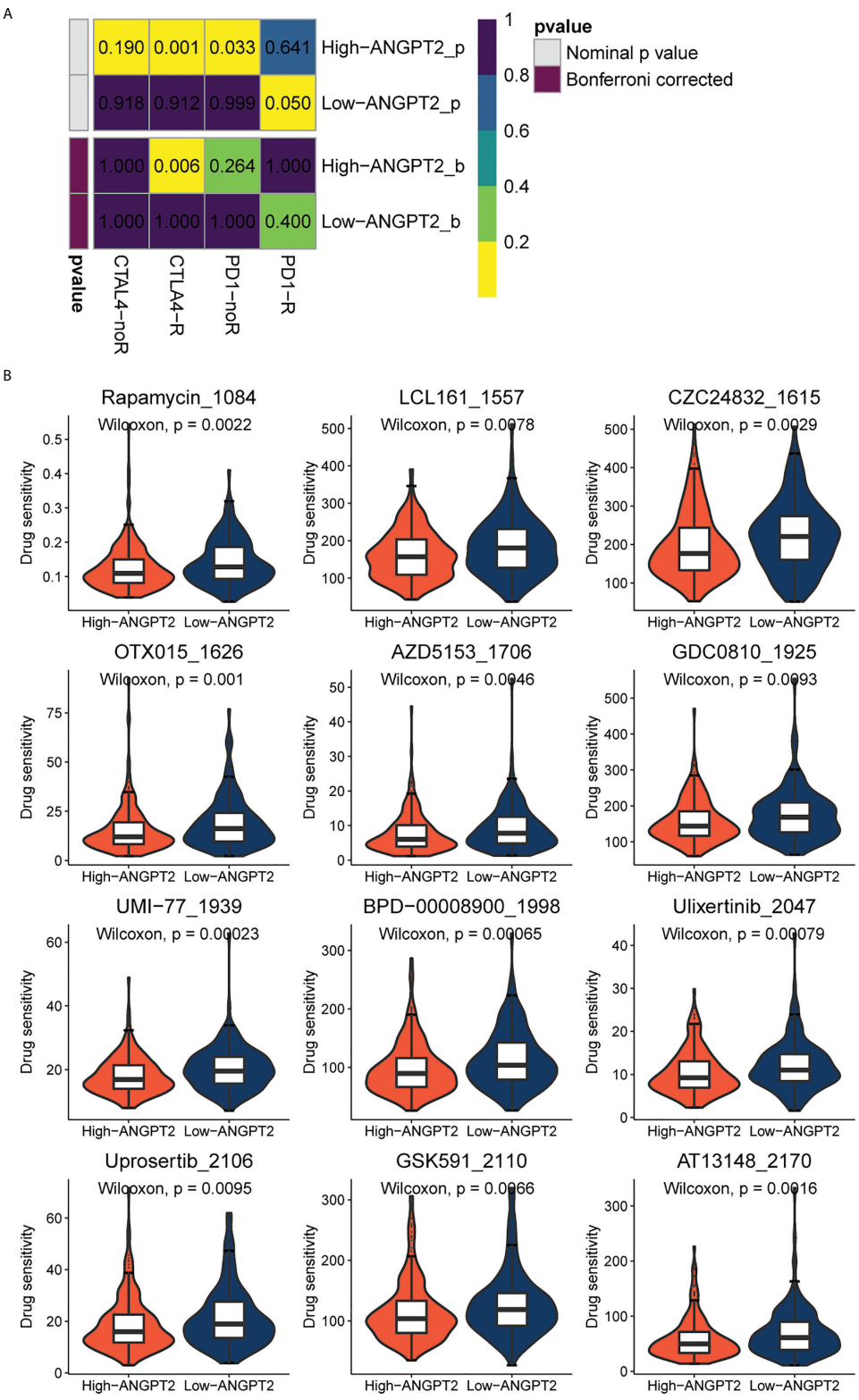


FIGURE 6
Immunotherapy and chemotherapy involved in ANGPT2 in TCGA-STAD. **(A)**Submap analysis of ANGPT2 levels in TCGA-STAD. **(B)** Box diagrams for estimating drug sensitivity of several GDSC chemotherapeutic drugs in the high ANGPT2 and low ANGPT2 groups.

Consistent with our previous experimental results, DNAse-1 could significantly weaken this phenomenon (Figures 7G, H). Through immunofluorescence staining, it was found that ANGPT2 expression increased, and CD31 expression decreased after HUVECs were stimulated with 0.4 μg/ml NETs, while DNAse-1 could weaken the effects of NETs (Figures 7I, J).

Discussion

An increasing number of studies have shown that the tumor microenvironment (TME) has important clinical significance in predicting prognosis and guiding immunotherapy. However, there is still a lack of systematic cell interaction analysis in TME. In our study, we analyzed and identified differentially expressed proteins in NET-stimulated HUVECs in GC with the help of TMT-MS/MS and PRM techniques. We found that NETs stimulation can increase ANGPT2 expression in HUVECs. Through comprehensive bioinformatic analysis, we found that the increased ANGPT2 expression is closely related to the poor survival of patients with GC. At the same time, ANGPT2 is related to the genomic changes of GC and TME. ANGPT2 can be used not only as a therapeutic indicator of immune checkpoint inhibitors but also as a target for accurate selection of chemotherapeutic drugs; thus, targeting the neutrophil/NETs-ANGPT2 axis may unveil new therapeutic targets in the future.

NETs are not only related to antimicrobial defense but also play a role in non-infectious diseases, including tumor occurrence and development, thrombosis, vasculitis, etch (28–30). Many NETs deposits have been found in the blood and tumor tissues of many solid tumors (31–33). Yang et al. found that neutrophils of patients with liver cancer, especially those with metastatic liver cancer, release more NETs; in animal experiments, they found that these NETs can capture liver cancer cells, induce cell death resistance, and enhance invasion (34). In our early studies, it has been confirmed that a large number of NETs deposit in GC tissue, significantly correlating with tumor staging. At the cellular level, NET deposits can promote the EMT process, invasion, and metastasis of GC cells, as well as the formation of a hypercoagulable state (35). In this study, we verified that NET deposition is more abundant in cancer tissues of patients with GC, accompanied by significant NET infiltration around miANGPT2cro vessels. Related studies have also confirmed that NETs can damage endothelial cells and recruit platelets to the injured site, leading to the formation of deep venous thrombosis (36, 37). This evidence shows that NETs can affect endothelial cells and trigger the onset and development of GC in the TME. Therefore, we should actively explore the specific mechanism of NETs on HUVECs. Due to the relatively low consistency between mRNA and protein expression in some tumor types, proteome analysis may outperform transcriptome analysis in detecting disease-related changes in cell activity and function (38). Therefore, in our study, proteomic technique was used to understand the changes in protein post-transcriptional translation levels after NETs stimulation of HUVECs. Thus, this approach allows us to explore the mechanism underlying the effects of NETs on HUVECs, providing a theoretical basis for the clinical treatment of GC.

Existing research results show that neovascularization in the TME provides nutrition and oxygen for tumors and is also a channel for tumor spread (39–41). Endothelial-derived SLIT2 protein and its receptor ROBO1 reportedly promote the migration and infiltration of cancer cells into endothelial tissue, whereas endothelial Slit2 knockout can inhibit tumor metastasis. In contrast, Slit2 knockout in the tumor can promote metastasis. Tumor cell-derived double-stranded RNA acts as an upstream signal to interact with RNA receptor TLR3 to induce endothelial SLIT2 expression (42). This study revealed that endothelial cells play a direct role in driving tumor metastasis and spread, proving that single genes from different cell sources can promote or inhibit cancer progression. In our study, we found that NETs act on HUVECs to trigger the release of various tumor-related factors, such as ANGPT2, ECM, CD40, and IL-8. Therefore, we have reason to believe that endothelial cells play a diverse role in the TME and can promote GC progression under the action of triggering factors.

In this study, we found that NETs promoted ANGPT2 release from HUVECs. Immunohistochemical staining showed large depositions of ANGPT2 in GC. The expression level in stage I was lower than that in stage II-III, which was positively correlated with the clinical stage, indicating that ANGPT2 overexpression occurred in the late stage of tumorigenesis. In murine subcutaneous tumor tissues, ANGPT2 expression of DNAse-1-treated mice was significantly decreased, which proved that inhibition of NETs could significantly reduce ANGPT2 expression. Consistent with this study, many studies have shown that ANGPT2 expression is significantly upregulated in breast cancer, pancreatic cancer, glioma, GC, colon cancer, liver cancer, melanoma, and other tumors. ANGPT2 plays a key role in tumor angiogenesis, tumor inflammation, and tumor metastasis (43–50). For instance, studies have shown that ANGPT2 and other factors are involved in the growth and metastasis of breast cancer, and their expression is related to the clinical stage of cancer, blood lymphatic, etc. Its high expression can lead to abnormal regulatory functions such as vascular repair and reconstruction in breast tissue, increasing the incidence of breast cancer (43). In colon cancer, an immunohistochemical study showed that the expression level of ANGPT2 protein was not related to the degree of differentiation and lymph node metastasis but to the depth of intestinal wall invasion, blood metastasis, and poor clinical prognosis (46). In normal intestinal mucosa, the expression of ANGPT2 is lower than that of intestinal adenoma, and there is a positive correlation between the

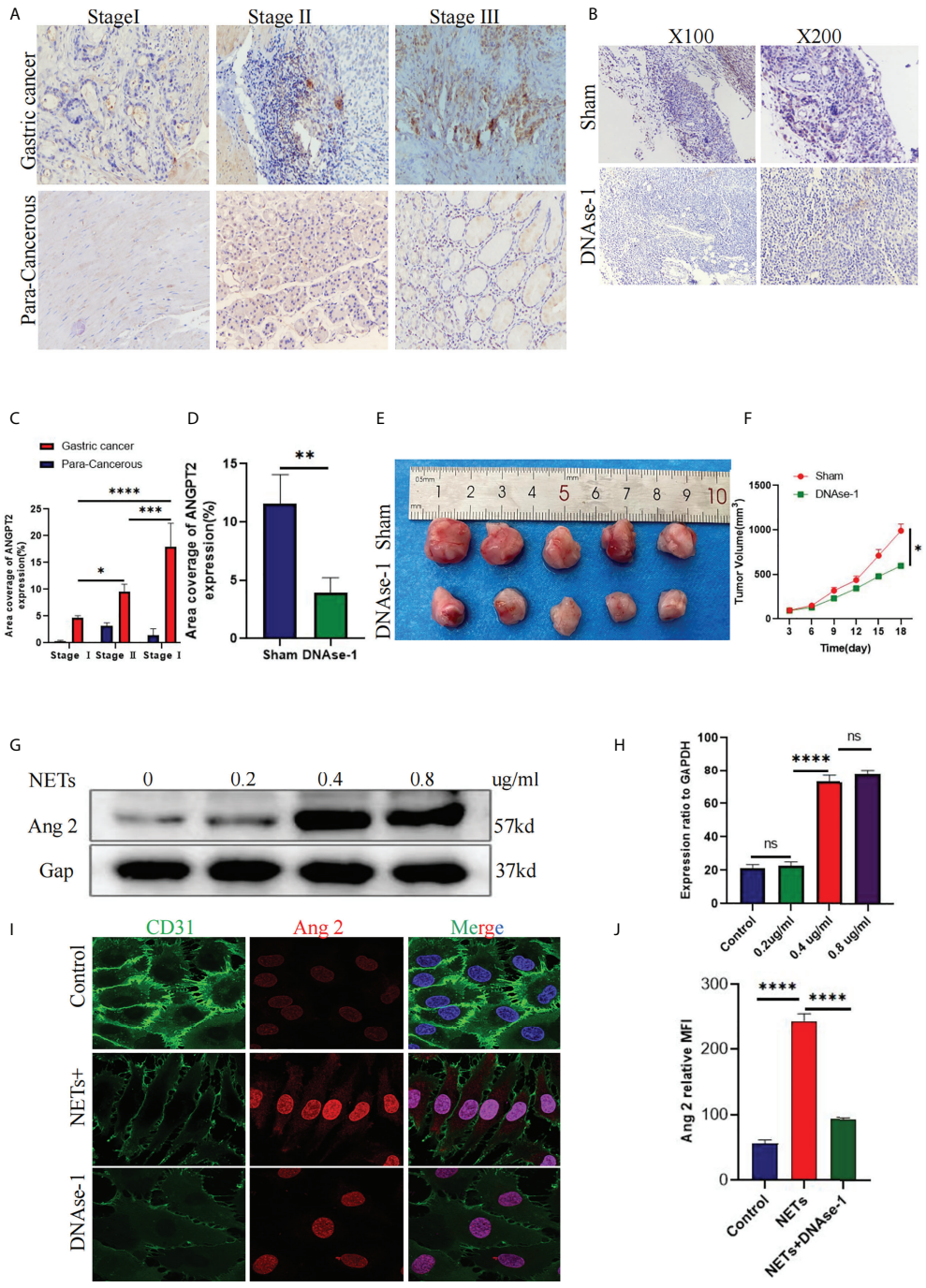


FIGURE 7
ANGPT2 expression in NET-stimulated HUVECs. **(A)** Immunohistochemical analysis of ANGPT2 expression in gastric cancer and paracancerous tissues of different stages. **(B)** ANGPT2 expression in mouse subcutaneous tumor tissue was analyzed by immunohistochemical staining. **(C)** Quantitative analysis of ANGPT2 in gastric cancer and paracancerous tissues. **(D)** Quantitative analysis of ANGPT2 in mouse subcutaneous tumor tissues. **(E)** Subcutaneous tumor samples from nude mice treated with DNAse-1 and control group, n=5. **(F)** Growth curve of subcutaneous tumor in nude mice, all values are mean ± SD. *p < 0.05 by two-way ANOVA. **(G)** Western blot analysis of ANGPT2 expression after HUVECs were stimulated by different concentrations of NETs. **(H)** Quantitative analysis of mRNA of ANGPT2. **(I)** ANGPT2 expression in HUVECs stimulated by PBS, NETs, and NETs+DNAse-1 was analyzed by immunofluorescence. **(J)** Tthe expression is indicated as the mean fluorescence intensity (MFI),analyzed with Image J software. All values are mean ± SD. ****P < 0.0001, ***P < 0.001,** P < 0.01, *P < 0.05 and ns, not significant by one-way ANOVA.

angiogenic factor and tumor cell proliferation activity. The ANGPT2 expression can promote colorectal cancer growth. Therefore, the intervention of the ANGPT2 mechanism system is considered a measure to treat tumors.

Conclusion

In this study, we used bioinformatics to predict the significance of ANGPT2 in GC. We found that ANGPT2 is a carcinogenic factor which is closely related to the tumor microenvironment in GC. It also has guiding significance with immunotherapy and chemotherapy, with a certain reference significance for clinical application.

Our method used three normal HUVECs samples and three NETs-stimulated HUVECs samples for proteomic analysis. The sample size selection was statistically significant. Our study established the relationship between NETs and HUVECs in the TME, formed a molecular network of neutrophils, NETs, and ANGPT2, and gave GC a new biomarker reference, ultimately providing a theoretical reference for the development of new therapeutic targets. These findings also support research to determine how ANGPT2-related biomarkers contribute to personalized GC treatment.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by this study was approved by the Internal Audit and Ethics Committee of the Second Affiliated Hospital of Harbin Medical University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by the Internal Audit and Ethics Committee of the Second Affiliated Hospital of Harbin Medical University.

Author contributions

SY, XZ and JL designed the study, completed the experiments, and drafted the manuscript. HY, AZ and YZ collected the patient clinical data and performed part of the experiments. LeZ and LiZ participated in the animal experiments. All authors have read and approved the final manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1010434/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The genomic alterations in high-ANGPT2 and low-ANGPT2 groups. **(A, B)** Amplification and deletions in gastric cancer with high and low ANGPT2 expression. **(C, D)** The detailed amplification and deletion of chromosome copy number variation in gastric cancer with high and low ANGPT2 expression.

SUPPLEMENTARY FIGURE 2
The somatic mutations in high-ANGPT2 and low-ANGPT2groups. **(A)** The forest plot illustrates the difference in mutation patterns between high-ANGPT2 and low-ANGPT2 groups. **(B, C)** The heatmap presents the somatic interaction in gliomas with high and low ANGPT2 levels.

SUPPLEMENTARY FIGURE 3
Analysis of the correlation between ANGPT2 and other cellular immune infiltration. The lollipop illustrates the difference in immune infiltration between high-ANGPT2 and low-ANGPT2 groups.

References

1. Lyons K, Le LC, Pham YT, Borron C, Park JY, Tran CTD, et al. Gastric cancer: epidemiology, biology, and prevention: A mini review. *Eur J Cancer Prev* (2019) 28:397–412. doi: 10.1097/CEJ.0000000000000480

2. Seidlitz T, Chen YT, Uhlemann H, Scholch S, Kochall S, Merker SR, et al. Mouse models of human gastric cancer subtypes with stomach-specific CreERT2-mediated pathway alterations. *Gastroenterology* (2019) 157:1599–1614.e2. doi: 10.1053/j.gastro.2019.09.026

3. Liu LP, Sheng XP, Shuai TK, Zhao YX, Li B, Li YM. Helicobacter pylori promotes invasion and metastasis of gastric cancer by enhancing heparanase expression. *World J Gastroenterol* (2018) 24:4565–77. doi: 10.3748/wjg.v27.i22.3138

4. Jiang L, Chen Y, Min G, Wang J, Chen W, Wang H, et al. Bcl2-associated athanogene 4 promotes the invasion and metastasis of gastric cancer cells by activating the PI3K/AKT/NF-kappaB/ZEB1 axis. *Cancer Lett* (2021) 520:409–21. doi: 10.1016/j.canlet.2021.08.020

5. Christofides A, Strauss L, Yeo A, Cao C, Charest A, Boussiotis VA. The complex role of tumor-infiltrating macrophages. *Nat Immunol* (2022) 23:1148–56. doi: 10.1038/s41590-022-01267-2

6. Zhao F, Li Z, Dong Z, Wang Z, Guo P, Zhang D, et al. Exploring the potential of exosome-related LncRNA pairs as predictors for immune microenvironment, survival outcome, and microbiotain landscape in esophageal squamous cell carcinoma. *Front Immunol* (2022) 13:918154. doi: 10.3389/fimmu.2022.918154

7. Peng R, Liu S, You W, Huang Y, Hu C, Gao Y, et al. Gastric microbiome alterations are associated with decreased CD8+ tissue-resident memory T cells in the tumor microenvironment of gastric cancer. *Cancer Immunol Res* (2022) 26: CIR-22-0107. doi: 10.1158/2326-6066.CIR-22-0107

8. Liu H, Ni S, Wang H, Zhang Q, Weng W. Charactering tumor microenvironment reveals stromal-related transcription factors promote tumor carcinogenesis in gastric cancer. *Cancer Med* (2020) 9:5247–57. doi: 10.1002/cam4.3133

9. Grunberg N, Pevsner-Fischer M, Goshen-Lago T, Diment J, Stein Y, Lavon H, et al. Cancer-associated fibroblasts promote aggressive gastric cancer phenotypes *via* heat shock factor 1-mediated secretion of extracellular vesicles. *Cancer Res* (2021) 81:1639–53. doi: 10.1158/0008-5472.CAN-20-2756

10. Quaas A, Pamuk A, Klein S, Quantius J, Rehkaemper J, Barutcu AG, et al. Sex-specific prognostic effect of CD66b-positive tumor-infiltrating neutrophils (TANs) in gastric and esophageal adenocarcinoma. *Gastric Cancer* (2021) 24:1213–26. doi: 10.1007/s10120-021-01197-2

11. Li S, Cong X, Gao H, Lan X, Li Z, Wang W, et al. Tumor-associated neutrophils induce EMT by IL-17a to promote migration and invasion in gastric cancer cells. *J Exp Clin Cancer Res* (2019) 38:6. doi: 10.1186/s13046-019-1168-1

12. Hiramatsu S, Tanaka H, Nishimura J, Yamakoshi Y, Sakimura C, Tamura T, et al. Gastric cancer cells alter the immunosuppressive function of neutrophils. *Oncol Rep* (2020) 43:251–9. doi: 10.3892/or.2019.7410

13. Mutua V, Gershwin LJ. A review of neutrophil extracellular traps (NETs) in disease: Potential anti-NETs therapeutics. *Clin Rev Allergy Immunol* (2021) 61:194–211. doi: 10.1016/j.intimp.2021.108516

14. Zhang R, Su L, Fu M, Wang Z, Tan L, Chen H, et al. CD177(+) cells produce neutrophil extracellular traps to promote biliary atresia. *J Hepatol* (2022). doi: 10.1016/j.jhep.2022.06.015

15. Masucci MT, Minopoli M, Del Vecchio S, Carriero MV. The emerging role of neutrophil extracellular traps (NETs) in tumor progression and metastasis. *Front Immunol* (2020) 11:1749. doi: 10.3389/fimmu.2020.01749

16. Kaltenmeier C, Yazdani HO, Morder K, Geller DA, Simmons RL, Tohme S. Neutrophil extracellular traps promote T cell exhaustion in the tumor microenvironment. *Front Immunol* (2021) 12:785222. doi: 10.3389/fimmu.2021.785222

17. Langseth MS, Opstad TB, Bratseth V, Solheim S, Arnesen H, Pettersen AA, et al. Markers of neutrophil extracellular traps are associated with adverse clinical outcome in stable coronary artery disease. *Eur J Prev Cardiol* (2018) 25:762–9. doi: 10.1177/2047487318760618

18. Goggs R, Jeffery U, LeVine DN, Li RHL. Neutrophil-extracellular traps, cell-free DNA, and immunothrombosis in companion animals: A review. *Vet Pathol* (2020) 57:6–23. doi: 10.1177/0300985819861721

19. Shahzad MH, Feng L, Su X, Brassard A, Dhoparee-Doomah I, Ferri LE, et al. Neutrophil extracellular traps in cancer therapy resistance. *Cancers (Basel)* (2022) 14(5):1359. doi: 10.3390/cancers14051359

20. Muller WEG, Ackermann M, Wang S, Neufurth M, Munoz-Espi R, Feng Q, et al. Inorganic polyphosphate induces accelerated tube formation of HUVEC endothelial cells. *Cell Mol Life Sci* (2018) 75:21–32. doi: 10.1007/s00018-017-2601-2

21. Langiu M, Palacios-Acedo AL, Crescence L, Mege D, Dubois C, Panicot-Dubois L. Neutrophils, cancer and thrombosis: The new Bermuda triangle in cancer research. *Int J Mol Sci 23* (2022) 23(3):1257. doi: 10.3390/ijms23031257

22. Zhang Y, Wang C, Li W, Tian W, Tang C, Xue L, et al. Neutrophil cyto-pharmaceuticals suppressing tumor metastasis *via* inhibiting hypoxia-inducible factor-1alpha in circulating breast cancer cells. *Adv Healthc Mater* (2022) 11: e2101761. doi: 10.1002/adhm.202101761

23. Mauracher LM, Posch F, Martinod K, Grilz E, Daullary T, Hell L, et al. Citrullinated histone H3, a biomarker of neutrophil extracellular trap formation, predicts the risk of venous thromboembolism in cancer patients. *J Thromb Haemost* (2018) 16:508–18. doi: 10.1111/jth.13951

24. Jary M, Hasanova R, Vienot A, Asgarov K, Loyon R, Tirole C, et al. Molecular description of ANGPT2 associated colorectal carcinoma. *Int J Cancer* (2020) 147:2007–18. doi: 10.1002/ijc.32993

25. Huang YH, Lian WS, Wang FS, Wang PW, Lin HY, Tsai MC, et al. MiR-29a curbs hepatocellular carcinoma incidence *via* targeting of HIF-1alpha and ANGPT2. *Int J Mol Sci 23* (2022) 23(3):1636. doi: 10.3390/ijms23031636

26. Zhao Y, Fu B, Chen P, Li Q, Ouyang Q, Zhang C, et al. Activated mesangial cells induce glomerular endothelial cells proliferation in rat anti-Thy-1 nephritis through VEGFA/VEGFR2 and Angpt2/Tie2 pathway. *Cell Prolif* (2021) 54:e13055. doi: 10.1111/cpr.13055

27. Karabid NM, Wiedemann T, Gulde S, Mohr H, Segaran RC, Geppert J, et al. Angpt2/Tie2 autostimulatory loop controls tumorigenesis. *EMBO Mol Med* (2022) 14:e14364. doi: 10.15252/emmm.202114364

28. Bergqvist C, Safi R, El Hasbani G, Abbas O, Kibbi A, Nassar D. Neutrophil extracellular traps are present in immune-complex-mediated cutaneous small vessel vasculitis and correlate with the production of reactive oxygen species and the severity of vessel damage. *Acta Derm Venereol* (2020) 100:adv00281. doi: 10.2340/00015555-3363

29. Cho JH, Fraser IP, Fukase K, Kusumoto S, Fujimoto Y, Stahl GL, et al. Human peptidoglycan recognition protein s is an effector of neutrophil-mediated innate immunity. *Blood* (2005) 106:2551–8. doi: 10.1182/blood-2005-02-0530

30. Demkow U. Neutrophil extracellular traps (NETs) in cancer invasion, evasion and metastasis. *Cancers (Basel)* (2021) 13(17):4495. doi: 10.3390/cancers13174495

31. Martins-Cardoso K, Almeida VH, Bagri KM, Rossi MID, Mermelstein CS, Konig S, et al. Neutrophil extracellular traps (NETs) promote pro-metastatic phenotype in human breast cancer cells through epithelial-mesenchymal transition. *Cancers (Basel)* (2020) 12(6):1542. doi: 10.3390/cancers12061542

32. Yang C, Wang Z, Li L, Zhang Z, Jin X, Wu P, et al. Aged neutrophils form mitochondria-dependent vital NETs to promote breast cancer lung metastasis. *J Immunother Cancer 9* (2021) 9(10):e002875. doi: 10.1136/jitc-2021-002875

33. Li T, Ma R, Zhang Y, Mo H, Yang X, Hu S, et al. Arsenic trioxide promoting ETosis in acute promyelocytic leukemia through mTOR-regulated autophagy. *Cell Death Dis* (2018) 9:75. doi: 10.1038/s41419-017-0018-3

34. Yang W, Wang P, Cao P, Wang S, Yang Y, Su H, et al. Hypoxic *in vitro* culture reduces histone lactylation and impairs pre-implantation embryonic development in mice. *Epigenet Chromatin* (2021) 14:57. doi: 10.1186/s13072-021-00431-6

35. Zhu T, Zou X, Yang C, Li L, Wang B, Li R, et al. Neutrophil extracellular traps promote gastric cancer metastasis by inducing epithelialmesenchymal transition. *Int J Mol Med 48* (2021) 48(1):127. doi: 10.3892/ijmm.2021.4960

36. McDonald B, Urrutia R, Yipp BG, Jenne CN, Kubes P. Intravascular neutrophil extracellular traps capture bacteria from the bloodstream during sepsis. *Cell Host Microbe* (2012) 12:324–33. doi: 10.1016/j.chom.2012.06.011

37. Wang C, Yu M, Zhou P, Li B, Liu Y, Wang L, et al. Endothelial damage and a thin intercellular fibrin network promote haemorrhage in acute promyelocytic leukaemia. *EBioMedicine* (2020) 60:102992. doi: 10.1016/j.ebiom.2020.102992

38. Chen Y, Quan L, Jia C, Guo Y, Wang X, Zhang Y, et al. Proteomics-based approach reveals the involvement of SERPINB9 in recurrent and relapsed multiple myeloma. *J Proteome Res* (2021) 20:2673–86. doi: 10.1021/acs.jproteome.1c00007

39. Onetti Y, Kalin RE, Pitter B, Hou M, Arribas V, Glass R, et al. Deletion of endothelial alpha-parvin inhibits tumour angiogenesis, reduces tumour growth and induces tumour cell apoptosis. *Angiogenesis* (2022) 25:155–8. doi: 10.1007/s10456-021-09829-y

40. Zhang K, Liu D, Zhao J, Shi S, He X, Da P, et al. Nuclear exosome HMGB3 secreted by nasopharyngeal carcinoma cells promotes tumour metastasis by inducing angiogenesis. *Cell Death Dis* (2021) 12:554. doi: 10.1038/s41419-021-03845-y

41. Lechertier T, Reynolds LE, Kim H, Pedrosa AR, Gomez-Escudero J, Munoz-Felix JM, et al. Pericyte FAK negatively regulates Gas6/Axl signalling to suppress tumour angiogenesis and tumour growth. *Nat Commun* (2020) 11:2810. doi: 10.1038/s41467-020-16618-6

42. Tavora B, Mederer T, Wessel KJ, Ruffing S, Sadjadi M, Missmahl M, et al. Tumoural activation of TLR3-SLIT2 axis in endothelium drives metastasis. *Nature* (2020) 586:299–304. doi: 10.1038/s41586-020-2774-y

43. Blanco-Gomez A, Hontecillas-Prieto L, Corchado-Cobos R, Garcia-Sancha N, Salvador N, Castellanos-Martin A, et al. Stromal SNAI2 is required for ERBB2 breast cancer progression. *Cancer Res* (2020) 80:5216–30. doi: 10.1158/0008-5472.CAN-20-0278

44. Schulz P, Fischer C, Detjen KM, Rieke S, Hilfenhaus G, von Marschall Z, et al. Angiopoietin-2 drives lymphatic metastasis of pancreatic cancer. *FASEB J* (2011) 25:3325–35. doi: 10.1096/fj.11-182287

45. Koga K, Todaka T, Morioka M, Hamada J, Kai Y, Yano S, et al. Expression of angiopoietin-2 in human glioma cells and its role for angiogenesis. *Cancer Res* (2001) 61:6248–54. doi: 10.1002/1097-0142(20010815)92:4+<1055::AIDCNCR1420>3.0.CO;2-E

46. Hacker UT, Escalona-Espinosa L, Consalvo N, Goede V, Schiffmann L, Scherer SJ, et al. Evaluation of angiopoietin-2 as a biomarker in gastric cancer: Results from the randomised phase III AVAGAST trial. *Br J Cancer* (2016) 114:855–62. doi: 10.1038/bjc.2016.30

47. Engin H, Ustundag Y, Tekin IO, Gokmen A, Ertop S, Ilikhan SU. Plasma concentrations of angiopoietin-1, angiopoietin-2 and tie-2 in colon cancer. *Eur Cytokine Netw* (2012) 23:68–71. doi: 10.1684/ecn.2012.0308

48. Chen Y, Wu Y, Zhang X, Zeng H, Liu Y, Wu Q, et al. Angiopoietin-2 (Ang-2) is a useful serum tumor marker for liver cancer in the Chinese population. *Clin Chim Acta* (2018) 478:18–27. doi: 10.1016/j.cca.2017.12.017

49. Helfrich I, Edler L, Sucker A, Thomas M, Christian S, Schadendorf D, et al. Angiopoietin-2 levels are associated with disease progression in metastatic malignant melanoma. *Clin Cancer Res* (2009) 15:1384–92. doi: 10.1158/1078-0432.CCR-08-1615

50. Srirajaskanthan R, Dancey G, Hackshaw A, Luong T, Caplin ME, Meyer T. Circulating angiopoietin-2 is elevated in patients with neuroendocrine tumours and correlates with disease burden and prognosis. *Endocr Relat Cancer* (2009) 16:967–76. doi: 10.1677/ERC-09-0089

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fimmu.2022.1007426

Check for updates

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Xuyu Gu,
Southeast University, China
Qi Zhang,
Huazhong University of
Science and Technology, China

*CORRESPONDENCE
Wenjie Song
wjsong@fmmu.edu.cn

SPECIALTY SECTION
This article was submitted to
Cancer Immunity
and Immunotherapy,
a section of the journal
Frontiers in Immunology

RECEIVED 30 July 2022
ACCEPTED 24 August 2022
PUBLISHED 14 September 2022

CITATION
Jia W, Yao Q, Wang Y, Mao Z,
Zhang T, Li J, Nie Y, Lei X, Shi W and
Song W (2022) Protective effect of
tertiary lymphoid structures against
hepatocellular carcinoma: New
findings from a genetic perspective.
Front. Immunol. 13:1007426.
doi: 10.3389/fimmu.2022.1007426

COPYRIGHT
© 2022 Jia, Yao, Wang, Mao, Zhang, Li,
Nie, Lei, Shi and Song. 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.

Protective effect of tertiary lymphoid structures against hepatocellular carcinoma: New findings from a genetic perspective

Weili Jia[1,2], Qianyun Yao[1], Yanfang Wang[1,2], Zhenzhen Mao[1,2], Tianchen Zhang[1,2], Jianhui Li[3], Ye Nie[2], Xinjun Lei[1,2], Wen Shi[1,2] and Wenjie Song[2]*

[1]Xi'an Medical University, Xi'an, China, [2]Department of Hepatobiliary Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China, [3]Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China

Background: Tertiary lymphoid structures (TLS) have an effect on hepatocellular carcinoma (HCC), but the underlying mechanism remains to be elucidated.

Methods: Intratumoral TLS (iTLS) was classified in the Cancer Genome Atlas-Liver Hepatocellular Carcinoma (TCGA-LIHC) cohort using pathological sections from the Cancer Digital Slide Archive. Univariate and multivariate Cox regression analyses were performed to validate the effect of iTLS on overall survival (OS), relapse-free survival (RFS), and disease-free survival (DFS). The genes differentially expressed between the iTLS-negative and iTLS-positive groups were analyzed in combination with sequencing data. Gene set enrichment analysis (GSEA) was used to explore the signaling pathways affected by these differentially expressed genes. The random forest algorithm was used to identify genes with the highest correlation with the iTLS in the training set. Multivariate logistic regression was used to build a model to predict iTLS in tissue samples. Spearman's correlation was used to analyze the relationship between TLS-associated chemokines and signature genes, and CIBERSORT was used to calculate immune infiltration scores. Copy number variation and its relationship with immune cell infiltration and signature genes were assessed using the gene set cancer analysis (GSCA). The Correlation R package was used for gene ontology (GO), disease ontology (DO), and gene mutation analyses. The GSCA was used for drug sensitivity analysis. LASSO regression was used to build prognostic models, and external data were used to validate the models.

Results: There were 218 positive and 146 negative samples for iTLS. iTLS was significantly associated with better RFS and DFS according to Cox regression analysis. Twenty signature genes that were highly associated with iTLS positivity were identified. GO and mutation analyses revealed that the signature genes were associated with immunity. Most signature genes were sensitive to

immune checkpoint inhibitors. Risk scores calculated using a characteristic gene-based prognostic model were found to be an independent prognostic factor for OS.

Conclusions: The improvement of RFS in HCC by iTLS was not limited to the early period as previously reported. iTLS improved DFS in patients. Characteristic genes are closely related to the formation of iTLS and TLS chemokines in HCC. These genes are closely related to immunity in terms of cellular infiltration, biological functions, and signaling pathways. Most are sensitive to immune checkpoint inhibitors, and their expression levels can affect prognosis.

KEYWORDS

tertiary lymphoid structures (TLS), hepatocellular carcinoma (HCC), immunotherapy, cancer prognosis, immune microenvironment (IME)

1 Introduction

As of 2020, hepatocellular carcinoma (HCC) was the sixth leading cause of cancer-related deaths, making it one of the world's leading public health problems (1). The incidence of HCC continues to increase annually. According to World Health Organization estimates, by 2040 the number of new cases and deaths will exceed 1.4 million and 1.3 million, respectively (2). Tertiary lymphoid structures (TLS), also known as tertiary lymphoid organs (TLO) and ectopic lymphoid structures (ELS), are aggregates of lymphocytes capable of providing ectopic hubs for the acquired immune response, and can affect various disease outcomes (3). TLS is acquired and often formed under the stimulation of chronic inflammation to address the invasion of various pathogenic factors (4). The first study on TLS was conducted on non-small-cell lung cancer (NSCLC). Seventy-four patients were studied using immunohistochemistry in the early stages (5). Subsequently, the amount of TLS-related literature has increased annually, and more scientists are involved in the research of TLS every year.

Most patients with HCC are not diagnosed until in the advanced stages, thus missing the optimal window for treatment (6). The insensitivity of HCC to conventional malignancy therapies has led to the emergence of immunotherapy as one of the most promising treatments (7). As an immune structure, TLS is gaining attention from researchers worldwide. There is growing evidence of its direct and indirect impact on HCC outcomes. While TLS in cancer is generally protective, two contrasting effects have been reported in HCC: intratumoral TLS (iTLS) can be protective against HCC, while peritumoral TLS (pTLS) can be detrimental for HCC. iTLS may be associated with sustained and effective anti-tumor immunity (8–12). Additionally, more mature iTLS could help improve patient prognosis (9). Conversely, there is evidence that pTLS can promote HCC development (13, 14). TLS found in excised non-neoplastic liver tissue surrounding HCC is associated with poor prognosis and increased prevalence (9, 14). In addition, studies have shown that depletion of TLS in non-neoplastic liver parenchyma can inhibit cancer progression (14). However, a recent study showed that patients with HCC and higher pTLS densities had better overall survival (OS) and relapse-free survival (RFS) (15).

Few studies have analyzed TLS at the genetic level, particularly in HCC. Therefore, we performed a bioinformatic analysis in combination with hematoxylin and eosin (HE) pathologically stained sections to explore and elucidate the genetic characteristics of iTLS in HCC and the prognostic implications of differential expression of genes related to iTLS in patients. Together with previous studies, we believe that our study will contribute to further understanding of the role of iTLS in HCC.

2 Materials and methods

2.1 Data acquisition and processing

XENA (https://xena.ucsc.edu) is an online discovery tool for public and private multi-omics and clinical/phenotypic data (16). We used this tool to obtain the liver hepatocellular carcinoma (LIHC) dataset from The Cancer Genome Atlas (TCGA) in the RNAseq count format. The data were processed as follows: 1) data from the same sample but from different tables were averaged; 2) data from different samples were combined to form a genomic

matrix; and 3) a log2(x+1) transformation was performed on all data. Data from the International Cancer Genome Consortium (ICGC) cohort were obtained from the official website. We downloaded the LIRI-JP data from the ICGC along with the corresponding clinical information of the cohort. Please refer to Supplementary Table 1 for a summary of the relevant data.

2.2 Evaluation of iTLS in pathological sections

We evaluated the density of lymphocyte infiltration by retrieving HE pathologically stained sections of the corresponding TCGA samples from the Cancer Digital Slide Archive (CDSA). Frozen sections and formalin-fixed paraffin-embedded (FFPE) tissue sections were used. This study used methods based on Clarice et al. (17) for counting all forms of iTLS as follows: 1) lymphocyte aggregates (Agg) with lymphocyte infiltration but no lymphoid follicle formation; 2) primary follicles (FL1), with well-defined clusters of round or oval lymphocytes or plasma cells (no germinal centers present); and 3) secondary follicles (FL2), with well-defined clusters of round or oval lymphocytes or plasma cells (germinal centers present). Based on the above groupings, two pathologists independently evaluated iTLS in all HCC samples. Subsequently, a third pathologist assisted in identifying the conflicting results. Then, based on the technique of Clarice et al. (17), samples with at least one occurrence of any form of iTLS (Agg, FL1, FL2) were categorized as the TLS-positive (TLS+) group and the samples without any occurrence of iTLS were categorized as the TLS-negative (TLS-) group for the next analysis.

2.3 Analysis of the impact of iTLS on the prognosis of HCC

Previously, Li et al. (11) and Calderaro et al. (8) found the significance of iTLS in HCC for better early stage RFS in patients, but further validation was needed to support this conclusion. Therefore, to verify their findings or to discover new information, we plotted the iTLS grouping status (positive or negative) in relation to OS, RFS, and disease-free survival (DFS) to determine the prognostic impact of iTLS after excluding patients with incomplete follow-up information. OS was defined as the time from the start of follow-up until the patient died or was lost to follow-up for any reason, RFS was defined as the time from the start of follow-up until the patient experienced a disease recurrence, and DFS was defined as the time from the start of follow-up until the patient died, was lost to follow-up, or experienced a disease recurrence for any reason. Univariate and multivariate Cox regression analyses were performed to determine whether the iTLS was an independent prognostic factor. Indicators with significant results ($p < 0.05$) in the univariate analysis were included in the multivariate analysis.

2.4 Analysis of gene expression differences

To investigate the differential gene expression between the two groups, we selected intra-tumor pathology samples with both HE stained sections and RNAseq expression data, and then performed differential expression analysis using the R package "edgeR" according to the TLS grouping (TLS+ or TLS-). First, samples with zero expression were excluded. Differential expression analysis was performed to explore differences between the two groups. Finally, gene set enrichment analysis (GSEA) was performed using the KEGG and Reactome databases to clarify the signaling pathways in the locations of the differential genes.

2.5 Identification of key genes associated with iTLS

The Boruta algorithm was used to identify the key differential genes affecting iTLS formation. Boruta is a feature selection algorithm, which is specifically a wrapper algorithm for random forests that filters out the set of all features correlated with the dependent variable (18). It can manage a large number of input variables, evaluate the importance of the variables during processing, and has been used in many studies in cancer-causative gene analysis (19, 20). To explore all potential genes that may have an impact on the formation of iTLS, we included all genes in which differential expression analysis showed significant results ($p < 0.05$) in the subsequent study.

We first used the Boruta algorithm to rank the differential genes from largest to smallest according to their effects on iTLS. The top 20 results were selected as iTLS signature genes, which were later used to build a logistic regression model for predicting the presence of iTLS. The specific steps were as follows: 1) the samples were randomly divided into training and test groups; the training group was used to select the key genes affecting the formation of iTLS and the test group was used to validate the results from the training group; 2) all differential gene expressions in the training group were characterized using the R package "Boruta" to identify the key categorical variables; 3) five 10-fold crossover validations were performed using the R package "caret" to find the most accurate mtry value (mtry refers to the number of variables randomly sampled when constructing decision tree branches in random forest modeling, and an appropriate mtry value can reduce the prediction error rate of the random forest model); and 4) the final selected model was extracted and receiver operating characteristic (ROC) curves were created using the validation set data to verify the prediction ability of the model situation to establish a multivariate logistic regression model. ROC curves were plotted for all samples to further validate the predictive ability of the model. Previous

studies used several genetic traits relevant to TLS to assess TLS in tissues. Results showed that these traits had different combinations, such as the 9-TLS trait (21), 12-TLS trait (22), 40-TLS trait (23), and 50-TLS trait (24). We selected these chemokines, developed correlation prediction models, and compared them with our model.

2.6 Immune cell infiltration analysis

iTLS belongs to the immune structure family. Therefore, one can hypothesize that there is a correlation between iTLS and immune cell infiltration. Hence, we calculated the immune infiltration score for each patient's cancer tissue using CIBERSORT. The relationship between the signature genes and the immune infiltration score was then analyzed using Spearman correlation analysis. Finally, the effect of copy number variation (CNV) on immune cell infiltration was explored using gene set cancer analysis (GSCA).

2.7 Biological functional analysis

The R package "clusterProfiler" supports the functional characterization of thousands of coding and non-coding genomic data with up-to-date gene annotations (25). It provides a unified interface for gene function annotations from a variety of sources and can therefore be applied to a variety of scenarios. We used this program for gene ontology (GO) biological processes, GO cellular composition, GO molecular function, and disease ontology (DO) analysis.

2.8 Single nucleotide polymorphism (SNP) and CNV analysis

To understand the mutation of signature genes within tumors, we investigated single nucleotide polymorphism (SNP) mutations in signature genes using the R package "maftools", while exploring the signaling pathways affected by the mutations. In addition, using GSCALite, a web-based platform for genomic cancer analysis (26), we investigated the CNV of the signature genes.

2.9 Drug sensitivity analysis

The ultimate goal of medical research is to facilitate clinical treatment; therefore, we explored the relationship between signature genes and drug sensitivity using GSCA. The tool contains data from both the GDSC and Clinical Trials Reporting Program (CTRP) databases. GDSC (27) characterized 1000 human cancer cell lines and screened them for more than 100 compounds; CTRP (28) has similar characterization and screening of data. We further downloaded gene expression and drug sensitivity data from CellMiner, a web-based suite of genomics and pharmacology tools (29). Subsequently, drug sensitivity data were screened by selecting drugs that were validated by clinical trials and Food and Drug Administration (FDA) approval. Finally, the expression data of the characterized genes were subjected to the Spearman correlation test with the drug sensitivity data to obtain and visualize the correlation data between them.

2.10 Construction and validation of prognostic model

To explore the prognostic value of signature genes, LASSO regression was used to further screen genes from the signature genes that clearly affect the prognosis of HCC. The optimal signature model was then constructed based on the Akira pooling information criterion (AIC), and the result with the lowest AIC value was used to construct the signature: risk score = expression (A) × c of (A) + expression(B) × c of (B) +. expression(n) × c of(n). OS, DFS, and RFS curves were plotted based on risk scores using the Kaplan-Meier method. The accuracy of the model was verified using ROC curves. Univariate and multivariate Cox regression analyses were used to identify the model as an independent influence on prognosis, and heatmaps of the risk score, relevant clinical indicators, and characteristic gene expression were plotted. Finally, we validated the predictive power of the model using the ICGC data as external data.

2.11 Statistical analysis

All data analysis, data visualization (graphical plots), and statistical analysis were performed using R Studio Desktop (version 4.1.2), unless otherwise specified. Logistic regression analysis was performed using the R software. Gene difference analysis was performed using R package "edgeR". The R package "GSVA" was used for the ssGSEA analysis. Correlation analysis was performed using the R package "Hmisc". SNP mutation analysis was performed using the R package "maftools". The R package "Boruta" was used for random forest analysis. Survival analysis was performed using the R package "survivor". The R package "survivalROC" was used to plot the time-dependent ROC curves. RFS analysis was performed using Kaplan-Meier plots and log-rank tests. Correlations between the two non-normal datasets were analyzed using Spearman's method. The Wilcoxon signed-rank test was used to evaluate between-group differences in pathological parameters. The cardinality test was used to analyze the relationship between the clinicopathological parameters and characteristics. P values below 0.05 were considered statistically significant if not otherwise stated.

3 Results

3.1 iTLS is associated with better RFS and DFS in patients

A total of 365 samples were observed; 137 samples were classified as the Agg group, 56 as the FL1 group, and 26 as the FL2 group. The remaining 146 samples had no lymphocytic infiltration (Figure 1). There were 218 samples in the final TLS+ group and 146 in the TLS- group (Supplementary Table 2). Results of OS, RFS, and DFS curves showed that the TLS+ group was associated with better RFS (p-values less than 0.001 at 1, 2, and 5 years), DFS ($p<0.001$, $p<0.001$, $p = 0.004$ at 1, 2, and 5 years, respectively), and 2-year OS ($p = 0.033$) (Figure 2).

Univariate Cox regression results for OS were not statistically significant ($p = 0.064$, Figure 3). Univariate and multivariate Cox regression results for both RFS and DFS showed that iTLS (TLS+) was a protective factor (Figure 4). Interestingly, in a slight departure from the results of previous studies (8, 11), we found that iTLS was not only associated with early RFS. Rather, at all periods (1, 2, and 5 years), iTLS was associated with better RFS. This further validates the beneficial effects of iTLS in patients with HCC.

3.2 Differential genes are associated with immune-dominated pathways

The results of the differential expression analysis showed that of the 1057 differential genes, most (625 genes) were downregulated, whereas the expression of the other 432 showed varying degrees of upregulation (Supplementary Table 3).

In the GSEA, the KEGG database showed that the differential genes were mainly distributed in "cell adhesion molecules", "chemokine signaling pathway", and "cytokine-cytokine receptor interaction" pathways (Figure 5A). Conversely, the Reactome database reported that the differentially expressed genes were mainly enriched in the "Adaptive Immune System", "Class A/l (Rhodopsin-like receptors)", and "Cytokine Signaling in Immune System" pathways. Clearly, most of these pathways are closely related to immunity (Figure 5B).

3.3 iTLS prediction model consisting of 20 genes

Using Boruta, we selected the most suitable mtry values (Figure 6A) and identified 24 important genes associated with iTLS (Figure 6B). A 10-fold-5 cross-validation was then performed on the training set, and the top 20 most important data points affecting the iTLS profile in the training set were selected to build the prediction model (Figure 6C). These 20 genes were: SYTL1, TMEM25, ARL4D, PITHD1, CCR7, LCK, CCDC88B, CCL21, CORO1A, RASAL3, LIMD2, COQ3, KCNE4, ITPRIP, DBT, CXCR3, SMIM3, CD3D, PSTPIP1, and PLAU.

The area under the curve (AUC) of the ROC curve obtained using the model in the validation set was 0.733 (Figure 6D). The AUC of the ROC curve of the iTLS prediction model built using multifactorial logistic regression was 0.782 for all samples

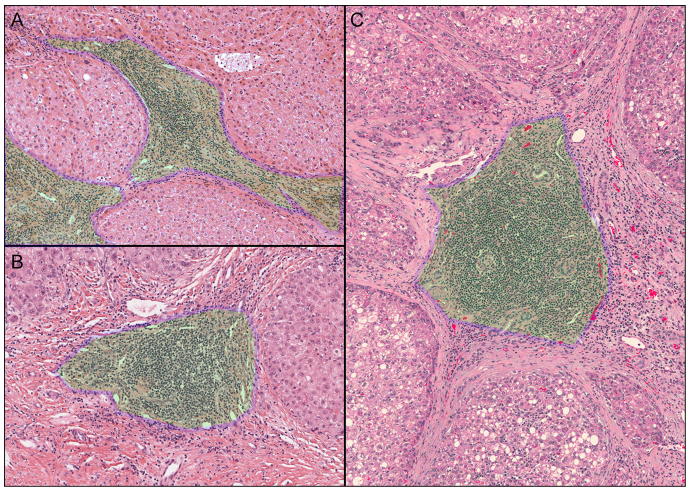


FIGURE 1
Observed iTLS images. The green area with the blue border marked in the figure is the iTLS. **(A)** Lymphocyte aggregation (Agg). **(B)** Primary lymphoid follicles (FL1). **(C)** Secondary lymphoid follicles (FL2).

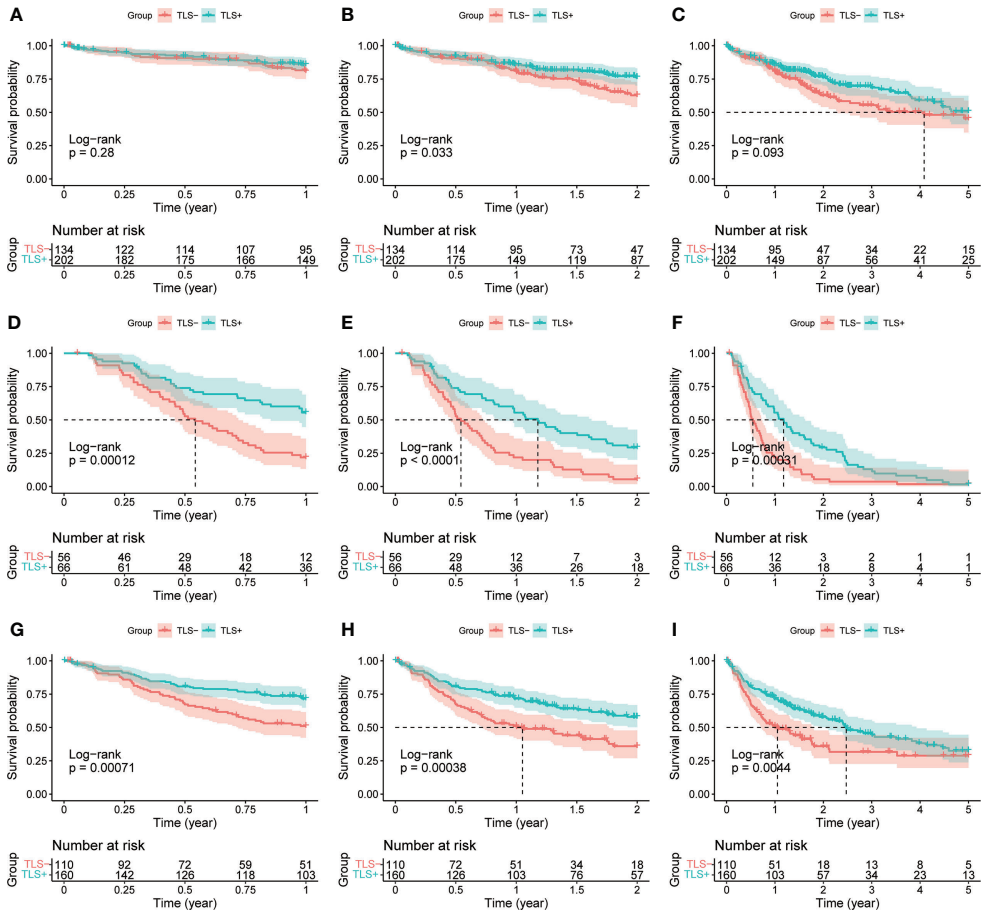


FIGURE 2
(A) 1-year OS curves of TLS-group vs. TLS+ group. **(B)** 2-year OS curves of TLS-group vs. TLS+ group. **(C)** 5-year OS curves of TLS-group vs. TLS+ group. **(D)** 1-year RFS curves of the TLS-group versus the TLS+ group. **(E)** 2-year RFS curves of the TLS-group versus the TLS+ group. **(F)** 5-year RFS curves of the TLS-group versus the TLS+ group. **(G)** 1-year DFS curves of the TLS-group versus the TLS+ group. **(H)** 2-year DFS curves of TLS-group vs. TLS+ group. **(I)** 5-year DFS curves of TLS-group vs. TLS+ group.

(Figure 6E). According to the TLS grouping (positive or negative), we plotted a gene expression heatmap of the characteristic genes in all HCC samples (Figure 6F). As the figure shows, with CCDC88B as the dividing line, the upper genes were concentrated in the TLS+ group with high expression, whereas the lower genes were concentrated in the TLS- group.

We also built prediction models using TLS-related chemokines to verify the accuracy of our models. The results showed that the AUCs of the 9-TLS, 12-TLS, 40-TLS, and 50-TLS prediction models were 0.668, 0.697, 0.771, and 0.792, respectively. The accuracy of our results is slightly higher than that of the 40-TLS and slightly lower than that of the 50-TLS. However, one problem that cannot be ignored is that when using the 40-TLS and 50-TLS features for logistic regression, the number of independent variables is too large. Therefore, the results may not be accurate. However, considering the above factors, the predictive ability of our model remains excellent.

3.4 Further exploration around signature genes

3.4.1 Signature genes are associated with TLS-associated chemokines

The Spearman correlation analysis using previously reported expression levels and the expression levels of our signature genes was performed to further validate our signature genes. The results showed that most of our signature genes had strong correlations with the aforementioned features, whether it was the classical 12-TLS feature or the 9-TLS feature proposed by Feng et al. (21), the 40-TLS feature proposed by Zhou et al. (23), or the 50-TLS feature used by Wu et al. (24) (Figure 7).

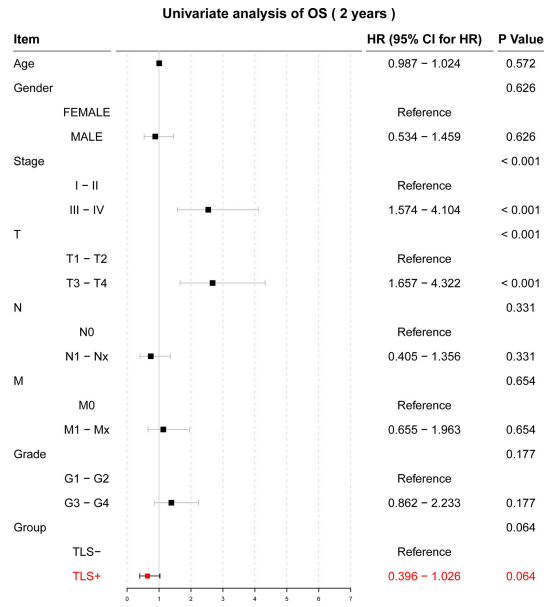


FIGURE 3
Results of the univariate Cox regression analysis with two-year OS as the outcome.

3.4.2 Signature genes are associated with immune infiltration

We plotted a heatmap of the correlation between the expression of the signature gene and calculated the immune score using CIBERSORT (Supplementary Table 4 and Figure 8A). Using CD4 T cells as the boundary, the level of immune cell infiltration above the boundary is roughly positively correlated with the signature gene, whereas the level of immune cell infiltration below the boundary is roughly negatively correlated with the signature gene (Figure 8A).

Serendipitously, using the GSCA (http://bioinfo.life.hust.edu.cn/GSCA/), we found a large number of CNV in the signature genes, most of which were heterozygous variants (Figure 8B). Among them, LIMD2 had the most heterozygous amplifications, whereas SYTL1 had the most heterozygous deletions. Based on the relationship between signature genes and immunity, we hypothesized that the CNV of signature genes might affect immune infiltration. Therefore, we explored the relationship between CNV and immune cell infiltration (Figure 8C). The figure shows that both gene amplification and deletion increase B-cell infiltration, whereas the infiltration levels of CD4_naïve, NK, and Macrophage are all diminished. The figure also shows that both amplification and deletion mutations cause a decrease in the immune infiltration score, which in turn affects immune infiltration (see infiltration score in Figure 8C).

In addition, we used the immune database ImmPort1 for further analysis to explore the relationship between signature genes and immune genes. The results showed that most of the immune genes correlated with our signature genes, with the strongest correlations being ARL4D and CCL (Supplementary Table 5 and Supplementary Figure 1).

3.4.3 Signature genes are associated with multiple biological functions and hepatobiliary diseases

GO enrichment analysis showed that, in terms of biological processes, the 20 signature genes were significantly associated with immune processes including "T cell activation", "positive regulation of T cell activation", and "positive regulation of leukocyte cell adhesion" ($p < 0.001$). Additionally, cell composition was associated with "immunological synapse", and molecular function was associated with "C-C chemokine receptor activity", "C-C chemokine binding", and "G protein-coupled chemoattractant receptor activity" ($p = 0.02$). The results of DO analysis were also associated with "hepatitis", "hepatitis C", and "primary biliary cirrhosis" (Figure 9).

3.4.4 Mutations in signature genes can affect cancer-related signaling pathways

SNP, mainly DNA sequence diversity caused by variants in a single nucleotide at the genomic level, can lead to the development of disease. We further investigated the SNP profiles of the characteristic genes and mapped the mutations between samples (Figure 10A).

The close relationship between cancer progression and signaling pathways is well known. Therefore, we investigated the signaling pathways affected by relevant signature genes in each group of samples (Figure 10B). RTK-RAS, a pathway known to influence cancer progression, ranked first among the affected pathways; therefore, we mapped the RTK-RAS pathway in terms of gene mutations (Figure 10C). Methylation is one of the first identified and most intensively studied epigenetic regulatory mechanisms that can influence the progression of many cancers. We explored the correlation between methylation of characteristic genes and mRNA expression between normal and cancer samples. The expression of CORO1A, LCK, and ARL4D showed a significant negative correlation with methylation (Figures 10D–F).

We further explored the relationship between signature genes and CNV using GSCALite. We also used this tool to investigate the effect of gene expression differences on pathway activation, where all mutations affected more than one signaling pathway (Figure 11).

3.4.5 Signature genes are sensitive to immune checkpoint inhibitors

Immune checkpoint inhibitors have been approved as conventional drugs for HCC, and the possibility of immunotherapy should be further investigated. We explored the correlation of signature genes with the sensitivity to GDSC and CTRP drugs in pancreatic cancer using the GSCA website. As seen in Figure 12, both the GDSC and CTRP databases showed that RASAL3 and COROIA

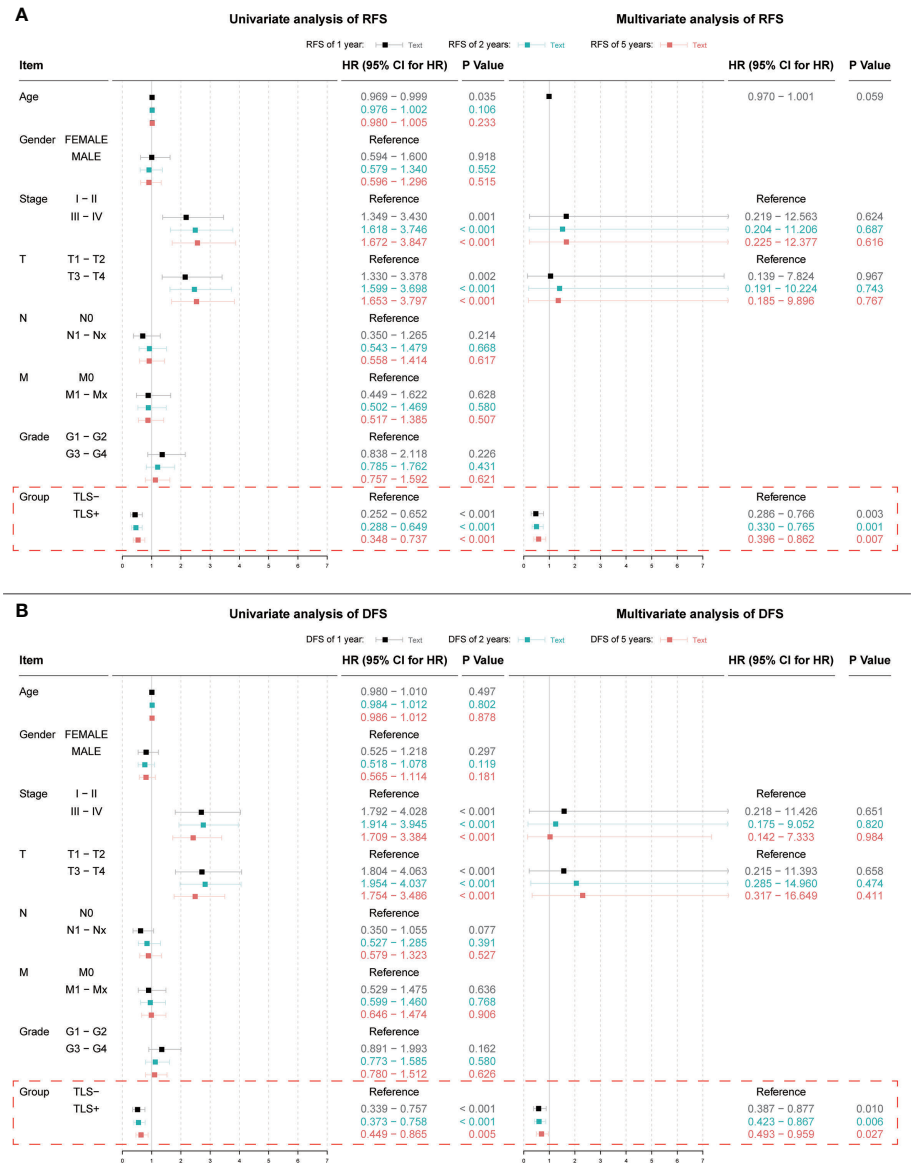


FIGURE 4
(A) Results of univariate and multivariate Cox regression analyses with 1, 2, and 5-year RFS as outcomes; **(B)** Results of univariate and multivariate Cox regression analyses with 1, 2, and 5-year DFS as outcomes. Indicators with significant ($p < 0.05$) results in the univariate Cox regression analysis are further included in the multivariate Cox regression analysis.

had the highest sensitivity to immune checkpoint inhibitors, suggesting that these genes may be potential therapeutic targets. LIMD2, PSTPIP1, CD3D, LCK, and CCR7 were significantly negatively correlated with the IC50 of most drugs, whereas CCDC88B, ITPRIP, SYTL1, PLAU, TMEM25, and ARL4D were correlated with some drugs to varying degrees. PITHD1 is not presented in the figure because of a lack of relevant data.

Using the CellMiner database, we evaluated the relationship between signature genes and drug IC50 (Supplementary Table 5). From this analysis, we selected the portion of the data presented in the figure with the most significant effect (Figure 13). The results showed that 20 signature genes may be promising potential drug targets for HCC and merit further in-depth study.

3.4.6 Signature genes have a significant impact on the prognosis of HCC

Given the protective effect of iTLS on HCC prognosis, we hypothesized that signature genes could be used to predict patient prognosis. Therefore, we used the LASSO regression to build a prognostic model based on a 20-trait gene screen (Figures 14A, B). Twelve genes were eliminated during the screening process and the final prognostic model was obtained as follows:

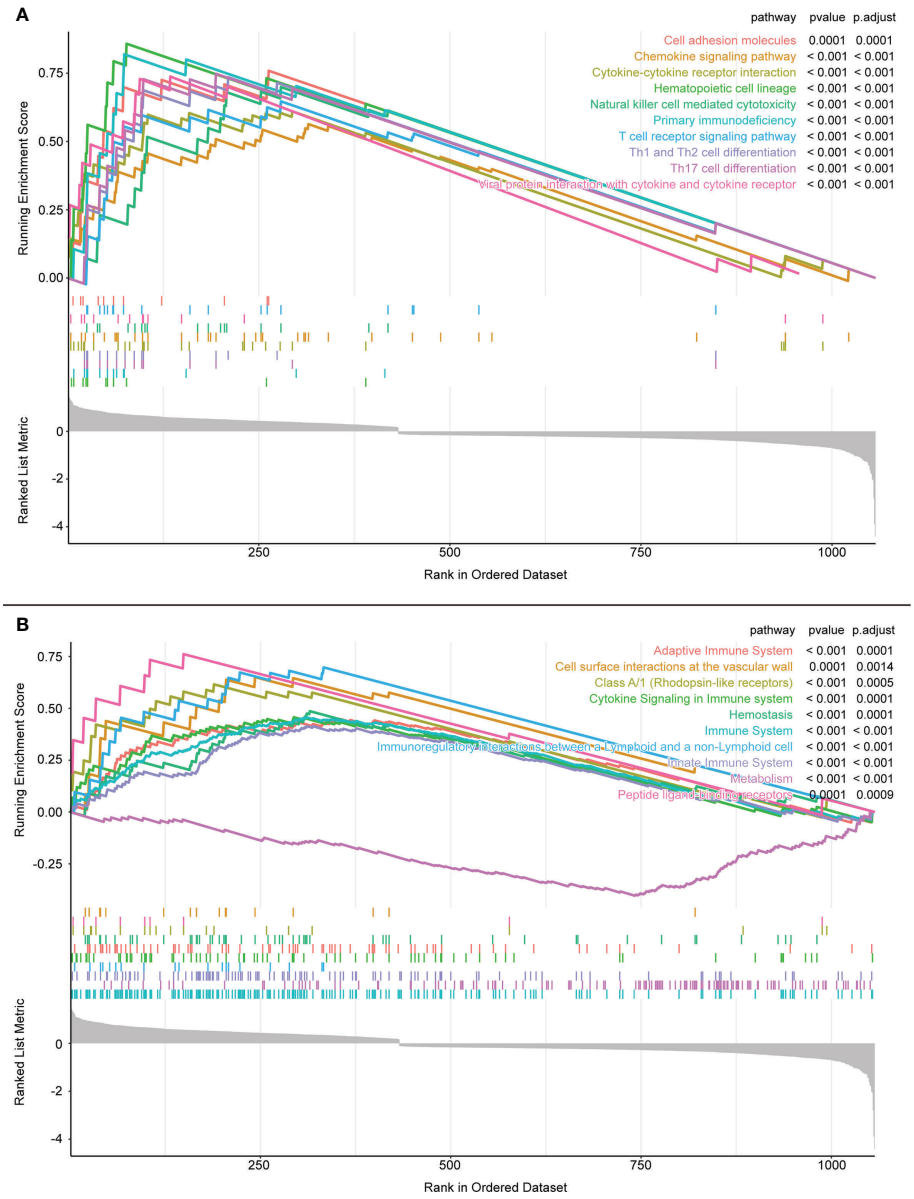


FIGURE 5
(A) GSEA enrichment analysis results of KEGG database. **(B)** GSEA enrichment analysis results of Reactome database.

0.109×exp(PITHD1)+-0.178×exp(RASAL3)+-0.053×exp(CCR7)+0.171×exp(COQ3)+-0.042×exp(PSTPIP1)+0.045×exp(KCNE4)+0.224×exp(CCDC88B)+0.106×exp(SMIM3)

After scoring each patient according to the prognostic model, the patients were divided into high-risk and low-risk groups according to the median score. The results showed statistically significant differences in OS, RFS, and DFS between patients in the high-risk and low-risk groups ($p < 0.0001$, $p = 0.0021$, and $p < 0.0001$, respectively; Figures 14C–E). The predictive ability of the model was assessed using time-dependent ROC curves, and the AUC was 0.654, 0.717, and 0.718 at 1, 3, and 5 years, respectively (Figure 14F). Univariate and multivariate Cox regression results also showed that the risk score was an independent prognostic factor ($p < 0.001$; Figure 15A).

To further validate the effect of this model, we performed an external validation using ICGC data. The results again showed a statistically significant difference in OS between patients in the high- and low-risk groups ($p = 0.0086$; Figure 14G). The time-dependent ROC curves had 1-, 3-, and 5-year AUCs of 0.723, 0.629, and 0.718, respectively; Figure 14H). Univariate Cox

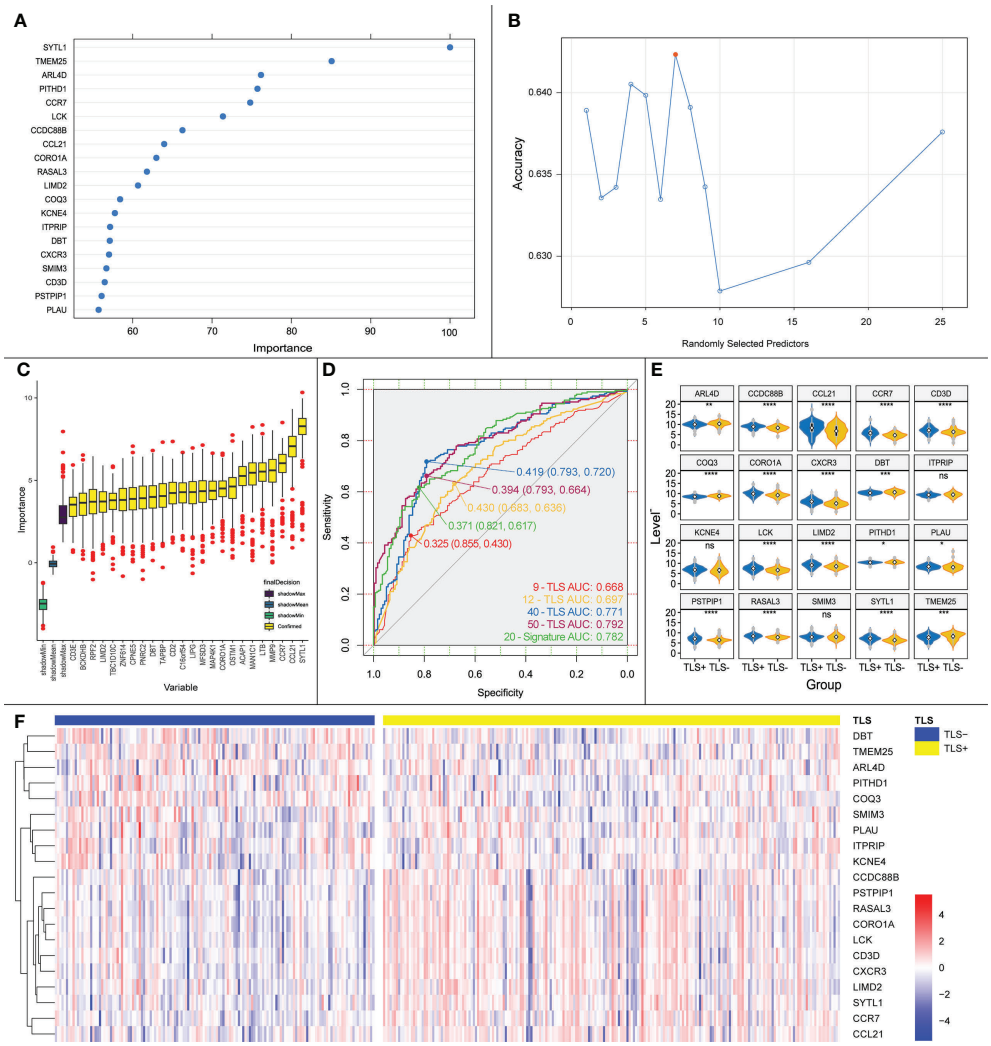


FIGURE 6
(A) Key genes influencing the presence or absence of iTLS screened by Boruta algorithm, horizontal coordinates are genes and vertical coordinates are the importance of the genes. **(B)** Schematic diagram of the screening process for mtry values, red points are the most appropriate values; horizontal coordinates are mtry values and vertical coordinates are the accuracy of the value. **(C)** Genes screened by Boruta's algorithm with the relative importance of the gene in the horizontal coordinate and the gene in the vertical coordinate. **(D)** ROC curves plotted using our proposed prediction model together with the previously reported TLS chemokines. **(E)** Expression of the 20 TLS signature genes obtained by screening in each group. Significance marks, ns: p > 0.05; *:p <= 0.05; **:p <= 0.01; ***:p <= 0.001; ****:p <= 0.0001. **(F)** Heatmap based on the expression of the 20 TLS signature genes obtained by screening in each sample, from which it can be seen that with CCDC88B as demarcation, the upper genes are up-regulated in expression in the TLS+ group, and the lower genes are the opposite.

regression results (p = 0.009) and multivariate Cox regression results (p = 0.014; Figure 15B) again showed that the risk score was an independent influencer of prognosis.

Finally, to explore the relationship between the risk score, prognostic model gene expression, clinicopathological parameters, and iTLS, we plotted a heatmap of gene expression in the prognostic model (Figure 14I). As seen in the figure, the risk score increases with higher gene expression of PITHD1, COQ3, KCNE4, SMIM3, and CCDC88B, whereas the opposite is true for CCR7, which may herald these genes as potential promising targets for immunotherapy. In addition, the figure clearly shows that the chance of iTLS emergence decreases with increasing risk scores.

4 Discussion

We replicated previous studies based on pathological tissue sections and survival data from patients with HCC; our findings differed slightly from that of previous studies. Previously, iTLS

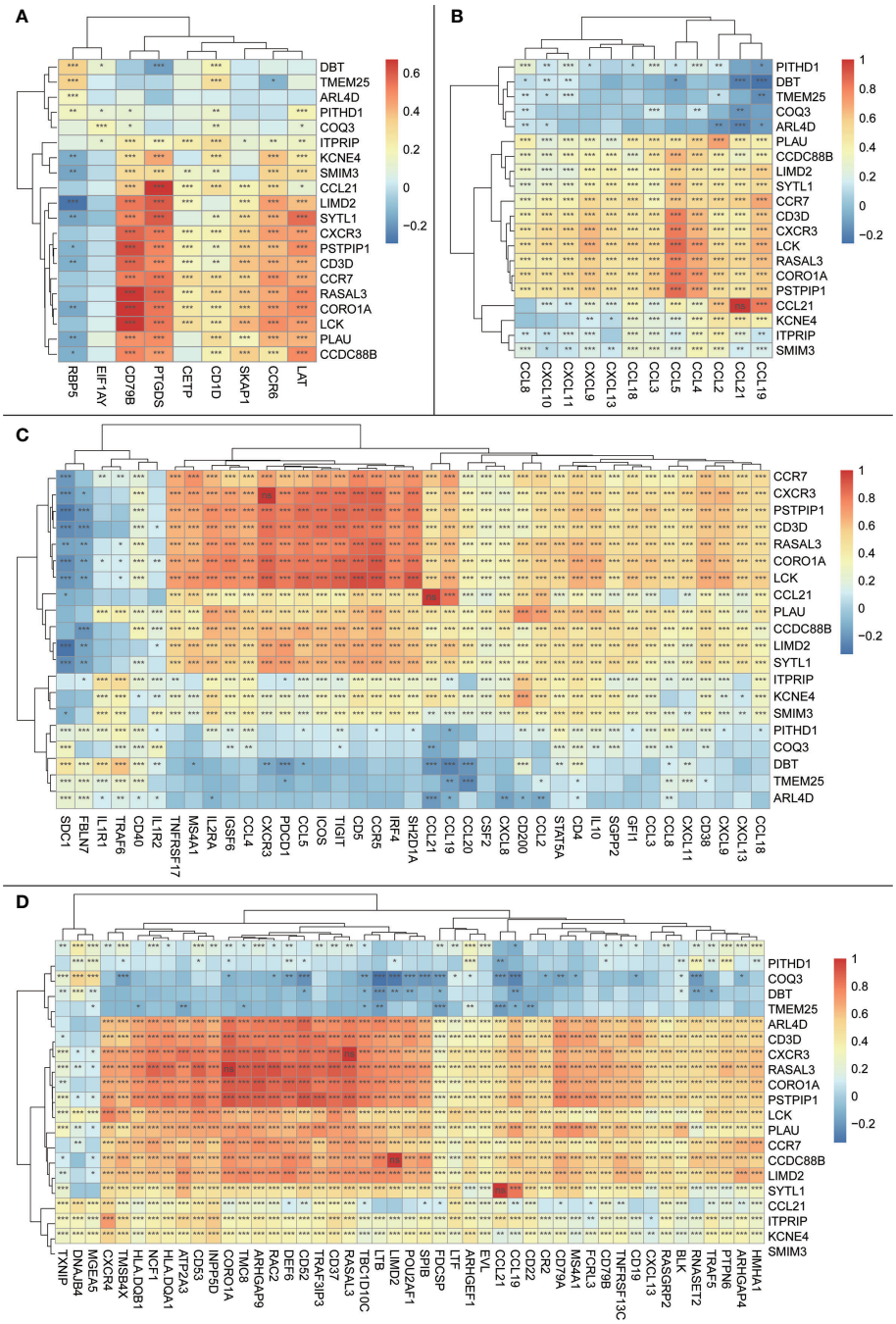


FIGURE 7
Heatmap of correlations between the trait genes and other previously reported TLS features; the vertical axis is the signature genes obtained in the current study, and the horizontal axis is other previously reported TLS signature genes; the magnitude of correlations is indicated by the block color; significance is shown in the block by symbols: ns: p > 0.05 or the two items tested for correlation are the same item; *:p <= 0.05; **:p <= 0.01. **(A)** Heatmap of correlations between trait genes and 9-TLS traits. **(B)** Heatmap of correlations between trait genes and 12-TLS traits. **(C)** Heatmap of correlation between the trait genes and 40-TLS traits. **(D)** Heatmap of correlations between the trait genes and 50-TLS features.

was thought to be associated with improved RFS in the early stages of HCC (8, 11). However, we found that this improvement is not limited to the early stages, but is reflected throughout the entire period (1, 2, and 5 years). In addition, we found a strong relationship between iTLS and improvement of DFS. However, consistent with our results, other recent studies on iTLS in HCC did not find improvement in patient OS, and controversy remains regarding the role of TLS in HCC Further studies are

10.3389/fimmu.2022.1007426

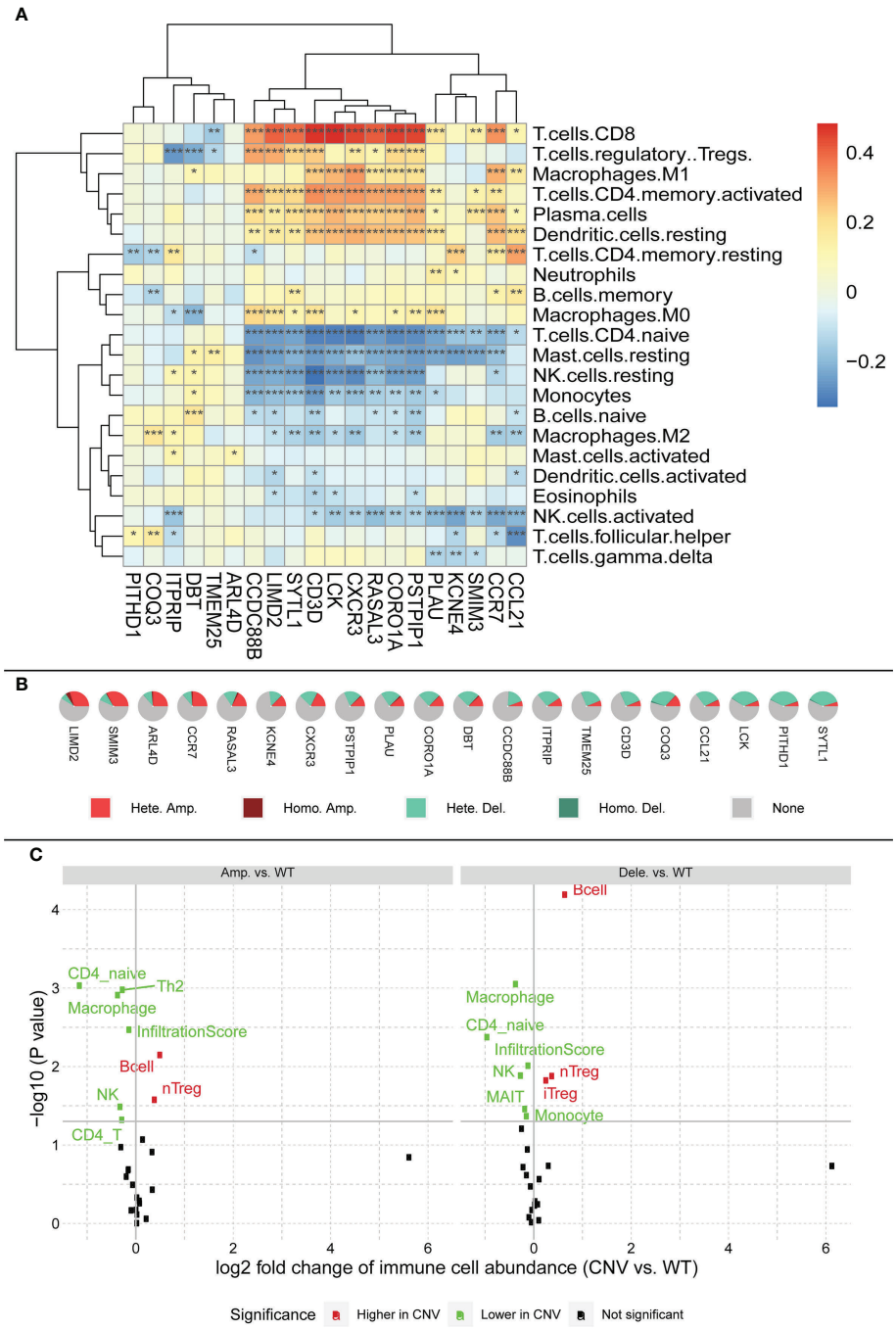


FIGURE 8
(A) Relationship between characteristic genes and immune cell infiltration, block colors represent the level of correlation. Significance markers: *p <= 0.05; **p <= 0.01; ***p <= 0.001. **(B)** CNV profile of feature genes in HCC; legend meanings are as follows: Hete amp, the percentage of samples with copy number heterozygous amplification; Hete dele, the percentage of samples with copy number heterozygous deletion; Homo amp, the percentage of samples with copy number homozygous amplification; Homo dele, the percentage of samples with copy number homozygous deletion. **(C)** Relationship between CNV and immune cell infiltration; black dots in the figure represent no effect of CNV on immune cell infiltration, red dots represent enhanced immune cell infiltration caused by CNV, and green dots represent diminished immune cell infiltration caused by CNV.

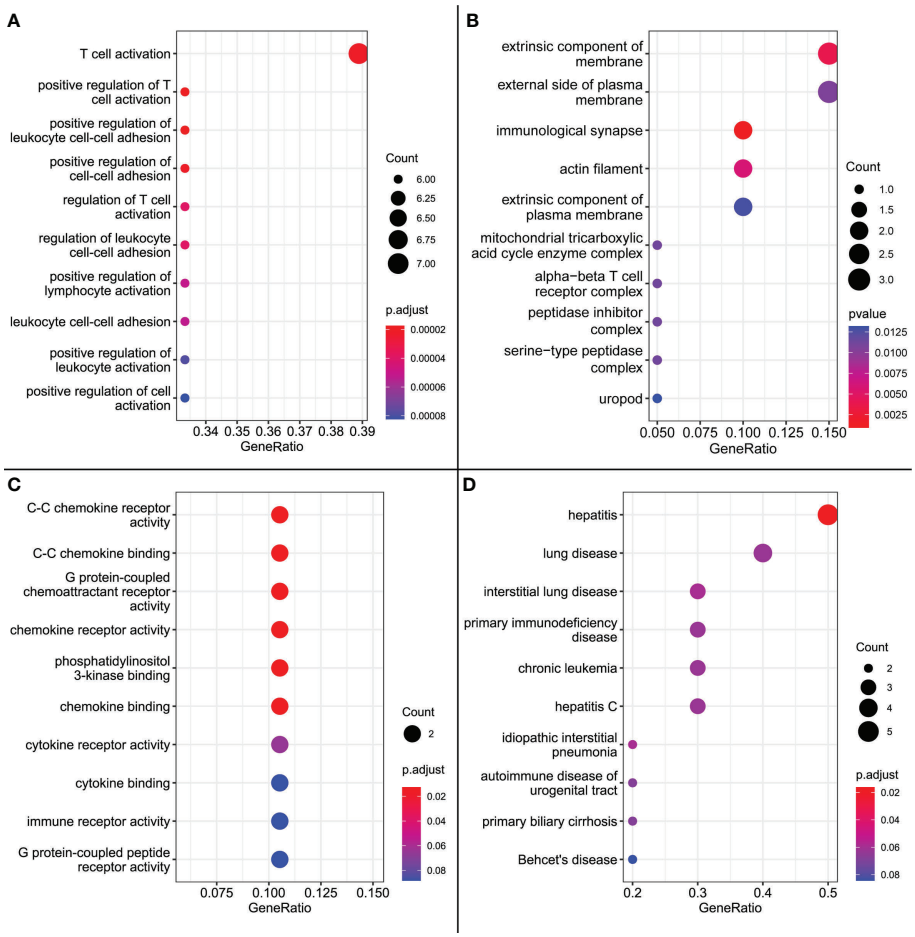


FIGURE 9
(A) GO enrichment results for biological processes. **(B)** GO enrichment results for cellular composition. **(C)** GO enrichment results for molecular functions. **(D)** DO enrichment results.

required to investigate and resolve this controversy. Combined with the gene expression data, we identified 20 genes that have an important relationship with iTLS formation in HCC. We demonstrated that these genes are closely related to the immune system in terms of cellular infiltration, biological functions, and signaling pathways. In addition, we found that most signature genes had some degree of sensitivity to immune checkpoint inhibitors. Considering these findings, we can speculate that signature genes may be promising targets for future HCC treatment and further demonstrate the protective effect of iTLS on HCC prognosis. However, unlike our results, most previous reports found that the number of patients in a TLS-positive group was less than that in a TLS-negative group. We speculate that these results were due to the inclusion of TIL in the TLS-positive group. In fact, there is no precise method for distinguishing between TLS and TIL, and the boundary between the two definitions is blurred. Therefore, to maximize the prognostic impact of TLS, we included TIL in the TLS-positive group, which may have led to the difference in results. Clarice et al., used the same method (17).

Our findings showed that the signature genes correlated with most of the previously reported TLS signatures, highlighting the accuracy of our signature genes. Additionally, a large number of immune-related genes were present in the signature genes, including SYTL1, ARL4D, PITHD1, CCR7, LCK, CCDC88B, CCL21, RASAL3, CXCR3, CD3D, PSTPIP1, and KCNE4. Among these, SYTL1, the gene with the highest importance in the random forest results, may play an important role in cytotoxic granule cytokinesis in lymphocytes (30, 31). There is evidence that ARL4D can control T effector function by limiting

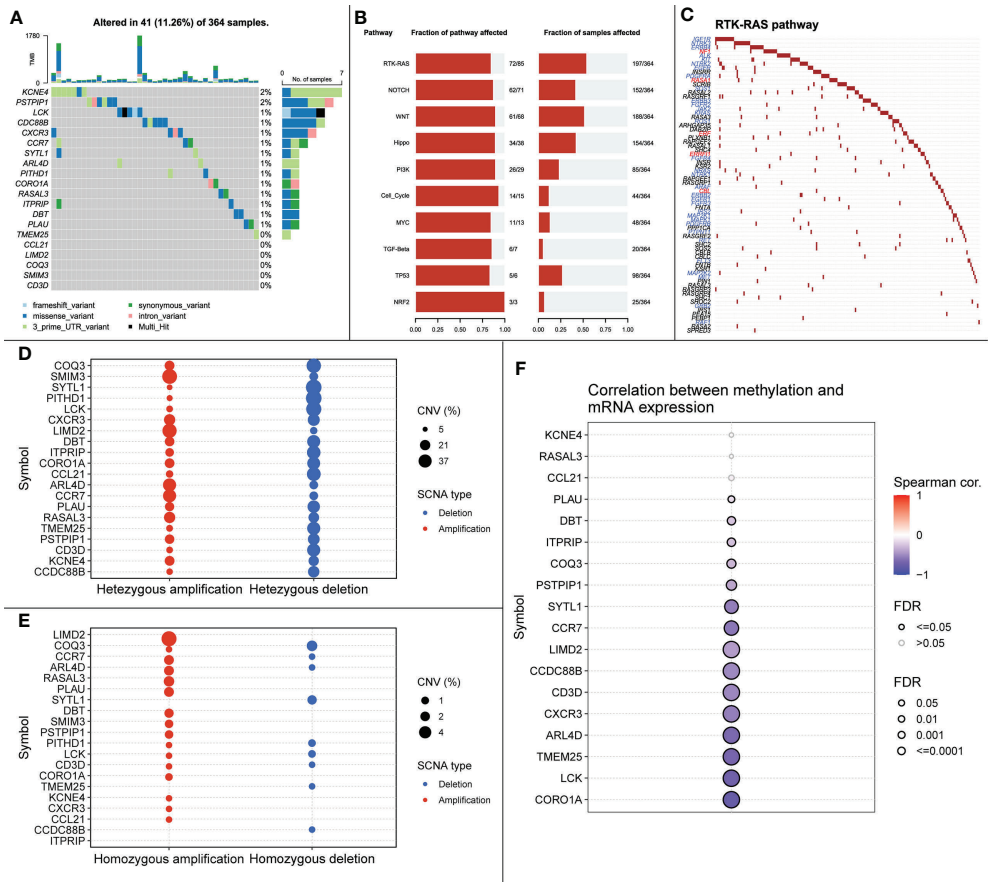


FIGURE 10
(A) SNP mutations in all samples. **(B)** Signaling pathways affected by feature genes in all samples. **(C)** Mutations in RTK-RAS pathway in all samples, blue font represents cancer-promoting mutations, red font represents cancer-suppressing mutations. **(D)** Heterozygous CNV mutations in the TCGA-LIHC cohort for the signature gene. **(E)** Pure-zygous CNV mutations in the TCGA-LIHC cohort for the signature gene. **(F)** Relationship between mRNA expression of the signature gene and methylation.

IL-2 production (32), and other genes, such as PITHD1, CCR7, LCK, and CCDC88B, have also been shown to have different effects on immunity (22, 33–36). These results corroborate the important influence of immunity on the development of HCC.

To our knowledge, we are the first to report the associations of TMEM25, COQ3, ITPRIP, DBT, and PLAU with HCC. TMEM25 has been identified as a member of the immunoglobulin superfamily, which is a target of pharmacogenomics in oncology and regenerative medicine (37). PLAU is of great importance in renal cell carcinoma (38). COQ3 has been shown to have an important role in the prognosis of esophageal cancer (39). ITPRIP has also been reported in patients with colon cancer (40). DBT has been shown to have an important effect on primary biliary cirrhosis (41).

As an immune structure, the tumor immune microenvironment to which TLS is directly exposed is important for its formation. We explored TLS signature genes within HCC tumors from the perspectives of signaling pathways and immune cell infiltration using GSEA, CIBERSORT, GSCA, and GO enrichment analysis. The results of GSEA, using both the KEGG and Reactome databases, demonstrated that differentially expressed genes between TLS+ and TLS-group samples are associated with a large number of immune pathways, which can be further verified by GO analysis. The immune infiltration score showed that our signature genes are closely associated with the infiltration of immune cells, such as CD8T, Tregs, CD4T, and NK. These cells play different roles in the progression of the disease during the fight against HCC

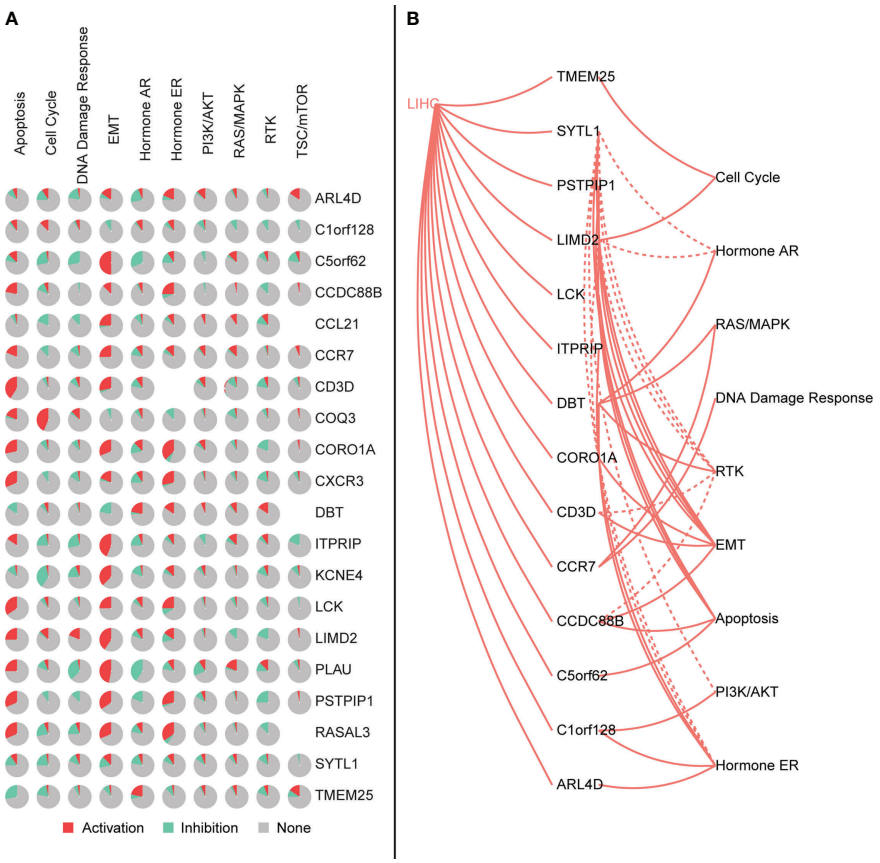


FIGURE 11
(A) Characteristic genes affect the activation or inhibition of the pathway, where red represents the pathway being activated and green represents the pathway being inhibited. **(B)** Interaction between the characteristic genes and the pathway, where the solid line is the pathway being activated and the dashed line is the pathway being inhibited.

(42, 43), which also implies that the signature genes can influence the development of HCC through immunity. We found that CNV with signature genes, regardless of type, caused a decrease in immune infiltration, which may demonstrate that signature genes have an important role in immune cell recruitment. Understanding gene mutations in different samples and exploring the upstream and downstream signaling pathways affected by mutated genes is important for the development of targeted cancer therapies. We discovered that most of the mutations in iTLS-related signature genes were concentrated in the RTK-RAS pathway, and that overactivation of this pathway was closely related to HCC (44). After exploring the drug sensitivity of the signature genes, we found that RASAL3 and CORO1A had strong sensitivity to most of the drugs. This finding may lead to potential improvement and/or new development of targeted therapies.

The purpose of medical research is ultimately clinical, and our prognostic model may improve the assessment of patient prognosis. More importantly, this study provides further support for the protective role of iTLS in HCC. However, for several reasons, we were unable to perform basic experiments to further validate our findings, which is one of the limitations of this study. Research of iTLS in HCC is currently at an early stage, and presently a lack of evidence is common. Therefore, more in-depth study is urgently needed to demonstrate the effect of iTLS on HCC. iTLS' positive effect in HCC has been repeatedly reported and was validated in this study. However, additional large multicenter studies and emphasis on the importance of continued investment in research are required to elucidate the specific functional mechanisms of iTLS. A key issue to consider before conducting further studies is to unify the evaluation criteria for TLS. Currently, scholars use different criteria to define TLS, which inevitably causes errors in research results. Predictably, in the near future, the development of artificial intelligence and improved computer technology will make standardization of TLS identification possible. With improved standardization, TLS will be a promising tool to add to the arsenal in the clinical fight against cancer.

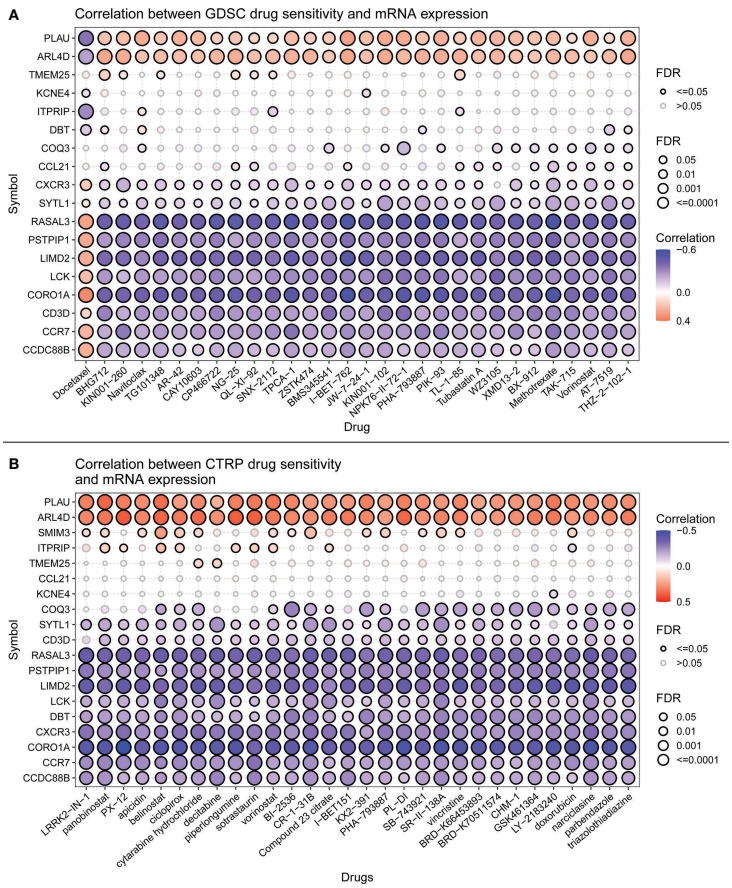


FIGURE 12
(A) Drug sensitivity of signature genes in GDSC. **(B)** Drug sensitivity of signature genes in CTRP.

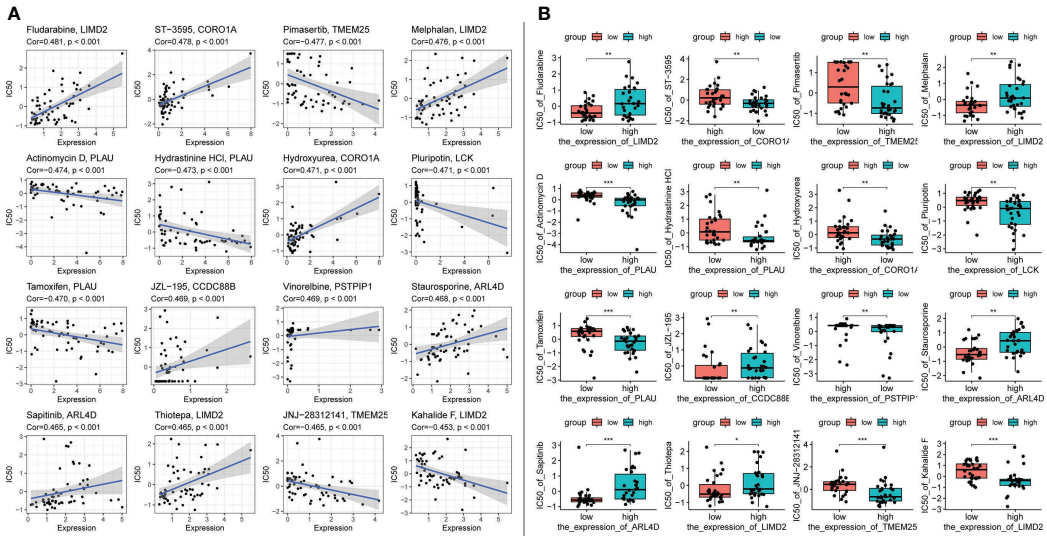


FIGURE 13
Top 16 drug effects with the highest correlation in the CellMiner database for characteristic genes, gene and drug names are identified in the figure by chart title or vertical coordinate. **(A)** Horizontal axis is gene expression, vertical axis is drug IC50. **(B)** Horizontal axis is gene expression along median dichotomous classification, vertical axis is drug IC50. significance markers, ns: p > 0.05; *p <= 0.05; **p <= 0.01; ***p <= 0.001.

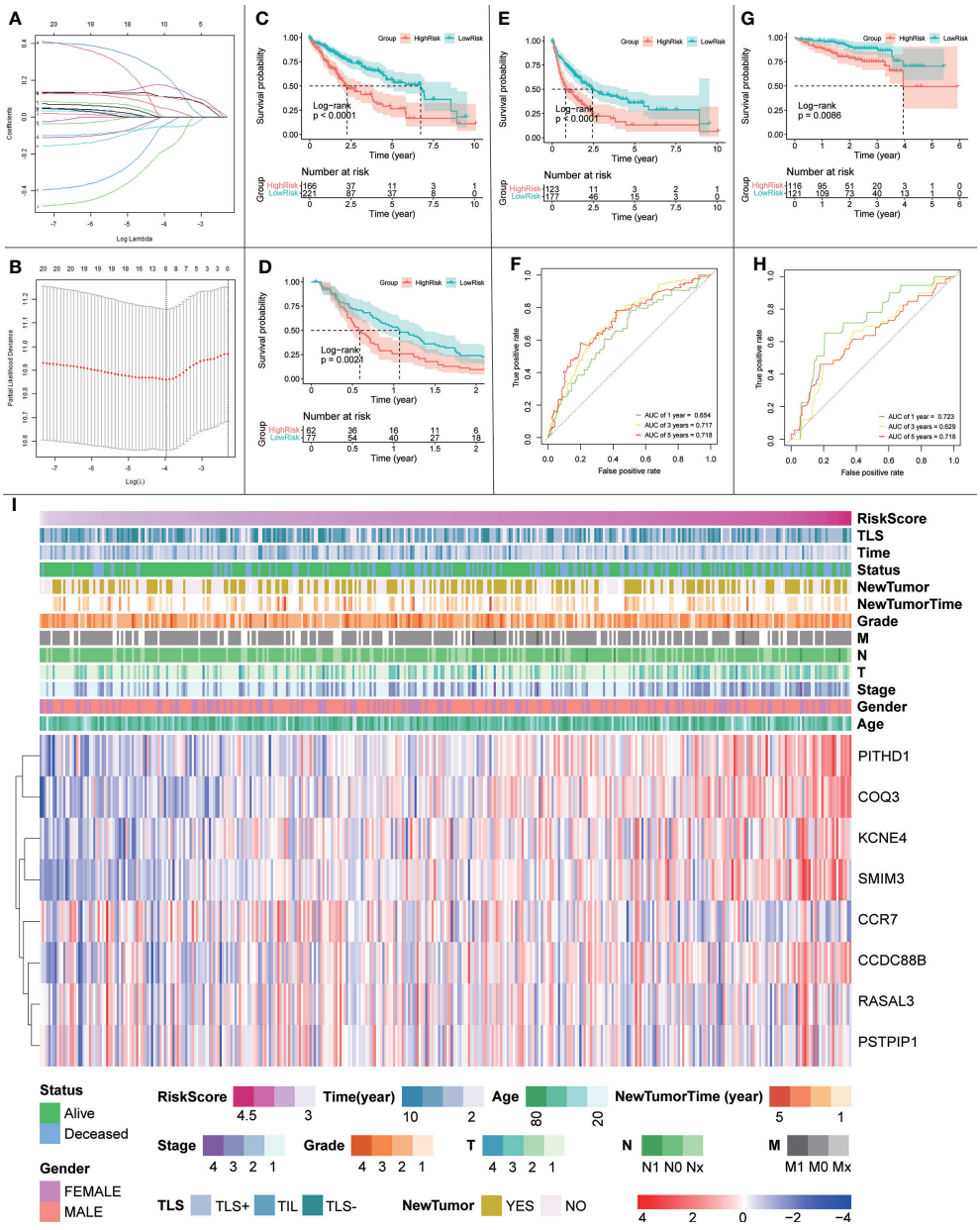


FIGURE 14
(A, B) Schematic diagram of the LASSO regression variable shrinkage screening process. **(C)** OS of the prognostic model in the TCGA cohort. **(D)** RFS of the prognostic model in the TCGA cohort. **(E)** DFS of the prognostic model in the TCGA cohort. **(F)** Time-dependent ROC curve of the prognostic model in the TCGA cohort. **(G)** OS curve of the prognostic model in the ICGC cohort. **(H)** Prognostic model in the ICGC cohort with time-dependent ROC curves. **(I)** Heatmap of risk scores, clinical features and expression of signature genes.

5 Conclusion

We found that the improvement of RFS in patients with HCC due to iTLS is not limited to the early period as previously reported but is reflected throughout the entire period. In addition, we found that iTLS could improve DFS. Combined with the gene expression data, we identified 20 genes that have an important relationship with iTLS formation in HCC. We demonstrated that these genes are closely related to immunity in terms of cellular infiltration, biological functions, and signaling pathways. In addition, we

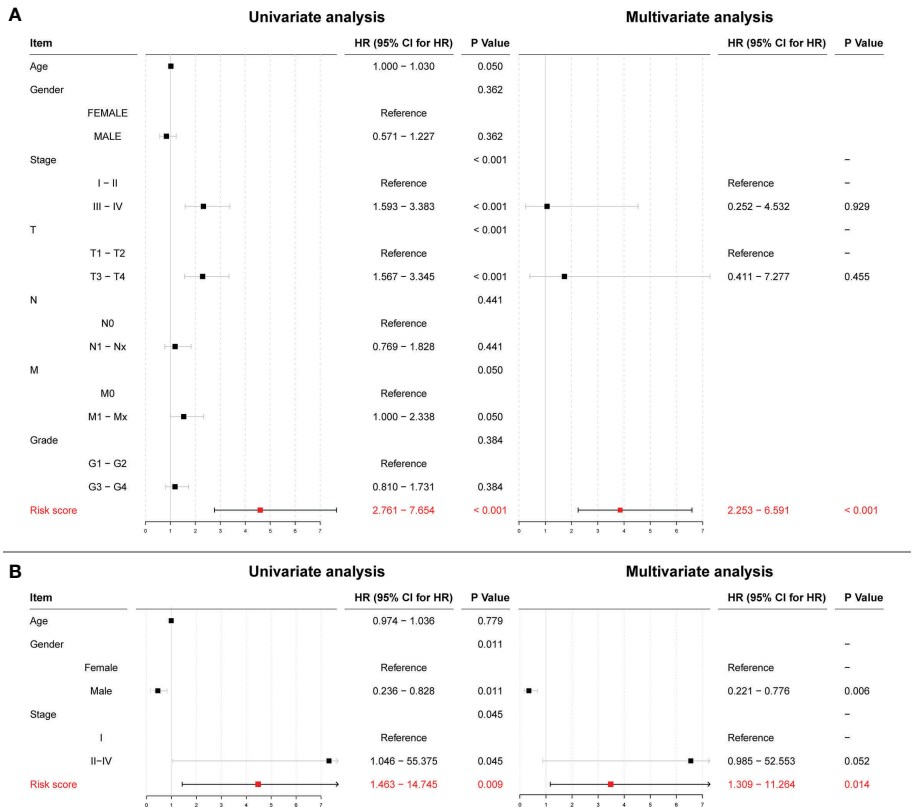


FIGURE 15
(A) Univariate and multivariate Cox regression results for risk scores in the TCGA cohort. **(B)** Univariate and multivariate Cox regression results for risk scores in the ICGC cohort.

found that the majority of the signature genes had some degree of sensitivity to immune checkpoint inhibitors. Considering these findings, we speculate that signature genes may be promising targets for future HCC therapy and further demonstrate the protective effect of iTLS on HCC prognosis.

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

WJ contributed to the concept and design of the study. QY carefully reviewed the first draft of the article. TZ, JL, YN, WS, and XL participated in the discussions. WJS oversaw all aspects of literature review design and manuscript writing. All authors contributed to the manuscript and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (81672716).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1007426/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660

2. Ferlay J, Laversanne M, Ervik M, Lam F, Colombet M, Mery L, et al. Global cancer observatory: Cancer tomorrow. *Cancer Tomorrow* (2020).

3. van de Walle T, Vaccaro A, Ramachandran M, Pietilä I, Essand M, Dimberg A. Tertiary lymphoid structures in the central nervous system: Implications for glioblastoma. *Front Immunol* (2021) 12:724739. doi: 10.3389/fimmu.2021.724739

4. Fridman WH, Meylan M, Petitprez F, Sun C-M, Italiano A, Sautès-Fridman C. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. *Nat Rev Clin Oncol* (2022), 1–17. doi: 10.1038/s41571-022-00619-z

5. Dieu-Nosjean M-C, Antoine M, Danel C, Heudes D, Wislez M, Poulot V, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J Clin Oncol* (2008) 26:4410–7. doi: 10.1200/JCO.2007.15.0284

6. Cheng W, Li H-L, Xi S-Y, Zhang X-F, Zhu Y, Xing Le, et al. Growth differentiation factor 1-induced tumour plasticity provides a therapeutic window for immunotherapy in hepatocellular carcinoma. *Nat Commun* (2021) 12:7142. doi: 10.1038/s41467-021-27525-9

7. Xing R, Gao J, Cui Q, Wang Q. Strategies to improve the antitumor effect of immunotherapy for hepatocellular carcinoma. *Front Immunol* (2021) 12:783236. doi: 10.3389/fimmu.2021.783236

8. Calderaro J, Petitprez F, Becht E, Laurent A, Hirsch TZ, Rousseau B, et al. Intra-tumoral tertiary lymphoid structures are associated with a low risk of early recurrence of hepatocellular carcinoma. *J Hepatol* (2019) 70:58–65. doi: 10.1016/j.jhep.2018.09.003

9. Shen Y-C, Hsu C-L, Jeng Y-M, Ho M-C, Ho C-M, Yeh C-P, et al. Reliability of a single-region sample to evaluate tumor immune microenvironment in hepatocellular carcinoma. *J Hepatol* (2020) 72:489–97. doi: 10.1016/j.jhep.2019.09.032

10. Luo M, Lin Y, Liang R, Li Y, Ge L. Clinical significance of the HHLA2 protein in hepatocellular carcinoma and the tumor microenvironment. *J Inflammation Res* (2021) 14:4217–28. doi: 10.2147/JIR.S324336

11. Li H, Wang J, Liu H, Lan T, Xu L, Wang G, et al. Existence of intratumoral tertiary lymphoid structures is associated with immune cells infiltration and predicts better prognosis in early-stage hepatocellular carcinoma. *Aging (Albany NY)* (2020) 12:3451–72. doi: 10.18632/aging.102821

12. Zhao H, Wang H, Zhou Q, Ren X. Insights into tertiary lymphoid structures in the solid tumor microenvironment: anti-tumor mechanism, functional regulation, and immunotherapeutic strategies. *Cancer Biol Med* (2021). doi: 10.20892/j.issn.2095-3941.2020.0569

13. Patman G. Ectopic lymphoid structures promote carcinogenesis in the liver. *Nat Rev Gastroenterol Hepatol* (2015) 12:671–1. doi: 10.1038/nrgastro.2015.192

14. Finkin S, Yuan D, Stein I, Taniguchi K, Weber A, Unger K, et al. Ectopic lymphoid structures function as microniches for tumor progenitor cells in hepatocellular carcinoma. *Nat Immunol* (2015) 16:1235–44. doi: 10.1038/ni.3290

15. Li H, Liu H, Fu H, Li J, Xu L, Wang G, et al. Peritumoral tertiary lymphoid structures correlate with protective immunity and improved prognosis in patients with hepatocellular carcinoma. *Front Immunol* (2021) 12:648812. doi: 10.3389/fimmu.2021.648812

16. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data *via* the xena platform. *Nat Biotechnol* (2020) 38:675–8. doi: 10.1038/s41587-020-0546-8

17. Groeneveld CS, Fontugne J, Cabel L, Bernard-Pierrot I, Radvanyi F, Allory Y, et al. Tertiary lymphoid structures marker CXCL13 is associated with better survival for patients with advanced-stage bladder cancer treated with immunotherapy. *Eur J Cancer* (2021) 148:181–9. doi: 10.1016/j.ejca.2021.01.036

18. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Software* (2010) 36:1–13. doi: 10.18637/jss.v036.i11

19. Luo Z, Wang W, Li F, Songyang Z, Feng X, Xin C, et al. Pan-cancer analysis identifies telomerase-associated signatures and cancer subtypes. *Mol Cancer* (2019) 18:106. doi: 10.1186/s12943-019-1035-x

20. Bao X, Zhang H, Wu W, Cheng S, Dai X, Zhu X, et al. Analysis of the molecular nature associated with microsatellite status in colon cancer identifies clinical implications for immunotherapy. *J Immunother Cancer* (2020) 8:e001437. doi: 10.1136/jitc-2020-001437

21. Feng H, Yang F, Qiao L, Zhou K, Wang J, Zhang J, et al. Prognostic significance of gene signature of tertiary lymphoid structures in patients with lung adenocarcinoma. *Front Oncol* (2021) 11:693234. doi: 10.3389/fonc.2021.693234

22. Tokunaga R, Nakagawa S, Sakamoto Y, Nakamura K, Naseem M, Izumi D, et al. 12-chemokine signature, a predictor of tumor recurrence in colorectal cancer. *Int J Cancer* (2020) 147:532–41. doi: 10.1002/ijc.32982

23. Zhou X, Li W, Yang J, Qi X, Chen Y, Yang H, et al. Tertiary lymphoid structure stratifies glioma into three distinct tumor subtypes. *Aging (Albany NY)* (2021) 13:26063–94. doi: 10.18632/aging.203798

24. Wu R, Guo W, Qiu X, Wang S, Sui C, Lian Q, et al. Comprehensive analysis of spatial architecture in primary liver cancer. *Sci Adv* 7:eabg3750. doi: 10.1126/sciadv.abg3750

25. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

26. Liu C-J, Hu F-F, Xia M-X, Han L, Zhang Q, Guo A-Y. GSCALite: a web server for gene set cancer analysis. *Bioinformatics* (2018) 34:3771–2. doi: 10.1093/bioinformatics/bty411

27. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41:D955–61. doi: 10.1093/nar/gks1111

28. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* (2016) 12:109–16. doi: 10.1038/nchembio.1986

29. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* (2012) 72:3499–511. doi: 10.1158/0008-5472.CAN-12-1370

30. McAdara Berkowitz JK, Catz SD, Johnson JL, Ruedi JM, Thon V, Babior BM. JFC1, a novel tandem C2 domain-containing protein associated with the

leukocyte NADPH oxidase. *J Biol Chem* (2001) 276:18855–62. doi: 10.1074/jbc.M011167200

31. Holt O, Kanno E, Bossi G, Booth S, Daniele T, Santoro A, et al. Slp1 and Slp2-a localize to the plasma membrane of CTL and contribute to secretion from the immunological synapse. *Traffic* (2008) 9:446–57. doi: 10.1111/j.1600-0854.2008.00714.x

32. Tolksdorf F, Mikulec J, Geers B, Endig J, Sprezyna P, Heukamp LC, et al. The PDL1-inducible GTPase Arl4d controls T effector function by limiting IL-2 production. *Sci Rep* (2018) 8:16123. doi: 10.1038/s41598-018-34522-4

33. Lu B, Sun X, Chen Y, Jin Q, Liang Q, Liu S, et al. Novel function of PITH domain-containing 1 as an activator of internal ribosomal entry site to enhance RUNX1 expression and promote megakaryocyte differentiation. *Cell Mol Life Sci* (2015) 72:821–32. doi: 10.1007/s00018-014-1704-2

34. Kennedy JM, Fodil N, Torre S, Bongfen SE, Olivier J-F, Leung V, et al. CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. *J Exp Med* (2014) 211:2519–35. doi: 10.1084/jem.20140455

35. Marth JD, Overell RW, Meier KE, Krebs EG, Perlmutter RM. Translational activation of the lck proto-oncogene. *Nature* (1988) 332:171–3. doi: 10.1038/332171a0

36. Salem A, Alotaibi M, Mroueh R, Basheer HA, Afarinkia K. CCR7 as a therapeutic target in cancer. *Biochim Biophys Acta Rev Cancer* (2021) 1875:188499. doi: 10.1016/j.bbcan.2020.188499

37. Katoh M, Katoh M. Identification and characterization of human TMEM25 and mouse Tmem25 genes in silico. *Oncol Rep* (2004) 12:429–33. doi: 10.3892/or.12.2.429

38. Wakita T, Hayashi T, Nishioka J, Tamaru H, Akita N, Asanuma K, et al. Regulation of carcinoma cell invasion by protein c inhibitor whose expression is decreased in renal cell carcinoma. *Int J Cancer* (2004) 108:516–23. doi: 10.1002/ijc.11594

39. King RJ, Qiu F, Yu F, Singh PK. Metabolic and immunological subtypes of esophageal cancer reveal potential therapeutic opportunities. *Front Cell Dev Biol* (2021) 9:667852. doi: 10.3389/fcell.2021.667852

40. Abdul Aziz NA, Mokhtar NM, Harun R, Mollah MMH, Mohamed Rose I, Sagap I, et al. A 19-gene expression signature as a predictor of survival in colorectal cancer. *BMC Med Genomics* (2016) 9:58. doi: 10.1186/s12920-016-0218-1

41. Nishio A, Van de Water J, Leung PS, Joplin R, Neuberger JM, Lake J, et al. Comparative studies of antimitochondrial autoantibodies in sera and bile in primary biliary cirrhosis. *Hepatology* (1997) 25:1085–9. doi: 10.1002/hep.510250506

42. Zhao H-Q, Li W-M, Lu Z-Q, Yao Y-M. Roles of tregs in development of hepatocellular carcinoma: A meta-analysis. *World J Gastroenterol* (2014) 20:7971–8. doi: 10.3748/wjg.v20.i24.7971

43. El-Rebey HS, Abdou AG, Sultan MM, Ibrahim SH, Holah NS. The profile and role of tumor-infiltrating lymphocytes in hepatocellular carcinoma: An immunohistochemical study. *Appl Immunohistochem Mol Morphol* (2021) 29:188–200. doi: 10.1097/PAI.0000000000000865

44. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer* (2006) 6:674–87. doi: 10.1038/nrc1934

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fimmu.2022.975762

Check for updates

OPEN ACCESS

EDITED BY
Jun Liu,
Yuebei People's Hospital, China

REVIEWED BY
Mengya Zhao,
China Pharmaceutical University, China
Zi Mei,
Peking University, China
Qingjia Chi,
Wuhan University of Technology, China

*CORRESPONDENCE
Genhao Zhang
wangshuya617@bjmu.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 22 June 2022
ACCEPTED 31 August 2022
PUBLISHED 15 September 2022

CITATION
Zhang G (2022) Regulatory T-cells-related signature for identifying a prognostic subtype of hepatocellular carcinoma with an exhausted tumor microenvironment.
Front. Immunol. 13:975762.
doi: 10.3389/fimmu.2022.975762

COPYRIGHT
© 2022 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.

Regulatory T-cells-related signature for identifying a prognostic subtype of hepatocellular carcinoma with an exhausted tumor microenvironment

Genhao Zhang*

Department of Blood transfusion, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Regulatory T-Cells (Tregs) are important in the progression of hepatocellular cancer (HCC). The goal of this work was to look into Tregs-related genes and develop a Tregs-related prognostic model. We used the weighted gene co-expression network analysis (WGCNA) to look for Tregs-related genes in the TCGA, ICGC, and GSE14520 cohorts and then used the non-negative matrix factorization (NMF) algorithm to find Tregs-related subpopulations. The LASSO-Cox regression approach was used to determine Tregs-related genes, which were then condensed into a risk score. A total of 153 overlapping genes among the three cohorts were considered Tregs-related genes. Based on these genes, two Tregs-associated clusters that varied in both prognostic and biological characteristics were identified. When compared with Cluster 1, Cluster 2 was a TME-exhausted HCC subpopulation with substantial immune cell infiltration but a poor prognosis. Five Tregs-related genes including *HMOX1, MMP9, CTSC, SDC3*, and *TNFRSF11B* were finally used to construct a prognostic model, which could accurately predict the prognosis of HCC patients in the three datasets. Patients in the high-risk scores group with bad survival outcomes were replete with immune/inflammatory responses, but exhausted T cells and elevated PD-1 and PD-L1 expression. The results of qRT-PCR and immunohistochemical staining (IHC) analysis in clinical tissue samples confirmed the above findings. Moreover, the signature also accurately predicted anti-PD-L1 antibody responses in the IMvigor210 dataset. Finally, *HMOX1*, *MMP9*, and *TNFRSF11B* were expressed differently in Hep3B and Huh7 cells after being treated with a PD1/PD-L1 inhibitor. In conclusion, our study uncovered a Tregs-related prognostic model that could identify TME-exhausted subpopulations and revealed that PD1/PD-L1 inhibitors could alter the expression levels of *HMOX1*, *MMP9*, and *TNFRSF11B* in Hep3B and Huh7 cells, which might help us better understand Tregs infiltration and develop personalized immunotherapy treatments for HCC patients.

KEYWORDS

HCC, prognosis, PD1/PD-L1, Tregs, TME, exhausted

Introduction

Chronic liver inflammation caused by HBV or HCV infection, alcoholism, or nonalcoholic fatty liver disease (NAFLD) can result in an aberrant concentration of immune cells in tumors and adjacent tissues, including T lymphocytes, macrophages, and dendritic cells (1). These immune cells, together with other non-immune components (fibroblasts, extracellular matrix), comprise the tumor microenvironment (TME) surrounding cancer cells (2). The dynamic interaction between cancer cells and TME can disrupt tumor cells' immune surveillance, accelerate tumor cell proliferation, clonal evolution, immune evasion, and treatment resistance, and play a key role in tumor genesis and progression (3). TME potentially causes tissue remodeling and functional impairment by generating local hypoxia in tumor tissue and ultimately promoting tumor metastasis (4). In addition, TME can impact the delivery of anticancer medications to the tumor location by interacting with mesenchymal stem cells (5). $CD4^+CD25^+FoxP3^+$ T regulatory lymphocytes (Tregs), as an important heterogeneous T cell subset, have been identified to participate in the development of HCC by promoting immune tolerance (6). Treg cells are abundant in tumors and can make up 10 to 50 percent of the $CD4^+$ T cells there (7). Notably, the proportion of Treg cells in the peripheral blood did not match the density of Treg cells in the TME, indicating that the study of the function of Treg cells in the TME is more crucial in the field of cancer immunology. Depletion of T-reg cells encourages the growth of high endothelial venules, which are crucial for lymphocyte recruitment (8). Tregs can block cytotoxic $CD8^+$ T lymphocytes (CTLs) cytotoxic and proliferative capabilities, aid in the creation of an immunosuppressive TME, and are linked to the advancement of HCC (9), while $CD8^+$ T cells specific for neoantigens are more resistant to Treg cell-mediated immune suppression (10). Crosstalk between Tregs and neutrophil extracellular traps promotes the transition of NAFLD to HCC (11). TGFβ-activated stromal cells reduce the recruitment of Tregs in TME, thereby regulating the balance between anti-tumor and pro-tumor immune cells (12). High Tregs infiltration is also linked to poor outcomes and recurrence in HCC patients (13, 14). In recent years, immunotherapy using immune checkpoint inhibitors (ICIs), such as anti-PD1 antibody nivolumab and anti-PD-L1 agents atezolizumab, has shown strong antitumor activity in a subset of HCC patients by blocking the interaction of PD1 with its ligands, thereby preventing exhaustion or dysfunction of effector T cells (15). Interestingly, there is a close and complex relationship between Tregs infiltration and PD1 expression. Tregs expressing PD1 in the TME can impact immunosuppressive function and are associated with progression in cancer patients (16). PD1 blockade induces enhanced $PD1^+$ Tregs-mediated immunosuppression (17). In addition, Lenvatinib can improve anti-PD1 effectiveness by reducing tumor PD-L1 levels and Tregs differentiation (18). Therefore, exploring them more deeply can help us gain a deeper understanding of the complex mechanisms of the TME in HCC development, and help clinicians to formulate strategies for the use of ICIs in cancer treatment.

Given the difficulty in collecting enough tumor tissue for tumor-infiltrating lymphocyte (TIL) assay analysis by flow cytometry (FCM) and the fact that crosstalk between Tregs and cancer is a complex process involving multiple genes, we built and validated a prognostic stratification model based on Tregs-related genes in public datasets that could be used to efficiently categorize HCC patients prognostically and predict their response to anti-PD-L1 immunotherapy.

Materials and methods

Acquisition of public datasets and clinical sample processing

Transcriptome expression data from HCC patients were gathered from three public databases, including the TCGA-LIHC (2022.04), the ICGC (LIRI-JP, 2019.11), and the GSE14520 (2010.12) cohorts. The clinical characteristics of HCC patients in the three cohorts were displayed in Table S1. Patients with incomplete survival data or who lived for less than one month were eliminated from the study. Clinically verified samples for qRT-PCR research were fresh frozen tumor biopsies and their surrounding normal tissues from 20 previously collected HCC patients. The primer sequences are shown in Table S2. Zhengzhou University's Ethics Committees gave its approval to this work. Written informed consent was taken.

Estimation of immune cell infiltration and TME scores

The relative abundance of 28 immune cell subtypes in the three datasets was assessed by the ssGSEA algorithm, and the immune cell abundance identifier (ImmuCellAI, 2020.02) was further utilized to specifically assess the abundance of comprehensive T cell subsets (19). ESTIMATE algorithm was used to calculate stromal and immunological scores in tumor tissue based on gene expression patterns of HCC samples to determine the quantity of stromal and immune cells inside the tumor (20). Immunohistochemical staining (IHC) was performed to explore the infiltration of Tregs and $CD8^+$ T cells in HCC tissues. Two competent pathologists performed

IHC findings assessment using a single-blind and uniform standard procedure.

Tregs-related genes identification by the weighted gene co-expression network analysis

The WGCNA was used to create a scale-free co-expression network based on transcriptome expression data from the three datasets to find the gene modules most relevant for Tregs infiltration abundance. Standard deviation (SD > 50%) was used to screen for highly variable genes. Module membership represented the link between module characteristic genes and gene expression patterns, whereas gene significance (GS) was utilized to assess the relationship between individual genes and Tregs infiltration abundance. The genes identified from the modules most linked with Tregs infiltration abundance were appraised as candidate genes using a p-value threshold of GS < 0.0001 and a significance level of univariate Cox regression of p < 0.01. The overlapping genes of the candidate genes in the three datasets were finally confirmed as Tregs-related genes and used for subsequent analysis.

Identification of prognostic molecular subtypes by the non-negative matrix factorization algorithm

Based on the Tregs-related genes obtained above, patients were clustered using the NMF algorithm, the standard was "brunet", and the iterations were 50 times. The number of clusters varies from 2 to 6, and the optimal number of clusters is determined based on cophenetic, dispersion, and contour. Kaplan-Meier survival analyses were further performed to assess differences in patients' survival rates between different subtypes.

DEG identification and functional enrichment analysis

Gene Expression Profile Interaction Analysis (GEPIA, 2017.07) (21) (http://gepia.cancer-pku.cn/) was used to investigate the expression levels of Tregs-related genes, and genes with statistically significant differences were defined as differentially expressed genes (DEGs) with a P-value < 0.05 and a $|\log_2 FC|$ cutoff criterion of ≥0.5. The Metascape database (22) (http://metascape.org/, 2019.04) was then used to investigate the functional annotation of DEGs for GO and KEGG pathway analysis. With a significant threshold of |normalized enrichment score|>1 and a nominal p-value of < 0.05, Gene set enrichment analysis (GSEA) was used to investigate changes in Hallmarks.

Formation and validation of Tregs-related prognostic risk scoring model

The correlation between Tregs-related genes and HCC patient survival outcomes was calculated using univariate Cox regression with a P-value < 0.01. The Tregs-related genes with prognostic significance were then investigated using the LASSO-Cox regression technique and a classifier linked with prognosis was established. The multivariate Cox relapse coefficient (β) was used to create a risk score based on the concept of directly mixing the equation below with the mRNA expression level. The risk score = ∑iCoefficient (mRNAi)*Expression (mRNAi). Due to the optimal hazard score edge, we divided the HCC patients into two categories. ROC analysis, Kaplan-Meier survival analysis, and cox relative risks relapse investigation were used to assess the prognostic signature's predictive autonomy. The ICGC and GSE14520 datasets were used as validation cohorts for validating our constructed Tregs-related signature.

Genetic alterations and drug susceptibility analysis

The R package "maftools" was used to assess the differences in genetic variations between various subgroups using the mutation and CNA data of 342 HCC patients acquired from the TCGA dataset. The association between anticancer drug sensitivity and mRNA molecules in our risk model was directly explored in the CellMiner database (2012.07) (23) with cutoff criteria of adjusted P-values <0.001 and Pearson's correlation coefficients >0.4.

Immunohistochemistry staining analysis

The paraffin samples were cut into 4 μm slides and soaked for the identification of Tregs and $CD8^+$ T cells. The tissue fragments were progressively hydrated in graded alcohol after being deparaffinized in xylene. By heating 0.01 mol/L citrate buffer in a steam cooker for 10 minutes, antigen retrieval was accomplished. To suppress endogenous peroxidase activity, slides were washed with PBS and then incubated for 20 minutes at 37°C with a 0.3 percent H_2O_2 solution. Slides were then blocked with bovine serum albumin (BSA), and continuously incubated with anti-FOXP3 and CD8 antibodies overnight at 4°C, respectively. After being washed with PBS, slides were continuously incubated with secondary antibodies coupled to horseradish peroxidase (HRP) for 30 minutes. Utilizing HRP's routine substrate detection, immune complexes were found. Slides were then dehydrated in graded alcohols and xylene after being stained with hematoxylin.

Cell culture and PD1/PD-L1 inhibitor treatment

Hep3B and Huh7 cells from the Cell Bank of the Shanghai Institute of Cell Research, Chinese Academy of Sciences (Shanghai, China) were cultured in the suggested DMEM medium (Sangon Biotech, China) containing 10% fetal bovine serum (FBS, Sangon Biotech, Shanghai, China) at 100% humidity, 37°C, and 5% CO2. Cells were incubated for 4 hours at room temperature in the DMEM medium containing 4 mg/mL PD1/PD-L1 inhibitor (Abcam, ab230369, UK) for PD1/PD-L1 blockade.

Statistical analysis

Categorical data were compared using Pearson's chi-square test or Fisher's exact test when appropriate, and quantitative data between two groups were compared using the t-test. The examination of data from three or more groups was done using a one-way analysis of variance (ANOVA). R software (Version 4.0.3) was used to analyze the prediction performance of survival outcomes using receiver operating characteristic (ROC) curve analysis and Kaplan-Meier survival analysis. The association between a prognostic classifier and survival outcomes was investigated using a Cox proportional model. When the P-value < 0.05, the results were considered statistically significant. The flowchart of this study is shown in Figure S1.

Results

Identification genes associated with Tregs infiltration

After removing outliers (Figure S2), 9, 11, and 17 non-grey modules were created in the three datasets, respectively, according to the results of WGCNA (Figure 1A). As shown in Figures 1B, C, the yellow module was the most significantly related to Tregs infiltration in the TCGA cohort (R^2 = 0.82, P = 2e−53), and the yellow module was the most significantly related to Tregs infiltration in the ICGC cohort (R^2 = 0.83, P = 2e−51), and the brown module was the most significantly related to Tregs infiltration in the GSE14520 cohort (R^2 = 0.95, P = 2e−70), respectively. 153 overlapping genes among the three cohorts were identified as Tregs-related genes (Figure 1D). The biological importance of these Tregs-related genes was mainly enriched in the immune-inflammatory response and regulation of lymphocytes (Figure 1E). When expression in normal tissues was considered, out of 153 genes, only 16 were identified as differentially expressed genes (DEGs) (Figure S3A). Then, a PPI network was performed to explore the potential interactions between these DEGs (Figure S3B), and the biological importance of these Tregs-related DEGs was mainly enriched in Cytokine signaling and regulation of leukocytes (Figure S3C).

Identification of prognostic molecular subtypes

To further explore the mechanism of these 153 Tregs-related genes in HCC, the NMF algorithm was performed. Due to the cophenetic, dispersion, and profile (Figure S4), number two was identified as the optimal number of clusters (Figure 2A). Patients in Cluster 2 had better survival outcomes when compared with patients in Cluster 1 (Figure 2B). Subsequently, we found that the mutation rates of mutated genes in the two subgroups were also significantly different. The most commonly mutated gene in the Cluster 1 was CTNNB1 (29%, Figure 2C), while it was TP53 (34%, Figure 2D) in Cluster 2. After the tumor mutation burden (TMB) was estimated, patients in Cluster 1 had a higher TMB value, compared with patients in Cluster 2 (Figure 2E). Finally, we found that patients in Cluster 2 with high TMB values had the worst survival outcomes when compared with others (Figure 2F).

Patients in Cluster 2 had an exhausted immune microenvironment

Patients in Cluster 2 had higher immune, stromal, and ESTIMATE scores compared with patients in Cluster 1, as shown in Figure 3A. According to the ssGSEA algorithm, almost all types of immune cells were higher in Cluster 2 than those in Cluster 1 except for CD56bright natural killer cell, CD56dim natural killer cell, eosinophil, neutrophil, and Type 17 T helper cell (Figure 3B). Interestingly, we found that patients in Cluster 2 had poor survival outcomes but a higher abundance of $CD8^+$ T cells infiltration. Considering that decreased infiltration levels of $CD8^+$ T cells were often associated with poor survival rates, therefore, we assumed that these $CD8^+$ T cells in Cluster 2 were exhausted T cells. To test this conjecture, we then analyzed genes involved in immune/inflammatory responses, including *CD8A*, *GZMB*, *IFNG*, *TBX2*, and *TNF*, and found that these genes were significantly up-regulated in Cluster 2 (Figure 4A). We also found that the expression of PD1, a marker of exhausted T cells, was also significantly up-regulated in Cluster 2 (Figure 4B), as was the expression level of PD-L1 (Figure 4C). In addition, exhausted T cells were significantly enriched in Cluster 2, according to the results of ImmuCellAI analysis (Figure 4D). Finally, the GSEA results indicated that Cluster 1 displayed an attenuated IFN-γ response (Figure 4E), which can directly increase PD-L1 expression and activate the PD-1/PD-L1 signaling axis. Together, the aforementioned findings showed that Cluster 2 had a robust immunological and inflammatory response, but the elevated PD1 and PD-L1 in this group might

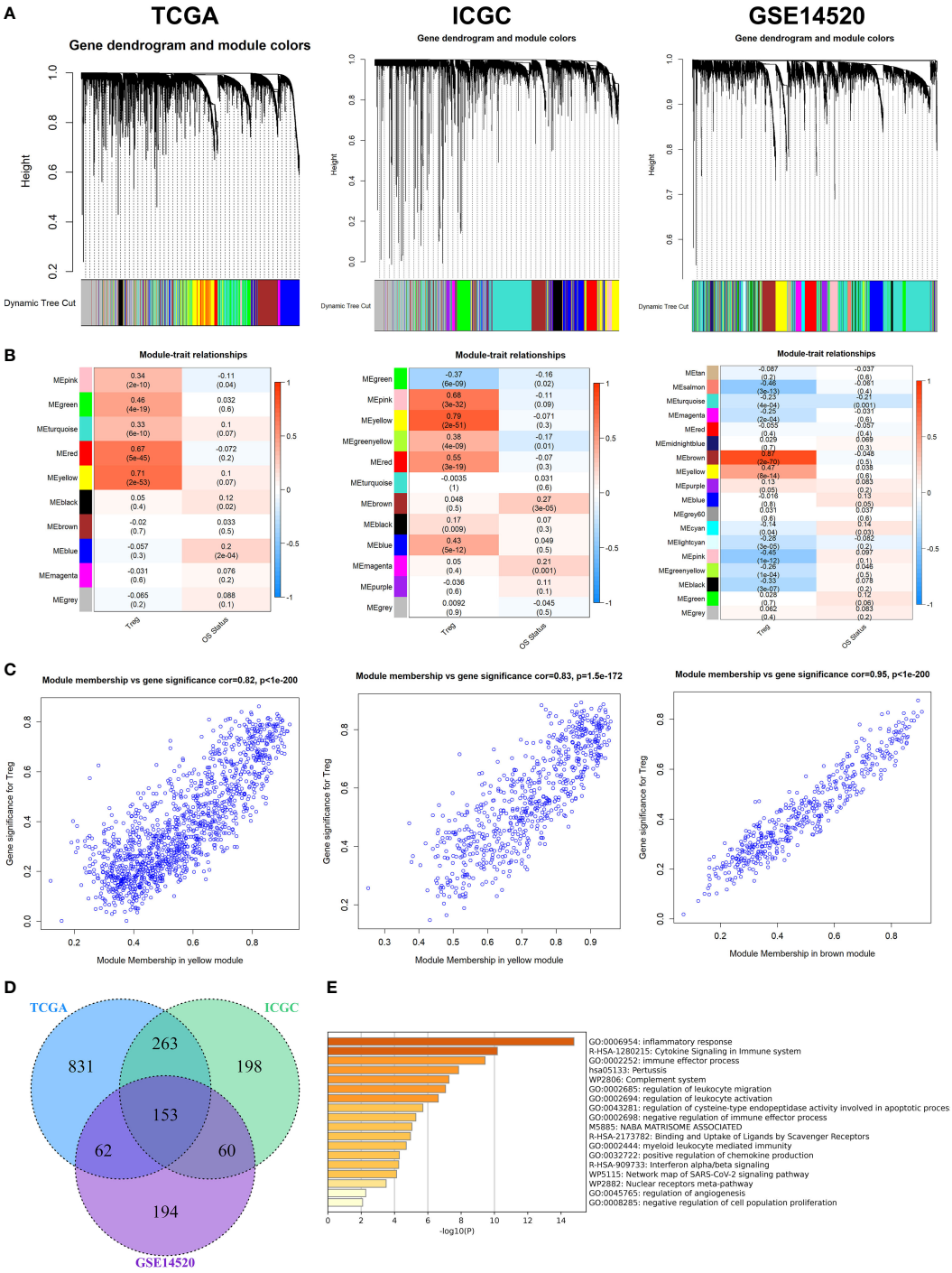


FIGURE 1
WGCNA for Tregs-related genes. **(A)** The coexpression network was established in the TCGA, the GSE14520, and the ICGC cohorts. **(B)** Heatmap demonstrating the correlation between module eigengenes and Tregs. **(C)** Determination of modules most significantly associated with Tregs infiltration. **(D)** 153 overlapping genes among the three cohorts were identified as Tregs-related genes. **(E)** The biological importance of these Tregs-related genes was mainly enriched in the immune-inflammatory response and regulation of lymphocytes.

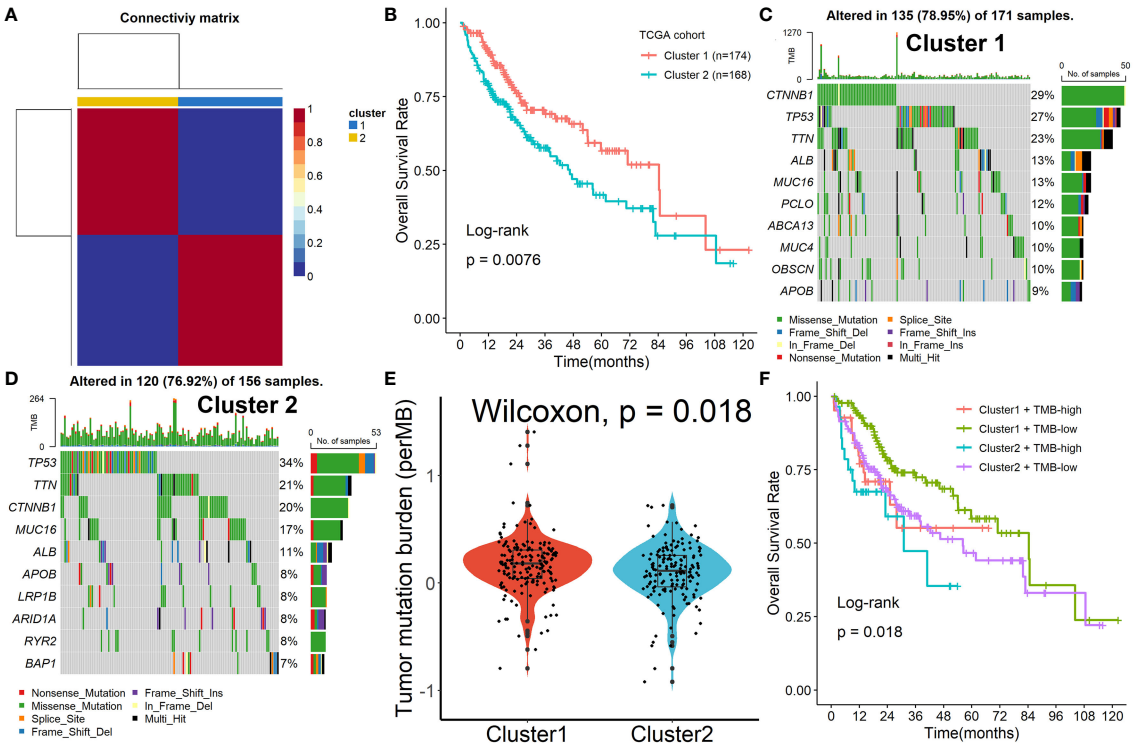


FIGURE 2
Identification of prognostic molecular subtypes by the NMF algorithm. **(A)** Number two was identified as the optimal number of clusters. **(B)** Patients in Cluster 2 had better survival outcomes when compared with patients in Cluster 1. **(C, D)** The mutation rates of mutated genes in the two subgroups. **(E)** Patients in Cluster 1 had a higher TMB value, compared with patients in Cluster 2. **(F)** Patients in Cluster 2 with high TMB values had the worst survival outcomes when compared with others.

result in an exhausted TME and eventually have a negative impact on the survival of HCC patients.

Formation of Tregs-related prognostic signature in HCC

Among the 153 overlapping genes in the three cohorts obtained by WGCNA, 15 prognosis-associated genes were identified by univariate Cox regression with a p-value less than 0.01 (Figure 5A). These genes were then selected by using the LASSO-Cox regression model based on the minimum value of λ (Figure 5B). Five genes including HMOX1, MMP9, CTSC, SDC3, and TNFRSF11B were screened out and were then put into a multivariate Cox proportional model, andfinally, a prognostic Tregs-related signature was formatted. Risk score = (0.16468758×*HMOX1*) + (0.04918601×*MMP9*) + (0.16592365×*CTSC*) + (0.06017538×*SDC3*) + (0.09164677×*TNFRSF11B*). Patients were divided into high- or low-risk scores subgroups with the median of scores after patients' risk scores were calculated with the above formula (Figure 5C). We found that patients with lower risk scores were remarkably relevant to better survival outcomes (Figure 5D) and this Tregs-related signature had a good prognostic performance with AUCs at 1-, 3-, 5-year of 0.698, 0.643, 0.680 (Figure 5E). In addition, patients in Cluster 2 had higher risk scores compared with patients in Cluster 1, as shown in Figure 5F. Finally, after controlling for other clinical parameters, this Tregs-related signature might be used as an independent predictive factor for HCC patients (HR=2.566, 95 percent CI 1.4401 -4.5742, P = 0.0013). Only MMP9 and CSTC were significantly correlated with overall survival rates, despite the five Tregs-associated gene expression levels in the GEPIA database varied between normal and malignant tissues (Figure S5). Additionally, the five Tregs-related genes' protein expression in both normal and HCC was examined in the Human Protein Atlas database (24) (HPA, www.proteinatlas.org), as shown in Figure S6.

Functional enrichment and genetic alterations analysis

GSEA analysis revealed that the immunological response, controlling lymphocyte activity, and production and metabolism

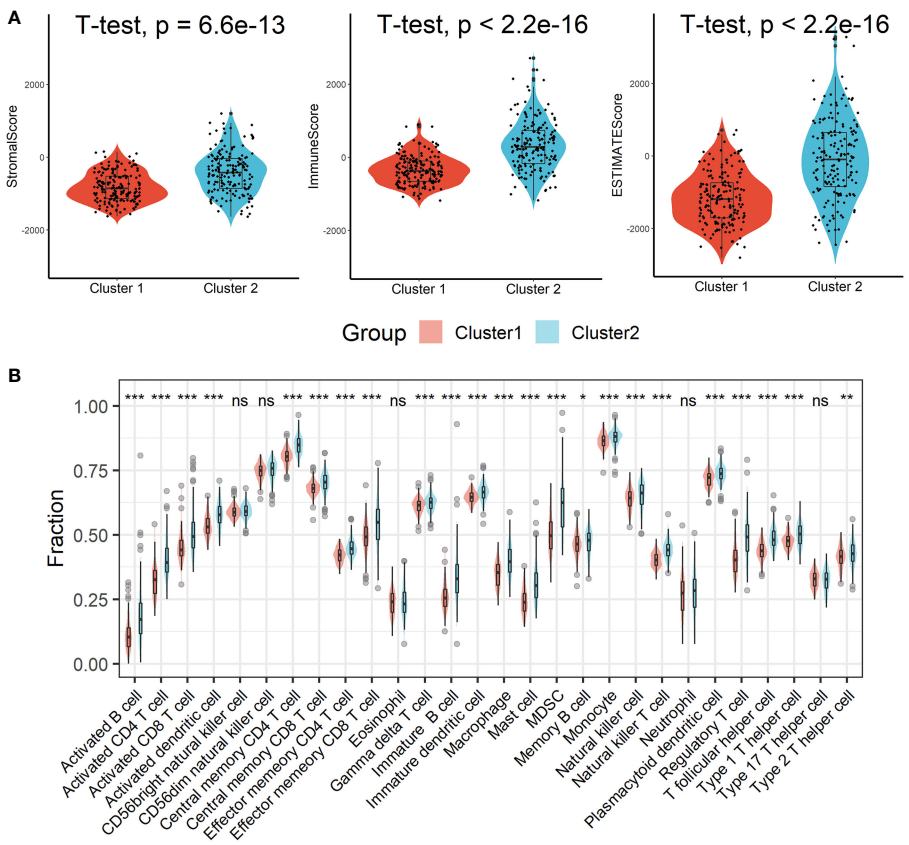


FIGURE 3
Estimation of immune cell infiltration in different clusters. **(A)** Patients in Cluster 2 had higher immune, stromal, and ESTIMATE scores compared with patients in Cluster 1. **(B)** Almost all types of immune cells were higher in Cluster 2 than those in Cluster 1 except for CD56bright natural killer cell, CD56dim natural killer cell, eosinophil, neutrophil, and Type 17 T helper cell. ns, not significant; *p < 0.05; **p <0.01; ***p < 0.001.

of cytokines were the three primary areas of changed GO and KEGG items between high- and low-risk score groups (Figure S7). We next discovered that the two subgroups had dramatically differing mutation rates for the affected genes. TP53 (40%) and CTNNB1 (30%) were the most frequently altered genes in the groups with high and low risk scores, respectively (Figure S8A). Finally, we discovered that patients with high TMB levels who had high risk scores had the lowest survival rates (Figure S8B).

Verification of the Tregs-related signature in external cohorts

The ICGC and GSE14520 datasets were used as validation cohorts to verify this Tregs-related signature. In the ICGC cohort (Figure 6A) and GSE14520 cohort (Figure 6E), patients' risk scores were computed using the same formula, and patients were then split into high- or low-risk categories. No matter whether we looked at the ICGC cohort (Figure 6B) or the GSE14520 cohort (Figure 6F), we discovered that patients in the later TNM stage had greater risk ratings than patients in the early stage. Additionally, both in the ICGC cohort (Figure 6C) and the GSE14520 cohort (Figure 6G), patients with lower risk scores were strongly associated with higher OS rates. ROC analysis showed that this Tregs-related signature had a good prognostic performance with AUCs at 1-, 2-, 3-year of 0.650, 0.591, 0.629 in the ICGC cohort (Figure 6D) and at 1-, 3-, 5-year of 0.620, 0.631, 0.673 in the GSE14520 cohort (Figure 6H), respectively. Finally, after controlling for other clinical parameters, this Tregs-related signature might be used as an independent predictive factor for HCC patients in the GSE14520 cohort (HR=1.608, 95%CI 1.006-2.569, $P = 0.046$) but not in the ICGC cohort, which may be related to tumor heterogeneity.

Patients in the high-risk score group had an exhausted immune microenvironment

Patients in the high-risk scores group had higher immune, stromal, and ESTIMATE scores compared with patients in the

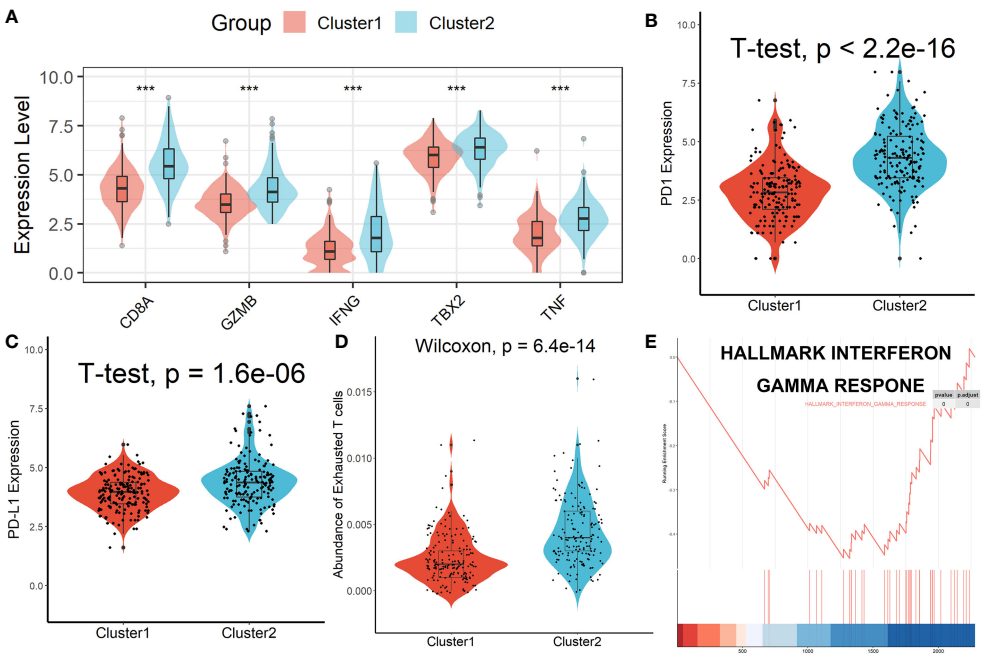


FIGURE 4
Patients in Cluster 2 had an Exhausted Immune Microenvironment. **(A)** Differential analysis of inflammation/immune response-related genes. **(B, C)** Differential analysis of PD1 and PD-L1 expression. **(D)** Exhausted T cells were significantly enriched in Cluster 2. **(E)** Cluster 1 displayed an attenuated IFN-γ response. ***p < 0.001.

low-risk scores group, as seen in Figure 7A. According to the ssGSEA algorithm, almost all types of immune cells were higher in the high-risk scores group than those in the low-risk scores group except for CD56bright natural killer cell, memory B cell, neutrophil, and eosinophil (Figure 7B). In addition, we also assessed the abundance of immune cell infiltration in HCC patients using various methods including TIMER (25), CIBERSORT (26), and MCP-counter (27) algorithms, and the results were consistent with the analysis results of the ssGSEA algorithm, as shown in Figure S9. Interestingly, we found that patients in the high-risk scores group had poor survival outcomes but a higher abundance of $CD8^+$ T cells infiltration. Considering that decreased infiltration levels of $CD8^+$ T cells were often associated with poor survival rates, therefore, we assumed that these $CD8^+$ T cells in the high-risk scores group were exhausted T cells. To test this conjecture, we then analyzed genes involved in immune/inflammatory responses, including CD8A, GZMB, IFNG, TBX2, and TNF, and found that these genes were significantly up-regulated in the high-risk scores group (Figure 8A). We also found that the expression of PD1 was significantly up-regulated in the high-risk scores group (Figure 8B), as was the expression level of PD-L1 (Figure 8C). In addition, exhausted T cells were significantly enriched in the high-risk scores group, according to the results of the ImmuCellAI analysis (Figure 8D). Finally, the GSEA results indicated that the low-risk scores group displayed an attenuated IFN-γ response (Figure 8E). Together, the aforementioned findings showed that the high-risk scores group had a robust immunological and inflammatory response, but the elevated PD1 and PD-L1 in this group might result in an exhausted TME and eventually have a negative impact on the survival of HCC patients.

Formation of a nomogram model and drug susceptibility analysis

A nomogram model was built in the TCGA dataset to investigate the coefficient prediction efficiency of this Tregs-related signature, and the results revealed that the nomogram with a C-index of 0.758 could help us provide a quantitative method for accurately predicting the 1-, 3-, and 5-year survival rate (Figure S10A). The calibration curves showed good agreement between the anticipated and actual probability of 1-, 3-, and 5-year survival rates (Figure S10B). We also uncovered 54 tumor-sensitive medications that target the five Tregs-related genes (Table S3), with the top 16 most important tumor-sensitive compounds indicated in Figure S11.

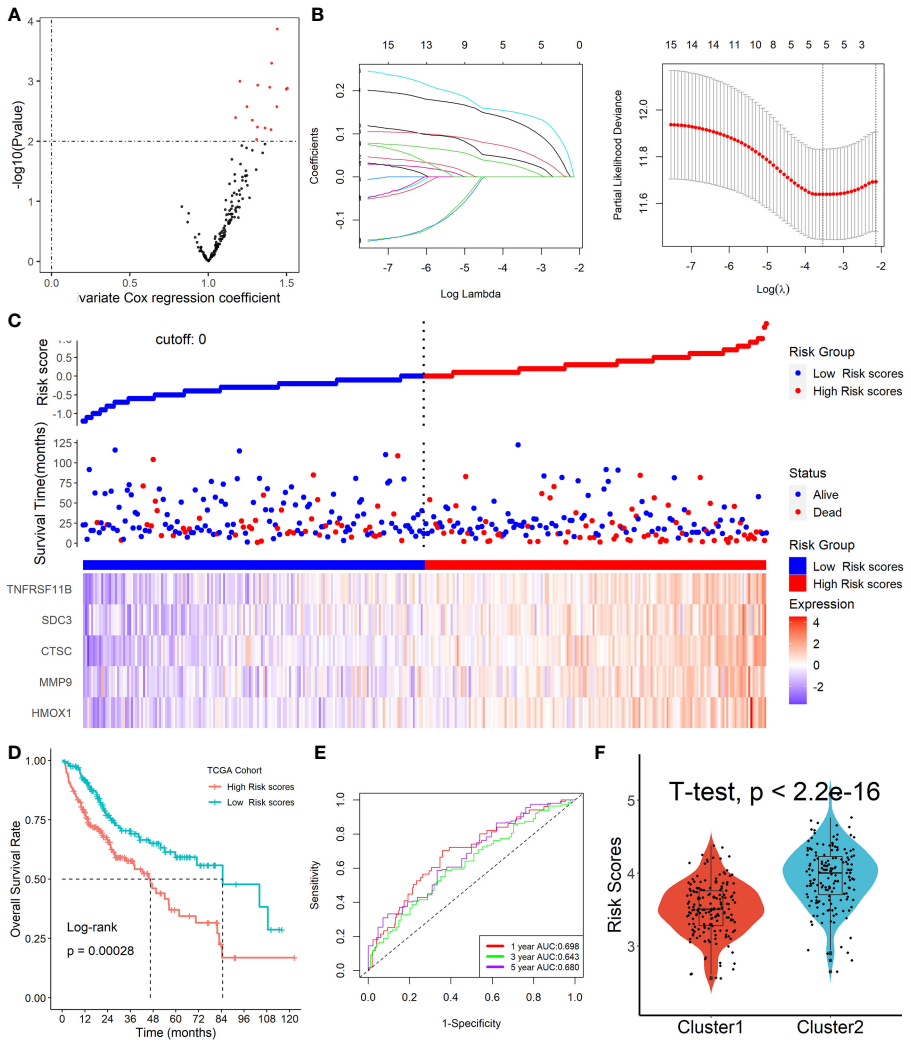


FIGURE 5
Formation of Tregs-related prognostic signature in HCC. **(A)** 15 prognosis-associated genes were identified by univariate Cox regression with a p-value less than 0.01. **(B)** Tregs-related genes were screened by the LASSO-Cox regression model. **(C)** Patients were divided into high- or low-risk scores subgroups with an optimal threshold after patients' risk scores were calculated with the above formula. **(D)** Patients with lower risk scores were remarkably relevant to better survival outcomes. **(E)** This Tregs-related signature had a good prognostic performance. **(F)** Patients in Cluster 2 had higher risk scores compared with patients in Cluster 1.

Forecasting response to anti-PD-L1 therapy using the Tregs-related signature

We discovered that the tumor immune dysfunction and exclusion (TIDE) scores in the high-risk scores group were much greater than that in the low-risk scores group using the TIDE algorithm (Figure 9A). T-cell exclusion scores did not differ significantly between the two groups (Figure 9B), but T-cell dysfunction scores were greater in the high-risk scores group than in the low-risk scores group (Figure 9C). In addition, due to a shortage of data on HCC patients undergoing anti-PD-L1 medication, the IMvigor210 database was utilized as an external anti-PD-L1 cohort to investigate the possible predictive usefulness of the Tregs-related signature. This research comprised 298 individuals who exhibited an objective response. We discovered that patients with low risk ratings had a significant survival advantage over those with high risk scores (Figure 9D). As demonstrated in Figure 9E, patients who had a complete response/partial response (CR/PR) had lower risk scores than patients who had stable disease/progressive disease (SD/PD). Finally, in the GSE109211 cohort, we looked at the link between Sorafenib treatment efficacy and risk scores and discovered that patients in the high-risk scores group had worse treatment results (Figure 9F).

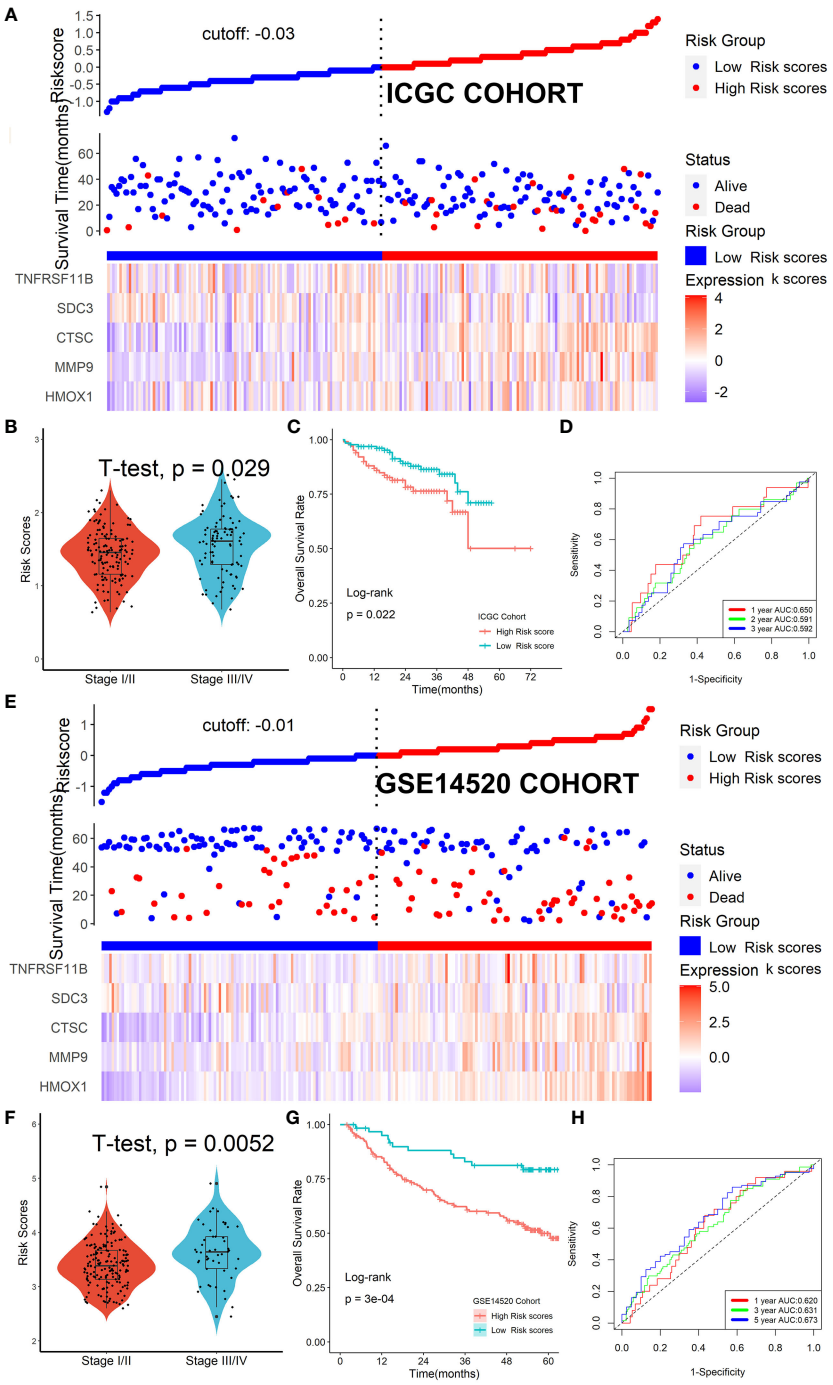


FIGURE 6
Verification of the Tregs-related signature in external cohorts. Patients were split into high- or low-risk categories in ICGC **(A)** and GSE14520 **(E)** cohorts. Patients in the later TNM stage had greater risk ratings than patients in the early stage, no matter whether we looked at the ICGC cohort **(B)** or the GSE14520 cohort **(F)**. Patients with lower risk scores were strongly associated with higher OS rates both in the ICGC cohort **(C)** and the GSE14520 cohort **(G)**. This Tregs-related signature had a good prognostic performance no matter in the ICGC cohort **(D)** or the GSE14520 cohort **(H)**.

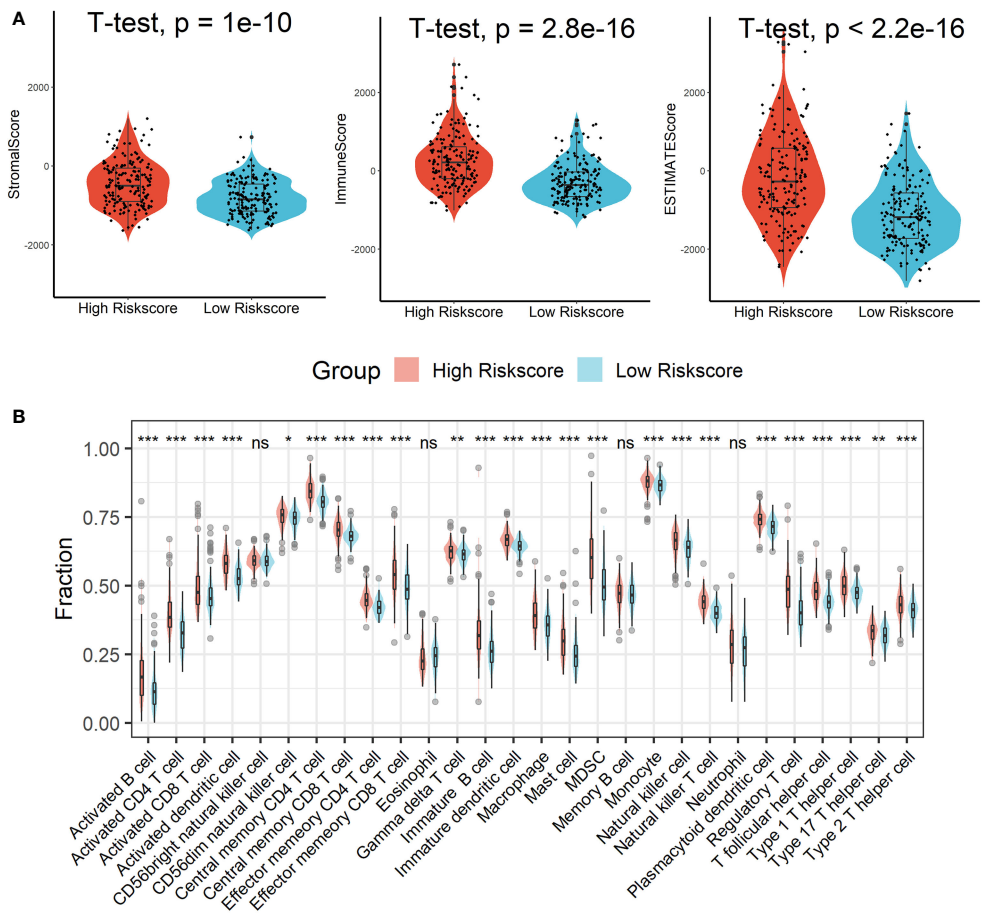


FIGURE 7
Estimation of immune cell infiltration in different risk scores groups. **(A)** Patients in the high-risk scores group had higher immune, stromal, and ESTIMATE scores compared with patients in the low-risk scores group. **(B)** Almost all types of immune cells were higher in the high-risk scores group than those in the low-risk scores group except for CD56bright natural killer cell, memory B cell, neutrophil, and eosinophil. ns, not significant; *p < 0.05; **p <0.01; ***p < 0.001.

Verification of the Tregs-related signature in clinical samples

All five Tregs-related genes revealed differential expression between normal and tumor tissues, according to the results of the qRT-PCR investigation (Figure 10A). After risk scores were determined using the same formula, patients were split into high-risk and low-risk groups according to the mean of the risk score. IHC was then used to analyze the infiltration of Tregs and $CD8^+$ T cells in the tissues of patients in high- and low-risk scores groups. Patients in the high-risk scores group had a higher abundance of Tregs and $CD8^+$ T cells infiltration (Figure 10B). Finally, expression levels of *CD8A*, *GZMB*, *IFNG*, *TBX2*, *TNF*, *PD1*, and *PD-L1* genes were analyzed and all of these genes were significantly up-regulated in the high-risk scores group (Figure 10C). Taken together, the aforementioned findings showed that the high-risk scores group had a robust immunological and inflammatory response, but the elevated PD1 and PD-L1 in this group might result in an exhausted TME and eventually have a negative impact on the survival of HCC patients.

Verification of the Tregs-related signature after PD1/PD-L1 blockade

As shown in Figure 11A, we found that all five genes in the Tregs-associated signature were significantly associated with PD1 and PD-L1 expression not only in the TCGA cohort but also in 20 clinical HCC samples, suggesting that these genes may be targets for anti-PD1/PD-L1 immunotherapy. To further explore their relationship, we treated cells with a small molecule PD1/PD-L1 inhibitor and found that the expression levels of *HMOX1*, *MMP9*, *CTSC*, and *TNFRSF11B* were significantly reduced in Hep3B cells, while only *HMOX1*,

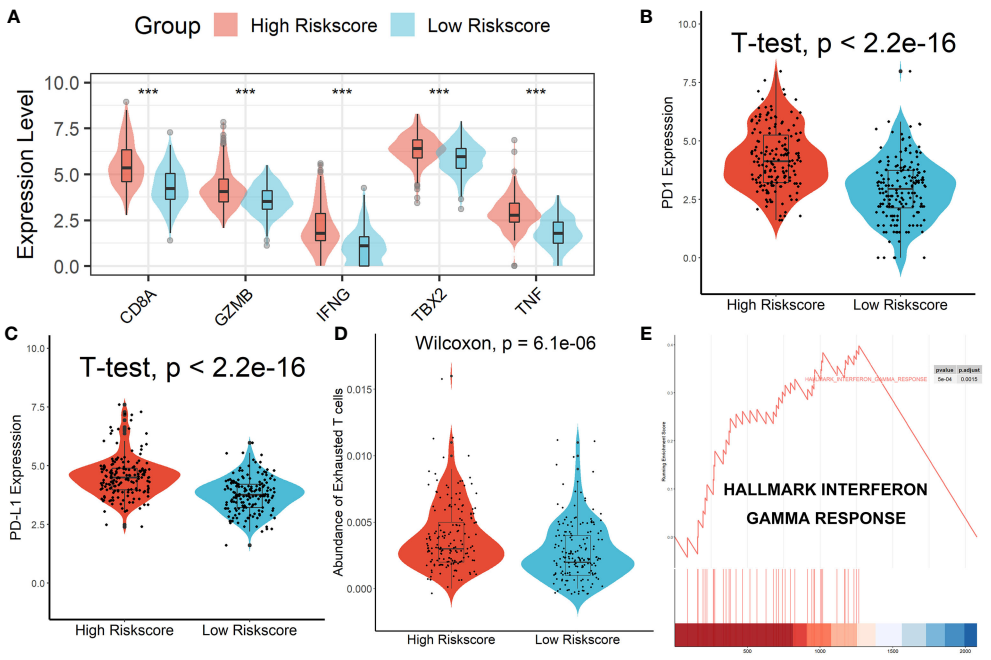


FIGURE 8
Patients in the high-risk score group had an Exhausted Immune Microenvironment. **(A)** Differential analysis of inflammation/immune response-related genes. **(B, C)** Differential analysis of PD1 and PD-L1 expression. **(D)** Exhausted T cells were significantly enriched in the high-risk scores group. **(E)** The low-risk scores group displayed an attenuated IFN-γ response. ***p < 0.001.

MMP9, and *TNFRSF11B* were expressed differently in Huh7 cells (Figure 11B). These confirmed our hypothesis that *HMOX1*, *MMP9*, and *TNFRSF11B* could be targeted for anti-PD1/PD-L1 immunotherapy.

Discussion

The interaction between tumors and TME has been a hot topic in recent years (3, 28–30). On the one hand, TME plays a role in immune surveillance and immune defense of tumor cells. On the other hand, tumor-related inflammation can cause abnormal infiltration of immune cells in tumor tissue and surrounding areas, resulting in an imbalance in the production of chemokines and cytokines, helping tumor cells to adapt to immune evasion, and ultimately promote tumor development. Increasing evidence suggests that TME is involved in the occurrence and progression of HCC, the development of drug resistance, and the efficacy of immunotherapy (31, 32). Therefore, a deeper understanding of the specific mechanism of TME in HCC progression is extremely important for planning and formulating targeted therapy for HCC. According to the proportion of immune cells in the TME, HCC patients can be divided into four different subgroups: immune desert type, immunogenic type, innate immune type, and mesenchymal type (33). Among them, the immune desert type has the best prognosis due to the lack of immune cell infiltration, while the innate immune type has the worst prognosis due to the presence of a large number of immune cell infiltration and activated immune suppression. Interestingly, although cytotoxic $CD8^+$ T lymphocytes (CTLs) have antitumor properties, which can induce apoptosis of tumor cells by recognizing tumor-specific antigens on target cells and releasing cytotoxic enzymes and cytokines, HCC patients with highly CTLs infiltration sometimes have poorer survival outcomes because these CTLs are exhausted (9). Various cancer cell-secreted metabolites, such as Kynurenine, S-adenosyl-L-methionine (SAM), and methylthioadenosine (MTA), have been reported to lead to T cell exhaustion (34–36). Exhaustion $CD8^+$ T cells may serve as a novel biomarker for efficacy monitoring during immunotherapy in HCC patients (37). Therefore, reducing the proportion of exhausted T cells in the TME or relieving the exhausted state of T cells may become the next frontier of HCC immunotherapy.

In this research, after the relative abundance of 28 immune cell subtypes was assessed by the ssGSEA algorithm in the TCGA, ICGC, and GSE14520 cohorts, we used WGCNA to create a scale-free co-expression network to find the gene modules most relevant for Tregs infiltration abundance. The overlapping genes of the candidate genes in the three datasets were finally confirmed as Tregs-related genes and mainly enriched in the immune-

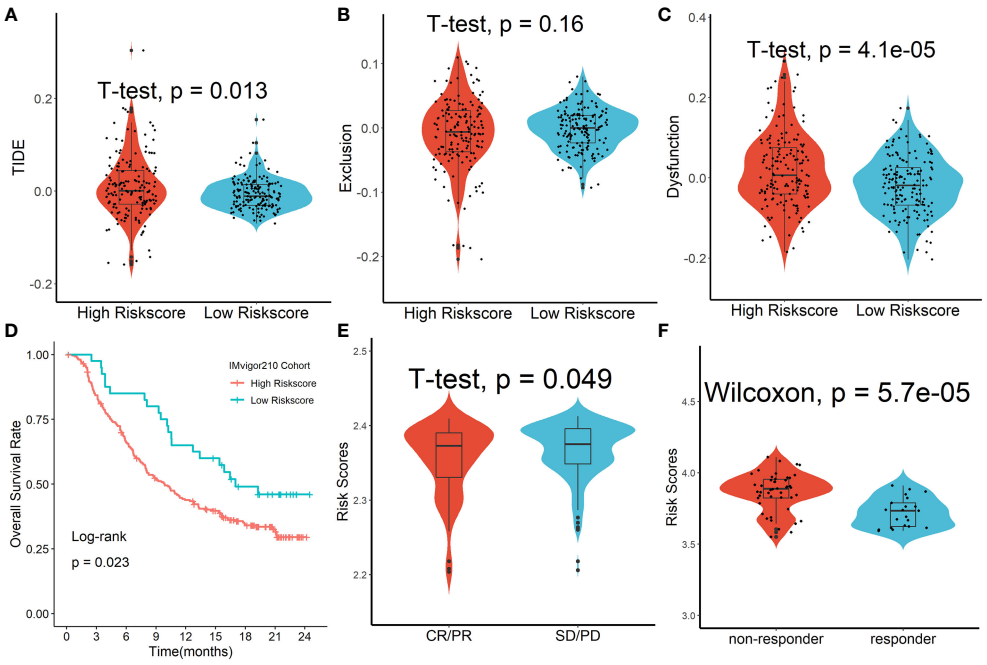


FIGURE 9
Forecasting response to anti-PD-L1 therapy using the Tregs-related signature. **(A)** TIDE scores in the high-risk scores group were much greater than that in the low-risk scores group. **(B)** T-cell exclusion scores did not differ significantly between the two groups. **(C)** T-cell dysfunction scores were greater in the high-risk scores group than in the low-risk scores group. **(D)** Patients with low risk ratings had a significant survival advantage over those with high risk scores. **(E)** Patients who had a complete response/partial response (CR/PR) had lower risk scores than patients who had stable disease/progressive disease (SD/PD). **(F)** Patients in the high-risk scores group had worse Sorafenib treatment efficacy.

inflammatory response and regulation of lymphocytes. Based on these Tregs-related genes, we divided patients into two clusters with differences in survival rates and mutation rates of mutated genes using the NMF algorithm. Compared with patients in Cluster 1, patients in Cluster 2 not only had higher immune, stroma, and estimated scores but also had higher proportions of almost all types of immune cells. Interestingly, we found that CD8+ T cells in Cluster 2 were exhausted T cells and subsequently confirmed this finding by analyzing the expression levels of PD1, PD-L1, and genes involved in immune/inflammatory responses and performing GSEA enrichment analysis. To further explore the specific mechanisms of these Tregs-related genes in HCC, we constructed a Tregs-related prognostic score model using LASSO-Cox regression. The model can not only stratify the prognosis of HCC patients well, but also effectively predict the 1-, 3-, and 5-year survival rates of patients. This Tregs-related signature was also verified in external ICGC and GSE14520 datasets. In addition, we found that CD8+ T cells in the high-risk scores group were exhausted T cells and subsequently confirmed this finding by analyzing the expression levels of PD1, PD-L1, and genes involved in immune/inflammatory responses and performing GSEA enrichment analysis. Finally, the Tregs-related prognostic score model was verified in collected fresh frozen tumor biopsies and their surrounding normal tissues by qRT-PCR and IHC analysis. Excitingly, we also observed an exhausted state of T cells in the tissue of HCC patients with high risk scores.

As a target gene of miRNA-15a-3p, heme oxygenase 1 (HMOX1) may play a role in the development and progression of HCC and is strongly correlated with the poor prognosis of HCC patients (38). HMOX1 worked in conjunction with genes involved in iron metabolism and the hypoxia phenotype to forecast patient outcomes and the effectiveness of immunotherapy (39, 40). In addition, HMOX1 has been linked to the recurrence of cancer in rats following ischemic liver transplantation (41) and can inhibit the immunomodulatory effect of Treg cells through carbon monoxide produced during metabolism (42). HMOX1 inhibitors enhance the anti-tumor effects of anti-PD-L1 antibodies in mouse melanoma and also reduce tumor size by abolishing resistance to anti-PD1 immunotherapy in female mice bearing E0771 mammary tumors (43). HMOX1 has also been implicated in PD1-involved exhausted T-cell metabolic regulation in melanoma (44). The overexpressed receptor tyrosine kinase c-Mett in renal cancer cells can inhibit cancer cell apoptosis by regulating the synergistic effect between HMOX1 and PD-L1 (45). The poor prognosis of NAFLD patients as well as HCC patients is impacted by MMP9 overexpression (46–48). Additionally, MMP9 can work with

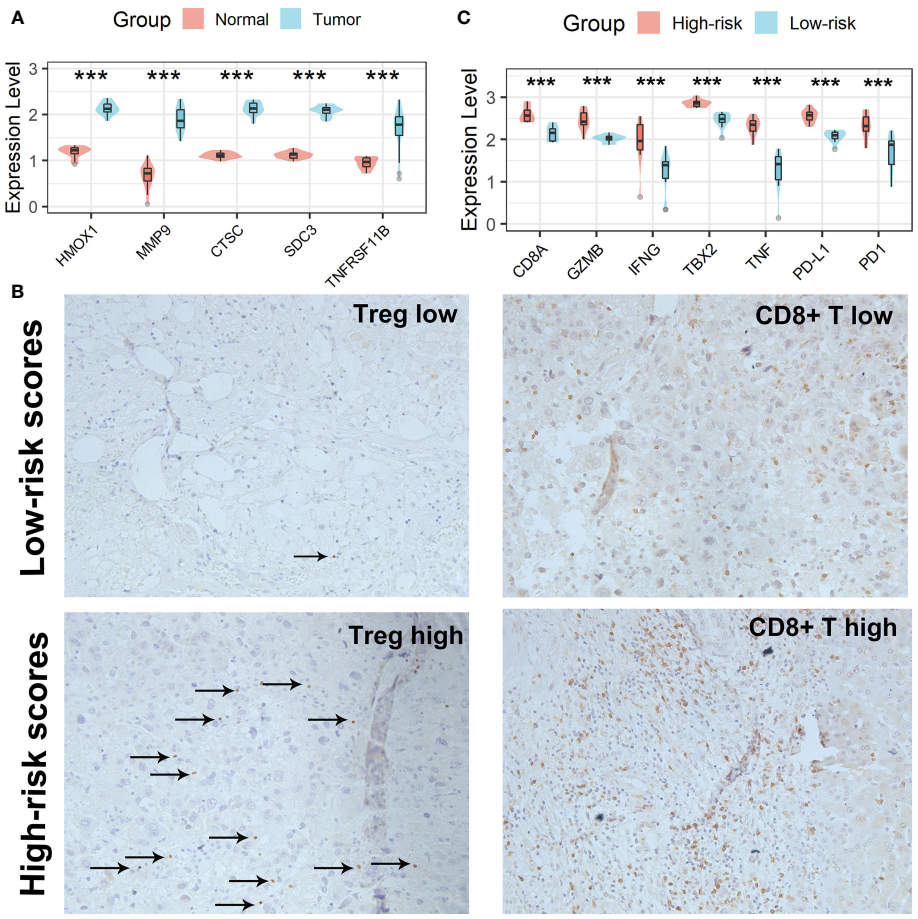


FIGURE 10
Verification of the Tregs-related signature in clinical samples. **(A)** All five Tregs-related genes revealed differential expression between normal and tumor tissues. **(B)** Patients in the high-risk scores group had a higher abundance of Tregs and $CD8^+$ T cells infiltration. **(C)** Differential analysis of inflammation/immune response-related genes. ***p < 0.001.

several signaling pathways to encourage the development and spread of HCC (49–52). As a crucial cytokine, MMP9 can play a role in the control of the Th17/Treg immunological imbalance (53). MMP9 was significantly positively correlated with PD-L1 and promoted poor prognosis in patients with tongue squamous cell carcinoma and colorectal cancer (54, 55). MMP9 can significantly increase PD-L1 expression by activating TGF-β-induced epithelial-to-mesenchymal (EMT) (56, 57). By raising CD8+ T cell cytotoxicity, MMP9 inhibitors can boost the therapeutic efficacy of PD-1 inhibition (58). Ayse identified significant changes in intratumoral MMP9 expression during anti-PD1 therapy in breast cancer patients using single-cell sequencing technology (59). Cathepsin C (CTSC), a lysosomal cysteine protease that is highly expressed in several tissues and a member of the papain superfamily, is essential for many biological activities. According to reports, CTSC speeds up the growth of some tumor types (60). Through the TNF-/MAPK (p38) pathway, up-regulated CTSCs in HCC have been demonstrated to promote HCC proliferation and metastasis (61). By controlling neutrophil infiltration and the development of neutrophil extracellular traps, CTSC facilitates breast cancer lung metastases (62). Gastric and colon cancer growth can be slowed down by CTSC silencing by promoting apoptosis (63, 64). A vital member of the SDC family, syndecan 3 (SDC3) is essential for cell adhesion, migration, and development. SDC3 expression is boosted by hypoxia in the tumor microenvironment, which influences pro-inflammatory reactions and the overall survival of melanoma patients (65). Additionally, SDC3 was linked to more dangerous tumors and a worse prognosis in prostate cancer (66). As a gene associated with dendritic cells, SDC3 is also important in developing a risk model for predicting the prognosis of HCC (67). The anti-apoptotic activity of TNF receptor superfamily member 11B (TNFRSF11B) can bind to and suppress TRAIL (TNF-related apoptosis-inducing ligand), which inhibits the spread of HCC and improves patient prognosis (68, 69). TNFRSF11B is

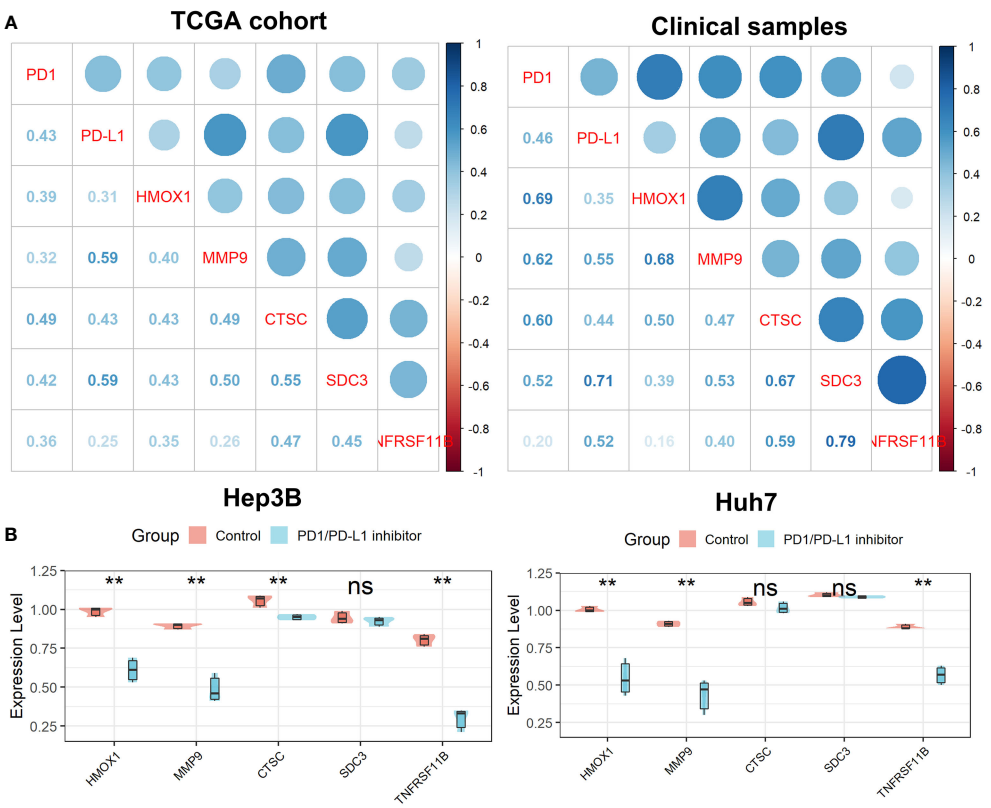


FIGURE 11
Verification of the Tregs-related signature after PD-L1 blockade. **(A)** All five Tregs-related genes were related to PD1 and PD-L1 expression. **(B)** Differential expression of the five Tregs-related genes after PD-L1 blockade. ns, not significant; **p <0.01.

significantly upregulated in peripheral blood mononuclear cells of chronic hepatitis C virus-infected patients and has been implicated in PD1-mediated T cell exhaustion and biological processes related to apoptotic signaling (70).

In this study, we found that all five genes can be acted as Treg cell-related genes to predict the prognosis and immunotherapy effect of HCC patients, and the expressions of HMOX1, MMP9, and TNFRSF11B were significantly reduced in both Hep3B and Huh7 cells after PD1/PD-L1 inhibitor treatment, suggesting that there is a certain synergy between these genes and anti- PD1/PD-L1 antibodies effect. Of course, more *in vitro* and *in vivo* studies are needed to verify the relationship and mechanism between HMOX1, MMP9, and TNFRSF11B and anti-PD1/PD-L1 therapy in HCC. In future work, we will construct a subcutaneous tumor model in C57 mice. In a nutshell, mice received subcutaneous injections of 1 x 10^5 Hep3B and Huh7 tumor cells. After that, mice with tumors measuring 100 mm^3 or larger were divided into four groups and given various treatments: control treatment with the PD1/PD-L1 inhibitor, treatment with HMOX1/MMP9/TNFRSF11B antibody, and the combination treatment of PD1/PD-L1 inhibitor and HMOX1/MMP9/TNFRSF11B antibody. Tumor development was monitored every three days while anti- PD1/PD-L1 therapies were given every three days and HMOX1/MMP9/TNFRSF11B antibody treatments were given every day until tumor capture on the ninth day. Finally, we observe the changes in tumor volume, the expression of inflammation-related genes, and tumor immune cell infiltration in the tumor to explore the synergistic mechanism of HMOX1/MMP9/TNFRSF11B and PD1/PD-L1 in HCC. Therefore, a deeper understanding of their mechanisms can help us dissect the complex relationship between the tumor microenvironment, the efficacy of anti-PD1/PD-L1antibody immunotherapy, and HCC.

Immune checkpoint blockade response in HCC patients can be predicted using the TIDE score. Patients in the low-risk score group had lower TIDE scores, which suggests that they may respond to ICIs better (71). In various cancer types, ICIs-related immunotherapy, particularly PD-1/PD-L1, has shown good therapeutic effectiveness in reversing local immunosuppression in the TME. For tumor patients with significant immune cell infiltration but compromised immunity, such as those in our study's low-risk score group, PD-1/PD-L1 inhibitors are appropriate. The IMvigor210 dataset was utilized as an external anti-PD-L1 cohort for this study to see if our created

Tregs-related risk score can predict patient response to anti-PD-L1 medication. When compared to patients in the high-risk scores group, we discovered that patients in the low-risk scores group had substantial clinical and survival advantages. Finally, we evaluated how well Tregs-related risk scores predicted the effectiveness of sorafenib in patients with HCC. Patients with low risk scores performed better with sorafenib, according to our research. Anti-PD1/PD-L1 with sorafenib may be a viable option for HCC patients in the low-risk score category to improve their prognosis.

The efficacy and prognosis of HCC patients receiving immunotherapy are significantly impacted by drug resistance, which is mostly caused by the complexity and diversity of TME components. Promising treatment now involves reducing the number of tumor-associated macrophages (TAMs) by preventing monocyte recruitment to the TME, eliminating invading TAMs, or re-educating TAMs to the more pro-inflammatory M1 subtype (72). By attracting macrophages and Tregs to the TME of HCC, tumor-associated neutrophils (TANs) cause sorafenib resistance in HCC patients, encouraging cancer development and post-treatment recurrence (73). Targeting Tregs may modify TME composition and speed tumor remission, but it may also result in significant systemic autoimmunity and inflammation (74). Immune-checkpoint molecules, such as PD1, are expressed by tumor-infiltrating Tregs at levels that rely on the TME, suggesting that PD1 inhibitors may have an impact on Tregs infiltration (75). It is clear that PD1 signaling lessens the immunosuppressive effect of Tregs since PD1-deficient Tregs or PD1 inhibition had enhanced immunosuppressive activity that was sufficient to reverse the auto-immune phenotype (76). Furthermore, human glioblastoma tissues with high PD1 expression levels in Tregs have a fatigued phenotype, which is linked to diminished immunosuppressive activity (77). When PD1 on Tregs interacts directly with PD-L1 on $CD8^+$ T cells, immunosuppressive effects are directly mediated, and PD1 inhibitors can drastically reduce these effects (78). Monitoring tumor-infiltrating Tregs alterations in patients taking PD1 inhibitors may be important since PD1 appears to have a detrimental influence on Tregs-mediated immunosuppression in tumors and anti-PD1 medication appears to increase Tregs activity (79). For HCC patients, comprehensive immunotherapy targeting both tumor cells and immunosuppressive cells in the TME may become the treatment of choice, and it will likely play a significant guiding role in the choice of patient-specific immunotherapy regimens in the future.

Undoubtedly, our study has certain flaws. First off, it was hard to fully investigate the effect of our Tregs-related signature on the prognosis of HCC patients due to the small number of HCC tissues we gathered and the paucity of survival data. To verify the precision of our prognostic model, we require a prospective multicenter investigation with a bigger sample size. Additionally, the outcomes of single-cell sequencing can aid in our understanding of how the Tregs-related genes have changed in the TME. Finally, to further understand the molecular processes by which Tregs-related genes influence HCC development, functional tests (*in vitro* and *in vivo*) should be carried out in the future.

Conclusions

In summary, our study uncovered and validated a Tregs-related prognostic model that could identify TME- exhausted subpopulations and revealed that PD1/PD-L1 inhibitors could alter the expression levels of HMOX1, MMP9, and TNFRSF11B in Hep3B and Huh7 cells, which might help us better understand Tregs infiltration and develop personalized immunotherapy treatments for HCC patients.

Data availability statement

We can find the datasets analyzed in this study at the https://xena.ucsc.edu, https://dcc.icgc.org/projects/LIRI-JP, and https://www.ncbi.nlm.nih.gov/geo. The accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committees of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

GZ designed and experimentally validated this study.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.975762/full#supplementary-material

References

1. Greten FR, Grivennikov SI. Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity* (2019) 51(1):27–41. doi: 10.1016/j.immuni.2019.06.025

2. Baghban R, Roshangar L, Jahanban-Esfahlan R, Seidi K, Ebrahimi-Kalan A, Jaymand M, et al. Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication Signaling CCS* (2020) 18(1):59. doi: 10.1186/s12964-020-0530-4

3. Neophytou CM, Panagi M, Stylianopoulos T, Papageorgis P. The role of tumor microenvironment in cancer metastasis: Molecular mechanisms and therapeutic opportunities. *Cancers* (2021) 13(9):2053. doi: 10.3390/cancers13092053

4. Schito L, Semenza GL. Hypoxia-inducible factors: Master regulators of cancer progression. *Trends Cancer* (2016) 2(12):758–70. doi: 10.1016/j.trecan.2016.10.016

5. Babajani A, Soltani P, Jamshidi E, Farjoo MH, Niknejad H. Recent advances on drug-loaded mesenchymal stem cells with anti-neoplastic agents for targeted treatment of cancer. *Front Bioengineering Biotechnol* (2020) 8:748. doi: 10.3389/fbioe.2020.00748

6. Motz GT, Coukos G. Deciphering and reversing tumor immune suppression. *Immunity* (2013) 39(1):61–73. doi: 10.1016/j.immuni.2013.07.005

7. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat Med* (2016) 22(6):679–84. doi: 10.1038/nm.4086

8. Colbeck EJ, Jones E, Hindley JP, Smart K, Schulz R, Browne M, et al. Treg depletion licenses T cell-driven HEV neogenesis and promotes tumor destruction. *Cancer Immunol Res* (2017) 5(11):1005–15. doi: 10.1158/2326-6066.CIR-17-0131

9. Sun Y, Wu L, Zhong Y, Zhou K, Hou Y, Wang Z, et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* (2021) 184(2):404–421.e416. doi: 10.1016/j.cell.2020.11.041

10. Maeda Y, Nishikawa H, Sugiyama D, Ha D, Hamaguchi M, Saito T, et al. Detection of self-reactive $CD8^+$ T cells with an anergic phenotype in healthy individuals. *Sci (New York NY)* (2014) 346(6216):1536–40. doi: 10.1126/science.aaa1292

11. Wang H, Zhang H, Wang Y, Brown ZJ, Xia Y, Huang Z, et al. Regulatory T-cell and neutrophil extracellular trap interaction contributes to carcinogenesis in non-alcoholic steatohepatitis. *J Hepatol* (2021) 75(6):1271–83. doi: 10.1016/j.jhep.2021.07.032

12. Dituri F, Mancarella S, Serino G, Chaoul N, Lupo LG, Villa E, et al. Direct and indirect effect of TGFβ on treg transendothelial recruitment in HCC tissue microenvironment. *Int J Mol Sci* (2021) 22(21):11765. doi: 10.3390/ijms222111765

13. Gao Q, Qiu SJ, Fan J, Zhou J, Wang XY, Xiao YS, et al. Intratumoral balance of regulatory and cytotoxic T cells is associated with prognosis of hepatocellular carcinoma after resection. *J Clin Oncol Off J Am Soc Clin Oncol* (2007) 25(18):2586–93. doi: 10.1200/JCO.2006.09.4565

14. Zhang C, Gao Y, Du C, Markowitz GJ, Fu J, Zhang Z, et al. Hepatitis b-induced IL8 promotes hepatocellular carcinoma venous metastasis and intrahepatic treg accumulation. *Cancer Res* (2021) 81(9):2386–98. doi: 10.1158/0008-5472.CAN-20-3453

15. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Sci (New York NY)* (2018) 359(6382):1350–5. doi: 10.1126/science.aar4060

16. Kamada T, Togashi Y, Tay C, Ha D, Sasaki A, Nakamura Y, et al. PD-1(+) regulatory T cells amplified by PD-1 blockade promote hyperprogression of cancer. *Proc Natl Acad Sci USA* (2019) 116(20):9999–10008. doi: 10.1073/pnas.1822001116

17. Kumagai S, Togashi Y, Kamada T, Sugiyama E, Nishinakamura H, Takeuchi Y, et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nat Immunol* (2020) 21(11):1346–58. doi: 10.1038/s41590-020-0769-3

18. Yi C, Chen L, Lin Z, Liu L, Shao W, Zhang R, et al. Lenvatinib targets FGF receptor 4 to enhance antitumor immune response of anti-programmed cell death-1 in HCC. *Hepatol (Baltimore Md)* (2021) 74(5):2544–60. doi: 10.1002/hep.31921

19. Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, et al. ImmuCellAI: A unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Advanced Sci (Weinheim Baden-Wurttemberg Germany)* (2020) 7(7):1902880. doi: 10.1002/advs.201902880

20. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45(W1):W98–102. doi: 10.1093/nar/gkx247

22. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6

23. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* (2012) 72(14):3499–511. doi: 10.1158/0008-5472.CAN-12-1370

24. Colwill K, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* (2011) 8(7):551–8. doi: 10.1038/nmeth.1607

25. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* (2017) 77(21):e108–10. doi: 10.1158/1538-7445.AM2017-108

26. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* (2015) 21(8):938–45. doi: 10.1038/nm.3909

27. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5

28. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. *Signal transduction targeted Ther* (2020) 5(1):28. doi: 10.1038/s41392-020-0134-x

29. Xu R, Wu Q, Gong Y, Wu Y, Chi Q, Sun D, et al. A novel prognostic target-gene signature and nomogram based on an integrated bioinformatics analysis in hepatocellular carcinoma. *Biocell* (2022) 46(5):1261–88. doi: 10.32604/biocell.2022.018427

30. Wang L, Luo Z, Sun M, Yuan Q, Zou Y, Fu D, et al. Identification of a three-gene signature in the triple-negative breast cancer. *Biocell* (2022) 46(3):595–606. doi: 10.32604/biocell.2022.017337

31. Wu Z, Li S, Zhu X. The mechanism of stimulating and mobilizing the immune system enhancing the anti-tumor immunity. *Front Immunol* (2021) 12:682435. doi: 10.3389/fimmu.2021.682435

32. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* (2020) 20(11):651–68. doi: 10.1038/s41577-020-0306-5

33. Gao X, Huang H, Wang Y, Pan C, Yin S, Zhou L, et al. Tumor immune microenvironment characterization in hepatocellular carcinoma identifies four prognostic and immunotherapeutically relevant subclasses. *Front Oncol* (2020) 10:610513. doi: 10.3389/fonc.2020.610513

34. Liu Y, Liang X, Dong W, Fang Y, Lv J, Zhang T, et al. Tumor-repopulating cells induce PD-1 expression in CD8(+) T cells by transferring kynurenine and AhR activation. *Cancer Cell* (2018) 33(3):480–494.e487. doi: 10.1016/j.ccell.2018.02.005

35. Rad Pour S, Morikawa H, Kiani NA, Yang M, Azimi A, Shafi G, et al. Exhaustion of CD4+ T-cells mediated by the kynurenine pathway in melanoma. *Sci Rep* (2019) 9(1):12150. doi: 10.1038/s41598-019-48635-x

36. Hung MH, Lee JS, Ma C, Diggs LP, Heinrich S, Chang CW, et al. Tumor methionine metabolism drives T-cell exhaustion in hepatocellular carcinoma. *Nat Commun* (2021) 12(1):1455. doi: 10.1038/s41467-021-21804-1

37. Barsch M, Salié H, Schlaak AE, Zhang Z, Hess M, Mayer LS, et al. T-Cell exhaustion and residency dynamics inform clinical outcomes in hepatocellular carcinoma. *J Hepatol* (2022) 77(2):397–409. doi: 10.1016/j.jhep.2022.02.032

38. Jiang C, He ZL, Hu XH, Ma PY. MiRNA-15a-3p inhibits the metastasis of hepatocellular carcinoma by interacting with HMOX1. *Eur Rev Med Pharmacol Sci* (2020) 24(24):12694–700. doi: 10.26355/eurrev_202012_24167

39. Shen Y, Li X, Zhao B, Xue Y, Wang S, Chen X, et al. Iron metabolism gene expression and prognostic features of hepatocellular carcinoma. *J Cell Biochem* (2018) 119(11):9178–204. doi: 10.1002/jcb.27184

40. Jiang HY, Ning G, Wang YS, Lv WB. Ahypoxia-related signature enhances the prediction of the prognosis in hepatocellular carcinoma patients and correlates with sorafenib treatment response. *Am J Trans Res* (2020) 12(12):7762–81.

41. Oldani G, Crowe LA, Orci LA, Slits F, Rubbia-Brandt L, de Vito C, et al. Pre-retrieval reperfusion decreases cancer recurrence after rat ischemic liver graft transplantation. *J Hepatol* (2014) 61(2):278–85. doi: 10.1016/j.jhep.2014.03.036

42. Brusko TM, Wasserfall CH, Agarwal A, Kapturczak MH, Atkinson MA. An integral role for heme oxygenase-1 and carbon monoxide in maintaining peripheral tolerance by CD4+CD25+ regulatory T cells. *J Immunol (Baltimore Md 1950)* (2005) 174(9):5181–6. doi: 10.4049/jimmunol.174.9.5181

43. Khojandi N, Kuehm LM, Piening A, Donlin MJ, Hsueh EC, Schwartz TL, et al. Oxidized lipoproteins promote resistance to cancer immunotherapy independent of patient obesity. *Cancer Immunol Res* (2021) 9(2):214–26. doi: 10.1158/2326-6066.CIR-20-0358

44. Hu X, Que W, Hirano H, Wang Z, Nozawa N, Ishii T, et al. 5-aminolevulinic acid/sodium ferrous citrate enhanced the antitumor effects of programmed cell death-ligand 1 blockade by regulation of exhausted T cell metabolism in a melanoma model. *Cancer Sci* (2021) 112(7):2652–63. doi: 10.1111/cas.14930

45. Balan M, Mier y Teran E, Waaga-Gasser AM, Gasser M, Choueiri TK, Freeman G, et al. Novel roles of c-met in the survival of renal cancer cells through the regulation of HO-1 and PD-L1 expression. *J Biol Chem* (2015) 290(13):8110–20. doi: 10.1074/jbc.M114.612689

46. Coilly A, Desterke C, Guettier C, Samuel D, Chiappini F. FABP4 and MMP9 levels identified as predictive factors for poor prognosis in patients with nonalcoholic fatty liver using data mining approaches and gene expression analysis. *Sci Rep* (2019) 9(1):19785. doi: 10.1038/s41598-019-56235-y

47. Xu Y, Tian H, Luan CG, Sun K, Bao PJ, Zhang HY, et al. Telocytes promote hepatocellular carcinoma by activating the ERK signaling pathway and miR-942-3p/MMP9 axis. *Cell Death Discov* (2021) 7(1):209. doi: 10.1038/s41420-021-00592-z

48. Sun SJ, Wang N, Sun ZW, Chen J, Cui HW. MiR-5692a promotes the invasion and metastasis of hepatocellular carcinoma *via* MMP9. *Eur Rev Med Pharmacol Sci* (2018) 22(15):4869–78. doi: 10.26355/eurrev_201808_15623

49. Gao X, Qiao X, Xing X, Huang J, Qian J, Wang Y, et al. Matrix stiffness-upregulated MicroRNA-17-5p attenuates the intervention effects of metformin on HCC invasion and metastasis by targeting the PTEN/PI3K/Akt pathway. *Front Oncol* (2020) 10:1563. doi: 10.3389/fonc.2020.01563

50. Lan S, Zheng X, Hu P, Xing X, Ke K, Wang F, et al. Moesin facilitates metastasis of hepatocellular carcinoma cells by improving invadopodia formation and activating β-catenin/MMP9 axis. *Biochem Biophys Res Commun* (2020) 524 (4):861–8. doi: 10.1016/j.bbrc.2020.01.157

51. Wen Y, Cai X, Chen S, Fu W, Chai D, Zhang H, et al. 7-Methoxy-1-Tetralone induces apoptosis, suppresses cell proliferation and migration in hepatocellular carcinoma *via* regulating c-met, p-AKT, NF-κB, MMP2, and MMP9 expression. *Front Oncol* (2020) 10:58. doi: 10.3389/fonc.2020.00058

52. Lin XL, Li K, Yang Z, Chen B, Zhang T. Dulcitol suppresses proliferation and migration of hepatocellular carcinoma *via* regulating SIRT1/p53 pathway. *Phytomedicine Int J Phytotherapy Phytopharmacology* (2020) 66:153112. doi: 10.1016/j.phymed.2019.153112

53. Fan Q, Liu Y, Rao J, Zhang Z, Xiao W, Zhu T, et al. Anti-atherosclerosis effect of angong niuhuang pill *via* regulating Th17/Treg immune balance and inhibiting chronic inflammatory on ApoE(-/-) mice model of early and mid-term atherosclerosis. *Front Pharmacol* (2019) 10:1584. doi: 10.3389/fphar.2019.01584

54. Furukawa K, Kawasaki G, Yoshida T, Umeda M. Clinicopathological and prognostic analysis of PD-L1 and PD-L2 expression in surgically resected primary tongue squamous cell carcinoma. *Anticancer Res* (2021) 41(1):101–11. doi: 10.21873/anticanres.14755

55. Zhang Z, Wu Q, Zhu D, He G, Feng Q, Xu J. Tumor microenvironment derived signature predicting relapse-free survival in I-III cancer and preliminary experiment verification. *Int Immunopharmacol* (2021) 91:107243. doi: 10.1016/j.intimp.2020.107243

56. Costanza B, Umelo IA, Bellier J, Castronovo V, Turtoi A. Stromal modulators of TGF-β in cancer. *J Clin Med* (2017) 6(1):7. doi: 10.3390/jcm6010007

57. Evanno E, Godet J, Piccirilli N, Guilhot J, Milin S, Gombert JM, et al. Tri-methylation of H3K79 is decreased in TGF-β1-induced epithelial-to-mesenchymal transition in lung cancer. *Clin Epigenet* (2017) 9:80. doi: 10.1186/s13148-017-0380-0

58. Ye Y, Kuang X, Xie Z, Liang L, Zhang Z, Zhang Y, et al. Small-molecule MMP2/MMP9 inhibitor SB-3CT modulates tumor immune surveillance by regulating PD-L1. *Genome Med* (2020) 12(1):83. doi: 10.1186/s13073-020-00780-z

59. Bassez A, Vos H, Van Dyck L, Floris G, Arijs I, Desmedt C, et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med* (2021) 27(5):820–32. doi: 10.1038/s41591-021-01323-8

60. Korkmaz B, Lamort AS, Domain R, Beauvillain C, Gieldon A, Yildirim A, et al. Cathepsin c inhibition as a potential treatment strategy in cancer. *Biochem Pharmacol* (2021) 194:114803. doi: 10.1016/j.bcp.2021.114803

61. Zhang GP, Yue X, Li SQ. Cathepsin c interacts with TNF-α/p38 MAPK signaling pathway to promote proliferation and metastasis in hepatocellular carcinoma. *Cancer Res Treat* (2020) 52(1):10–23. doi: 10.4143/crt.2019.145

62. Xiao Y, Cong M, Li J, He D, Wu Q, Tian P, et al. Cathepsin c promotes breast cancer lung metastasis by modulating neutrophil infiltration and neutrophil extracellular trap formation. *Cancer Cell* (2021) 39(3):423–437.e427. doi: 10.1016/j.ccell.2020.12.012

63. Kim S, Lee SI, Kim N, Joo M, Lee KH, Lee MW, et al. Decursin inhibits cell growth and autophagic flux in gastric cancer *via* suppression of cathepsin c. *Am J Cancer Res* (2021) 11(4):1304–20.

64. Khaket TP, Singh MP, Khan I, Kang SC. *In vitro* and *in vivo* studies on potentiation of curcumin-induced lysosomal-dependent apoptosis upon silencing of cathepsin c in colorectal cancer cells. *Pharmacol Res* (2020) 161:105156. doi: 10.1016/j.phrs.2020.105156

65. Prieto-Fernández E, Egia-Mendikute L, Bosch A, García Del Río A, Jimenez-Lasheras B, Antoñana-Vildosola A, et al. Hypoxia promotes syndecan-3 expression in the tumor microenvironment. *Front Immunol* (2020) 11:586977. doi: 10.3389/fimmu.2020.586977

66. Santos NJ, Barquilha CN, Barbosa IC, Macedo RT, Lima FO, Justulin LA, et al. Syndecan family gene and protein expression and their prognostic values for prostate cancer. *Int J Mol Sci* (2021) 22(16):8669. doi: 10.3390/ijms22168669

67. Huang C, Jiang X, Huang Y, Zhao L, Li P, Liu F. Identifying dendritic cell-related genes through a Co-expression network to construct a 12-gene risk-scoring model for predicting hepatocellular carcinoma prognosis. *Front Mol Biosci* (2021) 8:636991. doi: 10.3389/fmolb.2021.636991

68. Gao YB, Xiang ZL, Zhou LY, Wu ZF, Fan J, Zeng HY, et al. Enhanced production of CTGF and IL-11 from highly metastatic hepatoma cells under hypoxic conditions: an implication of hepatocellular carcinoma metastasis to bone. *J Cancer Res Clin Oncol* (2013) 139(4):669–79. doi: 10.1007/s00432-012-1370-4

69. Sekar D, K AP. Letter to the Editor: H19 promotes HCC bone metastasis by reducing osteoprotegerin expression in a PPP1CA/p38MAPK-dependent manner and sponging miR-200b-3p. *Hepatol (Baltimore Md)* (2021) 74(3):1713. doi: 10.1002/hep.31719

70. Barathan M, Gopal K, Mohamed R, Ellegård R, Saeidi A, Vadivelu J, et al. Chronic hepatitis c virus infection triggers spontaneous differential expression of biosignatures associated with T cell exhaustion and apoptosis signaling in peripheral blood mononucleocytes. *Apoptosis an Int J Programmed Cell Death* (2015) 20(4):466–80. doi: 10.1007/s10495-014-1084-y

71. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* (2019) 18 (3):197–218. doi: 10.1038/s41573-018-0007-y

72. Cendrowicz E, Sas Z, Bremer E, Rygiel TP. The role of macrophages in cancer development and therapy. *Cancers* (2021) 13(8):1946. doi: 10.3390/cancers13081946

73. Zhou SL, Zhou ZJ, Hu ZQ, Huang XW, Wang Z, Chen EB, et al. Tumor-associated neutrophils recruit macrophages and T-regulatory cells to promote progression of hepatocellular carcinoma and resistance to sorafenib. *Gastroenterology* (2016) 150(7):1646–58.e1617. doi: 10.1053/j.gastro.2016.02.040

74. Wang Z, He L, Li W, Xu C, Zhang J, Wang D, et al. GDF15 induces immunosuppression via CD48 on regulatory T cells in hepatocellular carcinoma. *J immunotherapy Cancer* (2021) 9(9):e002787. doi: 10.1136/jitc-2021-002787

75. Romano E, Kusio-Kobialka M, Foukas PG, Baumgaertner P, Meyer C, Ballabeni P, et al. Ipilimumab-dependent cell-mediated cytotoxicity of regulatory T cells ex vivo by nonclassical monocytes in melanoma patients. *Proc Natl Acad Sci USA* (2015) 112(19):6140–5. doi: 10.1073/pnas.1417320112

76. Zhang B, Chikuma S, Hori S, Fagarasan S, Honjo T. Nonoverlapping roles of PD-1 and FoxP3 in maintaining immune tolerance in a novel autoimmune pancreatitis mouse model. *Proc Natl Acad Sci USA* (2016) 113(30):8490–5. doi: 10.1073/pnas.1608873113

77. Lowther DE, Goods BA, Lucca LE, Lerner BA, Raddassi K, van Dijk D, et al. PD-1 marks dysfunctional regulatory T cells in malignant gliomas. *JCI Insight* (2016) 1(5):e85935. doi: 10.1172/jci.insight.85935

78. Park HJ, Park JS, Jeong YH, Son J, Ban YH, Lee BH, et al. PD-1 upregulated on regulatory T cells during chronic virus infection enhances the suppression of CD8+ T cell immune response *via* the interaction with PD-L1 expressed on CD8+ T cells. *J Immunol (Baltimore Md 1950)* (2015) 194(12):5801–11. doi: 10.4049/jimmunol.1401936

79. Togashi Y, Shitara K, Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat Rev Clin Oncol* (2019) 16(6):356–71. doi: 10.1038/s41571-019-0175-7

frontiers | Frontiers in Immunology

TYPE Review
PUBLISHED 23 September 2022
DOI 10.3389/fimmu.2022.985863

Check for updates

OPEN ACCESS

EDITED BY
Ping Zheng,
The University of Melbourne, Australia

REVIEWED BY
Joaquim Carreras,
Tokai University, Japan
Juha P. Väyrynen,
University of Oulu, Finland

*CORRESPONDENCE
Jie Fu
fujie74@sjtu.edu.cn

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 04 July 2022
ACCEPTED 07 September 2022
PUBLISHED 23 September 2022

CITATION
Li Z, Yu Q, Zhu Q, Yang X, Li Z and Fu J (2022) Applications of machine learning in tumor-associated macrophages. *Front. Immunol.* 13:985863. doi: 10.3389/fimmu.2022.985863

COPYRIGHT
© 2022 Li, Yu, Zhu, Yang, Li and Fu. 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.

Applications of machine learning in tumor-associated macrophages

Zhen Li[1†], Qijun Yu[2,3†], Qingyuan Zhu[1], Xiaojing Yang[1], Zhaobin Li[1] and Jie Fu[1*]

[1]Radiation Oncology Department, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China, [2]Department of Pulmonary and Critical Care Medicine, Ruijin Hospital, Shanghai Jiao Tong University, Shanghai, China, [3]Institute of Respiratory Diseases, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Evaluation of tumor-host interaction and intratumoral heterogeneity in the tumor microenvironment (TME) is gaining increasing attention in modern cancer therapies because it can reveal unique information about the tumor status. As tumor-associated macrophages (TAMs) are the major immune cells infiltrating in TME, a better understanding of TAMs could help us further elucidate the cellular and molecular mechanisms responsible for cancer development. However, the high-dimensional and heterogeneous data in biology limit the extensive integrative analysis of cancer research. Machine learning algorithms are particularly suitable for oncology data analysis due to their flexibility and scalability to analyze diverse data types and strong computation power to learn underlying patterns from massive data sets. With the application of machine learning in analyzing TME, especially TAM's traceable status, we could better understand the role of TAMs in tumor biology. Furthermore, we envision that the promotion of machine learning in this field could revolutionize tumor diagnosis, treatment stratification, and survival predictions in cancer research. In this article, we described key terms and concepts of machine learning, reviewed the applications of common methods in TAMs, and highlighted the challenges and future direction for TAMs in machine learning.

KEYWORDS

machine learning, tumor microenvironment, tumor-associated macrophages (TAMs), deep learning, artificial intelligence

1 Introduction

The tumor microenvironment (TME) is a complex system consisting of various components that would shape tumorigenesis, progression and metastasis. In addition to cancer cells, numerous innate immune cells reside within the TME, for instance, macrophages, dendritic cells, neutrophils, myeloid-derived suppressor cells, etc. In the complex environment, tumor-associated macrophages (TAMs), the major immune cells

infiltrating tumors, can orchestrate various aspects of tumor biology, such as tumor initiation, progression, metastasis, and even anti-tumor immunosuppression. As crucial drivers in fostering tumor progression, TAMs are standing out as promising targets for diagnosis and new treatments in malignant tumors.

Machine Learning (ML) is a group of data-analytical methods to build predictive models by summarizing past empirical or theoretical literature. Deep learning (DL) is considered an evolution of machine learning. It uses a programmable artificial neural network (ANN) which is inspired by a biological nervous system to make accurate decisions. Recently, ML, DL, in particular, has exhibited a remarkable development with the support of the rapid increase in the storage capacity and processing power of computers. In the era of big data, ML methods have come to attention as their extraordinary ability to process large and heterogeneous data sets in complex biological systems. As P4 (Predictive, Preventive, Personalized, and Participatory) and precision medicine are emerging and gaining traction (1), ML has become integral to modern biological research for its ability to solve challenges not well addressed by traditional methods. There have been many applications of ML in medical research ranging from cancer classification, subtyping, new biomarker discovery, and drug discovery (2–5). Considering the crucial role of TAMs in TME and tumor biology, ML has been widely employed in TAMs-related studies and has achieved successful outcomes.

This review is intended for readers with little knowledge of ML algorithms. Firstly, we briefly review the origins, types, and functions of TAMs. Secondly, we introduce the basic principles and key concepts needed to understand how ML methods could be applied and utilized in cancer research. Thirdly, we discuss the methods and applications at the intersection of ML and TME, especially TAMs. In the end, we highlight the current challenges in ML that need to be addressed, as well as the future directions that could be used to fully realize the potential applications in cancer therapy.

2 Origins and types of TAMs

TAMs comprise almost 50% of immune cells infiltrating tumors. They are highly heterogeneous cells that can be divided into two main origins: bone-marrow-derived macrophages (BMDMs) developing from hematopoietic stem cells and tissue-resident macrophages (TRMs) from progenitors seeded into tissues during embryonic development. For a long time, BMDMs have been considered the main effectors in TAMs, but nowadays, TRMs have emerged as an inseparable and essential component in TME (6).

In a simplified view, there are two distinct populations of polarized macrophages, the classical M1 [upon lipopolysaccharide (LPS) and IFNG stimulation] and the alternative M2 (upon IL4 or IL13 stimulation) phenotypes macrophages. Macrophages undergo polarization and get activated in multiple processes during physiological and disease processes (7, 8). M1 and M2 macrophages have different markers, including CD surface receptors, cytokines, chemokines, transcription profiles, etc. (Table 1). We have listed the characterized biomarkers, CDs, and cytokines for TAMs identification. M2 macrophages can be further classified into different subtypes, namely M2a (mediated by IL4 and IL13), M2b (mediated by immune complexes (IC) with LPS or IL1R ligand), M2c (mediated by TGFB1, IL10, and

TABLE 1 M1 and M2 macrophages markers.

Characteristics	M1 (classical)	Reference	M2(alternative)	Reference
Stimuli	LPS/IFNG/CSF2	(9, 10)	IL4/IL13/CSF1	(10, 11)
CDs and MHC	CD68, CD80, CD86, MHC-II	(12–14)	CD68, CD204, CD163, CD206	(15)
Cytokines and Chemokines	IL1B, IL6, IL12, TNF, IFNG CXCL9, CXCL10, CXCL11, CXCL16	(9, 11) (9, 16),	IL10, VEGFA/C, TGFB1 CCL17, CCL18, CCL22, CCL24	(9, 15)
Non-coding RNAs	miR-125b-2	(17)	miR-375	(18)
	miR-16	(19)	miR-34a	(20)
	miR-9	(21)	miR-301a	(22)
	lncRNA-PVT1	(23)	miR-934	(24)
	lncRNA-MEG8	(25)	miR-940	(26)
	lncRNA-GAS5	(27)	let-7b	(28)
	miR-155	(29, 30)	let-7c	(31)
	miR-142-3p	(32)	let-7d-5p	(33)
	miR-146a	(14)	miR-19b-3p	(34)
			lncRNA-MM2P	(35)
Others	NOS2, ROS, HMGB1	(11, 14, 36–38),	PD-1/PD-L1, MMP1/2/9, Arg1, Chil3, Retnla	(39–43)

glucocorticoids), and M2d (activated by tumor-associated factors, the major part of TAMs) (44, 45). In contrast to proinflammatory, antibacterial, and anti-angiogenic M1 macrophages, M2 macrophages suppress inflammation, facilitate tissue repair, remodeling, angiogenesis, and retain homeostasis under physiological conditions (46, 47).

In general, TAMs contain M2 and small populations of M1 cells (48). However, the distinction between the M1 and M2 states is less clear in TME since TAMs probably display phenotypes anywhere in between these two extremes. Moreover, the phenotype of TAMs dynamically changes with the development and progression of tumors. Each macrophage in TME might show anti- or pro-tumorigenic properties to form a plastic and heterogeneous tumor-promoting totality in response to diverse microenvironmental signals (a mixed M1–M2 phenotype). In a word, the M1 or M2 only phenotype is too simple to elucidate the intricate roles of TAMs in the TME (49–53).

3 Roles of TAMs in tumor

Macrophages are considered essential components in immune defense and immune sentinels combating tumor growth; however, accumulated evidence supports a new tumor-promoting role of macrophages as well. Different from the basic functions of phagocytizing pathogens and apoptotic cell debris, TAMs are equipped to execute a broad repertoire of pro-tumorigenic functions as heterogeneous effectors (Figure 1).

3.1 TAMs in tumor initiation and development

TAMs profusely infiltrate TME with the ability to suppress anti-tumoral immune surveillance. Accumulating evidence has suggested that TAMs can express a variety of immunosuppressive chemokines and factors which promote tumor cell proliferation and survival, including platelet-derived growth factor (PDGF), epithelial growth factor (EGF), and transforming growth factor beta 1 (TGFB1) (54, 55). The abovementioned chemokines and factors lead to immune cell–cell interactions as well. For instance, TAMs can inhibit anti-tumor immunity by restraining antigen presentation and blocking T cells function, in which case T cells lose their capacity in recognizing and even killing tumor cells (45). Usually, activated cytotoxic T lymphocytes (CTLs) can attack cancer cells to suppress tumor growth, while TAMs express immunosuppressive cytokines, chemokines, and growth factors like IL10 and TGFB1 to make CTLs hyporesponsive (6). As a distinct T-cell subpopulation, regulatory T cells (Tregs) are actively engaged in the maintenance of immunological self-tolerance (56). IL10 and TGFB1 from TAMs can also induce Tregs-mediated immunosuppression (57). Besides, TAMs are able to recruit Tregs *via* CCL22 production, which further suppresses the antitumor immune response of T-cells and fosters tumor growth (58). Moreover, it is worth noting that cancer cells can strongly induce TAMs into pro-tumorigenic phenotype by secreting colony-stimulated factor 1, mucins and exosomes (59–61). To sum up, all these factors work together and make the TME a hospitable site.

3.2 TAMs in tumor angiogenesis

Angiogenesis can be briefly defined as the formation of new capillaries from pre-existing blood vessels. It is generally accepted that tumor growth largely depends on angiogenesis since new vessels can supply fresh oxygen and nutrients as well as remove wastes and metabolites. Furthermore, angiogenesis is a vital event in hematogenous metastasis (62). Angiogenesis is activated when pro-angiogenic factors predominate over anti-angiogenic factors (63). As shown in Table 1, TAMs can produce diverse pro-angiogenic molecules (VEGF family, PDFG, TGFB1, etc.) and matrix metalloproteinases (MMP) to facilitate angiogenesis. In particular, developing tumors consume oxygen supply rapidly and tend to create an oxygen deficiency condition (hypoxia). It has been increasingly recognized that TAMs massively infiltrate hypoxic regions in tumors and hypoxic macrophages achieve a pro-angiogenic response by directly upregulating the abovementioned pro-angiogenic molecules through hypoxia-inducible factor-1 alpha (HIF1A) (64–67).

3.3 TAMs in tumor metastasis

TAMs demonstrate lots of essential functions in tumor biology. In tumor metastasis, it is still a puzzling question how TAMs facilitate tumor spread specifically, though TAMs get involved in almost every process of metastasis. Herein, we provide a quick summary of the fundamental mechanics. First, TAMs within the TME can enhance tumor cell migration and invasion, thereby enabling the escape of tumor cell from the confines of the basement membrane into the surrounding tissues. Second, TAMs are associated with tumor angiogenesis, which, as was previously mentioned, results in tumor intravasation and vasculature-based tumor spread (68). Third, in the immunosuppressive TME, cancer cells can escape from being killed by T cells and prolong cell survival, which make it easier to spread to farther tissues and organs (69). It should be highlighted that tumor metastasis is a process that starts at a very early stage rather than a late event initiated and shaped in advanced cancers. Distant organs are conducive to the survival and outgrowth of primary cancer cells before their arrival. Those 'primed' sites are known as 'pre-metastatic niches' (PMNs) (70) and special attention has been given to the key role of TAMs in

PMNs from clinical evidence (71). Upon the induction of many tumor-secreted factors, TAMs are recruited into the blood and then gather at the pre-metastatic sites (70, 72–74). Meanwhile, TRMs stemming from yolk sac progenitors, like cerebral microglia, liver Kupffer cells, pulmonary alveolar macrophages, and osteoclasts, have been resident in the distant sites before tumorigenesis and get involved in orchestrating PMNs formation following diverse stimulation as well. These macrophages guide circulating tumor cells (CTCs) into the PMNs through enhancing the expression of chemokines and remodeling the extracellular matrix (ECM) into more tumor-favorable structures (75).

3.4 TAMs enhance resistance to chemotherapy, radiotherapy and immunotherapy

Emerging cancer research depicts that a high proportion of TAMs infiltration in tumor samples is often associated with shortened survival and poor prognosis in many tumors (76–79). Furthermore, TAMs infiltration is thought to offset therapeutic response to radiotherapy, chemotherapy and targeted therapy, even leading to treatment failure (80, 81). Regarding underlying mechanisms, TAMs can reduce the efficacy of radiotherapy by triggering the anti-apoptotic programs in cancer cells that are resistant to radiotherapy. They also secrete a variety of cytokines and survival factors to mediate the resistance of the solid tumor to many chemotherapy drugs, including IL6 and milk-fat globule-epidermal growth factor-VIII (82, 83). Programmed death ligand 1 (PD-L1), which is thought to be carried by TAMs and is upregulated in response to stimulation of TME-derived factors, has been linked to immune exhaustion *via* the checkpoint ligand/receptor interaction. However, existing studies do not depict a comprehensive picture since another study comes to a contrary conclusion that PD-L1 expression on TAMs, instead of cancer cells, is positively associated with patients' overall survival (84). Thus, further studies addressing the precise mechanisms involved are urgently needed.

Considering all these functions of TAMs, it is essential to comprehend heterogeneous TAMs and their roles in tumor biology to create and enhance more potent treatments. To date, various molecular strategies against TAMs are currently in preclinical or clinical trials, trying to overcome the knotty problem of immune suppression, such as TAMs recruitment, TAMs depletion and TAMs reprogramming (85).

4 Basics of machine learning

The term machine learning was first coined in the 1950s by Arthur Samuel, a computer scientist at IBM (86). Since then, ML

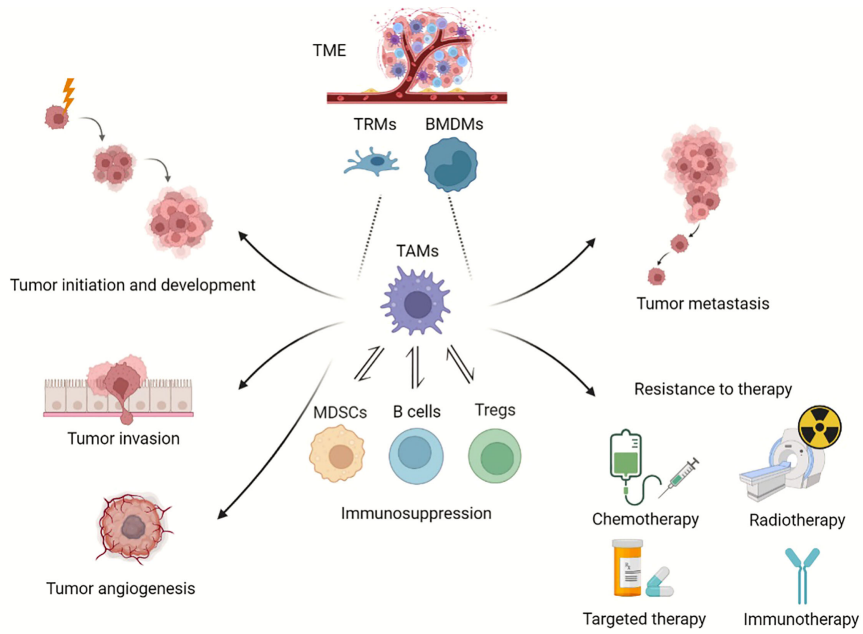


FIGURE 1
Roles of TAMs in tumor progression. Overview of TAMs in tumor progression. TAMs can derive from BMDMs and TRMs. TAMs provide a niche for tumor initiation and development, participate in angiogenesis, promote tumor metastasis, and enhance resistance to chemotherapy, radiotherapy and immunotherapy. (Created with BioRender.com).

TABLE 2 Common terminologies and explanations in ML.

Artificial Intelligence	Artificial intelligence is the capability of a computer to perform tasks that are generally completed by humans because they require human intelligence and conception.
Features	Features are the observable quantities and characteristics across all samples in the data set, either raw or mathematically transformed.
Feature selection	Feature selection is the process of selecting the most relevant features in developing a predictive model and can reduce the computational cost of modeling as well as improve the performance of the model.
Data augmentation	Data augmentation refers to techniques that can increase the diversity of training sets by applying random (but realistic) transformations, such as image rotation, flipping, scaling, etc.
Overfitting	Overfitting refers to a model that performs pretty well on the training data and fails to generalize and perform well in the case of unseen data scenarios.
Underfitting	Underfitting refers to a model that does not work correctly in the training data and also has poor performance in the new data.
Dimensionality reduction	Dimensionality reduction refers to techniques that reduce the number of random variables to the principal component of a data set.

has evolved considerably and now is playing a critical role in modern medical science. ML is a subdivision of artificial intelligence and can be briefly defined as enabling algorithms to make accurate predictions based on prior experiences (87). The boundary between conventional statistical techniques and ML is obscure, whilst some terms in ML have similar functions to statistical methods. Some conventional statistical techniques, such as ridge regression can be combined with ML algorithms for prediction (88). One key distinction between ML and traditional statistical methods is that conventional statistic methods focus on the relationship between variables (89). However, ML contributes to identifying patterns from massive data and then performing predictions. Moreover, ML aims to solve more complicated problems, often dealing with high dimensional variables with the technique of feature selection, pattern analysis and dimensionality reduction. As a result, it extends and supplements existing statistical methods by offering tools and algorithms to decipher patterns in enormous, intricate and heterogeneous data sets. Common terminologies and explanations in ML can be seen in Table 2.

In oncology studies, ML can analyze large-scale data in different format and combine them into predictions for tumor staging, cancer susceptibility, tumor recurrence, and patient survival (90). The process of ML is to extract knowledge from massive data sets, identify the underlying patterns, build predictive models, and finally make predictions on unseen data. A basic explanation of ML in cancer research can be achieved by considering the example of tumor recurrence prediction. Features from heterogeneous sources of data (clinical, imaging and genomic) are extracted by the ML algorithm. ML algorithm identify the combinations of specific features and tumor recurrence risk, and then build a prediction model. After that, when presented with a new case, the algorithm could provide the likelihood of recurrence for the new case.

TABLE 3 Categories of supervised learning and unsupervised learning for common algorithms.

Supervised Learning	Unsupervised Learning
Ordinary Least Square Regression	K-Means
Logistic Regression	Principal Component Analysis
Least Absolute Shrinkage Selection Operator Regression	Information Maximizing Component
Linear Discriminant Analysis	Self-organizing Maps
Ridge Regression	Topological Data Analysis
Elastic Net Regression	
Support Vector Machines	
Bayesian Networks	
Naïve Bayes Classifiers	
Random Forests	

4.1 Categories in machine learning

ML techniques can be generally categorized into three main groups based on whether the labels are required in the training data (91). Common categories of supervised and unsupervised learning can be found in Table 3.

4.1.1 Supervised learning

The term 'supervised' refers to the technique where a model is supplied with labels, which are desired outcomes of the learning target (e.g., correct segmentation or classification results) (92). Generally, supervised learning is used to build a model to predict or categorize future events. It primarily focuses on classification (e.g., classifying benign or malignant tumors) and regression (calculating the risk of tumor relapse, estimating individualized disease-free survival, or predicting the length of patient life) (88).

4.1.2 Unsupervised learning

Unsupervised learning is used when the input data has no labels. Hence, it learns the relationship between variables and uncovers patterns in unlabeled data. Supervised learning primarily addresses classification and regression issues, while unsupervised learning focuses more on dimensionality reduction and clustering (88). Clustering refers to identifying groups of similar cases within a data set based on some specific

features; dimensionality reduction is used to reduce the complexity and heterogeneity of features extracted from massive biomedical data sets.

4.1.3 Semi-supervised learning

Semi-supervised learning combines supervised and unsupervised ML. It can be helpful when only a tiny fraction of the data is labeled, or the labels on the input data are incomplete (93). A lack of sufficient labeled data frequently occurs in medical contexts because, given the complexity and variability of biomedical data, labeling information (e.g., correctly delineating the target in auto-segmentation) can be labor- and time-consuming. From this respective, semi-supervised learning can improve the efficiency and accuracy of information extraction for large data sets.

4.2 General workflow

4.2.1 Data preparation

ML workflow usually starts with data acquisition and pre-processing. Data sets are typically split into training, validation, and evaluation sets. The predictive model is constructed on the basis of the training set and tuned by the validation set; finally, the model performance is assessed by the held-out evaluation set (89). In practice, the training set usually accounts for a larger fraction of the data (70%), whereas validation and evaluation sets usually make up 15%, respectively.

The prerequisites of ML success are a sufficient number of samples and high-quality data. To make the most of ML, enough training data size should be ensured to extract more generic features from the whole data set without over-emphasizing the impact from a few certain samples. Besides, the data quality should be checked to ensure input data's appropriateness, reproducibility, and versatility. Specifically, for supervised learning, the correctness of the ground truth labels is also quite essential. Incorrect labels can significantly downgrade the model performance and are difficult to detect during training (86).

4.2.2 Training and validation

The proper performance of the model relies heavily on features across sample sets, and model refinement can be achieved using the technique of feature selection. Inappropriate feature selection would undermine the training performance by straining computational resources, including time and memory. For ML application in TAMs, thousands of features can be used to predict the output variables (94), e.g., cell morphology, the molecular feature of TAMs, immune-related gene-based novel subtypes, patient characteristics, tumor infiltration, etc. After feature selection, ML would search for the optimal parameters and translate the features into accurate predictions. The parameters are created through a complicated calculation process.

After that, a validation set is also needed to optimize the parameters of the algorithm. In validation, a preliminary estimate of the model's generalizability and accuracy is obtained; errors can be detected and corrected in this phase, and the process is then repeated (95). In other words, validation serves as a supplemental role in identifying the errors in a model in an early phase.

The input data is usually partitioned into k subsets of equal size. A single subset is retained as the validation set, and the remaining k-1 subsets are used as training data. The process of training and validation will continue until there is no further improvement in model performance.

4.2.3 Evaluation

The evaluation data is used to assess the performance of the final model on samples outside the input data set (training and validation set). This process aims to estimate the model performance in the real-world. The evaluation set should be utilized at the very end of the research, avoiding the model being tuned to fit the evaluation set (96). The performance of a specific model relies on many factors, such as the data size and quality of training data, as mentioned above. The complexity and the relationship between the input and output variables, as well as the computational resources such as available training time and memory, all play essential roles in achieving high model performance (94).

5 ML algorithms used in TAMs

In this section, we are going to introduce the most common utilized ML algorithms applied in cancer research, especially, TAMs. We also compared the advantages and disadvantages of different algorithms in Table 4 (97–101). Since the combination of ML and TAMs is an emerging cross-cutting research field, most studies were published in the last five years. All the matches were reviewed for suitability and significance for this review. Table 5 depicts the publications we found most pertinent to our topic. Cancer type, sample size, research purpose, as well as the ML applications are presented in the table.

5.1 Dimensionality reduction

Dimensionality reduction refers to techniques that transform data in high dimensions into a lower-dimensional form while preserving the relationships between the data points as much as possible. In a nutshell, it is a data preparation technique used for downsizing the input variables and performed before modeling. By far, Principal Component

TABLE 4 Pros and cos of common machine learning algorithms.

	Pros	Cons
Support Vector Machine	• Good performance with high dimensional data • Good performance when classes are separable	• Slow • Cannot deal with overlapped classes • Selecting appropriate hyperparameters is essential • Selecting the appropriate kernel
Principal Components Analysis	• Reduce overfitting • Improve visualization • Improve model performance	• Independent variables become less interpretable • Data standardization is necessary • Lose information
Naive Bayes	• Fast prediction • Insensitive to irrelevant features • Can be used for multi-class prediction • Perform well with high dimensional data • Less dependent to data size	• Independence of features does not hold • Relatively low prediction accuracy • Zero Frequency
Logistic Regression	• Simple to implement and interpret • Feature scaling is unnecessary • Perform well for linearly separable dataset • Tuning of hyperparameters is unnecessary • Fast at classifying unknown records	• Assumption of linearity between the dependent variable and the independent variables • Requires average or no multicollinearity between independent variables • High reliance on proper presentation of data
Random Forest	• Reduced error with high accuracy (balance the bias-variance well with multiple trees) • Good performance on imbalanced datasets • Can handle linear and non-linear relationships well • Little impact of outliers • Not prone to overfitting • Useful for feature selection	• Features need to have some predictive power • Predictions of the trees need to be uncorrelated • Not easily interpretable • Computationally intensive for large datasets • Black box nature
Decision Tree	• Normalization or data scaling is unnecessary • Can handle huge amount of data • Easy to explain • Easy visualization • Automatic Feature selection • Missing values does not affect building decision tree	• Prone to overfitting • A small change in data can cause large change in structure of decision tree • Long training time • Inadequate for applying regression and predicting continuous values
K-Nearest Neighbor	• Simple to understand and implement • No assumptions about data • Constantly evolving model • Can handle multi-class problem • One hyper-parameter(k)	• Slow • Poor performance on datasets with large number of features • Scaling is necessary • Imbalanced data causes problems • Outlier sensitivity • No capability of dealing with missing values
Artificial Neural Network	• High Efficiency • High accuracy • Multi-tasking • Able to deal with incomplete information • Having fault tolerance	• Hardware dependence • Black Box Nature • Complex algorithm compared to traditional machine learning algorithms • Need large data set

Analysis (PCA) is the most popular multidimensional data analysis technique (126). It reduces the dimensionality by eliminating less important components to omit the redundant dimensions and focusing only on the most important components that could best explain the heterogeneity in the data (Figure 2A). Other dimensionality reduction algorithms include t-distributed stochastic neighbor embedding and uniform manifold approximation and projection.

PCA is primarily applied to problems where there are a large number of features, which are referred to as high-dimensional problems (127). Generally, there are many important applications of PCA in cancer research because the input variables in oncology data are complex and massive. For example, PCA is used to extract principal components as signature score to calculate the patients' risk scores based on meaningful macrophage-related genes (105, 106). Zhang et al. performed PCA on 487 patients to reduce the feature dimensions and clearly distinguished high-risk and low-risk patients (107). Autoencoder in deep learning neural networks is another method to perform dimensionality reduction. Encoder is the part of the model prior to the bottleneck. It aims to compress the data dimension to a bottleneck layer that is much smaller than the initial input data. Shen et al. developed a deep learning model through self-supervised feature representation learning to characterize immune infiltration from transcriptome (116). The developed model was used to distill expression signatures of the transcriptome in brain tumor

TABLE 5 ML algorithms and their applications in TAMs.

Authors and Years	Cancer Types	Sample Size	ML Algorithms	Research Purposes	ML Applications
Chang et al. (102) (2021)	Ovarian cancer	1566	Cox, LASSO	To construct macrophage related prognostic model for ovarian cancer	identify multiple features related to survival (uni-and multi-variate Cox) and construct the macrophage-related prognostic model (LASSO)
Rostam et al. (103) (2017)	/		Orange Data Mining Toolbox	To identify different macrophage functional phenotypes	auto-identification of phenotypes based on cell size and morphology (Orange)
Zhu et al. (104) (2019)	Rectal cancer	46	SVM	To investigate the role of tumor-infiltrating leukocyte cell composition in the prognosis of radiotherapy for rectal cancer	classify responsive and non-responsive patients (SVM)
Zhang et al (105) (2021)	Glioma	2405	NN, SVM ER, PCA	To investigate the predictive value of monocytes in the immune microenvironment and prognosis in glioma patients	validate clustering results (NN, SVM) and calculate the risk scores of patients (ER, PCA)
Zhang et al. (106) (2021)	Glioma	2365	Pamr, NN, SVM ER, PCA	To build a prognostic model based on the molecular feature of TAMs for gliomas	validate the clustering results (Pamr, SVM, and NN), construct risk scores (ER, PCA) and further validate the clustering results (SVM, NN)
Zhang et al. (107) (2020)	Prostate cancer	487	LASSO, PCA	To build a model to predict the risk of prostate cancer based on immune-related gene-based novel subtypes	determine the properties of the subtypes (PCA) and build the risk predictive model (LASSO)
Yin et al. (108) (2022)	Cervical squamous cell carcinoma	78	Cox, LASSO, LR, GMM	To investigate the roles of TAMs in the development of cervical squamous cell carcinoma	select immune-related genes (Univariate Cox and LASSO), construct the risk score model (multi-variate Cox), build a diagnosis signature (LR), and then select the best models (GMM)
Yan et al. (109) (2020)	Ovarian cancer	365	Cox, LASSO, SVM, SVM-RFE	To explore prognostic genes associated with immune infiltration in ovarian cancer	identify the most valuable genes related to immune infiltration (LASSO, Cox), distinguish two different standards of immune infiltration (SVM), and work out the most valuable variables of immune infiltration (SVM-RFE)
Wu et al. (110) (2022)	Non-small cell lung cancer	681	RF	To develop a macrophages-based immune-related risk score model for relapse prediction in stage I–III non-small cell lung cancer	screen the robust prognostic markers and construct risk score to predict disease-free survival (RF)
Wei et al. (111) (2020)	Gastric cancer	407	SVM, LASSO, SVM-RFE	To investigate the effect of various components in gastric cancer TME and identify mechanisms exhibiting potential therapeutic targets	minimize the redundancy of features (LASSO) and rank the features (SVM, SVM-RFE)
Wang et al. (112) (2021)	Lung cancer	507	Mask R-CNN	To develop a prognostic model for the prediction of high- and low- risk lung adenocarcinoma	segment the nuclei of tumor, stroma, lymphocyte, macrophage, karyorrhexis and red blood cells (Mask R-CNN)
Vayrynen et al. (113) (2020)	Colorectal cancer	931	inform	To investigate the prognostic role of macrophage polarization in the colorectal cancer microenvironment	identify macrophages in tumor intraepithelial and stromal regions (inForm)
Ugai et al. (114) (2021)	Colorectal cancer	3092	inform	To investigate if the relationship between smoking and colorectal cancer incidence varies depending on macrophage infiltration	perform tissue category segmentation, cell segmentation, and cell type classification
Starosolski et al. (115) (2020)	Transgenic mouse models of neuroblastoma	16	Non-parametric neighborhood component analysis	To investigate if nanoradiomics can differentiate tumors based on TAM burden	radiomic features selection (the non-parametric neighborhood component method)
Shen et al. (116) (2021)	Brian tumor	3810	A self-developed deep learning algorithm based on contrastive learning	To stratify brain tumors for better clinical decision-making and prognosis prediction	distill expression signatures of transcriptome (DL)

(Continued)

TABLE 5 Continued

Authors and Years	Cancer Types	Sample Size	ML Algorithms	Research Purposes	ML Applications
Nakamura et al. (117) (2019)	Ovarian carcinoma	1656	SVM, RF, NN, LDA	To identify relationships between the expression of immune and inflammatory mediators and patient outcomes	classify ovarian cancer and normal tissue (SVM, RF, and NN) and map high-dimensional input data into a two-dimensional space (LDA)
Liang et al. (118) (2021)	Various cancers	9881	CART, LR, LDA, K-Neighbors Classifier, Gaussian Naive Bayes, SVM	To investigate the inflammasome signaling status to clarify its clinical and therapeutic significance	classify samples and validate gene set enrichment (all 6 ML methods)
Li et al. (119) (2021)	Bone-related malignancies	1675	RF	To investigate if a distinct immune infiltrative microenvironment exists in malignant bone-associated tumors and build a model for tumor diagnosis and prognosis	develop a bone-related tumor differential diagnosis model (RF)
Li et al. (120) (2022)	Gliomas	652	NN, LSTM, Cox, LASSO, RF	To predict survival and tumor-infiltrating macrophages in gliomas using MRI radiomics	extract significant radiomic features to construct a prediction model (NN, LSTM, Cox, LASSO, RF)
Kuang et al. (121) (2021)	Hodgkin lymphoma	130	LASSO, Cox, RF	To investigate potential markers for the diagnosis and prediction of classic Hodgkin lymphoma prognosis	identify prognostic genes and build a model for prognosis (LASSO, Cox, RF)
Hagos et al. (122) (2022)	Follicular lymphoma	32	ConCORDe-Net	To identify cell phenotypes and spatial distribution of immune cell subsets in the inter-follicular area of follicular lymphoma TME	detect different immune cells within and outside neoplastic follicles (ConCORDe-Net)
Guo et al. (123) (2021)	Pulmonary sarcomatoid carcinoma	97	Cox, RF	To build an immune-based risk-stratification system for prognosis in pulmonary sarcomatoid carcinoma	construct a predictive model and rank the predictive ability of each variable (Cox, RF)
Lange et al. (124) (2018)	Uveal melanoma	64	HCA, PCA	To study the immune environment and explore whether absolute T-cell quantification and expression profiles can dissect disparate immune components	reveal cell-specific expression patterns in gene selection (HCA, PCA)
Lin et al. (125) (2022)	Adamantinomatous craniopharyngioma (ACP)	57	RF, LASSO	To study the molecular immune mechanism in ACP and find potential biomarkers for the targeted therapy for ACP	screen diagnostic markers (RF, LASSO)

Cox, Cox Proportional-hazards Regression; LSSO, Least Absolute Shrinkage and Selection Operator; PCA, Principal Component Analysis; ER, Elastic Regression; SVM, Support Vector Machine; Pamr, Prediction Analysis for Microarrays; LR, Logistic Regression; GMM, Gaussian Mixture Model; LDA, Linear Discriminant Analysis; SVM-RFE, Support Vector Machine Recursive Feature Elimination; NN, Neural Network; RF, Random Forest; CART, Classification and Regression Trees; ConCORDe-Net, Cell Count Regularized Convolutional Neural Networks; HCA, Hierarchical Cluster Analysis; LSTM, Long short-term memory; MLP, Multi-layer perceptron; Weka, Waikato Environment for Knowledge Analysis; ROF, Rudin-Osher-Fatemi.

samples. The application of PCA in TAMs research could potentially be promising in enhancing predictive accuracy when input variables and their inter-connections are remarkably complicated.

5.2 Regression

Regression analysis is a method to mathematically describe the relationships between the outcome of interest (e.g., patient survival or relapse risk) and one or more features, also termed as variables (Figure 2B) (128). It answers the questions: Which variable is the most significant? What's the connection among these variables? And, perhaps most importantly, how certain are we about all of these variables? Regression analysis has been applied to cancer research for decades, from survival analysis with Cox's proportional hazard regression to Least Absolute Shrinkage Selection Operator Regression (LASSO) regression for significant feature selection.

Linear regression is the most common and simplest model for discovering how one or more explanatory variables determine the dependent variable (129). Logistic regression is extended by a linear regression model for classification problems. However, it differs from linear regression by being employed when the outcome variable is binary. Yin et al. built a diagnosis signature by logistic regression based on selected significant factors correlated with TAMs. They found that these factors were conducive to distinguish normal tissues from tumor (108). Cox proportional hazard is generally used when the outcome is the time to an occurrence (for example,

time to death, time to relapse). The results of Cox are explained in terms of a hazard ratio, indicating the risk of an event at a given time. Ridge regression and LASSO regression are variants of linear regression (linear regression appended with a regularization term) introduced for more accurate prediction. Ridge and LASSO are commonly used to reduce model complexity and prevent potential over-fitting. Typically, LASSO and Cox are combined together for disease prognosis. These studies generally use univariate Cox regression and LASSO regression to identify the significant characteristics and multivariate Cox regression to build risk score models (108).

Another variant of linear regression is elastic net regression. It integrates the LASSO and ridge regression methods by learning from their drawbacks to improve the regularization of statistical models. Thus, it achieves a more stable and better prediction than LASSO and ridge regression in less training samples. In two studies that intended to develop a prognostic model based on the molecular feature of TAMs, they both used elastic net to construct risk scores (105, 106). Especially, in Zhang et al's study, they found that glioma with higher risk scores is populated by macrophages comprising both the traditional M1 and M2 phenotypes, which further indicates that M0/M1/M2 is a continuum rather than two extremes (106).

5.3 Classification

5.3.1 Support vector machine

Support Vector Machine (SVM) is a powerful method that can be used for both regression and classification tasks (130). However, it mostly works as a classifier and aims to create a decision boundary, also termed as hyperplane, between two classes that distinctly classifies the data into different categories (131). The objective of SVM is to maximize the margin to select the best hyperplane, which offers some reinforcement so that subsequent data points can be classified with greater confidence. The margin is determined by a series of hyperplanes parallel to the decision boundary whose distance to the nearest data point is the largest in either the positive or negative class, as depicted in Figure 2C.

As a classifier, SVM is frequently used in TAMs. Patients can be classified into different groups based on the significant tumor-infiltrating immune cell proportions. For instance, patients with rectal cancer can be classified into responsive and non-responsive groups through the ML method based on the tumor-infiltrating immune cell composition and achieved an accuracy of 65% (104). Nakamura et al. applied SVM to discriminate between malignant and non-malignant tissues in ovarian cancer patients and malignant ovary samples through the immune signatures including M1 macrophage mediator signatures (117). Yan et al. used SVM to explore prognostic genes associated with immune infiltration and the classification accuracy reached as high as 0.934. Also of note, the high and low-risk groups exhibited significantly different proportions of TAMs (104). Some researchers used SVM to further validate the clustering results (105, 106). In an article by Liang, the authors applied six ML algorithms to predict inflammasome clusters, in which macrophages were the major immune cell population enriched in inflammasome complex^{Mid} and inflammasome complex^{High} clusters. In this paper, SVM achieved a highest prediction accuracy of 96% (118). Some researchers also use SVM-RFE, a feature selection algorithm that ranks the features according to the recursive feature deletion sequence, to identify prognostic genes associated with TAMs infiltration (109, 111).

The strength of SVM is that it can be used for complex data sets with many variables or dimensions. However, when it comes to high dimensions, SVM achieves a powerful model at the cost of easy interpretation of which features are influencing the model.

5.3.2 Random forest

Random forest (RF) is an ensemble decision tree classifier combining multiple tree predictors introduced by Leo Breiman (132). As an ML algorithm near the top of the classifier hierarchy, the RF classifier is capable of ranking the predictive ability of each variable and constructing a predictive model (110). Generally, RF is based on the aggregation of a large number of uncorrelated and weak decision trees, and each uncorrelated tree casts an individual prediction. The final decision is made by majority voting of all trees, which outperform any single classifier (Figure 2D). RF models are considered less vulnerable to overfit the training data set given the large number of trees built, making each tree an independent model. Given a large number of trees ensembled and each tree indicating an independent model, random forest models are thought to be less susceptible to overfitting. The ability of RF to precisely classify observations is extremely valuable in oncology applications, such as predicting patient death or tumor relapse. So far, RF has been applied to many TAMs studies for classification. They are generally used to screen TAMs-related markers and construct an immune-related risk score for risk prediction (110, 121, 123, 125). By utilizing RFs, a diagnostic model based on immune infiltration can accurately perform the differential diagnosis of bone-related malignancies (119). Nakamura et al. used RF to investigate whether genes identified by literature search or other analysis can distinguish between normal tissues and cancer tissues (117). In many studies, RFs also worked with other algorithms to screen the overlapping markers, e.g., LASSO (121, 125).

Overall, the advantage of RF is that it is an ensemble algorithm which has more accuracy than any individual prediction, especially when multi-modality variables are combined (133). However, the high dimension of all the features in cancer research and their complex interactions make it very difficult for humans to interpret the model and results.

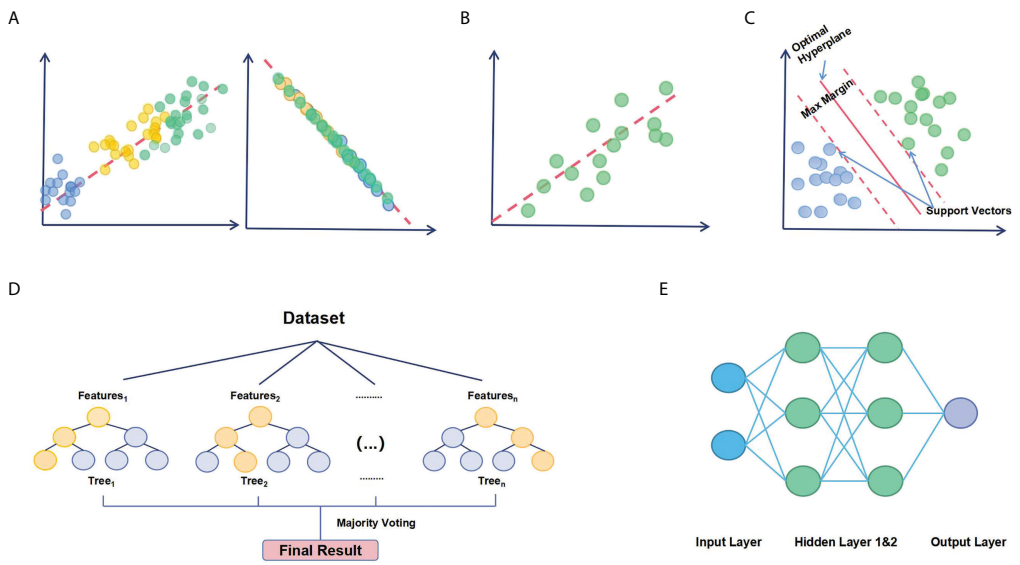


FIGURE 2
Basic principles of standard ML algorithms. **(A)** PCA reduces the dimensionality of a data set consisting of plenty of interrelated variables. **(A)** illustrates a series of data points viewed from another angle with approximately the same value on that dimension. It shows that the distinction between the data points can be represented by a principal component. **(B)** Regression analysis determines the relationship between factors and disease outcomes or identifies relevant prognostic factors for diseases. **(B)** illustrates regression estimating a mathematical formula that relates input variables to the output variable. **(C)** SVM generates a hyperplane in higher-dimensional feature space and maximizes the margin of error to select the best hyperplane. The best hyperplane would serve as a decision boundary for classification. **(D)** RF model ensembles a large number of small decision trees. Each tree is capable of making an individual prediction. **(E)** Neural networks tend to resemble the connections of neurons and synapses in human brain. The input data is assigned initial weights and transferred to output layers for classification. Hidden layers would tune the initial wrights to minimize the neural network's prediction error.

5.4 Neural networks and deep learning

Deep learning (DL) is a notable sub-class of ML which has a remarkable ability to learn patterns from raw, unstructured input data by incorporating artificial neural networks (ANN) (134). ANN is inspired by the structure and function of the brain. It attempts to use multiple layers of calculation units to imitate how the human brain processes input information. It is essentially a mathematical model consisting of an input layer, multiple hidden layers, and an output layer, as shown in Figure 2E. Each layer has multiple artificial neurons, also known as nodes in neural network. The nodes in input layers gather source material such as image pixels and numerical data. Hidden layers in the middle connect nodes to the next layer, creating non-linear representations between source data and the output layer (135).

Despite deriving from ANN, the DL framework differs from a straightforward neural network. Overall, DL networks are larger and consist of more layers and nodes, making it possible to reflect complicated interrelationships precisely. DL is able to process plenty of features across a large number of samples and derive neural network-based 'representations' quickly. Many specialized DL models have outperformed traditional ML models for various tasks, such as medical image segmentation and image-based tumor staging. Classical DL algorithms include Convolutional Neural Network (136), Recurrent Neural Networks (137), Radial Basis Function Networks (138), Long Short-Term Memory Networks (LSTMs) (139), Self-Organizing Maps (140), Autoencoders (141), etc., which have been proved to achieve state-of-the-art performance in specific applications (142–144).

Applications of neural networks and DL in TAMs focus more on classification and medical image segmentation. Li et al. developed an MRI radiomics approach to predict survival and tumor-infiltrating macrophages in gliomas (120). They used two neural network models and one long short-term memory DL model to divide patients into long and short-term survival clusters. In research conducted by Wang et al. (112), Mask R-CNN, a DL-based model, was applied to segment the nuclei of the tumor, lymphocyte, stroma, karyorrhexis, red blood cells and macrophage from pathology images. In addition to the existing segmentation algorithms, some studies developed their own DL segmentation models to characterize immune infiltration. Risom et al. segmented cell nuclei using Msmer, a DL-based algorithm developed in their lab (145), and Hagos et al. used ConCORDe-Net to detect cells in multiplex immunohistochemistry images (122). Meanwhile, commercial and Open-source software could also be used for segmentation in cancer research. For example, inForm

software package (Akoya Biosciences) has been applied in some studies to automatically perform tissue category segmentation, cell segmentation, and cell type classification (113, 114). InForm software is a powerful software that enables per-cell analysis of immunohistochemistry and immunofluorescence. It allows the separation and measurement of weak and spectrally overlapping markers and automatic detection and segmentation of specific tissues. Orange Data Mining Toolbox is another open-source software. Rostam et al. used it to automatically identify different macrophage functional phenotypes based on cell size and morphology (103).

Interest in DL models has grown in recent decades owing to rapid advances in high-performance computing infrastructure, such as cloud and GPU computing (146). However, it is still far from meeting the vast amounts of data needed for medical research. Developing deep neural networks and then training is time-consuming and computationally expensive compared with traditional ML methods.

6. Challenges

Despite such exciting research, various limitations or requirements must be addressed before ML can realize its full potential in the studies focusing on TAMs. As most ML models are data-driven, the most critical challenge is the requirement of tremendous and valuable data sets (147). Generally, data related to TAMs can be incredibly complex, with thousands of variables capturing different facets of the TME system. However, these data sets are still too small for ML modeling, especially for unsupervised learning. The lack of sample size might lead to poor model performance or overfitting. Deep neural networks are especially vulnerable to overfitting because they have thousands to millions of parameters.

Moreover, data quality and completeness are also challenging in the studies of tumor prognosis, in which patient follow-up might be irregularly collected or lost, and different institutions may use various standards of testing. In response to the challenge of massive clinical data acquisition, some cloud-based cancer repositories such as Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) have been created to enable cross-institution data sharing and data quality assurance. We hope with the emergence of more open-source data sets and data standardization, these restrictions will be less of an issue in the future.

Clinical translation is also a challenge for ML. Many trials are still in the stage of by-proof-test. Research groups and companies are facing the challenges of making their products more reliable and practical in large-scale implementations or even real usage scenarios. Similarly, many innovative solutions, generated from the frontiers of ML research and shown to be theoretically powerful, have yet to integrate into day-to-day clinical use. In modeling, most models take fixed training and testing data set, which is impractical in real clinical practice. Considering the rapid changes in tumor data, continuous updating and reevaluation are required to monitor the model performance and guarantee model consistency. In addition, most of the current ML-based tumor models are single-center studies. There are considerably fewer external validation studies of TAMs in the published papers. Future studies should involve external or cross-institution validation to ensure the test set is diversified enough with different clinic scenarios involved. We believe the robust external validation and improvements in interpretability and generalizability may boost clinician confidence in ML and facilitate further incorporation of ML models into clinical practice.

Furthermore, after reviewing papers combining ML with TAMs, we come to realize that the complexity and heterogeneity of TAMs in TME are far from being fully elucidated. As discussed above, the dichotomy of TAMs is too simple to clarify macrophage activation states *in vivo*. What should be noted is that M0/M1/M2 is a continuum *in vivo* instead of well-delineated categories. TAMs are characterized by its remarkable plasticity. The phenotypes can switch between the two extremes, while most existing studies still regard TAMs as two distinct extremes. Besides, subtypes of M2-TAMs can be further identified and classified as M2a, M2b, M2c (148, 149), and M2d in TME. Identifying complexity and heterogeneity of TAMs *in vivo* and the subtypes of M2 macrophages more precisely to reduce side effects of cancer therapy using ML methods can be challenging but promising. Therapies addressing the recruitment, depletion and repolarization of M2 are promising strategies for tumor treatment. With the help of ML, many studies are enabled to identify specific molecules involved in polarization of M0 macrophages towards M1/M2 macrophages and TAMs recruitment. However, the key biomarkers in depletion and repolarization of M2 based on ML have not received a lot of attention. By integrating more medical images and omics data, it is anticipated that ML will have broader prospects on exploring, validating and implementing critical genes in the repolarization of TAMs to further facilitate precision oncology.

7. Future directions

ML in cancer research is still in the early stage of exploration. More investigations and efforts are required to break through current limitations. In terms of reducing the need for a large data set, Generative Adversarial Networks (GAN) are receiving attention. GAN has two neural networks, which are generative and discriminator networks. They contest with each other in a zero-sum game and generate new and synthetic instances of data that can 'fool' the discriminator network.

Precision medicine is the future direction of cancer therapy, in which case patients can get optimized management and treatment to improve survival. An important part of precision oncology involves understanding cancer genomics, radiomics and the complex heterogeneity of TME. With the help of ML, scientists are able to

disentangle more cancer characteristics, enabling precision oncology. One of the popular and evolutionary directions in ML is reinforcement learning. It learns to achieve goals in an uncertain and complex environment. Due to the non-stationary tumor environment with changing conditions and stimuli, reinforcement learning has the potential to offer computer-guided decision support for personalized treatment. Currently, its applications in medicine are mainly focus on medical image analysis, disease screening and personalized treatment recommendations. In the future, we envision that it could be employed for dynamic cancer treatment regimens after personalized tumor prognosis, tailoring the treatment for each individual.

Overall, the combination of ML and TAMs is relatively young and far from fulfilling its potential in cancer research. The distinctive nature of cancer studies makes accuracy and interpretability extremely crucial. We still have a long way to go to uncover and harness the intricacies of ML and the complexities of TME. Hopefully, with ever-evolving algorithms, more potent supercomputers, and substantial investment being involved in this field, these applications will be more intelligent, cost-effective, and time-efficient. In the future, ML is expected to play a more critical role in TAMs analysis and precision oncology.

Author contributions

ZL and QY wrote the manuscript; QZ and XY reviewed and edited the paper; ZBL and JF offered technological guidance. All authors reviewed the results and approved the final version of the manuscript. ZL and QY are the co-first authors and JF is the corresponding author.

Funding

This study was supported by the Science and Technology Project of Shanghai Municipal Science and Technology Commission (No.22Y31900500).

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

1. König IR, Fuchs O, Hansen G., Mutius von E, M, Kopp V, et al. What is precision medicine? *Eur Respir J* (2017) 50(4). doi: 10.1183/13993003.00391-2017

2. Issa NT, Stathias V, Schürer S, Dakshanamurthy S. Machine and deep learning approaches for cancer drug repurposing. *Semin Cancer Biol* (2021) 68:132–42. doi: 10.1016/j.semcancer.2019.12.011

3. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* (2019) 20(5):e262–73. doi: 10.1016/S1470-2045(19)30149-4

4. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* (2019) 18(6):463–77. doi: 10.1038/s41573-019-0024-5

5. Lynch CM, Abdollahi B, Fuqua JD, Carlo AR, Bartholomai JA, Balgemann RN, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform* (2017) 108:1–8. doi: 10.1016/j.ijmedinf.2017.09.013

6. Pathria P, Louis TL, Varner JA. Targeting tumor-associated macrophages in cancer. *Trends Immunol* (2019) 40(4):310–27.

7. Wynn TA, Chawla A, Pollard JW. Macrophage biology in development, homeostasis and disease. *Nature* (2013) 496(7446):445–55.

8. Krenkel O, Tacke F. Liver macrophages in tissue homeostasis and disease. *Nat Rev Immunol* (2017) 17(5):306–21.

9. Shapouri-Moghaddam A, Mohammadian S, Vazini H, Taghadosi MS, Esmaeili A, Mardani F, et al. Macrophage plasticity, polarization, and function in health and disease. *J Cell Physiol* (2018) 233(9):6425–40.

10. Jaguin M, Houlbert N, Fardel O, Lecureur V, et al. Polarization profiles of human m-CSF-generated macrophages and comparison of M1-markers in classically activated macrophages from GM-CSF and m-CSF origin. *Cell Immunol* (2013) 281(1):51–61.

11. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Macrophage polarization: Different gene signatures in M1(LPS+) vs. classically and M2(LPS-) vs. alternatively activated macrophages. *Front Immunol* (2019) 10:1084. doi: 10.3389/fimmu.2019.01084

12. Trombetta AC, Soldano S, Contini P, Tomatis V, Ruaro B, Paolino S, et al. A circulating cell population showing both M1 and M2 monocyte/macrophage surface markers characterizes systemic sclerosis patients with lung involvement. *Respir Res* (2018) 19(1):186. doi: 10.1186/s12931-018-0891-z

13. Feng D, Huang WY, Niu XL, Hao S, Zhang LN, Hu YJ, et al. Significance of macrophage subtypes in the peripheral blood of children with systemic juvenile idiopathic arthritis. *Rheumatol Ther* (2021) 8(4):1859–70. doi: 10.1007/s40744-021-00385-x

14. Cunha C, Gomes C, Vaz AR, Brites D. Exploring new inflammatory biomarkers and pathways during LPS-induced M1 polarization. *Med Inflamm* (2016) 2016:6986175. doi: 10.1155/2016/6986175

15. Yamaguchi T, Fushida S, Yamamoto Y, Tsukada T, Kinoshita J, Oyama K, et al. Tumor-associated macrophages of the M2 phenotype contribute to progression in gastric cancer with peritoneal dissemination. *Gastric Cancer* (2016) 19(4):1052–65. doi: 10.1007/s10120-015-0579-8

16. Nakagawa M, Karim MR, Izawa T, Kuwamura M, Yamate J. Immunophenotypical characterization of M1/M2 macrophages and lymphocytes

in cisplatin-induced rat progressive renal fibrosis. *Cells* (2021) 10(2). doi: 10.3390/cells10020257

17. Su MJ, Aldawsari H, Amiji M. Pancreatic cancer cell exosome-mediated macrophage reprogramming and the role of MicroRNAs 155 and 125b2 transfection using nanoparticle delivery systems. *Sci Rep* (2016) 6:30110. doi: 10.1038/srep30110

18. Qiu Y, Xu J, Yang L, Zhao G, Ding J, Chen Q, et al. MiR-375 silencing attenuates pro-inflammatory macrophage response and foam cell formation by targeting KLF4. *Exp Cell Res* (2021) 400(1):112507. doi: 10.1016/j.yexcr.2021.112507

19. Jang JY, Lee JK, Jeon YK, Kim CW. Exosome derived from epigallocatechin gallate treated breast cancer cells suppresses tumor growth by inhibiting tumor-associated macrophage infiltration and M2 polarization. *BMC Cancer* (2013) 13:421. doi: 10.1186/1471-2407-13-421

20. Weng YS, Tseng HY, Chen YA, Shen PC, Haq Al AT, Chen LM, et al. MCT-1/miR-34a/IL-6/IL-6R signaling axis promotes EMT progression, cancer stemness and M2 macrophage polarization in triple-negative breast cancer. *Mol Cancer* (2019) 18(1):42. doi: 10.1186/s12943-019-0988-0

21. Tong F, Mao X, Zhang S, Xie H, Yan B, Wang B, et al. HPV + HNSCC-derived exosomal miR-9 induces macrophage M1 polarization and increases tumor radiosensitivity. *Cancer Lett* (2020) 478:34–44. doi: 10.1016/j.canlet.2020.02.037

22. Wang X, Luo G, Zhang K, Cao J, Huang C, Jiang T, et al. Hypoxic tumor-derived exosomal miR-301a mediates M2 macrophage polarization via PTEN/PI3Kγ to promote pancreatic cancer metastasis. *Cancer Res* (2018) 78(16):4586–98. doi: 10.1158/0008-5472.CAN-17-3841

23. Luo YY, Yang ZQ, Lin XF, Zhao FL, Tu HT, Wang LJ, et al. Knockdown of lncRNA PVT1 attenuated macrophage M1 polarization and relieved sepsis induced myocardial injury via miR-29a/HMGB1 axis. *Cytokine* (2021) 143:155509. doi: 10.1016/j.cyto.2021.155509

24. Zhao S, Mi Y, Guan B, Zheng B, Wei P, Gu Y, et al. Tumor-derived exosomal miR-934 induces macrophage M2 polarization to promote liver metastasis of colorectal cancer. *J Hematol Oncol* (2020) 13(1):156. doi: 10.1186/s13045-020-00991-2

25. Jiang M, Dai J, Yin M, Jiang C, Ren M, Tian L, et al. LncRNA MEG8 sponging miR-181a-5p contributes to M1 macrophage polarization by regulating SHP2 expression in henoch-schonlein purpura rats. *Ann Med* (2021) 53(1):1576–88. doi: 10.1080/07853890.2021.1969033

26. Chen X, Ying X, Wang X, Wu X, Zhu Q, Wang X, et al. Exosomes derived from hypoxic epithelial ovarian cancer deliver microRNA-940 to induce macrophage M2 polarization. *Oncol Rep* (2017) 38(1):522–8. doi: 10.3892/or.2017.5697

27. Chi X, Ding B, Zhang L, Zhang J, Wang J, Zhang W, et al. lncRNA GAS5 promotes M1 macrophage polarization via miR-455-5p/SOCS3 pathway in childhood pneumonia. *J Cell Physiol* (2019) 234(8):13242–51. doi: 10.1002/jcp.27996

28. Rong J, Xu L, Hu Y, Liu F, Yu Y, Guo H, et al. Inhibition of let-7b-5p contributes to an anti-tumorigenic macrophage phenotype through the SOCS1/STAT pathway in prostate cancer. *Cancer Cell Int* (2020) 20:470. doi: 10.1186/s12935-020-01563-7

29. Pasca S, Jurj A, Petrushev B, Tomuleasa C, Matei D. MicroRNA-155 implication in M1 polarization and the impact in inflammatory diseases. *Front Immunol* (2020) 11:625. doi: 10.3389/fimmu.2020.00625

30. Ge X, Tang P, Rong Y, Jiang D, Lu X, Ji C, et al. Exosomal miR-155 from M1-polarized macrophages promotes EndoMT and impairs mitochondrial function via activating NF-κB signaling pathway in vascular endothelial cells after traumatic spinal cord injury. *Redox Biol* (2021) 41:101932. doi: 10.1016/j.redox.2021.101932

31. Banerjee S, Xie N, Cui H, Tan Z, Yang S, Icyuz M, et al. MicroRNA let-7c regulates macrophage polarization. *J Immunol* (2013) 190(12):6542–9. doi: 10.4049/jimmunol.1202496

32. Xu S, Wei J, Wang F, Kong LY, Ling XY, Nduom E, et al. Effect of miR-142-3p on the M2 macrophage and therapeutic efficacy against murine glioblastoma. *J Natl Cancer Inst* (2014) 106(8). doi: 10.1093/jnci/dju162

33. Baer C, Squadrito ML, Laoui D, Thompson D, Hansen SK, Kiialainen A, et al. Suppression of microRNA activity amplifies IFN-γ-induced macrophage activation and promotes anti-tumour immunity. *Nat Cell Biol* (2016) 18(7):790–802. doi: 10.1038/ncb3371

34. Chen J, Zhang K, Zhi Y, Wu Y, Chen B, Bai J, et al. Tumor-derived exosomal miR-19b-3p facilitates M2 macrophage polarization and exosomal LINC00273 secretion to promote lung adenocarcinoma metastasis via hippo pathway. *Clin Transl Med* (2021) 11(9):e478. doi: 10.1002/ctm2.478

35. Cao J, Dong R, Jiang L, Gong Y, Yuan M, You J, et al. LncRNA-MM2P identified as a modulator of macrophage M2 polarization. *Cancer Immunol Res* (2019) 7(2):292–305. doi: 10.1158/2326-6066.CIR-18-0145

36. Mills CD. M1 and M2 macrophages: Oracles of health and disease. *Crit Rev Immunol* (2012) 32(6):463–88. doi: 10.1615/CritRevImmunol.v32.i6.10

37. De Santa F, Vitiello L, Torcinaro A, Ferraro E. The role of metabolic remodeling in macrophage polarization and its effect on skeletal muscle regeneration. *Antioxid Redox Signal* (2019) 30(12):1553–98. doi: 10.1089/ars.2017.7420

38. Wang J, Li R, Peng Z, Hu B, Rao X, Li J, et al. HMGB1 participates in LPS–induced acute lung injury by activating the AIM2 inflammasome in macrophages and inducing polarization of M1 macrophages via TLR2, TLR4, and RAGE/NF–κB signaling pathways. *Int J Mol Med* (2020) 45(1):61–80.

39. Loeuillard E, Yang J, Buckarma E, Wang J, Liu Y, Conboy C, et al. Targeting tumor-associated macrophages and granulocytic myeloid-derived suppressor cells augments PD-1 blockade in cholangiocarcinoma. *J Clin Invest* (2020) 130(10):5380–96. doi: 10.1172/JCI137110

40. Gordon SR, Maute RL, Dulken BW, Hutter G, George BM, McCracken MN, et al. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature* (2017) 545(7655):495–9. doi: 10.1038/nature22396

41. Chen J, Lin Z, Liu L, Zhang R, Geng Y, Fan M, et al. GOLM1 exacerbates CD8(+) T cell suppression in hepatocellular carcinoma by promoting exosomal PD-L1 transport into tumor-associated macrophages. *Signal Transduct Target Ther* (2021) 6(1):397. doi: 10.1038/s41392-021-00784-0

42. Arlauckas SP, Garren SB, Garris CS, Kohler RH, Oh J, Pittet MJ, et al. Arg1 expression defines immunosuppressive subsets of tumor-associated macrophages. *Theranostics* (2018) 8(21):5842–54. doi: 10.7150/thno.26888

43. Ren J, Han X, Lohner H, Liang R, Liang S, Wang H, et al. Serum- and glucocorticoid-inducible kinase 1 promotes alternative macrophage polarization and restrains inflammation through FoxO1 and STAT3 signaling. *J Immunol* (2021) 207(1):268–80. doi: 10.4049/jimmunol.2001455

44. Mohapatra S, Pioppini C, Ozpolat B, Calin GA. Non-coding RNAs regulation of macrophage polarization in cancer. *Mol Cancer* (2021) 20(1):24. doi: 10.1186/s12943-021-01313-x

45. Zhou J, Tang Z, Gao S, Li C, Feng Y, Zhou X, et al. Tumor-associated macrophages: Recent insights and therapies. *Front Oncol* (2020) 10:188. doi: 10.3389/fonc.2020.00188

46. Leibovich SJ, Ross R. The role of the macrophage in wound repair. a study with hydrocortisone and antimacrophage serum. *Am J Pathol* (1975) 78(1):71–100.

47. Polverini PJ, Cotran PS, Gimbrone MAJr, Unanue ER. Activated macrophages induce vascular proliferation. *Nature* (1977) 269(5631):804–6. doi: 10.1038/269804a0

48. Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. *Nat Rev Cancer* (2016) 16(7):447–62. doi: 10.1038/nrc.2016.54

49. Guillot A, Tacke F. Liver macrophages: Old dogmas and new insights. *Hepatol Commun* (2019) 3(6):730–43. doi: 10.1002/hep4.1356

50. Wu K, Lin K, Li X, Yuan X, Xu P, Ni P, et al. Redefining tumor-associated macrophage subpopulations and functions in the tumor microenvironment. *Front Immunol* (2020) 11:1731. doi: 10.3389/fimmu.2020.01731

51. Helm O, Mennrich R, Petrick D, Goebel L, Freitag-Wolf S, Röder C, et al. Comparative characterization of stroma cells and ductal epithelium in chronic pancreatitis and pancreatic ductal adenocarcinoma. *PloS One* (2014) 9(5):e94357. doi: 10.1371/journal.pone.0094357

52. Wu MF, Lin CA, Yuan TH, Yeh HY, Su SF, Gu CL, et al. The M1/M2 spectrum and plasticity of malignant pleural effusion-macrophage in advanced lung cancer. *Cancer Immunol Immunother* (2021) 70(5):1435–50. doi: 10.1007/s00262-020-02781-8

53. Cai H, Zhang Y, Wang J, Gu J. Defects in macrophage reprogramming in cancer therapy: The negative impact of PD-L1/PD-1. *Front Immunol* (2021) 12:690869. doi: 10.3389/fimmu.2021.690869

54. Yin M, Li X, Tan S, Zhou HJ, Ji W, Bellone S, et al. Tumor-associated macrophages drive spheroid formation during early transcoelomic metastasis of ovarian cancer. *J Clin Invest* (2016) 126(11):4157–73. doi: 10.1172/JCI87252

55. Pan Y, Yu Y, Wang X, Zhang T. Tumor-associated macrophages in tumor immunity. *Front Immunol* (2020) 11:583084. doi: 10.3389/fimmu.2020.583084

56. Tanaka A, Sakaguchi S. Targeting treg cells in cancer immunotherapy. *Eur J Immunol* (2019) 49(8):1140–6. doi: 10.1002/eji.201847659

57. Noy R, Pollard JW. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* (2014) 41(1):49–61. doi: 10.1016/j.immuni.2014.06.010

58. Wang D, Yang L, Yue D, Cao L, Li L, Wang D, et al. Macrophage-derived CCL22 promotes an immunosuppressive tumor microenvironment via IL-8 in malignant pleural effusion. *Cancer Lett* (2019) 452:244–53. doi: 10.1016/j.canlet.2019.03.040

59. Allavena P, Chieppa M, Bianchi G, Solinas G, Fabbri M, Laskarin G, et al. Engagement of the mannose receptor by tumoral mucins activates an immune

suppressive phenotype in human tumor-associated macrophages. *Clin Dev Immunol* (2010) 2010:547179. doi: 10.1155/2010/547179

60. Thibault B, Castells M, Delord JP, Couderc B. Ovarian cancer microenvironment: implications for cancer dissemination and chemoresistance acquisition. *Cancer Metastasis Rev* (2014) 33(1):17–39. doi: 10.1007/s10555-013-9456-2

61. Chen YC, Lai YS, Hsuuw YD, Chang KT. Withholding of m-CSF supplement reprograms macrophages to M2-like via endogenous CSF-1 activation. *Int J Mol Sci* (2021) 22(7). doi: 10.3390/ijms22073532

62. Folkman J. What is the evidence that tumors are angiogenesis dependent? *J Natl Cancer Inst* (1990) 82(1):4–6. doi: 10.1093/jnci/82.1.4

63. Baeriswyl V, Christofori G. The angiogenic switch in carcinogenesis. *Semin Cancer Biol* (2009) 19(5):329–37. doi: 10.1016/j.semcancer.2009.05.003

64. Biswas SK, Sica A, Lewis CE. Plasticity of macrophage function during tumor progression: regulation by distinct molecular mechanisms. *J Immunol* (2008) 180(4):2011–7. doi: 10.4049/jimmunol.180.4.2011

65. White JR, Harris RA, Lee SR, Craigon MH, Binley K, Price T, et al. Genetic amplification of the transcriptional response to hypoxia as a novel means of identifying regulators of angiogenesis. *Genomics* (2004) 83(1):1–8. doi: 10.1016/S0888-7543(03)00215-5

66. Werno C, Menrad H, Weigert A, Dehne N, Goerdt S, Schledzewski K, et al. Knockout of HIF-1α in tumor-associated macrophages enhances M2 polarization and attenuates their pro-angiogenic responses. *Carcinogenesis* (2010) 31(10):1863–72. doi: 10.1093/carcin/bgq088

67. Li N, Li Y, Li Z, Huang C, Yang Y, Lang M, et al. Hypoxia inducible factor 1 (HIF-1) recruits macrophage to activate pancreatic stellate cells in pancreatic ductal adenocarcinoma. *Int J Mol Sci* (2016) 17(6). doi: 10.3390/ijms17060799

68. Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer. *Nat Rev Drug Discov* (2018) 17(12):887–904. doi: 10.1038/nrd.2018.169

69. Wang HW, Joyce JA. Alternative activation of tumor-associated macrophages by IL-4: priming for protumoral functions. *Cell Cycle* (2010) 9 (24):4824–35. doi: 10.4161/cc.9.24.14322

70. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* (2017) 17(5):302–17. doi: 10.1038/nrc.2017.6

71. Sleeman JP. The lymph node pre-metastatic niche. *J Mol Med (Berl)* (2015) 93(11):1173–84. doi: 10.1007/s00109-015-1351-6

72. Kaplan RN, Riba RD, Zacharoulis S, Bramley AH, Vincent L, Costa C, et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* (2005) 438(7069):820–7. doi: 10.1038/nature04186

73. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer* (2009) 9(4):239–52. doi: 10.1038/nrc2618

74. Sceneay J, Smyth MJ, Möller A. The pre-metastatic niche: finding common ground. *Cancer Metastasis Rev* (2013) 32(3-4):449–64. doi: 10.1007/s10555-013-9420-1

75. Lin Y, Xu J, Lan H. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *J Hematol Oncol* (2019) 12 (1):76. doi: 10.1186/s13045-019-0760-3

76. Werner L, Dreyer JH, Hartmann D, Barros MHM, Büttner-Herold M, Grittner U, et al. Tumor-associated macrophages in classical Hodgkin lymphoma: hormetic relationship to outcome. *Sci Rep* (2020) 10(1):9410. doi: 10.1038/s41598-020-66010-z

77. Hwang I, Kim JW, Ylaya K, Chung EJ, Kitano H, Perry C, et al. Tumor-associated macrophage, angiogenesis and lymphangiogenesis markers predict prognosis of non-small cell lung cancer patients. *J Transl Med* (2020) 18(1):443. doi: 10.1186/s12967-020-02618-z

78. Nie Y, Huang H, Guo M, Chen J, Wu W, Li W, et al. Breast phyllodes tumors recruit and repolarize tumor-associated macrophages via secreting CCL5 to promote malignant progression, which can be inhibited by CCR5 inhibition therapy. *Clin Cancer Res* (2019) 25(13):3873–86. doi: 10.1158/1078-0432.CCR-18-3421

79. Kleinschmidt J, Zucker CL, Yazulla S. Neurotoxic action of kainic acid in the isolated toad and goldfish retina: II. mechanism of action. *J Comp Neurol* (1986) 254(2):196–208. doi: 10.1002/cne.902540205

80. Xiang X, Wang J, Lu D, Xu X. Targeting tumor-associated macrophages to synergize tumor immunotherapy. *Signal Transduct Target Ther* (2021) 6(1):75. doi: 10.1038/s41392-021-00484-9

81. Li D, Ji H, Niu X, Yin L, Wang Y, Gu Y. Tumor-associated macrophages secrete CC-chemokine ligand 2 and induce tamoxifen resistance by activating PI3K/Akt/mTOR in breast cancer. *Cancer Sci* (2020) 111(1):47–58. doi: 10.1111/cas.14230

82. Jinushi M, Chiba S, Yoshiyama H, Masutomi K, Kinoshita I, Dosaka-Akita H, et al. Tumor-associated macrophages regulate tumorigenicity and anticancer drug responses of cancer stem/initiating cells. *Proc Natl Acad Sci USA* (2011) 108 (30):12425–30. doi: 10.1073/pnas.1106645108

83. Kong L, Zhou Y, Bu H, Lv T, Shi Y, Yang J, et al. Deletion of interleukin-6 in monocytes/macrophages suppresses the initiation of hepatocellular carcinoma in mice. *J Exp Clin Cancer Res* (2016) 35(1):131. doi: 10.1186/s13046-016-0412-1

84. Liu CQ, Xu J, Zhou ZG, Jin LL, Yu XJ, Xiao G, et al. Expression patterns of programmed death ligand 1 correlate with different microenvironments and patient prognosis in hepatocellular carcinoma. *Br J Cancer* (2018) 119(1):80–8. doi: 10.1038/s41416-018-0144-4

85. Kimm MA, Klenk C, Alunni-Fabbroni M, Kästle S, Stechele M, Ricke J, et al. Tumor-associated macrophages-implications for molecular oncology and imaging. *Biomedicines* (2021) 9(4). doi: 10.3390/biomedicines9040374

86. MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome* (2021) 64(4):416–25. doi: 10.1139/gen-2020-0131

87. Deo RC. Machine learning in medicine. *Circulation* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

88. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H, et al. eDoctor: machine learning and the future of medicine. *J Intern Med* (2018) 284 (6):603–19. doi: 10.1111/joim.12822

89. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP, et al. Machine learning in cardiovascular medicine: are we there yet? *Heart* (2018) 104 (14):1156–64. doi: 10.1136/heartjnl-2017-311198

90. Cammarota G, Ianiro G, Ahern A, Carbone C, Temko A, Claesson MJ, et al. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat Rev Gastroenterol Hepatol* (2020) 17(10):635–48. doi: 10.1038/s41575-020-0327-3

91. Berry MW, Mohamed A, Yap BW. *Supervised and unsupervised learning for data science*. Springer (2019).

92. Cunningham P, Cord M, Delany SJ. Supervised learning. In: *Machine learning techniques for multimedia*. Springer (2008). p. 21–49.

93. Zhu X, Goldberg AB. Introduction to semi-supervised learning. *Synthesis Lectures Artif Intell Mach Learn* (2009) 3(1):1–130. doi: 10.1007/978-3-031-01548-9

94. Yousef M, Allmer J. *miRNomics: microRNA biology and computational analysis*. Springer (2014).

95. Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia* (2019) 60(10):2037–47. doi: 10.1111/epi.16333

96. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* (2022) 23(1):40–55. doi: 10.1038/s41580-021-00407-0

97. Gupta S. *Pros and cons of various machine learning algorithms* (2020). Available at: https://towardsdatascience.com/pros-and-cons-of-various-classification-ml-algorithms-3b5bfb3c87d6.

98. (2019). Available at: https://www.i2tutorials.com/what-are-the-pros-and-cons-of-the-pca/.

99. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol* (2019) 29(7):R231–r236. doi: 10.1016/j.cub.2019.02.034

100. Sugahara S, Ueno M. Exact learning augmented naive bayes classifier. *Entropy (Basel)* (2021) 23(12). doi: 10.3390/e23121703

101. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med* (2011) 18 (10):1099–104. doi: 10.1111/j.1553-2712.2011.01185.x

102. Chang H, Zhu Y, Zheng J, Chen L, Lin J, Yao J, et al. Construction of a macrophage infiltration regulatory network and related prognostic model of high-grade serous ovarian cancer.
J Oncol (2021) 2021:1331031. doi: 10.1155/2021/1331031

103. Rostam HM, Reynolds PM, Alexander MR, Gadegaard N, Ghaemmaghami AM. Image based machine learning for identification of macrophage subsets. *Sci Rep* (2017) 7(1):3521. doi: 10.1038/s41598-017-03780-z

104. Zhu M, Li X, Ge Y, Nie J, Li X. (2019). The tumor infiltrating leukocyte cell composition are significant markers for prognostics of radiotherapy of rectal cancer as revealed by cell type deconvolution, in: *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* , IEEE.

105. Zhang N, Dai Z, Wu W, Wang Z, Cao H, Zhang Y, et al. The predictive value of monocytes in immune microenvironment and prognosis of glioma patients based on machine learning. *Front Immunol* (2021) 12:656541. doi: 10.3389/fimmu.2021.656541

106. Zhang H, Luo YB, Wu W, Zhang L, Wang Z, Dai Z, et al. The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Comput Struct Biotechnol J* (2021) 19:4603–18. doi: 10.1016/j.csbj.2021.08.019

107. Zhang E, He J, Zhang H, Shan L, Wu H, Zhang M, et al. Immune-related gene-based novel subtypes to establish a model predicting the risk of

prostate cancer. *Front Genet* (2020) 11:595657. doi: 10.3389/fgene.2020.595657

108. Yin R, Zhai X, Han H, Tong X, Li Y, Deng K, et al. Characterizing the landscape of cervical squamous cell carcinoma immune microenvironment by integrating the single-cell transcriptomics and RNA-seq. *Immun Inflammation Dis* (2022) 10(6):e608. doi: 10.1002/iid3.608

109. Yan S, Fang J, Chen Y, Xie Y, Zhang S, Zhu X, et al. Comprehensive analysis of prognostic gene signatures based on immune infiltration of ovarian cancer. *BMC Cancer* (2020) 20(1):1205. doi: 10.1186/s12885-020-07695-3

110. Wu XR, Peng HX, He M, Zhong R, Liu J, Wen YK, et al. Macrophages-based immune-related risk score model for relapse prediction in stage I-III non-small cell lung cancer assessed by multiplex immunofluorescence. *Transl Lung Cancer Res* (2022) 11(4):523–42. doi: 10.21037/tlcr-21-916

111. Wei S, Lu J, Lou J, Shi C, Mo S, Shao Y, et al. Gastric cancer tumor microenvironment characterization reveals stromal-related gene signatures associated with macrophage infiltration. *Front Genet* (2020) 11:663. doi: 10.3389/fgene.2020.00663

112. Wang S, Rong R, Yang DM, Fujimoto J, Yan S, Cai L, et al. Computational staining of pathology images to study the tumor microenvironment in lung cancer. *Cancer Res* (2020) 80(10):2056–66. doi: 10.1158/0008-5472.CAN-19-1629

113. Väyrynen JP, Haruki K, Lau MC, Väyrynen SA, Zhong R, Costa Dias A, et al. The prognostic role of macrophage polarization in the colorectal cancer microenvironment. *Cancer Immunol Res* (2021) 9(1):8–19. doi: 10.1158/2326-6066.CIR-20-0527

114. Ugai T, Väyrynen JP, Haruki K, Akimoto N, Lau MC, Zhong R, et al. Smoking and incidence of colorectal cancer subclassified by tumor-associated macrophage infiltrates. *J Natl Cancer Inst* (2022) 114(1):68–77. doi: 10.1093/jnci/djab142

115. Starosolski Z, Courtney AN, Srivastava M, Guo L, Stupin I, Metelitsa LS, et al. A nanoradiomics approach for differentiation of tumors based on tumor-associated macrophage burden. *Contrast Media Mol Imaging* (2021) 2021:6641384. doi: 10.1155/2021/6641384

116. Shen X, Wang X, Shen H, Feng M, Wu D, Yang Y, et al. Transcriptomic analysis identified two subtypes of brain tumor characterized by distinct immune infiltration and prognosis. *Front Oncol* (2021) 11:734407. doi: 10.3389/fonc.2021.734407

117. Nakamura M, Bax HJ, Scotto D, Souri EA, Sollie S, Harris RJ, et al. Immune mediator expression signatures are associated with improved outcome in ovarian carcinoma. *Oncoimmunology* (2019) 8(6):e1593811. doi: 10.1080/2162402X.2019.1593811

118. Liang Q, Wu J, Zhao X, Shen S, Zhu C, Liu T, et al. Establishment of tumor inflammasome clusters with distinct immunogenomic landscape aids immunotherapy. *Theranostics* (2021) 11(20):9884–903. doi: 10.7150/thno.63202

119. Li GQ, Wang YK, Zhou H, Jin LG, Wang CY, Albahde M, et al. Application of immune infiltration signature and machine learning model in the differential diagnosis and prognosis of bone-related malignancies. *Front Cell Dev Biol* (2021) 9:630355. doi: 10.3389/fcell.2021.630355

120. Li G, Li L, Li Y, Qian Z, Wu F, He Y, et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain* (2022) 145 (3):1151–61. doi: 10.1093/brain/awab340

121. Kuang Z, Tu J, Li X. Combined identification of novel markers for diagnosis and prognostic of classic Hodgkin lymphoma. *Int J Gen Med* (2021) 14:9951–63. doi: 10.2147/IJGM.S341557

122. Hagos YB, Akarca AU, Ramsay A, Rossi RL, Pomplun S, Ngai V, et al. High inter-follicular spatial co-localization of CD8+FOXP3+ with CD4+CD8+ cells predicts favorable outcome in follicular lymphoma. *Hematol Oncol* (2022). doi: 10.1002/hon.3003

123. Guo H, Li B, Diao L, Wang H, Chen P, Jiang M, et al. An immune-based risk-stratification system for predicting prognosis in pulmonary sarcomatoid carcinoma (PSC). *Oncoimmunology* (2021) 10(1):1947665. doi: 10.1080/2162402X.2021.1947665

124. de Lange MJ, Nell RJ, Lalai RN, Versluis M, Jordanova ES, Luyten GPM, et al. Digital PCR-based T-cell quantification-assisted deconvolution of the microenvironment reveals that activated macrophages drive tumor inflammation in uveal melanoma. *Mol Cancer Res* (2018) 16(12):1902–11. doi: 10.1158/1541-7786.MCR-18-0114

125. Lin D, Zhao W, Yang J, Wang H, Zhang H. Integrative analysis of biomarkers and mechanisms in adamantinomatous craniopharyngioma. *Front Genet* (2022) 13:830793. doi: 10.3389/fgene.2022.830793

126. Giuliani A. The application of principal component analysis to drug discovery and biomedical data. *Drug Discovery Today* (2017) 22(7):1069–76. doi: 10.1016/j.drudis.2017.01.005

127. Liu C, Reynolds PM, Alexander MR, Gadegaard N, Ghaemmaghami AM. Partial least squares regression and principal component analysis: similarity and differences between two popular variable reduction approaches. *Gen Psychiatr* (2022) 35(1):e100662. doi: 10.1136/gpsych-2021-100662

128. Montgomery DC, Peck EA, Vining GG. *Introduction to linear regression analysis*. John Wiley & Sons (2021).

129. DeGregory KW, Kuiper P, DeSilvio T, Pleuss JD, Miller R, Roginski JW, et al. A review of machine learning in obesity. *Obes Rev* (2018) 19(5):668–85. doi: 10.1111/obr.12667

130. Noble WS. What is a support vector machine? *Nat Biotechnol* (2006) 24 (12):1565–7. doi: 10.1038/nbt1206-1565

131. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W, et al. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* (2018) 15(1):41–51.

132. Blanchet L, Vitale R, Vorstenbosch van R, Stavropoulos G, Pender J, Jonkers D, et al. Constructing bi-plots for random forest: Tutorial. *Anal Chim Acta 2020* (1131) p:146–55.

133. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: A systematic review. *Front Aging Neurosci* (2017) 9:329. doi: 10.3389/fnagi.2017.00329

134. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. *IEEE Access* (2019) 7:53040–65. doi: 10.1109/ACCESS.2019.2912200

135. Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol* (2020) 7(7):e541–50. doi: 10.1016/S2352-3026(20)30121-6

136. Sarıgül M, Ozyildirim BM, Avci M. Differential convolutional neural network. *Neural Netw* (2019) 116:279–87. doi: 10.1016/j.neunet.2019.04.025

137. Cossu A, Carta A, Lomonaco V, Bacciu D. Continual learning for recurrent neural networks: An empirical evaluation. *Neural Netw* (2021) 143:607–27. doi: 10.1016/j.neunet.2021.07.021

138. Orr MJ. Introduction to radial basis function networks. In: *Technical report, center for cognitive science*. University of Edinburgh (1996).

139. Tai KS, Socher R, Manning CD. Improved semantic representations from tree-structured long short-term memory networks. *arXiv* (2015). arXiv:1503.00075. doi: 10.3115/v1/P15-1150

140. Kohonen T. *Self-organizing maps* Vol. Vol. 30. Springer Science & Business Media (2012).

141. Doersch C. Tutorial on variational autoencoders. *arXiv* (2016). arXiv:1606.05908.

142. Carreras J, Kikuti YY, Miyaoka M, Hiraiwa S, Tomita S, Ikoma H, et al. A combination of multilayer perceptron, radial basis function artificial neural networks and machine learning image segmentation for the dimension reduction and the prognosis assessment of diffuse large b-cell lymphoma. *AI* (2021) 2(1):106–34. doi: 10.3390/ai2010008

143. Carreras J, Kikuti YY, Miyaoka M, Hiraiwa S, Tomita S, Ikoma H, et al. The use of the random number generator and artificial intelligence analysis for dimensionality reduction of follicular lymphoma transcriptomic data. *BioMedInformatics* (2022) 2(2):268–80. doi: 10.3390/biomedinformatics2020017

144. Carreras J, Kikuti YY, Miyaoka M, Hiraiwa S, Tomita S, Ikoma H, et al. Artificial intelligence analysis of the gene expression of follicular lymphoma predicted the overall survival and correlated with the immune microenvironment response signatures. *Mach Learn Knowledge Extraction* (2020) 2(4):647–71. doi: 10.3390/make2040035

145. Risom T, Glass DR, Averbukh I, Liu CC, Baranski A, Kagel A, et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* (2022) 185(2):299–310.e18. doi: 10.1016/j.cell.2021.12.023

146. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Diversity* (2021) 25(3):1315–60. doi: 10.1007/s11030-021-10217-3

147. Schelter S, Biessmann F, Januschowski T, Salinas D, Seufert S, Szarvas G, et al. On challenges in machine learning model management. (2018).

148. Arora S, Dev K, Agarwal B, Das P, Syed MA, et al. Macrophages: Their role, activation and polarization in pulmonary diseases. *Immunobiology* (2018) 223(4-5):383–96. doi: 10.1016/j.imbio.2017.11.001

149. Wang LX, Zhang SX, Wu HJ, Rong XL, Guo J, et al. M2b macrophage polarization and its roles in diseases. *J Leukoc Biol* (2019) 106(2):345–58. doi: 10.1002/JLB.3RU1018-378RR

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fimmu.2022.1019967

Check for updates

OPEN ACCESS

EDITED BY
Hao Zhang,
Xiangya Hospital Central South University, China

REVIEWED BY
Guo-Peng Wang,
Capital Medical University, China
Dhyan Chandra,
University at Buffalo, United States

*CORRESPONDENCE
Xin Chen
dr.chenxin@tjh.tjmu.edu.cn
Qiuhui Li
qiuhui.li@whu.edu.cn
Jing Mao
maojing@hust.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 15 August 2022
ACCEPTED 06 September 2022
PUBLISHED 23 September 2022

CITATION
Li Q, Chen W, Li Q, Mao J and Chen X (2022) A novel neutrophil extracellular trap signature to predict prognosis and immunotherapy response in head and neck squamous cell carcinoma. *Front. Immunol.* 13:1019967. doi: 10.3389/fimmu.2022.1019967

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

A novel neutrophil extracellular trap signature to predict prognosis and immunotherapy response in head and neck squamous cell carcinoma

Qilin Li[1,2,3], Weimin Chen[1,2,3], Qiuhui Li[4]*, Jing Mao[1,2,3]* and Xin Chen[5]*

[1]Department of Stomatology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]School of Stomatology, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Hubei Province Key Laboratory of Oral and Maxillofacial Development and Regeneration, Wuhan, China, [4]State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan, China, [5]Department of Oncology, Tongji Hospital, Tongji Medical School, Huazhong University of Science and Technology, Wuhan, China

Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant cancers, and patients with HNSCC possess early metastases and poor prognosis. Systematic therapies (including chemotherapy, targeted therapy, and immunotherapy) are generally applied in the advanced/late stages of HNSCC, but primary and acquired resistance eventually occurs. At present, reliable biomarkers to predict the prognosis of HNSCC have not been completely identified. Recent studies have shown that neutrophil extracellular traps (NETs) are implicated in cancer progression, metastasis and cancer immune response, and NET-related gene signatures are associated with the prognosis of patients with several human cancers. To explore whether NET-related genes play crucial roles in HNSCC, we have performed systematic analysis and reported several findings in the current study. Firstly, we identified seven novel NET-related genes and developed a NET-score signature, which was highly associated with the clinicopathological and immune traits of the HNSCC patients. Then, we, for the first time, found that NIFK was significantly upregulated in HNSCC patient samples, and its levels were significantly linked to tumor malignancy and immune status. Moreover, functional experiments confirmed that NIFK was required for HNSCC cell proliferation and metastasis. Altogether, this study has identified a novel NET-score signature based on seven novel NET-related genes to predict the prognosis of HNSCC and NIFK has also explored a new method for personalized chemo-/immuno-therapy of HNSCC.

KEYWORDS

head and neck squamous cell carcinoma, neutrophil extracellular traps, immune cells (ICs), immunotherapy, prognosis

Introduction

Head and neck cancer (HNC) ranks sixth in terms of malignancy worldwide, and about 90% of HNCs are classified as head and neck squamous cell carcinomas (HNSCC) (1). HNSCC possesses a high incidence of cervical lymph node metastases, increased capacities of invasive and recurrence, and contributes to the poor prognosis (2, 3). The main treatment options for HNSCC include surgery, chemotherapy, radiotherapy, molecular targeted therapy, and multimethod in conjunction with surgical excision of tumor tissue, but the effectiveness of these approaches is limited due to tumor heterogeneity (4, 5).

Cancer immunotherapy is based on harnessing the immune system to detect and eliminate tumor cells, and the field of cancer immunotherapy has been growing with an increasing rate in modern oncology since it was first mentioned in 1985 (6–8). Active immunotherapy, passive immunotherapy, and immune checkpoint blockade are the major strategies of cancer immunotherapy (9, 10). For HNSCC immunotherapy, two immune checkpoint blockade agents, pembrolizumab and nivolumab, have been applied in clinical trials for patients with platinum-refractory HNSCC (11, 12). However, most HNSCC patients are non-responders and have acquired drug resistance (13–15). Recent studies have indicated that cancer immunotherapy may be hindered by immunosuppressive cells of the tumor microenvironment (TME), leading to the failure of antitumor immunity (16, 17).

Neutrophils are the most abundant immune cells in the TME, and increased neutrophil infiltration and high neutrophil-to-lymphocyte ratios were reported to be associated with poor patient outcomes of the patients with HNSCC (18–20). In activating neutrophils, DNA fibers decorated with histones and antimicrobial proteins found originally within neutrophil granules are released as neutrophil extracellular traps (NETs) (21, 22). NETs have been found as a new form of innate immunity and mediate the response of the host as a first line of defense (23). The development of NETs is a potential mechanism that contributes to tumor progression. Additionally, tumor cells can also escape immune surveillance through NETs (24). While the pro-oncogenic evidence of NETs is growing, the role of NETs in cancer immunotherapy remains unclear, particularly in HNSCC.

In this study, we have developed a novel NETs-score signature consisting of seven NETs-related genes, and we have found that NETs-score could reflect the response of HNSCC patients to chemotherapy and immunotherapy. Lastly, we have further identified the NET-related gene NIFK as a potentially carcinogenic factor for patients with HNSCC.

Methods and material

HNSCC database handling

HNSCC-related clinical information has been downloaded and collected from the TCGA database (519 patients, https://xenabrowser.net/) and GEO database (ID: GSE41613 n = 97, GSE42743 n = 103, GSE65858 n = 270, https://www.ncbi.nlm.nih.gov/geo/). Transcriptional profiles of 989 HNSCC patients were obtained from four cohorts, and the patients with insufficient OS information were excluded (Table S1).

The Affymetrix and Illumina platforms were used to generate raw data from the TCGA and GEO databases. Background correction and normalization are achieved using a robust multi-chip averaging (RMA) algorithm. The TCGA database provides RNA sequencing data. The fragment per kilobase (FPKM) values were converted to transcripts per kilobase (TPM) values with signal intensities similar to the RMA treatment.

Establishment of NET enrichment score

According to a recent study, we obtained a list of published NET gene sets and the descriptions of the gene sets (25–27) (Tables S2, S3). This NET-related gene set has a total of 69-gene with NET-initial biomarkers. We first performed univariate cox analysis to screen out the NET gene set associated with the prognosis of HNSCC patients for subsequent enrichment score calculation. NET-enrichment-score was calculated with single sample Gene Set Enrichment Analysis (ssGSEA) for HNSCC patients using the NET gene set associated with prognosis for further analysis.

Establishment of a NET-related signature

We employed Spearman correlation analysis to identify genes that were significantly positively correlated with NET-enrichment-score (correlation coefficient >0.4 and P-value <0.05, termed NET-related genes) and were selected for further analysis. Subsequently, we performed univariate Cox regression analysis to identify NET-related genes associated with the prognosis of HNSCC patients (*P*-value <0.05). We then screened out more valuable NET-related genes with prognostic potential by applying machine learning algorithms through the R "CoxBoost" and "randomForestSRC" packages. The NET-related signature named NET-score was constructed from the list of NET-related genes with prognostic

potential and weighted by their estimated regression coefficients in the Lasso regression analysis. Finally, we verified the prognostic evaluation performance of the NET-score. We estimated the NETs-score of 519 patients in the TCGA–HNSCC dataset, and then divided the patients into high and low NET-score groups based on the *P* value of the best cut-off. Kaplan–Meier curve analysis of the association between OS and NET-score. Time-ROC was used to validate the efficiency and accuracy of the NET-score for 1-year, 3-year, and 5-year prognosis prediction. Univariate and multivariate cox regression analyses were performed on the NET-score.

Genomic alteration

Somatic mutation and somatic copy number variation (CNV) data were collected from the TCGA dataset. Genomic Identification of Important Targets in Cancer (GISTIC) analysis was used to assess genomic signatures. The CNV landscape and the copy number gain or loss of amplified or deleted peaks were assessed by GISTIC 2.0 analysis (https://gatk.broadinstitute.org).

Assessing the immunological profile of the TME

We first used the ESTIMATE (The Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression) algorithm to estimate the abundance of immune cells and the infiltration level of stromal cells in HNSCC tumor tissue, which were reflected by immune score, stromal score, and estimated score, respectively. The Tumor Immune Estimation Resource2.0 (TIMER2.0, http://timer.cistrome.org/) web server was used to comprehensively analyze the level of immune-infiltrating cells in HNSC. Then, the relative proportions of 10 immune cells in the tumor were estimated using the MCPcounter algorithm. The infiltration levels of the 28 immune cells were represented by the enrichment scores based on the corresponding features. Enrichment scores were calculated using Single-Sample Genomic Enrichment Analysis (ssGSEA) implemented using the R Genomic Variation Analysis (GSVA) package. The response of HNSCs to anti-PD1 and anti-CTLA4 therapy was assessed by the submap algorithm. Response to anti-immune checkpoint therapy was assessed by the TIDE algorithm (28).

Functional annotation of differently expressed NET-related genes

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets were downloaded from the MSigDB database (29). Gene Set Enrichment Analysis (GSEA) and Gene Set Enrichment Analysis (GSEA) and GSVA are implemented by the clusterProfiler R package and the GSVA R package (30).

Prediction of drug response

We first used the Pharmacogenomics Data of Cancer Drug Sensitivity Genomics (GDSC, https://www.cancerrxgene.org/) to predict drug susceptibility in the included HNSCC cases (31). Drug responses were calculated with the oncoPredict R software package for drug sensitivity (32).

Plate clone formation assay

The Cal27 and SCC25 cells were digested and then resuspended in serum-free medium, and the cells were seeded into a 6-well culture plate at a density of 10^3 cells per well. Fourteen days later, the cells were continually cultured. Every 3 days, cells and clones were observed microscopically and sub-cultured. After colony formation was completed, the colonies formed by cells were photographed under a microscope and washed three times with PBS. Then, add 1 ml of crystal violet staining solution to each well and stain for 10–20 min. Finally, the six-well plate that formed the clones was scanned.

Transwell assay

Cal27 and SCC25 cells were added to the upper chamber with 200 µl of serum-free medium. In the lower chamber, 650 µl of medium containing 10% fetal serum was added. In the upper chamber, the rest of the cells were removed with a cotton swab, and those on the surface of the lower chamber were treated with 4% paraformaldehyde for 15 min at room temperature and stained with 0.1% gentian violet for 30 min. Cells from the lower chamber (migrated cells) were imaged under an inverted microscope.

RNA interference assay

Short hairpin RNA (sh-RNA) sequences of NIFK were synthesized by RiboBio (Guangzhou, China), and the target sequences of sh-NIFK are as follows: sh-NIFK#1: CATCAGT GAAACGGTATAATC, sh-NIFK#2:CGGATGGAGGA GCGATTTAAA. Based on our previous study (30), lentivirus vectors including short hairpin RNA were used for the RNA interference assay.

Statistical analysis

The Wilcoxon test was used for data that did not conform to a normal distribution. A t-test was used for normally distributed data. Kaplan–Meier survival plots were used to estimate OS between the two groups using the R package "survminer." Cox

regression for survival analysis was performed using the R package "survival." Time-dependent receiver operating characteristic (ROC) curves were plotted using the R package "timeROC." All heatmaps were performed *via* the R "pheatmap" package. Data were primarily visualized using ggplot2 R software (v4.1.2). A P-value of <0.05 was considered statistically significant.

Results

Identification of NET-enrichment-scores for the patients with HNSCC

Previous studies have applied 69 genes as the neutrophil extracellular trap (NET)-initial biomarkers. To identify a NET-relevant signature for HNSCC, the 69-gene NET-initial biomarkers were applied in the uniCox regression analysis in the TCGA-HNSCC training set, and we found 12 NET-associated genes with prognostic potential in HNSCC, including KCNJ15, CREB5, MME, F3, IL6, CXCL8, SELP, VNN3, CTSG, KCNN3, SELPLG, and IL17A, where the hazard ratio (HR) originated from uniCox regression analysis for each gene was included (Figure 1A). The ssGSEA was then applied to the TCGA-HNSCC with the 12 NET-associated genes, and a NETs-enrichment-score was established on the basis of their expression levels. Moreover, the correlation analysis showed that there was a strong correlation between the 12 NET-related genes (Figure 1B). Furthermore, the Kaplan–Meier analysis of the 12 NET-related genes showed that the survival of HNSCC patients was inversely correlated with the NET-enrichment-score, implying that HNSCC patients with high levels of the NET-enrichment-score may have a worse prognosis (Figure 1C). Finally, a heatmap displayed the correlation between the NET-enrichment-scores and the clinical characteristics of HNSCC samples, referring to the clinical stages, grade, gender, and age (Figure 1D). The results showed that each of the NET-associated genes exhibited a strong correlation with NET-enrichment-scores, which correlated with the clinical characteristics of HNSCC patients.

Establishment of a 7-gene NET related signature for HNSCC

Spearman correlation analysis identified 38 NET-related genes were positively correlated with NET-enrichment-scores, based on the criteria with correlation coefficient >0.4 and *P*-value <0.05. Heatmap showed that these 38 NET-associated genes correlated with NET-enrichment-scores and clinical futures of the HNSCC patients (Figure 2A). To further screen out the more valuable NET-related genes, univariate Cox regression analysis was performed to further select out 34 NET-related genes of potential prognostic value for HNSCC patients (*P*-value <0.05), and the univariate Cox analysis forest plot showed the HR of each single gene of the 34 NET genes for their prognosis (Figure 2B). Moreover, we used machine learning algorithms by applying the R "CoxBoost" to further select out nine NET-related genes (Figure 2C). Interestingly, the randomForestSRC (RFC) enabled us to further screen out seven NET-related genes, including NIFK, LINC00460, NUTF2, LINC02454, ITGA5, TNFRSF12α, and PDGFα (Figure 2D). Finally, the Lasso regression analysis was performed to calculate new NET-scores using these seven prognostic and NET-related genes based on their estimated regression coefficients (Figure 2E). Each estimated regression coefficients of the prognostic-related NET gene were following, 0.1936 ∗ ITGA5 + 0.4588 ∗ LINC00460 + 0.0361 ∗ LINC02454 + 1.1349 ∗ NIFK + 0.4079 ∗ NUTF2 + 0.4611 ∗ PDGF α + 0.2251 ∗ TNFRSF12 α.

Validation of the NET-scores for clinical predicting the survival in HNSCC

To test if the 7-gene NET-score signature predicted the clinical characteristics of the HNSCC patients, Kaplan–Meier analysis revealed that HNSCC patients with higher NET-scores had poor survival curves (Figure 3A). Additionally, both univariate Cox and multivariate Cox regression analysis showed that the NET-score of HNSCC patients was an independent risk factor compared with other clinical factors, such as tumor grade and gender (Figure 3B). As shown in Figure 3C, the time-dependent ROC curves at 1 year, 3 years, and 5 years of OS had AUC values of 0.685, 0.712, and 0.746, respectively, and these results indicated that our NET-score signature was of prognostic value. Furthermore, we used three independent cohorts in the GEO database (ID: GSE41613, GSE42743, and GSE65858) and further verified that HNSCC patients with higher NET-scores had worse prognosis and survival disadvantages (Figures 3D–F).

The NET-score was relevant to distinct genomic alterations of HNSCC patients

To check if somatic mutations are linked to the NET-scores, we conducted the Genomic Identification of Important Targets in Cancer (GISTIC) analysis and the results showed that HNSCC patients with either high or low NET-scores manifested respective somatic mutations, including TP53, PIK3CA, NOTCH1, and MUC17 (Figures 4A, B). Nevertheless, HNSCC patients with high NET-scores appeared to have a higher trend of TP53 mutations as compared to patients with low NET-scores, i.e., 71% vs. 63%, respectively (Figures 4A, B). Moreover, analysis of the copy number alterations showed that HNSCC patients with either high or low NET scores displayed copy number changes significantly at multiple chromosome loci (Figure 4C), probably related to the clinicopathological features of the HNSCC patients.

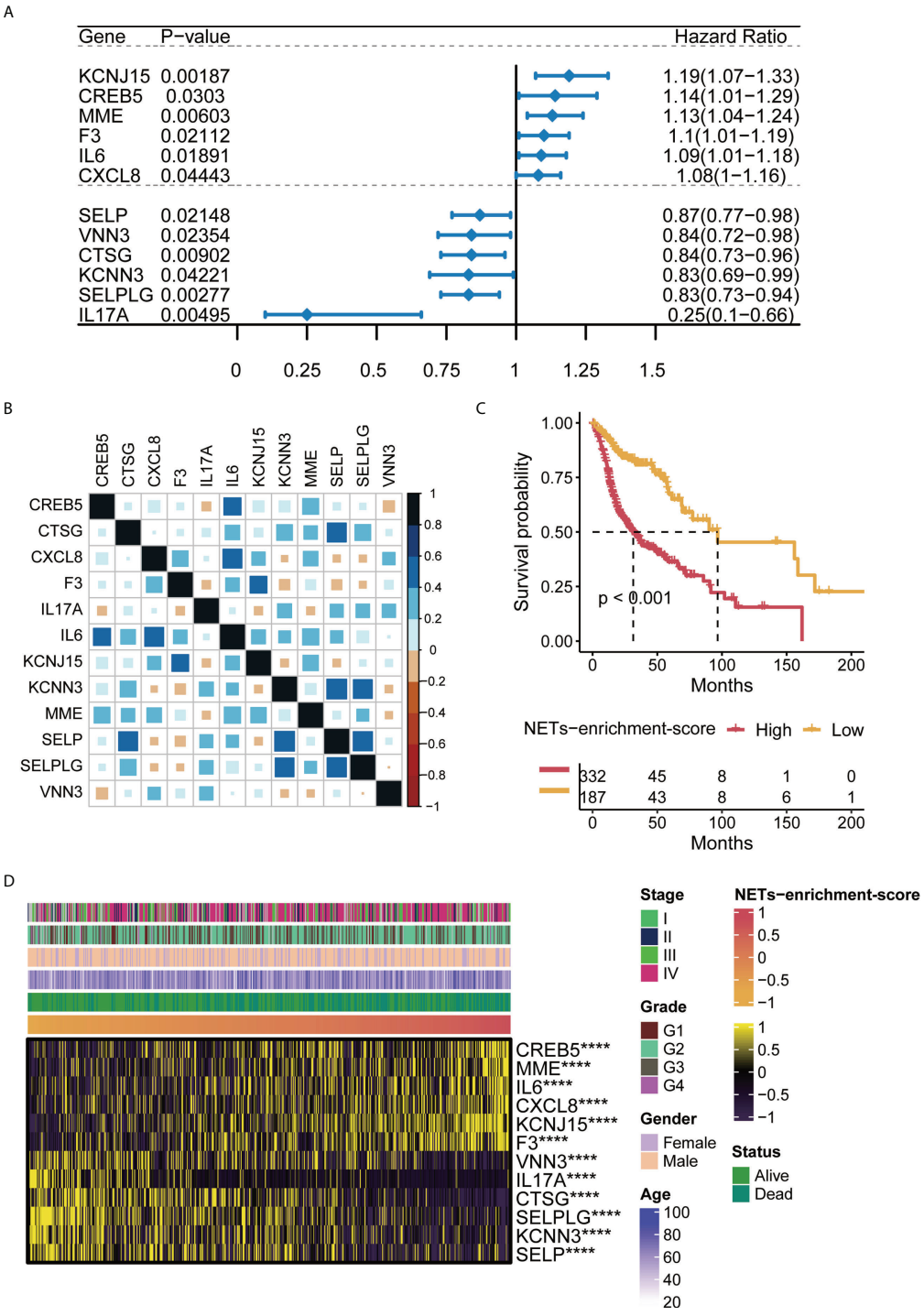


FIGURE 1
Characteristics of NET-enrichment in HNSC (TCGA). **(A)** Univariate Cox analysis results of 12 prognostic related NET genes in TCGA-HNSCC. **(B)** Correlation map of 12 prognosis-related NET gene sets in TCGA-HNSCC. **(C)** Kaplan–Meier curve showing the correlation of NET-enrichment-score with OS. **(D)** Heatmap showing the correlation of the NET-enrichment-score with 12 prognosis-related NET genes and their clinical features. ****,p<0.001.

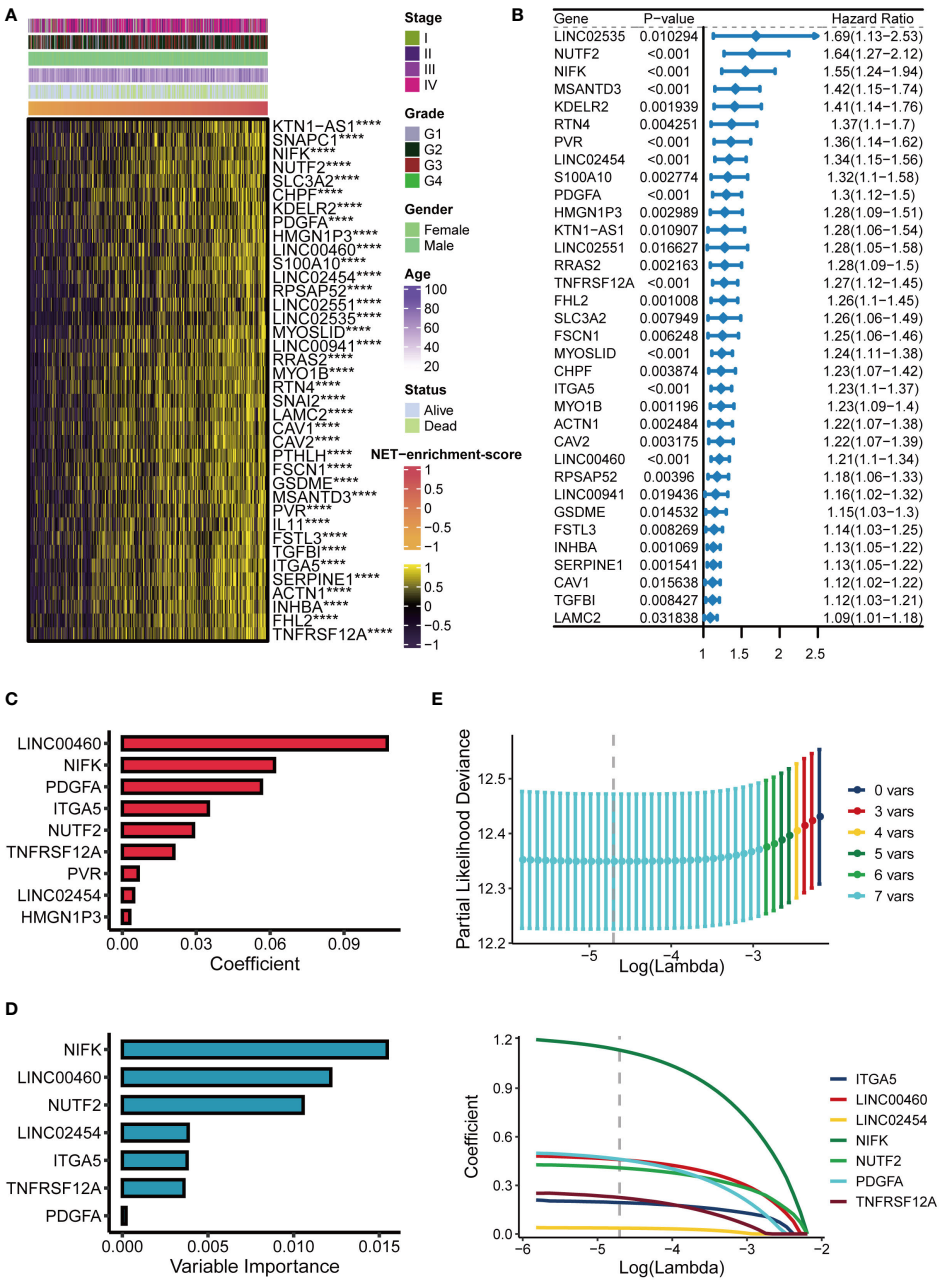


FIGURE 2
Establishment of the NET score signature. **(A)** Heatmap of 38 NET-related genes significantly positively correlated with NET-enrichment-score. **(B)** Univariate Cox analysis forest plot of 34 prognosis-related NET-related genes. **(C)** The machine learning method CoxBoost further screened NET-related genes (34 dimensionality reduction to 9). **(D)** Machine learning method for survival random forest to further screen NET related genes (reduced from 9 to 7). **(E)** Lasso regression method to calculate NET scores. ****,p<0.001.

The NET-scores are conversely related to immune infiltration for patients with HNSCC in TCGA cohorts

To examine the relationship between NET-scores and the immune status in patients with HNSCC in TCGA cohorts, we applied the ESTIMATE algorithm and found that HNSCC patients with low-NET scores had significantly higher ESTIMATE scores, higher immune scores, and higher stromal cells than those with high-NET scores (Figure 5A), indicating that the levels of NET scores are reversely correlated with immune status in HNSCC patients. To confirm this

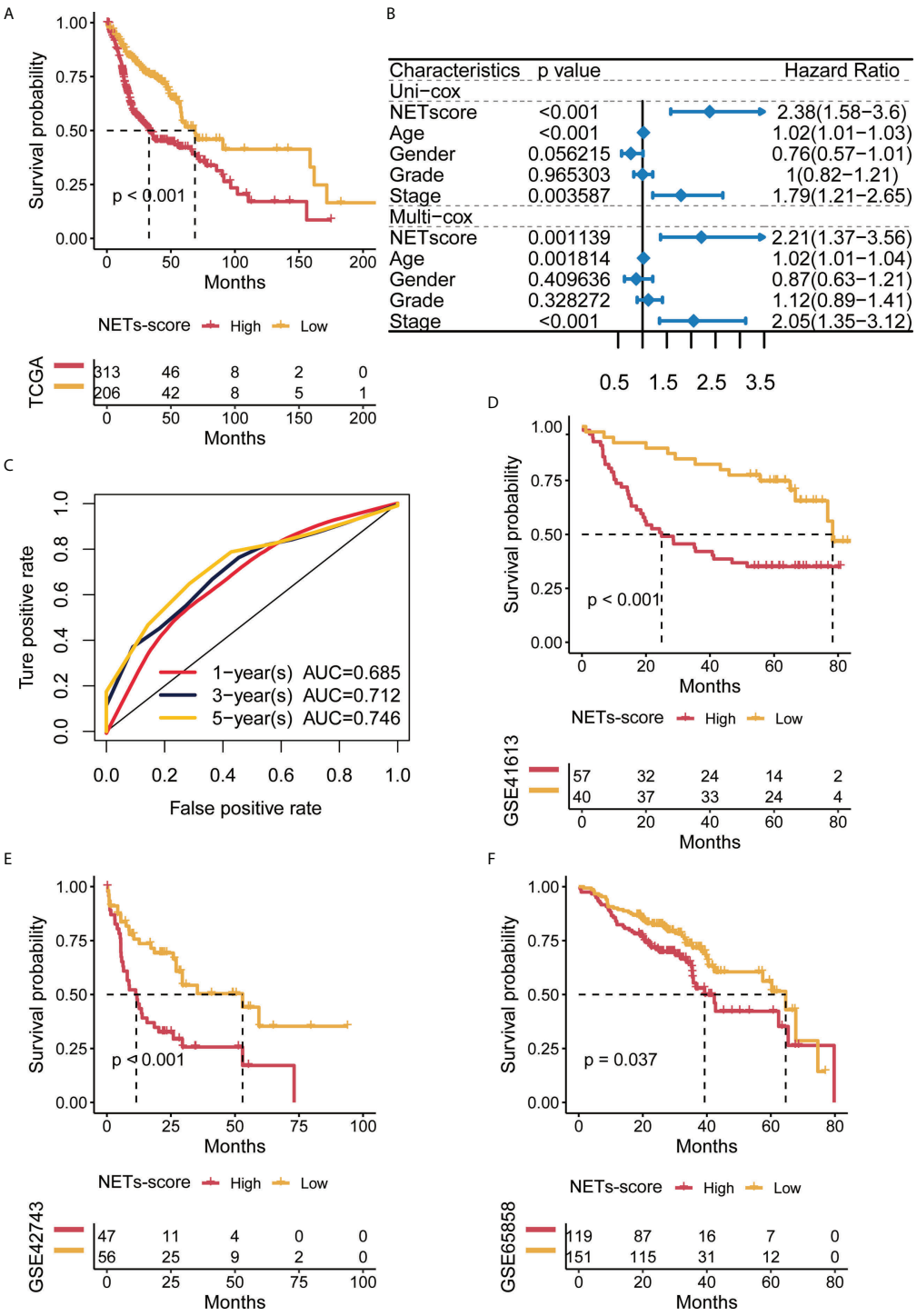


FIGURE 3
Predictive potential of the NET score for prognosis in HNSCC patients. **(A)** Kaplan–Meier curves of high and low NET scores in TCGA-HNSCC. **(B)** Forest plot of univariate and multivariate Cox regression of NET-score based on TCGA dataset. **(C)** Time-dependent ROC of NET-score in TCGA. **(D–F)** Kaplan–Meier curves of overall survival in HNSCC patients based on an external validation dataset.

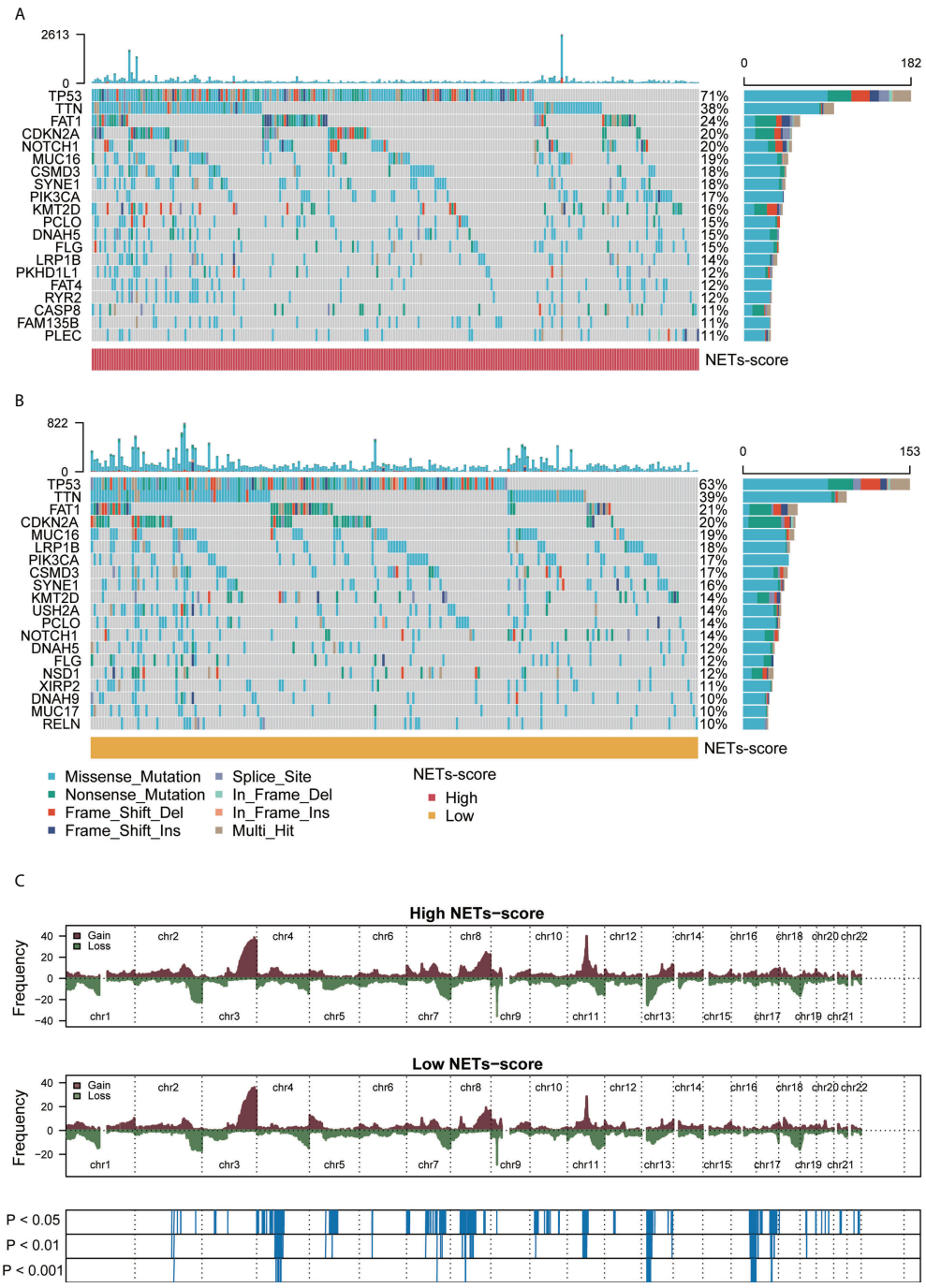


FIGURE 4
Genomic alterations associated with NET scores in HNSCC samples. **(A,B)** Waterfall plot of somatic mutations in HNSCCs between high and low NET-score groups. **(C)** Copy number changes of HNSCCs between NET-score high and low groups.

phenotype, the MCPcounter, ssGSEA, and TIMER algorithms were independently used to reveal the abundance of immune infiltrating cell populations based on the NET-scores, tumor stages, tumor grade, gender, and age (Figure 5B). As a result, the heatmap showed that many immune-infiltrating cells were enriched in HNSCC patients with low NET-scores, including CD8 T cells, cytotoxic lymphocytes, NK cells, and neutrophils (Figure 5B). Additionally, the correlation analysis implied that the NET-scores were negatively associated with the abundance of neutrophils in HNSCC patients (Figures 5C, D). Moreover,

10.3389/fimmu.2022.1019967

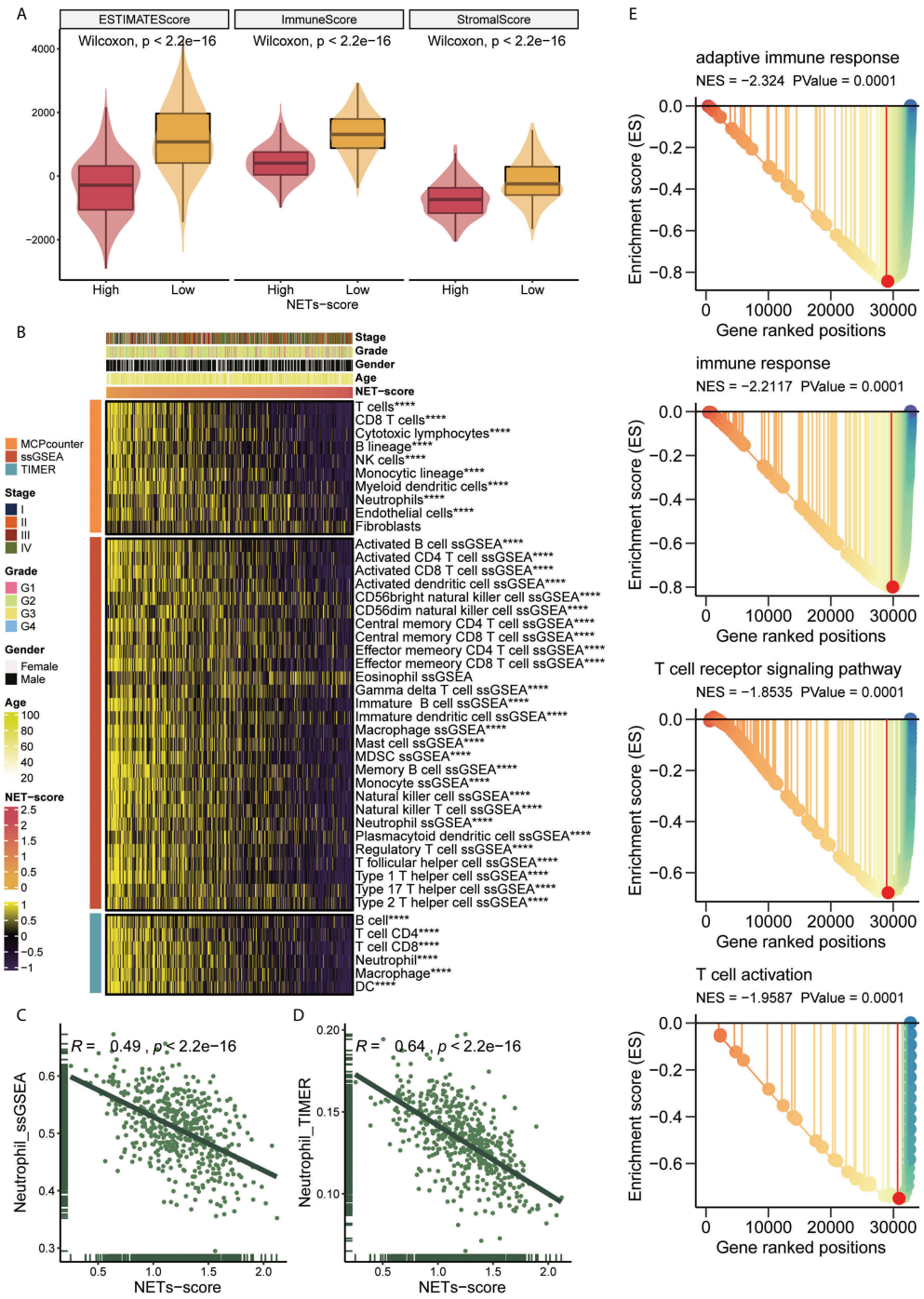


FIGURE 5
NET scores in relation to immunity in the TCGA cohort. **(A)** Changes in ESTIMATE among HNSCC patients with high and low NET scores. **(B)** Heatmap showing the abundance of infiltrating immune cell populations for different NET scores according to MCPcounter, ssGSEA, and TIMER algorithms. **(C,D)** Correlation of NET scores with Neutrophil_ssGSEA and Neutrophil_TIMER. **(E)** GSEA showing immune related pathways potentially related by NET-score. ****,p<0.001.

GSEA showed that several important immune-related pathways were more involved in HNSCC patients with low NET-scores, such as adaptive immune response, immune response, T-cell receptor signaling pathways, and T-cell activation (Figure 5E). Thus, our data revealed that the NET-scores for HNSCC patients may be highly linked to the tumor immune microenvironment.

The potential immunotherapeutic and chemotherapeutic response associated with NET-score of HNSCC patients

Recent developments in immunotherapy, particularly PD-1 inhibitors, have led to the outperformance of traditional chemotherapy in HNSCC at the recurrent and metastatic stages. Using chemo-immunotherapy, chemotherapy interacts with immune cell mechanisms to enhance current cancer treatment strategies (33). To explore the therapeutic responsiveness based on NET-scores, we first checked the correlation between the NET-scores and the immune checkpoint levels in HNSCC patients. The heatmap showed that HNSCC patients with low NET scores tended to have higher levels of immune checkpoints, including CD274 and CTLA4 (Figure 6A). Subsequent analysis of drug sensitivity revealed that HNSCC patients with low NET-scores were further enriched in the responders, but not in the non-responders, when immune checkpoint inhibitors were performed (Figures 6B, C). In contrast, patients with high NET-scores were probably more non-responders as the immunotherapies were applied (Figures 6B, C). Moreover, NET-scores were significantly linked to the targeted therapies, including afatinib, lapatinib, erlotinib, and ibrutinib, indicating that patients with low NET-scores had a better response to the targeted therapies (Figure 6D).

NIFK is a potentially prognostic factor and oncogene for HNSCC patients

As shown in Figure 2D, NIFK was ranked at the top of the 7-gene signature as per their variable importance. NIFK, a nucleolar protein interacting with the fork head associated (FHA) domain of Ki67, may play a role in cell cycle progression and mitosis. However, the function of NIFK in human cancer development is not clear. We therefore examined the expression of NIFK in HNSCC samples and their matched benign or normal tissues, and found that NIFK was significantly expressed in the tumorous samples as compared to their normal counterparts (Figure 7A). Further Kaplan–Meier analysis implied that NIFK level is reversely correlated with the survival of HNSCC patients using TCGA cohorts, showing that HNSCC patients with high a level of NIFK had a worse prognosis (Figure 7B). In addition, the GSVA heatmap showed several NIFK associated pathways, such as GO immune-related functions and KEGG tumor-related pathways (Figure 7C). For example, a high level of NIFK was closely linked to KEGG pathways regulating cell cycle, DNA replication, and proteasome, while a low level of NIFK was associated with immune-related pathways involving natural killer cell differentiation, leukocyte proliferation, and immune response (Figure 7C). Subsequent analysis hinted that low levels of NIFK were pertinent to levels of immune checkpoint in TCGA (Figure 7D). Additionally, to address the oncogenic role of NIFK in HNSCC, shRNA was used to knock down the mRNA expression levels of NIFK in two HNSCC cell lines that are widely used, such as Cal27 and SCC25. Using Transwell assay in Cal27/SCC25 control, sh-NIFK#1 and sh-NIFK#2 cells, and the results revealed that cell metastasis capacity in Cal27 and SCC25 cells was inhibited by NIFK shRNA (Figures 8A–D). Subsequently, to investigate the effect of NIFK on the proliferation of HNSCC, we conducted the plate cloning assay with Cal27/SCC25 control, sh-NIFK#1, and sh-NIFK#2 cells, and the results revealed that colony formation in Cal27 and SCC25 cells was inhibited by NIFK shRNA (Figures 8E–H). Thus, these data demonstrate that NIFK is a promising factor for predicting the prognosis of HNSCC patients.

Discussion

Head and neck squamous cell carcinomas (HNSCCs) are one of the most common malignant cancers (1, 34, 35). The mainstay treatments for HNSCC at the early stage are surgery and/or radiation, which benefit most patients with a good prognosis (4, 36, 37). For HNSCC patients at the advanced or late stages, systematic therapies are recommended, including chemotherapy, targeted therapy, and immunotherapy (12, 38, 39). Although many HNSCC patients at the advanced/late stages initially respond well to these treatments, most of them will eventually fail and progress to recurrent and/or metastatic diseases (2, 4, 40). For example, a randomized phase III clinical trial to compare the efficacy between different strategies of chemotherapy in advanced HNSCC showed that patients treated by cisplatin and fluorouracil (CF) had a median survival of 8.7 months, as compared to patients administered by cisplatin and paclitaxel (CP) with a median survival of 8.1 months (41). Further studies have revealed that cetuximab (a regimen of targeted therapy) plus platinum-fluorouracil chemotherapy significantly improved overall survival (OS) of recurrent or metastatic HNSCC to 10.1 months, as compared to the OS of patients of 7.4 months treated by platinum-fluorouracil chemotherapy alone (42). Recently, immunotherapy has been widely used in a variety of human cancers, including the recurrent or metastatic HNSCC (43, 44). Emerging evidence has demonstrated that the immune checkpoint targeting agent (such as Pembrolizumab) either alone or combined with chemotherapy significantly prolongs

10.3389/fimmu.2022.1019967

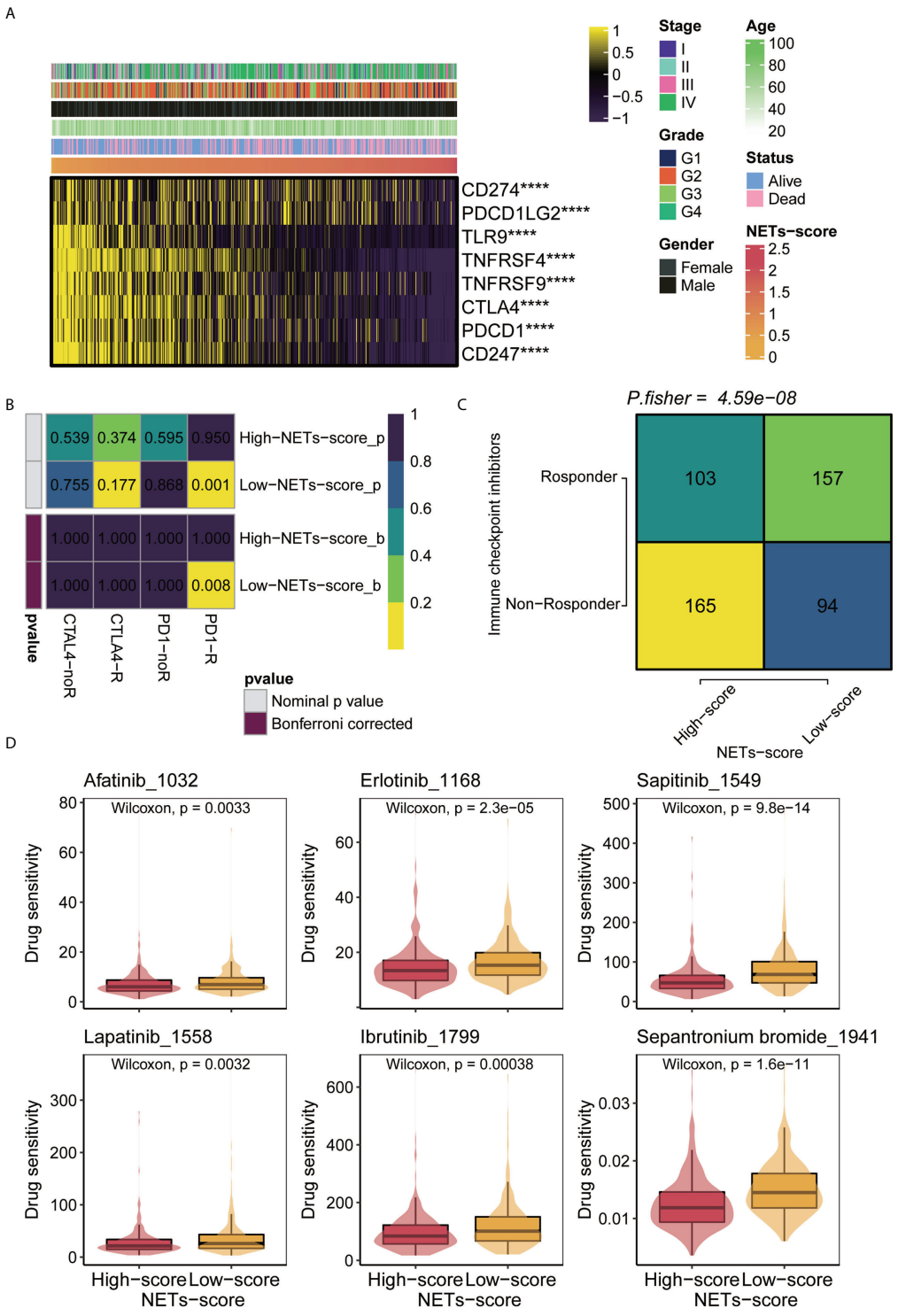


FIGURE 6
Immunotherapy and chemotherapy of NET scores involved in TCGA-HNSCC. **(A)** Correlation of NET scores and immune checkpoint levels in HNSCC. **(B)** Submap analysis of NET scores in TCGA-HNSCC. **(C)** TIDE analysis of NETs scores in TCGA-HNSCC. **(D)** Boxplots of estimated drug sensitivities for several GDSC chemotherapeutics in the high and low NET scores groups. ****,p<0.001.

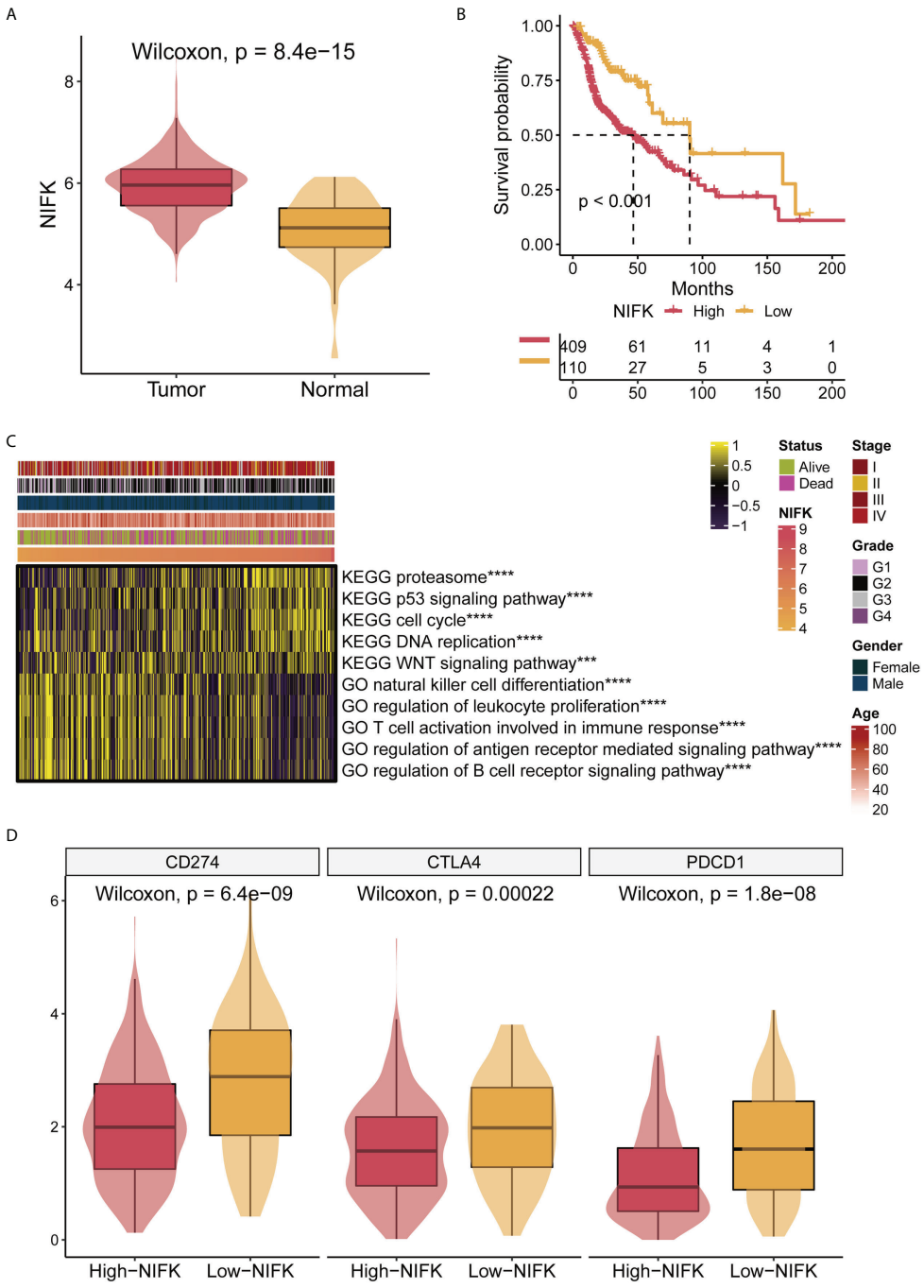


FIGURE 7
NIFK has an important role in TCGA-HNSCC. **(A)** NIFK levels in HNSCC samples grouped by cancer and para-cancerous status in TCGA. **(B)** Kaplan–Meier curves of TCGA high and low NIFK groups. **(C)** GSVA heatmap showing functional pathways significantly associated with NIFK in TCGA (GO immune-related functions and KEGG cancer-related pathways). **(D)** Association analysis showed that NIFK levels were related to immune checkpoints in TCGA. ***,p<0.001; ****,p<0.0001.

OS of the recurrent/metastatic HNSCC with a CPS (the PD-L1 combined positive score) of ≥20, as compared to cetuximab with chemotherapy (45, 46). Therefore, immunotherapy has been recommended as the first-line therapy for recurrent, unresectable, or metastatic head and neck cancers. Nevertheless, either primary or acquired resistance will eventually occur after treatment with immunotherapy. At present, several mechanisms have been proposed to explain

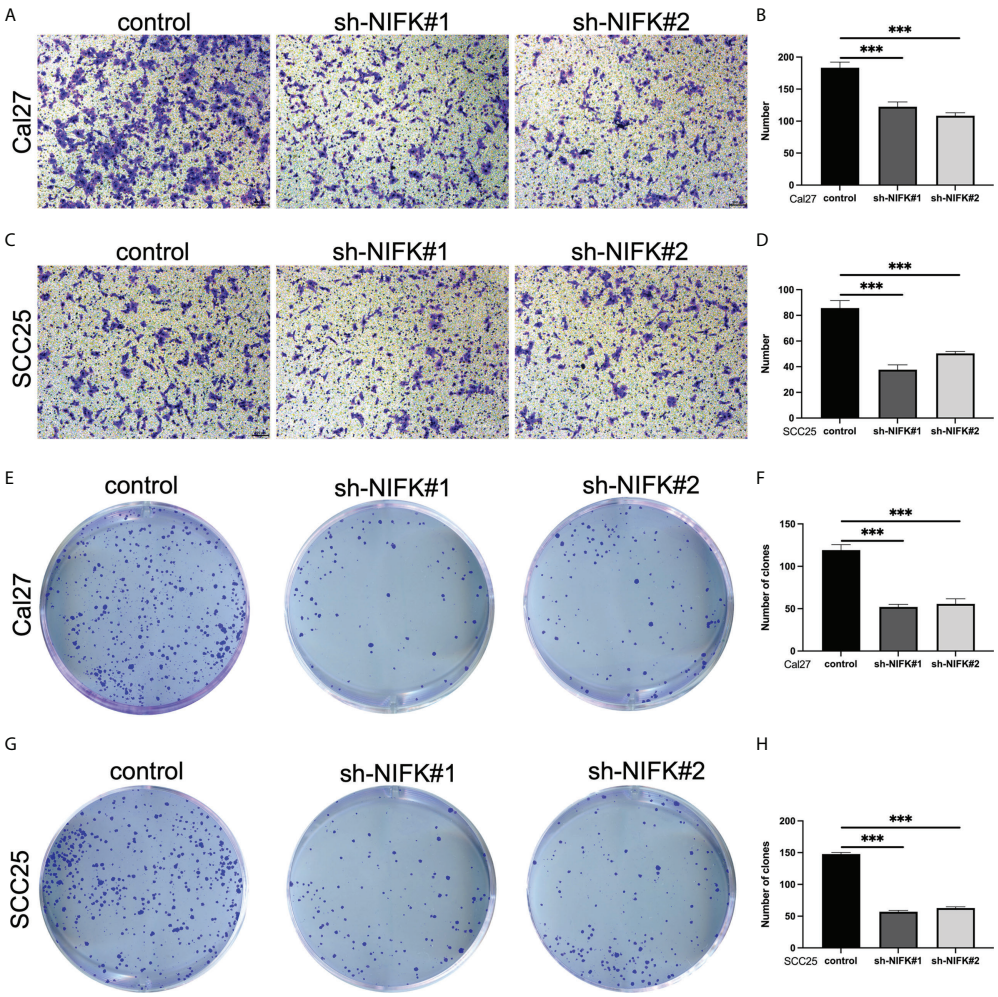


FIGURE 8
NIFK promotes tumor cell proliferation and metastasis in HNSCC. **(A)** Transwell migration assay in Cal27 cells silenced with control (sh-NC) or NIFK sh-RNA (#1 and #2). **(B)** Quantitative analysis of Transwell assay in Cal27 control and sh-NIFK#1 and sh-NIFK#2 cells. **(C)** Transwell assay in SCC25 control and sh-NIFK#1 and sh-NIFK#2 cells. **(D)** Quantitative analysis of Transwell assay in SCC25 control and sh-NIFK#1 and sh-NIFK#2 cells. **(E)** Plate cloning assay in Cal27 control and sh-NIFK#1 and sh-NIFK#2 cells. **(F)** Quantitative analysis of plate cloning assay in Cal27 control and sh-NIFK#1 and sh-NIFK#2 cells. **(G)** Plate cloning assay in SCC25 control and sh-NIFK#1 and sh-NIFK#2 cells. **(H)** Quantitative analysis of plate cloning assay in SCC25 control and sh-NIFK#1 and sh-NIFK#2 cells. ***P <0.001.

these resistant phenotypes, including tumors failing to produce robust T-cell infiltration or tumors excluding T cells. However, the exact mechanisms for the resistance are not completely understood, which will continue to be the future direction in the field (44, 47). Thus, a major hurdle emerging in the field of cancer immunotherapy is the lack of reliable and predictable biomarkers for many cancer patients, including HNSCC (48).

Neutrophil extracellular traps (NETs) participate in the regulation of neutrophil development. They are web-like structures consisting of chromatin and granule proteins (49, 50). NETs have been demonstrated to be linked with different conditions *via* distinct mechanisms, such as inflammation, cell damage, and vascular thrombosis (27, 51). Recently, increasing evidence has shown that neutrophil extracellular traps (NETs) play vital roles in tumor initiation, progression, recurrence, and metastasis (27, 44). In particular, NETs may play essential roles in the tumor microenvironment and are crucial to cancer immunotherapy (52, 53). Additionally, several prognostic models based on NETs have been shown in various human cancers. However, whether NETs are also implicated in HNSCC development and if NETs offer prognostic and predicative value for HNSCC is not understood. Li et al. have shown that oral squamous cell carcinoma (OSCC) patients with late stages (III/IV) had a higher level of NETs compared to early stages (I/II), and NETs dictate a procoagulant phenotype that can be partially dampened by DNase I (54). Moreover, a recent study has

identified a 6-gene signature associated with NETs, consisting of LTF, CYBB, SELPLG, GAPDH, ANXA3, and CSF2, which contributes to a clinical prognostic model for HNSCC (55).

To explore the prognostic biomarkers for HNSCC, we have conducted a series of analyses and validations in the current study, and our findings have novel points. First, we systematically applied 69 NET-initial biomarkers using TCGA-HNSCC datasets and identified 12 NET-related genes potentially predictive of prognosis for HNSCC. Second, further analysis has identified seven NET-related genes, including NIFK, LINC00460, NUTF2, and LINC02454, some of which are potentially predictive biomarkers for human cancers. For instance, NUTF2 has been reported to be highly expressed in HNSCC, associated with a poor prognosis and related to immune cells, which may serve as a potential biomarker and target for HNSCC (56–58). Additionally, previous studies have established a 12-gene signature based on the fatty acid metabolism to predict the prognosis of HNSCC, which contain the long non-coding RNA, LINC00460, indicating its predictive role for HNSCC (36, 59, 60). Also, another long non-coding RNA, LINC02454, is linked to predicting the prognosis of thyroid cancer (61, 62). In our study, we set up NET scores based on the seven prognostic-related NET genes, and HNSCC with low-NET scores was related to better prognosis and survival of patients. Importantly, our data hinted that the NET scores for HNSCC patients may be correlated with clinical traits for prognostic prediction (Figures 3, 4). Third, HNSCC patients with low-NET scores had higher immune scores, higher stromal cells, and immune-related pathways, which responded well to immunotherapy and targeted therapies (such as afatinib and lapatinib). Thus, our findings suggest that the seven NET-related gene signatures are predictive of prognosis for HNSCC.

In the current study, we identified that NIFK was highly upregulated in HNSCC patient samples as compared to normal tissues, and HNSCC patients with a high level of NIFK had a worse prognosis and a shortened life span, indicating that NIFK is a potential prognostic biomarker for HNSCC, although further functional validation is required. In Figure 7, our characterizations have found that levels of NIFK were related to cell cycle and DNA replication as well as WNT and P53 signaling pathways. In support of the previous reports showing that NIFK is vital for cell cycle progression *via* RNA recognition motif dependent pre-rRNA maturation (63). Nevertheless, how NIFK functions in human cancers is largely unknown. Recent studies have shown that NIFK is indispensable for lung cancer development through Ki-67 dependent cell proliferation and CK1α/β-catenin activated metastasis (64). Whether NIFK plays a similar role in HNSCC development is not clear, and more work is needed for its verification. Our present study has also found that NIFK involvement in HNSCC progression is linked with immune response and immune associated pathways (Figure 7), hinting that NIFK is also a potential therapeutic target for immunotherapy for HNSCC, although future work is required to validate this conjecture.

However, our current study has potential limitations. For instance, detailed experimental studies must be added to explore the possible mechanisms of NIFK regulation in HNSCC using cell lines, animal models, and human samples. Furthermore, our seven NET-related gene signatures and our NET scores must be validated in the clinics *via* large-cohort and multicenter studies. Moreover, there exist several gene signatures (including this study) to predict the prognosis of HNSCC. Future studies should be considered to compare the similarities and differences among these signatures and to select the representative targets for HNSCC 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

QLL conceived and designed the study. QLL and XC drafted the manuscript. QLL and WC did the statistical analysis, supervised by QHL, JM, and XC. All authors have reviewed, critically revised, and approved the manuscript.

Funding

This study was supported by grants from the National Natural Science Foundation of China (No. 81802973) and the general project of the Natural Science Foundation of Hubei Province (No. 2020CFB844) for QHL. JM was supported in part by the Key Project of the Health Commission of Hubei Province (No. WJ2019Z006). XC was supported, in part, by a grant from the National Natural Science Foundation of China (No. 81602592).

Acknowledgments

We thank colleagues at the Department of Stomatology and Department of Oncology at Tongji Hospital for support and suggestions for our manuscript. We apologize to the colleagues whose work was not cited due to space limitations.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1019967/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654

2. Santos-de-Frutos K, Segrelles C, Lorz C. Hippo pathway and YAP signaling alterations in squamous cancer of the head and neck. *J Clin Med* (2019) 8(12):2131. doi: 10.3390/jcm8122131

3. Chen Y, Li ZY, Zhou GQ, Sun Y. An immune-related gene prognostic index for head and neck squamous cell carcinoma. *Clin Cancer Res* (2021) 27:330–41. doi: 10.1158/1078-0432.CCR-20-2166

4. Frankart AJ, Breneman JC, Pater LE. Radiation therapy in the treatment of head and neck rhabdomyosarcoma. *Cancers (Basel)* (2021) 13(14):3567. doi: 10.3390/cancers13143567

5. Wang X, Wu S, Liu F, Ke D, Wang X, Pan D, et al. An immunogenic cell death-related classification predicts prognosis and response to immunotherapy in head and neck squamous cell carcinoma. *Front Immunol* (2021) 12:781466. doi: 10.3389/fimmu.2021.781466

6. Koury J, Lucero M, Cato C, Chang L, Geiger J, Henry D, et al. Immunotherapies: Exploiting the immune system for cancer treatment. *J Immunol Res* (2018) 2018:9585614. doi: 10.1155/2018/9585614

7. Finn OJ. Cancer immunology. *N Engl J Med* (2008) 358:2704–15. doi: 10.1056/NEJMra072739

8. Moskovitz JM, Ferris RL. Tumor immunology and immunotherapy for head and neck squamous cell carcinoma. *J Dent Res* (2018) 97:622–6. doi: 10.1177/0022034518759464

9. Wu Z, Li S, Zhu X. The mechanism of stimulating and mobilizing the immune system enhancing the anti-tumor immunity. *Front Immunol* (2021) 12:682435. doi: 10.3389/fimmu.2021.682435

10. Cillo AR, Kurten CHL, Tabib T, Qi Z, Onkar S, Wang T, et al. Immune landscape of viral- and carcinogen-driven head and neck cancer. *Immunity* (2020) 52:183–99.e9. doi: 10.1016/j.immuni.2019.11.014

11. Chow LQM, Haddad R, Gupta S, Mahipal A, Mehra R, Tahara M, et al. Antitumor activity of pembrolizumab in biomarker-unselected patients with recurrent and/or metastatic head and neck squamous cell carcinoma: Results from the phase ib KEYNOTE-012 expansion cohort. *J Clin Oncol* (2016) 34:3838–45. doi: 10.1200/JCO.2016.68.1478

12. Cramer JD, Burtness B, Ferris RL. Immunotherapy for head and neck cancer: Recent advances and future directions. *Oral Oncol* (2019) 99:104460. doi: 10.1016/j.oraloncology.2019.104460

13. O'Donnell JS, Long GV, Scolyer RA, Teng MW, Smyth MJ. Resistance to PD1/PDL1 checkpoint inhibition. *Cancer Treat Rev* (2017) 52:71–81. doi: 10.1016/j.ctrv.2016.11.007

14. O'Donnell JS, Smyth MJ, Teng MW. Acquired resistance to anti-PD1 therapy: checkmate to checkpoint blockade? *Genome Med* (2016) 8:111. doi: 10.1186/s13073-016-0365-1

15. Pitt JM, Vetizou M, Daillere R, Roberti MP, Yamazaki T, Routy B, et al. Resistance mechanisms to immune-checkpoint blockade in cancer: Tumor-intrinsic and -extrinsic factors. *Immunity* (2016) 44:1255–69. doi: 10.1016/j.immuni.2016.06.001

16. Guller M, Herberg M, Amin N, Alkhatib H, Maroun C, Wu E, et al. Nutritional status as a predictive biomarker for immunotherapy outcomes in advanced head and neck cancer. *Cancers (Basel)* (2021) 13(22):5772. doi: 10.3390/cancers13225772

17. Mei Z, Huang J, Qiao B, Lam AK. Immune checkpoint pathways in immunotherapy for head and neck squamous cell carcinoma. *Int J Oral Sci* (2020) 12:16. doi: 10.1038/s41368-020-0084-8

18. Watt S, Vasquez L, Walter K, Mann AL, Kundu K, Chen L, et al. Genetic perturbation of PU.1 binding and chromatin looping at neutrophil enhancers associates with autoimmune disease. *Nat Commun* (2021) 12:2298. doi: 10.1038/s41467-021-22548-8

19. Powell D, Lou M, Barros Becker F, Huttenlocher A. Cxcr1 mediates recruitment of neutrophils and supports proliferation of tumor-initiating astrocytes in vivo. *Sci Rep* (2018) 8:13285. doi: 10.1038/s41598-018-31675-0

20. Ho WJ, Yarchoan M, Hopkins A, Mehra R, Grossman S, Kang H. Association between pretreatment lymphocyte count and response to PD1 inhibitors in head and neck squamous cell carcinomas. *J Immunother Cancer* (2018) 6:84. doi: 10.1186/s40425-018-0395-x

21. Thomas GM, Carbo C, Curtis BR, Martinod K, Mazo IB, Schatzberg D, et al. Extracellular DNA traps are associated with the pathogenesis of TRALI in humans and mice. *Blood* (2012) 119:6335–43. doi: 10.1182/blood-2012-01-405183

22. Skoglund C, Appelgren D, Johansson I, Casas R , Ludvigsson J. Increase of neutrophil extracellular traps, mitochondrial DNA and nuclear DNA in newly diagnosed type 1 diabetes children but not in high-risk children. *Front Immunol* (2021) 12:628564. doi: 10.3389/fimmu.2021.628564

23. Kato Y, Nishida O, Kuriyama N, Nakamura T, Kawaji T, Onouchi T, et al. Effects of thrombomodulin in reducing lethality and suppressing neutrophil extracellular trap formation in the lungs and liver in a lipopolysaccharide-induced murine septic shock model. *Int J Mol Sci* (2021) 22(9):4933. doi: 10.3390/ijms22094933

24. Deng H, Kan A, Lyu N, He M, Huang X, Qiao S, et al. Tumor-derived lactate inhibit the efficacy of lenvatinib through regulating PD-L1 expression on neutrophil in hepatocellular carcinoma. *J Immunother Cancer* (2021) 9(6): e002305. doi: 10.1136/jitc-2020-002305

25. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol* (2016) 17:231. doi: 10.1186/s13059-016-1092-z

26. Zhang Y, Guo L, Dai Q, Shang B, Xiao T, Di X, et al. A signature for pan-cancer prognosis based on neutrophil extracellular traps. *J Immunother Cancer* (2022) 10(6):e004210. doi: 10.1136/jitc-2021-004210

27. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol* (2018) 18:134–47. doi: 10.1038/nri.2017.105

28. Brunner M, Thurnher D, Pammer J, Geleff S, Heiduschka G, Reinisch CM, et al. Expression of VEGF-A/C, VEGF-R2, PDGF-alpha/beta, c-kit, EGFR, her-2/Neu, mcl-1 and bmi-1 in merkel cell carcinoma. *Mod Pathol* (2008) 21:876–84. doi: 10.1038/modpathol.2008.63

29. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004

30. Li Q, Xia D, Wang Z, Liu B, Zhang J, Peng P, et al. Circadian rhythm gene PER3 negatively regulates stemness of prostate cancer stem cells *via* WNT/beta-catenin signaling in tumor microenvironment. *Front Cell Dev Biol* (2021) 9:656981. doi: 10.3389/fcell.2021.656981

31. Li S, Zhang N, Liu S, Zhang H, Liu J, Qi Y, et al. ITGA5 is a novel oncogenic biomarker and correlates with tumor immune microenvironment in gliomas. *Front Oncol* (2022) 12:844144. doi: 10.3389/fonc.2022.844144

32. Maeser D, Gruener RF, Huang RS. oncoPredict: an r package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* (2021) 22(6):bbab260. doi: 10.1093/bib/bbab260

33. Wolfl M, Schwinn S, Yoo YE, Ress ML, Braun M, Chopra M, et al. Src-kinase inhibitors sensitize human cells of myeloid origin to toll-like-receptor-induced interleukin 12 synthesis. *Blood* (2013) 122:1203–13. doi: 10.1182/blood-2013-03-488072

34. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers* (2020) 6:92. doi: 10.1038/s41572-020-00224-3

35. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer* (2018) 18:269–82. doi: 10.1038/nrc.2018.11

36. Li Q, Wang J, Meng X, Chen W, Feng J, Mao J. Identification of autophagy-related gene and lncRNA signatures in the prognosis of HNSCC. *Oral Dis* (2021). doi: 10.1111/odi.13889

37. Cramer JD, Burtness B, Le QT, Ferris RL. The changing therapeutic landscape of head and neck cancer. *Nat Rev Clin Oncol* (2019) 16:669–83. doi: 10.1038/s41571-019-0227-z

38. Miyauchi S, Kim SS, Pang J, Gold KA, Gutkind JS, Califano JA, et al. Immune modulation of head and neck squamous cell carcinoma and the tumor microenvironment by conventional therapeutics. *Clin Cancer Res* (2019) 25:4211–23. doi: 10.1158/1078-0432.CCR-18-0871

39. Moskovitz J, Moy J, Ferris RL. Immunotherapy for head and neck squamous cell carcinoma. *Curr Oncol Rep* (2018) 20:22. doi: 10.1007/s11912-018-0654-5

40. Garcia-Marin R, Reda S, Riobello C, Cabal VN, Suarez-Fernandez L, Vivanco B, et al. Prognostic and therapeutic implications of immune classification by CD8(+) tumor-infiltrating lymphocytes and PD-L1 expression in sinonasal squamous cell carcinoma. *Int J Mol Sci* (2021) 22(13):6926. doi: 10.3390/ijms22136926

41. Gibson MK, Li Y, Murphy B, Hussain MH, DeConti RC, Ensley J, et al. Randomized phase III evaluation of cisplatin plus fluorouracil versus cisplatin plus paclitaxel in advanced head and neck cancer (E1395): an intergroup trial of the Eastern cooperative oncology group. *J Clin Oncol* (2005) 23:3562–7. doi: 10.1200/JCO.2005.01.057

42. Vermorken JB, Mesia R, Rivera F, Remenar E, Kawecki A, Rottey S, et al. Platinum-based chemotherapy plus cetuximab in head and neck cancer. *N Engl J Med* (2008) 359:1116–27. doi: 10.1056/NEJMoa0802656

43. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science* (2018) 359:1350–5. doi: 10.1126/science.aar4060

44. Sharma P, Siddiqui BA, Anandhan S, Yadav SS, Subudhi SK, Gao J, et al. The next decade of immune checkpoint therapy. *Cancer Discovery* (2021) 11:838–57. doi: 10.1158/2159-8290.CD-20-1680

45. Burtness B, Harrington KJ, Greil R, Soulieres D, Tahara M, de Castro GJr., et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet* (2019) 394:1915–28. doi: 10.1016/S0140-6736(19)32591-7

46. Oliva M, Spreafico A, Taberna M, Alemany L, Coburn B, Mesia R, et al. Immune biomarkers of response to immune-checkpoint inhibitors in head and neck squamous cell carcinoma. *Ann Oncol* (2019) 30:57–67. doi: 10.1093/annonc/mdy507

47. Carlisle JW, Steuer CE, Owonikoko TK, Saba NF. An update on the immune landscape in lung and head and neck cancers. *CA Cancer J Clin* (2020) 70:505–17. doi: 10.3322/caac.21630

48. Solomon B, Young RJ, Rischin D. Head and neck squamous cell carcinoma: Genomics and emerging biomarkers for immunomodulatory cancer treatments. *Semin Cancer Biol* (2018) 52:228–40. doi: 10.1016/j.semcancer.2018.01.008

49. Jablonska E, Garley M, Surazynski A, Grubczak K, Iwaniuk A, Borys J, et al. Neutrophil extracellular traps (NETs) formation induced by TGF-beta in oral lichen planus - possible implications for the development of oral cancer. *Immunobiology* (2020) 225:151901. doi: 10.1016/j.imbio.2019.151901

50. Cedervall J, Hamidi A, Olsson AK. Platelets, NETs and cancer. *Thromb Res* (2018) 164 Suppl 1:S148–52. doi: 10.1016/j.thromres.2018.01.049

51. Masucci MT, Minopoli M, Del Vecchio S, Carriero MV. The emerging role of neutrophil extracellular traps (NETs) in tumor progression and metastasis. *Front Immunol* (2020) 11:1749. doi: 10.3389/fimmu.2020.01749

52. Li D, Shao J, Cao B, Zhao R, Li H, Gao W, et al. The significance of neutrophil extracellular traps in colorectal cancer and beyond: From bench to bedside. *Front Oncol* (2022) 12:848594. doi: 10.3389/fonc.2022.848594

53. Teijeira A, Garasa S, Ochoa MC, Villalba M, Olivera I, Cirella A, et al. IL8, neutrophils, and NETs in a collusion against cancer immunity and immunotherapy. *Clin Cancer Res* (2021) 27:2383–93. doi: 10.1158/1078-0432.CCR-20-1319

54. Li B, Liu Y, Hu T, Zhang Y, Zhang C, Li T, et al. Neutrophil extracellular traps enhance procoagulant activity in patients with oral squamous cell carcinoma. *J Cancer Res Clin Oncol* (2019) 145:1695–707. doi: 10.1007/s00432-019-02922-2

55. Chen N, He D, Cui J. A neutrophil extracellular traps signature predicts the clinical outcomes and immunotherapy response in head and neck squamous cell carcinoma. *Front Mol Biosci* (2022) 9:833771. doi: 10.3389/fmolb.2022.833771

56. Zhang R, Gao Y. Identification of NUTF2 as a candidate diagnostic and prognostic biomarker associated with immune infiltration in head and neck squamous cell carcinoma. *Onco Targets Ther* (2021) 14:5455–67. doi: 10.2147/OTT.S337469

57. Zhang T, Xi Y, Wu T, Liu J. Nuclear transporting factor 2 as a novel biomarker of head and neck squamous cell carcinoma and associated with T/B cell receptor signaling pathway. *BioMed Res Int* (2022) 2022:2885323. doi: 10.1155/2022/2885323

58. Li Y, Huang Y, Ren S, Xiao X, Cao H, He J. A pan-cancer analysis of the oncogenic role of nuclear transport factor 2 in human cancers. *Front Oncol* (2022) 12:829389. doi: 10.3389/fonc.2022.829389

59. Du P, Chai Y, Zong S, Yue J, Xiao H. Identification of a prognostic model based on fatty acid metabolism-related genes of head and neck squamous cell carcinoma. *Front Genet* (2022) 13:888764. doi: 10.3389/fgene.2022.888764

60. Jiang Y, Cao W, Wu K, Qin X, Wang X, Li Y, et al. LncRNA LINC00460 promotes EMT in head and neck squamous cell carcinoma by facilitating peroxiredoxin-1 into the nucleus. *J Exp Clin Cancer Res* (2019) 38:365. doi: 10.1186/s13046-019-1364-z

61. Shan Y, He R, Yang X, Zang S, Yao S, Gao M, et al. An autophagy-related lncRNA prognostic risk model for thyroid cancer. *Eur Arch Otorhinolaryngol* (2022) 279:1621–31. doi: 10.1007/s00405-021-07134-4

62. Tan J, Liu L, Zuo Z, Song B, Cai T, Ding D, et al. Overexpression of novel long intergenic noncoding RNA LINC02454 is associated with a poor prognosis in papillary thyroid cancer. *Oncol Rep* (2020) 44:1489–501. doi: 10.3892/or.2020.7712

63. Pan WA, Tsai HY, Wang SC, Hsiao M, Wu PY, Tsai MD. The RNA recognition motif of NIFK is required for rRNA maturation during cell cycle progression. *RNA Biol* (2015) 12:255–67. doi: 10.1080/15476286.2015.1017221

64. Lin TC, Su CY, Wu PY, Lai TC, Pan WA, Jan YH, et al. The nucleolar protein NIFK promotes cancer progression *via* CK1alpha/beta-catenin in metastasis and ki-67-dependent cell proliferation. *Elife* (2016) 5:e11288. doi: 10.7554/eLife.11288

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 27 September 2022
DOI 10.3389/fimmu.2022.1022149

Check for updates

OPEN ACCESS

EDITED BY
Nan Zhang,
Harbin Medical University, China

REVIEWED BY
Kui Zhang,
The University of Chicago,
United States
Kewa Gao,
University of California, Davis,
United States

*CORRESPONDENCE
Huan Xu
joyxu1027@163.com
Zhixin Li
saklizhixin@163.com

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 18 August 2022
ACCEPTED 12 September 2022
PUBLISHED 27 September 2022

CITATION
Gu X, Wei S, Li Z and Xu H (2022) Machine learning reveals two heterogeneous subtypes to assist immune therapy based on lipid metabolism in lung adenocarcinoma. *Front. Immunol.* 13:1022149. doi: 10.3389/fimmu.2022.1022149

COPYRIGHT
© 2022 Gu, Wei, Li 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.

Machine learning reveals two heterogeneous subtypes to assist immune therapy based on lipid metabolism in lung adenocarcinoma

Xuyu Gu[1], Shiyou Wei[2], Zhixin Li[3]* and Huan Xu[2]*

[1]School of Medicine, Southeast University, Nanjing, China, [2]Department of Anesthesiology, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China, [3]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China

Background: Lipid metabolism pivotally contributes to the incidence and development of lung adenocarcinoma (LUAD). The interaction of lipid metabolism and tumor microenvironment (TME) has become a new research direction.

Methods: Using the 1107 LUAD records from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, a comprehensive exploration was performed on the heterogeneous lipid metabolism subtypes based on lipid metabolism genes (LMGs) and immune-related genes (LRGs). The clinical significance, functional status, TME interaction and genomic changes of different subtypes were further studied. A new scoring system, lipid-immune score (LIS), was developed and validated.

Results: Two heterogeneous subtypes, which express more LMGs and show the characteristics of tumor metabolism and proliferation, are defined as lipid metabolism phenotypes. The prognosis of lipid metabolism phenotype is poor, and it is more common in patients with tumor progression. Expressing more IRGs, enrichment of immunoactive pathways and infiltration of effector immune cells are defined as immunoactive phenotypes. The immunoactive phenotype has a better prognosis and stronger anti-tumor immunity and is more sensitive to immunotherapy. In addition, KEAP1 is a driving mutant gene in the lipid metabolism subtype. Finally, LIS was developed and confirmed to be a robust predictor of overall survival (OS) and immunotherapy in LUAD patients.

Conclusion: Two heterogeneous subtypes of LUAD (lipid metabolism subtype and immune activity subtype) were identified to evaluate prognosis and immunotherapy sensitivity. Our research promotes the understanding of the interaction between lipid metabolism and TME and offers a novel direction for clinical management and precision therapy aimed to LUAD patients.

KEYWORDS
lung adenocarcinoma, machine learning, tumor microenvironment, lipid metabolism, immunotherapy

Introduction

As the most frequent malignancy, lung cancer causes the highest cancer-related deaths around the world (1). Lung cancer can appear in different histological types among which, non-small cell lung cancer (NSCLC) is the most common type with about 85% proportion of all lung cancer patients (2). Lung adenocarcinoma (LUAD) is the most abundant subtype of NSCLC, accounting for about 55% (3). LUAD is a heterogeneous disease with different clinical prognosis and drug response results. It is worth noting that despite the great progress in clinical diagnostic methods and multimodal treating approaches, the 5-year overall survival (OS) rate of patients with advanced lung cancer has remained very low (4). Therefore, LUAD patients are still in an urgent need for new early diagnosis and clinical intervention methods.

During cancer occurrence and progression, the immune system and the tumor cells are in complex interaction. On one hand, demand for local nutrition and oxygen highly increases due to fast proliferation of tumor cells. On the other hand, the same reason causes poor local vascularization, resulting in acidosis and hypoxia in the tumor microenvironment (TME) as well as local glucose deficient (5–7). Eventually, lipids existing in the TME begin to be used main as the alternative source of energy in both tumor tissues and immune cells to compensate for the energy shortages (8). Lipids also contribute to the biofilm formation, supplying the biomass production, and mediating some complex signaling pathways contributing to the growth and migration of cancer cells (9). In addition, the affinity of cancerous cells for lipids and cholesterol increases, directly leading to lipid accumulation in the TME and developing malignancy in the tumor tissues (10). Although, the lipid metabolic reprogramming and dysfunction as well as its dual impact in the TME and immune responses to tumor is not exactly recognized yet. Such further elaboration is essentially required for developing specific treatments based on the anti-tumor immune responses.

The present study aimed to survey the crosstalk between lipid metabolism and tumor immune response in LUAD patients and identified two heterogeneous subtypes (lipid metabolism subtype and immune activity subtype). These two subtypes show specific differences in clinical outcomes, biological functions, immune infiltration and genomic variation. In addition, a lipid-immune score (LIS) was developed and validated, which shows significant advantages in predicting prognosis and immunotherapy response. In conclusion, our work strengthens the understanding of the complex role between lipid metabolism and immune system in LUAD and provides a new perspective and reference for the accurate prediction and immunotherapy of LUAD patients.

Methods

Data extraction

Transcriptome RNA-seq data (HT-seq FPKM), mutation data (mutect2 tool), copy number variation (CNV), and their corresponding clinical information (from Cancer Genome Atlas-lung adenocarcinoma, tcga-LUAD queue) were obtained (https://portal.gdc.cancer.gov/repository). After excluding patients who lost follow-up and clinical information, 492 LUAD samples were collected. These data were used as discovery queues after Transcripts per million (TPM) standardization. In addition, three independent data sets from the GEO database were collected, including GSE30219 from GPL570 platform, GSE42127 from GPL6884 platform, and GSE72094 from GPL15048 platform. In order to prevent the batch effect of chips, three GEO data sets were combined through the combat function of the "sva" package and the data were log2 standardized (11). Finally, a total of 615 GEO meta queues containing LUAD samples with complete clinical information were used as external validation queues. Finally, two immunotherapy cohorts were collected to verify the model's prognostic power: NSCLC cohort GSE135222 receiving Programmed Death-1(PD-1) treatment, including 27 patients (12) and Imvigor210, a cohort of advanced urothelial carcinoma cases undergoing anti-Programmed Cell Death-Ligand 1 (PD-L1) immunotherapy, including 298 patients (13).

Identification of lipid and immune subtypes of LUAD

Lipid metabolism genes (LMGs) were obtained from the Molecular Signatures Database (MSigDB) (http://www.gsea-msigdb.org/gsea/index.jsp), containing 1426 LMGs (14). The immune-related genes (IRGs) were obtained from the "ImmPort" database (https://www.immport.org/resources) (15). It contains a total of 1638 IRGs defined as functional and

immune related. A detailed list in Table S1 to indicate the lipid genes and immune genes we used. First, LMGs and IRGs with independent prognostic efficacy were evaluated by univariate Cox regression analysis, and candidate genes were identified according to the threshold of $p < 0.05$. According to the transcriptional map of candidate genes, consensus clustering was conducted in the discovery queue and validation queue through the ConsensusClusterPlus package (16). Pam unsupervised clustering algorithm was adopted in this analysis, and 1000 iterations were carried out based on Euclidean distance. Eighty per cent of the samples were randomly selected in each iteration. The number of clusters was set to 2-5, and the optimal cluster number was jointly determined using the consensus matrix and cumulative distribution function (CDF).

Functional enrichment and immune infiltration analysis

Significant differentially expressed genes (DEGs) between subgroups were identified by 'limma' package in R program according to the threshold of False Discovery Rate (FDR) < 0.05 and fold change (FC) > 2. The functional enrichment of DEGs was achieved using metascape (www.metascape.org/) database. Gene Set Enrichment Analysis (GSEA) was conducted among subgroups and significantly altered pathways were selected using Kyoto Encyclopedia of Genes and Genomes (KEGG) by $p < 0.05$.

Based on the previously published molecular markers, ssGSEA analysis was performed using the 'gsva' package in R program to evaluate the biological pathway activity of the samples which included angiogenesis, epithelial-mesenchymal transition (EMT), myoid inflammation, and molecular markers of other immune related pathways (17–20). Molecular markers of hypoxia were collected from MSigDB (14). Detailed pathway gene markers were displayed in Table S2. The relative infiltration abundance of 22 different immune cell types was evaluated by 'cibersoft' package in R program (21). The immune activity and tumor purity of tumor samples were evaluated by Estimate algorithm (22). Finally, the microsatellite instability (MSI) score, indel neoantigens and SNV neoantigens of samples from the study of Thorsson et al. (23).

Analysis of the genome variation map between subgroups

The mutation data was processed with 'maftools' package in R package. First, the total number of mutations in the sample was measured, and then, the genes with the minimum mutation number > 30 were identified. The difference of mutation frequency of high-frequency mutation genes between the two subgroups was compared using the chi square test and visualized with maftools (24). CNV data were processed by Gistic 2.0 software. Based on the threshold of 0.2, significantly amplified and deleted chromosome segments were identified, and CNV differences on chromosome arms were evaluated. The CNV results were visualized by 'ggplot2' R package.

Constructing lipid-immune score

DEGs contained in all cohorts were selected for further analysis based on the above identified DEGs between the two subtypes. Univariate Cox regression analysis revealed the prognostic value of these genes. Subsequently, genes with statistical significance ($p < 0.05$) were incorporated into the Cox proportional hazard model with Least absolute shrinkage and selection operator (Lasso) penalty, and 300 iterative searches were carried out to find the most robust model. In order to prevent over fitting, five-cross validation was set up. The model with the highest frequency in 300 iterations was used as the final prognostic model and the lipid-immune score (LIS) was generated according to the formula: $LIS = \sum iCoefficient(mRNA_i) \times Expression(mRNA_i)$. The 'survcomp' package in R program was used to calculate the consistency C index and evaluate the prognostic value of the risk score (RS) in the training and verification sets. The higher C index indicates the more accurate prognostic power of the model (25). The high-risk group and low-risk group were divided based on their median FRS, and the prognostic value of the risk model was calculated using Kaplan-Meier (KM) survival curve, univariate and multivariate Cox regression, time-dependent ROC curve (tROC), and subgroup analysis system.

Predicting immunotherapy response

According to previous studies, the immunophenoscore (IPS) of the sample was calculated. Briefly, IPS is calculated from transcriptomic data of representative genes for different immunophenotypes and normalizes the final result to 0-10. Samples were positively weighted according to effective immune cells and negatively weighted according to suppressive immune cells, and then applied Z-score averaged. Z-score ≥ 3 was defined as IPS10 and Z-score ≤ 0 was defined as IPS0. The higher the IPS, the better the immunotherapy response (26). The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (http://tide.dfci.harvard.edu) was applied to predict the patients' response to the anti-PD-1 and anti-CTLA-4 treatments (27–30). Finally, the predictive power of LIS was evaluated in two external immunotherapy cohorts (GSE135222 and Imvigor210).

Statistical analysis

Pearson chi square or Fisher exact tests were applied to compare categorical variables. The continuous variables were compared between the two groups by Wilcoxon rank sum test. The KM curve was drawn by 'survminer' package and the tROC analysis was carried out by 'survivalROC' package both in R program. The univariate and multivariate Cox regression was completed by 'survival' package in R program. The 'rms' package in R was used to draw nomograms and calibration curves, and decision curve analysis (DCA) was carried out through DCA package (31). The ROC curve used to predict immunotherapy was performed by the 'pROC' package. Two tailed $p < 0.05$ was considered statistically significant unless otherwise specified.

Results

Parsing LMGs and IRGs in LUAD

The design of our study is shown in Figure S1. Univariate Cox regression analysis displayed 155 LMGs and IRGs with prognostic value ($p < 0.05$). The forest map showed the prognostic candidate genes of top15 (Figure 1A). Detailed Cox results are provided in Table S3. Figure 1B summarizes the mutation of top15 candidate genes. Specifically, the mutation type is single nucleotide mutation, and the genes with the highest mutation frequency are VEGFC (24%) and tnfrsf11a (10%). The waterfall diagram shows their mutation map in the tcga-LUAD cohort (Figure 1C). The histogram summarizes the CNV of the top15 candidate genes in tcga-LUAD, and the results show that

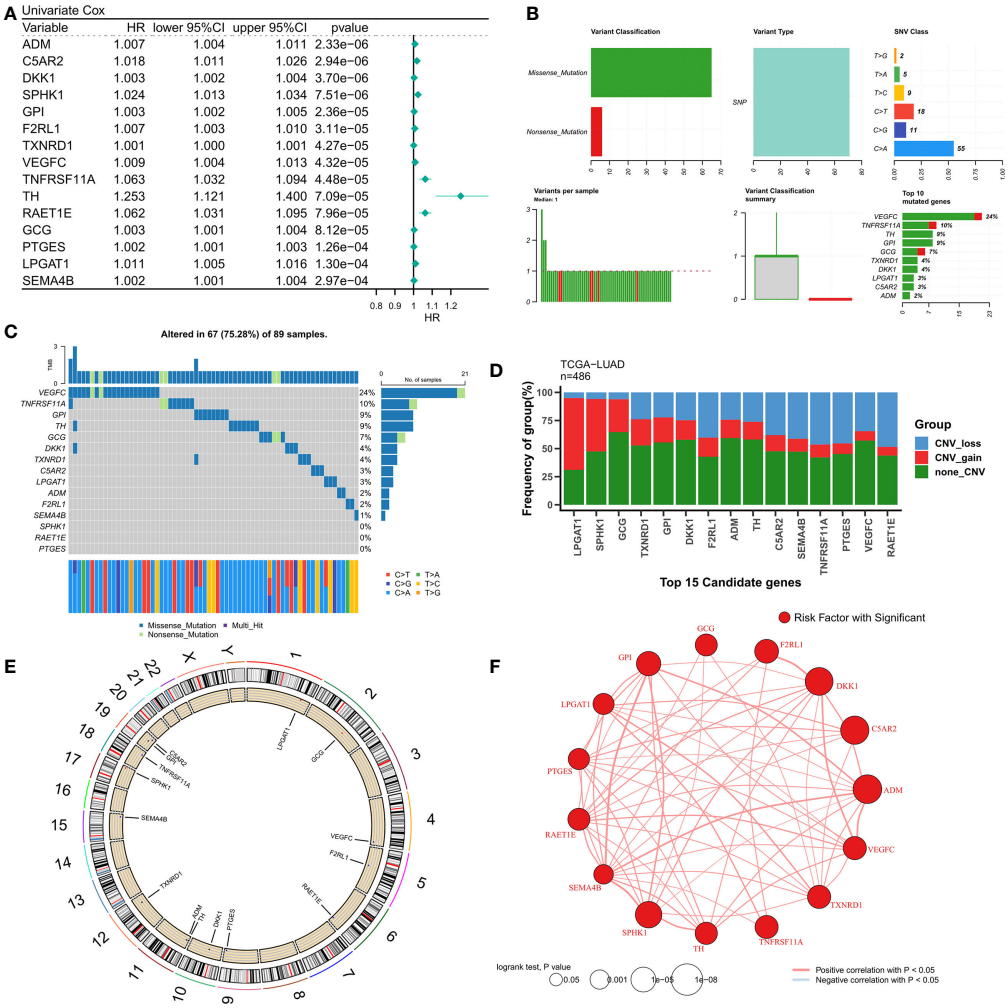


FIGURE 1
Genomic map of LMGs in LUAD. **(A)** univariate Cox regression analysis revealed the prognostic LMGs of Top15. **(B)** Summary of LMGs mutation events in tcga-LUAD. **(C)** Oncoplot showed the mutation map of LMGs. **(D)** Summary of CNV events of LMGs in tcga-LUAD. **(E)** The circle diagram shows the overall CNV characteristics of LMGs on chromosomes. **(F)** Correlation network of LMGs.

they have a wide range of CNV events. Lpgat1 was the most amplified gene, and raet1e was the most deleted gene (Figure 1D). The circle chart shows the overall CNV of the top15 candidate gene on the chromosome (Figure 1E). Finally, the interaction of top15 candidate genes was analyzed, and the correlation network showed that they were highly positively correlated (Figure 1F).

Identification of lipid and immune subtypes

Consensus clustering was performed on the discovery queue and GEO meta queue from tcga-LUAD by ConsensusClusterPlus. According to the CDF curve of consensus score, k = 2 was found to be the best choice (Figure 2A, Figure S2A). The consensus matrix also confirmed this result (Figure 2B, Figure S2B). Based on the transcriptional profiles of candidate LMGs and IRGs, lipid metabolism subtypes and immune activity subtypes were defined (Figure 2C, Figure S2C). IRGs were significantly increased in immunoactive subtypes, while LMGs were significantly increased in lipid metabolism subtypes. According to the survival analysis, lipid metabolism subtypes in the cohort was significantly worse compared to that of immune activity subtypes ($p = 0.001$, Figure 2D). A worse clinical outcome of lipid metabolism subtypes was confirmed in the validation cohort ($p < 0.001$, Figure S2D). In addition, the tcga-LUAD cohort had more detailed clinical follow-up information. There was a significant increase in patients with disease progression in the lipid metabolism subtype (Figures 2E, F).

Biological function difference between two subtypes

First, the DEGs between the two subtypes were identified by limma package. According to the threshold of FDR < 0.05 and FC > 2, a total of 1597 DEGs were identified, of which

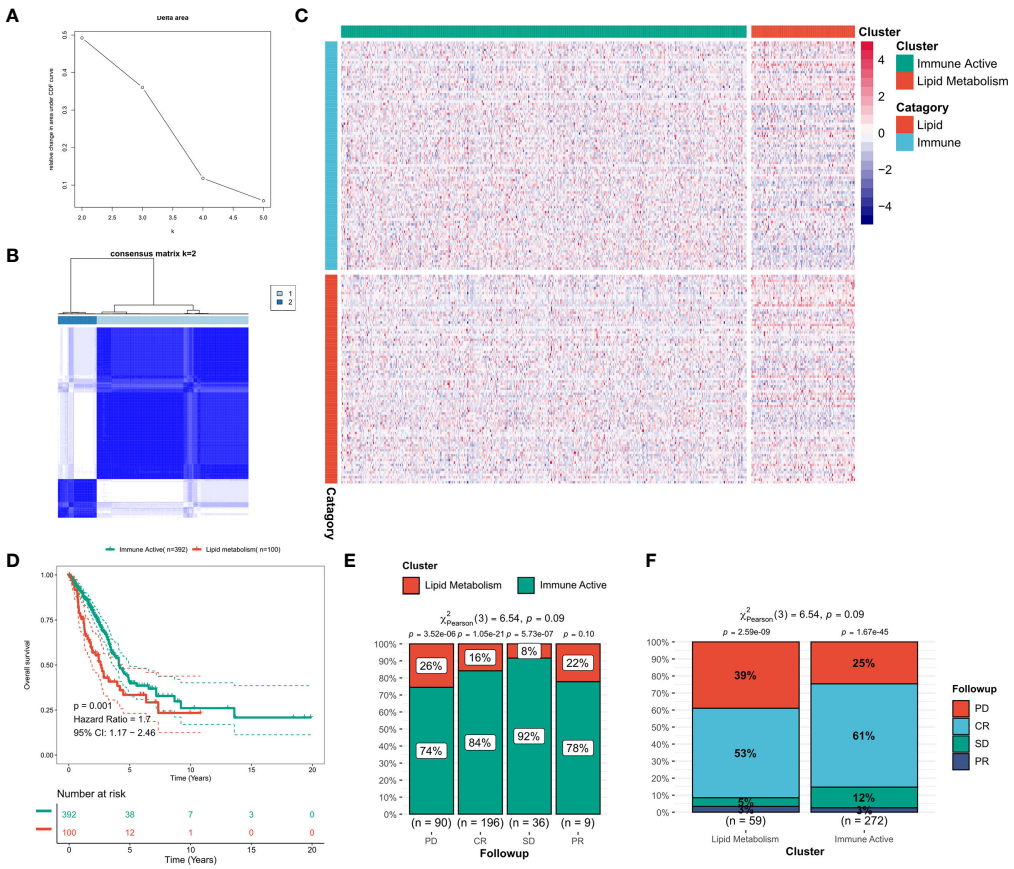


FIGURE 2
Identification of lipid and immune subtypes. **(A)** CDF curve of consensus matrix of different K **(B)** Consensus matrix when k = 2. **(C)** Expression heat maps of LMGs and IRGs in the two subtypes. **(D)** Survival curve of two subtypes in TCGA cohort. **(E)** The proportion of subtypes in LUAD patients with different outcomes. **(F)** Clinical outcomes of patients with different subtypes of LUAD.

1233 were up-regulated in immunoactive subtypes and 362 were up-regulated in lipid metabolism subtypes. Detailed results are provided in Table S4. Based on the functional enrichment analysis, the up-regulated genes in immunoactive subtypes mainly regulate cell activation, inflammatory response, cell adhesion and lymphocyte migration (Figure 3A), and Figure 3C showed the functional interaction network of immunoactive subtypes. The up-regulated genes in lipid metabolism subtypes mainly regulate biological oxidation, epithelial cell differentiation and glucose homeostasis (Figure 3B). Figure 3D shows the functional interaction network of lipid metabolism subtypes. GSEA analysis showed that the pathways enriched in immunoactive subtypes were mainly B-cells receptor, T-cells receptor, Toll-like receptor signal pathway and NK-cells killing activity (Figure 3E). The pathways enriched in lipid metabolism subtypes were fatty acid metabolism, protein secretion and TCA cycle pathway (Figure 3F). In conclusion, these results confirm that the immunocompetent subtype has stronger antitumor immune activity, while the tumor cells of the lipid metabolism subtype have stronger metabolic and proliferative activity, which may lead to the difference in prognosis between the two.

Difference of immune infiltration between two subtypes

The immune infiltration degrees were systematically compared between the two subtypes. First, the estimate algorithm showed that the immune activity subtype had a higher immune score, while the lipid metabolism subtype had a higher tumor purity (Figure 4A), which was confirmed in the validation queue (Figure S3A). The expression differences between five classical immune checkpoints and therapeutic targets (PD-L1, CD8A, CTLA-4, LAG-3, PD-1) were then examined. The results showed that the five checkpoints were significantly up-regulated in immunoactive subtypes (Figure 4B), and the validation cohort (Figure S3B). Through ssGSEA algorithm, we found that except for myeloid inflammation, other immune pathways were up-regulated in the immune activity subgroups. In particular, the activity of EMT pathway in the immunoactive pathway was also up-regulated (Figure 4C). Similar results were observed in the validation cohort. It is worth noting that the activity of angiogenesis pathway in lipid metabolism subtypes was up-regulated in the validation cohort (Figure S3C). Finally, cibersort results showed that NK cells, plasma cells and natural B cells increased in immunoactive subtypes, while Tregs increased

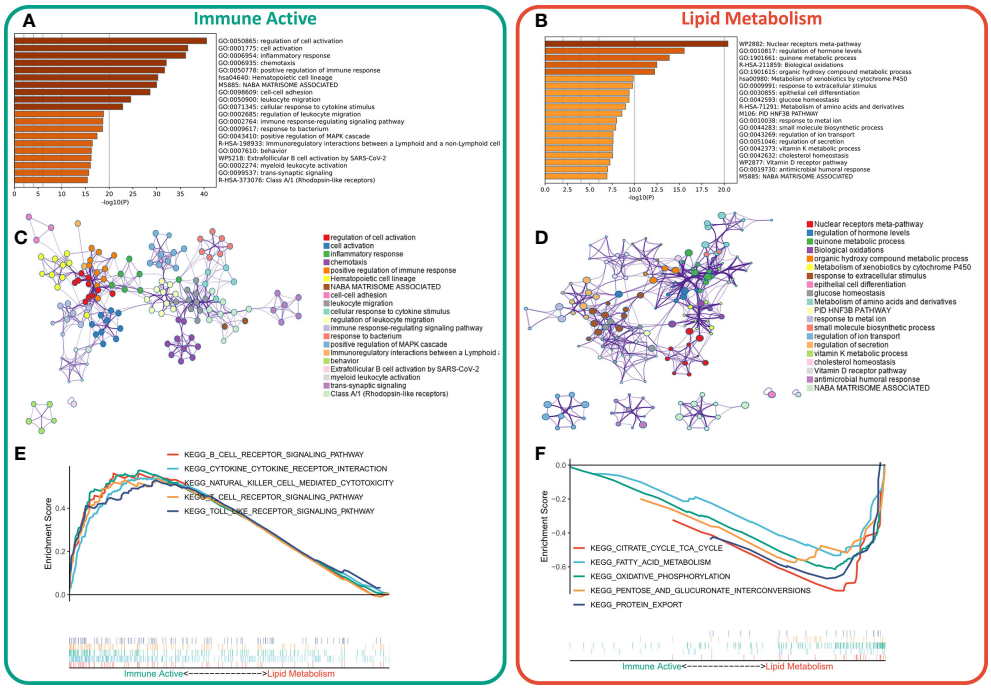


FIGURE 3
Functional enrichment of different subtypes. **(A)** Functional enrichment of characteristic genes of immune activity subtypes. **(B)** Functional enrichment of lipid metabolism subtype characteristic genes. **(C)** Functional network of immunoactive subtypes. **(D)** Functional network of lipid metabolism subtypes. **(E)** KEGG pathway enriched in immunocompetent subtypes. **(F)** KEGG pathway enriched in lipid metabolism subtypes.

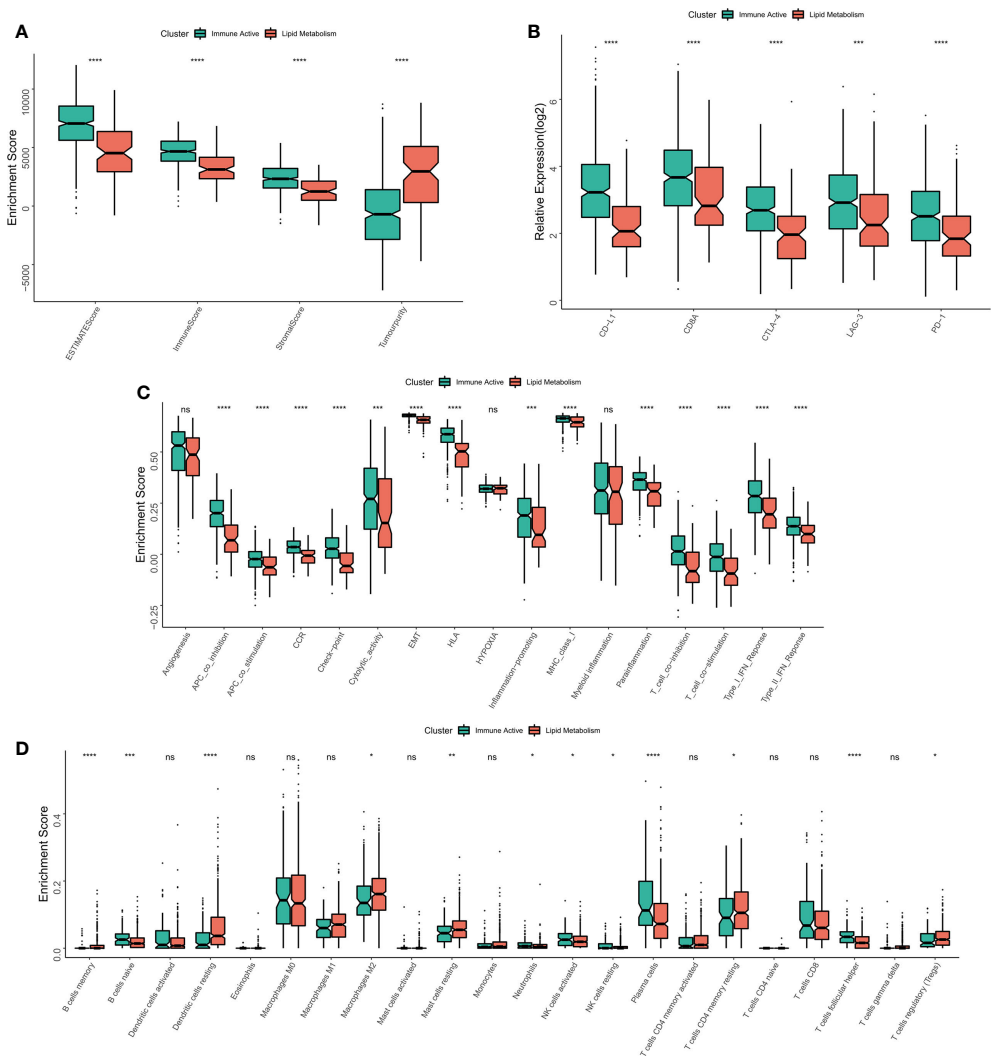


FIGURE 4
Immune infiltration analysis of different subtypes. **(A)** The difference of Estimate score between the two subtypes. **(B)** Differences in the expression of six typical immune checkpoints (PD-L1, CD8A, CTLA-4, LAG-3, PD-1) between the two subtypes. **(C)** Differences in the activity of immune related pathways between the two subtypes. **(D)** The difference of immune cell infiltration between the two subtypes. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; ns p > 0.05.

in lipid metabolism subtypes (Figure 4D). In addition, higher Tregs in lipid metabolism subtypes were also confirmed in the validation cohort (Figure S3D). In conclusion, these results convey that the immunoactive subtype has more antitumor immune activity and effector immune cells, while the lipid metabolism subtype is inhibited by higher Tregs infiltration.

Analysis of genome changes among subtypes

The original mutation data were processed with maftools package. Chi square test showed that the mutation frequency of KEAP1, KRAS and SPTA1 in lipid metabolism subtypes increased, especially KEAP1 (Figure 5A). The waterfall diagram shows the mutation map difference of a total of 32 high-frequency mutant genes between the two subtypes (Figure 5B). The TMB of each patient was calculated, and the results showed that the lipid metabolism subtype had a higher TMB, but the difference between the two subtypes was not significant (Figure 5C). CNV leads to chromosome variation in another way. We then evaluated the correlation between FRS and CNV and found that the amplification and deletion levels of immunoactive subtypes were significantly higher at the chromosome arm level (Figure 5D). The box diagram showed no significant difference in the total number of chromosome

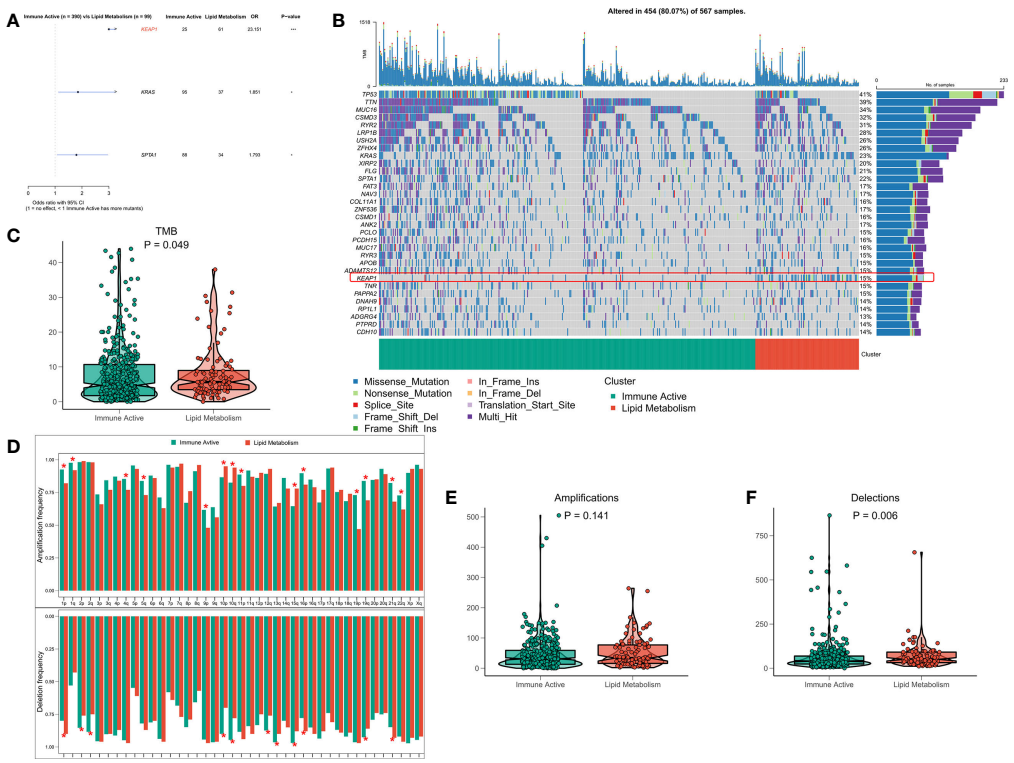


FIGURE 5
Genome driven events of different subtypes. **(A)** The forest map shows the genes with significant mutation differences between the two subtypes. **(B)** Oncoplot showed the mutation landscape of top25 mutation driver gene among subtypes. **(C)** The difference of TMB between the two subtypes. **(D)** The histogram analyzed the CNV events on the chromosome arm among subtypes. **(E)** Differences in overall amplification events among subtypes. **(F)** Differences in overall missing events among subtypes. *p < 0.05; ***p < 0.001;

amplification between the two subtypes (Figure 5E), and the number of chromosome deletions in the lipid metabolism subtype increased significantly (Figure 5F).

Immunoactive subtypes that are more sensitive to immunotherapy

The functional differences and immune landscape among subgroups suggest that patients with immunoactive subtypes may have better immune treatment response. According to the literature, better immunotherapeutic efficacy is in close relation to the increase in the number of neoantigens (32, 33). Therefore, we first evaluated the difference in the number of neoantigens between the two subtypes, and the results showed that the immunoactive subtypes have more SNV neoantigens and Indel neoantigens (Figures 6A, B). Recent studies have shown that MSI score is expected to become a new predictor of immunotherapy (34). However, there is no significant difference in MSI score between the two subtypes (Figure 6C). IPS can systematically evaluate the activity of effector immune cells and immune treatment response of patients. The discovery queue showed that IPS of immunoactive subtypes was significantly higher than that of lipid metabolism subtypes (Figure 6D), and the response rate of immunoactive subtypes to immunotherapy predicted by TIDE algorithm was higher than that of lipid metabolism subtypes (Figure 6E). Although there was no significant difference in IPS between the two subtypes in the validation cohort, the immunoactive subtypes in the validation cohort also had a higher response to immunotherapy (Figures 6E, F). In conclusion, our results suggest that immunoactive subtypes are more sensitive to immunotherapy.

Constructing and validating LIS

First, 1597 DEGs were analyzed by univariate Cox regression to identify the prognostic valuable DEGs. According to the threshold of $p < 0.001$, a total of 88 DEGs with prognostic significance were identified. Then, these 88 DEGs were recruited for Lasso regression to simplify the model. After 300 iterations, the model with 22 DEGs was the most stable showing a suitable efficacy in the training queue as well as the validation queue (C index > 0.6, Figure 7A). According to the best λ (0.02631), the

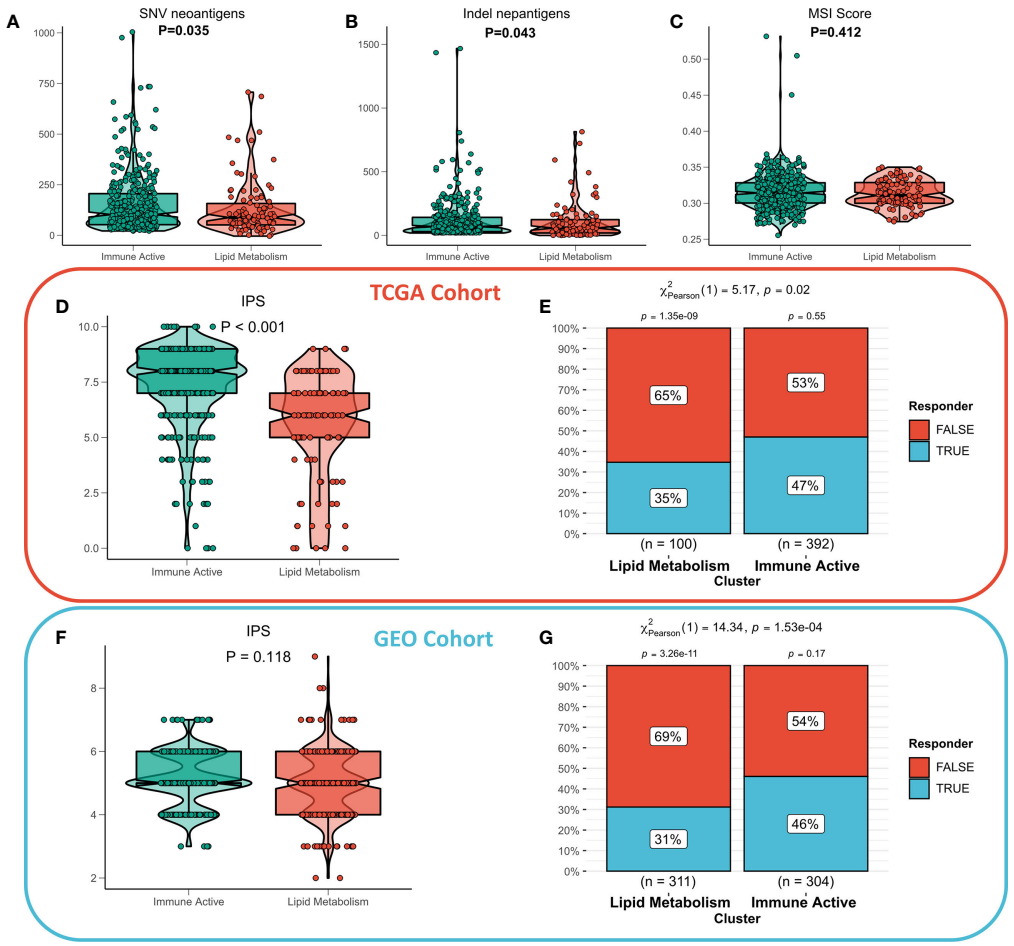


FIGURE 6
Immunotherapeutic sensitivity of different subtypes. **(A)** Differences in SNV neoantigens between the two subtypes. **(B)** The difference of indel neoantigens between the two subtypes. **(C)** The difference of MSI score between the two subtypes. **(D)** IPS differences between subtypes in TCGA cohort and **(F)** GEO cohort. The TIDE algorithm predicts the immune treatment responses of different subtypes in **(E)** TCGA queue and **(G)** GEO queue.

best model of 22 genes was identified (Figure 7B), and LIS was generated according to the formula: $LIS = \sum iCoefficient(mRNA_i) \times Expression(mRNA_i)$, detailed coefficients of 22 LIS genes can be found in Table S5. According to the survival analysis, patients with high LIS showed significantly less survival rate compared to the patients with low LIS ($p < 0.001$, Figure 7C), which was confirmed in the validation cohort ($p < 0.001$, Figure S4A). Based on the ROC analysis, the AUC values of the model in 1 year, 3 years, and 5 years were 0.792, 0.714, and 0.711, respectively (Figure 7D). In the external validation queue, LIS also had satisfactory prediction efficiency, specifically, 0.68 in 1 year, 0.69 in 3 years, 0.69 in 5 years and 0.71 in 8 years (Figure S4B). Figure 7E shows that the survival status of patients with high LIS were significantly worse compared that of patients with low LIS, and similar results were observed in the validation cohort (Figure S4C). TROC analysis showed that LIS was the best predictor of OS (Figure 7F), and the effectiveness of LIS and Stage was equivalent in the validation queue (Figure S4D). Finally, univariate Cox regression confirmed that LIS was an independent prognostic indicator in both training and validation sets ($p < 0.0001$, Figure 7G). Multivariate Cox regression showed that LIS was still an independent prognosticator for OS in the training and validation cohorts after correcting for other factors ($p < 0.0001$, Figure 7H).

Quantifying the risk of individual LUAD patients

Subgroup analysis showed that LIS in the training cohort showed excellent predictive ability in different clinical subgroups except patients in stage 3 and stage 4 ($p < 0.001$, Figure 8A). In the

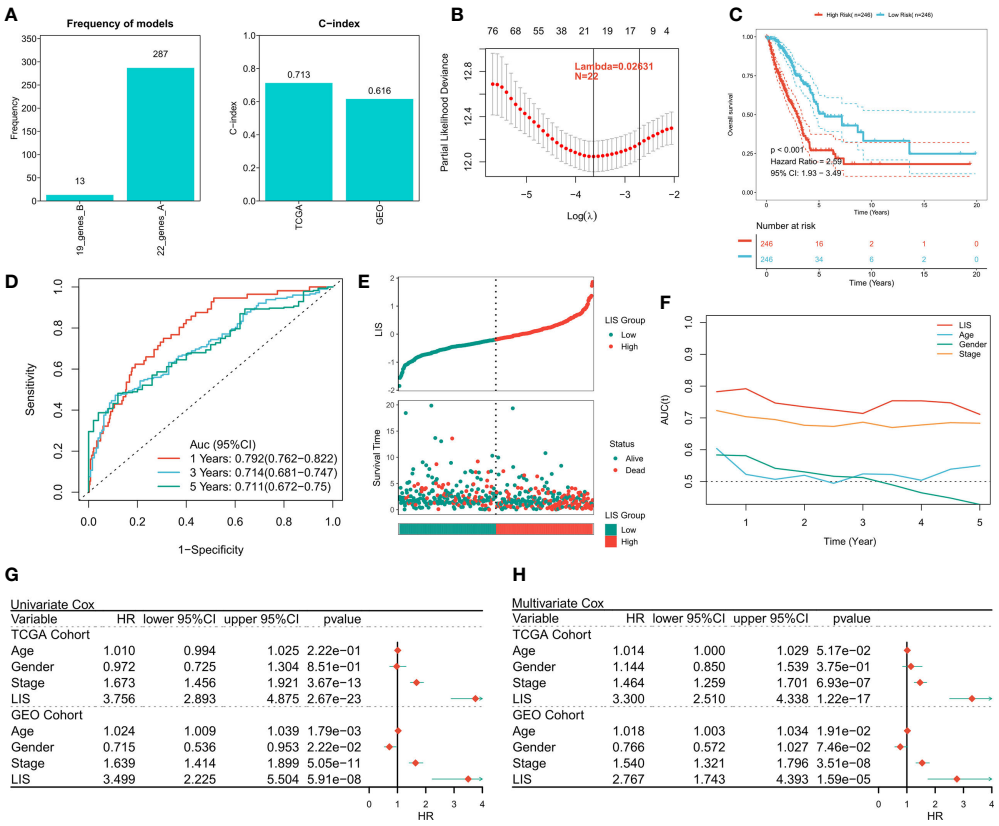


FIGURE 7
Building and verifying LIS. **(A)** Identifying the best Lasso model. Left: the frequency of different gene combinations in Lasso Cox regression model; Right: the best model is C-index in TCGA and GEO queues. **(B)** According to the best λ = 0.02631, the best 22-genes model was selected. **(C)** Survival curve of high and low LIS subgroups. **(D)** ROC analysis of LIS in TCGA queue. **(E)** The scatter plot shows the survival status of patients with different LIS in the TCGA cohort. **(F)** TROC curve of LIS in TCGA queue. **(G)** Univariate Cox regression analysis of OS in TCGA and GEO cohorts. **(H)** Multivariate Cox regression analysis of OS in TCGA and GEO cohorts.

validation cohort, LIS was able to distinguish patients with poor survival except for patients in stage 2-4 ($p < 0.05$, Figure S4E). These results suggest that LIS shows better performance in predicting early LUAD patients. For better quantifying of the death risk in individual LUAD patients, nomograms were constructed based on LIS (Figure 8B). Nomogram correction curve shows that nomogram model has good stability and accuracy in 1, 3 and 5 years (Figure 8C). TROC analysis showed that compared with clinical characteristics, nomogram model was the best predictor (Figure 8D). DCA was then performed to calculate the decision-making benefits of nomogram model. The results showed that nomogram was suitable for risk assessment of LUAD patients in 1, 3 and 5 years (Figure 8E).

LIS in predicting immunotherapy

First, TIDE was used to evaluate the difference of immunotherapy response between patients with high LIS and patients with low LIS. According to the results, patients with low LIS showed to be more benefitting from immunotherapy (Figure 9A, Figure S5A). Then five widely used immunotherapy biomarkers were calculated, including MDSC, MSI score, IFNG, CD8 and CD274. In the training cohort and validation cohort, LIS provided higher accuracy in predicting immunotherapy (Figure 9B, Figure S5B). Then, two immunotherapy cohorts were included to further study whether LIS could predict patients' response to immunotherapy. Consistent with the above, patients with high LIS showed worse survival in these two immunotherapy cohorts (Figures 9C, D). Finally, the relationship between LIS and neoantigens and TMB in Imvigor210 cohort was evaluated. The results showed that LIS had no strong correlation with neoantigens and TMB. However, patients with low LIS had higher neoantigens (Figures 9E, F). Overall, our study strongly confirms LIS as a prognosis factor for OS and immunotherapeutic response of patients, and is superior to widely used biomarkers.

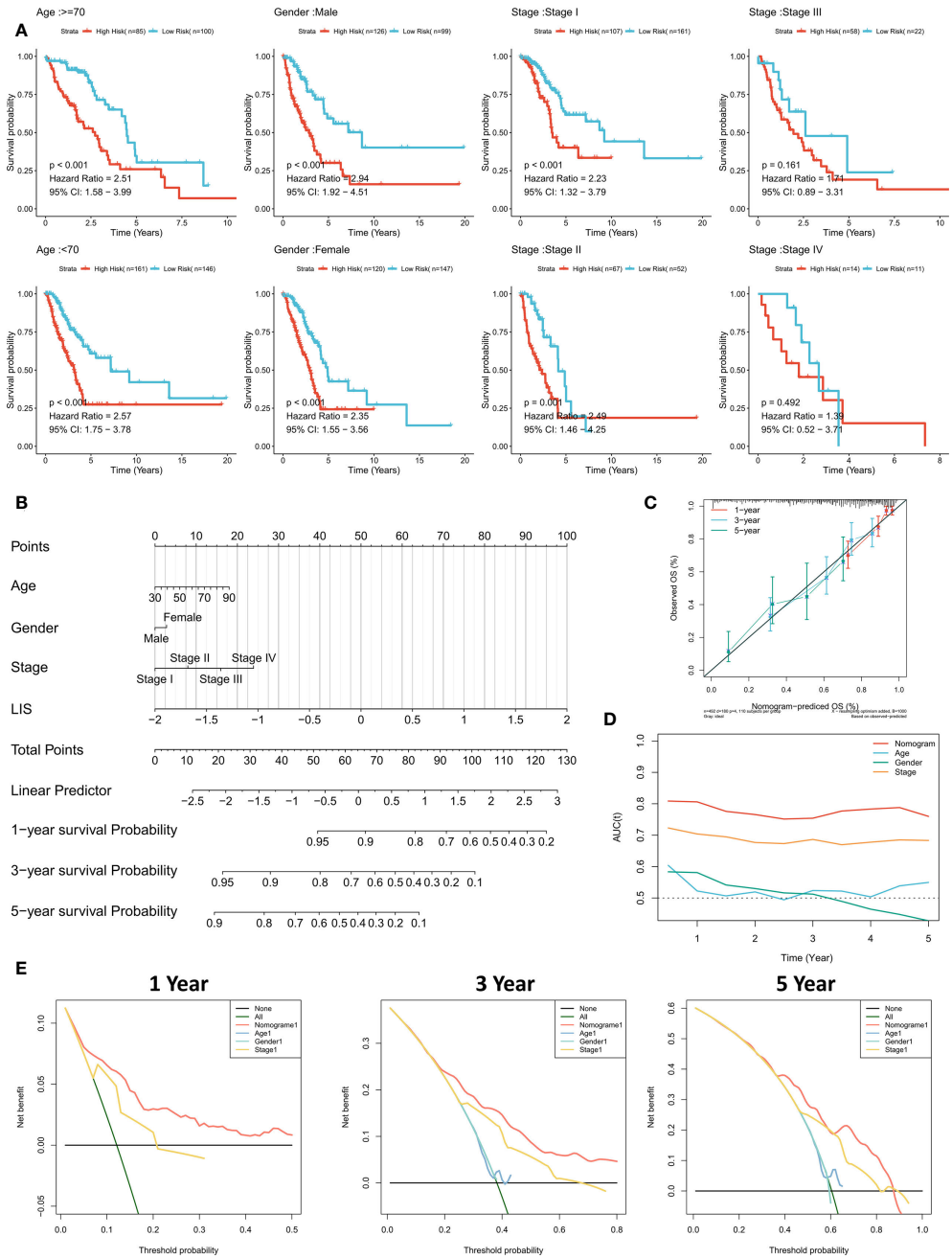


FIGURE 8
Quantifying individual LIS related risks. **(A)** Subgroup analysis of LIS in patients with different clinical characteristics. **(B)** LIS based nomograms were used to quantify individual patient risks. **(C)** Nomogram calibration curves at 1, 3 and 5 years. **(D)** TROC curve of nomogram. **(E)** Nomogram DCA curve in 1, 3 and 5 years.

Discussion

Lung cancer is the main cause of cancer-related death and LUAD is the most common histological subtype with the most patients at the advanced stage on the initial diagnosis (35, 36). Although a variety of targeted therapies and new chemotherapeutic drugs have been approved, the OS of advanced patients is still not ideal (4). Lipid metabolism has long been reported as the main energy source of cancer cells and is involved in the incidence and development of cancer (8). Recently, the dual regulation of lipid metabolism on immune response in TME has attracted extensive attention and has become a promising target for targeted therapy (6, 10). Our study

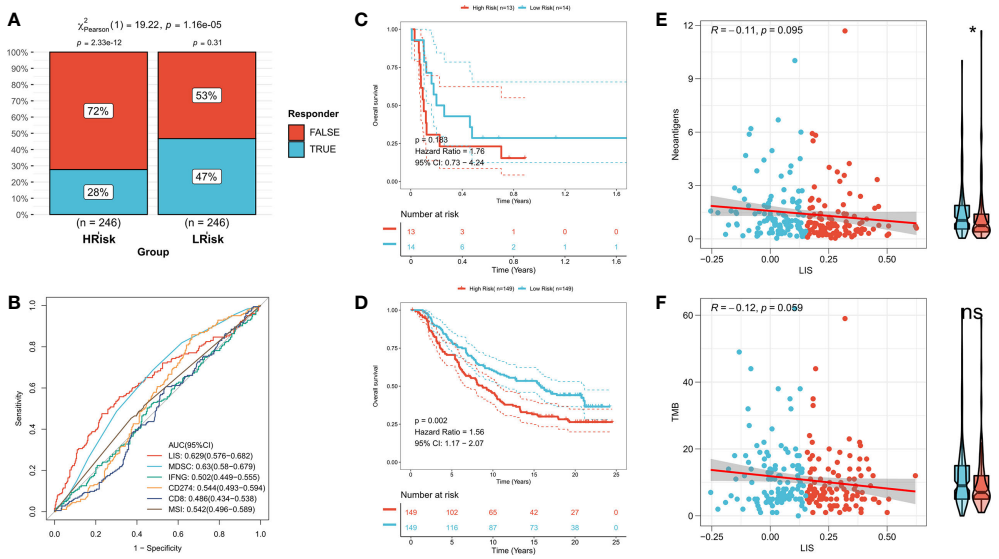


FIGURE 9
LIS in predicting immunotherapy. **(A)** TIDE algorithm predicted the response rate of immunotherapy with high LIS and low LIS. **(B)** ROC curve shows the prediction accuracy of LIS and different immune markers. **(C)** Survival curve of high LIS and low LIS subgroups in GSE135222 cohort. **(D)** Survival curve of high LIS and low LIS subgroups in Imvigor210 cohort. **(E)** Correlation between LIS and neoantigens in Imvigor210 queue. **(F)** Correlation between LIS and TMB in Imvigor210 queue. *p < 0.05; ns p > 0.05.

identified and verified two heterogeneous subtypes in LUAD, one of which was an effector immune cell with more expression of IRGs, enrichment of immunoactive pathways and high abundance, which was defined as an immunoactive subtype. Another kind of suppressive immune cells expressing more LMGs and high abundance, showing the characteristics of tumor metabolism and proliferation, was defined as lipid metabolism subtype. We verified the stability and repeatability of the two subtypes in a GEO meta cohort. These two subtypes also showed heterogeneity in genome driven events, clinical outcomes, and immunotherapy responses. In addition, a robust prognostic feature was proposed based on these two subtypes: LIS. Further analysis showed that LIS shows a leading advantage in predicting the immunotherapy of LUAD patients. These results promote the understanding of the interaction between lipid metabolism and TME and offer a new direction for clinical management and precision treatment of LUAD patients. These two subtypes showed different clinical characteristics. The survival of immunoactive phenotype was significantly better than that of lipid metabolism phenotype, and patients with more disease progression were in lipid metabolism phenotype. Functional enrichment indicated that metabolic related pathways and cell-cycle related pathways were enriched in the lipid metabolism phenotype, while effector immune cell receptor signaling pathways and immune related pathways were enriched in the immunoactive phenotype. In addition, immune infiltration analysis also suggested that there was higher effector immune cell infiltration in the immunoactive phenotype, and more tumor cells and inhibitory immune cells in the lipid metabolism subtype. These results suggest the hypermetabolism and proliferation of tumors in the lipid metabolism subtype and explain the worse survival rate and tumor progression of patients with this subtype (37). More effector immune cells and stronger immune activity in the immunoactive phenotype play an anti-tumor role, resulting in better survival and tumor remission of patients (38).

Next, in order to elaborate the molecular characteristics of the two subtypes, the genomic alterations of the two subtypes were compared. In general, there was a higher TMB in the lipid metabolism subtype. It is worth noting that the mutation frequency of KEAP1 gene in the lipid subtype was significantly increased compared to that in the immunoactive phenotype. KEAP1 is an essential regulator of cell homeostasis and antioxidant stimulation (39). Studies have reported that this mutation is common in NSCLC with close correlation to higher tumor growth and invasiveness (40). Additionally, tumors bearing KEAP pathway mutations have been reported in preclinical and clinical studies, which have stronger resistance to traditional treatment methods, such as chemotherapy, radiotherapy, and targeted therapy (41–43). In addition, we found that the amplification and deletion levels of immunoactive subtypes were significantly higher at the chromosome arm level, and the deletion levels of lipid metabolism subtypes were higher in general. The contradictory results may suggest that CNV does not seem to be playing a pivotal role in regulating the differences between subtypes. In general, the genomic changes of these two subtypes are mainly mediated by gene mutations, especially KEAP1, which may contribute to the heterogeneous response of subtypes to tumor

treatment, leading to different clinical outcomes. In addition, KEAP1 may also be a new target for drug development and clinical treatment of LUAD.

Finally, a prognostic feature called LIS was developed and validated in the TCGA cohort, GEO meta cohort, and two external immunotherapy cohorts. High LIS is an independent negative prognostic factor for OS, and subgroup analysis showed that LIS showed stronger performance in predicting early LUAD patients. Considering the heterogeneity of subtypes in immunotherapy, we also evaluated the effectiveness of LIS in predicting immunotherapy. The results showed that LIS also showed high accuracy in the immunotherapy cohort. In addition, LIS also showed better accuracy than commonly used biomarkers (MDSC, MSI score, IFNG, CD8 and CD274). Finally, we found that patients with low LIS may have more neoantigens, which may lead to stronger immunotherapy sensitivity in patients with low LIS. In conclusion, our results suggest that LIS is not only a robust prognostic marker, but also a promising predictive marker of immunotherapy.

We admit that our research also has some defects. First, we only use Bulk-seq data without considering the heterogeneity between cells. Secondly, the sequenced samples came from tumor tissue, which may lead to the fact that LIS is not suitable for peripheral blood samples, and the clinical application is limited. Finally, although we used algorithms and mature immunotherapy cohorts to evaluate the sensitivity of the two subtypes to immunotherapy, prospective clinical research cohorts are still needed for validation. In conclusion, our work identified and validated heterogeneous lipid metabolism subtypes and immune activity subtypes in LUAD, which showed heterogeneity in clinical outcomes, biological functions, immune infiltration, and genome driven events. In addition, we have developed a feature called LIS, which can be used as a reliable prognostic biomarker for predicting OS and immunotherapy response. These results promote the understanding of the interaction between lipid metabolism and TME and offer a new direction for clinical management and precision therapy of LUAD 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

XG performed the data analyses and wrote the manuscript. SW contributed significantly to analysis and manuscript preparation. ZL helped perform the analysis with constructive discussions. HX contributed to the conception of the study. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Programs of Shanghai Pulmonary Hospital (No. fkxr1902 and No.fkxr1904).

Acknowledgments

We greatly appreciate the analytical data provided by the TCGA and GEO databases.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1022149/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The workflow of this study. In this study, we aimed to survey the crosstalk between lipid metabolism and tumor immune response in LUAD patients and identified two heterogeneous subtypes (lipid metabolism subtype and immune activity subtype). These two subtypes show specific differences in clinical outcomes, biological functions, immune infiltration and genomic variation. In addition, a lipid-immune score (LIS) was developed and validated, which shows significant advantages in predicting prognosis and immunotherapy response. In conclusion, our work strengthens the understanding of the complex role between lipid metabolism and immune system in LUAD and provides a new perspective and reference for the accurate prediction and immunotherapy of LUAD patients.

SUPPLEMENTARY FIGURE 2
Validation of immune and lipid subtypes. **(A)** CDF curve of consensus matrix of different K in GEO queue. **(B)** Consensus matrix when k = 2 in GEO queue. **(C)** Expression heat maps of LMGs and IRGs in two subtypes in the GEO cohort. **(D)** Survival curve of two subtypes in GEO cohort.

SUPPLEMENTARY FIGURE 3
Verification of immune infiltration in subtypes. **(A)** Differences in Estimate scores between the two subtypes in the GEO cohort. **(B)** Differences in the expression of six typical immune checkpoints (PD-L1, CD8A, CTLA-4, LAG-3, PD-1) between the two subtypes in the GEO cohort. **(C)** Differences in immune related pathway activity between the two subtypes in the GEO cohort. **(D)** Differences in immune cell infiltration between the two subtypes in the GEO cohort.

SUPPLEMENTARY FIGURE 4
External verification of LIS. **(A)** Survival curves of high and low LIS subgroups in the GEO cohort. **(B)** ROC analysis of LIS in GEO queue. **(C)** The scatter plot shows the survival status of different LIS patients in the GEO cohort. **(D)** TROC curve of LIS in GEO queue. **(E)** Subgroup analysis of LIS in patients with different clinical characteristics in GEO cohort.

SUPPLEMENTARY FIGURE 5
External verification of LIS. **(A)** The TIDE algorithm in GEO queue predicted the response rate of immunotherapy with high LIS and low LIS. **(B)** ROC curves in the GEO cohort showed the predictive accuracy of LIS and different immune markers.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

2. Cheng TY, Cramb SM, Baade PD, Youlden DR, Nwogu C, Reid ME. The international epidemiology of lung cancer: Latest trends, disparities, and tumor characteristics. *J Thorac oncol: Off Publ Int Assoc Study Lung Cancer* (2016) 11 (10):1653–71. doi: 10.1016/j.jtho.2016.05.021

3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* (2018) 553(7689):446–54. doi: 10.1038/nature25183

4. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA: Cancer J Clin* (2018) 68(1):7–30. doi: 10.3322/caac.21442

5. Warburg O. On the origin of cancer cells. *Sci (New York NY)* (1956) 123 (3191):309–14. doi: 10.1126/science.123.3191.309

6. Yu W, Lei Q, Yang L, Qin G, Liu S, Wang D, et al. Contradictory roles of lipid metabolism in immune response within the tumor microenvironment. *J Hematol Oncol* (2021) 14(1):187. doi: 10.1186/s13045-021-01200-4

7. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab* (2016) 23(1):27–47. doi: 10.1016/j.cmet.2015.12.006

8. Wen YA, Xing X, Harris JW, Zaytseva YY, Mitov MI, Napier DL, et al. Adipocytes activate mitochondrial fatty acid oxidation and autophagy to promote tumor growth in colon cancer. *Cell Death Dis* (2017) 8(2):e2593. doi: 10.1038/cddis.2017.21

9. Yao CH, Fowle-Grider R, Mahieu NG, Liu GY, Chen YJ, Wang R, et al. Exogenous fatty acids are the preferred source of membrane lipids in proliferating fibroblasts. *Cell Chem Biol* (2016) 23(4):483–93. doi: 10.1016/j.chembiol.2016.03.007

10. Corn KC, Windham MA, Rafat M. Lipids in the tumor microenvironment: From cancer progression to treatment. *Prog Lipid Res* (2020) 80:101055. doi: 10.1016/j.plipres.2020.101055

11. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034

12. Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, et al. DNA Methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun* (2019) 10(1):4278. doi: 10.1038/s41467-019-12159-9

13. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* (2018) 554(7693):544–8. doi: 10.1038/nature25501

14. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinf (Oxford England)* (2011) 27(12):1739–40. doi: 10.1093/bioinformatics/btr260

15. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data* (2018) 5:180015. doi: 10.1038/sdata.2018.15

16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *Omics: J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

17. McDermott DF, Huseni MA, Atkins MB, Motzer RJ, Rini BI, Escudier B, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med* (2018) 24(6):749–57. doi: 10.1038/s41591-018-0053-3

18. Gibbons DL, Creighton CJ. Pan-cancer survey of epithelial-mesenchymal transition markers across the cancer genome atlas. *Dev dyn: an Off Publ Am Assoc Anat* (2018) 247(3):555–64. doi: 10.1002/dvdy.24485

19. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* (2017) 127(8):2930–40. doi: 10.1172/jci91190

20. Yu T, Tan H, Liu C, Nie W, Wang Y, Zhou K, et al. Integratively genomic analysis reveals the prognostic and immunological characteristics of pyroptosis and ferroptosis in pancreatic cancer for precision immunotherapy. *Front Cell Dev Biol* (2022) 10:826879. doi: 10.3389/fcell.2022.826879

21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12 (5):453–7. doi: 10.1038/nmeth.3337

22. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

23. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023

24. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28 (11):1747–56. doi: 10.1101/gr.239244.118

25. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. Survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinf (Oxford England)* (2011) 27(22):3206–8. doi: 10.1093/bioinformatics/btr511

26. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18 (1):248–62. doi: 10.1016/j.celrep.2016.12.019

27. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

28. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-Scale public data reuse to model immunotherapy response and resistance. *Genome Med* (2020) 12 (1):21. doi: 10.1186/s13073-020-0721-z

29. Wang C, Gu X, Zhang X, Zhou M, Chen Y. Development and validation of an E2F-related gene signature to predict prognosis of patients with lung squamous cell carcinoma. *Front Oncol* (2021) 11:756096. doi: 10.3389/fonc.2021.756096

30. Wang Y, Tan H, Yu T, Chen X, Jing F, Shi H. Potential immune biomarker candidates and immune subtypes of lung adenocarcinoma for developing mRNA vaccines. *Front Immunol* (2021) 12:755401. doi: 10.3389/fimmu.2021.755401

31. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inf decis mak* (2008) 8:53. doi: 10.1186/1472-6947-8-53

32. Desrichard A, Snyder A, Chan TA. Cancer neoantigens and applications for immunotherapy. *Clin Cancer res: an Off J Am Assoc Cancer Res* (2016) 22(4):807–12. doi: 10.1158/1078-0432.Ccr-14-3175

33. Hu Z, Ott PA, Wu CJ. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol* (2018) 18(3):168–82. doi: 10.1038/nri.2017.131

34. Abida W, Cheng ML, Armenia J, Middha S, Autio KA, Vargas HA, et al. Analysis of the prevalence of microsatellite instability in prostate cancer and response to immune checkpoint blockade. *JAMA Oncol* (2019) 5(4):471–8. doi: 10.1001/jamaoncol.2018.5801

35. Yin N, Liu Y, Khoor A, Wang X, Thompson EA, Leitges M, et al. Protein kinase cι and wnt/β-catenin signaling: Alternative pathways to Kras/Trp53-driven lung adenocarcinoma. *Cancer Cell* (2019) 36(2):156–67.e7. doi: 10.1016/j.ccell.2019.07.002

36. Justilien V, Ali SA, Jamieson L, Yin N, Cox AD, Der CJ, et al. Ect2-dependent rRNA synthesis is required for KRAS-TRP53-Driven lung adenocarcinoma. *Cancer Cell* (2017) 31(2):256–69. doi: 10.1016/j.ccell.2016.12.010

37. Vander Heiden MG, DeBerardinis RJ. Understanding the intersections between metabolism and cancer biology. *Cell* (2017) 168(4):657–69. doi: 10.1016/j.cell.2016.12.039

38. Madden MZ, Rathmell JC. The complex integration of T-cell metabolism and immunotherapy. *Cancer Discovery* (2021) 11(7):1636–43. doi: 10.1158/2159-8290.Cd-20-0569

39. Yamamoto M, Kensler TW, Motohashi H. The KEAP1-NRF2 system: a thiol-based sensor-effector apparatus for maintaining redox homeostasis. *Physiol Rev* (2018) 98(3):1169–203. doi: 10.1152/physrev.00023.2017

40. Hellyer JA, Padda SK, Diehn M, Wakelee HA. Clinical implications of KEAP1-NFE2L2 mutations in NSCLC. *J Thorac oncol: Off Publ Int Assoc Study Lung Cancer* (2021) 16(3):395–403. doi: 10.1016/j.jtho.2020.11.015

41. Jeong Y, Hellyer JA, Stehr H, Hoang NT, Niu X, Das M, et al. Role of KEAP1/NFE2L2 mutations in the chemotherapeutic response of patients with non-small cell lung cancer. *Clin Cancer res: an Off J Am Assoc Cancer Res* (2020) 26 (1):274–81. doi: 10.1158/1078-0432.Ccr-19-1237

42. Binkley MS, Jeon YJ, Nesselbush M, Moding EJ, Nabet BY, Almanza D, et al. KEAP1/NFE2L2 mutations predict lung cancer radiation resistance that can be targeted by glutaminase inhibition. *Cancer Discovery* (2020) 10(12):1826–41. doi: 10.1158/2159-8290.Cd-20-0282

43. Hellyer JA, Stehr H, Das M, Padda SK, Ramchandran K, Neal JW, et al. Impact of KEAP1/NFE2L2/CUL3 mutations on duration of response to EGFR tyrosine kinase inhibitors in EGFR mutated non-small cell lung cancer. *Lung Cancer (Amsterdam Netherlands)* (2019) 134:42–5. doi: 10.1016/j.lungcan.2019.05.002

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fimmu.2022.1027154

Check for updates

OPEN ACCESS

EDITED BY
Meng Zhou,
Wenzhou Medical University, China

REVIEWED BY
Fulong Yu,
Boston Children's Hospital and Harvard Medical School, United States
Xiufei Chen,
University of Oxford, United Kingdom

*CORRESPONDENCE
Hong Zhang
mingyi-hong@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

RECEIVED 24 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Wang X, Huang Y, Li S and Zhang H (2022) Integrated machine learning methods identify FNDC3B as a potential prognostic biomarker and correlated with immune infiltrates in glioma.
Front. Immunol. 13:1027154.
doi: 10.3389/fimmu.2022.1027154

COPYRIGHT
© 2022 Wang, Huang, Li 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.

Integrated machine learning methods identify FNDC3B as a potential prognostic biomarker and correlated with immune infiltrates in glioma

Xiao Wang[1,2†], Yeping Huang[2†], Shanshan Li[2] and Hong Zhang[2*]

[1]Department of Nephrology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2]Shanghai Diabetes Institute, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

Background: Recent discoveries have revealed that fibronectin type III domain containing 3B (FNDC3B) acts as an oncogene in various cancers; however, its role in glioma remains unclear.

Methods: In this study, we comprehensively investigated the expression, prognostic value, and immune significance of FNDC3B in glioma using several databases and a variety of machine learning algorithms. RNA expression data and clinical information of 529 patients from the Cancer Genome Atlas (TCGA) and 1319 patients from Chinese Glioma Genome Atlas (CGGA) databases were downloaded for further investigation. To evaluate whether FNDC3B expression can predict clinical prognosis of glioma, we constructed a clinical nomogram to estimate long-term survival probabilities. The predicted nomogram was validated by CGGA cohorts. Differentially expressed genes (DEGs) were detected by the Wilcoxon test based on the TCGA-LGG dataset and the weighted gene co-expression network analysis (WGCNA) was implemented to identify the significant module associated with the expression level of FNDC3B. Furthermore, we investigated the correlation between FNDC3B with cancer immune infiltrates using TISIDB, ESTIMATE, and CIBERSORTx.

Results: Higher FNDC3B expression displayed a remarkably worse overall survival and the expression level of FNDC3B was an independent prognostic indicator for patients with glioma. Based on TCGA LGG dataset, a co-expression network was established and the hub genes were identified. FNDC3B expression was positively correlated to the tumor-infiltrating lymphocytes and immune infiltration score, and high FNDC3B expression was accompanied by the increased expression of B7-H3, PD-L1, TIM-3, PD-1, and CTLA-4. Moreover, expression of FNDC3B was significantly associated with infiltrating levels of several types of immune cells and most of their gene markers in glioma.

Conclusion: This study demonstrated that FNDC3B may be involved in the occurrence and development of glioma and can be regarded as a promising prognostic and immunotherapeutic biomarker for the treatment of glioma.

KEYWORDS

FNDC3B, glioma, prognosis, immune infiltration, The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA)

Introduction

Glioma is the most common primary tumor of the central nervous system in adults and is characterized by high recurrence and mortality rates. According to the World Health Organization (WHO), glioma is typically divided into two principal subgroups: low-grade glioma (LGG; grade II and III) and glioblastoma multiforme (GBM; grade IV; most aggressive and lethal subtype) based on the malignant degree (1). The median survival is less than two years, and the overall prognosis is poor for glioma patients even after surgical resection, chemotherapy, and radiation therapy (CRT), especially for those with GBM (2). Even though many advances have been made in adjuvant therapy and surgery in the past few decades, the clinical outcomes have not been significantly improved for glioma patients. Apart from the traditional treatment (3), studies in recent years have revealed the use of novel and effective methods, such as immunotherapy to treat glioma owing to the success achieved from other solid tumors, including lung, bladder, and kidney cancers, and melanoma (4, 5). However, there is still an urgent need to identify additional immune biomarkers for combination therapy due to the resistance to monotherapy (6–8). Furthermore, this may elucidate the mechanism of tumorigenesis and help to identify new molecular targets for treatment.

Fibronectin type III domain containing 3B (FNDC3B, also named FAD104), which belongs to the FNDC3 family, was initially identified as a regulator of adipocyte and osteoblast differentiation (9). FNDC3B is an endoplasmic reticulum transmembrane protein with a single transmembrane domain at the C terminus preceded by nine repeated fibronectin type III domains. Its biological function remains largely unknown (10). Recent research demonstrated that FNDC3B plays a major role in cell adhesion, proliferation, and growth signaling due to the fibronectin type III domain, which has the ability to combine with various proteins (11). For the past few years, emerging evidence has demonstrated that FNDC3B was abnormally expressed in several types of human cancers, including hepatocellular carcinoma, acute myeloid leukemia, colorectal and cervical cancers (12–15). For instance, Han reported that FNDC3B expression was correlated with a worse prognosis in cervical cancer, while its carcinogenic effects are still unclear (15). Notably, a few studies have demonstrated that FNDC3B expression levels are correlated with glioblastoma. Wang and Xu reported that MiR-1225-5p and MiR-129-5p inhibit the malignant glioblastoma cells *via* targeting FNDC3B (16, 17). Furthermore, a newly integrated analysis of RNA binding proteins in glioma revealed that FNDC3B can not only serve as a useful prognostic biomarker but also promote glioma cell proliferation (18). However, the overall expression profile of FNDC3B and its potential role in the development and distinct clinical significance of glioma has not been fully elucidated. In previous study, Rajasagi et al. found that long-lived cytotoxic T-cell responses against peptides generated from personal tumor mutations in FNDC3B presented on chronic lymphocytic leukemia cells (19). Until now, there are very limited studies on the correlation between FNDC3B and tumor-infiltrating lymphocytes (TILs) in glioma.

Recent advancements in high-throughput sequencing technologies and large-scale cancer genomics databases have enabled a systematic and comprehensive analysis of genes from the perspective of machine learning (20–22). In the present study, we carried out an intensive analysis for the expression signature of FNDC3B using various publicly accessible databases, as well as applied data mining of the TCGA and CGGA datasets to explore its in-depth prognostic effect. We also investigated the correlations between FNDC3B expression and tumor immune microenvironment (TIM) in LGG patients in order to elucidate the underlying mechanisms and improve molecular diagnosis for glioma patients.

Materials and methods

Gene expression pattern based on ONCOMINE and GEPIA2

FNDC3B expression levels in various cancers were firstly explored by ONCOMINE database (https://www.oncomine.org/) (23), which is currently the largest public cancer microarray database and integrated data-mining platform. The threshold for ONCOMINE was set according to the default settings of $P < 0.0001$,

fold change > 2, and Gene Rank < Top 10%. Gene expression profiling interactive analysis 2 (GEPIA2) is another useful web-based tool (http://gepia.cancer-pku.cn/) that contains RNA sequencing data based on 9,736 tumor and 8,587 normal control samples from The Cancer Genome Atlas (TCGA) and The Genotype-Tissue Expression (GTEx) (24), providing differential expression analysis, correlation analysis, survival analysis, and custom data analysis. GEPIA2 contains 518 LGG samples, 163 GBM samples, and 207 normal brain samples. FNDC3B expression was compared between LGG or GBM and normal tissues by Student t-tests. Samples were considered to be significant with $p < 0.05$ and fold change > 2. Protein expression of FNDC3B in glioma and normal brain tissues were evaluated based on immunohistochemistry data from the Human Protein Atlas (HPA) (https://www.proteinatlas.org/). The mRNA expression of FNDC3B in various human cancer cell lines were obtained from Broad Institute Cancer Cell Line Encyclopedia (CCLE).

Data source and processing

Gene expression data and corresponding clinical information for glioma patients were downloaded from TCGA (LGG and GBM) and Chinese Glioma Genome Atlas (CGGA) (mRNAseq_325, mRNAseq_693 and mRNA_array_301) database. The data from the TCGA was applied to explore the prognostic role of FNDC3B in gliomas, and the three CGGA cohorts were used to validate the results. The DNA methylation along with transcriptional data of FNDC3B for patients with LGG from the TCGA database was downloaded *via* the cBio Cancer Genomics Portal (cBioPortal) website (http://www.cbioportal.org/). The detailed methodology of the study is shown in Figure 1.

Clinicopathological correlation and prognosis analysis

Coefficients of Cox regression were determined by data mining among 21 TCGA cancer types using OncoLnc (http://www.oncolnc.org/) to compare FNDC3B expression in different tumors. The gene expression data and corresponding clinical information from the TCGA and CGGA were used to evaluate the prognostic role of FNDC3B in gliomas. Patients with incomplete clinical information were eliminated. We explored the correlation between FNDC3B expression and various clinical features using the Wilcoxon test in R version 3.6.3. Kaplan-Meier analysis and log-rank test were implemented to evaluate

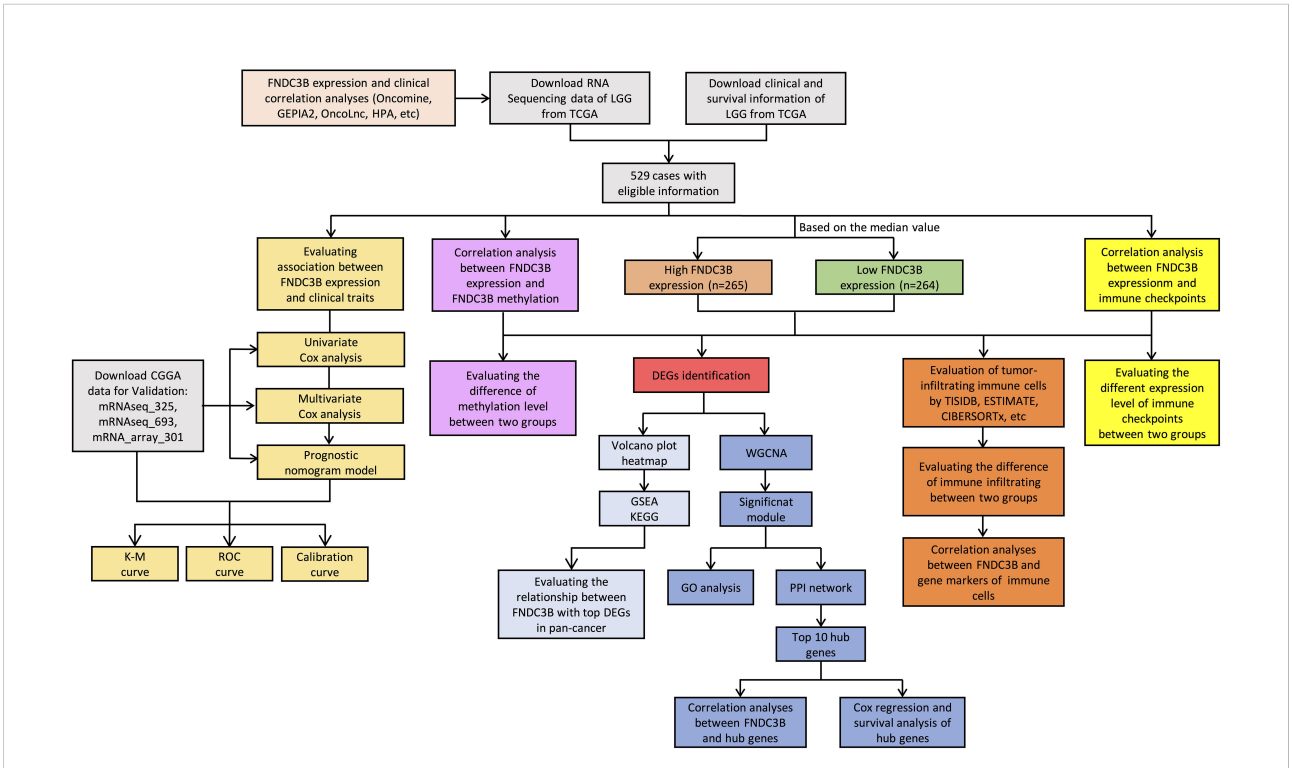


FIGURE 1
Flowchart of this study: data collection, processing, analysis, and validation. LGG, low-grade glioma; TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas; GEPIA2, gene expression profiling interactive analysis 2; HPA, Human Protein Atlas; K-M, Kaplan-Meier; ROC, receiver operating characteristic; DEGs, differentially expressed genes; WGCNA, weighted gene co-expression network analysis; GSEA, gene set enrichment analysis; KEGG, kyoto encyclopedia of genes and genomes; GO, gene ontology; PPI, protein-protein interaction.

the associations of FNDC3B and clinical variables with overall survival (OS). Univariate and multivariate Cox regression analyses were performed to further examine whether FNDC3B expression was a significant factor associated with OS when adjusted by clinical variables (age at initial pathological diagnosis, gender, neoplasm histologic grade, IDH mutation status, etc.). "Survival" and "survminer" packages in R software were used for the stepwise variable selection and Cox model construction.

Construction and evaluation of a nomogram

Nomogram is widely used as a predictive model for cancer patients and can provide prognostic risk individually and intuitively (25). Based on the TCGA-LGG dataset, we constructed a prognostic nomogram model to predict the probability of 2-, 3-, and 5- year OS using "rms" package in R. The nomogram combined the expression level of FNDC3B with traditional clinical parameters (age, grade, and IDH status) and formulated the scoring criteria for all the parameters in the regression equation based on their regression coefficients. Then, the summed score for each patient was converted into the probability of the outcome time by using the nomogram. The performance and prediction efficiency of the nomogram were evaluated *via* plotting the calibration, K-M and receiver operating characteristic (ROC) curves of the three CGGA validation datasets.

Screening of differentially expressed genes (DEGs)

The median value was used to create a categorical dependent variable based on FNDC3B expression level. For the TCGA-LGG dataset, DEGs between the high and low FNDC3B groups were screened using Wilcoxon test (screening criteria: $P < 0.01$, FDR < 0.05, and $|logFC| > 1$). Volcano plot of all DEGs was generated using R. For accurate results, we set the average expression level to at least 1 for the raw DEGs. Then, the filtered DEGs were selected for further analysis. Heatmap was generated by "ComplexHeatmap" package in R. To identify biological pathway differences between the high and low FNDC3B groups, gene set enrichment analysis (GSEA) and kyoto encyclopedia of genes and genomes (KEGG) were performed on the screened DEGs using "org.Hs.eg.db", "clusterProfiler", "enrichplot" and "ggplot2" R packages. The correlations between FNDC3B with DEGs in pan-cancer were obtained by using Gene_Corr module of TIMER2.0 (tumor immune estimation resource, version 2) (http://timer.cistrome.org/). $P < 0.05$ was considered to be a significant enrichment.

Co-expression network creation and hub genes identification

To reveal the correlation between genes and identify the ones with significant relationships, 2,099 DEGs between the high and low FNDC3B expression groups were used to construct a weighted co-expression network. Patients with incomplete expression information were removed, and the remaining samples were implemented to construct the network. In order to ensure the reliability of the network structure, we set gene expression value larger than 1 in at least 10% of all the samples, and the average expression level was at least 0.5 for the raw DEGs. A scale-free gene co-expression network based on 673 filtered DEGs was constructed using the R package "WGCNA". Hierarchical clustering tree was created based on a dissimilarity measure (1-TOM), and genes with similar expression patterns were merged into the same module. The most relevant module was revealed by calculating the correlation between modules and the FNDC3B group; the genes in the most significant module were extracted to determine the target genes. For the selected module, Gene Ontology (GO) was performed in R using the packages "org.Hs.eg.db", "AnnotationDbi", "enrichplot" and "ggplot2" with q-values less than 0.0001. The Search Tool for the Retrieval of Interacting Genes Database (STRING) version 11 was applied to generate the PPI network and the combined score > 0.4 was used as the cut-off criterion. Cytoscape version 3.8.2 was employed to visualize the molecular interaction networks and biological pathways. Hub genes were identified *via* CytoHubba plugin with the top 10 MCC values. We used GeneMANIA database (http://genemania.org/) to build the gene-gene interaction network for hub genes in terms of physical interactions, co-expression, shared protein domains, pathways, predicted interactions, and colocalization, as well as to predict their biological functions. The relationships of the expression levels between the hub genes and FNDC3B were shown by scatterplots using "ggplot2" R package.

Analysis of FNDC3B-associated immunomodulators

To investigate the immune infiltration of FNDC3B in different cancers, TISIDB database (http://cis.hku.hk/TISIDB/) was applied to infer the correlations between 28 types of tumor-infiltrating lymphocytes (TILs) and FNDC3B expression. TISIDB integrates a variety of data sources in tumor immunology, including abundant human cancer datasets from TCGA and text mining results from PubMed. Spearman correlation test was implemented to estimate the association between FNDC3B and TILs. Furthermore, Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression (ESTIMATE) algorithm was employed to evaluate

the stromal score, immune score, and ESTIMATE score for each sample using the downloaded data. Moreover, CIBERSORTx (https://cibersortx.stanford.edu/) was used to assess the relative variations of 22 types of tumor-infiltrating immune cells between the high and low FNDC3B expression groups in LGG. Packages "ggplot2", "ggpubr" and "ggExtra" in R were applied to investigate the correlation between FNDC3B and immune checkpoints based on TCGA-LGG dataset. The associations between gene markers of the significant immune cells and FNDC3B expression were also determined by the correlation analysis function in GEPIA2. Differences with a P-value < 0.05 were considered significant in all tests.

Statistical analysis

All statistical analysis was performed in R software and P < 0.05 was considered statistically significant. Wilcoxon test was used to compare the differences for clinical characteristics or immune scores grouped by FNDC3B expression. Univariate (Log-rank test) and multivariate Cox regression (cox proportional hazard, coxph) analyses were performed to assess clinical traits associated with OS.

Results

FNDC3B mRNA expression in various cancers

We first assessed the expression of FNDC3B in different tumors and normal tissues of multiple cancer types using the Oncomine database. The results showed that abnormal FNDC3B expression was retrieved from a total of 368 datasets. Among them, FNDC3B expression levels were significantly upregulated in tumor tissues in 43 datasets, including brain and central nervous system (CNS), head and neck, esophageal, kidney, cervical, bladder, and colorectal cancers, etc. (Figure 2A, P < 0.0001, Fold Change > 2, and Gene Rank < Top 10%). In addition, its expression in leukemia, lymphoma, breast cancer, and sarcoma lymphoma was shown to be downregulated in multiple datasets. In summary, FNDC3B is generally upregulated in several tumors. In the brain and CNS cancers dataset, there were seven studies on the upregulation of FNDC3B and no studies on its downregulation. Comparison of FNDC3B across the seven studies showed a Median Rank = 302, suggesting that FNDC3B was highly expressed in glioma tissues and concentrated both in LGG and GBM (Figure 2B, P <0.0001). As for the outlier analysis of FNDC3B, 822 different studies on FNDC3B have been included in Oncomine database. Among them, 11 studies indicated the upregulation of FNDC3B and four studies the downregulation in brain and CNS cancers (Figure 2A). Then, we evaluated the differentially expressed level of FNDC3B in TCGA pan-cancer data using GEPIA2. Our findings revealed that an elevated expression of FNDC3B was involved in a variety of tumors (Figure 2C). In TCGA LGG and GBM cohorts, FNDC3B expression was significantly higher in tumors compared to matched normal tissues (Figure 2D). Specifically, FNDC3B expression was 2.71-fold in LGG and 7.16-fold in GBM vs. normal brain tissue. Genetic alterations of FNDC3B in glioma patients were examined using cBioPortal (Supplementary Figure 1). Among 6216 samples from 5774 patients in 14 glioma datasets, the overall alteration frequency of FNDC3B gene is 1.5% (71/4774); amplification, mutations, and deep deletions were the most common types of alteration. Due to the low mutation rate, FNDC3B may not be a hypermutation gene in the glioma cohort. In our previous study, we had confirmed that methylation status of TERT was strongly correlated with its expression in hepatocellular carcinoma (HCC) (26). In the present research, the expression of FNDC3B was negatively correlated with FNDC3B DNA methylation; the methylation levels were reduced in the FNDC3B high group based on the TCGA-LGG dataset (Figures 2E, F). These results suggested that FNDC3B may be negatively regulated by epigenetic modification and lead to its high expression in glioma samples.

FNDC3B protein expression was explored using HPA database. The immunohistochemistry result revealed upregulated FNDC3B in glioma samples (Figure 2G). In human cancer tissues, FNDC3B protein expression in glioma was ranked as the top 12 out of 20 distinct cancer types (Supplementary Figure 2A). Furthermore, we detected the FNDC3B mRNA level in human normal tissues using the GTEx database, and FNDC3B expression was mainly found in lung, adipose tissue, thyroid gland, endometrium, and ovary. It is worth noting that the brain displayed the lowest expression levels of global FNDC3B transcript across all normal tissues (Supplementary Figure 2B). In addition, we systematically elucidated the expression levels of the FNDC3B in different cancer cells by querying the CCLE database and found that FNDC3B was highly expressed in glioma cell lines (Supplementary Figure 2C).

High expression of FNDC3B predicts poor prognosis of glioma

OncoLnc and GEPIA2 are free online resources and databases for the analysis and visualization of datasets from the TCGA and GTEx projects. To examine the function of FNDC3B on OS in various cancers, we used OncoLnc online tool to perform Cox regression analysis. We found that FNDC3B expression in LGG was ranked first among 21 different cancer types based on the FDR correction (Table 1). Moreover, we analyzed the relationships between FNDC3B expression and prognostic values in 33 types of cancer using GEPIA2 databases. As shown in Supplementary Figure 3, high FNDC3B expression levels were associated with poorer prognosis of OS and disease-free survival (DFS) in adrenocortical carcinoma (ACC), GBM,

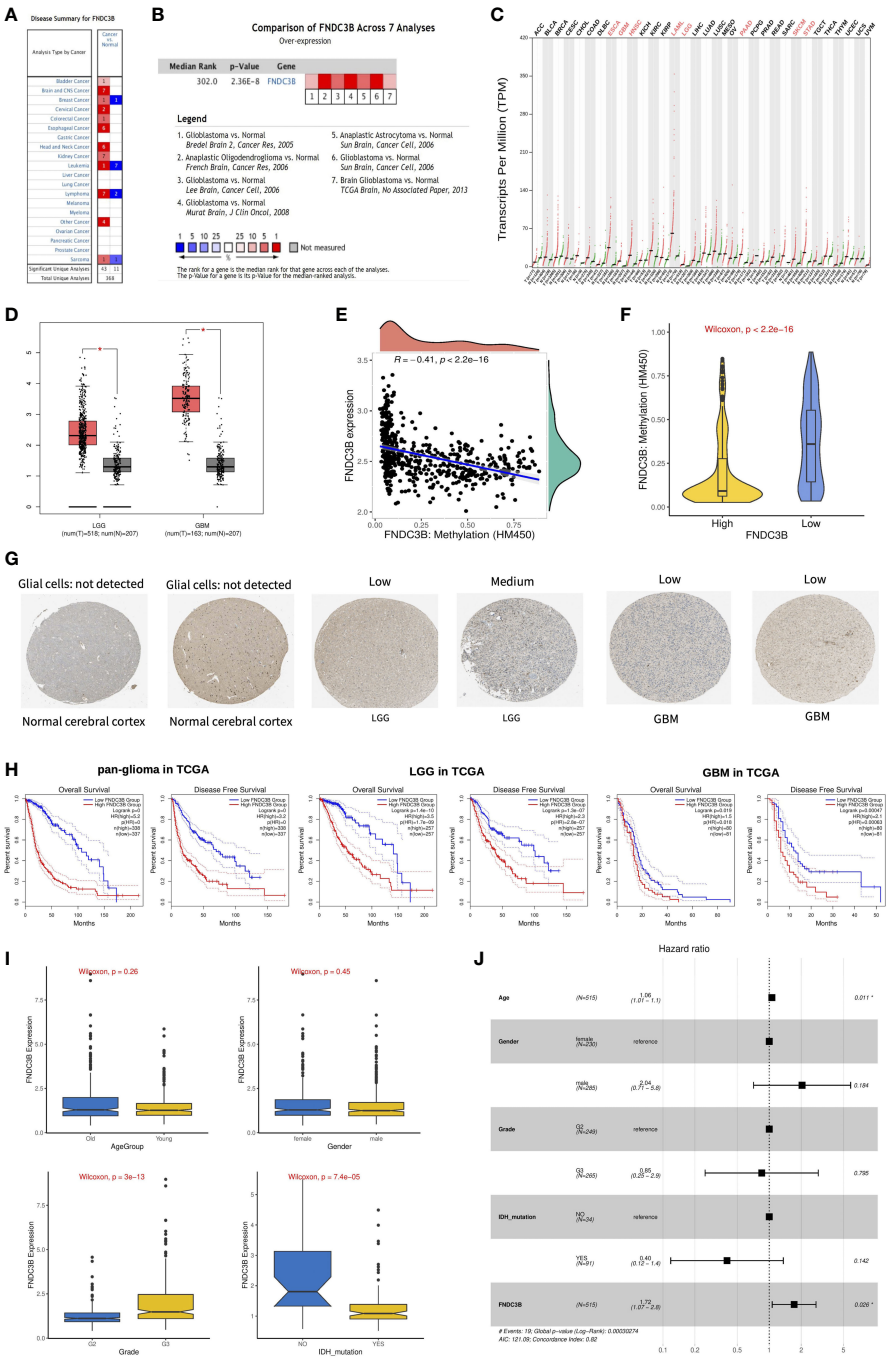


FIGURE 2
The expression profiles and prognostic value of FNDC3B in glioma. **(A)** In brain and CNS cancers, FNDC3B was significantly upregulated in seven studies. Red represents high expression and blue represents low expression. The darker the red color, the higher the gene expression level. The darker the blue color, the lower the gene expression level. **(B)** Comparison of FNDC3B expression across seven analyses, and red represents high expression. **(C)** FNDC3B was significantly upregulated in various tumors. **(D)** Expression level of FNDC3B in LGG and GBM compared to control. *, P < 0.05. **(E)** The expression of FNDC3B was negatively regulated by FNDC3B DNA methylation. **(F)** Different methylation levels of FNDC3B in the FNDC3B high- and low-expression groups in TCGA LGG samples. **(G)** Representative FNDC3B protein expression in normal and glioma tissues. Data were obtained from the Human Protein Atlas (HPA). **(H)** Kaplan-Meier analysis of OS and DFS based on FNDC3B high- vs. low-expression in pan-glioma, LGG, and GBM patients in the TCGA dataset. Red curve represents patients with high expression of FNDC3B, and blue curve represents low FNDC3B. **(I)** The correlations between FNDC3B expression and clinical characteristics based on TCGA LGG datasets: age, gender, grade, and isocitrate dehydrogenase (IDH) status. **(J)** Multiple Cox regression analysis of clinicopathological features (including FNDC3B expression) and OS in the TCGA LGG datasets.

10.3389/fimmu.2022.1027154

TABLE 1 Cox regression results of FNDC3B among 21 tumor types.

Cancer	Cox Coefficient	P-value	FDR	Rank	Median Expression
LGG	0.82	1.30E-13	2.43E-10	9	324.53
PAAD	0.472	2.00E-05	5.74E-03	57	2620.58
SKCM	-0.242	8.20E-04	1.23E-02	1073	1402.12
KIRP	0.513	1.50E-03	1.33E-02	1827	2117.32
CESC	0.618	2.30E-05	3.41E-02	11	1686.62
LUAD	0.162	3.10E-02	1.55E-01	3320	2465.66
LAML	-0.339	3.00E-03	1.72E-01	260	3845.4
LIHC	0.154	1.10E-01	3.33E-01	5095	1812.82
BLCA	0.113	1.30E-01	3.61E-01	5715	2083.57
BRCA	0.163	7.70E-02	4.29E-01	2968	1715.64
STAD	0.126	1.20E-01	4.70E-01	4198	2732.83
HNSC	0.078	2.60E-01	6.02E-01	7067	2235.09
SARC	0.084	4.40E-01	7.07E-01	10003	2106.74
KIRC	0.031	7.10E-01	7.98E-01	14796	3396.2
GBM	0.137	1.70E-01	8.17E-01	3379	862.45
COAD	-0.03	7.70E-01	9.27E-01	13553	1591.87
LUSC	-0.019	7.80E-01	9.50E-01	13845	2926.55
UCEC	-0.159	1.40E-01	9.64E-01	2309	2368.54
OV	-0.019	8.00E-01	9.65E-01	13931	2115.89
ESCA	-0.06	6.40E-01	9.84E-01	10805	3323.64
READ	-0.067	7.50E-01	9.86E-01	12457	1657.46

kidney Chromophobe (KICH), LGG, liver hepatocellular carcinoma (LIHC), mesothelioma (MESO), and pancreatic adenocarcinoma (PAAD); OS in cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) and lung adenocarcinoma (LUAD); DFS in colon adenocarcinoma (COAD) and uveal Melanoma (UVM). A high FNDC3B expression was correlated with a better prognosis of OS and DFS in skin cutaneous melanoma (SKCM) as well as OS in acute myeloid leukemia (LAML).

To assess the prognostic significance of FNDC3B in glioma patients from TCGA and CGGA, samples were first split into two groups according to the median expression of FNDC3B for each dataset. Among pan-glioma, LGG, and GBM in the TCGA datasets, patients with higher FNDC3B levels presented shorter OS and DFS (Figure 2H) compared to patients expressing low levels of FNDC3B. Similarly, high FNDC3B expression was significantly associated with poor prognosis in all the three CGGA datasets (seq_325, seq_693, array_301) (Supplementary Figure 4). In particular, highly expressed FNDC3B was significantly related to reduced DFS in GBM ($p<0.001$), however, just marginally correlated with worse OS ($P<0.05$). We postulate that the less obvious but significant results of GBM may be due to the insufficient statistical power of a small sample size. Furthermore, we investigated the associations between FNDC3B expression and clinical characteristics, such as age, gender, grade, and isocitrate dehydrogenase (IDH) status. FNDC3B expression was higher in high-grade and IDH wildtype patients; there was no significant difference between age and gender based on the TCGA datasets (Figure 2I). Our findings revealed that higher FNDC3B expression is closely correlated with the malignant clinical characters of gliomas.

Subsequently, univariate and multivariate Cox regression analyses were performed to identify whether FNDC3B expression represented an independent prognostic factor. Univariate Cox analysis showed that FNDC3B (HR = 1.64; 95% CI = 1.50-1.80; $P < 0.001$), grade (HR = 3.37; 95% CI = 2.28-4.98; $P < 0.001$) and age (HR = 1.06; 95% CI = 1.04-1.07; $P < 0.001$) were high-risk factors, and IDH mutation (HR = 0.18; 95% CI = 0.07-0.48; $P < 0.001$) was a low-risk factor (Table 2). In multivariate Cox regression analysis, FNDC3B was independently associated with overall survival, suggesting it could be an independent prognostic biomarker for glioma (HR = 1.72; 95% CI = 1.07-2.80; $P < 0.05$). In addition, age may also be an independent prognostic factor (Table 2 and Figure 2J).

Construction and validation of a prognostic nomogram

A quantitative prognostic nomogram model to predict individual survival chances was established based on the TCGA-LGG dataset using Cox regression (Figure 3A). According to the stepwise Cox multivariate regression analysis,

TABLE 2 Univariate and multivariate cox regression analyses of prognostic factors in 416 cases of low-grade glioma (LGG).

Parameter	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P-Value	HR (95% CI)	P-Value
Age (continuous, years)	1.06 (1.04-1.07)	**<0.001**	1.06 (1.01-1.10)	**0.011**
Gender (ref. Female)	1.11 (0.78-1.58)	0.574	2.01 (0.71-5.80)	0.184
Grade (ref. WHO II)	3.37 (2.28-4.98)	**<0.001**	0.85 (0.25-2.9)	0.795
IDH_mutation (ref. Wildtype)	0.18 (0.07-0.48)	**<0.001**	0.40 (0.12-1.40)	0.142
FNDC3B (continuous)	1.64 (1.50-1.80)	**<0.001**	1.72 (1.07-2.80)	**0.026**

Bold values means p<0.05.

age, grade, IDH status, and FNDC3B expression were features that were included in the nomogram, and the risk scores were calculated based on that model. The concordance index (C-index) for OS prediction was 0.775, indicating high predictive performance of the model. A calibration curve was implemented to reflect the degree of consistency between the predicted risk and actual occurrence risk, and it can be used to estimate the accuracy of the model in predicting the probability of an individual outcome in the future (27). In our study, the calibration curve showed acceptable agreement between nomogram-predicted and observed 2-, 3- and 5-year OS in the CGGA_325, CGGA_693, and CGGA_301 validation cohorts (Figure 3B).

The patients in the training dataset were divided into high-risk and low-risk groups according to the positive and negative values of the risk score, respectively. The K-M survival curve showed a good discriminating ability of the nomogram (P = 1.4e-4) (Figure 3C). Moreover, the area under the ROC curve for OS was 0.755, indicating a reliable predictive ability in the TCGA LGG dataset (Figure 3D). The CGGA_325, CGGA_693, and CGGA_301 datasets were used to validate the performance of the nomogram. A risk score for each patient was generated by the same method. Consistently, the patients in the high-risk group had a notably poorer prognosis in all the three validation datasets ($p < 0.0001$) (Figure 3C). The area under the curve (AUC) for CGGA_325, CGGA_693, and CGGA_301 OS was 0.8, 0.816, and 0.807, respectively (Figure 3D). In conclusion, these results indicated that the nomogram had adequate performance in predicting the OS of glioma patients.

DEGs identification and weighted co-expression network construction based on FNDC3B expression

To further elucidate the role of FNDC3B expression in the glioma microenvironment, the median expression value was used to create a categorical variable for the TCGA-LGG cohort. The DEGs between the two groups were detected using Wilcoxon test. After setting FDR < 0.05 and fold change ≥ 2 in either direction, a total of 2,099 DEGs, including 1,631 upregulated and 468 downregulated genes were screened out in FNDC3B highly expressed group compared to the low expressed group (Figure 4A). In order to obtain more reliable results, 670 DEGs were filtered after setting the mean expression value to 1 for the raw DEGs (Figure 4B). GSEA was conducted to assess the potential functions of these DEGs. Our results suggested that various immune-related gene signatures were enriched in LGG samples, such as response to cytokine, cytokine secretion, immune system process, and inflammatory response (Figure 4C). According to the KEGG analysis, DEGs were mainly concentrated in PI3K-Akt, p53, and MAPK signaling pathways (Figure 4D). We used TIMER2.0 to explore the relationship between FNDC3B with the top 10 most significantly up or downregulated DEGs in pan-cancer. The corresponding heatmap showed that the correlation was consistent with the gene expression direction of the 20 DEGs only for LGG (Figure 4E).

WGCNA was applied to build a co-expression network based on the 2,099 DEGs. Before constructing the co-expression network, we screened the DEGs by setting expression value larger than 1 in at least 10% of all the samples, and the average level at least 0.5. Finally, 673 DEGs were filtered and selected for subsequent analysis. Power eight was chosen as the appropriate soft threshold because it was the first value to make the degree of independence reach 0.90 and the corresponding average connectivity was close to zero. A total of five gene modules were excavated (genes in the grey module that were not co-expressed; Figure 4F). As shown in Figure 4G, the turquoise module was the most relevant in FNDC3B expression level (R = 0.51, p = 7e-35). Therefore, the 282 hub genes included in the turquoise module were extracted for further analysis. GO enrichment analyses of these genes indicated that the response to interferon-gamma, neutrophil activation and degranulation, regulation of immune effector process, antigen processing and presentation, T cell activation, leukocyte migration, lymphocyte proliferation, mononuclear cell proliferation, interleukin-8 production and acute inflammatory response was related to FNDC3B-mediated immune events. The top 30 GO items, as ranked by their P-values, are shown in Supplementary Figure 5. A total of 210 nodes and 1,061 edges were mapped for the turquoise module genes in the PPI network (Supplementary Figure 6). Using CytoHubba in

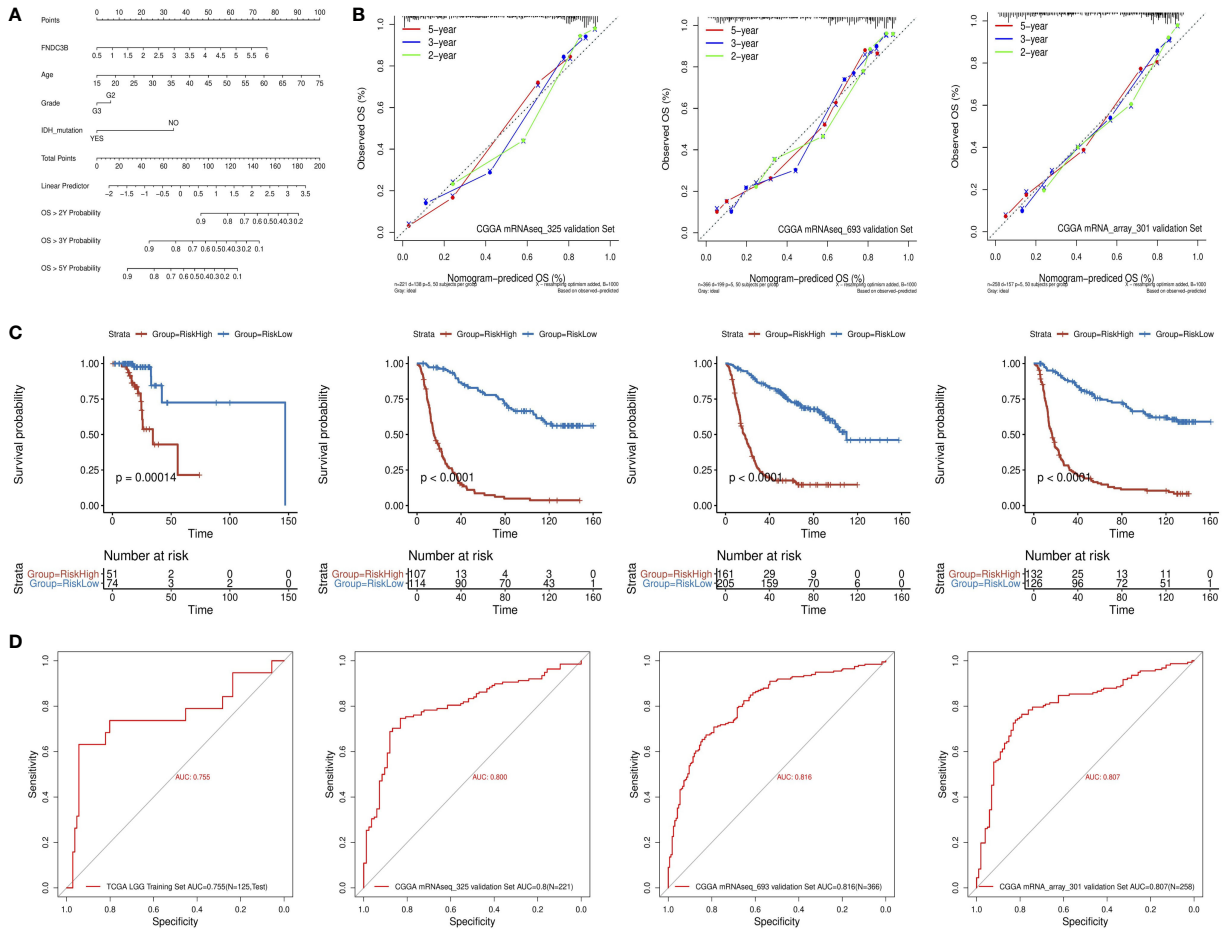


FIGURE 3
Construction and validation of the prognostic nomogram. **(A)** Prognostic nomogram for TCGA low-grade glioma (LGG) dataset. According to four variables (FNDC3B expression level, age, grade, and IDH status) in the model, four corresponding 'points' values can be obtained, and the 'total points' can be calculated by summing them. Therefore, the 2-/3-/5-year overall survival (OS) rate of patients can be predicted. **(B)** The calibration curves for predicting 2-/3-/5-year OS in the CGGA_325, CGGA_693, and CGGA_301 validation datasets, respectively. The nomogram-predicted probability of survival and actual survival are plotted on the x- and y-axes, respectively. The diagonal line represents a perfect prediction. **(C)** K-M curve of high-risk (red) and low-risk (blue) for TCGA low-grade glioma (LGG) training, CGGA_325, CGGA_693, and CGGA_301 validation datasets. **(D)** ROC curves for the risk score in the TCGA LGG training, CGGA_325, CGGA_693, and CGGA_301 validation datasets. TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas.

Cytoscape plug-in, we selected the top 10 genes ranked by the MCC method as hub genes, including TLR2, TLR7, PTPRC (CD45), CCR1, CCL5, TLR1, FN1, VCAM1 (CD106), CXCL10, and TLR6. They were immune-related genes and positively correlated with FNDC3B ($R > 0.3$ and $P < 0.0001$; Figure 4H).

A gene-gene interaction network for the 10 hub genes was built, and their functions were analyzed through the GeneMANIA database (Figure 4I). We found that toll-like receptor signaling pathways, ERK1 and ERK2 cascade and NIK/NF-kappaB signaling, were enriched in LGG. Then, OncoLnc online tool was applied to investigate the function of these hub genes on OS of LGG. The results showed that all hub genes were independent risk factors for evaluating OS (Table 3), and the K-M curves based on GEPIA2 displayed that higher expression of these genes predicted shorter OS in LGG (Supplementary Figure 7). Furthermore, the gene expression analysis of the LGG samples from TCGA showed that the combined expression of the 10 hub genes had a significant effect on overall survival (Figure 4J).

Relationship between FNDC3B expression and tumor immune infiltrates

Since previous studies have reported that TILs are independent predictors in cancers (28, 29), we used TISIDB database to infer the correlations between the expression of FNDC3B and the abundance of 27 types of TILs across TCGA pan-cancers. As shown in

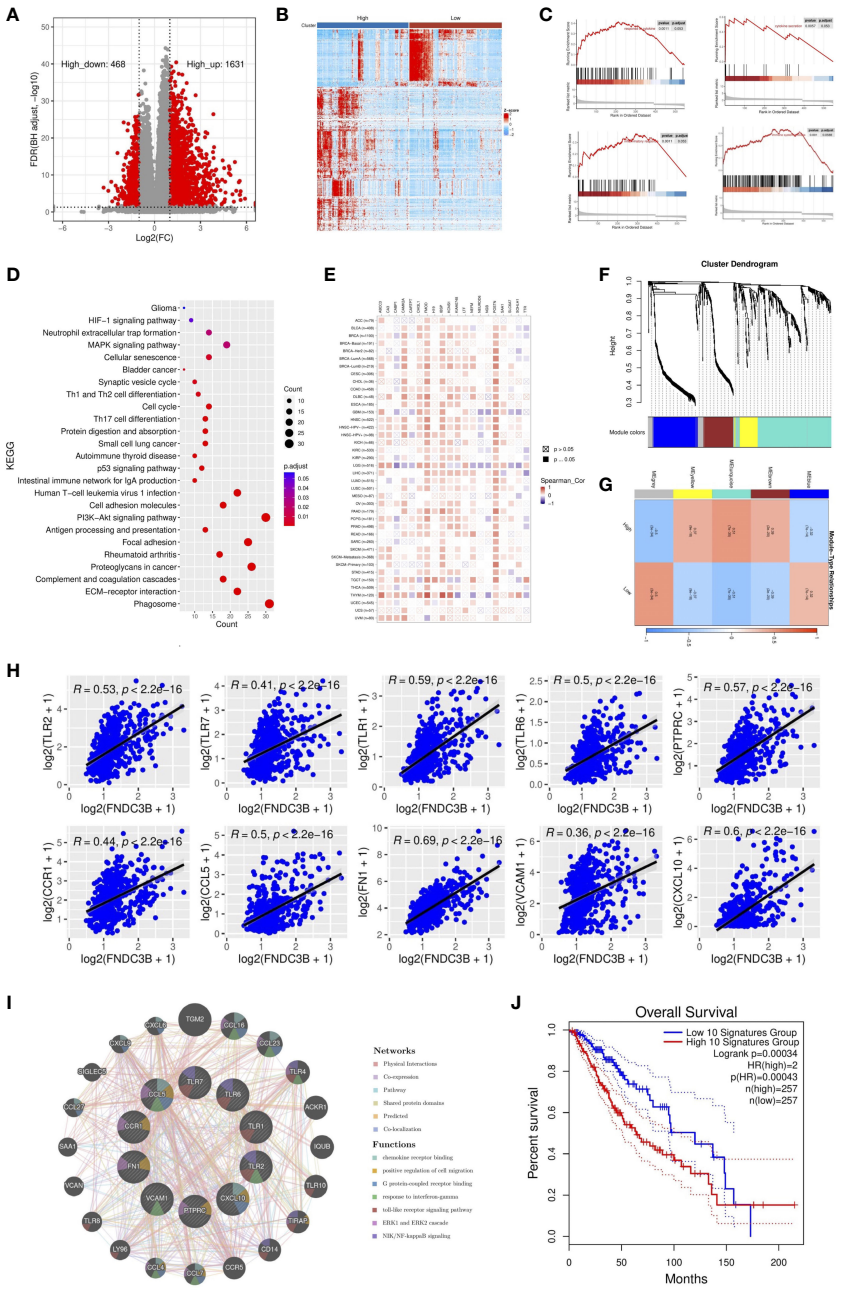


FIGURE 4
Identification and enrichment analysis of differentially expressed genes (DEGs) in TCGA low-grade glioma (LGG). **(A)** Volcano plot of all DEGs. **(B)** Heatmap of the 547 filtered DEGs. **(C)** Gene set enrichment analysis (GSEA) showed that FNDC3B is involved in the tumor immune microenvironment. **(D)** Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses of DEGs. **(E)** The correlation heatmap between FNDC3B with 20 DEGs in pan-cancer. TCGA: The Cancer Genome Atlas. **(F)** Co-expression network constructed with weighted gene co-expression network analysis (WGCNA), hierarchical clustering tree for DEGs based on a dissimilarity measure (1-TOM), genes with similar expression patterns were merged into the same module. **(G)** Correlation between modules and FNDC3B expression. The upper number in each grid represents the correlation coefficient of each module, and the lower number is the corresponding P-value. **(H)** Relationship of the expression level between the top 10 hub genes and FNDC3B. **(I)** GeneMANIA database analysis shows the interaction network among hub genes. Each node represents a gene. The node size displays the strength of interactions. The line color indicates the types of interactions and the node color represents the possible functions of each gene. **(J)** The survival times of TCGA low-grade glioma (LGG) patients with the highest and lowest expression of the top 10 hub genes were compared.

10.3389/fimmu.2022.1027154

TABLE 3 Cox regression results of 10 hub genes in low-grade glioma (LGG).

Gene	Cox coefficient	P-value	FDR
TLR2	0.400	7.30E-06	6.01E-05
TLR7	0.334	2.80E-04	1.18E-03
PTPRC	0.439	2.20E-06	2.30E-05
CCR1	0.306	5.90E-04	2.24E-03
CCL5	0.190	4.90E-02	9.17E-02
TLR1	0.444	2.90E-06	2.88E-05
FN1	0.309	1.20E-03	4.06E-03
VCAM1	0.553	2.40E-09	1.29E-07
CXCL10	0.373	6.40E-05	3.51E-04
TLR6	0.433	3.90E-06	3.65E-05

Figure 5A, FNDC3B expression was positively correlated with TILs in several human cancer types, especially in LGG. Moreover, we investigated the associations between FNDC3B expression and immune subtypes across human cancers, and the landscape of correlations between FNDC3B expression and immune subtypes in different types of cancer (Figure 5B). Among all cancer types, LGG showed the most significant results *via* Kruskal-Wallis test ($p = 1.26e-23$). In TISIDB, we further analyzed FNDC3B expression in different immune subtypes of LGG. We found FNDC3B was mainly expressed in three types, including C3 (inflammatory type), C4 (lymphocyte depleted type), and C5 (immunologically quiet type). FNDC3B expression was the highest in the C3 (inflammatory) type and the lowest in the C5 (immunologically quiet) type (Figure 5C). These results indicated that FNDC3B may play an important role in immune infiltration in glioma. Notably, FNDC3B expression was correlated with the abundance of central memory CD8 T cells ($r = 0.497$, $p < 2.2e-16$), effector memory CD8 T cells ($r = 0.412$, $p < 2.2e-16$), central memory CD4 T cells ($r = 0.468$, $p < 2.2e-16$), regulatory T cells ($r = 0.521$, $p < 2.2e-16$), natural killer (NK) cells ($r = 0.532$, $p < 2.2e-16$), natural killer T (NKT) cells ($r = 0.64$, $p < 2.2e-16$), memory B cells ($r = 0.64$, $p < 2.2e-16$), and macrophages ($r = 0.349$, $p < 2.2e-16$) in LGG (Figure 5D). The positive correlations between FNDC3B expression and TILs were also observed in GBM.

We then assessed the relationships between FNDC3B and eight genes previously reported to be targets of immune checkpoint inhibitors, including CD274 (PD-L1), PDCD1 (PD-1), CD152 (CTLA-4), CD276 (B7-H3), HAVCR2 (TIM-3), CD223 (LAG-3), TNFRSF4 (OX40), and VTCN1. There were significant positive correlations between FNDC3B with B7-H3 ($R = 0.69$, $p < 2.2e-16$), PD-L1 ($R = 0.58$, $p < 2.2e-16$), TIM-3 ($R = 0.43$, $p < 2.2e-16$), PD-1 ($R = 0.42$, $p < 2.2e-16$), CTLA-4 ($R = 0.34$, $p = 1.3e-15$), and OX40 ($R = 0.31$, $p = 1.3e-12$) (Figure 5E). Thus, the six genes were significantly upregulated in the FNDC3B high group compared with the low group (Figure 5F). In summary, these results suggested that FNDC3B was correlated with clinically relevant immune checkpoint molecules in glioma.

Subsequently, to investigate whether FNDC3B expression was correlated with immune infiltration patterns in LGG, we compared the degree of immune cell infiltration between high and low expression groups using the ESTIMATE algorithm. The immune, stromal, and ESTIMATE scores were higher in the high-expression group than in the low-expression group (Figure 5G). Furthermore, we explored the proportions of 22 types of immune cells for LGG using CIBERSORTx to acquire a deeper understanding of the relationship between FNDC3B expression and tumor immune infiltrates. Among the 529 TCGA-LGG samples, 265 samples were in the high expression group and 264 samples in the low expression group. Figure 5H shows the differences in the proportions of the 22 subpopulations of immune cells in these two groups. Naive B cells, plasma cells, naive CD4 T cells, resting memory CD4 T cells, activated memory CD4 T cells, follicular helper T cells (Tfh), M1 and M2 macrophages, activated dendritic cells, and neutrophils were the main immune cells affected by FNDC3B expression. Among them, there were more proportions of resting memory CD4 T cells ($p < 0.001$), activated memory CD4 T cells ($p < 0.01$), M1 macrophages ($p < 0.01$), M2 macrophages ($p < 0.01$), dendritic cells activated ($p < 0.01$), and neutrophils ($p < 0.0001$) in the high expression group. In contrast, the proportions of naive B cells ($p < 0.01$), plasma cells ($p < 0.0001$), naive CD4 T cells ($p < 0.05$), and follicular helper T cells ($p < 0.05$) were lower in the high expression group compared with the low expression group. The correlation matrix heatmap of 22 immune infiltration cells in LGG samples was shown in Figure 5I. Moreover, we analyzed the association between FNDC3B expression and gene markers of various TILs, including B cell, plasma cells, T cell, CD4+ T cell, Tfh, M1 and M2 macrophages, dendritic cells, and neutrophils (Table 4). Overall, FNDC3B expression was strongly positively correlated with gene markers of B cells, T cells, M1 and M2 macrophages, dendritic cells, and neutrophils for TCGA-LGG and TCGA-GBM.

Discussion

Glioma is the most common primary intracranial neoplasm, accounting for approximately 80% of malignant brain tumors.

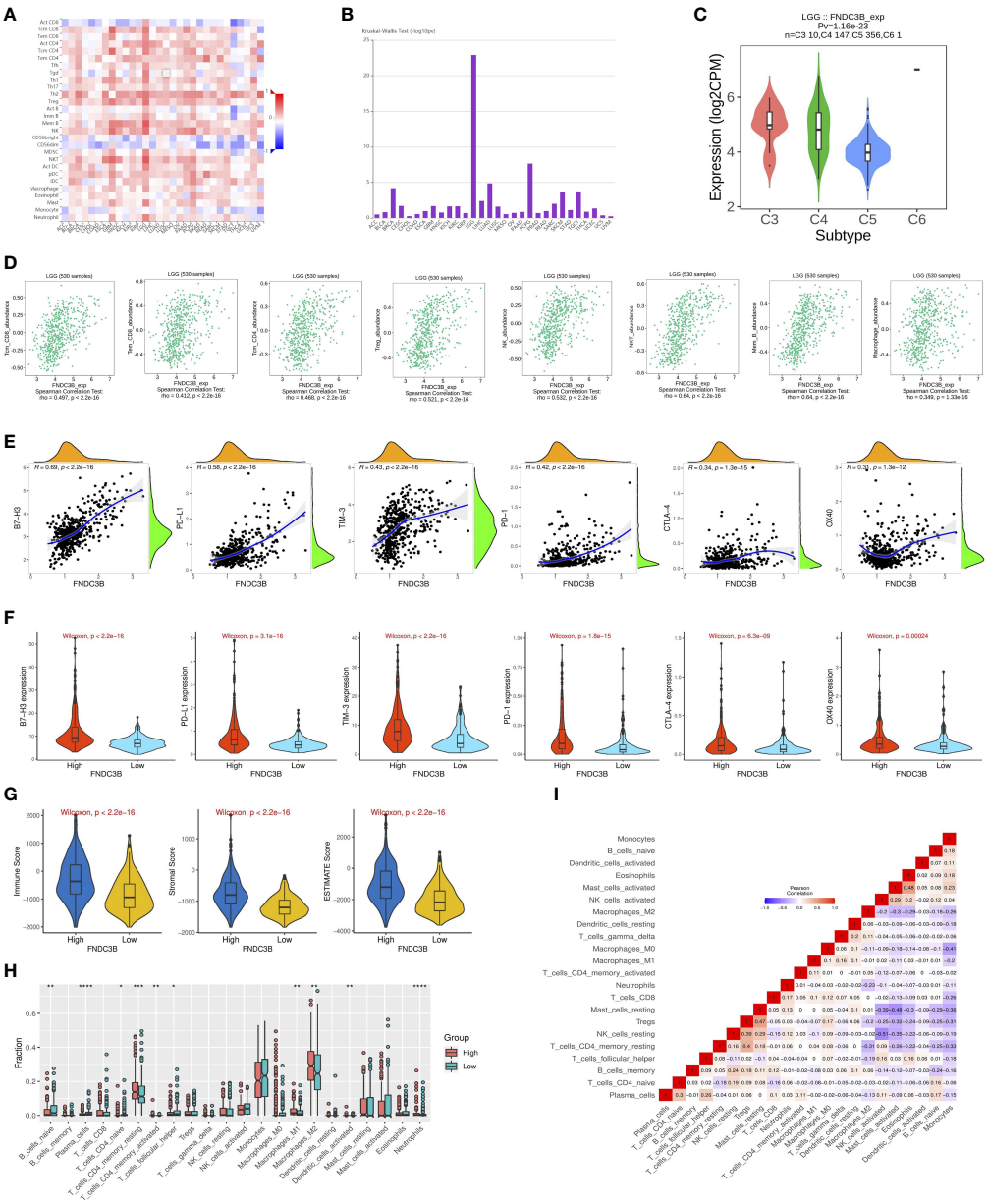


FIGURE 5
Correlation of FNDC3B expression with immune infiltration level in pan-cancer and TCGA low-grade glioma (LGG). **(A)** The landscape of correlation between FNDC3B expression and tumor-infiltrating lymphocytes (TILs) in pan-cancer (red is positive correlated and blue is negatively correlated). **(B)** Associations between FNDC3B expression and immune subtypes across human cancers. **(C)** Correlation of FNDC3B expression and immune subtypes in low-grade glioma (LGG). C3: inflammatory; C4: lymphocyte depleted; C5: immunologically quiet; C6: TGF-b dominant. **(D)** FNDC3B expression was positively closely related with infiltrating levels of central memory CD8 T cells, effector memory CD8 T cells, central memory CD4 T cells, regulatory T cells, natural killer cells, natural killer T cells, memory B cells, and M1 and M2 macrophages in LGG and glioblastoma multiforme (GBM). **(E)** The correlation between FNDC3B and immune checkpoint molecules (B7-H3, PD-L1, TIM-3, PD-1, CTLA-4, and OX40). **(F)** Different expression levels of the six immune checkpoint genes in the high and low FNDC3B expression groups in TCGA LGG samples. **(G)** Comparison of immune, stromal, and ESTIMATE scores between the FNDCB high- and low-expression groups. **(H)** Different proportions of 22 subtypes of immune cells in the FNDCB high- and low-expression groups in TCGA LGG dataset by CIBERSORTx. The proportions of naive B cells ($P < 0.01$), plasma cells ($P < 0.0001$), naive CD4 T cells ($P < 0.05$), resting memory CD4 T cells ($P < 0.001$), activated memory CD4 T cells ($P < 0.01$), follicular helper T cells ($P < 0.05$), macrophages M1 ($P < 0.01$), macrophages M2 ($P < 0.01$), dendritic cells activated ($P < 0.01$), neutrophils ($P < 0.0001$). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. **(I)** Correlation matrix heatmap of 22 immune infiltration cells in LGG samples.

10.3389/fimmu.2022.1027154

TABLE 4 Correlation analysis between FNDC3B expression and gene markers of immune cells in Gene Expression Profiling Interactive Analysis 2 (GEPIA2).

Immune cell types	Gene markers	LGG		GBM	
		Cor	P	Cor	P
B cells	CD2	0.42	***	0.3	***
	CD74	0.48	***	0.31	***
	CD27	0.24	***	0.3	***
Plasma cells	SPAG4	0.3	***	0.29	***
	PDK1	0.37	***	0.2	**
	MAST1	-0.31	***	−0.27	***
	MANEA	0.64	***	0.2	*
T cells	CD2	0.42	***	0.3	***
	CD3E	0.38	***	0.32	***
	CD3D	0.37	***	0.21	**
CD4+ T cells	CD4	0.47	***	0.42	***
Tfh	BCL6	0.16	***	0.28	***
	CD84	0.47	***	0.32	***
	IL6R	0.46	***	0.4	***
	IL21	0.22	***	0.19	*
M1 Macrophage	CD80	0.36	***	0.28	***
	IRF5	0.41	***	0.26	***
	IL6	0.19	***	0.42	***
	CD64	0.24	***	0.21	**
M2 Macrophage	CD163	0.4	***	0.43	***
	CD206	0.12	**	0.57	***
	VSIG4	0.34	***	0.3	***
	MS4A4A	0.35	***	0.36	***
Dendritic cell	HLA-DPB1	0.47	***	0.3	***
	HLA-DQB1	0.36	***	0.27	***
	HLA-DRA	0.52	***	0.24	**
	HLA-DPA1	0.47	***	0.27	***
	CD8A	0.43	***	0.25	**
	CD141	0.35	***	0.6	***
	NRP1	0.63	***	0.81	***
Neutrophils	CCR1	0.45	***	0.41	***
	CD11B	0.46	***	0.56	***
	CCR7	0.25	***	0.3	***
	SLC1A5	0.42	***	0.43	***
	CXCR2	0.38	***	0.22	**

*P < 0.05; **P < 0.01; ***P < 0.001.

The limitations of classical treatments lead to a poor OS (30). New progress in brain tumor research suggests that immunotherapy is a powerful tool for the treatment of gliomas (31–33). Therefore, the identification of novel effective biomarkers for early diagnosis and promising immune-related therapeutic targets for glioma patients has become imperative in clinical practice. Recently, several studies have reported that FNDC3B is an oncogene in various cancers, including glioma (18). To our knowledge, the expression pattern and biological function of FNDC3B in glioma have not been studied in detail, and its possible prognostic value in glioma remains to be explored. Based on the integrated machine learning methods, this is the first report to comprehensively analyze FNDC3B expression profiles and its correlation with immune infiltrates in gliomas.

In this study, we identified that FNDC3B was highly expressed in glioma tissues by mining multiple databases, and the expression levels of FNDC3B increased with the level of the malignant degree, which was also confirmed in another study (16). These results suggested that FNDC3B could serve as a promising molecular marker for predicting the degree of

malignancy in brain glioma. K-M plots indicated that patients with high FNDC3B expression had worse OS and DFS than those with low expression in gliomas. Furthermore, univariate and multivariate Cox analysis demonstrated a positive correlation between FNDC3B expression and poor prognosis of patients with glioma. To further apply FNDC3B in clinical treatments, a prognostic nomogram for personalized prediction was constructed, integrating the FNDC3B expression level with significant clinical parameters (age, grade, and IDH status). The C-index, AUC value, and calibration curve for TCGA training and three CGGA validation datasets showed that our nomogram was reliable and performed adequately. In general, we identified and validated the expression level of FNDC3B as a useful and independent prognostic biomarker for glioma. In the present study, we found that the expression of FNDC3B was negatively correlated with its DNA methylation and CNV was the most common type of alteration for FNDC3B gene. Compared with single-omics, multi-omics approaches can provide more deep insights on molecular changes for cancer subtyping (34). A recent study identified subtype-specific signatures *via* a computational framework for analyzing multi-omics profiles and patient survival and confirmed that subtype-specific signatures could be more feasible in the clinical practice (35). By combining multidimensional genomic measurements, a higher resolution of prognostic signatures will be available for different glioma subtypes in the future. Moreover, methylation (5mC) in cell-free DNA (cfDNA) have been widely observed in human diseases, regions with consistently altered 5mC levels for FNDC3B in circulating cfDNA during progression from low-grade glioma to glioblastoma could be used as markers for development of minimally invasive screening of early diagnosis and surveillance (36, 37).

To further investigate the functions of FNDC3B in glioma, we performed GSEA analysis using DEGs based on TCGA LGG data. The results showed that multiple immune-related pathways were enriched in the FNDC3B high-expression group, such as cellular response to cytokine stimulus, inflammatory response, and immune system process. In the KEGG analysis, PI3K-Akt, p53 and MAPK signaling pathways participated in tumor development. We found that the expression of FNDC3B correlated with that of multiple T cell markers (Th1, Th2, and Th17) in LGG. This suggested that FNDC3B may be involved in the regulation of T cell response in glioma. In addition, WGCNA was performed on the DEGs to find more valuable clues. Finally, 10 upregulated genes were identified as hub genes, including TLR2, TLR7, PTPRC, CCR1, CCL5, TLR1, FN1, VCAM1, CXCL10, and TLR6. Previous studies reported that T cells have the ability to directly recognize danger signals through the expression of toll-like receptors (TLRs) (38); interactions between CCR1 and CCL5 contribute to T-cell activation (39), and CXCL10 is usually considered to be a pro-inflammatory chemokine that enhances recruitment of CD8+ and Th1-type CD4+ effector T cells to infected or inflamed nonlymphoid tissues (40). FN1 can promote integrin β1 ubiquitination and degradation and its expression may be upregulated by the hyperactivation of ERK1/2 (41), considered to be critical mediators for T cell functions (42). VCAM1 induces T-cell antigen receptor-dependent activation of CD4+ T lymphocytes (43), and PTPRC is a well-known positive regulator of T-cell receptor signaling (44). Overall, these findings highlight the ability of FNDC3B to potentially regulate T cell responses in LGG.

TILs are independent predictors in cancers (29). Our findings showed that FNDC3B was strongly positively correlated with immune infiltration in LGG and GBM among all cancer types in the database, especially in the cytotoxic T cells and anti-tumor associated immune cells, such as central memory CD8 T cell, effector memory CD8 T cell, central memory CD4 T cell, regulatory T cell (Treg), natural killer cell (NK), natural killer T cell (NKT), memory B cell, and macrophage. The comparative analysis of FNDC3B gene expression in different immune subtypes in LGG suggested that FNDC3B may be strongly linked to immunological properties in the tumor microenvironment. A previous study reported that the immune microenvironment affected the gene expression of tumor tissues, and the degree of stromal and immune cell infiltration influenced prognosis (45). In our study, the FNDC3B high-expression group displayed higher values for immune, stromal, and ESTIMATE scores than the FNDC3B low-expression group, indicating that the high expression level of FNDC3B is positively related to immune infiltration in gliomas. Moreover, consistent with the TISIDB results, we found that memory CD4 T cell, macrophages M1 and M2, and neutrophils were enriched in the FNDC3B high group based on CIBERSORT analysis. Additionally, a relatively strong correlation between FNDC3B expression and gene markers of T cells, CD4+ T cell, follicular helper T cells, and dendritic cells indicated the potential role of FNDC3B in regulating T cell function in LGG and GBM.

The FNDC3B expression was positively correlated with genes of immune checkpoints, suggesting that FNDC3B could be a regulatory factor of various immune checkpoints in glioma. The correlation analysis showed that FNDC3B was mostly positively correlated with B7-H3, which was associated with a suppressive effect on T-cell activities in various tumors (46). In recent years, studies have shown that B7-H3 is a promising novel target for glioma immunotherapy (47, 48). Nehama and colleagues reported that B7-H3 is highly expressed in more than 70% of GBM samples and that B7-H3-redirected chimeric antigen receptor T (CAR-T) cells can effectively control tumor growth (31). Currently, tumor immunotherapy has attracted great attention. Hence, our findings indicated that whether we can participate in the immune checkpoint by inhibiting the expression of FNDC3B, and can FNDC3B be served as a powerful immune checkpoint blockade combination therapy to increase efficacy and reduce side effects? More investigations are required to get a full description and understanding of the

mechanisms in the future, this study provide a new insight for further exploration of the molecular mechanisms.

There are several limitations in the current research. First, it was mainly based on online public databases and computational methods. Nevertheless, integrated machine learning algorithms strengthen the conclusion of this study. Second, more investigations are needed to identify the expression and function of FNDC3B as well as their correlations with immune cell infiltration; thus, further clinical and experimental studies in the laboratory are required for verifying its role in glioma.

In conclusion, our comprehensive analysis revealed that FNDC3B was upregulated in glioma, while increased FNDC3B expression predicted an unfavorable prognosis. Moreover, FNDC3B is associated with the infiltration of various immune cells, and it may play a vital role in the tumor immune microenvironment of glioma. Therefore, we reported that FNDC3B is a possible prognostic biomarker and an immune-related therapeutic target for glioma, which will be useful for clinical applications.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of the Shanghai Jiao Tong University Affiliated Sixth People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Conception and design: XW and HZ. Acquisition and analysis of the data: XW, YH, SL, and HZ. Draft and revision of the manuscript: XW and HZ. All authors have read and approved the submitted version.

Funding

The current study was supported by the National NaturalScience Foundation of China (No. 81800708), the ShanghaiSixth People's Hospital Grant (No. ynhg202204 andynqn202105), the China Postdoctoral Science Foundation Grant (No. 2019M660088 and 2020T13408).

Acknowledgments

We greatly appreciate for the analytic data provided by the TCGA and CGGA databases. The draft was preprinted at https://www.researchsquare.com/article/rs-1102200/v1 2021.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1027154/full#supplementary-material

References

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol* (2016) 131(6):803–20. doi: 10.1007/s00401-016-1545-1

2. Jhaveri J, Liu Y, Chowdhary M, Buchwald ZS, Gillespie TW, Olson JJ, et al. Is less more? comparing chemotherapy alone with chemotherapy and radiation for high-risk grade 2 glioma: An analysis of the national cancer data base. *Cancer* (2018) 124(6):1169–78. doi: 10.1002/cncr.31158

3. Zheng H, Luo L, Zhao W. Factors associated with level III lymph nodes positive and survival analysis of its dissection in patients with breast cancer. *Laparoscopic Endoscopic Robotic Surg* (2020) 3(2):43–7. doi: 10.1016/j.lers.2020.03.001

4. Das S, Camphausen K, Shankavaram U. Cancer-specific immune prognostic signature in solid tumors and its relation to immune checkpoint therapies. *Cancers (Basel)* (2020) 12(9):2476. doi: 10.3390/cancers12092476

5. Zhang Z, Chen L, Chen H, Zhao J, Li K, Sun J, et al. Pan-cancer landscape of T-cell exhaustion heterogeneity within the tumor microenvironment revealed a

progressive roadmap of hierarchical dysfunction associated with prognosis and therapeutic efficacy. *EBioMedicine* (2022) 83:104207. doi: 10.1016/j.ebiom.2022.104207

6. Lei J, Zhou MH, Zhang FC, Wu K, Liu SW, Niu HQ, et al. Interferon regulatory factor transcript levels correlate with clinical outcomes in human glioma. *Aging (Albany NY)* (2021) 13(8):12086–98. doi: 10.18632/aging.202915

7. Bao S, Zhao H, Yuan J, Fan D, Zhang Z, Su J, et al. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform* (2020) 21(5):1742–55. doi: 10.1093/bib/bbz118

8. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L, Sun J, et al. An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol Neurobiol* (2018) 55(5):3684–97. doi: 10.1007/s12035-017-0572-9

9. Kishimoto K, Kato A, Osada S, Nishizuka M, Imagawa M. Fad104, a positive regulator of adipogenesis, negatively regulates osteoblast differentiation. *Biochem Biophys Res Commun* (2010) 397(2):187–91. doi: 10.1016/j.bbrc.2010.05.077

10. Fucci C, Resnati M, Riva E, Perini T, Ruggieri E, Orfanelli U, et al. The interaction of the tumor suppressor FAM46C with p62 and FNDC3 proteins integrates protein and secretory homeostasis. *Cell Rep* (2020) 32(12):108162. doi: 10.1016/j.celrep.2020.108162

11. Nishizuka M, Kishimoto K, Kato A, Ikawa M, Okabe M, Sato R, et al. Disruption of the novel gene fad104 causes rapid postnatal death and attenuation of cell proliferation, adhesion, spreading and migration. *Exp Cell Res* (2009) 315 (5):809–19. doi: 10.1016/j.yexcr.2008.12.013

12. Chen CF, Hsu EC, Lin KT, Tu PH, Chang HW, Lin CH, et al. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. *Hepatology* (2010) 52(5):1690–701. doi: 10.1002/hep.23847

13. Cheng CK, Wang AZ, Wong THY, Wan TSK, Cheung JS, Raghupathy R, et al. FNDC3B is another novel partner fused to RARA in the t(3;17)(q26;q21) variant of acute promyelocytic leukemia. *Blood* (2017) 129(19):2705–9. doi: 10.1182/blood-2017-02-767707

14. Li Y, Yang J, Wang H, Qiao W, Guo Y, Zhang S, et al. FNDC3B, targeted by miR-125a-5p and miR-217, promotes the proliferation and invasion of colorectal cancer cells via PI3K/mTOR signaling. *Onco Targets Ther* (2020) 13:3501–10. doi: 10.2147/OTT.S226520

15. Han B, Wang H, Zhang J, Tian J. FNDC3B is associated with ER stress and poor prognosis in cervical cancer. *Oncol Lett* (2020) 19(1):406–14. doi: 10.3892/ol.2019.11098

16. Wang GH, Wang LY, Zhang C, Zhang P, Wang CH, Cheng S, et al. MiR-1225-5p acts as tumor suppressor in glioblastoma via targeting FNDC3B. *Open Med (Wars)* (2020) 15(1):872–81. doi: 10.1515/med-2020-0156

17. Xu H, Hu Y, Qiu W. Potential mechanisms of microRNA-129-5p in inhibiting cell processes including viability, proliferation, migration and invasiveness of glioblastoma cells U87 through targeting FNDC3B. *BioMed Pharmacother* (2017) 87:405–11. doi: 10.1016/j.biopha.2016.12.100

18. Wang Z, Tang W, Yuan J, Qiang B, Han W, Peng X, et al. Integrated analysis of RNA-binding proteins in glioma. *Cancers (Basel)* (2020) 12(4):892. doi: 10.3390/cancers12040892

19. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* (2014) 124(3):453–62. doi: 10.1182/blood-2014-04-567933

20. Sun J, Zhang Z, Bao S, Yan C, Hou P, Wu N, et al. Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J Immunother Cancer* (2020) 8(1):e000110. doi: 10.1136/jitc-2019-000110

21. Bao S, Hu T, Liu J, Su J, Sun J, Ming Y, et al. Genomic instability-derived plasma extracellular vesicle-microRNA signature as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer. *J Nanobiotechnol* (2021) 19 (1):22. doi: 10.1186/s12951-020-00767-3

22. Zhang H, Ye J, Weng X, Liu F, He L, Zhou D, et al. Comparative transcriptome analysis reveals that the extracellular matrix receptor interaction contributes to the venous metastases of hepatocellular carcinoma. *Cancer Genet* (2015) 208(10):482–91. doi: 10.1016/j.cancergen.2015.06.002

23. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* (2004) 6(1):1–6. doi: 10.1016/S1476-5586(04)80047-2

24. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45(W1):W98–W102. doi: 10.1093/nar/gkx247

25. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* (2008) 26(8):1364–70. doi: 10.1200/JCO.2007.12.9791

26. Zhang H, Weng X, Ye J, He L, Zhou D, Liu Y. Promoter hypermethylation of TERT is associated with hepatocellular carcinoma in the han Chinese population. *Clin Res Hepatol Gastroenterol* (2015) 39(5):600–9. doi: 10.1016/j.clinre.2015.01.002

27. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128

28. Ohtani H. Focus on TILs: prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer. *Cancer Immun* (2007) 7:4. doi: 10.1158/1424-9634.DCL-4.7.1

29. Azimi F, Scolyer RA, Rumcheva P, Moncrieff M, Murali R, McCarthy SW, et al. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J Clin Oncol* (2012) 30(21):2678–83. doi: 10.1200/JCO.2011.37.8539

30. Goodenberger ML, Jenkins RB. Genetics of adult glioma. *Cancer Genet* (2012) 205(12):613–21. doi: 10.1016/j.cancergen.2012.10.009

31. Nehama D, Di Ianni N, Musio S, Du H, Patane M, Pollo B, et al. B7-H3-redirected chimeric antigen receptor T cells target glioblastoma and neurospheres. *EBioMedicine* (2019) 47:33–43. doi: 10.1016/j.ebiom.2019.08.030

32. Zhang H, Bi Y, Wei Y, Liu J, Kuerban K, Ye L. Blocking wnt/beta-catenin signal amplifies anti-PD-1 therapeutic efficacy by inhibiting tumor growth, migration, and promoting immune infiltration in glioblastomas. *Mol Cancer Ther* (2021) 20(7):1305–15. doi: 10.1158/1535-7163.MCT-20-0825

33. Anghileri E, Di Ianni N, Paterra R, Langella T, Zhao J, Eoli M, et al. High tumor mutational burden and T-cell activation are associated with long-term response to anti-PD1 therapy in lynch syndrome recurrent glioblastoma patient. *Cancer Immunol Immunother* (2021) 70(3):831–42. doi: 10.1007/s00262-020-02769-4

34. Sun J, Yan C, Xu D, Zhang Z, Li K, Li X. Immuno-genomic characterisation of high-grade serous ovarian cancer reveals immune evasion mechanisms and identifies an immunological subtype with a favourable prognosis and improved therapeutic efficacy. *Br J Cancer* (2022) 126(11):1570–80. doi: 10.1038/s41416-021-01692-4

35. Yu F, Quan F, Xu J, Zhang Y, Xie Y, Zhang J, et al. Breast cancer prognosis signature: linking risk stratification to disease subtypes. *Brief Bioinform* (2019) 20 (6):2130–40. doi: 10.1093/bib/bby073

36. Yu F, Li K, Li S, Liu J, Zhang Y, Zhou M, et al. CFEA: a cell-free epigenome atlas in human diseases. *Nucleic Acids Res* (2020) 48(D1):D40–4. doi: 10.1093/nar/gkz715

37. Hlady RA, Zhao X, Pan X, Yang JD, Ahmed F, Antwi SO, et al. Genome-wide discovery and validation of diagnostic DNA methylation-based biomarkers for hepatocellular cancer detection in circulating cell free DNA. *Theranostics* (2019) 9(24):7239–50. doi: 10.7150/thno.35573

38. Salerno F, Freen-van Heeren JJ, Guislain A, Nicolet BP, Wolkers MC. Costimulation through TLR2 drives polyfunctional CD8(+) T cell responses. *J Immunol* (2019) 202(3):714–23. doi: 10.4049/jimmunol.1801026

39. Choi SW, Hildebrandt GC, Olkiewicz KM, Hanauer DA, Chaudhary MN, Silva IA, et al. CCR1/CCL5 (RANTES) receptor-ligand interactions modulate allogeneic T-cell responses and graft-versus-host disease following stem-cell transplantation. *Blood* (2007) 110(9):3447–55. doi: 10.1182/blood-2007-05-087403

40. Peperzak V, Veraar EA, Xiao Y, Babala N, Thiadens K, Brugmans M, et al. CD8+ T cells produce the chemokine CXCL10 in response to CD27/CD70 costimulation to promote generation of the CD8+ effector T cell pool. *J Immunol* (2013) 191(6):3025–36. doi: 10.4049/jimmunol.1202222

41. Wang Y, Li C, Shi L, Chen X, Cui C, Huang J, et al. Integrin beta1D deficiency-mediated RyR2 dysfunction contributes to catecholamine-sensitive ventricular tachycardia in arrhythmogenic right ventricular cardiomyopathy. *Circulation* (2020) 141(18):1477–93. doi: 10.1161/CIRCULATIONAHA.119.043504

42. Crawford TQ, Hecht FM, Pilcher CD, Ndhlovu LC, Barbour JD. Activation associated ERK1/2 signaling impairments in CD8+ T cells co-localize with blunted polyclonal and HIV-1 specific effector functions in early untreated HIV-1 infection. *PloS One* (2013) 8(10):e77412. doi: 10.1371/journal.pone.0077412

43. Damle NK, Aruffo A. Vascular cell adhesion molecule 1 induces T-cell antigen receptor-dependent activation of CD4+T lymphocytes. *Proc Natl Acad Sci U.S.A.* (1991) 88(15):6403–7. doi: 10.1073/pnas.88.15.6403

44. Porcu M, Kleppe M, Gianfelici V, Geerdens E, De Keersmaecker K, Tartaglia M, et al. Mutation of the receptor tyrosine phosphatase PTPRC (CD45) in T-cell acute lymphoblastic leukemia. *Blood* (2012) 119(19):4476–9. doi: 10.1182/blood-2011-09-379958

45. Winslow S, Lindquist KE, Edsjo A, Larsson C. The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. *BMC Cancer* (2016) 16(1):841. doi: 10.1186/s12885-016-2864-2

46. Kontos F, Michelakos T, Kurokawa T, Sadagopan A, Schwab JH, Ferrone CR, et al. B7-H3: An attractive target for antibody-based immunotherapy. *Clin Cancer Res* (2021) 27(5):1227–35. doi: 10.1158/1078-0432.CCR-20-2584

47. Zhang C, Zhang Z, Li F, Shen Z, Qiao Y, Li L, et al. Large-Scale analysis reveals the specific clinical and immune features of B7-H3 in glioma. *Oncoimmunology* (2018) 7(11):e1461304. doi: 10.1080/2162402X.2018.1461304

48. Tang X, Zhao S, Zhang Y, Wang Y, Zhang Z, Yang M, et al. B7-H3 as a novel CAR-T therapeutic target for glioblastoma. *Mol Ther Oncolytics* (2019) 14:279–87. doi: 10.1016/j.omto.2019.07.002

Frontiers in Immunology

Explores novel approaches and diagnoses to treat immune disorders.

The official journal of the International Union of Immunological Societies (IUIS) and the most cited in its field, leading the way for research across basic, translational and clinical immunology.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact